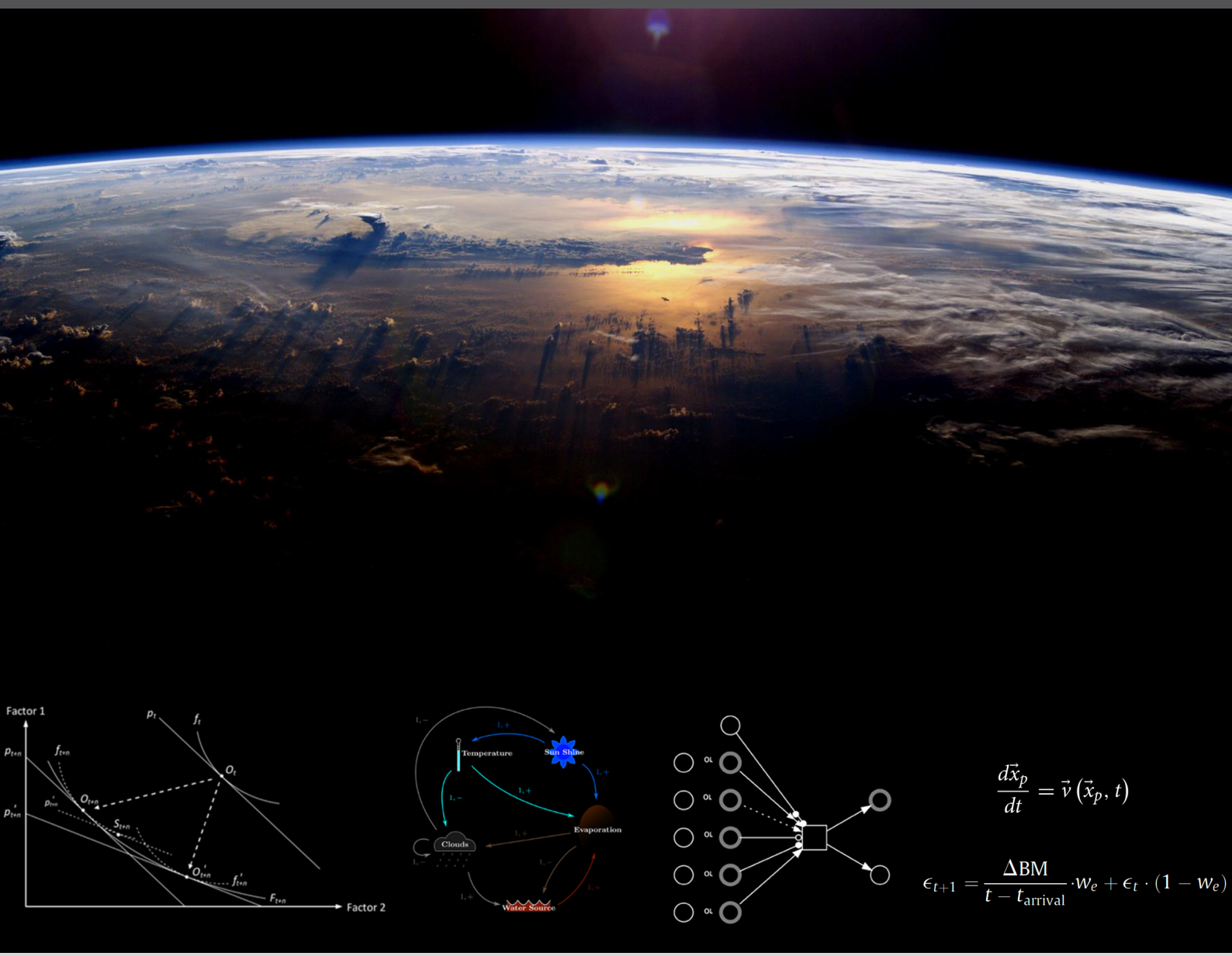
HYBRID SOLUTIONS FOR THE MODELLING OF COMPLEX ENVIRONMENTAL SYSTEMS

EDITED BY: Christian E. Vincenot, Stefano Mazzoleni and Lael Parrott
PUBLISHED IN: Frontiers in Environmental Science



$$\frac{d\vec{x}_p}{dt} = \vec{v}\left(\vec{x}_p, t\right)$$

$$\epsilon_{t+1} = \frac{\Delta \mathrm{BM}}{t - t_{\mathrm{arrival}}} \cdot w_e + \epsilon_t \cdot (1 - w_e)$$

frontiers Research Topics

frontiers

ISSN 1664-8714
ISBN 978-2-88945-055-8
DOI 10.3389/978-2-88945-055-8

About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to Quality

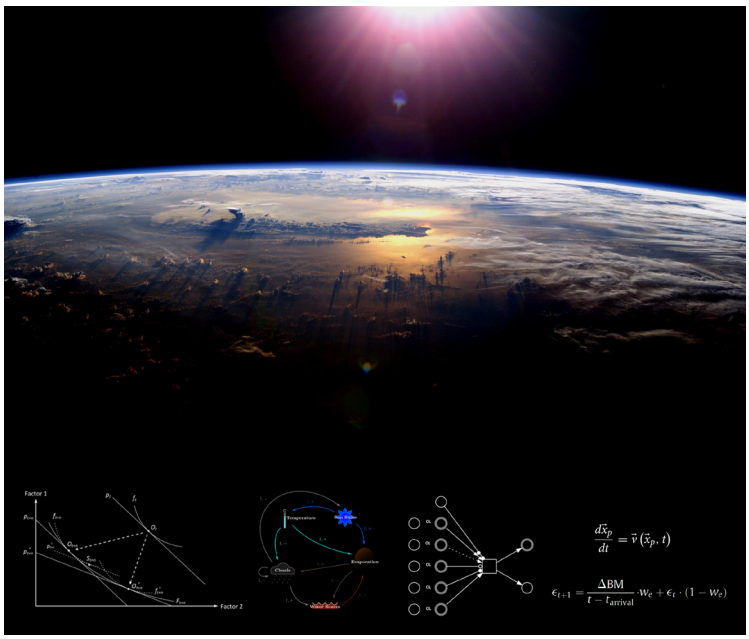
Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.
Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view.
By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: **researchtopics@frontiersin.org**

HYBRID SOLUTIONS FOR THE MODELLING OF COMPLEX ENVIRONMENTAL SYSTEMS

Topic Editors:
Christian E. Vincenot, Kyoto University, Japan
Stefano Mazzoleni, University of Naples, Italy
Lael Parrott, University of British Columbia, Canada



Composite image. This sunset view of the Earth horizon was taken onboard the International Space Station (ISS) and is property of NASA (ref. ISS007-E-10807, available at: https://eol.jsc.nasa.gov/SearchPhotos/photo.pl?mission=ISS007&roll=E&frame=10807).

Systems studied in environmental science, due to their structure and the heterogeneity of the entities composing them, often exhibit complex dynamics that can only be captured by hybrid modeling approaches. While several concurrent definitions of "hybrid modeling" can be found in the literature, it is defined here broadly as the approach consisting in coupling existing modelling paradigms to achieve a more accurate or efficient representation of systems. The need for hybrid models generally arises from the necessity to overcome the limitation of a single modeling technique in terms of structural flexibility, capabilities, or computational efficiency. This book brings together experts in the field of hybrid modelling to demonstrate how this approach can address the challenge of representing the complexity of natural systems. Chapters cover applied examples as well as modeling methodology.

Citation: Vincenot, C. E., Mazzoleni, S., Parrott, L., eds. (2016). Hybrid Solutions for the Modelling of Complex Environmental Systems. Lausanne: Frontiers Media. doi: 10.3389/978-2-88945-055-8

Table of Contents

frontiers in Environmental Science

EDITORIAL
published: 29 August 2016
doi: 10.3389/fenvs.2016.00053

Editorial: Hybrid Solutions for the Modeling of Complex Environmental Systems

Christian E. Vincenot[1]*, *Stefano Mazzoleni*[2] *and Lael Parrott*[3]

[1] Biosphere Informatics Laboratory, Department of Social Informatics, Kyoto University, Kyoto, Japan, [2] Laboratory of Applied Ecology and System Dynamics, Department of Agricultural Sciences, University of Naples Federico II, Portici, Italy, [3] Okanagan Institute for Biodiversity, Resilience, and Ecosystem Services, The University of British Columbia, Kelowna, BC, Canada

Keywords: integration, multiscale, simulation, agent, system dynamics, machine learning, network, coupled

The Editorial on the Research Topic

Hybrid Solutions for the Modeling of Complex Environmental Systems

Hybrid modeling is the combination of simulatory approaches aimed at the accurate mechanistic modeling of complex dynamic systems (Parrott, 2011; Vincenot et al., 2011). It basically consists in the coupling—to a varying degree ranging from simple comparison to bridging to total fusion—of models based on different existing modeling techniques. This concept may therefore, in some regards, seem trivial and more relevant to engineering than scientific research. Yet seeing this field of study through a purely technical eye produces a limited view of the complexity and open challenges relevant to the design and use of hybrid models. This shall be demonstrated throughout this book, in which we engage in showing the potential of hybrid modeling in environmental science in particular. The latter is a strongly interdisciplinary field devoted to the study of heterogeneous, highly dynamical, often non-linear, complex systems made of numerous different entities—rational or not—interacting in a changing environment.

Reasons underlying the use of hybrid models are actually multiple. Hybrid modeling, in the widespread meaning of hybrid system modeling, has been delineated in engineering as a purely technical solution to the challenge of representing concurrently discrete events and continuous processes, which are common in automated systems. This issue has been dominating also in environmental science, in which systems often exhibit the same constraints (see Vincenot et al., 2015 for a discussion of this aspect). However, hybrid modeling can have a range of motivations other than accurate time representation. In the study of ecological systems in particular, it actually stems from a philosophical root in its ability to merge reductionist and holistic views and thereby acknowledge the ontology of Nature (Vincenot et al., 2011). In microbiology, the organization of cells, which are discrete interacting entities of variable size, is dependent on exchanges of molecules, which are diffusing as continuous flows in space. Coupling both aspects requires a hybrid view of the system with both reductionist agents (i.e., cells) and aggregated flows (i.e., intercellular signaling processes) (Hay Mele et al.). In oceanography, Eulerian–Lagrangian approaches that couple aggregated biogeochemical models and physical spatial transport models are used to simulate in three dimensions tracer dynamics in moving fluids and thereby understand spatial patterns of plankton under oceanic currents (Chenillat et al.). On a different topic, Vincenot et al. (2016) similarly showed how the integration inside of a System Dynamics (SD)—Individual-based (IB) hybrid model of realistic plant-level processes

OPEN ACCESS

Edited by:
Marco Casazza, Parthenope University of Naples, Italy

Reviewed by:
Laura Uusitalo, University of Helsinki, Finland

****Correspondence:***
Christian E. Vincenot
christian@vincenot.biz

Specialty section:
This article was submitted to Environmental Informatics, a section of the journal Frontiers in Environmental Science

Received: *29 May 2016*
Accepted: *02 August 2016*
Published: *29 August 2016*

Citation:
Vincenot CE, Mazzoleni S and Parrott L (2016) Editorial: Hybrid Solutions for the Modeling of Complex Environmental Systems. Front. Environ. Sci. 4:53. doi: 10.3389/fenvs.2016.00053

(i.e., continuous metabolism as well as mechanistic seed dispersal events) on top of a basic hydrological submodel provides insights on the demographic response and spatial distribution of vegetation facing climatic stress. Using a comparable conceptual framework, Bradhurst et al. demonstrate that the simultaneous use of individual/agent-based and equation-based formulation of the pathogen transmission process at different levels leads to increased accuracy in epidemiological modeling.

Frequently, hybrid modeling arises naturally from the need to couple several existing models, each describing one part of a system in the most suitable manner with respect to modeling goals. Cross-scale ecological simulations are particularly prone to such settings. Girard et al. report on such a hybrid model featuring interactions between regional-scale climate, landscape-scale hydrology, and individual animals, each of them calling for the use of a dedicated modeling approach. The integration of economic or social dynamics in environmental models also typically leads to the same situation. For example, Drogoul et al. show here how they could improve the accuracy of governmental land-use change forecasts of the Mekong Delta in Vietnam by resorting to a hybrid model including social, ecological, and economic dimensions.

Hybrid modeling is often concurrently undertaken to reconciliate bottom-up and top-down views of a system. Ghersi summarizes the long-term work done at the Centre International de Recherche sur l'Environnement et le Développement (CIRED) on the coupling of bottom-up energy models with the IMACLIM-S macroeconomic model. Martin and Schlüter discuss a procedure to tackle the same methodological issue, which occurred in this case when linking the dynamics emerging from individual-level human decisions (micro-scale) and ecosystem-wide processes (macro-scale) inside of a socio-ecological model supporting lake restoration.

Hybrid modeling is not solely a technical means to build large models through a combinatory approach. For the modeling of even a single component, the differences in characteristics or features exhibited by different paradigms give rise to hybrid concepts unlocking beneficial tradeoffs at runtime, between precision and performance for instance, by dynamically changing the representation of model components. This was termed *dynamic model swapping* (Vincenot et al., 2011) and discussed more specifically in the case of epidemiological modeling (Vincenot and Moriya, 2011) before being implemented by Gray and Wotherspoon (2012). The same authors detail here the unique methodological questions surfacing in such adaptive hybrid models, and explain how a submodel selection strategy can be effectively performed by an intelligent monitor agent (Gray and Wotherspoon).

While hybrid models are not uncommon, this emerging discipline stands in need of further research establishing appropriate design and analysis methodologies. At the same time, there is still a clear necessity to explore new modeling avenues for environmental science. Advancements in the field of engineering especially can provide new simulative approaches and opportunities for model coupling to support a better rendering of complex environmental systems. Among others, Generalized Hybrid Petri Nets (GHPN) offer a framework to help formalize and visualize processes. The value of this approach is illustrated here in a study of Chagas disease transmission (Herajy and Heiner). In a different manner, machine learning, which is being increasingly applied across scientific disciplines, can be integrated in deterministic models to include data-driven components (Goldstein and Coco). Bayesian approaches can also serve to infer model structure and support the parameterization of individual-based models through data integration (Parise et al.). Agents can thereby be enhanced with human decision-making capabilities to study socio-ecological issues (Pope and Gimblett; an applied study on the effects of riparian corridors in the Sonoran desert).

The non-linear, self-organized dynamics of complex systems, including environmental ones, makes their behavioral trends extremely difficult to predict. Hybrid models, which typically represent the dynamics of a system over a wide parameter space, can be used to explore the envelope of possible, probable and plausible system trajectories. This is exemplified in the work presented by Tareen et al., who show how the simplification of the relationships between interacting components in terms of feedbacks in the frame of a parametric linear hybrid automaton allows for the isolation and study of particular system trajectories.

Through developments such as the ones presented in this book, hybrid modeling is expected to ultimately shift the fundamental question of environmental modeling from the usual "With this technique, how can this system be represented?" to "Which combination of paradigms would best render this set of processes?" Doing so will bring the discipline closer to understanding and successfully reproducing the complex dynamics of environmental systems through models.

AUTHOR CONTRIBUTIONS

All authors listed, have made substantial, direct and intellectual contribution to the work, and approved it for publication.

REFERENCES

Gray, R., and Wotherspoon, S. (2012). Increasing model efficiency by dynamically changing model representations. *Environ. Model. Softw.* 30, 115–122. doi: 10.1016/j.envsoft.2011.08.012

Parrott, L. (2011). Hybrid modelling of complex ecological systems for decision support: recent successes and future perspectives. *Ecol. Inform.* 6, 44–49. doi: 10.1016/j.ecoinf.2010.07.001

Vincenot, C. E., Carteni, F., Mazzoleni, S., Rietkerk, M., and Giannino, F. (2016). Spatial self-organization of vegetation subject to climatic stress—insights from

a system dynamics—individual-based hybrid model. *Front. Plant Sci.* 7:636. doi: 10.3389/fpls.2016.00636

Vincenot, C. E., Giannino, F., Rietkerk, M., Moriya, K., and Mazzoleni, S. (2011). Theoretical considerations on the combined use of System Dynamics and individual-based modeling in ecology. *Ecol. Modell.* 222, 210–218. doi: 10.1016/j.ecolmodel.2010.09.029

Vincenot, C. E., Mazzoleni, S., Moriya, K., Cartenì, F., and Giannino, F. (2015). How spatial resource distribution and memory impact foraging success: a hybrid model and mechanistic index. *Ecol. Complex.* 22, 139–151. doi: 10.1016/j.ecocom.2015.03.004

Vincenot, C. E., and Moriya, K. (2011). Impact of the topology of metapopulations on the resurgence of epidemics rendered by a new multiscale hybrid modeling approach. *Ecol. Inform.* 6, 177–186. doi: 10.1016/j.ecoinf.2011.04.002

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Environmental Science

ORIGINAL RESEARCH
published: 19 November 2015
doi: 10.3389/fenvs.2015.00073

Cell-Based Models in Plant Developmental Biology: Insights into Hybrid Approaches

Bruno Hay Mele[1], *Francesco Giannino*[1], *Christian E. Vincenot*[2], *Stefano Mazzoleni*[1]* and *Fabrizio Cartení*[1]

[1] Laboratory of Applied Ecology and System Dynamics, Department of Agricultural Sciences, University of Naples Federico II, Napoli, Italy, [2] Department of Social Informatics, Kyoto University, Kyoto, Japan

Computer models are nowadays part of the biologist's toolbox for studying biological dynamics and processes. Tissue development and functioning results from extremely complicated dynamics, that usual analysis does not come very far in terms of understanding the processes underlying those dynamics. In this context, mathematical and numerical models can help to disentangle complex interactions and to analyze non-intuitive dynamics that drives tissue development and functioning. Since these are multi-scale processes, both in time and space, there is the need to develop an appropriate modeling approach. The most promising one is hybrid modeling, that is a synthesis of the differential equation based reaction-diffusion approach at molecular and chemical continuous scales, and the Individual-Based modeling approach for simulating the mechanical and behavioral interactions of the cell ensemble constituting the tissue. Such an approach has been often used in developmental biology, both for plants and animals. In this paper, a brief history of hybrid modeling approaches and tools will be reviewed, and a simple example of its application to a current problem in plant developmental biology (the appearance of vascular patterning during plant growth) will be illustrated, showing the intuitiveness and the strength of such an approach.

Keywords: system dynamics, individual-based modeling, differential equations, mathematical models, numerical simulation, pattern formation, procambium differentiation, primary vascular structure

OPEN ACCESS

Edited by:
Alexander Kokhanovsky, EUMETSAT, Germany

Reviewed by:
Roeland M. H. Merks, Centrum Wiskunde & Informatica, Netherlands
Guennady Ougolnitsky, Southern Federal University, Russia
Gerardo Toraldo, University of Naples Federico II, Italy

***Correspondence:**
Stefano Mazzoleni
mazzolen@unina.it

Specialty section:
This article was submitted to Environmental Informatics, a section of the journal Frontiers in Environmental Science

Received: *08 May 2015*
Accepted: *02 November 2015*
Published: *19 November 2015*

Citation:
Hay Mele B, Giannino F, Vincenot CE, Mazzoleni S and Cartení F (2015) Cell-Based Models in Plant Developmental Biology: Insights into Hybrid Approaches. Front. Environ. Sci. 3:73. doi: 10.3389/fenvs.2015.00073

1. INTRODUCTION

During the first decade of the twenty-first century, biology has been profoundly transformed. The technological advances of this period contributed to produce a tremendous amount of biological data, which in turn has made computers fundamental tools in biology research.

Nowadays, researchers have means to investigate cell biophysical, biological and kinetic properties, providing large and very detailed amount of information to the scientific community. On the other hand, knowledge on how cell processes combine and give rise to tissue and organ properties is still meager, mainly because genetic analysis is time consuming when the number of interacting factors is large. As time passes, and data gathers, it is becoming extremely complicated to identify the networks underlying the regulation of cell activity, while considering all the parallel interactions that underlie cellular morphogenesis (Merks and Glazier, 2005).

Traditional research organizes results and develops new hypotheses about the behavior of gene networks using static schemes, an approach that is unfruitful in all but the simplest cases

(Merks and Glazier, 2005). This suggests that reconstructing the dynamics of the genetic regulatory networks with wet research it is not sufficient. Furthermore, gene networks seem just one part of the story: insights on the role of mechanical and physical interactions are needed if one wants to truly elucidate emergent properties linked to tissues (Dupuy et al., 2008). During the development of multi-cellular organisms, cells are capable of interacting with each other through a range of biological and physical mechanisms. A description of these networks of interactions is essential for understanding how it is possible for tissues and organs to co-ordinate cellular activity (Dupuy et al., 2008).

Working at tissue-level many cellular dynamics have to be considered, such as cell growth, cell elongation and cell division: all these are still subjected to intense study and discussions scientific community. This is particularly felt with stem cells, whose ability to differentiate is intrinsically linked to specific biological functions in multi-cellular organisms. A widespread idea is to "interpret stem cells as non-hierarchical self-organizing dynamical systems" (De Matteis et al., 2013) and "stemness" not as an explicit cellular property: stem cells would be "dynamically selected and modified in response to cell-cell and cell-environment interactions on the basis of their potential and flexibility, rather than being specialized a priori" (De Matteis et al., 2013).

The difficulty of working with living tissues, together with the aforementioned multi-scale complexity, is a major limitation to describe such systems, and computer modeling appears particularly helpful to characterize the behavior of multi-cellular systems: working with theoretical models and numerical simulations has proven to be effective in disentangling the relationship between cellular processes and tissue-level arrangement (Jönsson and Krupinski, 2010).

2. MODELING IN PLANT DEVELOPMENTAL BIOLOGY

The first pivotal moment in the history of mathematical modeling in developmental biology was the publication of D'Arcy Thompson "On growth and form" (Thompson, 1942), who spawned a geometrically-oriented approach to the problem. In plant biology, this path has been followed and extended by Lindenmayer with its L-system (Lindenmayer, 1975), a modeling framework for representing 1D linear and branching structures (e.g., cells, leaves, or shoots) in form of a sequence of elements. The "geometrical" approach has subsequently been adapted for describing 2D structures as a graph rotation system, and coded in vv-system (Smith, 2006) (for more information see Prusinkiewicz and Runions, 2012).

A second strong contribution to the formalization of developmental processes in biology came from the works of Hofmeister (1863), Sachs (1877), and Errera (1886): their mechanics theories on cells are the foundation of the modern approach to cell division, even if their use bring some discrepancy between models and observations (for a solution to this, see Besson and Dumais, 2011).

Another decisive contribution on theoretic developmental biology has been made by Turing, with his paper on biological pattern formation (Turing, 1952), whose concepts have been extensively used for showing that molecular-level interactions may lead to morphogenesis and differentiation. Turing used a system of partial differential equations (PDE) in his work, a mathematical tool widely used in biological modeling, for describing substance diffusion and reaction. This approach evolved to this day in plant science as a chemical/molecular-oriented one, mainly thanks the works of Meinhardt (1982).

Since the diffusion of these ideas, the contributions of mathematical modeling to developmental biology have grown substantially and contributed to many interesting results (for an in-depth review see Prusinkiewicz and Runions, 2012), but the issues with modeling multi-cellular systems have remained.

2.1. History of Cellular Based Models in Biology

During the last decade, more and more developmental plant biologists turned to mathematical models to explore their hypotheses, and specialized packages appeared for facilitating model construction. Following the realization that developmental processes depend on both geometrical and molecular dynamics, a synthesis of these two approaches has been performed, and some hybrid tools appeared on the scene. This hybrid approach has a deep conceptual meaning, since joining the geometrical and the molecular point of view means working simultaneously at different temporal and spatial scales (i.e., it is a multi-scale approach). A key point in this kind of approach is to establish a certain degree of simplification of cell processes, since using complex molecular models of the cell for simulating tissue-related dynamics (Krul et al., 2003) has been found very computationally demanding and too complex for gathering insights. It is possible to assimilate cells to well-mixed compartments and single cell functions can be represented in a very "simplified" way, still capturing the complexity of their spatial interactions. Considering a tissue as an ensemble of such "simplified" cells facilitates the search for the emergence of tissue organization due to the collective behaviors of the single cells; these behaviors are dependant on both cell internal processes and interactions with neighboring cells. Since chemical substances are signals regulating internal cell dynamics, reaction-diffusion processes, are key players in tissue modeling.

This approach is called cell-based modeling. As stated by Palm and Merks (2015): "The inputs to a cell-based model are the behavioral rules that cells follow. The output of a cell-based model is the tissue morphogenesis that follows indirectly from the collective behavior of the individual cells"

Spatially explicit, cell-based paradigms can be broadly classified according to the cells being part of a grid (in-lattice models) or not (off-lattice models).

A widespread in-lattice paradigm in animal developmental biology modeling is the Cellular Potts Model (CPM; Glazier and Graner, 1993), a modeling approach that considers single cells as agents trying to minimize their internal energy while growing, dividing, and interacting with chemical fields.

The CPM represent cells as a groups of neighboring pixels, distinguishing between boundary and non-boundary pixels in order to define interaction sites, and uses an energy-based approach for simulating growth, cell-cell interaction, and for maintaining cell shapes. Molecular details, i.e., substance production and diffusion, are handled by ODE and PDE (Ordinary and Partial Differential Equations) solvers coupled with geometric informations and with the growth process, that also permit to consider other continuous (eventually spatially-explicit) processes. The CPM approach has been successfully applied to a vast array of biological problems: here we will cite tissue patterning (Savill and Sherratt, 2003; Zeng et al., 2004), morphogenesis (Zajac et al., 2003), tumorigenesis (Turner and Sherratt, 2002), and vasculogenesis (Scianna and Preziosi, 2012).

While some models of plant systems have been produced within the CPM paradigm, it is considered unfit for simulating plant tissue dynamics. These are strongly influenced by the presence of a cell-wall, which is responsible for maintaining cell geometry, for preventing cell motility, and it is involved in substance (e.g., auxin) transport. The CPM lacks a proper way to simulate cell wall dynamics, and the paradigm chosen for cell geometry makes cell shape and motility (absent in almost all plant cells, due to the cell wall) difficult to control.

Merks et al. (2011) proposes an appropriate solution, starting from the CPM but using for geometry an off-grid modeling approach that describes cells as polygons delimited by cell walls, considered as separate entities and shared among adjacent cells. Walls are assimilated to mechanical springs, so that growth and mechanical interactions could be computed by means of a Markovian relaxation algorithm. Additional differential equations model diffusive transport across the two cell membranes and across the cell wall separating adjacent cells. As in the original CPM, sets of ordinary differential equations in each cell describe dynamics of biochemical networks and genetic regulatory networks.

Since the hybrid approach has proven itself fruitful, other variations have been developed, changing the way cell geometry and/or topology are defined, e.g., using L-system (Wabnik et al., 2013) or considering cells as polygonal (Dupuy et al., 2008) or Voronoi meshes (Mebatsion et al., 2006), as well as using different ways to describe cell walls and to render mechanical interactions, such as assimilating the structure to a spring (Shapiro et al., 2013) or a viscous fluid (Dupuy et al., 2008), and use appropriate physical models for simulating their behavior.

Since the diffusion of mathematical models as plant developmental biology tools, a few general purpose tools have been made, in order to allow non-programmers to enter the field.

The CPM has been integrated in the CompuCell3D software environment (Izaguirre et al., 2004—http://www.compucell3d.org), a python-based, extensible general-purpose framework used in biology for simulating in 2D and 3D a range of biological dynamics. Compucell3D has been used mainly for studies in human tumorigenesis (Boghaert et al., 2014; Swat et al., 2015) and animal tissue development (Dias et al., 2014), albeit other processes have been simulated (Popławski et al., 2008; Zhang et al., 2011; Giverso and Preziosi, 2014). It is not particularly suited for plant tissues, mainly because it is based on the CPM, although a model of root growth has been developed with this method (Grieneisen et al., 2007).

The plant-specific approach of Merks, based on the CPM, has been implemented in the VirtualLeaf modeling framework (Merks et al., 2011—https://biomodel.project.cwi.nl/software) for 2D tissue simulation. VirtualLeaf is written in C++, and comes with pre-programmed modules simulating many accepted results. It is possible to write new modules in order to test new hypotheses.

CellModeller (Dupuy et al., 2008—https://haselofflab.github.io/CellModeller/) is another stand-alone environment build with a modular approach and written in python, for modeling large-scale multi-cellular systems in 2D. As in other tools cell topologies are graph rotation system (i.e., composed of nodes and edges), and inter- and intra-cellular chemical interactions are formalized as PDEs and ODEs. Mechanical interactions and growth are simulated through physical laws regulating rearrangement of the cell nodes following strain/stress of the cell wall, itself rendered as a viscous fluid. This tool has been used for modeling microbial biofilms (Rudge et al., 2012), but his potential for modeling plant cells related dynamics has been often stated (Dupuy et al., 2008; Liu and Stewart, 2015).

Some other tools have been programmed as extension for pre-existing software environment, like CellZilla (Shapiro et al., 2013—http://www.cellzilla.info/) that extends the xlr8r (Shapiro et al., 2003) Mathematica package capabilities of solving chemical reactions as sets of ODE to 2D tissues. The topology of the cell is described as a set of nodes (vertexes), with the segments joining them (edges) being the cell wall. Using xlr8r, inter- and intra-cellular chemical interactions are formulated as differential equations; growth and mechanical interactions are modeled considering the walls subject to springs whose behavior is regulated by Hooke's law. CellZilla has been used for simulating auxin-driven development in the apical meristem of *Arabidopsis thaliana* L. (Nikolaev et al., 2013).

Finally, VPlants (https://team.inria.fr/virtualplants/) is a set of packages belonging the OpenAlea software environment that permit to analyze, model and simulate tissue dynamics (for detail on OpenAlea see Pradal et al., 2008). Its "Tissue" package permits to simulate tissues growth and division, starting from a single cell, or after reconstructing a digital tissue from a photograph. The cell is geometrically described as a polygon, and the user may choose different algorithms for growth; division is similarity coded on the basis of the user's needs. VPlants has been used for simulating, among other things, vascular development in *A. thaliana* (Muraro et al., 2014).

3. EXAMPLE OF HYBRID MODEL APPROACH

Vascular plants are characterized by a pervasive, specialized system of vessels, composed of two types of tissues: xylem, cells that transport water and mineral nutrient, and phloem cells that carry sugars and other organic molecules. The spatial position and differentiation of such vascular cells is established during early phases of tissue development. In the plant embryo,

four cells buried inside the cell mass are activated as an undifferentiated tissue named procambium: they will grow and divide following specific procedures, increasing the number of cells and forming a cylinder of procambial tissue. As the embryo matures, two zones will start to act as "stock" of undifferentiated cells: one on top, called shoot apical meristem, and the other, called root apical meristem, on the lower part of the embryo. Adjacent to the two apical meristems, some procambial cells will further divide, generating precursors of vascular cells (Jouannet et al., 2015). Peculiar procambium arrangement patterns will emerge during this phase, that will be propagated by the two meristems as the plant grows in height.

These arrangement patterns vary among species (Beck et al., 1982) but also between the stem and the root of the same individual.

How these arrangements arise, how they are maintained, and what causes their diversity are open questions that have been puzzling plant scientists and modelers for a long time (Jeffrey, 1903; Sieburth and Deyholos, 2006; Muraro et al., 2014). From an experimental standpoint, studies in plant vascular morphogenesis have followed the reductionist route, as the key players shifted from structures (Esau, 1960) to cells, and from those to genes (Caño-Delgado et al., 2010) and proteins (Sieburth and Deyholos, 2006), with many regulatory elements discovered recently (Donner et al., 2009), most of them in the species *A. thaliana*. At the moment there is only an approximate knowledge of the underlying dynamics (Jouannet et al., 2015), derived mainly from studies on plant hormones like auxins (Scarpella et al., 2006), brassinosteroids (Vert and Chory, 2006), and cytokinines (Mähönen et al., 2006); most of these focus on single aspects, and there is the need for a clarification of the interaction of these aspects both spatially and temporally.

3.1. Existing Models and Limitations

Whilst mathematical models have proven to be a good tool in helping to understand the connections between cells and tissues (Jönsson and Krupinski, 2010), they do so focusing mainly on sub-cellular processes related to plant hormones, especially auxins (Prusinkiewicz and Runions, 2012). The first steps were laid by Sachs (1969), with the so-called 'canalization hypothesis', used by various molecular models exploring the vascularization processes (Mitchison, 1980; Feugier et al., 2005; Bayer et al., 2009).

The first to approach plant primary vascular structure specification and arrangement were Muraro et al. (2014), with a multicellular model that explores the hypothesis of a vascular patterning mechanism dependent on hormones interaction in a very specific context (the roots of *A. thaliana* in a steady-state condition). Starting from a set of experimentally determined factors (i.e., gene products and hormones), the authors build a model formalizing the interaction between these factors and propose a minimal regulatory network capable of maintaining a stable vascular pattern in Arabidopsis root without predefined positional information (Muraro et al., 2014). This model used VPlants from the OpenAlea environment to simulate the tissue geometry and auxin diffusion.

The theory of an auxin-dependant patterning in *A. thaliana* has been further explored with the model of De Rybel et al. (2014), which combined experimental evidence and modeling work to investigate the interactions between auxin and cytokinin and their involvement in early embryo development. The model, formulated using VirtualLeaf, shows how these hormones contribute to cell division and differentiation of vascular tissues providing positional cues for the establishment of a stable spatial pattern within a growing domain.

Another model, based on reaction-diffusion dynamics, has been proposed by Cartenì et al. (2014) for simulating the differentiation and the spatial patterning of procambium, phloem and xylem. This theoretical model was formulated as a set of activator–substrate systems (Meinhardt, 1982) describing the dynamics of nine diffusible morphogens whose interactions lead to the differentiation of vascular tissues and the emergence of their spatial patterns. This model, implemented in MATLAB (MathWorks Inc.—https://www.mathworks.com/products/matlab/) successfully recreated a broad range of vascular arrangements, working with a fixed domain and without considering individual cell dynamics.

In order to provide a procedural example of development of a hybrid model, we built a simple model of *A. thaliana* root and stem early vascular differentiation. The hybrid approach permits to take into account both the molecular and geometric perspectives (**Figure 1**), thus providing the chance to capture emergent properties linked to the interaction of these two phenomena The use of VirtualLeaf approximates spatial interactions among cells in a computationally convenient way (Merks and Glazier, 2005), even if it considers stochastic-based dynamics not necessarily linked with cells biological properties (Merks et al., 2011).

3.2. Model Description

As mentioned in the introduction, it is possible to work with tissue-related dynamics using a hybrid modeling approach (Vincenot et al., 2011), i.e., coupling an Individual-Based Model (IBM) with a continuous PDE/ODE mathematical model. Under this paradigm, the IBM will account for internal cell rules and cell-cell interactions, whilst a set of differential equations will regulate substance physics and continuous processes.

Vincenot et al. (2011) proposed a conceptual framework comprising a set of reference cases representing combination patterns of SD and IBM sub-models, which shall serve as building blocks for hybrid models. The model described here fits into case 2b of the framework, since cells are represented as IBM individuals and a SD sub-model is embedded in each cell to compute growth and substance dynamics. Single cells are then networked depending on the adjacency of their walls, forming thereby a tissue.

Starting from Cartenì et al. (2014), we built a hybrid, growing-domain model of cell growth and differentiation in the roots and stems of *A. thaliana*; this model considers simplified tissue dynamics, and the production of four diffusible, cross-reacting substances (S_0, S_1, S_2, and A_P) which are responsible for

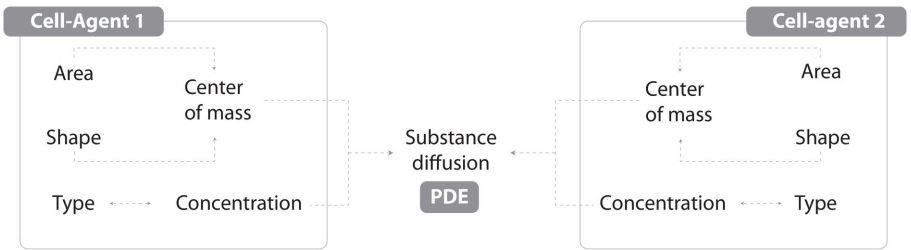


FIGURE 1 | An individual agent is described by its geometric properties (Area, Shape and derived Center of mass) and cell type. Each type is characterized by a specific metabolism (ODE) with products whose concentration is dynamically linked with diffusion processes (PDE).

procambium differentiation and proliferation. In this particular case, the PDE mathematical formulation is a simplified version of the Reaction/Diffusion model presented in Cartenì et al. (2014).

The model is not meant to provide a one-to-one correspondence with specific gene products or plant hormones or insights into the molecular mechanisms of vascular differentiation. It was rather selected for the simplicity of formulation and for the fact that to date it is the only one that provides a general framework (although only theoretical) for the spontaneous emergence of different spatial arrangements of vascular bundles. For more detail on the models underlying assumptions, see Cartenì et al. (2014).

The model domain is a transversal section of growing *A. thaliana* tissue. Each one of the cells composing the tissue is a complex agent made by the cell itself and by a polygonal cell wall, shared with neighbor cells. The cell agent properties are the position, the type (**Table 1**), the area, and the concentration of all substances considered in the model.

The model itself considers four substances:

- S_0, produced by the cell walls contacting the medium/outside. It represents the signal involved in the creation of a radial gradient that determines the establishment of pith, inner, and outer layers of cells.
- S_1 and S_2, produced respectively by the inner or the outer cells. These represent signal substrates regulating the production of the activator A_P.
- A_P, produced by the reaction between S_1 and S_2. It represents the procambium activator that triggers the differentiation of procambial cells.

All substances diffuse between cells but not outwards the tissue, and all four are considered homogeneous in the individual cell space.

In the model, cells grow at a constant rate (controlled by the parameter k_a, see **Table 2**), dividing if their size reach twice the original size. A theoretical substance (S_0), is then produced in the outermost layer of cells, contributing to the creation of an inside/outside gradient; and the other substances are produced according to cell type (**Table 1**). Each chemical species inside the cell diffuse through the tissue and react with each other according to the following equations:

TABLE 1 | Differentiation and substance production rules.

ID	Type	Differentiation condition		Substance produced
		On S_0	ON A_P	
0	Pith	$[S_0] < \overline{S_0}$	Any	None
1	Inner	$\overline{S_0} < [S_0] < \widehat{S_0}$	$[A_P] < \overline{A_P}$	S_1, A_P
2	Inner procambium	$\overline{S_0} < [S_0] < \widehat{S_0}$	$[A_P] > \overline{A_P}$	S_1, A_P
3	Outer	$[S_0] > \widehat{S_0}$	$[A_P] < \overline{A_P}$	S_2, A_P
4	Outer procambium	$[S_0] > \widehat{S_0}$	$[A_P] > \overline{A_P}$	S_2, A_P

$$\frac{\partial S_0}{\partial t} = \sigma_0 - \mu_0 S_0 + D_{S_0}\nabla^2 S_0 \quad (1)$$

$$\frac{\partial S_1}{\partial t} = \overline{\sigma_1}\left(1 - \frac{S_1}{1 + k_S A_P}\right) - \rho_S A_P^2 S_1 S_2 + D_S \nabla^2 S_1 \quad (2)$$

$$\frac{\partial S_2}{\partial t} = \overline{\sigma_2}\left(1 - \frac{S_2}{1 + k_S A_P}\right) - \rho_S A_P^2 S_1 S_2 + D_S \nabla^2 S_2 \quad (3)$$

$$\frac{\partial A_P}{\partial t} = \sigma_{A_P} + \rho_{A_P} A_P^2 S_1 S_2 - \mu_{A_P} A_P + D_{A_P}\nabla^2 A_P \quad (4)$$

with:

$$\overline{\sigma_1} = \begin{cases} \sigma_S & \text{if cell is not type 0 and is (type 1 or type 2)} \\ 0 & \text{else} \end{cases} \quad (5)$$

$$\overline{\sigma_2} = \begin{cases} \sigma_S & \text{if cell is not type 0 and is (type 3 or type 4)} \\ 0 & \text{else} \end{cases} \quad (6)$$

Algorithmically, at the beginning of each simulation time step the concentration of the substances inside each cell is checked and, if any of the threshold values in **Table 1** are met, the cell type attribute is updated; then the cell grows taking into account mechanical constraints due to cell-cell interactions, and the area and shape attributes are updated. After growth the cell size is checked for evaluating if to divide or not. The model then computes the reaction/diffusion module, based on its type (**Table 1**), and all substances first diffuse and then react among each other following the aforementioned system of Partial Differential Equations, taking into account geometry and topology. Then, the model enters the next time step.

In order to further explore the hybrid-model capabilities of VirtualLeaf, another simple assumption has been implemented

TABLE 2 | List of parameters values.

Module	Parameter	Description	Parameter value
IBM	$\overline{S_0}$	Threshold value for pith cell differentiation	0.02
IBM	$\widehat{S_0}$	Threshold value for internal cell differentiation	0.2
IBM	$\overline{A_P}$	Threshold value for procambium cell differentiation	0.5
PDE	σ_0	S_0 production	1
PDE	μ_0	S_0 consumption rate	0.2
PDE	D_{S_0}	S_0 diffusion coefficient	0.1
PDE	σ_S	S basic production rate	0.1
PDE	k_S	S production saturation constant	20
PDE	ρ_S	S cross-reaction coefficient	0.8
PDE	D_S	S diffusion coefficient	0.5
PDE	σ_{A_P}	A_P basic production rate	0.001
PDE	ρ_{A_P}	A_P cross-reaction coefficient	0.3
PDE	μ_{A_P}	A_P removal rate	0.02
PDE	D_{A_P}	A_P diffusion coefficient	0.001

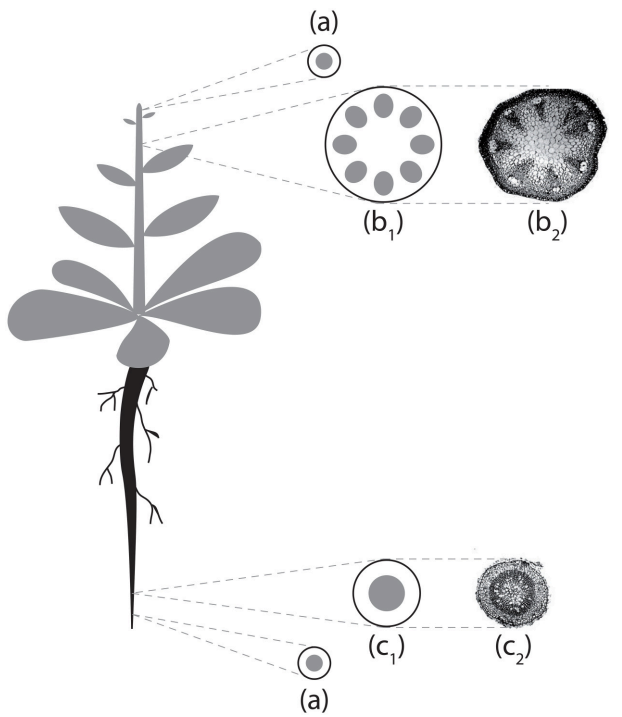


FIGURE 2 | Schematization of *Arabidopsis thaliana* vascular structures in stem and primary root. (a) Stem/root tip; (b_1) stem section scheme; (b_2) stem section micrograph; (c_1) root section scheme; (c_2) root section micrograph. Micrographs reproduced with permission of Avsian-Kretchmer et al. (2002).

for the description of cell growth and division. In this case cells still grow at a fixed rate, but the division threshold is modeled as inversely proportional to the S_0 concentration. Cell size increases with the distance from the outermost layer of cells.

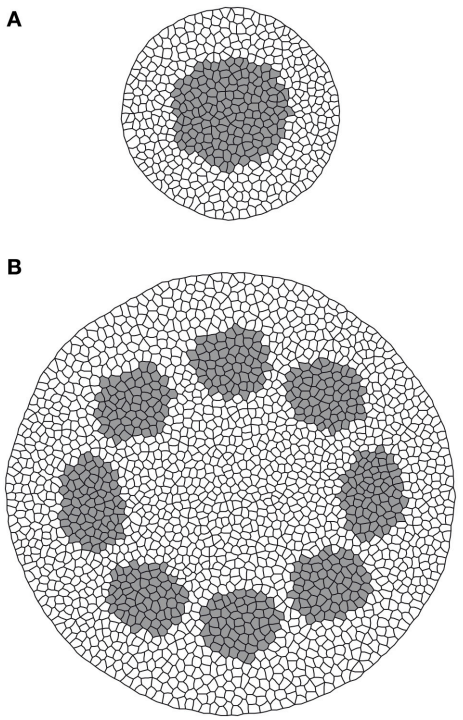


FIGURE 3 | Model output for root (A) and stem (B) sections. Gray is procambium, white is non-procambial tissue.

Parameter definitions and values for both the variants of the model are the same; they are presented in **Table 2**.

3.3. Model Output and Discussion

A set of simulations was performed to test the capabilities of the model to reproduce the provascular patterns observed in *A. thaliana* plants (**Figure 2**).

Figure 3 shows the output of two model simulations run for 1785 (**Figure 3A**) and 11505 (**Figure 3B**) time steps. All parameter values for the two simulations were the same.

From these results it is possible to observe how procambium arrangement is an emergent property dependent both on the reaction/diffusion dynamics that lead to the formation of Turing patterns (Turing, 1952) and on the tissue size, that influence the aforesaid dynamics and contribute to the differences between the stem (bigger domain) and the root (smaller domain) patterns. This model thus confirm, in a growing domain, the results presented by Cartenì et al. (2014). that the shift between a protostelic arrangement (**Figure 3A**) and a eustelic arrangement (**Figure 3B**) could be attributed only to the size of the domain in which the molecular dynamics occur. Moreover, both simulated sections show roughly the same cell number and the same total area of the reference photomicrographs (**Figures 2**c_1,b_2). However, in the real system, cells are heterogeneous in dimensions and the difference between the model and the real system can be reduced if a further assumption for growth and division dynamics is added to the model.

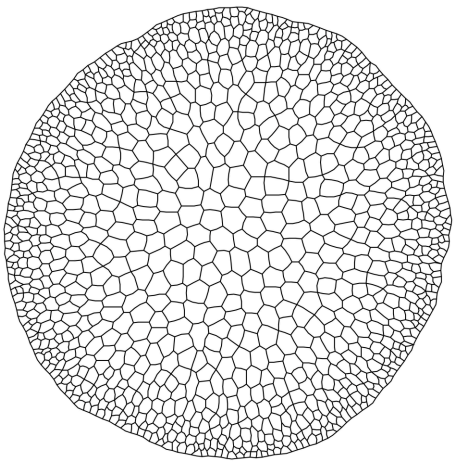


FIGURE 4 | Output from alternative growth model.

The output from the third model simulation (**Figure 4**) shows how a simple rule can affect the system state producing a difference in terms of size among the cell rows, linked to the concentration of S_0 inside the cells. Cells leaving meristematic zones often enlarge to hundreds of times their original size and the final size depends on the position they occupy in the radial arrangement. Cell expansion usually results from the combination of two processes: the increase in cell ploidy by endoreplication (DNA replication with no mitosis), and the complex process of cell expansion, which is driven by internal turgor pressure and restricted by the ability of cell walls to extend (Perrot-Rechenmann, 2010). These processes are under the control of several stimuli that can be spatially heterogeneous. In the presented simulation, this was achieved inserting the inverse of the concentration of a certain substance (S_0) as a term in the algorithm that triggers cell division. The condition may simulate the dynamics of a substance that inhibits cell growth, whose biological counterpart may be a sugar or a plant hormone (e.g., Evans et al., 1994).

The proposed solution effectively reproduced the radial pattern of cell sizes observed in Arabidopsis stems (**Figure 1**). This result is an example of how a simple individual rule makes a complex collective behavior emerge and illustrate the potential of VirtualLeaf for exploring spatial-explicit tissue dynamics.

The spatial position of plant vascular cells is established during early phases of development, when some cells are activated as procambium, and it is maintained along the plant life-span. Procambium arrangement patterns vary among different species but also between the stem and the root of the same individual: how they arise, how they are maintained, and what causes their diversity are open questions. Modeling the dynamics of the system in a manner appropriate to answer these questions is considered a complex challenge, since the system analyzed is multi-scale and characterized by a vast number of processes. IBMs and PDEs are unsuitable for this task, mainly because taken separately, these approaches are not able to efficiently describe multi-scale systems: the former cannot account for continuous dynamics and the latter cannot work with discrete elements.

In our case, a purely IBM approach would not efficiently model the diffusion of substances among cells (a continuous process) while a PDE model would fail to represent the properties of the single cells, mainly because it considers the tissue as a whole characterized by average properties (for the meaning of whole in this context, see Vincenot et al., 2011). A hybrid modeling approach, instead, may be suitable because it permits to analyze multi-scale systems like tissues without oversimplifying them. Beside preventing conceptual approximations, this approach may became necessary in order to inspect properties emerging from the interaction of processes that cannot be modeled with the same paradigm, as those linking individual properties with diffusion dynamics. In order to test this hypothesis we built a model using a hybrid approach, i.e., individual-based rules for modeling networks and interactions, and differential equations to model growth and diffusion.

4. CONCLUSIONS

This paper presented a brief review of cellular-based modeling in plant developmental biology, pointing out its strength and showing available tools. The present state is that models addressing the dynamics of root and stem vascular differentiation are few, and almost all of them work with static-domains, albeit the modeling tools proposed are more than capable to handle growing domains.

The paper also presents an example on how to analyze tissue-related dynamics using a hybrid modeling approach. The presented modeling process showed that this approach represents an efficient way to simulate tissue-level dynamics, because it provides an intuitive manner to aggregate individual based dynamics with continuous processes like diffusion. A model built with this technique permits to observe emergent processes linked to the interaction of these discrete/continuous elements. A simple tissue growth model with a process-based condition of cell division shows how VirtualLeaf is capable to link spatial individual properties and process-based internal processes.

REFERENCES

Avsian-Kretchmer, O., Cheng, J. C., Chen, L., Moctezuma, E., and Sung, Z. R. (2002). Indole acetic acid distribution coincides with vascular differentiation pattern during Arabidopsis leaf ontogeny. *Plant Physiol.* 130, 199–209. doi: 10.1104/pp.003228

Bayer, E. M., Smith, R. S., Mandel, T., Nakayama, N., Sauer, M., Prusinkiewicz, P., et al. (2009). Integration of transport-based models for phyllotaxis and midvein formation. *Genes Dev.* 23, 373–384. doi: 10.1101/gad.497009

Beck, C. B., Schmid, R., and Rothwell, G. W. (1982). Stelar morphology and the primary vascular system of seed plants. *Bot. Rev.* 48, 691–815. doi: 10.1007/BF02860874

Besson, S., and Dumais, J. (2011). Universal rule for the symmetric division of plant cells. *Proc. Natl. Acad. Sci. U.S.A.* 108, 6294–6299. doi: 10.1073/pnas.1011866108

Boghaert, E., Radisky, D. C., and Nelson, C. M. (2014). Lattice-based model of ductal carcinoma *in situ* suggests rules for breast cancer progression to an invasive state. *PLoS Comput. Biol.* 10:e1003997. doi: 10.1371/journal.pcbi.1003997

Caño-Delgado, A., Lee, J.-Y., and Demura, T. (2010). Regulatory mechanisms for specification and patterning of plant vascular tissues. *Annu. Rev. Cell Dev. Biol.* 26, 605–637. doi: 10.1146/annurev-cellbio-100109-104107

Cartenì, F., Giannino, F., Schweingruber, F. H., and Mazzoleni, S. (2014). Modelling the development and arrangement of the primary vascular structure in plants. *Ann. Bot.* 114, 619–627. doi: 10.1093/aob/mcu074

De Matteis, G., Graudenzi, A., and Antoniotti, M. (2013). A review of spatial computational models for multi-cellular systems, with regard to intestinal crypts and colorectal cancer development. *J. Math. Biol.* 66, 1409–1462. doi: 10.1007/s00285-012-0539-4

De Rybel, B., Adibi, M., Breda, A. S., Wendrich, J. R., Smit, M. E., Novák, O., et al. (2014). Integration of growth and patterning during vascular tissue formation in Arabidopsis. *Science* 345, 1255215. doi: 10.1126/science.1255215

Dias, A. S., de Almeida, I., Belmonte, J. M., Glazier, J. A., and Stern, C. D. (2014). Somites without a clock. *Science* 343, 791–795. doi: 10.1126/science.1247575

Donner, T. J., Sherr, I., and Scarpella, E. (2009). Regulation of preprocambial cell state acquisition by auxin signaling in Arabidopsis leaves. *Development* 136, 3235–3246. doi: 10.1242/dev.037028

Dupuy, L., Mackenzie, J., Rudge, T., and Haseloff, J. (2008). A system for modelling cell–cell interactions during plant morphogenesis. *Ann. Bot.* 101, 1255–1265. doi: 10.1093/aob/mcm235

Errera, L. (1886). Sur une Condition Fondamentale d'équilibre des Cellules Vivantes. *C. R. Hebd. Séances Acad. Sci.* 103, 822–824.

Esau, K. (1960). Anatomy of seed plants. *Soil Sci.* 90, 149. doi: 10.1097/00010694-196008000-00031

Evans, M. L., Ishikawa, H., and Estelle, M. A. (1994). Responses of Arabidopsis roots to auxin studied with high temporal resolution: comparison of wild type and auxin-response mutants. *Planta* 194, 215–222. doi: 10.1007/BF01101680

Feugier, F. G., Mochizuki, A., and Iwasa, Y. (2005). Self-organization of the vascular system in plant leaves: inter-dependent dynamics of auxin flux and carrier proteins. *J. Theor. Biol.* 236, 366–375. doi: 10.1016/j.jtbi.2005.03.017

Giverso, C., and Preziosi, L. (2014). "Using mathematical modelling as a virtual microscope to support biomedical research," in *Mathematical Models and Methods for Planet Earth,* eds A. Celletti, U. Locatelli, T. Ruggieri, and E. Strickland (Zurich: Springer International Publishing), 59–71.

Glazier, J. A., and Graner, F. (1993). Simulation of the differential adhesion driven rearrangement of biological cells. *Phys. Rev. E* 47, 2128. doi: 10.1103/physreve.47.2128

Grieneisen, V. A., Xu, J., Marée, A. F. M., Hogeweg, P., and Scheres, B. (2007). Auxin transport is sufficient to generate a maximum and gradient guiding root growth. *Nature* 449, 1008–1013. doi: 10.1038/nature06215

Hofmeister, W. (1863). Zusätze und berichtigungen zu den 1851 veröffentlichten untersuchungen der entwicklung höherer kryptogamen. *Jahrb. Wiss. Bot.* 3, 259–293.

Izaguirre, J. A., Chaturvedi, R., Huang, C., Cickovski, T., Coffland, J., Thomas, G., et al. (2004). Compucell, a multi-model framework for simulation of morphogenesis. *Bioinformatics* 20, 1129–1137. doi: 10.1093/bioinformatics/bth050

Jeffrey, E. C. (1903). The structure and development of the stem in the Pteridophyta and Gymnosperms. *Philos. Trans. R. Soc. B* 195, 119–146. doi: 10.1098/rstb.1903.0004

Jönsson, H., and Krupinski, P. (2010). Modeling plant growth and pattern formation. *Curr. Opin. Plant Biol.* 13, 5–11. doi: 10.1016/j.pbi.2009.10.002

Jouannet, V., Brackmann, K., and Greb, T. (2015). (pro) cambium formation and proliferation: two sides of the same coin? *Curr. Opin. Plant Biol.* 23, 54–60. doi: 10.1016/j.pbi.2014.10.010

Krul, T., Kaandorp, J. A., and Blom, J. G. (2003). "Modelling developmental regulatory networks," in *Computational Science—ICCS 2003*, eds P. M. A. Sloot, D. Abramson, A. V. Bogdanov, Y. E. Gorbachev, J. J. Dongarra, and A. Y. Zomaya (Berlin; Heidelberg: Springer-Verlag), 688–697.

Lindenmayer, A. (1975). Developmental algorithms for multicellular organisms: a survey of l-systems. *J. Theor. Biol.* 54, 3–22. doi: 10.1016/S0022-5193(75)80051-8

Liu, W., and Stewart, C. N. (2015). Plant synthetic biology. *Trends Plant Sci.* 20, 309–317. doi: 10.1016/j.tplants.2015.02.004

Mähönen, A. P., Bishopp, A., Higuchi, M., Nieminen, K. M., Kinoshita, K., Törmäkangas, K., et al. (2006). Cytokinin signaling and its inhibitor ahp6 regulate cell fate during vascular development. *Science* 311, 94–98. doi: 10.1126/science.1118875

Mebatsion, H. K., Verboven, P., Verlinden, B. E., Ho, Q. T., Nguyen, T. A., and Nicolaï, B. M. (2006). Microscale modelling of fruit tissue using voronoi tessellations. *Comput. Electron. Agricul.* 52, 36–48. doi: 10.1016/j.compag.2006.01.002

Meinhardt, H. (1982). *Models of Biological Pattern Formation*. London: Academic Press.

Merks, R. M. H., and Glazier, J. A. (2005). A cell-centered approach to developmental biology. *Phys. A Stat. Mech. Appl.* 352, 113–130. doi: 10.1016/j.physa.2004.12.028

Merks, R. M. H., Guravage, M., Inzé, D., and Beemster, G. T. S. (2011). Virtualleaf: an open-source framework for cell-based modeling of plant tissue growth and development. *Plant Physiol.* 155, 656–666. doi: 10.1104/pp.110.167619

Mitchison, G. J. (1980). A model for vein formation in higher plants. *Proc. R. Soc. Lond. B Biol. Sci.* 207, 79–109. doi: 10.1098/rspb.1980.0015

Muraro, D., Mellor, N., Pound, M. P., Lucas, M., Chopard, J., Byrne, H. M., et al. (2014). Integration of hormonal signaling networks and mobile micrornas is required for vascular patterning in Arabidopsis roots. *Proc. Natl. Acad. Sci. U.S.A.* 111, 857–862. doi: 10.1073/pnas.1221766111

Nikolaev, S. V., Zubairova, U. S., Penenko, A. V., Mjolsness, E. D., Shapiro, B. E., and Kolchanov, N. A. (2013). Model of structuring the stem cell niche in shoot apical meristem of *Arabidopsis thaliana*. *Dokl. Biol. Sci.* 452, 316–319. doi: 10.1134/S0012496613050104

Palm, M. M., and Merks, R. M. H. (2015). "Large-scale parameter studies of cell-based models of tissue morphogenesis using compucell3d or virtualleaf," in *Tissue Morphogenesis*, ed C. M. Nelson (New York, NY: Springer), 301–322.

Perrot-Rechenmann, C. (2010). Cellular responses to auxin: division versus expansion. *Cold Spring Harbor Perspect. Biol.* 2:a001446. doi: 10.1101/cshperspect.a001446

Popławski, N. J., Shirinifard, A., Swat, M., and Glazier, J. A. (2008). Simulation of single-species bacterial-biofilm growth using the glazier-graner-hogeweg model and the compucell3d modeling environment. *Math. Biosci. Eng. MBE* 5:355. doi: 10.3934/mbe.2008.5.355

Pradal, C., Dufour-Kowalski, S., Boudon, F., Fournier, C., and Godin, C. (2008). Openalea: a visual programming and component-based software platform for plant modelling. *Funct. Plant Biol.* 35, 751–760. doi: 10.1071/FP08084

Prusinkiewicz, P., and Runions, A. (2012). Computational models of plant development and form. *New Phytol.* 193, 549–569. doi: 10.1111/j.1469-8137.2011.04009.x

Rudge, T. J., Steiner, P. J., Phillips, A., and Haseloff, J. (2012). Computational modeling of synthetic microbial biofilms. *ACS Syn. Biol.* 1, 345–352. doi: 10.1021/sb300031n

Sachs, J. (1877). *Über die Anordnung der Zellen in Jüngsten Pflanzentheilen*. Würzburg: Stahel'schen.

Sachs, T. (1969). Polarity and the induction of organized vascular tissues. *Ann. Bot.* 33, 263–275.

Savill, N. J., and Sherratt, J. A. (2003). Control of epidermal stem cell clusters by notch-mediated lateral induction. *Dev. Biol.* 258, 141–153. doi: 10.1016/S0012-1606(03)00107-6

Scarpella, E., Marcos, D., Friml, J., and Berleth, T. (2006). Control of leaf vascular patterning by polar auxin transport. *Genes Dev.* 20, 1015–1027. doi: 10.1101/gad.1402406

Scianna, M., and Preziosi, L. (2012). Multiscale developments of the cellular potts model. *Multiscale Model. Simul.* 10, 342–382. doi: 10.1137/100812951

Shapiro, B. E., Levchenko, A., Meyerowitz, E. M., Wold, B. J., and Mjolsness, E. D. (2003). Cellerator: extending a computer algebra system to include biochemical

arrows for signal transduction simulations. *Bioinformatics* 19, 677–678. doi: 10.1093/bioinformatics/btg042

Shapiro, B. E., Meyerowitz, E. M., and Mjolsness, E. (2013). Using cellzilla for plant growth simulations at the cellular level. *Front. Plant Sci.* 4:408. doi: 10.3389/fpls.2013.00408

Sieburth, L. E., and Deyholos, M. K. (2006). Vascular development: the long and winding road. *Curr. Opin. Plant Biol.* 9, 48–54. doi: 10.1016/j.pbi.2005.11.008

Smith, C. (2006). *On Vertex-vertex Systems and Their Use in Geometric and Biological Modelling*. Ph.D. thesis, Calgary, Alta., Canada.

Swat, M. H., Thomas, G. L., Shirinifard, A., Clendenon, S. G., and Glazier, J. A. (2015). Emergent stratification in solid tumors selects for reduced cohesion of tumor cells: a multi-cell, virtual-tissue model of tumor evolution using compucell3d. *PLoS ONE* 10:e0127972. doi: 10.1371/journal.pone.0127972

Thompson, D. W. (1942). *On Growth and Form*. Cambridge: Cambridge University Press.

Turing, A. M. (1952). The chemical basis of morphogenesis. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 237, 37–72. doi: 10.1098/rstb.1952.0012

Turner, S., and Sherratt, J. A. (2002). Intercellular adhesion and cancer invasion: a discrete simulation using the extended potts model. *J. Theor. Biol.* 216, 85–100. doi: 10.1006/jtbi.2001.2522

Vert, G., and Chory, J. (2006). Downstream nuclear events in brassinosteroid signalling. *Nature* 441, 96–100. doi: 10.1038/nature04681

Vincenot, C. E., Giannino, F., Rietkerk, M., Moriya, K., and Mazzoleni, S. (2011). Theoretical considerations on the combined use of system dynamics and individual-based modeling in ecology. *Ecol. Model.* 222, 210–218. doi: 10.1016/j.ecolmodel.2010.09.029

Wabnik, K., Robert, H. S., Smith, R. S., and Friml, J. (2013). Modeling framework for the establishment of the apical-basal embryonic axis in plants. *Curr. Biol.* 23, 2513–2518. doi: 10.1016/j.cub.2013.10.038

Zajac, M., Jones, G. L., and Glazier, J. A. (2003). Simulating convergent extension by way of anisotropic differential adhesion. *J. Theor. Biol.* 222, 247–259. doi: 10.1016/S0022-5193(03)00033-X

Zeng, W., Thomas, G. L., and Glazier, J. A. (2004). Non-turing stripes and spots: a novel mechanism for biological cell clustering. *Phys. A Stat. Mech. Appl.* 341, 482–494. doi: 10.1016/j.physa.2004.03.089

Zhang, Y., Thomas, G. L., Swat, M., Shirinifard, A., and Glazier, J. A. (2011). Computer simulations of cell sorting due to differential adhesion. *PLoS ONE* 6:e24999. doi: 10.1371/journal.pone.0024999

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Environmental Science

ORIGINAL RESEARCH
published: 21 July 2015
doi: 10.3389/fenvs.2015.00051

Modeling cross-scale relationships between climate, hydrology, and individual animals: generating scenarios for stream salamanders

Philippe Girard[1], *Jana Levison*[2], *Lael Parrott*[3]*, *Marie Larocque*[4], *Marie-Audray Ouellet*[4] *and David M. Green*[5]

[1] *Département de Géographie, Université de Montréal, Montreal, QC, Canada,* [2] *School of Engineering, University of Guelph, Guelph, ON, Canada,* [3] *Okanagan Institute for Biodiversity, Resilience, and Ecosystem Services, The University of British Columbia, Kelowna, BC, Canada,* [4] *Département des Sciences de la Terre et de l'Atmosphère, Centre de Recherche Pour l'Étude et la Simulation du Climat à l'Échelle Régionale, Université du Québec à Montréal, Montréal, QC, Canada,* [5] *Redpath Museum, McGill University, Montreal, QC, Canada*

OPEN ACCESS

Edited by:
Marco Casazza, 'Parthenope' University of Naples, Italy

Reviewed by:
Ke-Seng Cheng, National Taiwan University, Taiwan
Jagdish Krishnaswamy, Ashoka Trust for Research in Ecology and the Environment, India

***Correspondence:**
Lael Parrott, Okanagan Institute for Biodiversity, Resilience, and Ecosystem Services, The University of British Columbia, Okanagan Campus, Science building, 1177 Research Road, Kelowna, BC V1V 1V7, Canada
lael.parrott@ubc.ca

Specialty section:
This article was submitted to Environmental Informatics, a section of the journal Frontiers in Environmental Science

Received: *16 March 2015*
Accepted: *03 July 2015*
Published: *21 July 2015*

Citation:
Girard P, Levison J, Parrott L, Larocque M, Ouellet M-A and Green DM (2015) Modeling cross-scale relationships between climate, hydrology, and individual animals: generating scenarios for stream salamanders. Front. Environ. Sci. 3:51. doi: 10.3389/fenvs.2015.00051

Hybrid modeling provides a unique opportunity to study cross-scale relationships in environmental systems by linking together models of global, regional, landscape, and local-scale processes, yet the approach is rarely applied to address conservation and management questions. Here, we demonstrate how a hybrid modeling approach can be used to assess the effect of cross-scale interactions on the survival of the Allegheny Mountain Dusky Salamander (*Desmognathus ochrophaeus*) in response to changes in temperature and water availability induced by climate change at the northern limits of its distribution. To do so, we combine regional climate modeling with a landscape-scale integrated surface-groundwater flow model and an individual-based model of stream salamanders. On average, climate scenarios depict a warmer and wetter environment for the 2050 horizon. The increase in average annual temperature and extended hydrological activity time series in the future, combined with a better synchronization with the salamanders' reproduction period, result in a significant increase in the long-term population viability of the salamanders. This indicates that climate change may not necessarily limit the survivability of small, stream-dwelling animals in headwater basins located in cold and humid regions. This new knowledge suggests that habitat conservation initiatives for amphibians with large latitudinal distributions in Eastern North America should be prioritized at the northern limits of their ranges to facilitate species migration and persistence in the face of climate change. This example demonstrates how hybrid models can serve as powerful tools for informing management and conservation decisions.

Keywords: hybrid ecological modeling, climate change, habitat conservation, hydrological modeling, individual-based modeling, stream salamanders

Introduction

While the body of literature about the impacts of climate change on a variety of species at global or regional scales is growing (Enquist, 2002; Midgley et al., 2002; Thomas et al., 2004; Malcolm et al., 2006), few studies have attempted to predict the fate of local ecosystems facing global warming. This is a major shortcoming as most conservation and management strategies are implemented locally. There are many reasons why the local scale has been mostly ignored, but one essential factor is the difficulty of relating broad scale climate change impacts to natural mechanisms affecting individuals at a given place and time (Wilbanks and Kates, 1999; Russell et al., 2009). Climate change involves complex, cross-scale effects acting concurrently upon numerous dynamic processes, both biological and environmental (Hulme, 2005). Predicting the impacts of such complex interactions is challenging, especially when data are lacking and knowledge gaps add uncertainty to climate predictions.

The potential impacts of climate change on amphibian populations provide a telling example of complex, cross-scale effects. Amphibians are believed to be highly vulnerable to global environmental change (Blaustein et al., 2001; Wake and Vredenburg, 2008) and major population declines have been documented worldwide over the past 30 years (Pounds et al., 1997; Green, 2003; D'Amen and Bombi, 2009). Amphibians are poikilotherms, which means that ambient temperature has a significant influence on these species' physiological, metabolic, and life-history traits (Carey and Alexander, 2003). In addition to these direct temperature effects, the potential of climate change to impact available water (Oki and Kanae, 2006) may also be expected to affect their habitat. Generally dependent upon moisture to varying degrees (Petranka, 1998), amphibians can be extremely vulnerable to changes in, for example, surface runoff, stream flow, groundwater levels, and associated spring activity (Alvo et al., 2003). In headwater basins associated with near-surface fractured bedrock aquifers, groundwater is likely to be vulnerable to climate change due to steep slopes, thin soil cover, and low permeability favoring runoff at the expense of infiltration (Kosugi et al., 2006). Climate change impacts on groundwater resources at a regional scale are increasingly studied (Jyrkama and Sykes, 2007; Scibek et al., 2007), but the magnitude of these changes is highly uncertain (Green et al., 2011). In particular, headwater basins have been the focus of only limited research (Levison et al., 2014a) and the impacts of climate change on habitat suitability for spring-restricted animals, such as many salamanders, remain largely unknown. However, depending on the extent and timing of the changes with respect to the species' life cycle, the impact could be very different from one part of the species' geographic range to another.

Given the uncertainty about the magnitude and timing of changing hydrological events, and their potential impacts on stream-dwelling amphibians, it is unclear how these species may be affected by climate change in the northern limits of their distributions. Although the increased temperatures may be beneficial, changes in the timing of the spring freshet or severity of summer droughts could have potentially devastating effects on these animals. In this work, climate models, a groundwater flow model and an individual-based model of salamander distribution and abundance are combined to develop a succession of models linking physical and ecological processes across global, regional and local spatial scales, with the aim of predicting the fate of the Allegheny Mountain Dusky Salamander (*Desmognathus ochrophaeus*) in the Covey Hill Natural Laboratory (Larocque et al., 2006) of southern Quebec. This approach, by integrating the linkages between regional-scale climate change, local hydrology, and salamander life cycles, may yield realistic predictions concerning the long-term survival of salamander populations, and estimates of associated uncertainty, for use in planning conservation strategies.

Materials and Methods

Species and Sampling Site

The Allegheny Mountain Dusky Salamander, *D. ochrophaeus* (Family Plethodontidae), inhabits seepage areas, small intermittent streams, and groundwater-fed springs (Sharbel and Bonin, 1992; Lannoo, 2005). The species is highly dependent on water availability for survival, reproduction and dispersal. The species can live for 4–5 years, going through a life cycle that involves a series of stages, including egg, larval, juvenile, and adult phases. Eggs are usually laid in spring in wet crevices, under logs, on moss in seepage areas or near small streams, and usually hatch in late summer. The resulting larvae must develop in water and typically metamorphose the following spring (Petranka, 1998). Adults and juveniles may disperse into adjacent wooded areas in wet weather and are often abundant on wet rock faces (Petranka, 1998). Though listed by the IUCN as "Least Concern" (Hammerson, 2004), the species' status at the northern edge of its distribution is precarious. In Canada, known occurrences are restricted to two locations: the Niagara Escarpment in southern Ontario and Covey Hill in southern Quebec. Both populations are listed as endangered federally and provincially (COSEWIC, 2009). For stream salamanders, temperature and moisture are the primary resources that limit their distribution, as well as the ability of eggs and larvae to survive, and for this reason these two variables have been chosen as the focus of this modeling study.

The Covey Hill Natural Laboratory (Larocque et al., 2006) is located immediately north of the Canada-USA border, in the northernmost extension of the Adirondack Mountains (**Figure 1**). The hill is composed of Potsdam Group sandstones, deformed and fractured during the Appalachian orogeny (Globensky, 1986). The absence of surface deposits over large areas near the top of the hill shows the extent of erosion during the last glaciation (12 Ma). The sandstone aquifer is generally unconfined over much of the study area. Groundwater flows mainly through fractures in the sandstone (Nastev et al., 2008). Where these fractures intersect the ground surface, groundwater discharge in the form of springs may be found, creating habitats well-suited for stream-dwelling salamanders. The landscape surrounding Covey Hill is a peri-urban agricultural region and is thus largely inhospitable to the Mountain Dusky Salamander. Maintaining appropriate habitat for the species on Covey Hill is therefore crucial to its survival in Southern Quebec.

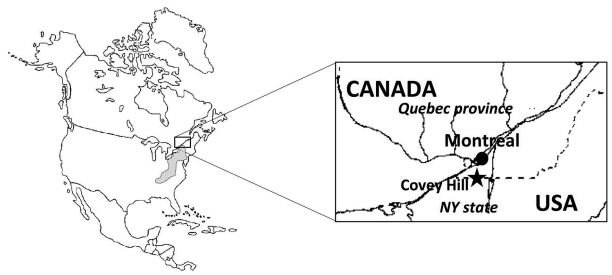


FIGURE 1 | Location of the study site in the Covey Hill area. The hill is on the northern edge of the putative distribution of the Allegheny Mountain Dusky Salamander (shaded zone).

Covey Hill is an important recharge area for the Châteauguay River regional aquifer (Croteau et al., 2010), which extends over 2500 km^2. The monthly average temperature observed between 1971 and 2000 in the Covey Hill area ranges between −9.6°C in January and 20.6°C in July with minima down to −30.5°C in winter and maxima up to 30.3°C in summer. The average annual temperature is 6.4°C. Precipitation is relatively evenly distributed throughout the year, ranging between 40.6 and 93.3 mm in February and July, respectively. The average monthly precipitation is 72 mm for a total annual average of 872.4 mm. Weather data are based on Environment Canada measurements at the Hemmingford, QC weather station.

Modeling Approach

Over the last decade, spatially-explicit individual-based models linked to landscape-scale process models have emerged as powerful tools in conservation, management, and planning (Grimm and Railsback, 2005; Stillman and Goss-Custard, 2010; McClain et al., 2012; Metcalfe et al., 2012; Parrott et al., 2012). By considering individual variations, entire life cycles, interactions among individuals and interactions between individuals and their immediate environment, this approach allows for accurate predictions about local viability of species in response to land use change or modification of landscape-scale environmental processes. The approach was used here to predict the fate of Mountain Dusky salamanders on Covey Hill in response to changes in temperature and water availability induced by climate change.

Three types of models operating at various spatial and temporal resolutions were combined (**Figure 2**). First, daily temperature and precipitation scenarios were obtained from an ensemble of regional climate models. Second, these outputs were input into an integrated surface-groundwater flow model to simulate the impact of these climate changes on groundwater-fed spring activity. Third, simulated daily time series of spring activity based on flow output from the groundwater model and daily temperature time series from the climate models were used as input data for an individual-based population model built specifically for stream-dwelling salamanders (Girard et al., 2015).

Climate Scenarios

The climate change scenarios are derived from three Regional Climate Models (RCMs) driven by five General Circulation Models (GCMs). This form of dynamic downscaling provides a better representation of both average conditions and extremes than other methods over the study area. Future RCM scenarios were further downscaled to a spatial resolution of 45 km and a temporal resolution of 24 h using the daily translation bias correction method (Mpelasoka and Chiew, 2009) to remove the biases between simulated and observed temperature and precipitation variables.

Although the ability of climate models to reproduce observed climate conditions during a reference period does not automatically guarantee more reliable results for future projections (Chiew et al., 2009; Gosling et al., 2011; Ledbetter et al., 2011) model performance remains nevertheless an important aspect to verify before any interpretation of climate simulations. Discrepancies found between observed and simulated climate variables for the reference period can be used to remove the consequent bias from future climate simulations to obtain simulations that are more in line with reality (*cf.* Samuel et al., 2012). In this study, bias correction was performed on RCM climate datasets (daily temperature and daily precipitation) following Mpelasoka and Chiew (2009). This method has proven effective not only in reproducing mean changes of temperature and precipitation, but in reproducing the full distribution of these variables, i.e., changes in extremes.

The modeled data were compared with weather data recorded at Hemmingford Four Winds meteorological station (Climate ID: 7023075) for a reference period of 1970–2000. The mean absolute deviation for observed average monthly temperature and modeled average monthly temperature over the 30 year reference period is 0.4° C (0.2°C in January and 0.2°C in July, with slightly higher deviations for the months of February is 1.4°C and September is 0.6°C). The mean absolute deviation for observed average monthly rainfall and modeled average monthly rainfall over the same 30-year period amounts to 4.6 mm or 5.5% (5.5 mm or 10.7% in January and 1.1 mm or 1.2% in July). Keeping in mind that the comparison of modeled values with corresponding values estimated from available station data must take into consideration the fact that the spatial scales related to these estimates are different (areal vs. point estimates), these values are within the acceptable margin of error which can be found in the literature (Moberg and Jones, 2004; Mailhot et al., 2007; Turco et al., 2013).

Nine climate projections were selected from the 25 dynamically downscaled simulations available for the Covey Hill area (see **Table 1**). Most of the simulations are outputs of the Canadian Regional Climate Model (CRCM) (Music and Caya, 2007) and were generated and supplied by the Ouranos Consortium on Regional Climatology and Adaptation to Climate Change. The remaining simulations are from the North American Regional Climate Change Assessment Program (NARCCAP). All projections are for the 2041–2070 climate. The nine simulations account for 85% of the future climate variability projected for the study site as established by a cluster analysis carried out on the range of available RCM scenarios. The simulations include Intergovernmental Panel on Climate Change emissions scenarios A1B and A2 (IPCC, 2000).

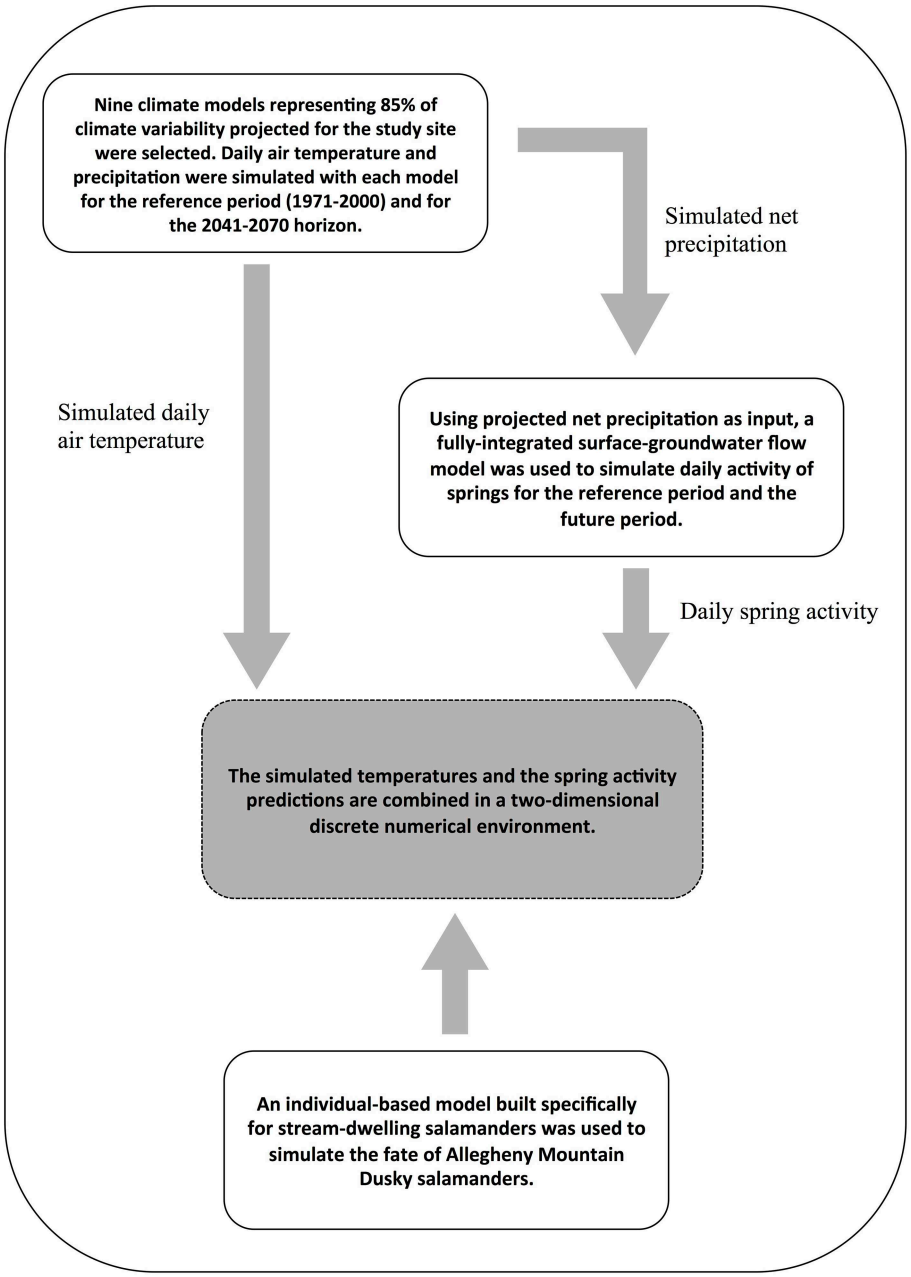


FIGURE 2 | Overview of the methodology used to predict the viability of stream salamanders in a changing climate.

Altogether, climate scenarios depict a warmer and wetter environment for the 2050 horizon (**Table 1**). The climate model ensemble predicts an increase in annual temperature ranging from 2 to 3.5°C over the 2041–2070 period, compared to the 1971–2000 reference period. An increase between 3 and 17% in annual precipitation is expected during the same period. This increase varies depending on the month of the year, although variability remains relatively equal throughout the year. Consensus among models over precipitation change is however less apparent. On average, models forecast an increase in precipitation for all months except in June. Although most models predict an increase in precipitation during the winter, the range of values is wide (from −3 to +47%). Conversely, no real consensus can be observed during the summer and autumn, where precipitation changes vary from −19 to +33%. More details on the climate scenarios can be found in Levison et al. (2014a).

Following Oudin et al. (2005), the increased temperatures predicted by the climate scenarios induce a 15 to 21% increase in predicted potential evapotranspiration as compared to the reference period. Stemming from these changes, net precipitation (i.e., precipitation—potential evapotranspiration)

TABLE 1 | Details for the nine climate projections selected for this study out of the group of 25 available for the region: emission scenario, simulated air temperature and precipitation changes, and derived potential evapotranspiration changes (derived from Levison et al., 2014b).

	RCM	GCM	Run	Domain	Emission scenario	Air temperature change (°C)	Precipitation (% change)	Net precipitation (% change)
1	CRCM4.2.3	CGCM3	5	AMNO	A2	3.1	8	13
2	CRCM4.2.3	CGCM3	2	AMNO	A2	3.2	9	12
3	CRCM4.2.3	ECHAM5	1	AMNO	A2	2.2	12	15
4	CRCM4.2.3	ECHAM5	2	AMNO	A2	2.5	10	6
5	CRCM4.2.3	Arpège UnifS2	–	AMNO	A1B	1.9	7	8
6	CRCM4.2.0	CGCM3	4	AMNO	A2	2.8	6	5
7	CRCM	CCSM	–	N. Amer.	A2	3	2	−4
8	ECP2	GFDL	–	N. Amer.	A2	2.6	12	11
9	RCM3	CGCM3	–	N. Amer.	A2	2.7	3	0

Model acronyms refer to the specific regional climate model (RCM) or global climate model (GCM) used and follow the naming convention of Mearns et al. (2012). Emission scenarios A2 and A1B are defined by the Intergovernmental Panel on Climate Change (IPCC).

varies from a 4% decrease (CRCM_CCSM) to a 15% increase (CRCM4.2.3_ECHAM#1). The majority of net precipitation scenarios agree on the sign of change: seven out of nine predict an increase in mean net precipitation.

Hydrogeological Model

The daily flow of groundwater through bedrock fractures to springs on the northeastern face of Covey Hill was simulated by Levison et al. (2014b) using HydroGeoSphere software (HGS; Therrien et al., 2012). The numerical integrated surface-groundwater flow model simulated four springs at elevations of 140, 150, 162, and 177 m. Levison et al. (2014b) have shown that for the 1971–2000 reference period, the lower elevation spring (140 m) is the one that flows most often during the year (on average 282 days, more than 75% of the year) (**Figure 3**). Spring activity decreases sharply at 150 m and at 162 m and increases slightly at the highest spring located at 177 m. The greater activity at the lowest spring is explained by its position near the base of the hill. Flow period decreases for the 150 and 162 m elevations because of reduced water pressure from above as elevation increases. The greater spring activity period at the highest spring occurs because this spring is located at the intersection of an important sub-horizontal fracture and the ground surface, as opposed to the two middle-elevation springs which discharge from vertical fractures. Applying the net precipitation for the climate change scenarios (2041–2070) to the model, Levison et al. (2014b) has shown that increases in precipitation predicted by the climate scenarios during the 2041–2070 period has direct consequences on groundwater recharge, and thus on the spring activity time series. In the future, an increase in the length of spring activity periods is expected at all elevations. Spring activity increases on average by 7 days per year at 140 m, 9 days at 150 m, 6 days at 162 m, and 5 days at 177 m. This increase is statistically significant for all four spring elevations.

For each spring, the number of days when the spring is active, the mean length of activity periods, the seasonal partitioning of spring activity and the magnitude of flow were calculated for both the reference (1970–2000) and the future (2041–2070) periods. Binary spring activity time series were built according to whether the spring was active or inactive. Only the binary spring activity time series were used in the salamander model. Results from these simulations were used to calibrate the salamander model (see below).

Salamander Model

The individual-based model developed in this work has been previously described in Girard et al. (2015). The model simulates the life cycle (growth and development) and movement of individual salamanders at a daily time step in response to environmental variables (temperature and humidity). The biological functions implemented in the model are restricted to growth, life cycle stage transition, dispersal, reproduction, and mortality. Growth and life cycle stage transition functions are temperature dependent functions. Reproduction requires water during the reproductive period in order to occur, and eggs and larvae require water to survive. Similarly, adult salamanders will die if they cannot find water after an extended period of time. Dispersal is oriented toward humid areas and is density dependent for juveniles. The salamander model has been validated and calibrated based on field data for the Allegheny Mountain Dusky Salamander and provides reasonably accurate life cycle and distribution patterns in response to spatiotemporal temperature and humidity data for the environment (Girard et al., 2015).

In the model, the individual salamanders live and move about in a spatial environment consisting of a two-dimensional landscape describing water features and a daily air temperature profile that is uniform across the landscape. The effects of precipitation are implicitly taken into account through their effects on the spring activity time series, which affects the presence or absence of water features on the landscape. All other habitat-related aspects such as the availability of resources, shelter or the presence of predators are ignored. The landscape is represented by a grid with a resolution of 1 m^2. Each grid cell may have one of three states: dry, humid or wet, with wet cells being those located directly in streams or springs. The cells at the edges of the grid are in a fourth state (exit cell) that allows individuals to leave the numerical landscape (permanently). Thus, the

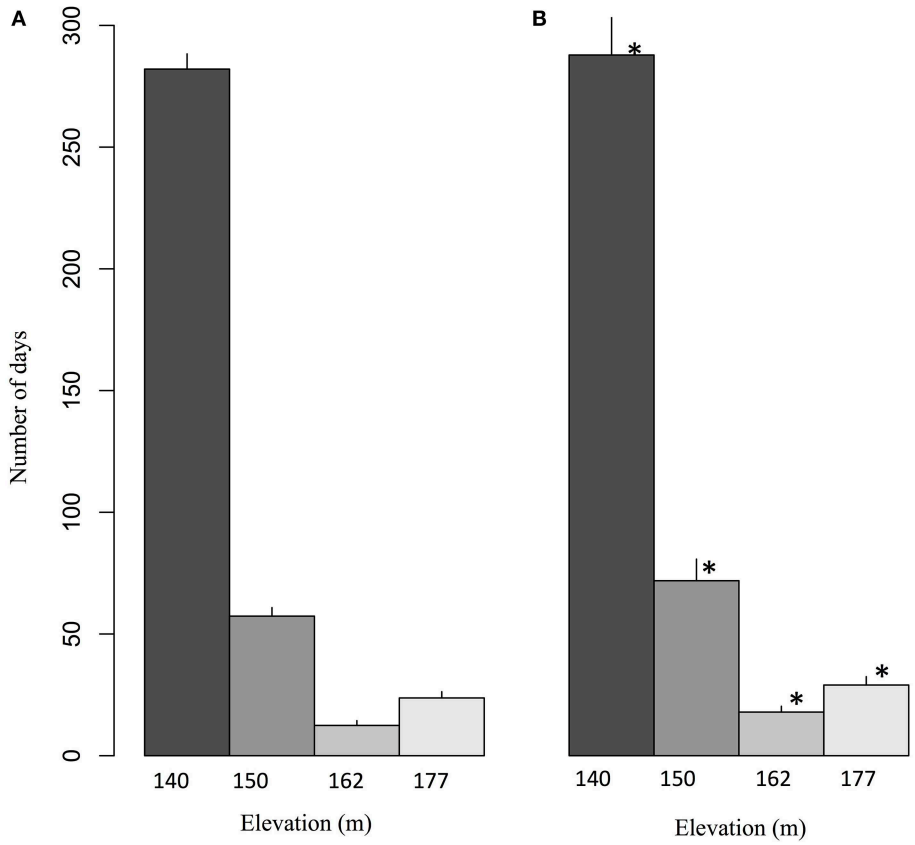


FIGURE 3 | Average number of days per year for which springs at different elevations are expected to flow during (A) the 1971–2000 reference period and (B) the 2041–2070 future period. The values correspond to ensemble averages and error bars represent the standard deviation. The stars correspond to significant differences ($\alpha = 0.05$) according to a Wilcoxon-Mann-Whitney test for paired samples (Adapted from Levison et al., 2014b).

numerical environment allows emigration, but immigration from the exterior does not occur. While the environment is represented as a discrete grid, salamanders move, and disperse in a continuous space.

The locations of the intermittent streams and groundwater springs on Covey Hill are coarsely known at best, and the density of the Allegheny Mountain Dusky Salamander population is unknown in most parts of the site. To overcome these limitations, the Covey Hill environment was represented by a simplified landscape that reproduces the expected temperature conditions and hydrodynamics predicted by the climate models and HGS model for generic springs on the hill. The simulated landscape thus represents a square-shaped spring of 1, 4, or 9 m^2 located in the center of a 50 × 50 m grid. This spring alternates between two states, active or inactive, following the activity time series derived from the hydrological simulations at elevations of 140, 150, 162, or 177 m on Covey Hill (**Figure 4A**). When the spring is active, its surface is composed of wet cells surrounded by 2 m of humid cells. The rest of the grid is composed of dry cells (**Figure 4B**). When the spring is inactive, its surface is composed of humid cells and the remainder of the grid is composed of dry cells (**Figure 4C**). Daily temperature time series are derived from the climate scenarios (**Figure 4D**). For each of these models, the reference period (1971–2000) is compared to the future period (2041–2070). In total, 108 pairs of numerical environments (three areas × four elevations × nine climate models) were tested. Although they might not be representative of specific locations on Covey Hill, these simplified environments nonetheless provide indications about the general fate of salamanders in typical habitats at different elevations of this hill and can be considered representative of similar regions in eastern North America. As described in Girard et al. (2015), the model's predictions of salamander life cycles and population densities were calibrated for the reference period using available environmental data for the region and known salamander densities estimated from field data.

Challenges in Linking the Models Across Scales

Linking models that operate at different temporal and spatial scales was one of the main challenges of this modeling study. The hydrogeological model, for example, predicted water table height over time at different elevations in our study area; this was more information than was required for the salamander model. The hydrological time series was thus simplified to retain only the necessary information, which is when the spring is active (i.e., the water table height permits spring flow) or not-active (i.e., the

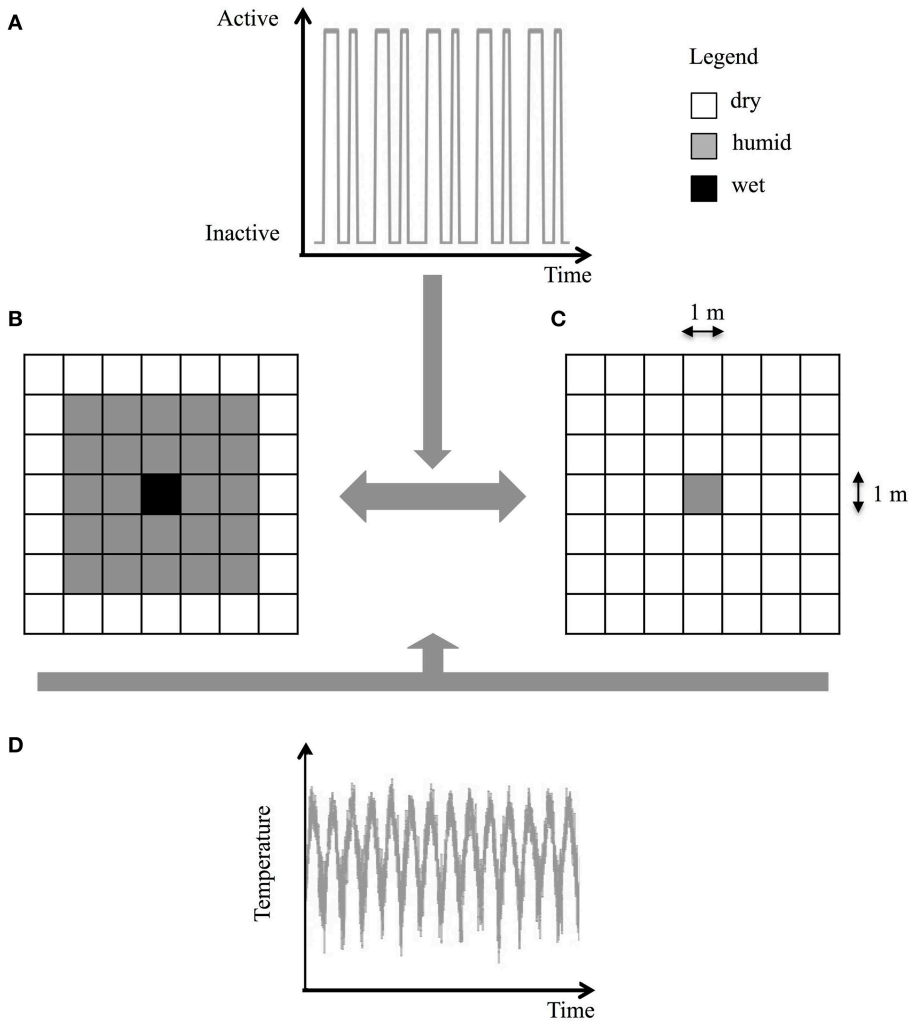


FIGURE 4 | Numerical environment of a single isolated spring located at the center of a 50 × 50 m landscape (for display purposes, the size of the landscape in (B) and (C) was reduced to 7 × 7 m). (A) Activity time series derived from the groundwater flow model, **(B)** active spring cell, **(C)** inactive spring cell, and **(D)** temperature air profile.

water table is too low for the spring to flow). The salamander and hydrological models were thus directly coupled, however the salamander model did not use all of the information output by the hydrological model. The salamander model thus did not directly use precipitation data from the regional climate model. It did use daily air temperature from the climate scenarios as a direct input for the 30-year reference period followed by the 30-year future period (2041–2070). Any uncertainty in predicted daily temperatures is thus translated into predictions of salamander viability and in particular the influence of temperature on the salamander life cycle.

Coupling the hydrogeological model to the climate data also posed challenges. The hydrogeological model requires the use of net precipitation (which is precipitation minus evapotranspiration) as an atmospheric water input. As noted in Levison et al. (2014b), the model directs net precipitation to surface flow (runoff) and infiltration (recharge). Predicted net precipitation and evapotranspiration values were derived from the regional climate scenarios described above. Although the modeled climate data satisfactorily matched historical climate records for the area, any uncertainty in predictions of the future climate have been transferred to the hydrological model and indirectly to the salamander model.

Predicting Climate Change Impacts on Salamander Populations

The impacts of climate change were assessed through a series of simulations using the calibrated salamander model implemented in the 108 pairs of numerical environments. Each simulation lasted 21,000 days. Simulations began on January 1st with 20 adults randomly distributed in the area of influence of the spring cells (wet and humid cells combined). The sex of these individuals was set randomly to ensure a balanced sex ratio. The first 10,950 days (corresponding to the 30-year reference period) were performed in springs showing constant activity to allow the model to reach a stable salamander population.

Activity and temperature profiles predicted by the hydroclimatic simulations were then implemented on the 10,951st day and run for 30 years. For each of the 216 simulated environments, 10 replicate simulations were run. At the end of each simulation, the abundance of adults over the last 5000 days was calculated. The probability of extinction through the 10 repetitions of a given numerical environment was also evaluated.

Results

The increase in the spring activity time series is not homogeneous during the year. Values presented in **Table 2** represent the interannual mean of the total number of flowing days for each spring per season. The number of flowing days during the winter increases at all four elevations, a potential indication for an earlier onset of spring freshet. In the spring season, the number of days of activity also generally increases while it decreases during the summer and fall. Activity changes for the future period are statistically significant during winter and spring seasons according to a Wilcoxon-Mann-Whitney test for paired samples ($\alpha = 0.05$), but are not significant during summer and fall. These results suggest an intensification of the hydrological activity in the Covey Hill springs for the 2050 horizon. This increased activity is likely due to a shorter winter period and to earlier spring snowmelt.

The simulations predict a strong elevation effect on the viability of salamander populations for the reference period. While stable population sizes were obtained at lower elevation springs (140 and 150 m; **Figure 5A**), the springs located at 162 and 177 m were not able to sustain salamander populations, which went locally extinct with a probability close to 1 (**Figure 5B**). However, the increase in average annual temperature and extended hydrological activity time series in the future resulted in a significant improvement for salamander populations. This improvement occurred on two fronts. First, at elevations which could sustain salamander populations during the reference period (140 and 150 m), an increase in population size of about 20% was predicted by the model (**Figure 5A**). Second, the increased hydrological activity at higher elevations (162 and 177 m) resulted in a significant decrease in the probability of extinction (**Figure 5B**). Interestingly, the increased viability of the population both in terms of abundance and extinction probability is better at 162 m than at 177 m despite the fact that spring activity is generally lower at 162 m. This result suggests that water availability is not the only factor influencing salamander population viability. Unlike the increase projected for the 162 m spring, no increase in hydrological activity is projected for the 177 m spring over the spring season (i.e., during an important part of the reproduction period). This observed difference in seasonal activity suggests that the timing of the increased spring activity exerts as great an influence as its magnitude on salamander viability.

The warmer future climate has a significant predicted impact on some critical eco-physiological traits of the salamanders such as life cycle duration and range of the reproduction period (**Figure 6**). Rising temperatures significantly accelerate the life cycle. According to the model, during the recent past, an average of 4.5 years separated oviposition from sexual maturation. This period was reduced by 5 months on average in the future due to the temperature increase (**Figure 6A**). Reproduction itself was also favored by the temperature increase since the oviposition season was extended by a dozen days on average, as a consequence of an earlier onset (**Figure 6B**).

In the salamander individual-based model, oviposition cannot be achieved if a female is not in water. Therefore, in an inactive spring, reproduction is prevented, even if ambient temperatures are adequate for reproduction. The overlap between the spring activity period and the reproduction season thus defines an effective reproduction period. During the reference period, springs were active for less than 2% of the breeding season (**Figure 7A**). However, the longer spring activity in the future, combined with a better synchronization with the reproduction period, increased this overlap to nearly 5% on average (**Figure 7A**). Simulations show that the number of years where spawning was not possible due to spring inactivity during the reproduction season decreased significantly during the future period at higher elevations (**Figure 7B**).

Discussion

The significant population improvements projected for the Allegheny Mountain Dusky Salamanders on Covey Hill for the 2050 horizon is an example demonstrating the intricate, cross-scale effects of global climate change on ecological systems and the importance of studies that link global and regional scale changes to local-scale processes. Such studies can provide

TABLE 2 | Percentage distribution of spring activity per season for the reference (REF) and the future periods (FUT), and for the four spring elevations (Adapted from Levison et al., 2014b).

Spring elevation	140 m		150 m		162 m		177 m	
	REF	FUT	REF	FUT	REF	FUT	REF	FUT
Winter (Dec, Jan, and Feb)	23.1†	25.1*	13.8	17.5*	7.9	10.8*	7.2	12.1*
Spring (Mar, Apr, and May)	25.4	27.7*	59.1	61.8*	76.9	83.4*	78.4	77.8*
Summer (Jun, Jul, and Aug)	28.1	25.7*	11.8	7.8	3.2	0.3	2.3	0.4
Fall (Sep, Oct, and Nov)	23.4	21.4	15.2	12.9	12.0	5.5	12.2	9.7

† *All values correspond to ensemble averages.*

**The stars correspond to significant differences ($\alpha = 0.05$) according to a Wilcoxon-Mann-Whitney test for paired samples.*

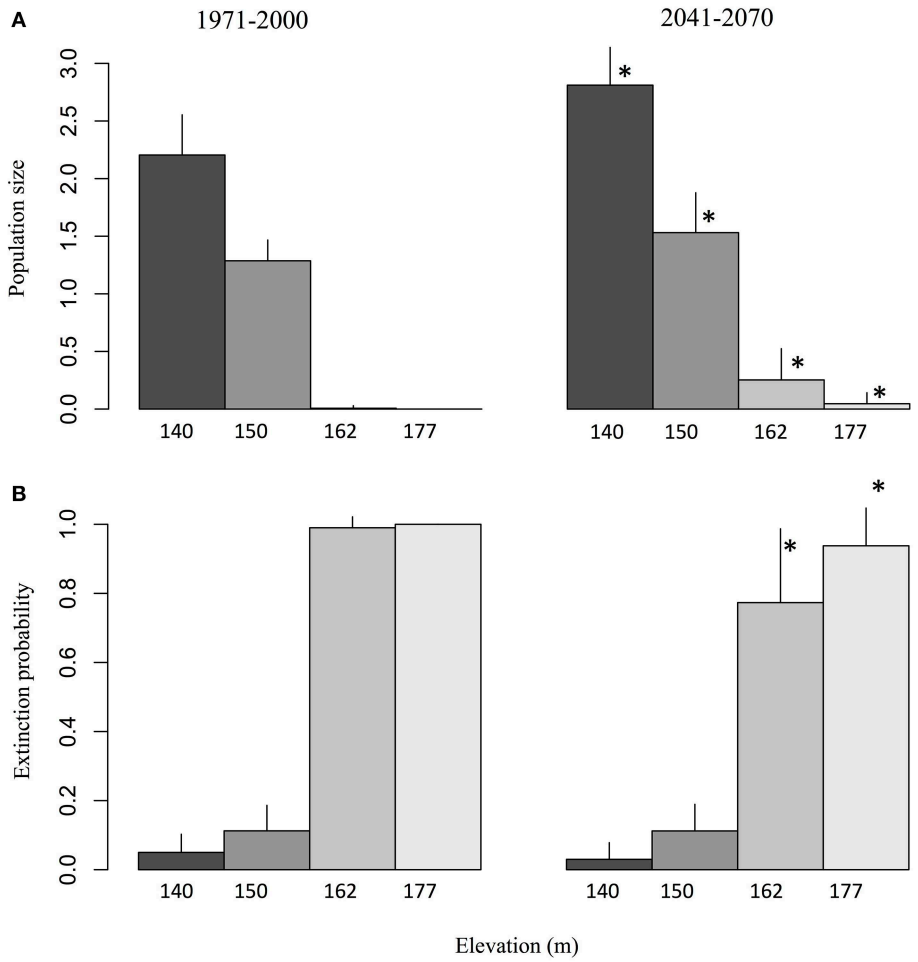


FIGURE 5 | (A) Salamander population size and **(B)** extinction probability observed over a spring of 4 m^2 during the reference and future periods, given the projected activity time series for elevations of 140, 150, 162, and 177 m. The values correspond to ensemble averages and error bars represent the standard deviation. The stars correspond to significant differences ($\alpha = 0.05$) between the reference climate (1971–2000) and the projected climate (2041–2070) according to a Wilcoxon-Mann-Whitney test for paired samples.

valuable insight for conservation initiatives that are typically carried out at this scale.

According to our model, the increased viability of the salamander population may be largely explained by the impacts of climate change on breeding phenology. The effects of global warming on amphibians' breeding have been abundantly documented (Blaustein et al., 2001; Corn, 2005; Li et al., 2013). Indeed, the timing of amphibian reproduction is primarily driven by environmental factors such as temperature and moisture (Carey and Alexander, 2003). As such, their breeding may be directly affected by global warming, resulting in a potentially earlier onset of the reproduction period (Beebee, 1995). The salamander model presented herein is in accordance with this hypothesis. However, while this effect was identified as a positive impact on the Covey Hill salamander population, some authors have suggested that it may rather increase breeder mortality due to early season frosts or floods resulting from snow melt (Corn, 2000, 2005; Inouye et al., 2000). It is acknowledged that the breeding mechanisms included in the salamander model are simplified and that neither frost nor flood-related causes of mortality are modeled. Nevertheless, oviposition is predicted to begin on average 12 days earlier in the future, corresponding approximately to the first week of May. In the context of a warming climate, the adverse effects of snow melt and early frost are rather unlikely in eastern North America at that point in the spring.

Most studies relating amphibian decline to climate change projected an increase of droughts (Araujo et al., 2006; McMenamin et al., 2008). This trend has not been detected for the Covey Hill region. As a result of the predicted increase in precipitation, it is expected that spring activity should intensify. This increase in spring flow is expected to favor salamander reproduction by prolonging the duration of the effective reproduction season. This is a major factor explaining the presence of salamanders at 162 and 177 m springs in the future. In the recent past (1971–2000), modeled springs were active for less than 2% of the oviposition period at these elevations, rendering reproduction virtually impossible. The intensification

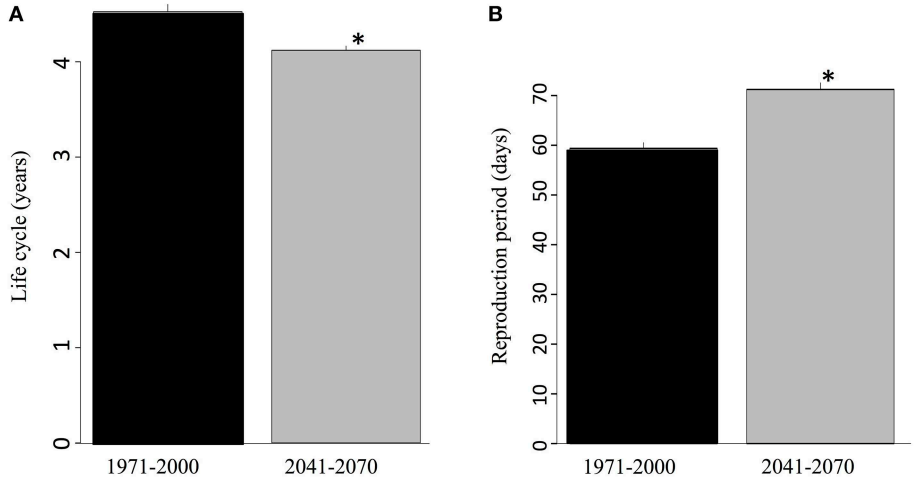


FIGURE 6 | (A) Average number of years for an individual to undergo its entire life cycle (from egg to sexual maturity) for the reference and future periods and **(B)** length of the breeding season according to the ambient temperature expected for the reference and future periods. The values correspond to ensemble averages and error bars represent the standard deviation. The stars correspond to significant differences ($\alpha = 0.05$) according to a Wilcoxon-Mann-Whitney test for paired samples.

of the hydrological activity during both the effective reproduction season and the reduction of interannual variability may thus serve to create new habitats suited for salamander viability, allowing these species to migrate northwards as the climate changes.

The effects of warmer temperatures on the speed of developmental processes could also be considered as a major benefit for salamander viability on Covey Hill. According to the model proposed here, climate change is expected to substantially accelerate life cycle duration, with a projected reduction of 5 months on average in the future. In nature, the immature stages (eggs, larvae, and juveniles) are by far the most vulnerable (Petranka, 1998). Any acceleration of the development of these phases thus greatly improves the chances of survival to the adult stage (Werner, 1986; Bruce, 2005) and consequently the likelihood that an individual will contribute to reproduction. The increased number of breeders combined with a lengthened reproductive season thus explains the increase of population size projected for the period 2041–2070. Given that Covey Hill lies at the northern limit of the Allegheny Mountain Dusky Salamander's geographical distribution, the projected improvement in conditions for the species with increased temperatures should not be so surprising. The salamander's current range includes much warmer climates to the south of the study area and thus the species should adapt well to a warming environment in the north, especially if it is correlated with an increase of available water.

These predictions must however be considered in the light of the model's limitations. For instance, the individual-based model used in this study ignores all interspecific interactions. Such a restriction is not without consequences. Interspecific competition is recognized as having impacts on the community structure of stream salamanders (Hairston, 1949; Jaeger, 1971; Smith and Pough, 1994), and their occupation of aquatic and terrestrial habitats (Krzysik, 1979). Moreover, spatial segregation or reduction of ecological niches has been observed empirically when stream-dwelling salamander species coexist (Organ, 1961; Krzysik, 1979; Keen, 1982; Grover, 2000). Predation by fish (especially of larvae), snakes, birds, small mammals, and other salamander species, has also been reported several times, thus playing a role in population size regulation (Petranka, 1998). Not only are these dynamics not taken into account by the model, but climate change may potentially alter them in an unknown manner. Furthermore, possible changes to land use on Covey Hill have not been taken into account in the future scenarios. Land use change, including potential increased groundwater withdrawals, may have significant, immediate and prolonged effects on salamander habitat quality. This commends caution: the increased viability of the Allegheny Mountain Dusky Salamander in a context of climate change at Covey Hill may prove to be too optimistic if other anthropocentric pressures are considered.

The positive impacts of climate change predicted for salamander populations on Covey Hill are important for future land conservation efforts, not because they limit the need for habitat protection, but because they reduce uncertainty about probable future conditions—that is to say that according to these results, climate change is not expected to limit the survival of the Allegheny Dusky Salamander on Covey Hill. This conclusion provides a focus for local conservation initiatives to prioritize efforts toward mitigating other identified threats to the species. It also provides support for habitat conservation initiatives for other similar amphibian species at the northern limits of their geographical distributions in the northeastern United States and eastern Canada. The Allegheny Dusky Salamander is likely not an isolated case. The results of these simulations suggest that many amphibious species may be favored by climate change in these regions, becoming abundant where they are currently scarce.

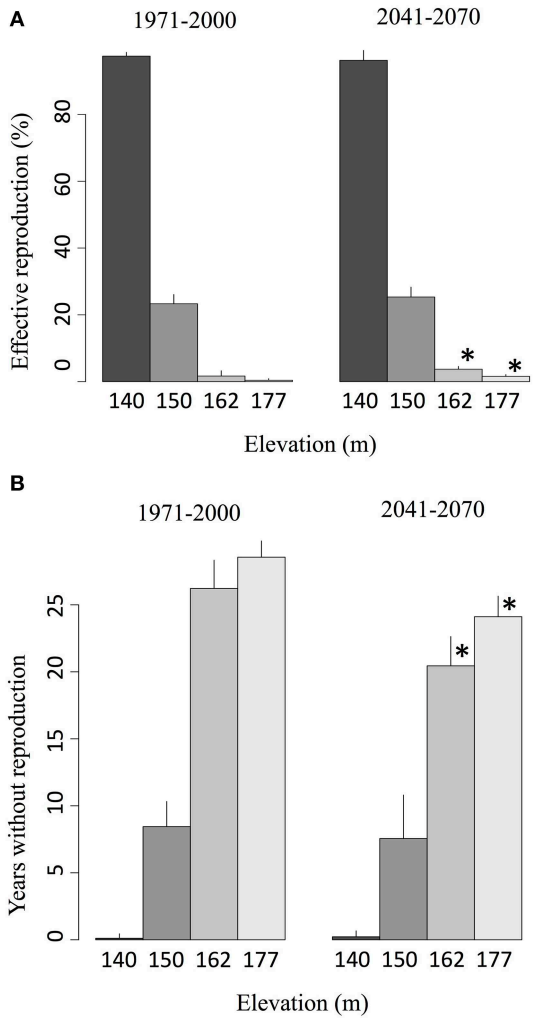


FIGURE 7 | (A) Average proportion of the reproduction season when a spring is active (effective reproduction window) given the projected activity time series for 140, 150, 162, and 177 m elevations and **(B)** number of years during which reproduction is impossible due to the inactivity of the spring for each elevation. The values correspond to ensemble averages and error bars represent the standard deviation. The stars correspond to significant differences ($\alpha = 0.05$) between the reference climate (1971–2000) and the projected climate (2041–2070) according to a Wilcoxon-Mann-Whitney test for paired samples.

Local conservation initiatives should thus seek to conserve this habitat so as to increase the chances for survival of these species in the future and to facilitate likely northern migration of the species.

It is likely that climate change will not be the only driver of change in the Covey Hill region. Other concomitant anthropogenic pressures, such as intensification of land use or groundwater withdrawal, may act to aggravate or counteract the impacts of climate change on the hydrological regime. Indeed, drainage of springs and redirection of streams for agricultural or domestic uses have been reported repeatedly. This can have significant, immediate and sustained effects on the salamander habitat quality. Caution is obviously a prerequisite when aiming at predicting the local resilience of a species facing multiple scale threats. For this reason, the precautionary principle must prevail. In fact, sound biodiversity conservation strategies are generally not dependent on whether the climate is changing or not and common conservation goals thus remain valid and include: habitat protection, dispersal corridors where habitats are fragmented, management of populations according to their carrying capacity, promotion of genetic, and ecosystem diversity. Strengthening existing land conservation measures (e.g., salamander habitat protection and conservation of preferential infiltration and recharge areas) and developing new initiatives to enhance ecosystem resilience are to be prioritized.

However, even if declining trends have been observed among amphibian populations through field studies (Pounds and Crump, 1994; Lips, 1998; Lips et al., 2006), the results obtained in this current study suggest strongly that generalization can be hazardous. There are no doubts that global warming is likely to have profound impacts on amphibians. However, the long-term effects of climate change on amphibians are difficult to predict. It is generally assumed that climate change is a threat to natural ecosystems, but it is quite possible that some species may prosper in a changed climate in certain locations. The significant population improvement projected for the Allegheny Mountain Dusky Salamanders on Covey Hill for the period 2040–2070 is an example demonstrating that the effects of global warming can be considerably more complex than one may expect, especially when looking at specific locations.

Acknowledgments

This project was funded by the Consortium Ouranos on regional climatology and adaptation to climate change, as part of the "Fonds vert" for the implementation of the Quebec Government Action Plan 2006–2012 on climate change and its measure (Grant #554007-107). The authors thank the Nature Conservancy of Canada and the Ministère des Ressources naturelles du Quebec for their support.

References

Alvo, R., Bonin, J., and Ostiguy, D. (2003). *Rapport sur la Situation de la Salamandre Sombre des Montagnes (Desmognathus ochrophaeus) au Québec*. Quebec City, QC: Société de la faune et des parcs du Québec.

Araujo, M. B., Thuiller, W., and Pearson, R. G. (2006). Climate warming and the decline of amphibians and reptiles in Europe. *J. Biogeogr.* 33, 1712–1728. doi: 10.1111/j.1365-2699.2006.01482.x

Beebee, T. J. C. (1995). Amphibian breeding and climate. *Nature* 374, 219–220. doi: 10.1038/374219a0

Blaustein, G. A., Belden, L. K., Olson, D. H., Green, D. M., Root, T. L., and Kiesecker, J. M. (2001). Amphibian breeding and climate change. *Conserv. Biol.* 15, 1804–1809. doi: 10.1046/j.1523-1739.2001.00307.x

Bruce, R. C. (2005). Theory of complex life cycles: application in plethodontid salamanders. *Herpetol. Monogr.* 19, 180–207. doi: 10.1655/0733-1347(2005)019[0180:TOCLCA]2.0.CO;2

Carey, C., and Alexander, M. A. (2003). Climate change and amphibian declines: is there a link? *Divers. Distrib.* 9, 111–121. doi: 10.1046/j.1472-4642.2003.00011.x

Chiew, F. H. S., Teng, J., Vaze, J., and Kirono, D. G. C. (2009). Influence of global climate model selection on runoff impact assessment. *J. Hydrol.* 379, 172–180. doi: 10.1016/j.jhydrol.2009.10.004

Corn, P. S. (2000). "Amphibian declines: review of some current hypotheses," in Ecotoxicology of *Amphibians* and *Reptiles*, eds D. W. Sparling, C. A. Bishop, and G. Linder (Pensacola, FL: Society of Environmental Toxicology and Chemistry), 663–696.

Corn, P. S. (2005). Climate change and amphibians. *Anim. Biodivers. Conserv.* 28, 59–67. doi: 10.1046/j.1472-4642.2003.00011.x

COSEWIC (2009). *Canadian Wildlife Species at Risk. 96*. Available online at: http://www.cosewic.gc.ca/eng/sct0/rpt/rpt_csar_e.cfm.

Croteau, A., Nastev, M., and Lefebvre, R. (2010). Groundwater recharge assessment in the Châteauguay River watershed. *Can. Water Resour. J.* 35, 451–468. doi: 10.4296/cwrj3504451

D'Amen, M., and Bombi, P. (2009). Global warming and biodiversity: evidence for climate-linked amphibian declines in Italy. *Biol. Conserv.* 142, 3060–3067. doi: 10.1016/j.biocon.2009.08.004

Enquist, C. A. F. (2002). Predicted regional impacts of climate change on the geographical distribution and diversity of tropical forests in Costa Rica. *J. Biogeogr.* 29, 519–534. doi: 10.1046/j.1365-2699.2002.00695.x

Girard, P., Parrott, L., Caron, C.-A., and Green, D. M. (2015). Effects of temperature and surface water availability on spatiotemporal dynamics of stream salamanders using pattern-oriented modelling. *Ecol. Modell.* 296, 12–23. doi: 10.1016/j.ecolmodel.2014.09.026

Globensky, Y. (1986). *Géologie de la Région de Saint-Chrysostome et de Lachine (sud)*. Quebec City, QC: Ministère de l'énergie et des ressources.

Gosling, S. N., Taylor, R. G., Arnell, N. W., and Todd, M. C. (2011). A comparative analysis of projected impacts of climate change on river runoff from global and catchment-scale hydrological models. *Hydrol. Earth Syst. Sci.* 15, 279–294. doi: 10.5194/hess-15-279-2011

Green, D. M. (2003). The ecology of extinction: population fluctuation and decline in amphibians. *Biol. Conserv.* 111, 331–343. doi: 10.1016/S0006-3207(02)00302-6

Green, T. R., Taniguchi, M., Kooi, H., Gurdak, J. J., Allen, D. M., Hiscock, K. M., et al. (2011). Beneath the surface of global change: impacts of climate change on groundwater. *J. Hydrol.* 405, 532–560. doi: 10.1016/j.jhydrol.2011.05.002

Grimm, V., and Railsback, S. F. (2005). *Individual-Based Modeling and Ecology*. New York, NY: Princeton University Press. doi: 10.1515/9781400850624

Grover, M. C. (2000). Determinants of salamander distribution along moisture gradients. *Copeia* 2000, 156–168. doi: 10.1643/0045-8511(2000)2000[0156:DOSDAM]2.0.CO;2

Hairston, N. G. (1949). The local distribution and ecology of the Plethodontid salamanders of the southern Appalachian. *Ecol. Monogr.* 19, 47–73. doi: 10.2307/1943584

Hammerson, G. (2004). "Desmognathus ochrophaeus," in *IUCN 2012. IUCN Red List of Threatened Species. Version* 2012.2, (Geneva).

Hulme, P. E. (2005). Adapting to climate change: is there scope for ecological management in the face of a global threat? *J. Appl. Ecol.* 42, 784–794. doi: 10.1111/j.1365-2664.2005.01082.x

Inouye, D. W., Barr, B., Armitage, K. B., and Inouye, B. D. (2000). Climate change is affecting altitudinal migrants and hibernating species. *Proc. Natl. Acad. Sci. U.S.A.* 97, 1630–1633. doi: 10.1073/pnas.97.4.1630

IPCC (2000). *Special report on emissions scenarios (SRES): A Special Report of Working Group III of the Intergovernmental Panel on Climate Change*. Cambridge, UK: Cambridge University Press.

Jaeger, R. G. (1971). Competitive exclusion as a factor influencing the distributions of two species of terrestrial salamanders. *Ecology* 52, 632–637. doi: 10.2307/1934151

Jyrkama, M. I., and Sykes, J. F. (2007). The impact of climate change on spatially varying groundwater recharge in the grand river watershed (Ontario). *J. Hydrol.* 338, 237–250. doi: 10.1016/j.jhydrol.2007.02.036

Keen, W. H. (1982). Habitat selection and interspecific competition in two species of plethodontid salamanders. *Ecology* 63, 94–102. doi: 10.2307/1937035

Kosugi, K., Katsura, S., Katsuyama, M., and Mizuyama, T. (2006). Water flow processes in weathered granitic bedrock and their effects on runoff generation in a small headwater catchment. *Water Resour. Res.* 42:W02414. doi: 10.1029/2005WR004275

Krzysik, A. J. (1979). Resource allocation, coexistence, and the niche structure of a streambank salamander community. *Ecol. Monogr.* 49, 173–194. doi: 10.2307/1942512

Lannoo, M., ed. (2005). *Amphibian Declines: The Conservation Status of United States Species*. Berkeley, CA: University of California Press. Available at: http://books.google.com/books

Larocque, M., Leroux, G., Madramootoo, C., Lapointe, F. J., Pellerin, S., and Bonin, J. (2006). Mise en place d'un Laboratoire naturel sur le mont Covey Hill (Québec, Canada). *VertigO* 7, 1–11. doi: 10.4000/vertigo.2118

Ledbetter, R., Prudhomme, C., and Arnell, N. (2011). A method for incorporating climate variability in climate change impact assessments: sensitivity of river flows in the Eden catchment to precipitation scenarios. *Clim. Change* 113, 803–823. doi: 10.1007/s10584-011-0386-0

Levison, J., Larocque, M., Fournier, V., Gagné, S., Pellerin, S., and Ouellet, M. A. (2014a). Dynamics of a headwater system and peatland under current conditions and with climate change. *Hydrol. Process.* 28, 4808–4822. doi: 10.1002/hyp.9978

Levison, J., Larocque, M., and Ouellet, M. A. (2014b). Modeling low-flow bedrock springs providing ecological habitats with climate change scenarios. *j. Hydrol.* 515, 16–28. doi: 10.1016/j.jhydrol.2014.04.042

Li, Y., Cohen, J., and Rohr, J. (2013). Review and synthesis of the effects of climate change on amphibians. *Integr. Zool.* 8, 145–161. doi: 10.1111/1749-4877.12001

Lips, K. (1998). Decline of a tropical montane amphibian fauna. *Conserv. Biol.* 12, 106–117. doi: 10.1046/j.1523-1739.1998.96359.x

Lips, K., Brem, F., Brenes, R., Reeve, J. D., Alford, R. A., Voyles, J., et al. (2006). Emerging infectious disease and the loss of biodiversity. *Proc. Natl. Acad. Sci. U.S.A.* 103, 3165–3170. doi: 10.1073/pnas.0506889103

Mailhot, A., Duchesne, S., Caya, D., and Talbot, G. (2007). Assessment of future change in intensity–duration–frequency (IDF) curves for Southern Quebec using the Canadian Regional Climate Model (CRCM). *J. Hydrol.* 347, 197–210. doi: 10.1016/j.jhydrol.2007.09.019

Malcolm, J. A., Canran, L., Neilson, R. P., Hansen, L., and Hannah, L. (2006). Global warming and extinctions of endemic species from biodiversity hotspots. *Conserv. Biol.* 20, 538–548. doi: 10.1111/j.1523-1739.2006.00364.x

McClain, C. R., Stegen, J. C., and Hurlbert, A. H. (2012). Dispersal, environmental niches and oceanic-scale turnover in deep-sea bivalves. *Proc. R. Soc. B* 279, 1993–2002. doi: 10.1098/rspb.2011.2166

McMenamin, S. K., Hadly, E. A., and Wright, C. K. (2008). Climatic change and wetland desiccation cause amphibian decline in Yellowstone National Park. *Proc. Natl. Acad. Sci. U.S.A.* 155, 16988–16993. doi: 10.1073/pnas.0809090105

Mearns, L. O., Arritt, R., Biner, S., Bukovsky, M. S., McGinnis, S., Sain, S., et al. (2012). The north american regional climate change assessment program : overview of Phase I results. *Bull. Amer. Meteor. Soc.* 93, 1337–1362. doi: 10.1175/BAMS-D-11-00223.1

Metcalfe, J. D., Le Quesne, W. J. F., Cheung, W. W. L., and Righton, D. A. (2012). Conservation physiology for applied management of marine fish: an overview with perspectives on the role and value of telemetry. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 367, 1746–1756. doi: 10.1098/rstb.2012.0017

Midgley, G. F., Hannah, L., Millar, D., Rutherford, M. C., and Powrie, L. W. (2002). Assessing the vulnerability of species richness to anthropogenic climate change in a biodiversity hotspot. *Glob. Ecol. Biogeogr.* 11, 445–451. doi: 10.1046/j.1466-822X.2002.00307.x

Moberg, A., and Jones, P. D. (2004). Regional climate model simulations of daily maximum and minimum near-surface temperatures across Europe compared with observed station data 1961–1990. *Clim. Dyn.* 23, 695–715. doi: 10.1007/s00382-004-0464-3

Mpelasoka, F. S., and Chiew, F. H. S. (2009). Influence of rainfall scenario construction methods on runoff projections. *J. Hydrometeorol.* 10, 1168–1183. doi: 10.1175/2009JHM1045.1

Music, B., and Caya, D. (2007). Evaluation of the hydrological cycle over the mississippi river basin as simulated by the Canadian Regional Climate Model (CRCM). *J. Hydrometeorol.* 8, 969–988. doi: 10.1175/JHM627.1

Nastev, M., Morin, R., Godin, R., and Rouleau, A. (2008). Developing a conceptual hydrological model for Potsdam sandstones in southwestern Quebec, Canada. *Hydrogeol. J.* 16, 373–388. doi: 10.1007/s10040-007-0267-9

Oki, T., and Kanae, S. (2006). Global hydrological cycles and world water resources. *Science* 313, 1068–1072. doi: 10.1126/science.1128845

Organ, J. A. (1961). Studies of the local distribution, life history, and population dynamics of the salamander genus Desmognathus in Virginia. *Ecol. Monogr.* 31, 189–220. doi: 10.2307/1950754

Oudin, L., Hervieu, F., Michel, C., Perrin, C., Andréassian, V., Anctil, F., et al. (2005). Which potential evapotranspiration input for a lumped rainfall–runoff model? *J. Hydrol.* 303, 290–306. doi: 10.1016/j.jhydrol.2004.08.026

Parrott, L., Chion, C., Gonzales, R., and Latombe, G. (2012). Agents, individuals and networks: modeling methods to inform natural resource management in regional landscapes. *Ecol. Soc.* 17, 32. doi: 10.5751/es-04936-170332

Petranka, J. W. (1998). *Salamanders of the United States and Canada*. Washington; London: Smithsonian Institution Press.

Pounds, J. A., and Crump, M. L. (1994). Amphibian declines and climate disturbance: the case of the golden toad and the harlequin frog. *Conserv. Biol.* 8, 72–85. doi: 10.1046/j.1523-1739.1994.08010072.x

Pounds, J. A., Fogden, M. P. L., Savage, J. M., and Gorman, G. C. (1997). Tests of null models for amphibian declines on a tropical mountain. *Conserv. Biol.* 11, 1307–1322. doi: 10.1046/j.1523-1739.1997.95485.x

Russell, B. D., Thompson, J.-A. I., Falkenberg, L. J., and Connell, S. D. (2009). Synergistic effects of climate change and local stressors: CO_2 and nutrient-driven change in subtidal rocky habitats. *Glob. Chang. Biol.* 15, 2153–2162. doi: 10.1111/j.1365-2486.2009.01886.x

Samuel, J., Coulibaly, P., and Metcalfe, R. A. (2012). Evaluation of future flow variability in ungauged basins: validation of combined methods. *Adv. Water Resour.* 35, 121–140. doi: 10.1016/j.advwatres.2011.09.015

Scibek, J., Allen, D. M., Cannon, A. J., and Whitfield, P. H. (2007). Groundwater–surface water interaction under scenarios of climate change using a high-resolution transient groundwater model. *J. Hydrol.* 333, 165–181. doi: 10.1016/j.jhydrol.2006.08.005

Sharbel, T. F., and Bonin, J. (1992). Northernmost record of *Desmognathus ochrophaeus*: biochemical identification in the Chateauguay River Drainage Basin, Queìbec. *J. Herpetol.* 26, 505–508. doi: 10.2307/1565133

Smith, E. M., and Pough, F. H. (1994). Intergeneric aggression among salamanders. *J. Herpetol.* 28, 41–45. doi: 10.2307/1564678

Stillman, R. A., and Goss-Custard, J. D. (2010). Individual-based ecology of coastal birds. *Biol. Rev.* 85, 413–434. doi: 10.1111/j.1469-185X.2009.00106.x

Therrien, R., McClaren, R. G., Sudicky, E. A., and Panday, S. M. (2012). HydroGeoSphere - A *Three*-dimensional *Numerical Model Describing Fully*-integrated *Subsurface* and *Surface Flow* and *Solute Transport.* Waterloo, ON: University of Waterloo.

Thomas, C. D., Cameron, A., Green, R., Bakkenes, M., Beaumont, L. J., Collingham, Y. C., et al. (2004). Extinction risk from climate change. *Nature* 427, 145–148. doi: 10.1038/nature02121

Turco, M., Sanna, A., Herrera, S., Llasat, M.-C., and Gutiérrez, J. M. (2013). Large biases and inconsistent climate change signals in ENSEMBLES regional projections. *Clim. Change* 120, 859–869. doi: 10.1007/s10584-013-0844-y

Wake, D. B., and Vredenburg, V. T. (2008). Are we in the midst of the sixth mass extinction? A view from the world of amphibians. *Proc. Natl. Acad. Sci. U.S.A.* 105, 11466–11473. doi: 10.1073/pnas.0801921105

Werner, E. E. (1986). Amphibian metamorphosis: growth rate, predation risk, and the optimal size at transformation. *Am. Nat.* 128, 319–341. doi: 10.1086/284565

Wilbanks, T. J., and Kates, R. W. (1999). Global change in local places: how scale matters. *Clim. Change* 43, 601–628. doi: 10.1023/A:1005418924748

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Environmental Science

ORIGINAL RESEARCH
published: 19 March 2015
doi: 10.3389/fenvs.2015.00017

A hybrid modeling approach to simulating foot-and-mouth disease outbreaks in Australian livestock

Richard A. Bradhurst[1]*, *Sharon E. Roche*[2], *Iain J. East*[2], *Paul Kwan*[1] and *M. Graeme Garner*[2]

[1] Discipline of Computer Science, School of Science and Technology, University of New England, Armidale, NSW, Australia,
[2] Epidemiology and One Health Program, Animal Health Policy Branch, Department of Agriculture, Canberra, ACT, Australia

Foot-and-mouth disease (FMD) is a highly contagious and economically important viral disease of cloven-hoofed animals. Australia's freedom from FMD underpins a valuable trade in live animals and animal products. An outbreak of FMD would result in the loss of export markets and cause severe disruption to domestic markets. The prevention of, and contingency planning for, FMD are of key importance to government, industry, producers and the community. The spread and control of FMD is complex and dynamic due to a highly contagious multi-host pathogen operating in a heterogeneous environment across multiple jurisdictions. Epidemiological modeling is increasingly being recognized as a valuable tool for investigating the spread of disease under different conditions and the effectiveness of control strategies. Models of infectious disease can be broadly classified as: population-based models that are formulated from the top-down and employ population-level relationships to describe individual-level behavior; individual-based models that are formulated from the bottom-up and aggregate individual-level behavior to reveal population-level relationships; and hybrid models which combine the two approaches into a single model. The Australian Animal Disease Spread (AADIS) hybrid model employs a deterministic equation-based model (EBM) to model within-herd spread of FMD, and a stochastic, spatially-explicit agent-based model (ABM) to model between-herd spread and control. The EBM provides concise and computationally efficient predictions of herd prevalence and clinical signs over time. The ABM captures the complex, stochastic and heterogeneous environment in which an FMD epidemic operates. The AADIS event-driven hybrid EBM/ABM architecture is a flexible, efficient and extensible framework for modeling the spread and control of disease in livestock on a national scale. We present an overview of the AADIS hybrid approach, a description of the model's epidemiological capabilities, and a sample case study comparing two strategies for the control of FMD that illustrates some of AADIS's functionality.

Keywords: AADIS, FMD, epidemiological model, hybrid model, spatiotemporal model

OPEN ACCESS

Edited by:
Christian E. Vincenot, Kyoto University, Japan

Reviewed by:
Jagdish Krishnaswamy, Ashoka Trust for Research in Ecology and the Environment, India
Fabrizio Carteni, University of Naples Federico II, Italy

***Correspondence:**
Richard A. Bradhurst, Discipline of Computer Science, School of Science and Technology, University of New England, Armidale, NSW 2351, Australia
rbradhur@myune.edu.au

Specialty section:
This article was submitted to Environmental Informatics, a section of the journal Frontiers in Environmental Science

Received: *05 December 2014*
Paper pending published: *14 January 2015*
Accepted: *24 February 2015*
Published: *19 March 2015*

Citation:
Bradhurst RA, Roche SE, East IJ, Kwan P and Garner MG (2015) A hybrid modeling approach to simulating foot-and-mouth disease outbreaks in Australian livestock. Front. Environ. Sci. 3:17. doi: 10.3389/fenvs.2015.00017

Introduction

An outbreak of foot-and-mouth disease (FMD) in Australia would have a major economic and social impact. This includes disruption of the domestic market for livestock and products, loss

of access to international markets, severe production and income losses in livestock and related industries, and the financial, political and social pressures of eradicating the disease (Carpenter et al., 2011; Matthews, 2011; Rushton et al., 2012). The present value of total direct economic losses from an outbreak of FMD in Australia are estimated at $5.6 to $52.2 billion AUD over 10 years, depending on the size of the outbreak and the effectiveness of control (Buetre et al., 2013).

Disease managers are faced with a number of challenges when responding to incursions of serious disease such as FMD. These include: what control measures to adopt; trade and economic implications of different control measures; how to manage resources such as personnel, equipment and vaccine; access to appropriate technology such as diagnostic tools; animal welfare issues; consumer concerns, and possible public health ramifications (Garner et al., 2007). The choice of control measures can be a compromise between the requirement for large-scale implementation and what is logistically and economically feasible (Tildesley et al., 2006). Disease models are increasingly being employed as decision support tools for outbreak planning and response (Garner and Hamilton, 2011). Models are especially useful when a country has not recently experienced the disease of concern (Bates et al., 2003b), for example, the last outbreak of FMD in Australia occurred in 1872 (Bunn et al., 1998).

Models of disease spread range from simple deterministic mathematical models (Haydon et al., 1997), through to complex spatially-explicit stochastic microsimulations (Garner and Beckett, 2005; Harvey et al., 2007; Stevenson et al., 2013). Models can be distinguished on the basis of how they handle time (discrete/continuous), space (spatially-explicit/non-spatial), and chance and uncertainty (deterministic/stochastic) (Taylor, 2003). Another way of classifying models is whether they are population-based, individual-based or a combination of both. A population-based model is formulated from the top-down and employs population-level relationships to describe individual-level behavior. An example is a traditional equation-based model (EBM) which uses a system of ordinary differential equations (ODEs) to prescribe ratios of infection states in a population over time (Keeling and Rohani, 2008). The model concisely and efficiently describes how a population 'flows' between Susceptible, Exposed, Infectious and Recovered (SEIR) compartments. Compartmental EBMs carry a general assumption of homogeneous contact rates and susceptibility, i.e., individuals mix uniformly and randomly, and have an equal likelihood of contracting a disease.

While simple mathematical models can provide useful insights into disease dynamics and epidemic behavior they tend to ignore the spatial, environmental, and social dimensions of epidemiology (Perez and Dragicevic, 2009). Assumptions of homogeneous mixing of the population and model parameters not varying over the solution interval understate the complexity of an epidemic. From a disease manager's perspective, outbreaks occur in a physical, economic, technological, management, and socio-political context (Garner and Hamilton, 2011). An epidemic environment is irregular and subject to probabilistic events that dynamically reshape the spread of disease (Bansal et al., 2007; Garner and Hamilton, 2011). Spatial effects, population heterogeneity, contact structures and social behavior all influence the course of an outbreak (Caraco et al., 2001; Hagenaars et al., 2004; Galvani and May, 2005; Lloyd-Smith et al., 2005; Bansal et al., 2007; James et al., 2007).

An individual-based model is formulated from the bottom-up and aggregates individual-level behavior to reveal population-level dynamics. Relationships between individuals emerge over time, as opposed to a population-based model where the relationships are prescribed as inputs. An example is a spatially-explicit agent-based model (ABM) where autonomous individuals with independent infection states interact within an environment. In this case, the emergent behavior of the model is the spatio-temporal spread of disease across a population. Individual-based models are well-suited to complex environmental systems due to their affinity for capturing heterogeneity, stochasticity, spatiality, social systems, and policy (Hare and Deadman, 2004), and subtle interactions between individuals that are especially important during the initial and final stages of an outbreak (Germann et al., 2006; Bansal et al., 2007). A data-driven, individual-based, modeling approach has proven popular in the field of animal health policy development with stochastic, spatially-explicit, state-transition microsimulations such as AusSpread (Garner and Beckett, 2005), InterSpread Plus (ISP) (Stevenson et al., 2013), and the North American Animal Disease Spread Model (NAADSM) (Harvey et al., 2007).

In this paper, we describe the Australian Animal Disease Spread (AADIS) model. AADIS is a national-scale hybrid model of livestock disease spread and control designed to support emergency animal disease planning in Australia. For this context, we narrow the definition of a hybrid model to one that employs both population-based and individual-based modeling techniques. AADIS models within-herd spread with a deterministic EBM and between-herd spread with a spatially-explicit stochastic ABM. The model architecture and software architecture were specifically developed for efficient handling of the national livestock population. Computational efficiency is especially important for large-scale stochastic models. It is often desirable to re-run a particular scenario hundreds, possibly thousands, of times to

Abbreviations: AADIS, Australian Animal Disease Spread model; ABARES, Australian Bureau of Agricultural and Resource Economics and Sciences; ABM, Agent-based model; ABS, Australian Bureau of Statistics; AUD, Australian dollar; AUSVETPLAN, Australian Veterinary Emergency Plan (Animal Health Australia, 2014a); CA, Control Area—a controlled area enclosing an RA and subject to lower levels of movement restrictions than those applied in RAs; CSV, Comma-Separated Values; DADS, Davis Animal Disease Simulation model (Bates et al., 2003a); DCP, Dangerous Contact Premises—a premises that, based on a risk assessment, is considered highly likely to contain an FMD-infected animal(s) or contaminated animal products, equipment or other material.; DTU-DADS, Technical University of Denmark—Davis Animal Disease Simulation model (Boklund et al., 2013); EBM, Equation-based model; FMD, Foot-and-mouth disease; GIS, Geographic Information System; GSAM, Global-Scale Agent Model (Parker and Epstein, 2011); HPAI, Highly Pathogenic Avian Influenza; HPC, High-Performance Computing; ISP, InterSpread Plus (Stevenson et al., 2013); IP, Infected Premises—a premises where infection has been confirmed.; NAADSM, North American Animal Disease Spread Model (Harvey et al., 2007).; NLIS, National Livestock Identification System; ODE, Ordinary Differential Equation; RA, Restricted Area—a controlled area surrounding an IP and subject to the highest level of movement restrictions; SEIR, Susceptible Exposed Infectious Recovered; SO, Stamping Out; SORV, Stamping Out plus suppressive Ring Vaccination; SQL, Structured Query Language.

see if patterns emerge from the underlying stochastic processes. Although the AADIS architecture supports any pathogen, FMD is the development test-case.

Modeling Context

Foot-and-Mouth Disease

FMD is an acute, highly contagious viral disease of domestic and wild cloven-hoofed animals. The disease is clinically characterized by the formation of vesicles and erosions in the mouth and nostrils, on the teats, and on the skin between and above the hoofs (Meyer and Knudsen, 2001; Animal Health Australia, 2014a). The FMD virus spreads between hosts through direct contact (e.g., movement of live animals between farms, and between farms and markets), indirect contact (e.g., livestock products, byproducts, and fomites), and aerosol (Meyer and Knudsen, 2001). Australia is a significant livestock producer and a major exporter of livestock, livestock products, and livestock genetic material. An outbreak of FMD would have severe economic consequences for the economy, in particular the loss of export markets (Buetre et al., 2013). Because of the serious consequences of an FMD outbreak, Australia invests considerable resources in prevention and planning.

Australia's approach to managing an incursion of FMD is described in the Australian Veterinary Emergency Plan—AUSVETPLAN (Animal Health Australia, 2014a). In brief, the policy is to eradicate the disease as quickly as possible using stamping out, which involves culling and disposal of infected and exposed animals. Standard zoo-sanitary measures and movement restrictions are also applied, with a minimum 3-day national livestock stand-still and the establishment of control zones around infected premises (IPs) and dangerous contact premises (DCPs). Vaccination is identified as an option under some circumstances in AUSVETPLAN.

Livestock Production Systems

Livestock production in Australia is largely based on extensive grazing and is dominated by wool, sheepmeat, beef, and dairy. Australia also has smaller intensive pig and poultry industries (Animal Health Australia, 2014b). The livestock industry is diverse and extends from the beef cattle areas of tropical north Queensland to the sheep areas of temperate southern Tasmania, and from the dairying areas of coastal New South Wales to the merino wool producing areas of Western Australia (Garner et al., 2002). The main industries that would be directly affected by FMD are beef, dairy, wool, sheepmeat, and pigs. Australia has approximately 74 million sheep, 28 million cattle and 2 million pigs on approximately 78,000 commercial farms (Australian Bureau of Statistics, 2012), with a further estimated 104,000 smallholder/lifestyle farms. From a disease transmission perspective, the key unit of interest is a herd, defined as a group of animals of the same species that is managed as a single group. Commercial farms in Australia can be large and may consist of more than one herd of the same or different species, e.g., sheep-beef farms.

Australia is a federation made up of six states and two mainland territories. The Australian Government is responsible for quarantine, disease reporting, export certification, and trade government. State and territory governments are responsible for animal health services within their respective jurisdictions. This means that while there are national policies for managing diseases like FMD, the actual control measures are administered by the jurisdictions under their own legislation. For disease control purposes, it is the farming enterprise rather than the herd that is the key unit.

AADIS Hybrid Approach

To study FMD in Australia on a national scale, a model needs to handle approximately 240,000 herds across a variety of species and production systems, as well as incorporating regional heterogeneity in disease transmission and jurisdictional variations in control measures and resourcing. To address this complexity, AADIS employs a hybrid model architecture that combines population-based modeling with individual-based modeling. This approach provides computationally efficient within-herd spread and captures the rich heterogeneous environment in which between-herd spread operates. For modeling purposes, the Australian livestock population has been categorized into 11 herd types and 10 farm types (**Table 1**). A herd has static attributes such as type, size, location, jurisdiction, and local government area, and dynamic attributes such as infection status. A farm is made up of one or more herds. Spatially, a farm and its constituent herds are defined as a point identified by latitude and longitude. The herd population is synthesized from agricultural census data (Australian Bureau of Statistics, 2012) and industry reports and data.

Equation-Based Modeling of Within-Herd Disease Spread

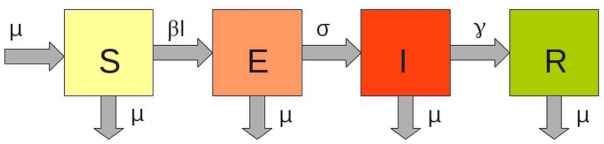
AADIS considers a herd to be well-mixed from a disease transmission point of view, i.e., all members of the herd are equally likely to contract a disease. This is a reasonable simplification for modeling the spread of disease at a national-scale, and one that lends itself to a population-based approach (Bradhurst et al., 2013). AADIS employs a non-spatial, deterministic SEIR compartmental EBM to represent within-herd spread of FMD (**Figure 1**). The approach is similar to that described by Keeling and Rohani (2008).

Each herd has its own ODE system customized for the herd type and the pathogen under study (**Figure 2**). AADIS simplifies a herd's size by considering it to be constant in the absence of culling. When a susceptible herd becomes infected the ODE system is solved numerically via a fourth-order Runge Kutta method (Cash and Karp, 1990), to yield the SEIR compartmental ratios over time. The EBM generates curves describing the prevalence and clinical signs of the infected herd. The EBM approach is computationally efficient as the solution remains in place up until an external asynchronous event acts upon the herd. If a herd is vaccinated and immunity levels increase, the EBM reacts by resolving the ODE system to yield updated SEIR compartment ratios from that point in time onwards (**Figure 3**). The EBM thus adapts and provides a dynamic representation of the within-herd infection state and presence of clinical signs.

TABLE 1 | Herd and farm types used in AADIS.

Farm type	Number of farms	Mean farm population size (min–max)	Herd type	Number of herds
Extensive beef	1331	1909 (1200–46, 575)	Extensive beef	3993
Intensive beef	51,383	280 (30 – 7436)	Intensive beef	51,383
Feedlot	508	1825 (100–39, 963)	Feedlot	508
Mixed beef/sheep	21,556	242 (30 – 5700)	Mixed beef	21,556
			Mixed sheep	21,556
Dairy	8675	298 (40 – 2742)	Dairy	8675
Small pigs	1873	244 (40 – 4850)	Small pigs	1873
Large pigs	333	4922 (1000–17, 896)	Large pigs	333
Sheep	22,150	1649 (20–44, 000)	Sheep	22,150
Small holder	103,641	5 (1 – 14)	Small holder	103,641
Total	202,775			235,668

$$\frac{dS}{dt} = \mu - \beta I S - \mu S$$
$$\frac{dE}{dt} = \beta I S - \mu E - \sigma E$$
$$\frac{dI}{dt} = \sigma E - \mu I - \gamma I$$
$$\frac{dR}{dt} = \gamma I - \mu R$$



$where\ S$ = proportion of the herd that are susceptible
E = proportion of the herd that are exposed
I = proportion of the herd that are infectious
R = proportion of the herd that are recovered
$\frac{1}{\mu}$ = average natural lifespan of the host, (μ = birth rate = natural mortality rate)
β = effective contact rate (contact rate × transmission probability)
$\frac{1}{\sigma}$ = average duration of the latent period, (σ = progression rate from exposed to infectious)
$\frac{1}{\gamma}$ = average duration of the infectious period, (γ = recovery rate)

FIGURE 1 | ODE system used by AADIS to model within-herd spread of FMD.

Agent-Based Modeling of Between-Herd Disease Spread and Control

Whilst a herd is viewed as a population for within-herd disease spread, it is somewhat paradoxically also viewed as an individual for between-herd spread (**Figure 2**). Aggregated herd-level infectious, latent (exposed), and clinical prevalence generated by the EBM, are inputs for modeling disease spread between herds. This is a sensible simplification for a model of national-scale, especially for a highly contagious disease such as FMD that when introduced into a susceptible herd will typically progress unchecked (Meyer and Knudsen, 2001; Carpenter et al., 2003). A herd is thus viewed as an atomic agent participating in an ABM for the purposes of between-herd spread.

While within-herd disease spread is deterministic and non-spatial, between-herd disease spread is highly stochastic and spatially-explicit. This is achieved through a rich ABM environment comprising disease spread pathways and control measures. The spread of disease is modeled with the following pathways:

- *Direct contact* – movement of live animals between premises,
- *Market/saleyard spread* – movement of live animals in and out of markets/saleyards,
- *Indirect contact* – movement of animal products, byproducts or fomites between herds,
- *Local spread* – proximity-based contact, e.g., over a boundary fence shared by adjoining premises,
- *Airborne transmission* – virus excreted by animals in aerosol form that remains viable in the air.

Each spread pathway has an algorithm that determines on any given simulation day whether disease transfers from an infectious herd to susceptible herd(s). AADIS introduces stochasticity through Monte Carlo sampling of probability distribution

functions (Vose, 2008). The spread of disease between heterogeneous herds is well-suited to an individual-based model such as an ABM.

The control of disease is also part of the ABM environment. This includes movement restrictions, surveillance and tracing, IP operations, resource management and vaccination. The emergent behavior of the ABM is the spatiotemporal spread of disease across the population and the subsequent activities to control and eradicate the disease. The disease spread pathways and control measures can be thought of as *components* of the ABM environment. A component has autonomous logic, its own thread of execution and a blocking queue for receiving asynchronous events. Each component of the AADIS ABM environment operates independently and concurrently.

AADIS Functional Description

Transmission Pathways

AADIS models five independent means by which FMD can be transmitted between herds:

Direct Contact Spread

Direct contact spread is driven by the movements of live animals between herds. The expected number, size and destination of daily movements into and out of herds, stratified by month, is derived from various reports and industry sources (AusVet Animal Health Services, 2005, 2006; Hassall and Associates, 2006; Kokic and Mues, 2006; East and Foreman, 2011; East et al., 2014). AADIS only models movements from infected herds since it

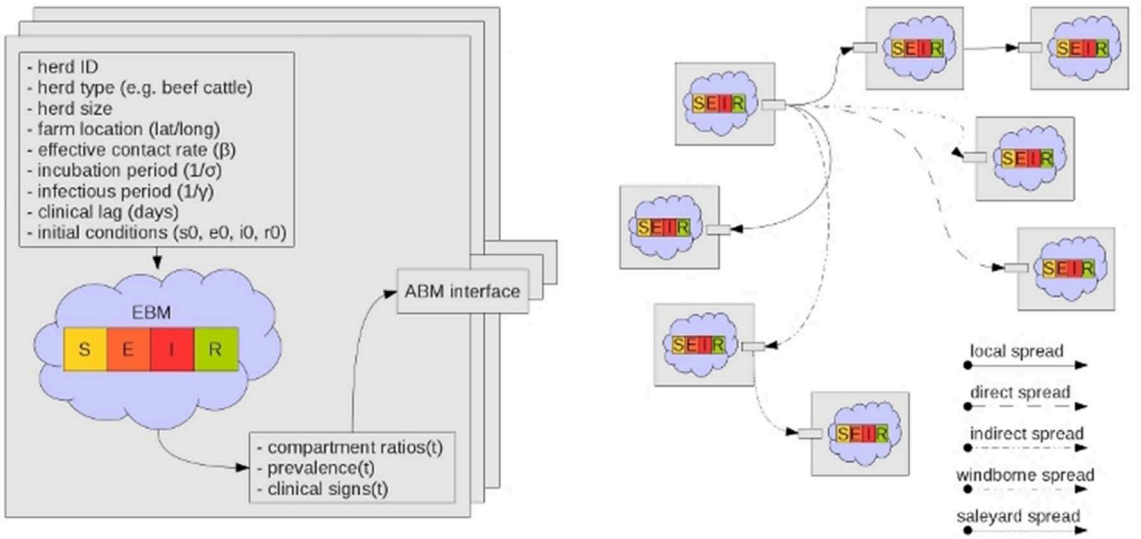


FIGURE 2 | Each herd instance has a customized SEIR ODE-based EBM. The ABM stochastically establishes infection connection paths between herds (Bradhurst et al., 2013).

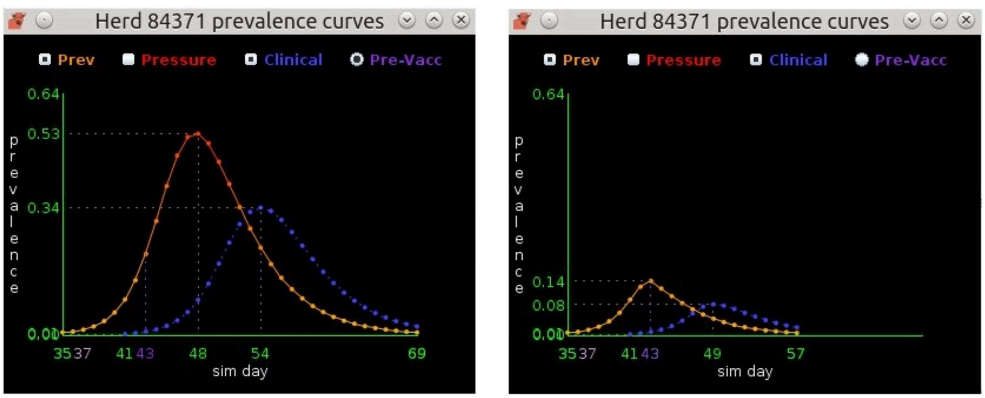


FIGURE 3 | The prevalence curves on the left depict the EBM solution for a non-vaccinated beef herd. Infection starts on day 35, peaks at 53% prevalence on day 48, and ends on day 69. The prevalence curves on the right show the same herd, but with vaccination occurring on day 37. Susceptible animals in the herd achieve immunity on day 43 resulting in a greatly diminished peak prevalence of 14% and an earlier end to infection on day 57.

would be computationally prohibitive to consider all movements from all herds. For each infected herd, the daily likelihood of a movement, the type of herd the movement is directed to, and the movement distance and direction is determined stochastically. This is based on configuration data that includes movement frequencies, distance distributions and contact matrices.

Transmission depends on the prevalence of infection in the source herd and the consignment size. The probability that a consignment contains at least one exposed or infectious animal is given by:

$$p_i = 1 - \left[1 - p(t)\right]^n \quad (1)$$

where, p_i = probability of infection, $p(t)$ = prevalence of infection in the source herd at time *t*, where prevalence is defined as the proportion of infectious and exposed animals in the herd (per the EBM), n = consignment size.

When a susceptible herd becomes infected an EBM is created and solved with initial conditions based on the proportion of infectious and exposed animals in the consignment, and the size of the destination herd.

Saleyard/Market Spread

Markets and saleyards have the potential to greatly amplify an outbreak prior to the disease being recognized and controls implemented (Gibbens et al., 2001; Mansley et al., 2003). This is because disease transmission is greatly facilitated by the large number of susceptible animals, and the degree of mixing and partitioning of animals into consignments. Outgoing consignments may then be moved to multiple, widely dispersed locations.

AADIS provides two options for simulating saleyard spread, depending on the availability of data. A simplified module takes into account the frequency and destination of consignments from different herd types. On any given day the likelihood that an infected herd sends animals to a saleyard is determined stochastically. Each infected consignment to a saleyard generates multiple infected outgoing consignments based on beta pert distributions. A more explicit representation of saleyard spread is available which takes into account the type, frequency and timing of livestock sales. This approach is driven by specific buying and selling patterns at individual livestock sales (Hassall and Associates, 2007). AADIS models plausible sale events during a simulation. If a sale happens to involve an infected herd, then a series of stochastic decisions are made to determine the number of outgoing infected consignments, the consignment destination types (herd, feedlot or abattoir), and destination locations. Infection is transmitted to the destination herds with a force relative to the viral load in the consignment.

Indirect Contact Spread

Indirect contact transmission arises from the movement between herds of contaminated animal products, byproducts, and fomites such as equipment, people and vehicles. Potential sources include veterinarians, shearing contractors, artificial insemination technicians, milk tankers, and feed delivery vehicles. Indirect contacts can be categorized as high, medium or low according to their potential for transmitting infection (Nielen et al., 1996; Bates et al., 2001; Sanson, 2005; Noremark et al., 2013). In the interests of computational efficiency, AADIS only uses a single category of indirect contacts with a specified average (baseline) probability of transmission. The user can parameterize this to represent different risk profiles. Compared to direct contacts, there is limited data on indirect contacts. The type and location of exposed herds is determined stochastically using a contact matrix and distance distributions by herd type.

If a herd is exposed through indirect contact, the probability of transmission depends on the viral load of the source herd, the relative infectiousness of the source herd (based on species and herd size), environmental conditions that influence virus survival, biosecurity practices, and relative susceptibility of the exposed herd (based on species and herd size).

$$p_i = P_b\, p(t)\; w_i\, w_s\, w_b\, w_x \quad (2)$$

where, p_i = probability that an indirect contact results in an infection, P_b = baseline probability that any indirect contact results in infection, $p(t)$ = normalized prevalence of the source herd at time *t*, w_i = infectivity weight of the source herd, w_s = susceptibility weight of the destination herd, w_b = biosecurity weight of the destination herd, w_x = seasonal weight.

Local Spread

Local spread covers the transmission of disease from an infected herd to susceptible herds in close proximity (default within 3 km). The actual means of transmission is poorly understood and could involve local aerosol spread across fences, straying of stock, vehicles, people, run off, sharing of equipment between neighbors, etc. (Gibbens et al., 2001). AADIS uses a spatial kernel approach to represent local spread, with all susceptible herds inside the local spread radius at risk. The probability of transmission for at-risk herds is decided stochastically taking into account infectious prevalence in the source herd, infectivity of the source herd (based on species and size), susceptibility of the destination herd, biosecurity measures in place at the destination premises, and distance between the source and destination herds. The influence of distance between the source herd and target herds is described by a linear decay function—the closer a herd is to the source, the greater the probability of transmission. Local spread can also occur between herds that are co-resident on the same farm, with the probability of transmission increased to reflect the higher potential for local contact between herds managed on the same farm.

$$p_i = P_b\, p(t)\; w_i\, w_s\, w_b\, w_x\, w_d\, w_n \quad (3)$$

where p_i = probability that a local contact results in an infection, P_b = baseline probability that a local contact between farms results in infection, $p(t)$ = normalized prevalence of the source herd at time *t*, w_i = infectivity weight of the source herd, w_s = susceptibility weight of the destination herd, w_b = biosecurity weight of the destination herd, w_x = seasonal weight, w_d = distance weight, w_n = detection weight.

Airborne Spread

Airborne spread is the infection of susceptible animals by virus conveyed on the wind. Pigs pose the greatest threat for airborne

spread because of their potential to excrete large quantities of virus relative to other species. Airborne spread requires a concentrated source of virus, appropriate weather conditions and susceptible animals downwind (Donaldson and Alexandersen, 2002). AADIS assumes that only pig herds are capable of transmitting FMD by airborne spread beyond a distance of 3 km. Aerosol transmissions within 3 km are captured by the local spread pathway. For each simulation day, the weather station closest to each infected pig herd is queried for conditions conducive to airborne spread (Garner et al., 2006). If suitable, a sector is constructed at the infected herd's location in the prevailing wind direction, subtended by a configurable angle (**Figure 4**). The sector radius represents the extent of the viral plume on that day and is determined by the number of infectious pigs in the source herd (Donaldson et al., 2001). Susceptible herds within the sector are identified, excluding those within 3 km. The probability of transmission takes into account the susceptible herd species, the size of the herd, and the distance of the susceptible herd from the infected herd.

$$p_i = \left[1 - \left(1 - P_{sp}\right)^n\right] w_d \quad (4)$$

where, p_i = probability that a susceptible herd will become infected, P_{sp} = probability that a single animal of the susceptible species will become infected, n = size of the susceptible herd, w_d = distance weight.

The distance weight models the diffusion of the plume over distance from the source herds, and hence the diminishing risk of transmission. Distance weight is configurable as either a linear or exponential decay function.

Disease Control

Australia's FMD policy is to eradicate the disease in the shortest possible time using a combination of strategies, while minimizing economic impact, (Animal Health Australia, 2014a). Mandatory control strategies include:

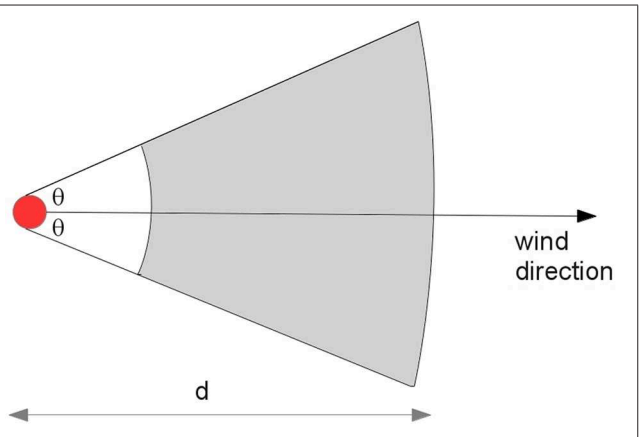


FIGURE 4 | Downwind extent of airborne spread for an infected pig herd.

- quarantine and movement controls of animals, animal products and fomites in declared areas in order to minimize the spread of infection
- tracing and surveillance to determine the source and extent of infection
- valuation and destruction of animals on infected premises and potentially on dangerous contact premises
- disposal of destroyed animals and infected animal products, and decontamination of depopulated premises.

Optional control strategies include:

- vaccination to reduce susceptibility of animals to infection and clinical disease, and potentially reduce virus excretion
- pre-emptive destruction of susceptible animals in order to minimize the spread of infection
- zoning and/or compartmentalization (to support trade)
- risk-based movement controls.

The farm is the population unit of interest for disease control. An AADIS farm has static attributes such as type and constituent herds, and dynamic attributes such as premises classification and declared area. The main simulated control strategies are movement restrictions, surveillance, tracing, IP operations and vaccination. The control and eradication phase of an outbreak commences after the declaration of the first infected premises. The day of first detection is either determined stochastically (using pre-configured probabilities of reporting by herd type, and clinical prevalence), or occurs on a fixed day at a specific or randomly selected farm.

Movement Restrictions

A national livestock standstill (minimum of 3 days), is implemented immediately following detection of the first IP. AADIS models livestock standstill by restricting the direct and saleyard spread pathways. The level of restriction depends on standstill status, type of control area, and the spread pathway being throttled. A compliance percentage for each pathway is defined in the AADIS configuration data to allow for the possibility of illegal movements during the standstill. The AADIS configuration data defines the length of the national standstill by jurisdiction. This reflects how individual jurisdictions may extend a standstill beyond the initial 3-day national period.

Controlled areas are established around each infected premises in order to restrict the movement of livestock, products and other material. The controlled areas are defined and enforced per-jurisdiction, and may be designated areas (local government, state/territory), or radius-based per IP. There are two levels of control: Restricted Areas (RAs) that immediately enclose IPs, and Control Areas (CAs) that enclose RAs. RAs have the highest level of control while CAs have a lower level of control (Animal Health Australia, 2014a). AADIS models the imposition of controlled areas in a staged manner. Larger controlled areas are enforced at the start of an outbreak. As the control program progresses, the dimensions of the controlled areas are reduced according to per-jurisdictional preferences. A radius-based controlled area is clipped to fall within the jurisdictional boundaries of the subject

IP. When IPs are clustered a meta-RA and meta-CA are formed from the union of the constituent RAs and CAs.

Surveillance and Tracing

Surveillance is the process by which new infections are identified and declared. During an FMD outbreak, surveillance is used to detect new outbreaks, define the extent of infection, and demonstrate freedom in uninfected areas (Animal Health Australia, 2014a).

AADIS allows for reporting of suspect cases on an *ad hoc* basis by owners/inspectors or others. This represents one of the most important mechanisms for finding new IPs (McLaws et al., 2007). AADIS commences suspect case reporting the day after the first IP has been declared, and allows for both true positive and false positive reports. False positive reports identify herds that are exhibiting symptoms but are not actually infected with FMD. True positive reports are generated stochastically based on an infected herd's clinical prevalence, the probability of reporting and the expected time to report. The latter two parameters are defined per herd-type in the AADIS configuration data. The number of false positive reports generated is proportional to an *n*-day (default 3), moving average number of true positive reports. The default ratio of false to true reports is 2.34 based on McLaws et al. (2007). The modeling of both true and false reports facilitates more realistic modeling of surveillance, i.e., team resources are consumed regardless of whether a surveillance visit yields a positive assessment or not. AADIS also models the active inspection of premises within RAs. All farms within a designated distance of IPs are subject to a configurable inspection schedule (number and frequency of inspections).

Tracing is the identification of movements onto and off IPs in order to ascertain where infection may have come from, or gone to. AUSVETPLAN provides minimum periods over which tracing should be carried out (Animal Health Australia, 2014a). Tracing includes animals, products, equipment, vehicles and people. Traced premises may be true cases (and thus infected), or false (not infected). AADIS can readily identify true traces by following infection chains during a simulation, allowing for variable tracing effectiveness by herd type and pathway (direct contact vs. indirect contact), and tracing duration. False forward traces are obtained by applying the direct and indirect spread pathways to a premises of interest within the forward tracing window. False backward traces are obtained by reversing the direct and indirect spread pathways over the backwards tracing window (i.e., modeling movements onto the premises of interest). This approach results in a set of plausible false traces, i.e., premises of a suitable type and location that could well have been sources or destinations of movements of concern.

Premises that require visits by surveillance teams are identified through tracing, active inspection of premises within RAs, and reporting of suspect premises. Laboratory samples are taken when needed. Surveillance visits are prioritized according to risk (Animal Health Australia, 2014a). AADIS maintains a resource-constrained dynamic queue of premises awaiting a surveillance visit. Visits are prioritized according to a configurable scheme that takes into account premises classification, declared area and herd type. If multiple premises have the same priority, then arbitration is based on how long a premises has been waiting for a visit. The visit duration (based on herd type), visit frequency (based on priority), and overall surveillance period are configurable.

IP Operations

IP Operations is comprised of the valuation, destruction and disposal of animals (stamping out), and decontamination of premises. Stamping out is Australia's default initial policy for controlling an outbreak of FMD (Animal Health Australia, 2014a). It is considered the fastest way to reduce viral excretions on IPs and thus dampen spread. Stamping out is implemented on all IPs, and potentially on DCPs, subject to risk assessment.

Premises undergoing IP Operations transition through the following states: cull pending, cull in progress, disposal pending, disposal in progress, decontamination pending, decontamination in progress, and resolved. Each jurisdiction has separate pools of teams for culling, disposal and decontamination. When a pool is exhausted (i.e., all of the teams are on assignment), pending jobs are held in a queue. Visits to premises are prioritized based on premises classification, herd/species priority, herd size, time in queue, and proximity to an IP. The times required for a premises to undergo culling, disposal and decontamination are defined by herd type in the AADIS configuration data.

Vaccination

Vaccination is one of the available options to support stamping out of an FMD outbreak. The decision to vaccinate and the specific role of vaccination in an FMD response varies according to the specific outbreak scenario (Animal Health Australia, 2014a). Vaccination strategies include:

- *Suppressive* – vaccination is carried out inside known infected areas (RAs) in order to suppress virus production in at-risk and exposed herds to reduce further spread.
- *Protective* – vaccination is carried out outside known infected areas in order to protect susceptible animals from infection.
- *Mass* – vaccination is carried out across a broad area to large numbers of animals. This strategy could be applied if an outbreak is not under control and there is a risk of spread escalating.

AADIS provides two triggers for commencing a vaccination program: on a configurable day into the control program, or once a configurable number of IPs has been declared. AADIS models all vaccination policies with an annulus of configurable inner and outer radii. The inner radius is set to zero for suppressive and mass vaccination. A vaccination annulus is established around each target IP, and eligible premises inside the annulus are scheduled for vaccination. The user can select to only vaccinate around IPs found on or after the day the vaccination program begins, or around all new and previously identified IPs. The vaccination candidates inside each annulus are prioritized according to herd type, herd size, and proximity to the nearest IP. It is also possible to omit certain herd types from vaccination. The direction of vaccination (from the outside in, or from the inside out), is set in the AADIS configuration data.

The effect of vaccination is to increase herd immunity (i.e., reduce a herd's susceptibility to infection) over time. When a partially immune herd is exposed to infection, the virus production profile generated by the EBM reflects that some of the animals have protective immunity.

As with surveillance and IP operations, the ability to implement a vaccination program depends on the availability of resources. Each jurisdiction has a separate pool of vaccination teams. When a pool is exhausted (i.e., all of the teams are on assignment), pending jobs are held in a queue. Visits to premises are prioritized according to herd type, herd size, time in queue, and proximity to an IP. The time required for a premises to undergo vaccination is defined by herd type in the AADIS configuration data.

Resourcing

The resources required to manage an emergency animal disease outbreak include personnel (e.g., veterinarians, animal health officers, control center staff), equipment (e.g., vehicles), facilities (e.g., laboratories) and consumables (e.g., vaccine, disinfectant). Some aspects of disease control and eradication are resource-intensive and the lack of resources can severely hamper the response to an outbreak (Matthews, 2011; Roche et al., 2014). AADIS models the personnel resources required for key operational activities: surveillance, culling, disposal, decontamination, and vaccination. As state and territory governments are responsible for emergency animal disease management within their own boundaries (Animal Health Australia, 2014a), the teams are organized into pools by jurisdiction, i.e., each jurisdiction has five pools. It is anticipated that resource levels ramp up over time, so initially the pools are small and increase in a linear manner up to the maximum size. The starting point, duration of the ramp-up and maximum pool size are defined in the AADIS configuration data by resource type and by jurisdiction. AADIS tracks the availability and allocation of resources to provide immediate feedback as to whether/where the control program is resource constrained (**Figure 5**).

Model Implementation

AADIS is implemented in Java (Oracle, 2014) and employs open-source products such as PostgreSQL (PostgreSQL, 2014) and OpenMap (BBN, 2014). AADIS runs under either Linux™ or Windows™ and has an asynchronous software architecture with concurrency achieved through Java threads. This takes good advantage of the inexpensive parallelism available on the quad-core x64 target machine. Although C and C++ are perhaps more typical language choices for computationally intense applications, Java offers the advantage of platform-independence and a rich collection of utility libraries. Parker and Epstein (2011) describe how the Java-based GSAM pandemic model scales up to 6.5 billion agents distributed across a 32-node high-performance computing (HPC) cluster.

AADIS is able to run complex disease spread and control scenarios across the entire Australian population of FMD-susceptible herds efficiently on a single desktop platform. For example, a 100-day national outbreak with all disease spread pathways enabled, all control measures deployed, dynamic resourcing, report writing, and real-time visualization takes around 10 s to complete on a quad-core laptop with 16 GB RAM.

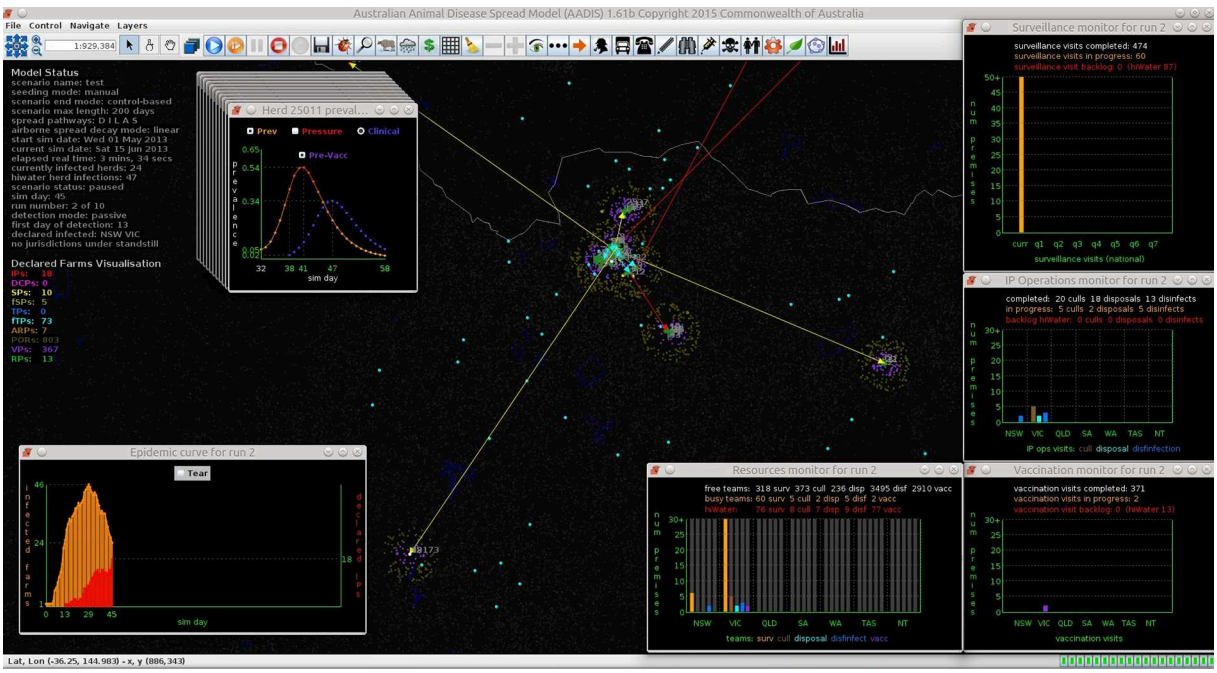


FIGURE 5 | Dynamic visualization of outbreak dynamics.

This is achieved through several strategies including the hybrid model architecture, asynchronous software architecture, a grid-based spatial indexing system (in lieu of geospatial Structured Query Language (SQL) exchanges with the database server), and lightweight agents. Implementation and performance details of interest will feature in a future paper. **Figure 6** provides a summary of the AADIS subsystems and the main relationships between them.

Simulation Flow

AADIS operates in discrete time steps of a day. At the start of a simulation day, the disease spread components and the control components access the herd/farm initial conditions for the day. All components then independently and concurrently proceed with their daily processing, making various stochastic decisions on the spread and control of disease. As each component finishes its daily processing, a set of herd/farm update requests are sent

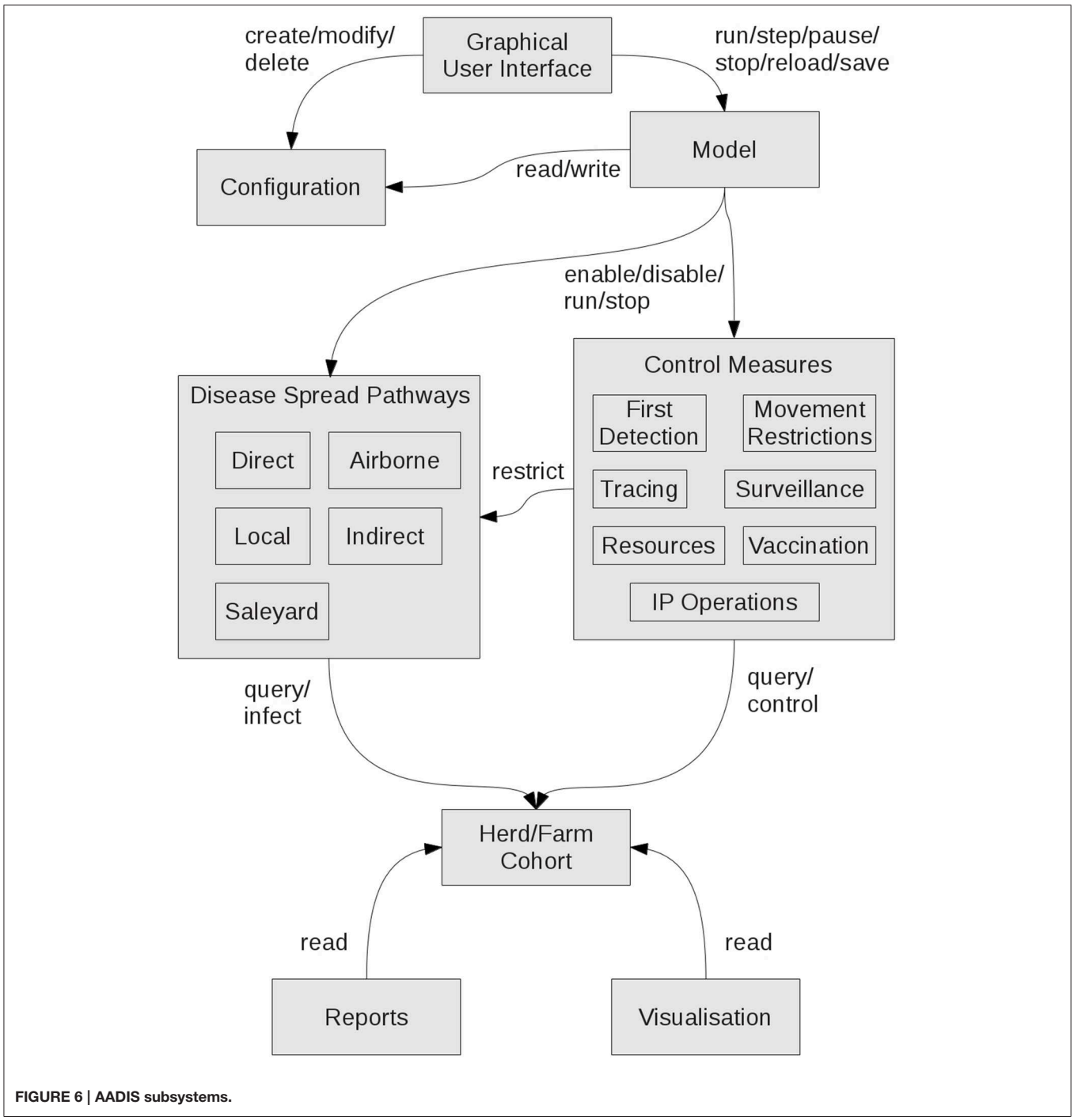


FIGURE 6 | AADIS subsystems.

asynchronously to the ABM component scheduler where they are queued. When all updates have been received, they are collated and submitted to the cohort of herd/farm agents. The new herd/farm reality is then available for all components at the start of the next simulation day. This concurrent approach is computationally efficient and reflects the reality that spread and control proceed independently and in parallel during an outbreak. The scheduler arbitrates whenever one or more components attempt to act upon the same herd/farm on the same day. The arbitration may be random or rule-based. For example, if the direct and indirect spread pathways both attempt to infect the same herd on the same day, then the scheduler randomly selects one pathway to succeed. On the other hand, if for example, the IP Operations component and the Vaccination component both attempt to control the same farm on the same day, then the scheduler always gives priority to IP Operations.

Configuration Data

AADIS has three levels of configuration:

- Project data for the study at hand includes the herd population, weather data, movement patterns and pathogen specific parameters. This data typically does not change often and may be large with cross-dependencies. The project data is stored in a relational database (PostgreSQL, 2014), and any changes require a database rebuild (which ensures referential integrity).
- Scenario-specific data is stored in a Java properties file that is persistent across multiple invocations of the model. The data is stored outside the database and so changes do not trigger a database rebuild.
- The graphical user interface can be used to make short-term adjustments of selected configuration data. The changes only last for the current invocation of the model.

Outputs

As AADIS is a stochastic model, it generates a range of possible outcomes when run with the same starting parameters. When modeling a specific scenario, multiple model runs (iterations) are used to generate a probability distribution of potential outcomes. Results of individual runs and summaries of groups of runs are produced. AADIS provides both tabular and graphical outputs.

Tabular

The formal outputs of an AADIS scenario run are comma-separated values (CSV) files. These contain a range of metrics at the herd, farm and scenario level, and are used for subsequent epidemiological analysis.

Graphical

AADIS provides three modes for visualizing an outbreak in progress:

- *Within-herd spread (EBM)* – infected herds are represented as heat-colored dots reflecting the viral load.
- *Between-herd spread (ABM)* – infected herds are represented as color-coded dots reflecting the particular pathway that triggered the infection. There is an option of displaying each pathway connection as a vector – thus depicting the entire infection network.
- *Control (ABM)* – farms are represented as color-coded dots reflecting the current premises classification.

AADIS provides a range of graphical utilities for the dynamic display of herd prevalence curves, epidemic curves, convex hull of infection, controlled areas, traces, resource usage, resource backlog, and peak resource levels (**Figure 5**).

Sample Case Study

The following simple epidemiological case study is provided to illustrate how AADIS can be used to address policy issues.

Outbreak Scenario

The south-east of Australia is an agriculturally intensive area that has previously been identified as vulnerable to an FMD outbreak (East et al., 2013). The Goulburn Valley is a 14,287 km^2 sub-region of Victoria with significant cattle and horticultural sectors (Regional Development Victoria, 2010). The dairy industry in this region comprises around 3000 farms and accounts for approximately 13% of Australia's milk production (Department of Environment and Primary Industries, 2015). Other livestock-based sectors in the region include beef, wool, sheepmeat, and pigs.

We assume FMD is introduced into the Goulburn Valley with the primary case occurring on a pig farm with a population of 3209 pigs. The farm has 20 neighboring farms within a 3 km radius. The outbreak occurs in May when the usual cool weather favors the survival of FMD virus outside a host. Detection of the index case occurs 21 days after the primary infection. **Table 2** lists some of the key EBM parameter values.

Two strategies for controlling the outbreak were assessed:

(1) Stamping out of infected premises (SO),
(2) Stamping out of infected premises plus suppressive ring vaccination (SORV).

TABLE 2 | Selected EBM ODE parameters.

Herd type	Effective contact rate	Latent period (days)	Infectious period (days)
Extensive beef	0.7	2	4
Intensive beef	2	2	4
Feedlot	8	2	4
Mixed beef	2	2	4
Mixed sheep	0.8	2	7
Dairy	6	2	4
Small pigs	6	1	4
Large pigs	8	1	4
Sheep	0.8	2	7
Small holder	2	2	5

Selected parameter settings for the control strategies are provided in **Table 3**. Note that this is a simplification of the model setup as only a subset of key parameters are described.

Method

The simulation was run 500 times for each control strategy and the following outputs compared:

- duration of the outbreak (defined as the number of days from when the index case was declared to when the last infected premises was resolved)
- cumulative number of infected premises
- cumulative number of culled premises
- cumulative number of culled animals
- cumulative number of vaccinated premises
- cumulative number of vaccinated animals

TABLE 3 | Selected control program parameter settings.

Control parameter	Value
National livestock standstill	3 days
Restricted area (RA)	Circle of 3 km radius enclosing each IP
Controlled area (CA)	Circle of 10 km radius enclosing each IP
Num days to report suspect premises after clinical signs	0–19 days (herd type-dependent)
Probability of reporting suspect premises	70–100% (herd type-dependent)
Ratio of false suspect premises reports to true reports	2.34:1
Forward tracing window	14 days
Backward tracing window	14 days
Time needed for a direct trace	0–4 days (species-dependent)
Time needed for an indirect trace	1–5 days (species-dependent)
Effectiveness of direct tracing	70–100% (species-dependent)
Effectiveness of indirect tracing	70–90% (species-dependent)
Non-compliance with direct movement controls inside RA	2%
Non-compliance with direct movement controls inside CA	2%
Reduction of indirect movements inside RA	15%
Reduction of indirect movements inside CA	50%
Surveillance visit duration	0.5 day (herd type-dependent)
Max number of surveillance teams	20 Per jurisdiction
Max number of culling teams	20 Per jurisdiction
Max number of disposal teams	20 Per jurisdiction
Max number of decontamination teams	20 Per jurisdiction
Max number of vaccination teams	200 Per jurisdiction
Days to cull a herd	0.5–14 (herd type-dependent)
Days to dispose a herd	0.5–18 (herd type-dependent)
Days to decontaminate a premises	1–28 (herd type-dependent)
Start of vaccination program	Seventh day of the control program
Days to vaccinate a herd	0.5–7 (herd type-dependent)
Vaccination annulus radii (km)	1, 3
Vaccination direction	Outside-in

In addition, a simple sensitivity analysis was carried out on selected parameters under strategy SO.

- time to first detection (7, 14, 21, 28 days)
- duration of the national standstill (0, 3, 7, 10 days)

The test hardware platform was a quad-core laptop with 16 GB RAM running 64-bit Kubuntu Linux™.

The Stata/IC statistical package (Stata, 2014) was used to analyse the distributions of the key model outputs. Data sets were imported into Stata and checked for normality. Non-parametric statistical methods were used throughout this analysis as some data sets were non-normal and could not be transformed to normality by standard transformation techniques. The number of infected premises, outbreak duration, number of culled animals, number of vaccinated premises and number of vaccinated animals were analyzed using the Kruskal-Wallis test for comparison of multiple independent groups of data. *Post hoc* analysis to identify differences between strategies was conducted using the Kruskal-Wallis test with the significance level adjusted per the Bonferroni correction for multiple pairwise comparisons. Model outcomes were expressed as medians with 90% confidence intervals.

Results

Figures 7–**9** provide visualization snapshots at day 21 of run number 1 (of 500) of the baseline stamping out scenario. **Figure 7** shows how within-herd spread is represented as heat-colors reflecting prevalence levels generated by the EBM of each infected herd. **Figure 7** also illustrates the optional display of the convex hull area of infection, in this case 33 km^2. **Figure 8** shows the infection network generated by the ABM, with color-coded vectors reflecting the particular spread pathway that triggered. At this stage in the outbreak there is only local (green) and airborne (cyan) spread emanating from the primary case pig herd. **Figure 9** shows the outbreak from a disease management point of view. Despite there being 13 infected herds on day 21, there is only one known infected premises (red). **Figure 9** also shows two optional popup windows: the prevalence curves for a herd (in this case the index case), and the epidemic curve depicting declared infected premises vs. actual infected premises. These outputs demonstrate the potential of AADIS as a training tool that provides various visualizations of disease transmission, and also contrasts a disease manager's incomplete view of an outbreak (what is known), with the physical reality (infected herds in the population).

In this study we investigated the effect of incorporating suppressive ring vaccination into the control program for an FMD outbreak. Strategy SORV was effective in reducing both the size and duration of an outbreak when compared to the baseline SO strategy. There were significantly less IPs, significantly shorter outbreaks, and significantly less culled animals than stamping out alone ($p < 0.05$) (**Figure 10** and **Table 4**). SORV was particularly effective in reducing the likelihood of a very large outbreak, which could be an important consideration for a disease manager.

The sensitivity analyses showed that findings are significantly influenced by the time to first detection. Varying the time to detection for strategy SO produced strongly correlated changes

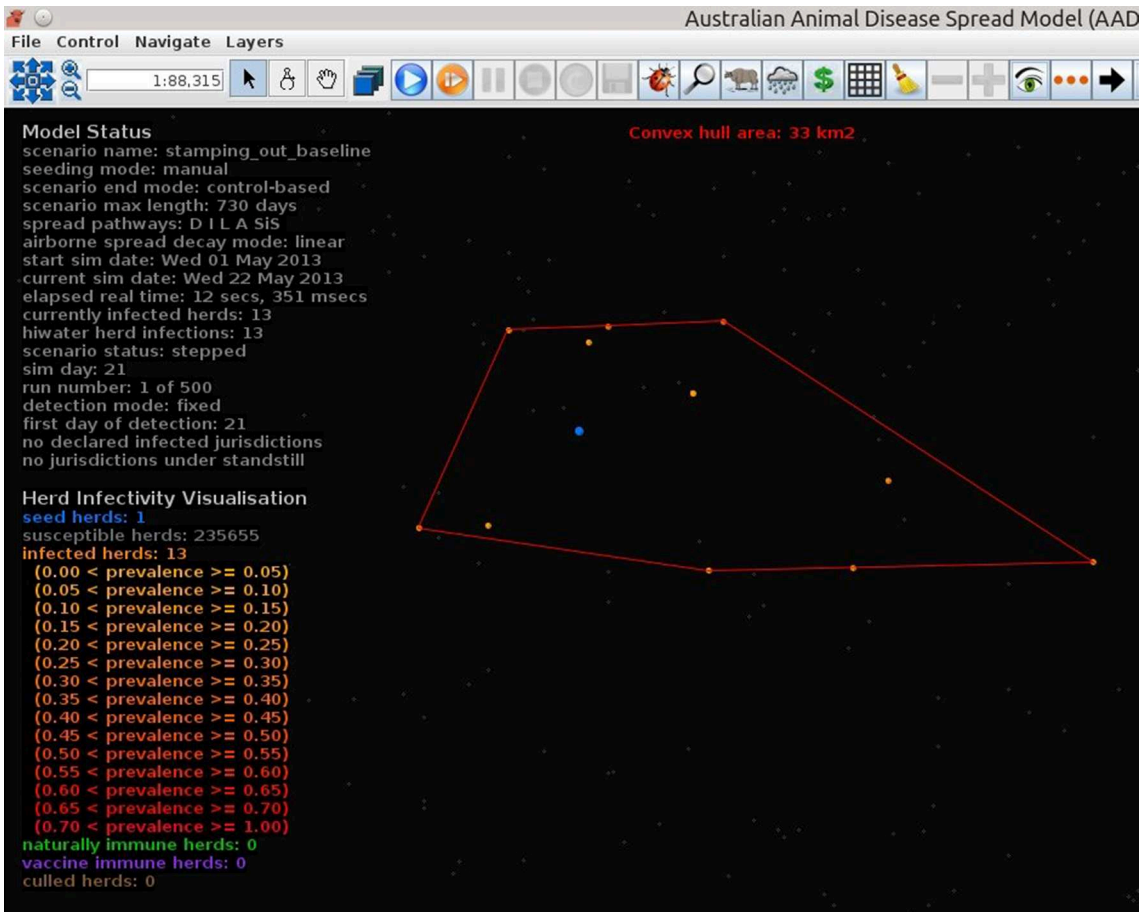


FIGURE 7 | Visualization of within-herd prevalence as heat-colors.

to the number of IPs, outbreak duration, and number of culled animals ($p < 0.05$) (**Table 4**). The findings were less sensitive to the duration of the national livestock standstill with only a 0-day standstill and a 10-day standstill producing significantly different outcomes ($p < 0.05$). This suggests that for the outbreak scenario, there is perhaps not a significant advantage in extending the default 3-day standstill.

Discussion

An incursion of FMD into Australia would have severe economic consequences and Australia thus invests heavily in prevention and contingency planning. The control and eradication of FMD is challenging due to the complexities of a highly contagious and multi-host pathogen operating in a heterogeneous environment across multiple jurisdictions. Models of disease spread and control are increasingly recognized as valuable tools for informing policy. Population-based approaches and individual-based approaches have both been used to model the spread and control of FMD and inform policy makers and disease managers. The hybrid approach of AADIS combines the advantages of population-based and individual-based approaches in order to efficiently model the spread and control of FMD on a national scale.

Advantages and Disadvantages of Population-Based Modeling

Livestock epidemics can occur in highly heterogeneous environments. Take for example, an outbreak of FMD within an extensive beef production system in a northern Australian jurisdiction, compared to one in an intensive dairy production system in a southern Australian jurisdiction. Despite the same pathogen and the same host species, there are significant differences in livestock density, farming practices, market systems and climate. The probability of disease detection and reporting varies with the level of contact between owners/inspectors and livestock. State/territory jurisdictions are responsible for their own disease control policies and resourcing. This results in distinct disease spread dynamics and control environments between the two regions.

Population-based models carry a general assumption of homogeneous contact rates and susceptibility. In the case of a compartmental SEIR EBM, individuals within any given compartment are indiscernible. The subtle contributions of specific

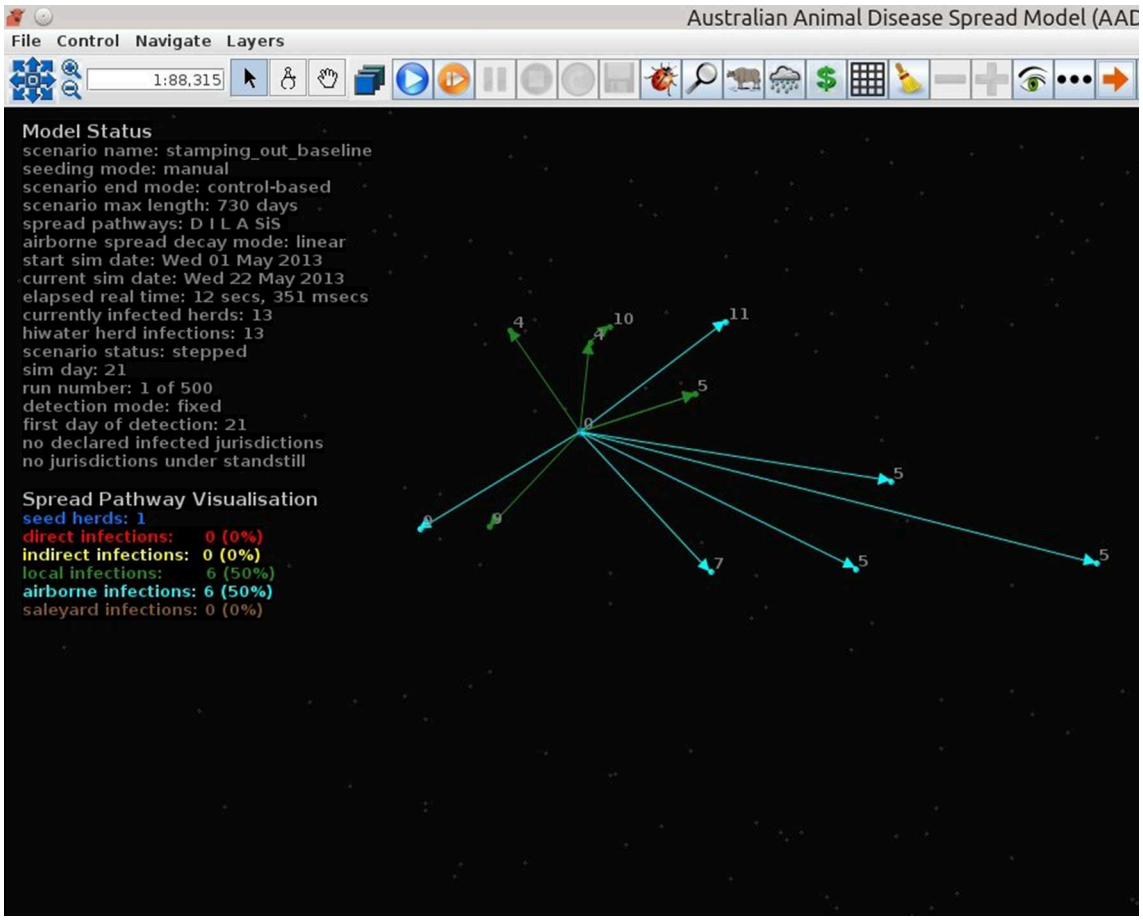


FIGURE 8 | Visualization of between-herd spread as an infection network.

individuals to the dynamics of an outbreak are thus lost in a population-based model. This is a limitation if the population and environment being modeled is heterogeneous. Complex environmental systems are typically multi-scale, non-linear and heterogeneous—characteristics that are ill-suited to an aggregated population-based modeling approach (Bansal et al., 2007; D'Souza et al., 2009; Parker and Epstein, 2011; Vincenot et al., 2011a). Although computationally efficient, an EBM can become complex and less tractable as more variables are factored into the mathematical abstraction (Miller, 1976; Parunak et al., 1998; Bobashev et al., 2007).

Advantages and Disadvantages of Individual-Based Modeling

Individual-based models are better suited to complex environmental systems due to their natural affinity for capturing heterogeneity, stochasticity, spatial relationships, social systems and policy (Hare and Deadman, 2004). The ability to distinguish between individuals in a population is especially important during the initial and final stages of an outbreak (Germann et al., 2006; Bansal et al., 2007). A data-driven, individual-based modeling approach has proven popular in the field of veterinary epidemiology with stochastic, spatially-explicit, state-transition models such as AusSpread, ISP and NAADSM. Individual-based models tend to be complex with a large number of parameters for which data may not always be available. Individual-based models may not scale well for large populations. Consider individual-based models of human pandemics in populations of millions or even billions. Such models have considerable computational requirements and typically require highly parallel platforms such as HPC clusters (Carley et al., 2006; Germann et al., 2006) or general purpose computing on graphics processing units (D'Souza et al., 2009) and custom software implementations (Parker and Epstein, 2011).

Advantages of Hybrid Models

Hybrid epidemiological models incorporate a population-based approach and an individual-based approach into a single model. Epidemics across a meta-population are multi-scale in the sense that the mechanisms and rates of within-site spread are distinct from those of between-site spread. In the case of a livestock epidemic, once infection is introduced into a farm the rate of within-farm propagation is dependent on the specifics of the pathogen and the farm. Factors include the host species, livestock

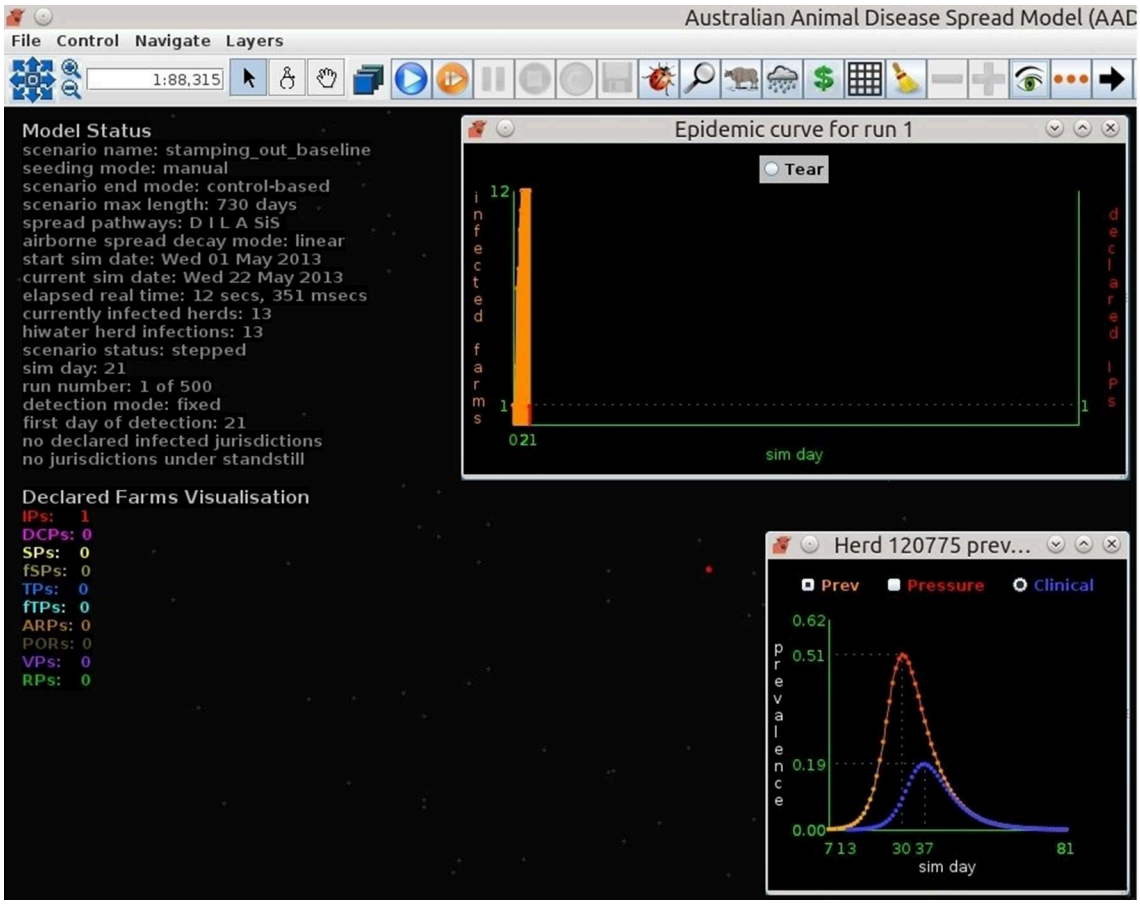


FIGURE 9 | Visualization of controlled premises.

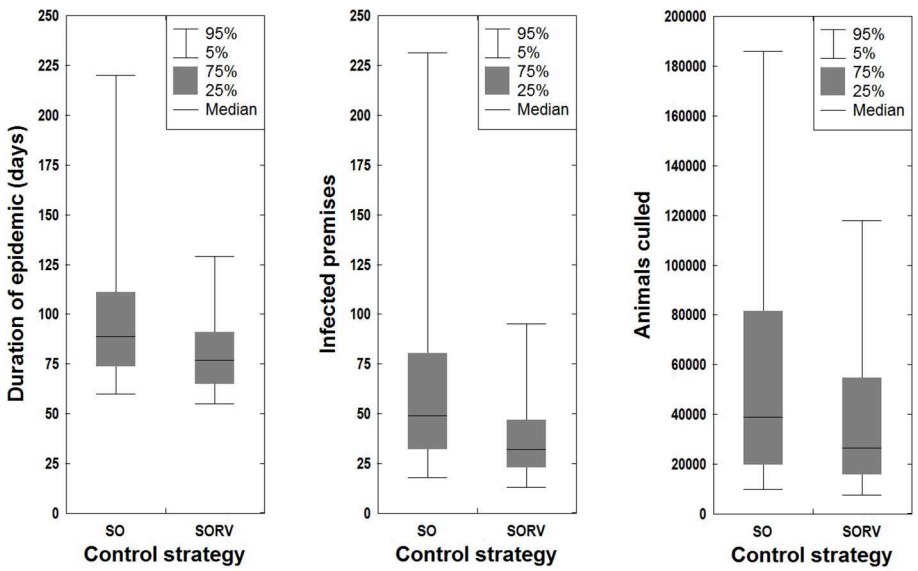


FIGURE 10 | Effect of control strategy on outbreak duration, outbreak size, and number of culled animals.

TABLE 4 | Case study results.

Control strategy	Detection day	Standstill duration (days)	Outbreak duration (days)[3,4,5]	Number of IPs[4,5]	Number of culled animals[4,5]	Number of vaccinated farms[4]	Number of vaccinated animals[4]	Scenario runtime (s)[4]
SO[1]	21	3	89 (60–220)[a]	49 (18–231)[a]	38, 875 (9838–185, 996)[a]	0	0	5.5 (3.3–16.1)
SORV[2]	21	3	77 (55–129)[b]	32 (13–95)[b]	26, 388 (7688–118, 036)[b,c]	192 (70–561)	51, 102 (15, 442–153, 972)	4.8 (3.1–9.6)
SENSITIVITY ANALYSIS OF THE TIME TO DETECTION								
SO	7	3	73 (51–107)[b]	20 (9–48)[b]	13, 385 (7201–61, 081)[b]	0	0	3.0 (2.0–4.8)
SO	14	3	78 (57–146)[b]	32 (14–98)[b]	19, 323 (7448–92, 393)[b]	0	0	4.3 (2.9–8.6)
SO[1]	21	3	89 (60–220)[a]	49 (18–231)[a]	38, 875 (9838–185, 996)[a]	0	0	5.5 (3.3–16.1)
SO	28	3	103 (63–380)[c]	78 (23–732)[c]	72, 275 (13, 314–738, 018)[d]	0	0	7.5 (4.0–46.6)
SENSITIVITY ANALYSIS OF THE DURATION OF THE NATIONAL LIVESTOCK STANDSTILL								
SO	21	0	90 (61–215)[a]	53 (19–241)[a]	45, 683 (10, 163–208, 485)[a]	0	0	5.6 (3.3–17.0)
SO[1]	21	3	89 (60–220)[a]	49 (18–231)[a]	38, 875 (9838–185, 996)[a]	0	0	5.5 (3.3–16.1)
SO	21	7	86 (59–184)[a]	48 (17–184)[a]	37, 111 (9919–165, 532)[a]	0	0	6.1 (3.9–15.1)
SO	21	10	85 (60–189)[a]	46 (19–176)[a]	39, 130 (10, 754–148, 028)[a]	0	0	5.2 (3.4–13.7)

[1] *Baseline stamping out policy.*
[2] *Baseline vaccination policy (stamping out plus suppressive ring vaccination).*
[3] *Time from detection of index case to resolution of final IP.*
[4] *Median (90% confidence interval).*
[5] *Within each column, values with a different superscript are significantly different.*

density, livestock numbers, production system, and biosecurity measures. The spread of disease between farms is influenced by more irregular factors such as contact networks between farms (direct and indirect), market practices, distance between farms and environmental conditions (including weather).

Bobashev et al. (2007) describe a stage-based hybrid model of global human influenza that dynamically switches between an ABM and an EBM based on the number of cases. Within-city spread is initially simulated by an ABM in order to capture subtle interactions between individuals early in an epidemic. When a cases threshold is reached, the ABM is halted and a snapshot of agent states is used as initial conditions for an EBM. Although the granularity of modeling decreases to population-level, it occurs at a point in the outbreak when the number of cases is sufficient to support a population-averaged approach. Moreover, the overall performance of the model is maintained due to the computationally efficient EBM. When the number of cases in a city falls below a threshold value, the model switches back to an ABM in order to capture subtle interactions between individuals as the epidemic wanes.

Network-based hybrid models employ a multi-scale approach to modeling the spread of disease across a meta-population. A population-based model handles the spread of disease within each meta-population site while an individual-based model handles the spread of disease between sites. An example is provided by Vincenot and Moriya (2011b) where a system dynamics-based EBM is used for within-site spread and a contact network is used for between-site spread. A compartment-based EBM is a good match for a closed homogeneous site while a data-driven spatially-explicit individual-based model captures heterogeneity in the epidemic environment. The edges of a contact network topology are formed from the potential conduits of disease across the meta-population. Network-based hybrid models have proven tractable in the study of human pandemics—driven by local and international mobility patterns derived from such sources as census data, surveys and the International Air Transport Association (IATA) database (Bansal et al., 2007; Balcan et al., 2009; Yu et al., 2010; Parker and Epstein, 2011; Van den Broeck et al., 2011; Yoneyama et al., 2012). Hybrid models used in the field of veterinary epidemiology include:

- The Davis Animal Disease Simulation Model (DADS) (Bates et al., 2003a) and DTU-DADS (Boklund et al., 2013) where within-herd spread of FMD is modeled with a Reed-Frost EBM (Fine, 1977), and between-herd spread modeled with a stochastic spatially-explicit contact network.
- The Netherlands FMD model (Backer et al., 2012) where within-herd spread is handled by an SEIR-based EBM and between-herd spread is modeled with a spatial kernel driven by probabilities derived from the 2001 outbreak in the Netherlands.
- Nickbakhsh et al. (2013) where within-flock spread of highly pathogenic avian influenza (HPAI) is modeled with a SEIR-based EBM and between-flock spread with a stochastic contact network.
- LaBute et al. (2014) where within-herd spread of FMD and HPAI are modeled with a compartmental SIR EBM and between-herd spread with a spatially-explicit contact network.

How AADIS Differs from other Hybrid Models

AADIS extends the network-based hybrid approach by employing an event-driven ABM in lieu of a contact network. The meta-population under study is heterogeneous, reflecting the multiple species of domestic cloven-hoofed animals that are susceptible to FMD. The network by which meta-population sites (i.e., herds) can 'connect' is multi-layer, reflecting how FMD

spreads via direct contact (animal movements between herds and between herds and markets), indirect contact (livestock products, by-products and fomites) and aerosols.

The AADIS ABM also models the control and eradication of FMD. Each disease spread pathway and control measure operates as an autonomous concurrent 'component' of the ABM environment. The decoupled component approach is robust, flexible and extensible. Components can be added/removed/modified with minimal impact on other components.

How AADIS Differs from Other Major Models of FMD Spread and Control

The AADIS EBM predicts a herd's prevalence and clinical signs over time based on the pathogen, herd type and herd size. These values dynamically feed into ABM decisions on the spread of disease between herds, the probability of detection, and the control of disease. In its role as an agent in the ABM, a herd reacts to environmental events such as culling and vaccination by resolving the EBM ODE system which in turn yields updated predictions for prevalence and clinical signs (**Figure 3**). The decoupling of within-herd spread and between-herd spread reflects the multi-scale nature of livestock epidemics (Carpenter et al., 2003; Keeling, 2005). Stochastic state-transition microsimulations such as AusSpread, ISP and NAADSM simplify intra-farm transmission as transitions through atomic infection states according to durations sampled from probability distributions. A state-transition approach to within-herd spread doesn't naturally capture the dynamics of intra-herd transmission. A simple herd state of 'infectious' doesn't distinguish between the infectiousness of a herd with 1% of the animals infected and that of a herd with 100% of the animals infected. This leads to a loss of prevalence information that is relevant to between-herd spread, and a loss of information on clinical signs that influences the detection and control of disease (Carpenter et al., 2003). It is possible to augment infection states with transmission probabilities that vary over time (Stevenson et al., 2013). However, this is less intuitive than the AADIS organic EBM approach. An architectural advantage of decoupling within-herd spread and between-herd spread is that alternative EBMs can be readily employed as required for the specific pathogen under study. This is awkward to accomplish when intra-herd spread and inter-herd spread are tightly coupled in a pure individual-based model such as a state-transition microsimulation.

Other distinguishing functional features of AADIS include:

- The configuration and deployment of control measures are decentralized to the separate state/territory jurisdictions. This permits realistic modeling of an epidemic that may spread across borders and require control at the jurisdictional level.
- The resources available for disease control and eradication are configurable by jurisdiction. This improves model realism as resource levels and priorities may vary considerably between jurisdictions. The AADIS ABM also allows resource requirements to emerge from a scenario as opposed to a top-down modeling approach that prescribes resourcing levels ahead of time. The inclusion of false positive suspect premises reports and traces provides more realistic modeling of surveillance as it reflects how resources are consumed regardless of the result of a surveillance visit.
- AADIS provides detailed graphical visualization modes that allow a user to dynamically view an outbreak unfolding in 'real' time. The graphical user interface allows a user to interact with an epidemic, for example to pause a scenario and view details of any herd/farm in the population. It is also possible to manually adjust the declared state of any farm. AADIS has potential as not only a predictive tool that informs policy, but also as a vivid training tool for disease managers.
- The multi-threaded asynchronous AADIS architecture offers significant performance improvements over a single-threaded state-transition approach. As all AADIS spread and control tasks proceed concurrently the length of a simulation day is only limited by the longest individual task. Computational efficiency is an important consideration for a stochastic model of national-scale as complex scenarios are re-run hundreds if not thousands of times to allow trends to emerge.
- Most of the current microsimulation models use a farm as the epidemiological unit of interest. AADIS's use of the herd captures heterogeneity in the spread of disease involving farms with co-located but separately managed herds, for example mixed beef/sheep farms.
- The national set of herds can be viewed abstractly as nodes in a network (Dube et al., 2011; Noremark et al., 2011). A network topology forms over time when spread pathways trigger and create edges (**Figures 2**, **8**). The topology takes the form of a directed acyclic graph, until such time as recovered herds lose their immunity. Network paths can subsequently be traversed forward to determine the downstream impact of an infected herd, and backward to trace the historical infection route. The network topology thus captures the spatiotemporal history of the simulated epidemic. The infection network can be mathematically analyzed to identify topological features of interest such as sinks and spreaders.

It should be noted that specific functional advantages of one model over another can be short-lived. Models such as AusSpread, ISP, NAADSM, DADS, DTU-DADS, and the Netherlands model (Backer et al., 2012) are active and continue to evolve. The principle innovation of AADIS is perhaps architectural, i.e., the movement away from the state-transition microsimulation approach of AusSpread, ISP and NAADSM to a hybrid EBM/ABM model. Network-based hybrid models tend to have single species meta-population and single layer contact network. AADIS expands this genre of models to a multi-species meta-population and a multi-layer contact network.

Limitations

The realism of data-driven models of disease spread hinges on the quality of the underlying data. This includes population data, contact structures, environmental data and pathogen data. Inadequate data can be replaced with assumptions/expert opinion but this has the potential to introduce bias into a model. In countries such as Australia where agriculture is of great importance to the economy, there is increasing availability

of spatially-based data on livestock and livestock products. An example is the National Livestock Identification System (NLIS) which tracks livestock from property of birth to place of slaughter/export (Meat and Livestock Australia, 2011). The NLIS database is a rich source of livestock movement data and takes into account species, production system and region. For AADIS to be used in a jurisdiction with a paucity of data, the spread and control components would need to be simplified. For example, a complex spread pathway based on animal movement data could be replaced with a simple distance kernel-based spread module.

AADIS has extensive configuration data spread across 40 tables in a relational database and three ASCII configuration files. This allows for detailed configuration of a heterogeneous environment and population, and pathogen under study. A result of this complexity is that the parameterization of the model requires a good understanding of the epidemiological system being modeled.

An artifact of the concurrent architecture adopted by AADIS is that thread scheduling arbitrarily influences the order that components request random numbers. This means that it is not possible to replay scenarios by specifying the pseudo-random number generator seed (and thus control the stream of random numbers used to sample from probability distributions). The ability to control the random number stream makes a stochastic model temporarily deterministic, and allows specific aspects of a scenario to be isolated. For example, a specific control measure such as vaccination can be varied and the impact on the scenario outcome directly observed (in the absence of variability introduced through stochasticity). The implication of this for AADIS is that a greater number of scenario runs may be required before results converge.

Concluding Remarks

Disease managers have to take into account technical, socio-political, economic and logistical issues when developing policies for disease control. Often there are conflicting objectives to balance, for example, to eradicate the disease as soon as possible and regain export markets, while minimizing the costs of control and compensation, and reducing impacts on other industries. Epidemiological modeling is emerging as an important contributor to the complex task of policy development.

Population-based models represent the spread of disease in a closed homogeneous population in a concise and computationally efficient manner. Individual-based models have a natural affinity for incorporating stochasticity, population heterogeneity, spatial effects, social factors and jurisdictional differences. This flexibility and realism has led to a strong interest in microsimulations for the purposes of informing official policy on disease control. Hybrid models have the modeling advantages of an individual-based approach but are also computationally efficient, which is particularly important when dealing with large livestock populations. The AADIS assumption that a herd is homogeneous is reasonable given that livestock are typically managed as single species cohorts that share a single contact network whilst on a farm. The AADIS SEIR-based EBM provides computationally efficient and adaptive predictions of herd prevalence and clinical signs over time. The AADIS ABM is well-suited to the complex, stochastic and heterogeneous environment in which an FMD epidemic operates.

There is an increasing availability of livestock movement and marketing data (including spatially-referenced data), through livestock identification, and tracing systems. This allows data-driven disease models such as AADIS to realistically simulate production system dynamics and contact structures.

The AADIS asynchronous hybrid EBM/ABM architecture has thus far shown itself as a flexible, efficient and extensible framework for modeling the spread and control of FMD in livestock on a national scale.

Acknowledgments

AADIS is a joint research venture between the Australian Department of Agriculture and the University of New England (UNE). The authors acknowledge both organizations for their support of the project. The authors would like to thank Dr. A.S.M. Sajeev who was a strong supporter of the project whilst Professor of Computer Science at UNE. The authors would also like to thank Dr. Rachel Iglesias of the Department of Agriculture for providing helpful comments on the manuscript. This work is funded under the Australian Government's Animal Biosecurity Response and Reform Program.

References

Animal Health Australia. (2014a). *Disease Strategy: Foot-and-Mouth Disease (Version 3.4). Australian Veterinarian Emergency Plan (AUSVETPLAN), 3rd Edn.* Canberra, ACT: Standing Council on Primary Industries.

Animal Health Australia. (2014b). *Animal Health in Australia 2013*. Canberra, ACT: Animal Health Australia.

Australian Bureau of Statistics. (2012). Agricultural Commodities, Australia. Available online at: http://www.abs.gov.au/ausstats/abs@.nsf/mf/7121.0

AusVet Animal Health Services. (2005). *A Review of the Structure and Dynamics of the Australian Dairy Cattle Industry*. Report prepared for the Department of Agriculture, Fisheries and Forestry, Canberra, ACT, Australia.

AusVet Animal Health Services. (2006). *A Review of the Structure and Dynamics of the Australian Beef Cattle Industry*. Report prepared for the Department of Agriculture, Fisheries and Forestry, Canberra, ACT, Australia.

Backer, J. A., Hagenaars, T. J., Nodelijk, G., and van Roermund, H. J. W. (2012). Vaccination against foot-and-mouth disease I: epidemiological consequences. *Prev. Vet. Med.* 107, 27–40. doi: 10.1016/j.prevetmed.2012.05.012

Balcan, D., Colizza, V., Goncalves, B., Hu, J., Ramasco, J. J., and Vespignani, A. (2009). Multiscale mobility networks and the spatial spreading of infectious diseases. *Proc. Natl. Acad. Sci. U.S.A.* 106, 21484–21489. doi: 10.1073/pnas.0906910106

Bansal, S., Grenfell, B. T., and Ancel Meyers, L. (2007). When individual behaviour matters: homogeneous and network models in epidemiology. *J. R. Soc. Interface* 4, 879–891. doi: 10.1098/rsif.2007.1100

Bates, T. W., Thurmond, M. C., and Carpenter, T. E. (2001). Direct and indirect contact rates among beef, dairy, goat, sheep and swine herds in three California counties, with reference to control of potential foot-and-mouth disease transmission. *Am. J. Vet. Res.* 62, 1121–1129. doi: 10.2460/ajvr.2001.62.1121

Bates, T. W., Thurmond, M. C., and And Carpenter, T. E. (2003a). Description of an epidemic simulation model for use in evaluating strategies to control an outbreak of foot-and-mouth disease. *Am. J. Vet. Res.* 64, 195–204. doi: 10.2460/ajvr.2003.64.195

Bates, T. W., Thurmond, M. C., and And Carpenter, T. E. (2003b). Results of epidemic simulation modeling to evaluate strategies to control an outbreak of foot-and-mouth disease. *Am. J. Vet. Res.* 64, 205–210. doi: 10.2460/ajvr.2003.64.205

BBN. (2014). *OpenMap*. BBN Technologies. Available online at: http://openmap.bbn.com/

Bobashev, G. V., Goedecke, D. M., Yu, F., and Epstein, J. M. (2007). "A hybrid epidemic model: combining the advantages of agent-based and equation-based approaches," in *Proceedings of the 2007 Winter Simulation Conference* (Washington, DC), 1532–1537.

Boklund, A., Halasa, T., Christiansen, L. E., and And Enoe, C. (2013). Comparing control strategies against foot-and-mouth disease: will vaccination be cost-effective in Denmark? *Prevent. Vet. Med.* 111, 206–219. doi: 10.1016/j.prevetmed.2013.05.008

Bradhurst, R. A., Roche, S. E., Garner, M. G., Sajeev, A. S. M., and Kwan, P. (2013). "Modelling the spread of livestock disease on a national scale: the case for a hybrid approach," in *Proceedings of the 20th International Congress on Modelling and Simulation (MODSIM2013), Modelling and Simulation Society of Australia and New Zealand* (Adelaide, SA), 345–351.

Buetre, B., Wicks, S., Kruger, H., Millist, N., Yainshet, A., Garner, G., et al. (2013). *Potential Socio-Economic Impacts of an Outbreak of Foot-and-Mouth Disease in Australia*. Research Report 13.11, October 2013, Australian Bureau of Agricultural and Resource Economics and Sciences (ABARES).

Bunn, C. M., Garner, M. G., and Cannon, R. M. (1998). The 1872 outbreak of foot-and-mouth disease in Australia – why didn't it become established? *Aust. Vet. J.* 76, 262–269. doi: 10.1111/j.1751-0813.1998.tb10157.x

Caraco, T., Duryea, M. C., and Glavavakov, S. (2001). Host spatial heterogeneity and the spread of vector-borne infection. *Theor. Popul. Biol.* 59, 185–206. doi: 10.1006/tpbi.2000.1517

Carley, K. M., Fridsma, D. B., Casman, E., Yahja, A., Altman, N., Chen, L., et al. (2006). BioWar: Scalable agent-based model of bioattacks. *IEEE Trans. Syst. Man Cybern. A Syst. Hum.* 36, 252–265. doi: 10.1109/TSMCA.2005.851291

Carpenter, T. E., Bates, T. W., and Thurmond, M. C. (2003). "The importance of including intra-herd transmission in a foot-and-mouth disease model," in *Proceedings of the 10th Symposium of the International Society for Veterinary Epidemiology and Economics, Vina del Mar, Chile, November 2003, New Concepts – Modelling Session*," 249–251.

Carpenter, T. E., O'Brien, J. M., Hagerman, A. D., and McCarl, B. A. (2011). Epidemic and economic impacts of delayed detection of foot-and-mouth disease: a case study of a simulated outbreak in California. *J. Vet. Diagn. Invest.* 23, 26–33. doi: 10.1177/104063871102300104

Cash, J. R., and Karp, A. R. (1990). A variable order Runge-Kutta method for initial value problems with rapidly varying right-hand sides. *ACM Trans. Math. Softw.* 16, 201–222. doi: 10.1145/79505.79507

Department of Environment and Primary Industries. (2015). *Goulburn Valley Regional Overview*. State Government of Victoria. Available online at: http://www.depi.vic.gov.au/agriculture-and-food/food-and-fibre-industries/region-overviews/goulburn

Donaldson, A. I., Alexandersen, A., Sorensen, J. H., and Mikkelsen, T. (2001). Relative risks of the uncontrollable (airborne) spread of FMD by different species. *Vet. Rec.* 148, 602–604. doi: 10.1136/vr.148.19.602

Donaldson, A. I., and Alexandersen, A. (2002). Predicting the spread of foot and mouth disease by airborne spread. *Rev. Sci. Tech.* 21, 569–575.

D'Souza, R. M., Marino, S., and Kirschner, D. (2009). "Data-parallel algorithms for agent-based model simulation of tuberculosis on graphics processing units," in *Proceedings of the 2009 Spring Simulation Multiconference (SpringSim '09), Society for Computer Simulation International* (San Diego, CA), 1–12.

Dube, C., Ribble, C., Kelton, D., and McNab, B. (2011). Introduction to network analysis and its implications for animal disease modelling. *Rev. Sci. Tech.* 30, 425–436.

East, I. J., and Foreman, I. (2011). The structure, dynamics and movement patterns of the Australian sheep industry. *Aust. Vet. J.* 89, 477–489. doi: 10.1111/j.1751-0813.2011.00852.x

East, I. J., Wicks, R. M., Martin, P. A. J., Sergeant, E. S. G., Randall, L. A., and Garner, M. G. (2013). Use of a multi-criteria analysis framework to inform the design of risk based general surveillance systems for animal disease in Australia. *Prev. Vet. Med.* 112, 230–247. doi: 10.1016/j.prevetmed.2013.09.012

East, I. J., Davis, J., Sergeant, E. S. G., and Garner, M. G. (2014). Structure, dynamics and movement patterns of the Australian pig industry. *Aust. Vet. J.* 92, 52–57. doi: 10.1111/avj.12141

Fine, P. E. M. (1977). A commentary on the mechanical analogue to the Reed-Frost epidemic model. *Am. J. Epidemiol.* 106, 87–100.

Galvani, A. P., and May, R. M. (2005). Dimensions of superspreading. *Nature* 438, 293–294. doi: 10.1038/438293a

Garner, M. G., and Beckett, S. D. (2005). Modelling the spread of foot-and-mouth disease in Australia. *Aust. Vet. J.* 83, 758–766. doi: 10.1111/j.1751-0813.2005.tb11589.x

Garner, M. G., Dube, C., Stevenson, M. A., Sanson, R. L., Estrada, C., and Griffen, J. (2007). Evaluating alternative approaches to managing animal disease outbreaks – the role of modelling in policy formulation. *Vet. Ital.* 43, 285–298.

Garner, M. G., and Hamilton, S. A. (2011). Principles of epidemiological modelling. *Rev. Sci. Tech.* 30, 407–416.

Garner, M. G., Hess, G. D., and Yang, X. (2006). An integrated modelling approach to assess the risk of wind-borne spread of foot-and-mouth disease virus from infected premises. *Environ. Model. Assess.* 11, 195–207. doi: 10.1007/s10666-005-9023-5

Garner, M. G., Murray, J. G., and Fisher, B. S. (2002). Economic aspects of foot and mouth disease: perspectives of a free country, Australia. *Rev. Off. Int. Epizoot.* 21, 625–635. doi: 10.1016/j.prevetmed.2013.07.013

Germann, T. C., Kadau, K., Longini, I. M., and And Macken, C. A. (2006). Mitigation strategies for pandemic influenza in the United States. *Proc. Natl. Acad. Sci. U.S.A.* 103, 5935–5940. doi: 10.1073/pnas.0601266103

Gibbens, J. C., Sharpe, C. E., Wilesmith, J. W., Mansley, L. M., Michalopoulou, E., Ryan, J. B., et al. (2001). Descriptive epidemiology of the 2001 foot-and-mouth disease epidemic in Great Britain: the first five months. *Vet. Rec.* 149, 729–743.

Hagenaars, T. J., Donnelly, C. A., and Ferguson, N. M. (2004). Spatial heterogeneity and the persistence of infectious diseases. *J. Theor. Biol.* 229, 349–359. doi: 10.1016/j.jtbi.2004.04.002

Hare, M., and Deadman, P. (2004). Further towards a taxonomy of agent-based simulation models in environmental management. *Math. Comput. Simul.* 64, 25–40. doi: 10.1016/S0378-4754(03)00118-6

Harvey, N., Reeves, A., Schoenbaum, M. A., Zagmutt-Vergara, F. J., Dubé, C., Hill, A. E., et al. (2007). The North American Animal Disease Spread Model: a simulation model to assist decision making in evaluating animal disease incursions, *Prevent. Vet. Med.* 82, 176–97. doi: 10.1016/j.prevetmed.2007.05.019

Hassall and Associates. (2006). *The Structure and Dynamics of Australia's Sheep Population*. Report prepared for the Department of Agriculture, Fisheries and Forestry, Canberra, ACT, Australia.

Hassall and Associates. (2007). *A Review and Analysis of Saleyard Marketing in Australia*. Report prepared for the Department of Agriculture, Fisheries and Forestry, Canberra, ACT, Australia.

Haydon, D. T., Woolhouse, M. E., and Kitching, R. P. (1997). An analysis of foot-and-mouth-disease epidemics in the UK. *IMAJ. Math. Appl. Med. Biol.* 14, 1–9. doi: 10.1093/imammb/14.1.1

James, A., Pitchford, J. W., and Plank, M. J. (2007). An event-based model of superspreading in epidemics. *Proc. R. Soc. B Biol. Sci.* 274, 741–747. doi: 10.1098/rspb.2006.0219

Keeling, M. J. (2005). Models of foot-and-mouth disease. *Proc. R. Soc. B Biol. Sci.* 272, 1195–1202. doi: 10.1098/rspb.2004.3046

Keeling, M. J., and Rohani, P. (2008). *Modeling Infectious Diseases in Humans and Animals*. Princeton; Oxford. Princeton University Press.

Kokic, P., and Mues, C. (2006). *Cattle Movements*. Research Report. Australian Bureau of Agricultural and Resource Economics and Sciences (ABARES).

LaBute, M. X., McMahon, B. H., Brown, M., Manore, C., and Fair, J. M. (2014). A flexible spatial framework for modeling spread of pathogens in animals with biosurveillance and disease control applications. *ISPRS Int. J. Geo-Inform.* 3, 638–661. doi: 10.3390/ijgi3020638

Lloyd-Smith, J. O., Schreiber, S. J., Hopp, P. E., and Getz, W. M. (2005). Super-spreading and the effect of individual variation on disease emergence. *Nature* 438, 355–359. doi: 10.1038/nature04153

Mansley, L. M., Dunlop, P. J., Whiteside, S. M., and Smith, R. G. H. (2003). Early dissemination of FMD virus through sheep marketing in February 2001. *Vet. Rec.* 153, 43–50. doi: 10.1136/vr.153.2.43

Matthews, K. (2011). *A Review of Australia's Preparedness for the Threat of Foot-and-Mouth Disease*. Australian Government Department of Agriculture, Fisheries and Forestry, Canberra, ACT. Available online at: http://www.agriculture.gov.au/animal-plant-health/pests-diseases-weeds/animal/fmd/review-foot-and-mouth-disease

McLaws, M., Ribble, C., Stephen, C., McNab, B., and Barrios, P. R. (2007). Reporting of suspect cases of foot-and-mouth disease during the 2001 epidemic in the UK, and the herd sensitivity and herd specificity of clinical diagnosis. *Prevent. Vet. Med.* 78, 12–23. doi: 10.1016/j.prevetmed.2006.09.001

Meat and Livestock Australia. (2011). *National Livestock Identification Scheme Database*. Available online at: http://www.mla.com.au/Meat-safety-and-traceability/National-Livestock-Identification-System

Meyer, R. F., and Knudsen, R. C. (2001). Foot-and-mouth disease: a review of the virus and the symptoms. *J. Environ. Health* 64, 21–23.

Miller, W. M. (1976). "A state-transition model of epidemic foot-and-mouth disease," in *Proceedings of the International Symposia on Veterinary Epidemiology and Economics (ISVEE 1), New Techniques in Veterinary Epidemiology and Economics*. Reading: University of Reading, Descriptive Modelling Session, 56–72.

Nickbakhsh, S., Matthews, L., Dent, J. E., Innocent, G. T., Arnold, M. E., Reid, S. W. J., et al. (2013). Implications of within-farm transmission for network dynamics: consequences for the spread of avian influenza. *Epidemics* 5, 67–76. doi: 10.1016/j.epidem.2013.03.001

Nielen, M., Jalvingh, A. W., Horst, H. S., Dijkhisen, A. A., Maurice, H., Schut, B. H., et al. (1996). Quantification of contacts between Dutch farms to assess the potential risk of foot-and-mouth disease spread. *Prevent. Vet. Med.* 28, 143–158. doi: 10.1016/0167-5877(96)01042-2

Noremark, M., Hakansson, N., Sternberg Lewerin, S., Lindberg, A., and Jonsson, A. (2011). Network analysis of cattle and pig movements in Sweden: measures relevant for disease control and risk based surveillance. *Prevent. Vet. Med.* 99, 78–90. doi: 10.1016/j.prevetmed.2010.12.009

Noremark, M., Frossling, J., and Sternberg Lewerin, S. (2013). A survey of visitors on Swedish livestock farms with reference to the spread of animal diseases. *BMC Vet. Res.* 9:184. doi: 10.1186/1746-6148-9-184

Oracle. (2014). *Java*. Oracle Corporation. Available online at: http://www.java.com/en/

Parker, J., and Epstein, J. M. (2011). A distributed platform for global-scale agent-based models of disease transmission. *ACM Trans. Model. Comput. Simul.* 22:2. doi: 10.1145/2043635.2043637

Parunak, H. V. D., Savit, R., and Riolo, R. L. (1998). "Agent-based modeling vs. equation-based modeling: a case study and users' guide," in *Proceedings of Multi-Agent Systems and Agent-Based Simulation* (MABS'98), 10–25.

Perez, L., and Dragicevic, S. (2009). An agent-based approach for modeling dynamics of contagious disease spread. *Int. J. Health Geogr.* 8:50. doi: 10.1186/1476-072X-8-50

PostgreSQL. (2014). *PostgreSQL*. The PostgreSQL Global Development Group. Available online at: http://www.postgresql.org/

Regional Development Victoria. (2010). Goulburn Valley sub-regional plan – The Hume Strategy for sustainable communities 2010-2010. State Government of Victoria. Available online at: http://www.rdv.vic.gov.au/victorian-regions/hume

Roche, S. E., Garner, M. G., Wicks, R. M., East, I. J., and de Witte, K. (2014). How do resources influence control measures during a simulated outbreak of foot and mouth disease in Australia? *Prevent. Vet. Med.* 113, 436–446. doi: 10.1016/j.prevetmed.2013.12.003

Rushton, J., Knight Jones, T. J. D., Donaldson, A. I., de Leeuw, P. W., Ferrari, G., and Domenech, J. (2012). *The Impact of Foot and Mouth Disease – Supporting Document N° 1*. Paper prepared for the FAO/OIE Global Conference on Foot and Mouth Disease Control, Bangkok (Thailand), 27 to 29 June 2012.

Sanson, R. L. (2005). A survey to investigate movements off sheep and cattle farms in New Zealand, with reference to the potential transmission of foot-and-mouth disease. *N. Z. Vet. J.* 53, 223–233. doi: 10.1080/00480169.2005.36550

Stata. (2014). *STATA Intercooled Software v11.0*. Stata Corporation. Available online at: http://www.stata.com/products/which-stata-is-right-for-me/#IC

Stevenson, M. A., Sanson, R. L., Stern, M. W., O'Leary, B. D., Sujau, M., Moles-Benfell, N., et al. (2013). InterSpread Plus: a spatial and stochastic simulation model of disease in animal populations. *Prevent. Vet. Med.* 109, 10–24. doi: 10.1016/j.prevetmed.2012.08.015

Taylor, N. (2003). *Review of the Use of Models in Informing Disease Control Policy Development and Adjustment*. School of Agriculture, Policy and Development, The University of Reading.

Tildesley, M. J., Savill, N. J., Shaw, D. J., Deardon, R., Brooks, S. P., Woolhouse, M. E. J., et al. (2006). Optimal reactive vaccination strategies for a foot-and-mouth outbreak in the UK. *Nature* 440, 83–86. doi: 10.1038/nature04324

Van den Broeck, W., Gioannini, C., Gonçalves, B., Quaggiotto, M., Colizza, V., and Vespignani, A. (2011). The GLEaMviz computational tool, a publicly available software to explore realistic epidemic spreading scenarios at the global scale. *BMC Infect. Dis.* 11:37. doi: 10.1186/1471-2334-11-37

Vincenot, C. E., Giannino, F., Rietkerk, M., Moriya, K., and Mazzoleni, S. (2011a). Theoretical considerations on the combined use of System Dynamics and individual-based modeling in ecology. *Ecol. Modell.* 222, 210–218. doi: 10.1016/j.ecolmodel.2010.09.029

Vincenot, C. E., and Moriya, K. (2011b). Impact of the topology of metapopulations on the resurgence of epidemics rendered by a new multiscale hybrid modeling approach. *Ecol. Inform.*, 6, 177–186. doi: 10.1016/j.ecoinf.2011.04.002

Vose, D. (2008). *Risk Analysis: A Quantitative Guide, 3rd Edn*. Chichester: John Wiley and Sons, Ltd.

Yoneyama, T., Das, S., and Krishnamoorthy, M. (2012). A hybrid model for disease spread and an application to SARS pandemic. *J. Artif. Soc. Soc. Simul.* 15, 5.

Yu, B., Wang, J., McGowan, M., Vaidyanathan, G., and Younger, K. (2010). "Gryphon: A hybrid agent-based modeling and simulation platform for infectious diseases. Advances in Social Computing," in *Third International Conference on Social Computing, Behavioural Modeling and Prediction, SBP 2010, Proceedings, LNCS 6007* (Bethesda, MD), 199–207.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Environmental Science

ORIGINAL RESEARCH
published: 13 October 2015
doi: 10.3389/fenvs.2015.00066

Combining system dynamics and agent-based modeling to analyze social-ecological interactions—an example from modeling restoration of a shallow lake

Romina Martin[1,2]* *and Maja Schlüter*[2]

[1] Ecosystem Research, Leibniz-Institute of Freshwater Ecology and Inland Fisheries, Berlin, Germany, [2] Stockholm Resilience Centre, Stockholm University, Stockholm, Sweden

OPEN ACCESS

Edited by:
Christian E. Vincenot,
Kyoto University, Japan

Reviewed by:
Silvie Daniels,
Hasselt University, Belgium
Birgit Kopainsky,
University of Bergen, Norway

***Correspondence:**
Romina Martin,
Stockholm Resilience Centre,
Stockholm University, Kräftriket 2B,
11419 Stockholm, Sweden
romina.martin@su.se

Specialty section:
This article was submitted to Environmental Informatics, a section of the journal Frontiers in Environmental Science

Received: 31 May 2015
Accepted: 22 September 2015
Published: 13 October 2015

Citation:
Martin R and Schlüter M (2015) Combining system dynamics and agent-based modeling to analyze social-ecological interactions—an example from modeling restoration of a shallow lake. Front. Environ. Sci. 3:66. doi: 10.3389/fenvs.2015.00066

Modeling social-ecological interactions between humans and ecosystems to analyze their implications for sustainable management of social-ecological systems (SES) has multiple challenges. When integrating social and ecological dynamics, which are often studied separately, one has to deal with different modeling paradigms, levels of analysis, temporal and spatial scales, and data availabilities in the social and ecological domains. A major challenge, for instance, is linking the emergent patterns from individual micro-level human decisions to system level processes such as reinforcing feedbacks determining the state of the ecosystem. We propose building a hybrid model that combines a system dynamics with an agent-based approach to address some of these challenges. In particular, we present a procedure for model development and analysis that successively builds up complexity and understanding of model dynamics, particular with respect to feedbacks between the social and ecological system components. The proposed steps allow for a systematic increase of the coupling between the submodels and building confidence in the model before deploying it to study the coupled dynamics. The procedure consists of steps for (i) specifying the characteristics of the link between the social and ecological systems, (ii) validating the decoupled submodels, (iii) doing sensitivity analysis of the decoupled submodels with respect to the drivers from the respective other subsystem and, finally (iv) analyzing the coupled model. We illustrate the procedure and discuss opportunities and limitations of hybrid models against the background of an archetypical SES case study, namely the restoration of a turbid lake. Our approach exemplifies how a hybrid model is used to unpack SES complexity and analyze interactions between ecological dynamics and micro-level human actions. We discuss the benefits and challenges of combining a system dynamics models as an aggregated view with an agent-based model as a disaggregated view to improve social-ecological system understanding.

Keywords: social-ecological system, human decision making, lake restoration, agent-based modeling, system dynamics

Introduction

Humans depend on many services provided by ecosystems such as food and clean water, nutrient retention, protection from flooding or recreational opportunities (MEA, 2005). At the same time human activities such as resource exploitation, construction or industrial production have altered ecosystems at and across all scales (e.g., Steffen et al., 2015). Approaches for understanding and sustainably managing ecosystems need to take these two way interactions between people and ecosystems into account (Berkes and Folke, 1998). Modeling, which is a widely used tool for ecosystem and natural resource management, thus has to go beyond studying humans as a simple driver of ecosystem dynamics or natural resources as a simple input to a production process by explicitly considering feedbacks between the ecological and social domains (Schlüter et al., 2012). Integrating human individual and collective behavior with complex ecological dynamics, however, comes with many challenges. These include amongst others the need to connect processes at different levels of aggregation used to represent the social and ecological systems, the different temporal and spatial scales, differences in data availability, as well as different modeling paradigms associated with different fields or levels of analysis. In this paper we present an example of a hybrid modeling approach that combines system dynamics with agent-based modeling to address some of these challenges. We present a procedure for the development and analysis of the hybrid model and discuss opportunities and limitations against the background of an archetypical social-ecological systems (SES) case study, namely the restoration of a turbid lake to a more desirable clear water state.

Most commonly, lake management is investigated as an external intervention in the biophysical properties of a lake (Carpenter, 2005; Pers, 2005; Kara et al., 2011), rather than as the result of the interplay between ecological dynamics and social processes. Important social processes are decision making about ecosystem management measures and human activities that affect future states of the ecosystem. To understand the multiple drivers for lake restoration, it is necessary to analyze the lake management system as a coupled SES, where humans are embedded in the ecosystems they affect and depend on (Folke et al., 2005). One approach to take these interdependencies into account is to identify key social-ecological interactions that determine human responses and ecosystem change (Ostrom, 1990). Taking this perspective, we account in this study explicitly for responses of managers that monitor the lake state and regulate human interactions with the lake as well as decision making by actors whose behavior directly affects ecological dynamics of the lake. In doing so we face the challenge of linking emergent patterns from individual, micro-level decisions of humans to system level ecological processes such as nonlinear, reinforcing feedbacks determining the lake state in order to model individual and collective human responses to ecological change and vice versa ecological change resulting from human action. Different approaches have commonly been used to model these different social and ecological processes.

The ecological dynamics of shallow lakes, particularly so called regime shifts that shift a lake from a clear to a turbid state, are commonly investigated as dynamical systems using analytical minimal models (Scheffer, 1989, 1990; Biggs et al., 2009) or simulation models (Carpenter et al., 1999; Mooij et al., 2009). Regime shifts are abrupt shifts in the state of a lake that persist over time. They are often caused by an external driver such as nutrient increases (eutrophication) that is amplified by ecological feedbacks (Scheffer, 1990). Dynamical systems can undergo a fold bifurcation which can be detected using bifurcation analyses to identify the range of driver values over which bistable behavior can be expected. Some systems show hysteretic behavior, which is when different stable states can be reached depending on the initial state. Existing minimal models of regime shifts in lakes show such hysteresis (Scheffer, 1990). The phenomenon can be explained by feedback loops that keep the lake system in either the clear or the turbid state. The moment that the driver crosses a threshold, the dominant reinforcing feedback switches and pushes the system into a different state. While these models are good at analyzing general system characteristics and important ecological feedbacks (e.g., Biggs et al., 2009) or aggregated effects from human decisions in land-use (e.g., Haase et al., 2012), they do not take into account potential human adaptive responses to changing ecological dynamics which may affect the regime shift or its reversal, e.g., through active restoration. They thus do not allow investigating particular social-ecological interactions such as the development of regulations as a macro-level response to ecological degradation or the lack thereof that may affect the dynamics of the whole social-ecological system.

The social processes related to the use and management of ecosystem services such as monitoring, regulating, and enforcing new rules or adopting new technologies have not yet been modeled much in the context of SES. Traditional ecological models rarely take into account human decision-making (but see as an exception Janssen, 2001; Carpenter and Brock, 2004), despite its importance for management outcomes. Agent-based models (ABM) as they have been used in social simulation (Balke and Gilbert, 2014) as well as in land use (Janssen et al., 2000; Le et al., 2008) or marine sciences explicitly describe human decision making and range on a continuum from rather simple interactions among agents to deliberative/cognitive representations (see for reviews Bousquet and Le Page, 2004; Hare and Deadman, 2004). In ABM's, processes of decision making involving multiple factors are often implemented using sets of conditional rules, but optimization routines that determine the optimal choice under given constraints are also common. Fewer ABM's address multiple, organizational levels of human interactions and how macro-level social behavior emerges from micro-level interaction.

In summary, building models for coupled social-ecological systems involve several challenges which comprise the explicit representation of feedbacks on multiple levels of aggregation and links between conceptually different subsystem models operating on different scales. One approach, that looks promising to us to address those challenges, is to develop and analyze a hybrid model in a stepwise approach to take advantage of different modeling paradigms. In this paper, we reflect on these

particular steps and illustrate a potentially generic approach for social-ecological model analysis with the example case of lake restoration. The purpose of the stepwise procedure and the reflection upon the model development is to allow for (i) making explicit design choices related to the selection of the modeling approach for each subsystem, and (ii) rigorous testing of submodels using specific tools of each approach. The single steps also help to successively build up complexity of the model and its analysis for (i) systematically investigating the effect of feedbacks between the social and ecological systems, (ii) exploring the coupled model to build understanding for full model analysis. For the hybrid model, we propose combining system dynamics with agent-based modeling to link micro-level human action with system-level ecological dynamics. Such an integrated hybrid model allows studying social-ecological interactions over time to understand their implications for the future development of the coupled SES and its management.

In this paper, we briefly review the two modeling approaches and suggest ways to combine them in practice. After introducing the case study, we present a stepwise approach to systematically develop and test a hybrid model. We illustrate our approach with a social-ecological model on the restoration of a turbid lake. The model has been developed to study the interactions between lake managers, private house owners, and a turbid lake to estimate the potential of improved sewage water treatment for lake restoration, where the aim is to better understand the critical time lags between social and ecological responses during the restoration process. We finally discuss the benefits as well as the conceptual and methodological challenges of combining an aggregated, system level with a disaggregated, agent-based modeling perspective.

Characteristics of Agent-Based and System Dynamics Models and Ways to Combine Them

This section provides a brief overview of characteristics of agent-based and system dynamics models that were particularly acknowledged in the past (**Table 1**). The purpose is to identify both advantages and challenges of those approaches. The presented model paradigms are iconic in the way that they are often presented as exclusive alternatives to analyze complex systems either from a top-down/aggregate or bottom-up/disaggregate perspective (Vincenot et al., 2011; Swinerd and McNaught, 2012). Finally, we discuss examples how they have been or could potentially be combined in a hybrid model.

The choice of a modeling paradigm is primarily determined by the research question and objectives. Agent-based modeling is a common approach to study complex adaptive systems, i.e., systems that are characterized by self-organization, emergence and adaptation (Levin, 1998). ABM approaches generally focus on micro-level interactions that may explain emergent patterns such as transient dynamics on a system level. By this, one can identify which mechanisms are important, taking into account heterogeneity of entities, spatial and temporal heterogeneity of processes, and stochasticity. The output can be time series of system-level output variables, spatial patterns or the statistical analysis of system patterns that can be validated against empirical data. Rather than reproducing micro-level processes realistically, the aim in agent-based modeling is usually to simulate the minimum of necessary functions that are able to reproduce a certain system level behavior or pattern as an emergent outcome.

While agent-based models are used to describe disaggregated parts of a system, system dynamics models represent the aggregated system in the form of stocks and flows. Here, the main focus lies on describing system-level behavior such as exponential growth through system -level structures such as feedbacks and non-linearities rather than letting the behavior emerge from micro-level interactions. Dynamics on the system-level are often represented by a set of differential equations and analyzed with the tools of dynamical systems theory. Here the aim is to describe the fixed points/stable states of the system (stability analysis) and identify how system structure drives system behavior (e.g., bifurcation analysis). System dynamics as a particular approach within dynamical systems applies causal-loop diagrams as a graphical approach to represent the main system-level interactions. This is usually done to investigate the nature of feedbacks (balancing, reinforcing) and estimate the overall dynamics of the system for example with stakeholders for policy analyses in environmental problem settings (van den Belt, 2004). It becomes necessary to implement the equations for the causal-loop diagram, however, to examine which feedbacks in the system are dominant and drive the development over time. Outputs are often stable states of the system, bifurcation diagrams or time-series describing system level behavior. Such simple equation-based models, so called minimal models, can still be studied analytically to identify system properties such as stability.

For more complex analyses of both, system dynamics and agent-based models, simulations with varied parameters and multiple repetitions (in case of stochastic models) are evaluated to generate average system-level outcomes. These outcomes can be used to formulate new or more specific hypotheses that can be tested for instance with other model simulations.

If a research question targets specifically the link between two parts of the system under study, it may be useful to define subsystem models and their interactions separately. Linking different subsystem models requires, firstly, specifying the particular variables, the aggregation level, and time scales that define the interface between subsystems. While the modeling paradigm suggests that agent-based models are rather suitable for representing the micro-level and system dynamics the system level processes, it does not mean that this order determines the way that they are connected (Swinerd and McNaught, 2012). There exist multiple ways to combine different types of simulation models ranging from subsequently coupled (such as in Gaube et al., 2009) to dynamically coupled (Shanthikumar and Sargent, 1983; Swinerd and McNaught, 2012). Often, a system dynamics representation is chosen to simulate internal processes of an entity while the interaction among entities is mapped via an agent-based approach (Bradhurst et al., 2015). In other cases, processes on the system level affect individual entities top-down (downward-causation) and in turn interacting individuals

TABLE 1 | Characteristics of agent-based and system dynamics models, and how they can be combined within a fully integrated hybrid model.

	Agent-based model	System dynamics model	Hybrid model (proposed)
Characteristic question	How do emergent system-level patterns develop from micro-level interaction (e.g., spatially, between individuals)?	- How do stocks change or stabilize? (given that rates are constant) - Which process/feedback is dominating?	- How do changing process rates (impacted by decisions) affect dynamics? - How do changing stocks affect agent states/the distribution of traits?
Purposes			
In general for all: improve system understanding rather than prediction or forecasting (Kelly(Letcher) et al., 2013).	- To identify mechanisms (specific interactions) that are responsible for emerging system-level patterns (disaggregated). - Generate hypotheses, exploration of micro-level behavior (Epstein, 2008)	Investigate system-level dynamics (aggregated), stability properties of the system, loop dominance, explaining temporal dynamics, projection into the future	Investigating different micro- or system level mechanisms that drive certain dynamics. Generate hypotheses of systems state-change (when does dominance of feedbacks change?) or structural development over time (when does an average trait of agents change?)
Focus	Micro-level interactions between entities, network structure (heterogeneous characteristics of individuals/actors, temporal discrete behavior), transient dynamics	Processes driving accumulation in stocks at (sub-)system level, stable-states, feedbacks (balancing, amplifying), non-linearities	Process of restructuring in a system which can focus either on a structure affecting the processes, or processes affecting the structure
Tests for model calibration	Statistical pattern matching—can the model grow patterns that are found in reality?	Stability analysis—under which parameter setting can fixed points/equilibria occur? How stable are they?	Separate sub-system tests (paradigm specific) and qualitative check for the coupled version
Suitable and traditional analysis tools, typical experiments	Only through simulations, often with multiple repetitions because of stochastic elements: plotting group/system level characteristics over time (average), evaluating a limited parameter range, describing transient dynamics	Simple models through analytical tools (basins of attraction, bifurcation analysis, overall stability), and more complex through simulations (state space plots from simulations, evaluating stable-states, equilibria)	Through simulations with a focus on either 1. Change in structure/parameters: how does it affect the dynamics? 2. Change in dynamics: how does it affect the structure?
Type of outcome	Emerging spatial/agent patterns, scenario comparison between structurally different model versions, system properties such as the average state of a population	Aggregated system properties in terms of stability, loop dominance	Time series of emerging state-transitions

have emergent properties that play out on the system level (e.g., Haase et al., 2012). We present a hybrid model herein (Section A Stepwise Approach to Develop, Test and Analyse a Hybrid Social-ecological Model) where the two subsystems can be presented side-by-side with a dynamic feedback between them.

Case Introduction: Lakes Analyzed as Social-Ecological System

Regime shifts in shallow lakes from the clear to the turbid state are typically driven by an increase of nutrient inputs into the lake from streams, channels and groundwater. The affluent availability of nutrients, in particular phosphorus, enhances the growth of planktonic algae, which reduces light penetration and thus causes submerged plants to disappear that dominate in clear waters. The density of algae also depends on the presence of its predator zooplankton and their predator whitefish such as bream (*Abramis brama L.*). A typical turbid lake is characterized by high densities of planktonic algae, low zooplankton biomass and diversity, high abundance of whitefish, a low density of piscivorous fish (such as pike—*Esox Lucius L.*), and reduced amounts of macrophytes (Ekvall et al., 2014). Conversely, a clear lake contains limited algae, a rich and diverse zooplankton community, few whitefish, a higher density of pike and macrophytes. A range of minimal models exist that are able to reproduce lake regime shifts that originate in different parts of a lake's foodweb (Scheffer, 1990). While the causes for those regime shifts are often diffuse, accumulated over long time, the consequences appear relatively abrupt, are clearly visible and affect multiple ecosystem services (Biggs et al., 2012b).

For the case of lakes, multiple actors potentially contribute to eutrophication, namely farmers using fertilizer for agriculture, municipal sewage treatment plants, and private house owners who are not connected to municipal sewage treatment. Through the water framework directive in the EU, particular rules for fertilizer use in agriculture and sufficient purification procedures for sewage water were formulated which reduced inflow from agriculture and commercial sewage treatment plants. However, single house owners were not much addressed before as their contribution to nutrients in water catchments was not seen as significant as today. Recent evidence, however, shows that in

some regions such as the Baltic Sea untreated sewage water from private households (mainly holiday homes) can be a major polluting source (Kininmonth et al., 2015). The decision is made by individual house owners about whether to upgrade their private sewage system to withhold more of the nutrients. This decision can be influenced by cost-benefit considerations, the presence of sanctions such as fines, or social considerations, such as social norms or social pressure within a social network. Individual house owners, for example, are obliged to invest into a private, legal sewage treatment system, but the benefits therefrom are communally shared, probably lag in time and are difficult to see which incites to delay the private investment. The way how house owners are informed and connected among them plays a crucial role in the catchment-wide adoption of sewage treatment upgrades (Wallin, 2012). To drive a desired regime shift in the lake from the turbid to the clear state, multiple stakeholders (municipalities, farmers and responsible persons for sewage water treatment) are required to take effective and coordinated action.

A Stepwise Approach to Develop, Test and Analyse a Hybrid Social-ecological Model

Here, we present our general stepwise approach of developing and analyzing a hybrid model representing a SES (**Figure 1**). Developing one of the submodels might involve a model development cycle in itself (Schmolke et al., 2010) but here we focus solely on considerations for coupling existing submodels that build on different paradigms or operating on different aggregation levels. The same accounts for results of analyses of the decoupled subsystems which for the purpose of this coupled model are elements of the model building process. We consider the outcomes from the coupled model that answer our research question about the impact of social time lags for lake restoration our model results.

The first step starts on the conceptual level with the aim to identify the main social-ecological interactions in the target SES that need to be defined for the hybrid model. This includes a specification of the links between the ecological and social systems, as well as the submodels representing each of the subsystems, their aggregation level, temporal, and spatial scales. From those specifications, one can also derive which modeling paradigm is suitable to represent each submodel.

The second and third steps are performed for each submodel separately and involve tests to verify or validate the submodel by itself and with respect to the other model it will be coupled with. Step two can involve a newly developed or an existing model that is now adapted and analyzed for the purpose of the hybrid model. Adapting an existing model might involve a transfer of the model to a different modeling environment and tests of characteristic submodel behavior. For step three, we propose a sensitivity analysis for expected parameter changes that might become relevant when coupling the two models. So, each submodel is tested for its behavior driven by the output expected from the other respective submodel (but without a feedback to the other model). This provides a better understanding of one-directional influences before the feedback between submodels is fully integrated. The final step four goes back to the fully integrated social-ecological view and explores the behavior of the coupled model implementation. Prior submodel analyses can be very helpful to interpret results from the fully implemented feedback.

To illustrate the proposed steps, the remaining parts of this section present the actual description and analysis of our hybrid model on lake restoration. Thus firstly, we provide an overview on the model structure including the links between the social and ecological systems, assumptions, scales, and the included processes (Step 1). Secondly, we focus on the ecological submodel and reproduce an existing minimal lake regime shift model using the tool Matlab Grind to identify stable state values for the fish populations and the critical nutrient levels (Step 2b). These are

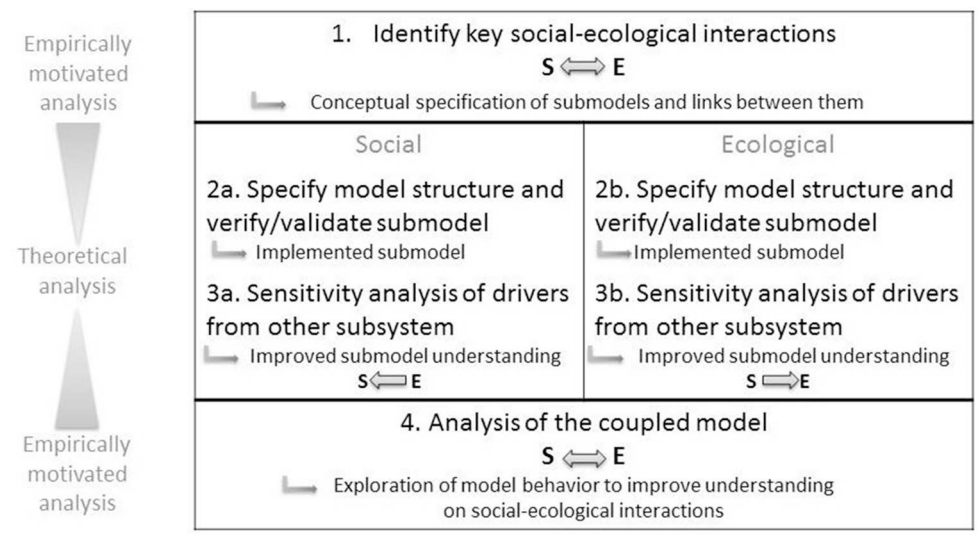


FIGURE 1 | Main steps for developing and analyzing a hybrid model for a SES.

used to test the reimplementation in NetLogo, an environment that is capable and comfortable for developing agent-based and system dynamics models. Note that in our case for the social submodel this step (2a) is less relevant because we build on a model that has already been developed in a SES context (Tavoni et al., 2012). The social submodel testing therefore takes place mainly in step 3a. Thirdly, we test the behavior of both submodels with respect to changing drivers from the other respective submodel (Step 3a and b) and finally, we explore the coupled model (Step 4).

A Hybrid Model to Examine Lake Restoration: Limnoses (Step 1)

The purpose of our hybrid model ("LimnoSES," stands for "limnological social-ecological system") is to investigate human responses to a turbid lake that might promote or delay a desired regime shift back to the clear state. The shift from a clear to a turbid state has been frequently described and analyzed (Scheffer, 1990; Carpenter and Brock, 2004; Mooij et al., 2009) but the decisions and social processes that are necessary for lake restoration have been much less investigated. We aim to get a better system understanding about the relevant social-ecological interactions and therefore develop submodels for the social and ecological subsystem respectively.

The social-ecological system of lake management as modeled in LimnoSES consists of a lake subsystem and a social subsystem that are connected through two social-ecological interactions

Figure 1 Main steps for developing and analyzing a hybrid model for a SES.

Figure 2 One interaction is the monitoring activity where the responsible authority oversees regular measures of indicators of the state of the lake. They report annually data on water quality, nutrients, composition and density of planktonic algae, amount of fish and submerged plants (macrophytes). For the purpose of this model, we chose the density of pike as a system-level indicator of the lake's state because it is also a commercially interesting fish species. As only the aggregated number of fish rather than individual pike characteristics play a role in determining the lake state, this gives a first indication for the suitability of a system dynamics perspective for the ecological submodel.

The second interaction is the pollution of the lake caused by insufficient treatment of sewage water from private house owners in the catchment. We assume that the increase of nutrients in the lake in the past was mainly caused by fertilizer use in agriculture and historically insufficient municipal sewage treatment. But since the use of fertilizer has been regulated and municipal sewage treatment has been technically upgraded, private house owners and their aggregated impact have now a significant influence on lake nutrient levels. They are thus important actors to decrease today's nutrient input to the lake by upgrading their onsite-sewage systems (OSS) (Wallin et al., 2013b). Before the aggregated impact from private sewage treatment affects the lake, house owners decide individually whether or not to upgrade their private system and their decision is for instance influenced by the current legislation. These two levels of individual and collective social processes indicate that an agent-based perspective is suitable for the social submodel.

Starting from the focal social-ecological interactions, we determined which entities and processes need to be represented in the ecological and the social submodels in order to model the restoration process. The ecological submodel, for which

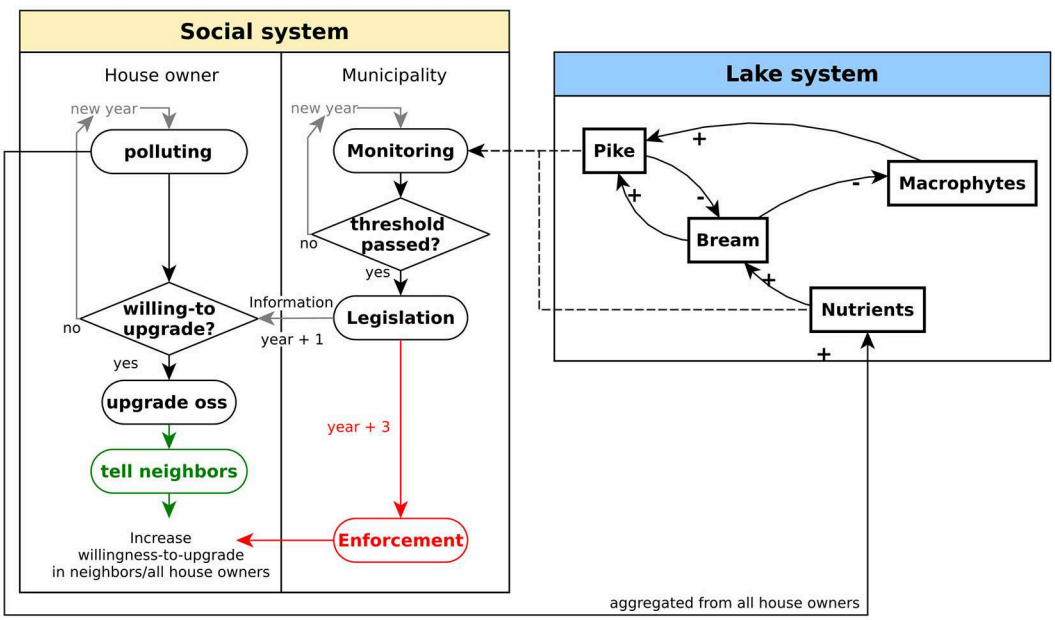


FIGURE 2 | Conceptual graph of our hybrid model with the social system represented by a flow chart and the lake system by a causal-loop diagram. The subsystems are connected via the monitoring of the municipality and the nutrient release from private house owners with insufficient onsite-sewage systems (OSS). The colored processes for the social actors show optional, additional responses that are explored and compared in our model analysis.

we choose the minimal model by Scheffer (1989), represents two fish populations, namely pike (piscivore fish) and bream (whitefish) which interact as predator and prey respectively. The fish populations are also influenced by the amount of nutrients and macrophytes in the lake. The minimal model operates on a daily time scale and shows bistable behavior for the fish densities in response to present nutrient amounts.

The causal-loop diagram for this minimal lake model (**Figure 2**, right) consists of two feedback loops. A balancing loop describes the predator-prey relationship between pike and bream which enables these two populations to equilibrate around a stable state. The reinforcing loop that connects pike and bream via the amount of macrophytes can on the other hand force the two fish populations away from the stable state they were previously in. These two feedback loops are the main structural elements that enable bistable behavior of the lake system and therewith regime shifts between two stable states.

The social submodel represents the regulation of private sewage treatment that is triggered by the municipality when a worsening state of the lake is monitored (**Figure 2**, left). We assumed a simple regulation response on an annual time scale where the municipality acts upon drops of the nutrient level below an assumed threshold and informs the private house owners in the following year about the need to upgrade their on-site sewage system (OSS). Agents represent individual house owners ($n = 100$) that together release nutrients to the lake while using their old OSS. After house owners have been informed about the need to upgrade their OSS, they individually decide whether they make the investment for the OSS upgrade. The outcome of this decision depends in reality on a multitude of economic and social factors (such as capital, social norms and more; Wallin, 2012) that we combine for the purpose of this study within one individual probability named "willingness-to-upgrade."

The problem of OSS upgrading of individuals in the community for the benefit of all represents a social dilemma where individual and collective benefits are not aligned and the incentive to free-ride and not contribute to the public good of a clean lake is high. There is a rich literature and many models exploring possible mechanisms that enable cooperation for achieving the socially optimal outcome. The development of the social submodel has been inspired by the model of Tavoni et al. (2012) which investigates the role of social disapproval for successful cooperation. We explore two different mechanisms of how this willingness to upgrade could be influenced based on evidence from a recent survey among Swedish house owners (Wallin et al., 2013b). When confronting house owners with the request to upgrade their sewage system, they find themselves in a high-cost low-benefit situation. As a consequence, house owners tend to avoid timely upgrade of their OSS and the question is how their decision to upgrade can be supported. In some cases, inspectors were sent out by municipalities to check the current installation and support upgrading by reminding house owners about their legal obligation ("central enforcement"). But in many municipalities, inspectors could not be afforded and successful OSS upgrade was found to depend on horizontal information exchange among house owners and their trust in the governance system ("social engagement").

The general schedule of how the two submodels are implemented, linked and processed is as follows. Each model year starts with the calculation of the aggregated release of nutrients by the house owners. The amount of nutrients is used to drive the daily dynamics of the fish and macrophyte densities. Once a year, the nutrient level in the lake is monitored by the municipality. If the threshold is crossed, the municipality initiates a process of legislation and informs all house owners in the following year about a new regulation that requires the upgrade of their OSSs. Depending on the house owners' willingness-to-upgrade, they upgrade their OSS and therewith eventually stop releasing nutrients into the lake. Two alternative mechanisms are assumed that potentially increase the house owners willingness-to-upgrade, one is through horizontal enforcement ("social engagement":"tell neighbors" in **Figure 2**) among house owners and the other through vertical enforcement from the municipality ("central enforcement": "enforcement" in **Figure 2**). More details on the model description can be found in the Supplementary Material where we used the protocol ODD+D that was specifically designed to describe human decisions in agent-based models (Müller et al., 2013).

Our hybrid model simulation fully couples the agent-based and system dynamics submodel with a sustained feedback on the system level in both directions which can be called an integrated hybrid simulation (Swinerd and McNaught, 2012 or class II-type after Shanthikumar and Sargent, 1983). Thus, former external drivers of the submodel, such as the nutrient flow, became integrated and result from the aggregated decisions of individual house owners. LimnoSES was implemented in NetLogo (**Figure 3**).

Validating Patterns from the Minimal Lake Regime Shift Model (Step 2b)

The lake submodel consists mainly of two differential equations that are able to reproduce shifts between a clear and a turbid state indicated by the pike/bream-relationship (Scheffer, 1989). The densities of the two modeled fish populations span a state space for this ecological submodel. Typical patterns for regime shifts are bistability in the state space and hysteresis, which should be reproduced by the ecological submodel and are relevant for the analysis of our hybrid lake model. The stability analyses are usually performed on continuous model implementations. Within our hybrid model, however, a discrete representation is required because of discrete events in the social system. We calculated the values for the stable states of the continuous lake submodel at different nutrient levels in order to use them later to verify the discrete implementation.

We implemented the given equations and parameter settings in Matlab Grind[1] to visualize the stable states within a state-space diagram. Depending on the nutrient level, the zero-isocline of bream has different positions and determines whether the lake system has one or two stable states (**Figure 4**). For the bistable case, one state shows a very high density of bream with a very low

[1]http://www.sparcs-center.org/grind.

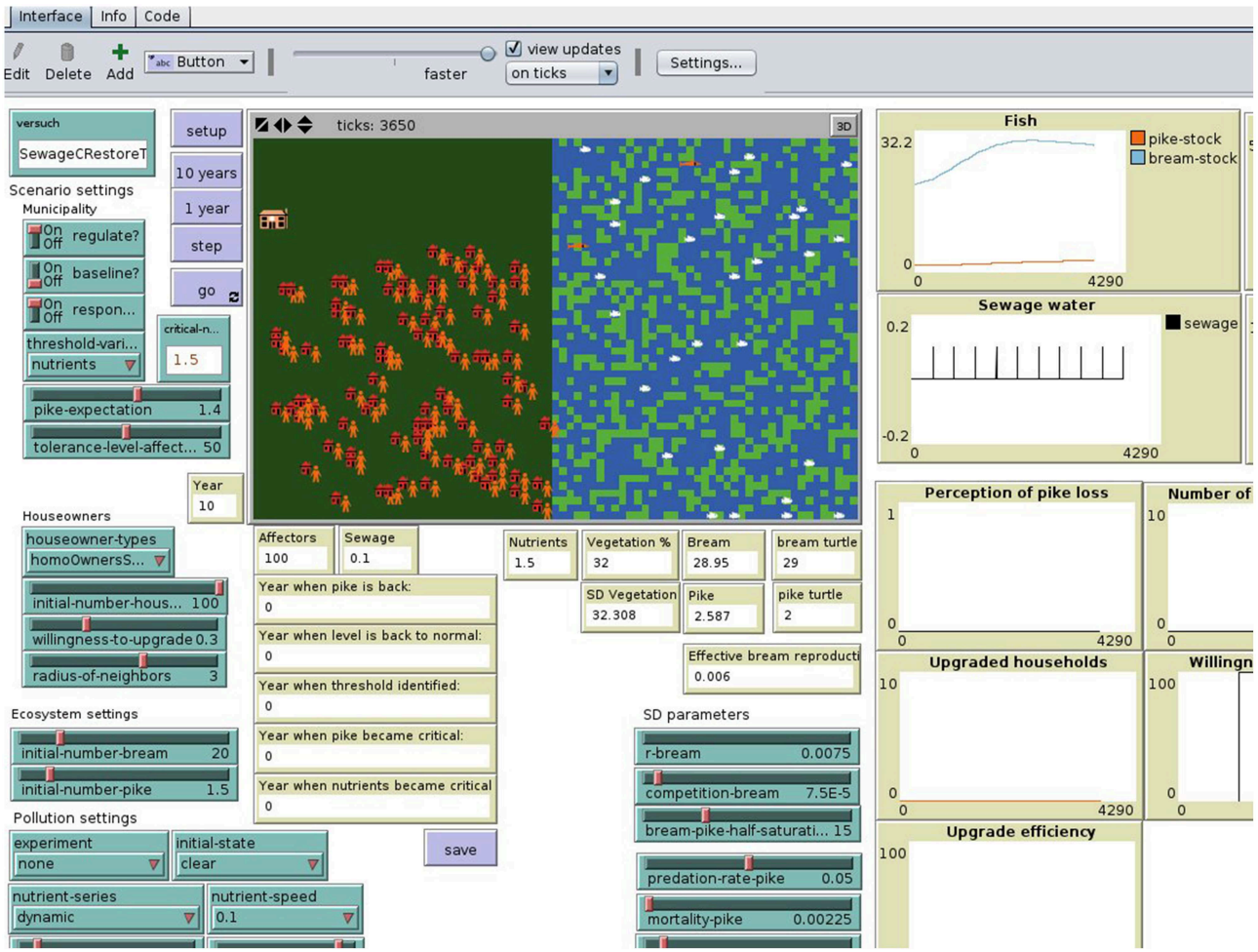


FIGURE 3 | Part of the graphical interface from our "LimnoSES" implementation in NetLogo showing in the main visualization panel house owner agents on the left side and representative agents for the fish stocks on the right side. Buttons, sliders, information boxes and diagrams are used for testing single scenarios before they were executed in a batch mode for the analyses.

density of pike which represents the undesired state where the lake is turbid. The desired state instead has a lower bream density and a relatively high pike density. This analysis also shows that the desired state cannot be reached if the nutrient level is too high. The stable state values for the clear water stat (high pike density) and turbid stat (low pike density) are later used to initialize a lake eutrophication or restoration scenario respectively.

Another convenient tool to investigate system-dynamics models within Matlab Grind is a bifurcation analysis to determine the parameter range for driver values (here: nutrients) where bistability can be expected. The values for the stable states together with the parameter range for nutrients allow us to formulate and test hypotheses about the expected behavior of this model when it runs in a different implementation. So, for instance, at a nutrient level of 1.5 (dimension-less), stable states can be expected around a population density of 25.8 $g \cdot m^{-2}$ for bream and 2.6 $g \cdot m^{-2}$ for pike in the clear state and at 84 $g \cdot m^{-2}$bream and 0.04 $g \cdot m^{-2}$ pike in the turbid state. The hysteretic behavior of fish densities in response to nutrients ranges between 0.9 and 2.1. This means that, above this range, the lake system ultimately runs into the turbid state, whereas below this range, only the clear state is reached. Within the bistable range of nutrients, it depends on the initial values of pike and bream whether the lake runs into a clear or turbid state.

We decided to implement the prototype of our hybrid model in NetLogo since the main interest in this study is to test different mechanisms of human decision-making in response to the lakes state. This environment is primarily designed to implement and test agent-based models but it has an interface to implement system dynamics models as well. But since the numeric representation of such models is always discrete, we had to test the suitable step size, meaning the respective time scale, at which bistability and regime shifts between those states can be reproduced.

The following calibration tests were performed to ensure that the NetLogo implementation reproduces the specified patterns above for the ecological subsystem:

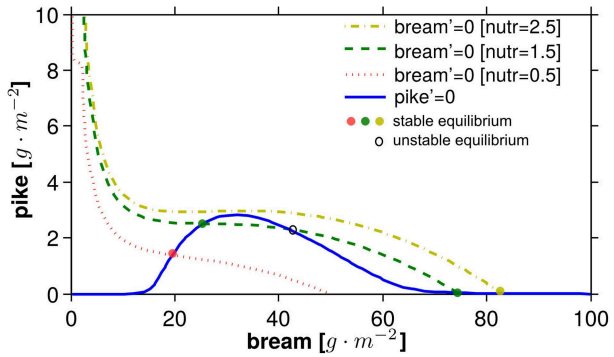


FIGURE 4 | State-space diagram showing zero-isoclines for bream and pike with different stable states depending on the nutrient level that influences the position of the zero-isocline for bream. The zero-isocline for pike is unaffected by the nutrient level. Reimplementation is based on the model developed by Scheffer (1989).

- Can bistable behavior be reproduced?—Yes, increase or decrease in nutrients respectively cause the fish stocks to run into different states.
- What are suitable initial states for fish populations to stay in the clear or turbid state? Does the time-scale (annual, daily) play a role?—While the time scale is not relevant for continuous models, it actually has an effect in time discrete simulations. Although the original model operates with parameters on the daily time scale, we checked whether it is possible to transform it to an annual scale to simplify the link to the annually operating social submodel. But, the annual scale for the system dynamics was found to limit the initial conditions that are suitable to lead to the clear state considerably. The daily scale was chosen to be more suitable thereafter as the state space for fish stocks was more equally divided between the clear and turbid state.

Sd-submodel Test: How does the Lake State Respond to a Dynamic Driver from the Social System? (Step 3b)

Since the coupled model is intended for the investigation of changing nutrient values affecting the lake state, we are particularly interested in the transient dynamics. This is in contrast to a common bifurcation analysis where simulations run with constant parameters until a stable state is reached. We aim to explore the transient dynamics in the ecological submodel caused by in- or decreasing nutrient values because lake managers are confronted with situations where transient dynamics may cause different outcomes in terms of the lake state. We also want to estimate in what way the response time of house owners to upgrade their sewage system may affect nutrient change and therewith the lake state.

For the following experiment, we run scenarios of nutrient increases at three different rates and respective time spans (10, 20, and 40 years) so that they end at the same nutrient level. We selected three suitable nutrient end levels which were right below, in the middle and above the hysteresis range (**Figure 5**, left) For this experimental setup, we expected that simulations with the highest nutrient level result in the turbid state and simulations running to the lowest nutrient level at the clear state. But it was unclear how the outcome will be when the medium nutrient level is targeted. The result confirms that under a low and high nutrient regime, the lake reaches a clear and turbid state, respectively, independent from the rate of nutrient changes over time. The lake reaches a clear state once the nutrient level is below the hysteresis range, as indicated by bream and pike densities in the state space (relatively high pike and low bream densities, **Figure 5**, right). Further, the lake reaches a turbid state once the nutrient level goes beyond the hysteresis range independent of the rate of nutrient change. In contrast to these extreme scenarios, the medium nutrient level scenario shows that the end state of pike and bream densities can also depend on the previous rate of nutrient increments. The highest rate caused the fish populations to end in the turbid state, while lower rates enabled leveling off into the clear state.

The rate dependent behavior in transient simulations extends previous equilibrium analyses that solely examined stable states. The rate dependent shift to another basin of attraction was called r-tipping before (Ashwin et al., 2012) but it is much less recognized than bifurcation or driver related tipping points. Interestingly, we could not find this rate dependent behavior in an experiment where nutrients were reduced in the same way as they were increased in this experiment, which would simulate the restoration process at different rates.

ABM-submodel Test: How House Owners Respond to a Lake Becoming Turbid? (Step 3a)

Since we had no particular pattern to test our social submodel against, we explored the general effect from three potential response scenarios of house owners that need to upgrade their sewage system. To this end, all house owner agents were initialized homogeneously with a value for their willingness-to-upgrade. We compared scenarios of homogeneous house owners without interaction among them as a baseline with two scenarios that describe social interactions with which the initial willingness-to-upgrade can be increased over the simulation time. For the baseline scenario, the rate is static with which house owners upgrade their OSS. In the "central enforcement" scenario it is assumed that, after a fixed time lag following the municipal legislation for upgrading OSS, inspectors visit individual house owners which promotes the individual willingness-to-upgrade. In the "social engagement" scenario, we assume that direct knowledge exchange with neighbors and the house owners sense of responsibility to follow the law improves the willingness-to-upgrade (Wallin et al., 2013a). So, house owners are more likely to follow the example of their neighbors who already upgraded their OSS. From this assumption follows that at low values for the willingness-to-upgrade, upgrading activities will take off quite slowly before they accelerate and reach every house owner. But at higher values for the willingness-to-upgrade, a much faster upgrade by all house owners can be expected. In contrast to this scenario, the "central enforcement" scenario explores how after a fixed time lag (assumed as a buffer time before a municipality finds time and resources to check the implementation of the new

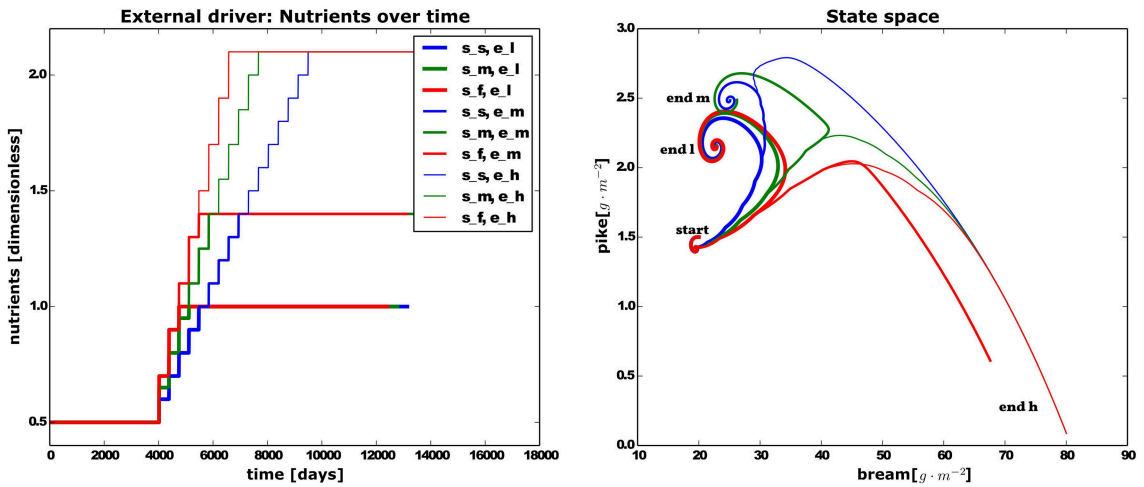


FIGURE 5 | Left: Nutrient time series as driver for scenarios, at three rates (speed: slow, medium, high), and three different destination levels (end: low, medium, high), showing results from nine simulations in total. Right: The resulting fish populations in the state space. At the lowest nutrient levels, the resulting fish state is independent of the rate of change. But in the intermediate nutrient level, fish populations driven by the fastest increase (red) experience pike reduction while at slower rates pike increases can be observed. At the highest nutrient level, all fish simulations result in the low-pike-state.

law) all house owners are inspected which is assumed to increase their willingness-to-upgrade.

In **Figure 6** we present the average time lag between the year that house owners were informed about new sewage water regulations and the year that they actually upgraded their sewage system. We aggregated the individual responses and show only the average time lag for the total population in each scenario. Along the x-axis with increasing values for the willingness-to-upgrade, the time-lag for OSS upgrade decreases exponentially in all three scenarios as expected. But further, the scenario comparison revealed that the "central enforcement" worked best, in terms of a short time lag, for a house owner population with an extremely low willingness-to-upgrade in the beginning. But as soon as this initial willingness was higher, "social engagement" performed better in terms of lowering the average time lag to upgrade.

Coupled Model Analysis: How Does Ecological Restoration Time Relate to Social Response Lags? (Step 4)

To analyze the coupled simulation model, we explored how the ecological and the social system respond to each other in terms of time lags that emerge from their respective dynamics. The purpose of this experiment is to relate the time necessary for ecological restoration (in terms of sufficiently high pike densities) to the preceding time lag in human responses to critical nutrient levels (through legislation and upgrades of private sewage treatment). For the ecological time lag, we calculated the time interval between the year when pike first drops below the expected threshold and when pike returns to levels above this threshold. As in the previous experiment, the social time lag is calculated as the interval between the detection of a surpassed nutrient threshold and the time step when the house owners, in average, upgraded their OSS. From previous experiments with the decoupled models, we expect that for low values of willingness-to-upgrade, the social time lag is lowest for the "central enforcement" scenario. In contrast, higher willingness-to-upgrade values are expected to be more effective in the "social engagement" scenario to result in short social time lags. We further hypothesize that ecological time lags increase nonlinearly with increments in social response lags due to the reinforcing feedbacks identified in the lake system.

Results shown in **Table 2** confirm those hypotheses. The mechanism of central enforcement seems to be effective to shorten social responses to critical nutrient levels under conditions of low willingness-to-upgrade among house owners. Contrasting to that, the mechanism of social engagement is more effective when the initial willingness-to-upgrade is slightly higher. The relation between social and ecological time lags reveals that under all scenarios increases in the social time lag causes a multiple increase in the time lag for ecological restoration.

Discussion of Challenges and Benefits from Combining Agent-based and System Dynamics Models

Analyzing complex social-ecological systems is often approached with models that need to deal with multiple sources of complexity: two-way interactions or feedbacks between the social and ecological systems, processes on multiple scales, links between different levels of aggregation, and emergent as well as downward-causation processes. To tackle those challenges, we proposed a stepwise approach to develop and analyze a hybrid model that links micro-level human behavior and system level ecological dynamics with the aim to systematically investigate two-way social-ecological interactions. Using our hybrid model

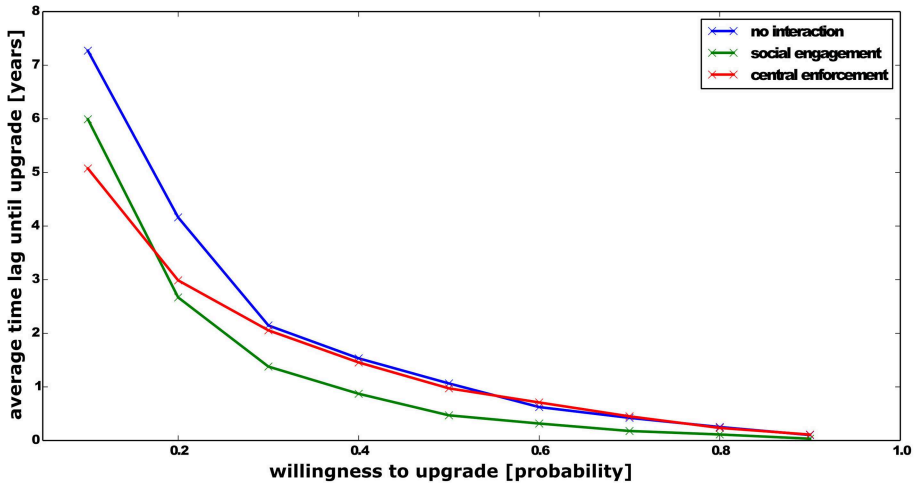


FIGURE 6 | Three scenarios showing the aggregated, average time that is necessary for private house owners to upgrade their sewage system over a range of initial values for "willingness-to-upgrade."

TABLE 2 | Calculated time lags from simulations under different scenarios.

Scenario	Willingness-to-upgrade	Social lag [years]	Ecological lag [years]
No Interaction	0.1	9.6 ± 0.9	34.6 ± 2.8
	0.2	4.2 ± 0.4	9.7 ± 4.8
Social engagement	0.1	5.7 ± 0.4	21.4 ± 2.5
	0.2	**2.5 ± 0.3**	**3.9 ± 0.3**
Central enforcement	0.1	**4.9 ± 0.3**	**19.9 ± 2.1**
	0.2	3 ± 0.3	4.8 ± 1.9

Bold entries mark the minimal time lag compared to the alternative interaction scenarios tested.

as an illustrative example, we focus here on the conceptual and methodological challenges of combining an agent-based and a system dynamics modeling approach to explore social-ecological interactions and outline the advantages therefrom.

Agent-based and system dynamics models were often viewed as antagonistic approaches to represent complex systems as they operate at different levels of aggregation and capture different, complementary aspects of the system under study (Parunak et al., 1998). Recently, the common and synergistic features were highlighted for coupled ecological models (Vincenot et al., 2011) but implementations of different types of submodels in one integrated model are rare and mainly in single domains (see for examples Hudjetz et al., 2014; Bradhurst et al., 2015; Vincenot et al., 2015). Here, we want to stress how beneficial it is to couple agent-based and system dynamics models in a hybrid model to investigate SES particularly for bridging between such distinct domains as ecology and society.

A Stepwise Approach to Design and Analyze Hybrid Models for SES

It was acknowledge before that ecosystem models need to integrate more aspects of complexity (Gray and Wotherspoon, 2012) but we found that a general strategy to build coupled models for SES that helps navigate the increasing conceptual and analytical complexity was lacking so far. To not reinvent the wheel, it is of great help to build on already existing and fully analyzed models in specialized domains (Mooij et al., 2010), choosing carefully among available approaches (Kelly(Letcher) et al., 2013), and follow established procedures for document their development and application (Schmolke et al., 2010; Müller et al., 2013). Building on this valuable literature, we propose a procedure to couple agent-based and system dynamics models to explore SES. Dividing hybrid model development and analysis tasks into distinctive steps as outlined in our procedure proved useful to approach system complexity incrementally and we reflect on this procedure with illustrations from developing a model on lake restoration.

Conceptualizing Social-Ecological Interactions and Specifying Submodels (Steps 1 + 2)

To start the conceptualization and development of a coupled model representing a SES, we consider the identification of key social-ecological interactions that explain the phenomena of interest, e.g., the restoration of shallow lakes, as a helpful way to identify the main links between the social and ecological subsystems. It further helps to specify the research question and draw suitable system boundaries. In specific, we focused on how authorities monitor the lake and respond to the turbid state through regulating private installations for sewage water treatment. Those social-ecological interactions relate to specific aggregation levels (e.g., the community, individual house owners, turbidity of the lake), temporal and spatial scale (e.g., daily/annual), occur in response to certain events or trigger other processes. Those considerations help to draft first submodel characteristics, expected submodel

patterns or behavior, and their boundaries to finally select a suitable paradigm for the submodels. If the pattern of interest is homogeneous, deterministic and aggregated over space and entities, representation through an average, system-level value for this characteristic might be sufficient. But in case individual processes depend on historical or environmental context, stochastic events and local interactions of heterogeneous entities, they are better represented by an agent-based approach. Translating an agent-based model version into a system dynamics version and vice versa can help to identify the sensitivity of outcomes toward the implemented modeling paradigm (Wilson, 1998). We suggest that the choice of modeling paradigm should be an informed decision taking the research question, system characteristics, and the expected outcome into account asking e.g., whether heterogeneity matters for the research question. Rather than striving to represent every process within a single paradigm, we argue here that in some cases a greater system understanding can be achieved when combining different types of submodels and using the traditional, specialized analysis tools that are available for the respective approaches.

In our case, ecological lake dynamics were modeled on the systems level at a daily time scale and the social processes were based on micro-level decisions and operated on an annual time scale. Linking those contrasting submodels can be done in multiple ways (Swinerd and McNaught, 2012): (i) the link can be uni- or bidirectional, (ii) system level variables effect individuals in the same way or differently, (iii) individuals affect system variables individually or aggregated. For instance, fishing can be done by individual fishers with specific strategies on a fish population that is modeled on the aggregated system level. The polluting activity in our case, however, is rather an aggregated, collective impact from individual house owner decisions on the lake's system level. We identified three general model design questions that can support the choice of how to specify and link the required submodels: (i) are submodel processes best described continuously or discrete? (ii) which different time scales apply in the submodels to be coupled?, (iii) is the overall model outcome expected to analyze transient or stable state dynamics?

For step two in our procedure, we asked for submodel verification to reproduce characteristic model behavior or patterns within a given model paradigm and potentially a validation against empirical data. Using existing, fully analyzed models as submodels is desirable but great care is necessary for transferring them into a new context and probably a different implementation environment. Examples exist for how this can be achieved for system dynamics and agent-based models (Macal, 2010) but it seems to be rarely done for SES models so far.

Strategies for Calibrating, Testing, and Analyzing Hybrid Models (Step 3 + 4)

For step three, we explored the behavior of each submodel with respect to drivers expected as output from the other respective submodel (but without a feedback to the other model).

For the lake model, we analyzed the stable states at different rates of gradual nutrient increases (**Figure 4**). They led to different stable states than those obtained from common bifurcation analyses (Janse et al., 2008), which proved that not only the absolute value of the driver but also its transient dynamics can determine the lake state. This corresponds to previous tipping point analyses with analytical models where different mechanisms causing regime shifts were described (Ashwin et al., 2012). Our example shows that the choice between analyzing equilibria or transient dynamics can result in qualitatively very different outcomes. This has implications e.g., for models that aim to provide estimates of critical nutrient loadings for lake managers (Janse et al., 2008) and supports the argument that multiple modeling approaches are required to create an integrative view on system dynamics (Mooij et al., 2010).

Unfortunately it is often the case for SES models that data and knowledge on the social subsystem structure and dynamics are much more limiting than for the ecological subsystem. It therefore becomes crucial to determine the robustness of model results to model assumptions and parameterization. This should include testing for structural uncertainties such as different mechanisms for human decision making. Our analysis of the "central enforcement" and "social engagement" scenarios showed that the mechanisms of how rules on sewage treatment are implemented determine the social response time to critical lakes states. Empirical evidence on the actual duration of policy making and adaptation of new rules is important to make these model processes more realistic. Insights from theoretical models that study prototypical social situations such as commons dilemmas such as the one underlying human individual and collective behavior in our lake restoration case can also be valuable starting points to identify relevant social mechanisms.

During the hybrid model analysis in step 4, we encountered two challenges that particularly relate to agent-based and system dynamics implementations. One mechanism that is crucial to understand for managing regime shifts is the switch of dominant feedbacks in the system. This can hardly be analyzed in the coupled model implementation that is based on simulation outcome. Here we suggest an iterative procedure of comparing experiments using traditional analysis tools for dynamical systems (as we exemplified with stable state and bifurcation analyses in step 2) with systematic time series analyses from the coupled model (step 4). This allows a step-wise selection of relevant parameter ranges and specifying further hypotheses.

Methodological Benefits from Building Hybrid Models

Current challenges of developing models to improve understanding of coupled SES involve the integration of existing models without "reinventing the wheel" (Mooij et al., 2010) and combining multiple levels of aggregation (Schlüter et al., 2012). We have integrated a minimal model that was studied extensively before (Scheffer, 1989) into a hybrid model which clearly supported the interpretation of results from the coupled model simulation. Particularly the assessment how feedback loops take effect, which is the special focus in system dynamics

models and may become hidden in coupled simulations, enabled a better understanding of why time lags in the social system can be amplified through ecological dynamics. We observed in those scenarios that delays in the social response to the turbid lake state can increase the time for potential restoration of the clear lake in a non-linear way. While this exemplifies the added value in generating system understanding from a system-level perspective, we also found benefits from integrating the micro-level for social processes. To improve the relevance of models for supporting environmental decision-making and increase trust in scenario simulations, it is crucial to link model outcome to empirical patterns (Grimm et al., 2005). But since empirical evidence for individual decisions on the adoption of new sewage treatment technologies are still lacking, the response time for lake restoration cannot simply be subsumed under one parameter. The agent-based approach for the social submodel revealed the sensitivity of the outcome, here the lake restoration time, toward these micro-level decisions.

We further believe that the integration of existing models and particularly the development of hybrid models strongly benefits from their documentation with broadly applied protocols (Müller et al., 2013, 2014). These comprehensive documentations increase awareness about alternative model design options and support reasoning about the underlying assumptions more explicitly. This is particularly relevant if the process of developing models becomes an interdisciplinary activity.

Suitability and Added Value of Hybrid Models to Understand Social-ecological Systems

Models of social-ecological systems do not necessarily have to be developed such that the ecological part is represented by system dynamics and the social part by an agent-based approach. However, such a separation suits well as an illustrative example where the benefits of using a hybrid approach become particularly visible.

Social-ecological systems research uses multiple perspectives and multiple methods in order to understand the complexity of these systems and develop integrative approaches (Ragin, 1987; Poteete et al., 2010; Biggs et al., 2012a). Specific modeling approaches that emphasize either micro-level interactions that lead to system level patterns or aggregated, system level processes are commonly used. Of special interest are models that create a better understanding of alternative stable states (Scheffer et al., 2009) and transient dynamics in social-ecological systems (Schlüter et al., 2012).

We propose that a hybrid modeling approach enables a "multi-scopic" view on the system, integrating micro- and system level processes and a broader set of analytical tools. Beyond that, it encourages the modeler to reason and be more explicit about the model assumptions and its suitability for the purpose of the study. Developing and combining alternative representations can lead to a more in depth and nuanced understanding of a given phenomenon taking multiple possible explanations and methods into account (Poteete et al., 2010). This, however, requires a greater flexibility in the tool use by model developers. Finally, through the integration of existing minimal, equation-based models, with an agent-based approach, it is possible to link theory to empirical patterns that are currently explored on social-ecological research frontiers. With our presented procedure to model interacting system- and micro-level processes, we hope to motivate more hybrid models being developed and analyzed for social-ecological research.

Acknowledgments

We are very grateful for insightful comments on the earlier manuscript from our colleagues Steven Lade, Nanda Wijermans, and Björn Schulte-Herbrüggen. The comments by two anonymous reviewers and editor Christian E. Vincenot are very much appreciated and helped particularly to sharpen the focus and structure of the paper. RM and MS acknowledge support from the EU-project LIMNOTIP funded under the FP7 ERA-Net Scheme (BiodivERsA, 01LC1207A), MS acknowledges funding by the European Research Council under the European Union's Seventh Framework Programme (FP/2007-2013)/ERC grant agreement no. 283950 SES-LINK and a core grant to the Stockholm Resilience Centre by Mistra.

Supplementary Material

The Supplementary Material for this article can be found online at: http://journal.frontiersin.org/article/10.3389/fenvs.2015.00066

References

Ashwin, P., Wieczorek, S., Vitolo. R., and Cox. P. (2012). Tipping points in open systems: bifurcation, noise-induced and rate-dependent examples in the climate system. *Philos. Tran. R. Soc. Lond. A Math. Phys. Eng. Sci.* 370, 1166–1184. doi: 10.1098/rsta.2011.0306

Balke, T., and Gilbert, N. (2014). How do agents make decisions? A Survey. *J. Artif. Soc. Soc. Simul.* 17, 13. Available online at: http://jasss.soc.surrey.ac.uk/17/4/13.html

Berkes, F., and Folke, F. (ed.) (1998). *Linking Sociological and Ecological Systems: Management Practices and Social Mechanisms for Building Resilience*. New York, NA: Cambridge University Press.

Biggs, R., Carpenter, S. R., and Brock, W. A. (2009). Turning back from the brink: detecting an impending regime shift in time to avert it. *Proc. Natl. Acad. Sci. U.S.A.* 106, 826–831. doi: 10.1073/pnas.0811729106

Biggs, R., Schlüter, M., Biggs, D., Bohensky, E. L., BurnSilver, S., Cundill, G., et al. (2012a). Toward principles for enhancing the resilience of ecosystem services. *Annu. Rev. Environ. Resour.* 37, 421–448. doi: 10.1146/annurev-environ-051211-123836

Biggs, R. O., Thorsten, B., Carl, F., Line, G., Albert, N., Magnus, N., et al. (2012b). "Regime Shifts," in *Encyclopedia of Theoretical Ecology,* eds A. Hastings and L. Gross (Ewing, NJ: University of California Press), 609–617.

Bousquet, F, and Le Page, C. (2004). Multi-Agent simulations and ecosystem management: a review. *Ecol. Model.* 176, 313–332. doi: 10.1016/j.ecolmodel.2004.01.011

Bradhurst, R. A., Roche, S. E., East, I. J., Kwan, P., and Garner, M. G. (2015). A hybrid modelling approach to simulating foot-and-mouth disease outbreaks

in Australian livestock. *Front. Environ. Sci.* 3:17. doi: 10.3389/fenvs.2015.00017

Carpenter, S., Brock, W., and Hanson, P. (1999). Ecological and social dynamics in simple models of ecosystem management. *Conserv. Ecol.* 3, 4. Available online at: http://www.consecol.org/vol3/iss2/art4/

Carpenter, S. R., and Brock, W. A. (2004). Spatial complexity, resilience, and policy diversity? fishing on lake-rich landscapes. *Ecol. Soc.* 9, 8. Available online at: http://www.ecologyandsociety.org/vol9/iss1/art8/

Carpenter, S. R. (2005). Eutrophication of aquatic ecosystems: bistability and soil phosphorus. *Proc. Natl. Acad. Sci. U.S.A.* 102, 10002–10005. doi: 10.1073/pnas.0503959102

Ekvall, M. K., Urrutia-Cordero, P., and Hansson, L.-A. (2014). Linking cascading effects of fish predation and zooplankton grazing to reduced cyanobacterial biomass and toxin levels following biomanipulation. *PLoS ONE* 9:e112956. doi: 10.1371/journal.pone.0112956

Epstein, J. M. (2008). Why Model? *J. Artif. Soc. Soc. Simul.* 11, 12. Available online at: http://jasss.soc.surrey.ac.uk/11/4/12.html

Folke, C., Hahn, T., Olsson, P., and Norberg, J. (2005). Adaptive governance of social-ecological systems. *Annu. Rev. Environ. Resour.* 30, 441–473. doi: 10.1146/annurev.energy.30.050504.144511

Gaube, V., Kaiser, C., Wildenberg, M., Adensam, H., Fleissner, P., Kobler, J., et al. (2009). Combining agent-based and stock-flow modelling approaches in a participative analysis of the integrated land system in reichraming, Austria. *Landsc. Ecol.* 24, 1149–1165. doi: 10.1007/s10980-009-9356-6

Gray, R., and Wotherspoon, S. (2012). Increasing model efficiency by dynamically changing model representations. *Environ. Model. Softw.* 30, 115–122. doi: 10.1016/j.envsoft.2011.08.012

Grimm, V., Revilla, E., Berger, U., Jeltsch, F., Mooij, W. M., Railsback, S. F., et al. (2005). Pattern-oriented modeling of agent-based complex systems: lessons from ecology. *Science* 310, 987–991. doi: 10.1126/science.1116681

Haase, D., Haase, A., Kabisch, N., Kabisch, S., and Rink, D. (2012). Actors and factors in land-use simulation: the challenge of urban shrinkage. *Environ. Model. Softw.* 35, 92–103. doi: 10.1016/j.envsoft.2012.02.012

Hare, M., and Deadman, P. (2004). Further towards a taxonomy of agent-based simulation models in environmental management. *Math. Comput. Simul.* 64, 25–40. doi: 10.1016/S0378-4754(03)00118-6

Hudjetz, S., Lennartz, G., Kräemer, K., Roß-Nickoll, M., Gergs, A., and Preuss, T. G. (2014). Modeling wood encroachment in abandoned grasslands in the Eifel National Park - model description and testing. *PLoS ONE* 9:e113827. doi: 10.1371/journal.pone.0113827

Janse, J. H., De Senerpont Domis, L. N., Scheffer, M., Lijklema, L., Van Liere, L., Klinge, M., et al. (2008). Critical Phosphorus loading of different types of shallow lakes and the consequences for management estimated with the ecosystem model PCLake. *Limnol. Ecol. Manag. Inland Waters* 38, 203–219. doi: 10.1016/j.limno.2008.06.001

Janssen, M. A., Walker, B. H., Langridge, J., and Abel, N. (2000). An adaptive agent model for analysing co-evolution of management and policies in a complex rangeland system. *Ecol. Model.* 131, 249–268. doi: 10.1016/S0304-3800(00)00256-8

Janssen, M. A. (2001). An exploratory integrated model to assess management of lake eutrophication. *Ecol. Model.* 140, 111–124. doi: 10.1016/S0304-3800(01)00260-5

Kara, E. L., Heimerl, C., Killpack, T., Bogert, M. C., Yoshida, H. and Carpenter, S. R. (2011). Assessing a decade of phosphorus management in the Lake Mendota, Wisconsin Watershed and scenarios for enhanced phosphorus management. *Aquat. Sci.* 74, 241–253. doi: 10.1007/s00027-011-0215-6

Kelly(Letcher), R.A., Jakeman, A. J., Barreteau, O., Borsuk, M. E., ElSawah, S., Hamilton, S. H., et al. (2013). Selecting among Five Common Modelling Approaches for Integrated Environmental Assessment and Management. *Environ. Model. Softw.* 47, 159–181. doi: 10.1016/j.envsoft.2013.05.005

Kininmonth, S., Bergsten, A., and Bodin, Ö. (2015). Closing the collaborative gap: aligning social and ecological connectivity for better management of interconnected wetlands. *Ambio* 44, 138–148. doi: 10.1007/s13280-014-0605-9

Le, Q. B, Park, S. J., Vlek, P. L. G., and Cremers, A. B. (2008). Land-Use Dynamic Simulator (LUDAS): a multi-agent system model for simulating spatio-temporal dynamics of coupled human-landscape system. I. structure and theoretical specification. *Ecol. Inform.* 3, 135–153. doi: 10.1016/j.ecoinf.2008.04.003

Levin, S. A. (1998). Ecosystems and the biosphere as complex adaptive systems. *Ecosystems* 1, 431–436. doi: 10.1007/s100219900037

Macal, C. M. (2010). "To agent-based simulation from system dynamics," in *Proceedings of the 2010 Winter Simulation Conference, IEEE Xplore Conference,* eds B. Johansson, S. Jain, J. Montoya-Torres, J. Hugan, and E. Yücesan (Argonne, IL).

MEA. (2005). *Ecosystems and Human Well-Being: Synthesis*. Ecosystem Assessment Millenium. Washington, DC: Island Press.

Mooij, W. M., De Senerpont Domis, L. N., and Janse, J. H. (2009). Linking species- and ecosystem-level impacts of climate change in lakes with a complex and a minimal model. *Ecol. Model.* 220, 3011–3020. doi: 10.1016/j.ecolmodel.2009.02.003

Mooij, W. M., Trolle, D., Jeppesen, E., Arhonditsis, G., Belolipetsky, P. V., Chitamwebwa, D. B. R., et al. (2010). Challenges and opportunities for integrating lake ecosystem modelling approaches. *Aquat. Ecol.* 44, 633–667. doi: 10.1007/s10452-010-9339-3

Müller, B., Balbi, S., Buchmann, C. M., de Sousa, L., Dressler, G., Groeneveld, J., et al. (2014). Standardised and transparent model descriptions for agent-based models: current status and prospects. *Environ. Model. Softw.* 55, 156–163. doi: 10.1016/j.envsoft.2014.01.029

Müller, B., Bohn, F., Dreßler, G., Groeneveld, J., Klassert, C., Martin, R., et al. (2013). Describing human decisions in agent-based models - ODD+D, an extension of the ODD protocol. *Environ. Model. Softw.* 48, 37–48. doi: 10.1016/j.envsoft.2013.06.003

Ostrom, E, (1990). *Governing the* Commons: *The Evolution of Institutions for Collective Action*. Cambridge: Cambridge University Press.

Parunak, H. V. D., Savit, R., and Riolo, R. L. (1998). "Multi-Agent systems and agent-based simulation," *Proceedings of the First International Workshop of Multi-Agent Systems and Agent-Based Simulation* (Berlin; Heidelberg: Springer-Verlag).

Pers, B.C. (2005). Modeling the response of eutrophication control measures in a Swedish Lake. *Ambio* 34, 552–558. doi: 10.1579/0044-7447-34.7.552

Poteete, A. R, Janssen, M. A., and Ostrom, E. (eds.). (2010). *Working Together: Collective Action, the Commons, and Multiple Methods in Practice*. Princeton, NJ: Princeton University Press.

Ragin, C. (1987). *The Comparative Method: Moving beyond Qualitative and Quantitative Strategies*. Berkeley, CA: University of California Press.

Scheffer, M., Bascompte, J., Brock, W. A., Brovkin, V., Carpenter, S. R., Dakos, V., et al. (2009). Early-Warning signals for critical transitions. *Nature* 461, 53–9. doi: 10.1038/nature08227

Scheffer, M. (1989). Alternative stable States in Eutrophic, shallow freshwater systems: a minimal model. *Hydrobiol. Bull.* 23, 73–83. doi: 10.1007/BF02286429

Scheffer, M. (1990). Multiplicity of stable states in freshwater systems. *Hydrobiologia* 200–201, 475–486. doi: 10.1007/BF02530365

Schlüter, M., McAllister, R. R. J., Arlinghaus, R., Bunnefeld, N., Eisenack, K., Hölker, F., et al. (2012). New horizons for managing the environment: a review of coupled social-ecological systems modeling. *Nat. Resour. Model.* 25, 219–272. doi: 10.1111/j.1939-7445.2011.00108.x

Schmolke, A., Thorbek, P., DeAngelis, D. L., and Grimm, V. (2010). Ecological models supporting environmental decision making: a strategy for the future. *Trends Ecol. Evol.* 25, 479–486. doi: 10.1016/j.tree.2010.05.001

Shanthikumar, J. G., and Sargent, R. G. (1983). A unifying view of hybrid simulation/analytic models and modeling. *Oper. Res.* 31, 1030–1052.

Steffen, W., Richardson, K., Rockström, J., Cornell, S. E., Fetzer, I., Bennett, E. M., et al. (2015). Planetary boundaries: guiding human development on a changing planet. *Science* 347:1259855. doi: 10.1126/science.1259855

Swinerd, C., and McNaught, K. R. (2012). Design classes for hybrid simulations involving agent-based and system dynamics models. *Simul. Model. Pract. Theory* 25, 118–133. doi: 10.1016/j.simpat.2011.09.002

Tavoni, A., Schlüter, M., and Levin, S. (2012). The survival of the conformist: social pressure and renewable resource management. *J. Theor. Biol.* 299, 152–61. doi: 10.1016/j.jtbi.2011.07.003

van den Belt, M. (2004). *Mediated Modeling - A System Dynamics Approach To Environmental Consensus Building*. Washington, DC: Island Press.

Vincenot, C. E., Giannino, F., Rietkerk, M., Moriya, K., and Mazzoleni, S. (2011). Theoretical considerations on the combined use of system dynamics and individual-based modeling in ecology. *Ecol. Model.* 222, 210–218. doi: 10.1016/j.ecolmodel.2010.09.029

Vincenot, C. E., Mazzoleni, S., Moriya, K., Cartenì, F., and Giannino, F. (2015). How spatial resource distribution and memory impact foraging success: a hybrid model and mechanistic index. *Ecol. Complexity* 22, 139–151. doi: 10.1016/j.ecocom.2015.03.004

Wallin, A., Zannakis, M., Johansson, L.-O., and Molander, S. (2013a). Influence of interventions and internal motivation on Swedish Homeowners' change of on-site Sewage systems. *Resour. Conserv. Recycling* 76, 27–40. doi: 10.1016/j.resconrec.2013.04.004

Wallin, A., Zannakis, M., and Molander, S. (2013b). On-site sewage systems from good to bad to…? Swedish experiences with institutional change and technological dependencies 1900 to 2010. *Sustainability* 5, 4706–4727. doi: 10.3390/su5114706

Wallin, A. (2012). *Factors Influencing Actors at the Interface between the Socio-Technical and the Ecological Systems: The Case of on-Site Sewage Systems and Eutrophication*. Division of Environmental Systems Analysis, Department of Energy and Environment, Gothenburg.

Wilson, W. (1998). Resolving discrepancies between deterministic population models and individual-based simulations. *Am. Nat.* 151, 116–134. doi: 10.1086/286106

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Environmental Science

ORIGINAL RESEARCH
published: 07 August 2015
doi: 10.3389/fenvs.2015.00055

Linking Bayesian and agent-based models to simulate complex social-ecological systems in semi-arid regions

Aloah J. Pope and Randy Gimblett*

School of Natural Resources and the Environment, College of Agriculture and Life Sciences, University of Arizona, Tucson, AZ, USA

Interdependencies of ecologic, hydrologic, and social systems challenge traditional approaches to natural resource management in semi-arid regions. As a complex social-ecological system, water demands in the Sonoran Desert from agricultural and urban users often conflicts with water needs for its ecologically-significant riparian corridors. To explore this system, we developed an agent-based model to simulate complex feedbacks between human decisions and environmental conditions in the Rio Sonora Watershed. Cognitive mapping in conjunction with stakeholder participation produced a Bayesian model of conditional probabilities of local human decision-making processes resulting to changes in water demand. Probabilities created in the Bayesian model were incorporated into the agent-based model, so that each agent had a unique probability to make a positive decision based on its perceived environment at each point in time and space. By using a Bayesian approach, uncertainty in the human decision-making process could be incorporated. The spatially-explicit agent-based model simulated changes in depth-to-groundwater by well pumping based on an agent's water demand. Changes in depth-to-groundwater feedback to influence agent behavior, as well as determine unique vegetation classes within the riparian corridor. Each vegetation class then provides varying stakeholder-defined quality values of ecosystem services. Using this modeling approach allowed us to examine effects on both the ecological and social system of semi-arid riparian corridors under various scenarios. The insight provided by the model contributes to understanding how specific interventions may alter the complex social-ecological system in the future.

Keywords: Bayesian cognitive mapping, agent-based modeling, sonoran desert, social-ecological systems, hybrid modeling

OPEN ACCESS

Edited by:
Lael Parrott, The University of British Columbia, Canada

Reviewed by:
Saumitra Mukherjee, Jawaharlal Nehru University, India
Gunter Spöck, Alpen Adria Universität Klagenfurt, Austria

****Correspondence:***
Aloah J. Pope, School of Natural Resources and the Environment, College of Agriculture and Life Sciences, University of Arizona, 1064 E Lowell St., Tucson, AZ 85719, USA
aloahpope@gmail.com

Specialty section:
This article was submitted to Environmental Informatics, a section of the journal Frontiers in Environmental Science

Received: *23 April 2015*
Accepted: *20 July 2015*
Published: *07 August 2015*

Citation:
Pope AJ and Gimblett R (2015) Linking Bayesian and agent-based models to simulate complex social-ecological systems in semi-arid regions. Front. Environ. Sci. 3:55. doi: 10.3389/fenvs.2015.00055

Introduction

As complex social-ecological systems, water demands in semi-arid deserts from agricultural and urban users often conflicts with water needs for its ecologically significant riparian corridors. Riparian corridors in semi-arid regions are oases of biological diversity; however, their appealing features also attract urban and agricultural developments. River diversions and groundwater pumping have significantly induced riparian vegetation change,

altered agricultural and urban behaviors, as well as diminished ecosystem services.

Traditional approaches to modeling semi-arid systems use hydrogeological models, water balance equations, and simplified water demand functions. Only recently has dynamic human behavior been incorporated into groundwater models through agent-based modeling (Feuillette et al., 2003; Pertoghese et al., 2013). Agent-based models (ABMs) are simulations of autonomous entities (agents) that respond heterogeneously to their environment (patches). Since agent-based modeling is a bottom-up approach, researchers can explore how intricacies in micro-scale behavior can influence macro-scale patterns (Schlüter et al., 2012). Agent-based modeling provides an excellent conduit in which to study coupled social-ecological systems since they are capable of addressing issues that make traditional approaches to studying social-ecological systems difficult, such as spatial and temporal complexity, non-linearity, and uncertainty (Schlüter et al., 2012). This is especially true for common pool resources, such as groundwater, since dynamic change is the result of interactions between social and physical systems.

A remaining concern in social-ecological system modeling is that uncertainty in social systems was not being addressed (Schlüter et al., 2012). One approach to address social uncertainty is the use of Bayesian cognitive mapping for modeling human decision-making. Drawing from the fields of Bayesian probability theory and cognitive mapping, Bayesian cognitive mapping attempts to create a probability of the likelihood a certain decision will be made (Sedki and de Beaufort, 2012). Cognitive mapping, or developing a network-based representation of an expert's cognition, produces a qualitative and static representation of decisions (Eden et al., 1992). Bayesian networks use probability theory to represent expert knowledge in situations in which knowledge is ambiguous or incomplete (Spiegelhalter et al., 1993). By embedding cognitive maps with probabilities, researchers can create a quantitative acyclic graph of the decision-making process, in which each decision relationships between components are expressed as conditional probabilities. By employing a Bayesian cognitive map as a decision model, researchers can incorporate inference and uncertainty in the complex decision-making process.

Linking Bayesian cognitive mapping with agent-based modeling techniques is a new approach in social-ecological system modeling. To understand how future climatic conditions may influence both social and physical systems, we developed a hybrid model of ranchers' livelihoods in Sonora, Mexico, where conflict over water use is a debated topic. A great unknown is how ranchers will react and adapt to challenges associated with impending climate change in the region.

Materials and Methods

Case Study

The Rio Sonora is a semi-arid watershed with north-south drainage among transverse mountains. Precipitation varies widely both annually and seasonally, averaging around 500 mm annually (Lizarrago-Celaya et al., 2010). Most of the precipitation occurs during the monsoon season in the form of thunderstorms in July and August. The remaining rain is dropped during winter as small events. Between the winter and monsoon rain events lies a dry, hot summer period. The basin experiences large variability in climate, with periods of drought and pluvial that can last several years to decades (Sheppard et al., 2002). While general circulation models predict a more arid climate in the future (Seager et al., 2007; Diffenbaugh et al., 2008), a dynamically downscaled model predicts increased precipitation (Robles-Morua et al., 2014), highlighting the uncertainties surrounding the effects of climate change for the Rio Sonora Basin.

Within the Rio Sonora's riparian corridor lay the lushest vegetation in the region, predominately cottonwoods and fodder crops. Outside the riparian corridor, the landscape is marginal (Liverman, 1990), including cacti, desert shrubs, and mesquite trees. In a harsh semi-arid climate with limited water resources, ranching remains one of the few remaining options to support livelihoods (West, 1993). The Rio Sonora Watershed produces one of the greatest volumes of livestock, primarily cattle, in Mexico. The carrying capacity, or number of cattle per hectare a rancher can successfully produce on his land, depends on the composition of its vegetation. Lush riparian vegetation can sustain more cattle per hectare than desert shrubs, while variation in precipitation can significantly alter the amount of natural fodder produced within a land cover type. In a drought year, reduced rainfall results in reduced carrying capacity. In such a case, a rancher must make decisions to prevent cattle loss, such as selling off cattle early or purchasing additional feed.

Water resources are an important component to ranching operations. In the past 30 years, groundwater has become vital, producing water when rainwater is limited. During that time however, overexploitation has reduced the groundwater so much that many wells have gone dry. If water is in short supply and forage is minimal on the range, a rancher may be forced to sell part of his herd. Under more favorable water and forage conditions, a rancher can wait to sell his cattle until prices peak in the beef market. The method a rancher adopts to manage his herd is one of many decisions heavily influenced by both water and money. In an uncertain political and environmental climate, concern rises about the future of the valued subculture of ranching in Mexico (Jordan, 1993).

As a case study for the Rio Sonora Watershed, a series of workshops were held in the city of Rayon, Sonora, Mexico. Founded in 1638, Rayon is one of the largest towns in the San Miguel River basin (**Figure 1**), a sub-basin of the Rio Sonora, with a population of approximately 1600 people (INEGI, 2010). For the city of Rayon, ranching is the most important component of the local economy. The purpose of this study was to assess the effects of rancher decision-making on groundwater pumping, depth-to-groundwater, riparian vegetation and ecosystem services along a 21 km stretch of the Rio San Miguel using a hybrid modeling approach.

Description of the Decision Model

Bayesian causal networks are composed of three features: nodes, links, and probabilities. Nodes represent system variables and can

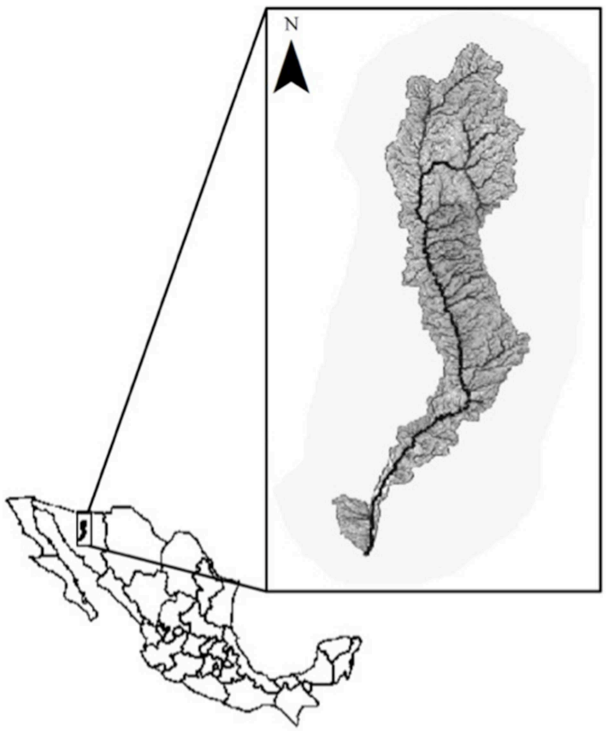


FIGURE 1 | The Rio Sonora Watershed lies within the Mexican state of Sonora.

be either discrete or continuous. For example, a discrete amount of rainfall could be below average, average, or above average whereas a continuous amount of rainfall can be binned across a full range of rainfall values. Links represent causal relationships between two nodes. A link, for example, from rainfall to crop yield would describe the effect of crop yield caused by the amount of rainfall. In Bayesian networks, the "effect" node is described as a child of the parent "cause" node. Within each node lies a set of probabilities that defines the relationships between parents and their children. A unique probability specifies the likelihood a child will be in a certain state for each combination of parent states. The set of probabilities are called conditional probability tables (CPTs). The incorporation of conditional probabilities tables transforms the structure into a functioning Bayesian network. By changing the state of a parent node, one can observe the model updating the probabilities of all decisions throughout the network using Bayes' theorem. Bayes' theorem is stated as the following equation:

$$P(A|B) = \frac{P(B|A)\,P(A)}{P(B)},$$

where A and B are events. $P(A)$ and $P(B)$ are the probabilities of A and B without regard to one another, $P(A|B)$ is the conditional probability of A given that B is true, and $P(B|A)$ in the conditional probability of B given that A is true. Because perfect knowledge of a parent's state is rare, probability distributions across a parent's state can be used as input in Bayesian cognitive maps, incorporating additional uncertainty.

The construction of a Bayesian cognitive map requires stakeholder involvement in two steps: development of an acyclic graph structure and data collection (Cain, 2001). A Bayesian cognitive map was developed in two stakeholder workshops (January 2013, 2014) held in Rayon, Sonora, Mexico after informed consent and Institutional Review Board (IRB) approval.

Development of the acyclic graph starts by choosing key variables. Among ranchers in Sonora, Mexico, the two critical issues are livelihood sustainability (profit) and water security (water availability). Next, states must be chosen for key variables. In this model, we chose below normal, normal, and above normal states for nodes "Profit" and "Water Availability" to encompass the entire range of conditions a stakeholder may perceive each year. For each key variable, children nodes were identified by stakeholders through unstructured surveys as important decisions affected by changing profit and water availability: "Purchase Additional Feed," "Change Herd Size," "Change Careers," "Increase Water Efficiency," or "Apply to Dig a New Well." An additional child node, "New Well Approved," was created to assess whether the National Water Commission would approve the new well application. Stakeholders were also asked to define what influenced their state of profit and water availability. Low profit was defined as when natural forage production was so low they must invest and purchase additional feed, while normal and high profit were defined by average and above average productivity of forage, respectively. Stakeholders defined low water availability when rainfall was below average, normal water availability when rainfall was average, and high water availability when rainfall was above average. High water availability was also perceived when wells were deeper than the depth-to-groundwater. Based on these responses, parent nodes of "Invest," "Rainfall," and "Well water" were added to the acyclic graph. With this information, we have the completed acyclic graph structure of Sonoran Ranchers decisions.

In order to populate the conditional probability table, a structured survey was distributed to ranchers to assess the likelihood a positive decision would be made within each iteration of contributing variables (below average/average/above average) along a Likert scale, from "Not At All Likely" to "Completely Likely." Each response on the Likert scale then translated to a probability from 0 to 100%. The average probability within each iteration of contributing variables was used to calculate the conditional probability table for each decision. Once compiled, the Bayesian cognitive map automatically calculates the probability a positive decision will be made based on user-input on parent nodes "Invest," "Rainfall," and "Well water." The resulting Bayesian cognitive map from rancher experiencing unknown profit, average rainfall, and a dry well is demonstrated in **Figure 2**. More detail on the development of the rancher Bayesian Cognitive map can be found in Pope and Gimblett (in press). The Bayesian cognitive models were developed using GeNIe (GeNIe, 2013), a Bayesian network software tool (http://genie.sis.pitt.edu/).

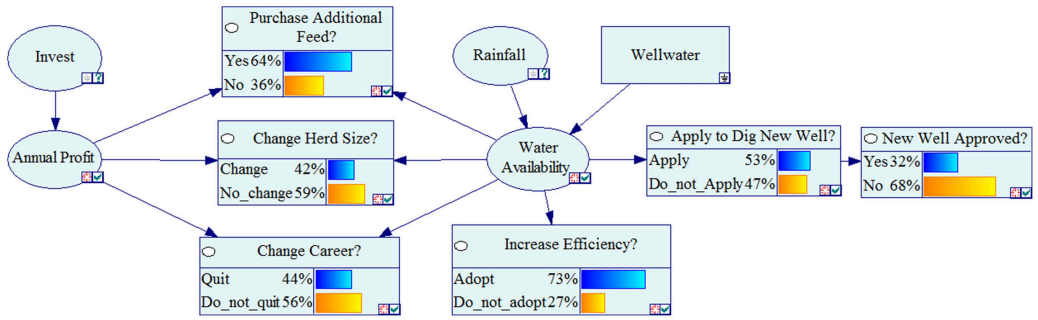


FIGURE 2 | Bayesian cognitive map of important decisions when an individual rancher's annual profit is unknown, his well is dry, and potential rainfall is under normal conditions (Pope and Gimblett, in press).

Bayesian cognitive maps represent a snapshot in time. If we are concerned about the future of a coupled natural-human system, we must find a way to incorporate temporal dynamics. One way to do so is to apply the same Bayesian cognitive map to consecutive years, but allow the variables at the top of the network to change. This could be natural change in a coupled natural-human system, such as fluctuating rainfall patterns, or changing node states based on the results of the Bayesian cognitive map from the previous time step.

Since our Bayesian cognitive map was developed from participation with individual ranchers, each agent in our agent-based model will also represent an individual rancher. Our Bayesian cognitive map also tells us that each rancher's value of water availability, annual profit, and herd size is independent of other ranchers. These variables are defined as rancher attributes. Each of these rancher attributes can be in three states: below normal, normal, and above normal. Value of water availability is determined by a combination of rainfall and well conditions. The conditions of the "Rainfall" node as well as the "Well water" node are used as inputs into the Bayesian cognitive map. If a rancher decides to dig a new well, the node "Well water" switches to below water table and the resulting new probabilities are calculated. The value of annual profit is determined by the output of the previous time step. If water availability is considered below normal and a rancher agent chooses to purchase feed, annual profit is set to below normal. The new value of profit is used as input into the next year's Bayesian cognitive map. Herd size is determined by the output of the Bayesian cognitive map. If a rancher decides to change herd size and water availability is below normal, herd size drops to below normal. If a rancher decides to change herd size and water availability is above normal, herd size improves to above normal. Each time an agent state changes, the Bayesian cognitive map updates to calculate new probabilities. All attributes are set to normal at initialization of the model.

The output of the Bayesian cognitive maps are used as the probability of approving the following decisions in the model: purchasing additional feed, changing herd size, and digging a new well. In order to incorporate uncertainty via Bayesian probabilities, all decisions use the following equations:

IF $P_R < P_D$, THEN $D = 1$;

IF $P_R > P_D$, THEN $D = 0$;

where P_R equals a random probability between 0 and 100 and P_D equals the calculated probability from the Bayesian cognitive map for making decision D. If the model calculates $D = 1$, a positive decision has been made.

Description of the Agent-based Model

The agent-based model consists of three interacting sub-models: social, hydrological, and ecological. The social sub-model includes urban, farmer, and rancher decision-making. The hydrological sub-model simulates spatial changes in depth-to-groundwater. The ecological sub-model demonstrates changes in riparian vegetation and subsequently, ecosystem services, along the Rio San Miguel. The ABM was developed in Netlogo, a multi-agent programmable modeling environment, because it is freely available and user-friendly for non-modelers (Wilensky, 1999, http://ccl.northwestern.edu/netlogo/).

Social Sub-model

Water demand is calculated for each agent type. Social agents function under the same basic formula:

Number X Per Capita Water Use = Water Demand,

wherein number equals the number of cattle and people for rancher and urban agents, respectfully; however, the calculation of each input differs.

In rancher agents, number of cattle, or herd size, is estimated by calculating the stocking rate of an agent's pasture. In order to make this calculation, we need three pieces of information: a land cover map, spatially-delineated pastures, and local values of animal unit months. A land cover map, or a categorical map of the physical material at the surface of the earth, was previously developed for 2004 by Hunt et al. (2004). Categories of land cover and land use are water, bare soil, shrubland, grassland, riparian mesquite, riparian woodland, sparse woodland, evergreen, subtropical shrub, and agriculture. Since spatial data on individual pastures in Rayon is not available, Theissen polygons were created in open ranchland surrounding wells that were designated for livestock use. Although a rough approximation, this approach allows us to estimate that spatial distribution of herd sizes across the landscape. Within each polygon, the total area of each land cover type was calculated using ArcGIS (ESRI, 2011). The total area of each land cover

type was then multiplied its annual animal unit month, or the number of animal units (1000 lb cow) a land cover type can sustain in a year, to calculate the total number of cows each different land cover area can support (Cuningham and Sampson, 2000; California: Forest and Range Assessment, 2003). Adding the number of cattle that should be stocked for each land cover type calculates the herd size of each pasture. Departure from the calculated herd size will be determined by the results of the Bayesian decision model. Herd size was multiplied with per capita water demand (12 gallons/head/day), and the number of days in each time period to estimate the total water demand for each rancher agent in each time period.

In urban agents, number is the population size (1599) of Rayon (INEGI, 2010). Population growth projections are used to simulate changes in population size over time. In Rayon, the last 10 years has seen a minimal change in population size (average annual growth rates = 0.03%). Population size was multiplied by per capita water demand and the number of days in each time period to estimate the total water demand in each time period.

In farmer agents, the number of hectares of farmland was calculated by digitizing fields in ArcGIS (ESRI, 2011). Individual fields were divided amongst wells that were designated for farmland based on proximity. Field area was multiplied by water demand/acre (1–4 acre ft/yr), percentage of the field that was irrigated, and the total number of days in each time period to estimate the total water demand by the field. Based on stakeholder interviews, farmers cease irrigation if there is little to no rainfall in the second half of the year. Therefore, farmer agents have no water demand if rainfall is modeled as "below average" in the second half of the year.

Hydrological Sub-model

Water demands calculated from the social sub-model are then used as inputs into the hydrological sub-model. The hydrological model uses this information to adjust the depth-to-groundwater in patches using the Theis (1935) equations:

$$s = \frac{Q}{4\pi T} W(u)$$

$$u = \frac{r^2 S}{4Tt}$$

$$W(u) = -0.577216 - \ln(u) + u - \frac{u^2}{2 \times 2!} + \frac{u^3}{3 \times 3!} - \frac{u^4}{4 \times 4!} + \ldots$$

where s is the drawdown, u is a dimensionless time parameter, Q is the pumping rate of the well, T and S are the transmissivity and storativity of the aquifer, r is the distance from the pumping well, t is time since pumping began, and $W(u)$ is the exponential integral. Hydrologic coefficients were parameterized to the Rio San Miguel Watershed. By using the Theis equation, the model can simulate the change in depth-to-groundwater for each agent's well within its radius of influence, or the spatial extent of influence from the pumping well. If the depth-to-groundwater falls below the depth of a well, the well becomes inactive and agents can no longer pump.

Ecological Sub-model

Changes in depth-to-groundwater calculated in the hydrological sub-model were used as inputs into the ecological sub-model. The relationships between depth-to-groundwater and riparian land-cover classes based on rooting depth were used to predict which riparian vegetation class could exist under changing patch conditions (Hough et al., in preparation). Specifically, marsh vegetation could be found if depth-to-groundwater was <1 m, young cottonwood in <3 m, adult cottonwood in <6 m, *Tamarisk* in <10 m, and mesquite in <53 m.

Although most research on habitat degradation has focused on ecological impacts, literature on human impacts has shown an increasing trend. Dubbed ecosystem services, human and ecological benefits from natural systems can be an instrumental influence on decision-making (Millennium Ecosystem Assessment, 2005; Reyers et al., 2009). Human-caused land-cover change has been recognized as one of the most important drivers of ecosystem change (Vitousek et al., 1997); however, little is known on how land-cover change influences quality of ecosystem services locally. In order to assess changes in ecosystems services, data was collected in a series of stakeholder workshops (January 2013; 2014). During the first stakeholder workshop, individuals listed important ecosystem services. From this list, the three most common responses that could be assessed on a cell-by-cell basis were chosen to be modeled in the ABM: agriculture, recreation, and aesthetics. At the second workshop, stakeholders ranked the quality of each ecosystem service on a five point Likert scale, five being the highest, for each riparian vegetation class (**Table 1**).

Initialization of the ABM

Initialization of the model creates the spatial environment for the social-ecological system model (**Figure 3**). Initial conditions of the spatial distribution of depth-to-groundwater were interpolated using 13 sample sites collected in December 2008. The riparian corridor is defined as a 120 m buffer along the Rio San Miguel. At the beginning of the simulation, the riparian corridor consisted of 4948.2 ha of mesquite, 281.7 ha of *Tamarisk*, and 434.7 ha of old cottonwood. Each pumping wells' attributes include location, well depth, the type of water user, the water user's water demand equation, and their water concession, or the maximum amount of water that can be pumped from that well

TABLE 1 | Quality ranking (1–5) of the top three ecosystems services for each riparian vegetation class.

Ecosystem service	Mesquite	*Tamarisk*	Adult cottonwood	Young cottonwood	Marsh vegetation
Agriculture	2	3	3	3	4
Recreation	2	3	4	4	4
Aesthetics	3	3	4	4	4

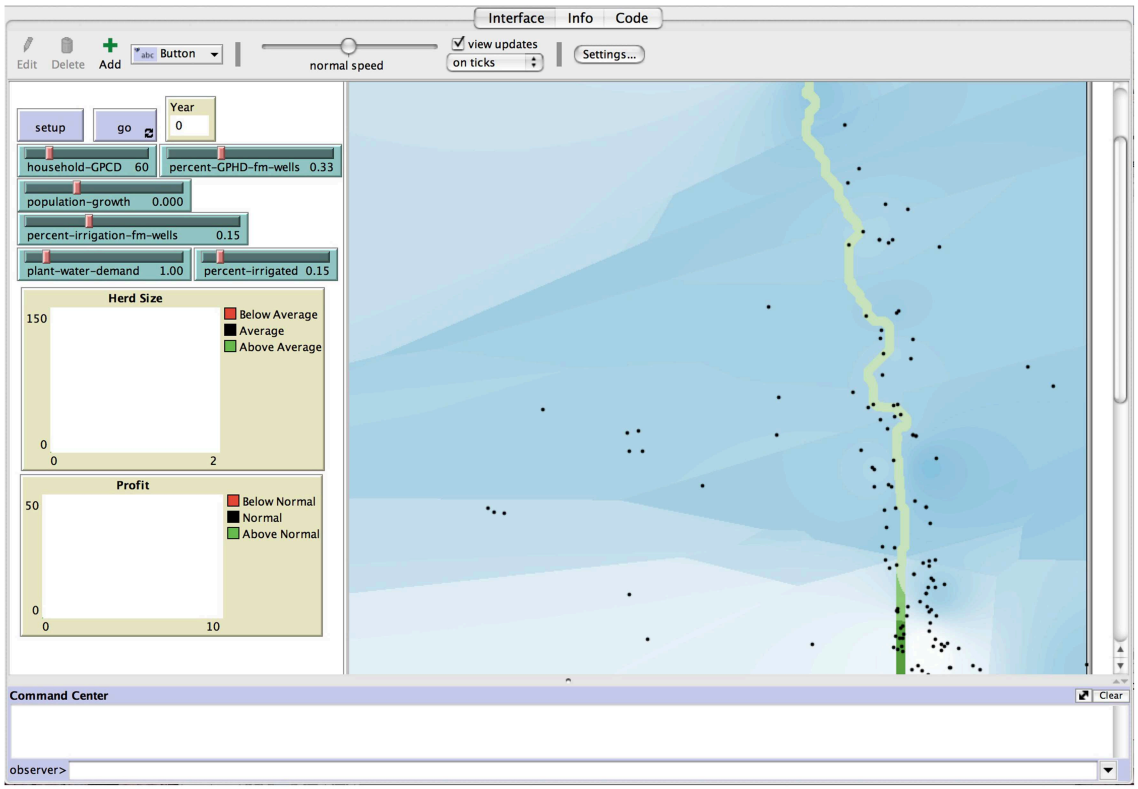


FIGURE 3 | Initialization of the social-ecological system model. Spatial distribution of depth-to-groundwater (light blue-shallowest, dark blue-deepest), riparian vegetation (light green-Mesquite, medium green-*Tamarisk*, dark green-Old Cottonwood), and pumping wells (black circles). Sliders bar present potential scenarios for stakeholder exploration of the system.

(Public Registry of Water Rights, Mexico.) At the initialization of the model, 10.3% of wells were dry.

Variations in rainfall were measured at the Rayon rain gauge for 1974–2012. Rainfall values that fell within the first quartile of the long-term rainfall dataset were considered "below average," while those that fell within the fourth quartile were considered "above average." Rainfall conditions were always considered "below average" for the first half of the year, since minimal rainfall combined with high evapotranspiration rates result in little to no infiltration before August.

The ABM has a spatial resolution of 30 × 30 m and a temporal resolution of 6 months.

Scenarios

Since the future of the local climate, specifically water security, is one of the greatest concerns in the region, the agent-based model is used to test how a potential extended drought may impact riparian vegetation and its ecosystem services. Using Netlogo's BehaviorSpace Tool, a total of 10 repetitions were simulated for two treatment scenarios: a period of average rainfall and a period of below average rainfall. The model was initialized at the end of 2008 and used measured rainfall values until the end of 2012, followed by 6 years of the treatment conditions. The effects of the social-ecological system model on rancher well-being was also monitored. At the end of the simulation, the distribution of profit and herd size (below average/average/above average) and the number of dry wells were recorded.

Results

Ecological

At the end of the simulation, the compositions of riparian vegetation classes were significantly different between rainfall scenarios (**Table 2**). The total of mesquite was significantly greater in the average rainfall scenario (5050.9 ± 2.8 ha) than in the below average scenario (4980.5 ± 6.3; $p < 0.01$). Likewise, the total are of *Tamarisk* was also greater in the average rainfall scenario (448.7 ± 2.3 ha) than in the below average scenario (410.2 ± 10.5 ha; $p < 0.01$). Significantly greater area of adult cottonwood was found in the below average scenario (273.9 ± 12.6 ha) than in the average rainfall scenario (164.8 ± 1.6 ha; $p < 0.01$). No riparian area was observed with either young cottonwood or marsh vegetation (**Figure 4**).

Social

The rainfall scenarios also had a significant effect on ecosystem service quality (1-lowest, 5-greatest; **Table 3**). Although both rainfall scenarios only produced below average (2) to average (3) quality agricultural area, the total area (ha) of average quality agricultural area was significantly greater in the below average

TABLE 2 | Total area (ha) of riparian vegetation classes produced in the average and below average rainfall treatment scenarios.

Vegetation class	Rainfall treatment		*p*-values
	Average	Below average	
Mesquite	5050.9 (2.8)	4980.5 (6.3)	<0.01
Tamarisk	448.7 (2.3)	410.2 (10.5)	<0.01
Adult Cottonwood	164.8 (1.6)	273.9 (12.6)	<0.01
Young Cottonwood	0 (0)	0 (0)	N/A
Marsh Vegetation	0 (0)	0 (0)	N/A

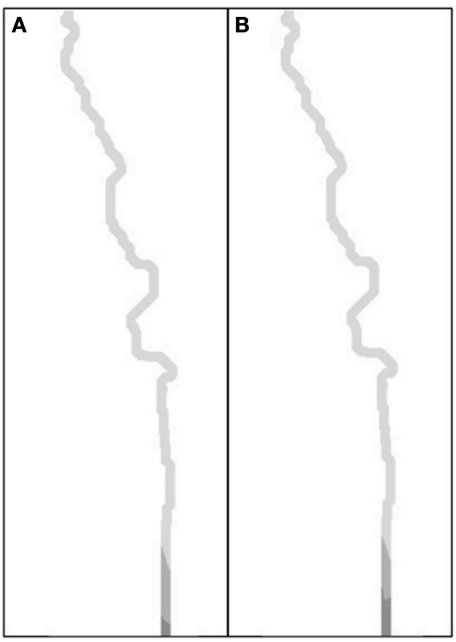


FIGURE 4 | Spatial distribution of riparian vegetation classes, Mesquite (light grey), *Tamarisk* (medium grey), and Old Cottonwood (dark grey), along the Rio San Miguel for average (A) and below average (B) rainfall treatments.

rainfall scenario (684.1 ± 6.3 ha) than in the average scenario (613.6 ± 2.7 ha; $p < 0.01$). The total area of below average quality was significantly greater in the average rainfall scenario (5050.9 ± 2.8) than in the below average scenario (4980.5 ± 6.3 ha; $p < 0.01$). Recreation area was found in below average (2), average (3), and above average (4) quality in both rainfall scenarios. The total area of above average quality of recreation was significantly greater in the below average rainfall scenario (273.9 ± 12.6 ha) than in the average rainfall scenario (164.8 ± 1.6 ha; $p < 0.01$). Significantly greater area of average quality recreation area was found in the average rainfall scenario (448.7 ± 2.3 ha) than in the below average scenario (410.2 ± 10.5 ha; $p < 0.01$). Significantly greater area of average quality recreation area was also found in the average rainfall scenario (4980.5 ± 6.3 ha) than in the below average scenario (5050.9 ± 2.8 ha; $p < 0.01$). Both rainfall scenarios produced average (3) and above average (4) quality of aesthetic area. Total area of above quality of aesthetics

TABLE 3 | Total area (ha) of differing quality values (1-lowest, 5-greatest) for agriculture, recreation, and aesthetics ecosystem services produced in the average and below average rainfall treatment scenarios.

Ecosystem service	Rainfall treatment		*p*-values
Quality	Average	Below average	
AGRICULTURE			
1	0 (0)	0 (0)	N/A
2	5050.9 (2.8)	4980.5 (6.3)	<0.01
3	613.6 (2.7)	684.1 (6.3)	<0.01
4	0 (0)	0 (0)	N/A
5	0 (0)	0 (0)	N/A
RECREATION			
1	0 (0)	0 (0)	N/A
2	5050.9 (2.8)	4980.5 (6.3)	<0.01
3	448.7 (2.3)	410.2 (10.5)	<0.01
4	164.8 (1.6)	273.9 (12.6)	<0.01
5	0 (0)	0 (0)	N/A
AESTHETICS			
1	0 (0)	0 (0)	N/A
2	0 (0)	0 (0)	N/A
3	5499.7 (1.6)	5390.7 (12.6)	<0.01
4	164.8 (1.6)	273.9 (12.6)	<0.01
5	0 (0)	0 (0)	N/A

was significantly greater in the below average rainfall scenario (273.9 ± 12.6 ha) than in the average rainfall scenario (164.8 ± 1.6 ha, $p < 0.01$). Significantly greater area of average aesthetic quality was found in the average rainfall scenario (5499.7 ha ± 1.6) than in the below average scenario (5390.7 ± 12.6; $p < 0.01$).

At the end of the simulation, a significantly greater percentage of wells in the average rainfall period (40.2%) than in the below average rainfall period (34.3%) were dry ($p < 0.01$). Herd sizes of ranchers between the below average and average rainfall scenarios were significantly different (**Figure 5A**). A significantly greater number of ranchers experienced below average herd sizes in the below average rainfall scenario (27.1 ± 1.1) than in the average rainfall scenario (13.5 ± 0.7; $p < 0.01$). Average herd sizes in the average rainfall scenario (48.1 ± 1.8) were significantly greater than in the below average rainfall scenario (44.5 ± 0.8; $p < 0.05$). Above average herd sizes in the average rainfall scenario (21.4 ± 1.4) were also significantly greater than in the below average rainfall scenario (11.4 ± 1.1; $p < 0.01$).

Profit of ranchers between the below average and average rainfall scenarios were significantly different (**Figure 5B**). A significantly greater number of ranchers experienced below average profit in the below average rainfall scenario (33.3 ± 1.9) than in the average rainfall scenario (13.1 ± 0.7; $p < 0.01$). Average profit in the below average rainfall scenario (32.7 ± 0.9) was also significantly greater than in the average rainfall scenario (28.6 ± 1.3; $p < 0.05$). Above average profit in the average rainfall scenario (28.6 ± 1.3) was significantly greater than in the below average rainfall scenario (17 ± 1.7; $p < 0.01$).

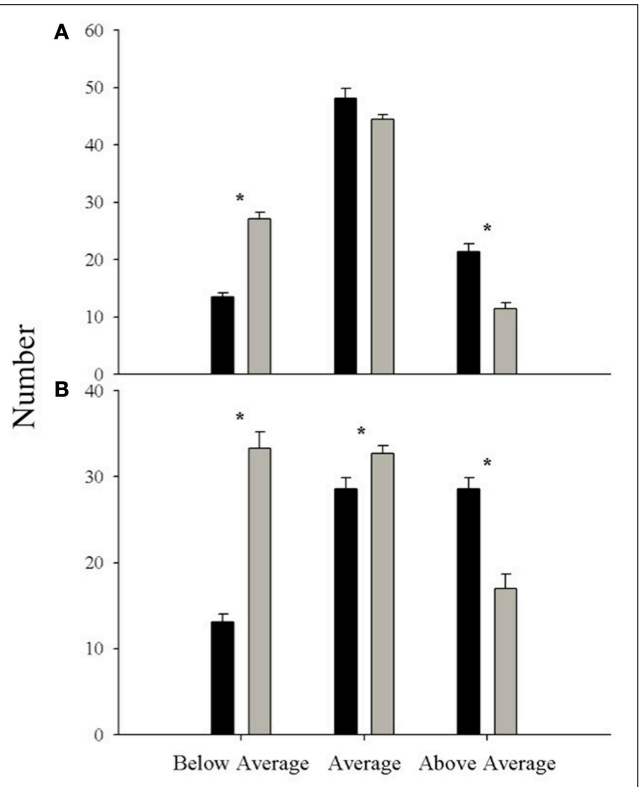


FIGURE 5 | Number of ranchers who observed below normal, normal, or above normal states of (A) herd size and (B) profits for a decade of average (black bars) and below average (gray bars) conditions (*$p < 0.05$).

Model Validation

In December 2012, depth-to-groundwater was collected at 12 sample sites. A Chi-squared Goodness of Fit test was used to compare modeled depth-to-groundwater with measured depth-to-groundwater at these sites. No significant difference was observed between modeled and measured depth-to-groundwater values ($\chi^2 = 10.96, p > 0.05$).

Discussion

The preservation of more hydrophilic riparian vegetation classes during a dryer period compared to a period of normal rain is counterintuitive. Traditional thinking suggests that in a dryer period of time, a farmer would increase irrigation on his land to maintain productivity. Conversely, farmers in Rayon cease irrigation if rainfall is unsubstantial because their crop is unsalvageable. The decreased irrigation efforts resulted in reduced pumping near river, maintaining sufficient depth-to-groundwater for hydrophilic vegetation. Although the model found a dry period to be beneficial to riparian vegetation, the same results may not be true at a geologic timescale since rains are needed for groundwater recharge. Limited irrigation resulting in fewer effects on the water table coincides with previous work conducted downstream (Palma et al., 2015). It should be highlighted however that in either scenario, the total area of old cottonwood decreased, while xeric vegetation classes of mesquite and *Tamarisk* increased.

Following a similar trend to the changes observed in riparian vegetation composition, all three ecosystem services experienced a greater shift to lower quality after normal rainfall compared to low rainfall. The purpose of translating riparian vegetation composition to ecosystem services is to demonstrate to stakeholders the effects of scenarios, in this case rainfall patterns, on the system with an anthropocentric unit. This model can be a much more powerful tool if a stakeholder can see how their decisions can feedback to impact themselves, either through the drying of wells or reduction in quality of agriculture. According to the stakeholders, most of the agricultural area in the riparian corridor is already below average. Continued deterioration of the riparian corridor could be disastrous for a community that relies on agriculture as its main economic activity.

Unlike the positive effect a dry period had on vegetation composition and ecosystems services, we found that a dry period resulted in an increased percentage of dry wells, a shift toward smaller herd sizes and an increased number of ranchers with below average profit. As is common with many other ecological problems, the model exposes a potential conflict with a scenario that benefits the environment at the expense of rancher well-being. Ideally, this model could be used to explore multiple scenarios, such as policy decisions, that either minimize or eradicate this conflict.

A novel approach in this study was the development of a hybrid model that utilized a Bayesian cognitive map as the decision model in the agent-based model. Instead of a traditional deterministic decision model, the Bayesian Cognitive Map elicited different probabilities of making a positive decision, which allowed us to incorporate real-world uncertainty of the human decision-making process in the agent-based model. This paper attempted to simulate the dynamic nature between water users and the water table by developing a tightly-coupled agent-based model of groundwater pumping by ranchers, farmers, and municipal water users. Since external stressors and internal conflicts are likely to worsen if current trends of increasing water demand and decreasing water supply in semi-arid regions continue, these types of models can be extremely useful tools for decision-makers, increasing the ability to make more-informed decisions to improve resilience of their systems.

Acknowledgments

We would like to thank the staff, students, and PIs of The University of Arizona, Universidad de Sonora, Colegio de Sonora, and University of Copenhagen working on the "Strengthening the Resilience of Arid Region Riparian Corridors" project. We particularly thank Dr. Alan Navarro, América Lutz, and Ryan Lee for help with stakeholder workshop coordination and execution, and Dr. Miguel Rangel Medina and Anho Antonio Taylor for help with developing the depth-to-groundwater map. This research was sponsored by a NSF Dynamics of Coupled Human-Natural Systems grant (DEB-1010495).

References

Cain, J. (2001). *Planning Improvements in Natural Resources Management; Guidelines for using Bayesian Networks to Manage Development Projects*. Wallingford: Institute of Hydrology.

California Department of Forestry and Fire Protection, Forest and Range Assessment Program. (2003). *The Changing California: Forest and Range 2003 Assessment*. Sacramento, CA: State of California Resources Agency.

Cuningham, G. L., and Sampson, M. W. (2000). *Grasshopper Integrated Pest Management user Handbook*. 1st. Edn. Washington, DC: USDA Animal and Plant Health Inspection Service.

Diffenbaugh, N. S., Giorgi, F., and Pal, J. S. (2008). Climate change hotspots in the United States. *Geophys. Res. Lett*. 35, L16709. doi: 10.1029/2008GL035075

Eden, C., Ackermann, F., and Cropper, S. (1992). The analysis of cause maps. *J. Manag.Stud*. 29, 309–323. doi: 10.1111/j.1467-6486.1992.tb00667.x

ESRI. (2011). *ArcGIS Desktop: Release 10*. Redlands, CA: Environmental Systems Research Institute.

Feuillette, S., Bousquet, F., and Goulven, P. L. (2003). SINUSE: a multi-agent model to negotiate water demand management on a free access water table. *Environ. Model. Softw*. 18, 213–427. doi: 10.1016/S1364-8152(03)00006-9

GeNIe. (2013). *GeNIe. Decision Systems Laboratory*, University of Pittsburgh. Available online at: http://genie.sis.pitt.edu/

Hunt, E. R. Jr., Jackson, T. J., and Yilmaz, M. T. (2004). *SMEX04 Land Cover Classification for Arizona*. Sonora.

INEGI. (2010). INEGI *Censo General de Poblacion y Vivienda*, Sonora: Instituto Nacional de Estadistica, Geografia e Informatica.

Jordan, T. (1993). *North American Cattle Ranching Frontiers: Origins, Diffusion and Differentiation*. Alberquerque, NM: University of New Mexico Press.

Liverman, D. (1990). *Drought Impacts in Mexico: Ritual, Capital, and Environmental Degradation in the Cordillera of Northern Luzon, 1900-1986*. Berkley, CA: University of California Press.

Lizarrago-Celaya, C., Watts, C., Rodrigues, J., Garatuza-Payan, J., Scott, R., and Saiz-Hernandez, J. (2010). Spatio-temperal variations in surface characteristics over the North American Monsoon region. *J. Arid Environ*. 74, 540–548. doi: 10.1016/j.jaridenv.2009.09.027

Millennium Ecosystem Assessment. (2005). *Ecosystems and Human Well-being: Synthesis*. Washington, DC: Island Press.

Palma, A., Gonzalez, F., and Cruickshank, C. (2015). Manageed aquifer recharge as a key element in Sonora River Basin management, Mexico. *J. Hydrol. En*. 20, B4014004. doi: 10.1061/(ASCE)HE.1943-5584.0001114

Pertoghese, I., D'Agostino, D., Giordano, R., Scardigno, A., Apollonio, C., and Vurro, M. (2013). An integrated modeling tool to evaluate the acceptability of irrigation constraint measure for groundwater protection. *Environ. Model. Softw*. 46, 90–113. doi: 10.1016/j.envsoft.2013.03.001

Pope, A., and Gimblett, R. (in press). "Linking participatory, Bayesian, and agent-based modeling techniques to simulate coupled natural-human system: a case study with ranchers in Sonora Mexico," in *Environmental Modeling with Stakeholders: Theory, Methods and Applications*, eds S. A. Gray, M. J. Paolisso, S. R. J. Gray and R. C Jordan (New York, NY: Springer Public Registry of Water Rights).

Reyers, B., O'Farrell, P. J., Cowling, R. M., Egoh, B. N., Le Maitre, D. C., and Vlok, J. H. J. (2009). Ecosystem services, land-cover change, and stakeholders: finding a sustainable foothold for a semiarid biodiversity hotspot. *Ecol. Soc*. 14:38. Available online at: http://www.ecologyandsociety.org/vol14/iss1/art38/

Robles-Morua, A., Che, D., Mayer, A. S., and Vivoni, E. R. (2014). Hydrologic assessment of proposed reservoirs in the Sonora River Basin, Mexico, under historical and future climate scenarios. *Hydrol. Sci. J*. 60, 50–66. doi: 10.1080/02626667.2013.878462

Schlüter, M., McAllister, R., Arlinghaus, R., Bunnefeld, N., Eisenack, K., Hölker, F., et al. (2012). New horizons for managing the environment: a review of coupled social-ecological systems modeling. *Nat. Res. Model*. 25, 219–272. doi: 10.1111/j.1939-7445.2011.00108.x

Seager, R., Ting, M., Held, I., Kishnir, Y., Lu, J., Vecchi, G., et al. (2007). Model Projections of an imminent transition to a more arid climate in southwestern North America. *Science* 316, 1181–1184. doi: 10.1126/science.1139601

Sedki, K., and de Beaufort, L. (2012). "Cognitive maps and Bayesian networks for knowledge representation and reasoning," in *24th International Conference on Tools with Artificial Intelligence* (Greece).

Sheppard, P. R., Comrie, A. C., Packin, G. D., Angersbach, K., and Hughes, M. K. (2002). The climate of the US Southwest. *Clim. Res*. 21, 219–238. doi: 10.3354/cr021219

Spiegelhalter, D., Dawid, A., Lauritzen, S., and Cowell, R. (1993). Bayesian analysis in expert systems. *Stat. Sci*. 8, 219–247.

Theis, C. V. (1935). The relation between the lowering of the piezometric surface and the rate and duration of discharge of a well using groundwater storage. *Am. Geophys.Union Trans*. 165, 519–524. doi: 10.1029/TR016i002p00519

Vitousek, P., Mooney, H., Lubechenco, J., and Melillo, J. (1997). Human domination of earth's ecosystems.*Science* 277, 494–499. doi: 10.1126/science.277.5325.494

West, R. (1993). *Sonora: Its Geographical Personality*. Austin, TX: University of Texas Press.

Wilensky, U. (1999). *NetLogo*. Evanston, IL: Center for Connected Learning and Computer-Based Modeling, Northwestern University. Available online at: http://ccl.northwestern.edu/netlogo/

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Environmental Science

ORIGINAL RESEARCH
published: 30 March 2016
doi: 10.3389/fenvs.2016.00019

Coupling Environmental, Social and Economic Models to Understand Land-Use Change Dynamics in the Mekong Delta

Alexis Drogoul[1,2]*, *Nghi Q. Huynh*[3,4] *and Quang C. Truong*[3,5]

[1] UMI 209 UMMISCO, IRD/UPMC/Sorbonne Universités, Paris, France, [2] ICT Lab, University of Science and Technology of Hanoi, Hanoi, Vietnam, [3] PDI-MSC, IRD/UPMC/Sorbonne Universités, Paris, France, [4] CTU/IRD JEAI DREAM, CICT, Can Tho University, Can Tho, Vietnam, [5] CTU/IRD JEAI DREAM, CENRES, Can Tho University, Can Tho, Vietnam

The Vietnamese Mekong Delta has undergone in recent years a considerable transformation in agricultural land-use, fueled by a boom of the exportation, an increase of population, a focus on intensive crops, but also environmental factors like sea level rise or the progression of soil salinity. These transformations have been, however, largely misestimated by the 10-year agricultural plans designed at the provincial levels, on the predictions of which, though, most of the large-scale investments (irrigation infrastructures, protection against flooding or salinity intrusion, and so on) are normally planned. This situation raises the question of how to explain the divergence between the predictions used as a basis for these plans and the actual situation. Answering it could, as a matter of fact, offer some insights on the dynamics at play and hopefully allow designing them more accurately. The dynamics of land-use change at a scale of a region results from the interactions between heterogeneous actors and factors at different scales, among them institutional policies, individual farming choices, land-cover and environmental changes, economic conditions, social dynamics, just to name a few. Understanding its evolution, for example, in this case, to better support agricultural planning, therefore requires the use of models that can represent the individual contributions of each actor or factor, and of course their interactions. We address this question through the design of an integrated hybrid model of land-use change in a specific and carefully chosen case study, which relies on the central hypothesis that the main force driving land-use change is actually the individual choices made by farmers at their local level. Farmers are the actors who decide (or not) to switch from one culture to another and the shifts observed at more global levels (village, district, province, region) are considered, in this model, as a consequence of the aggregation of these individual decisions. The central component of our hybrid model is then an agent-based model of farmers, provided with a sophisticated mechanism of decision-making that is influenced, at different degrees, by their perception of the contexts in which they act or interact with other actors. The economic context, accessible by them through the market prices of crops, plays a role, as well as the changes observed or forecasted in their physical context (land-cover changes, salinity rise) or the decisions made by others in their social

OPEN ACCESS

Edited by:
Christian E. Vincenot,
Kyoto University, Japan

Reviewed by:
Costica Nitu,
University Politehnica of Bucharest, Romania
Qing Tian,
George Mason University, USA

***Correspondence:**
Alexis Drogoul
alexis.drogoul@ird.fr

Specialty section:
This article was submitted to Environmental Informatics, a section of the journal Frontiers in Environmental Science

Received: *29 June 2015*
Accepted: *02 March 2016*
Published: *30 March 2016*

Citation:
Drogoul A, Huynh NQ and Truong QC (2016) Coupling Environmental, Social and Economic Models to Understand Land-Use Change Dynamics in the Mekong Delta. Front. Environ. Sci. 4:19. doi: 10.3389/fenvs.2016.00019

context (neighbors, family members, opinion leaders). The model of farmers is coupled, through this decision-making mechanism, with other independent sub-models, each of them carrying out a realistic description of one of these contexts. Since the dynamics depicted in these sub-models obey to different logics, operate at different scales and rely on different data, they are represented using appropriate modeling techniques: the spatial model is based on GIS information on parcels, soils, and rivers; a cellular automaton is used to account for the evolution of land-cover changes and the diffusion of salinity; an aggregated mathematical model represents the fluctuation of prices on the regional and national markets; and a graph-based social network model is used to represent familial networks of influence. Beyond the descriptions of these models, the paper is organized around a discussion about the two main outcomes of this research work. The first one is applicative: the way we have calibrated, coupled together, and experimented in different scenarios these five models is presented and we show that some findings obtained with the resulting hybrid model could not have been obtained with more traditional techniques. The second one is methodological: the underlying co-modeling architecture we used for declaring and running this assembly of heterogeneous models, implemented in the GAMA modeling and simulation platform, is presented and we show how it can be generalized to arbitrary hybridizations of models.

Keywords: integrated model, land-use change, hybrid modeling, agent-based modeling, models coupling, cognitive model, Mekong Delta

1. INTRODUCTION

1.1. Context: Land-Use Change in the Vietnamese Mekong Delta

The region of the Vietnamese Mekong Delta (VMD), composed of 13 provinces including a municipality and was home of approximately 18 million of inhabitants, was by far the most productive region of Vietnam in agriculture and aquaculture in 2014. In terms of rice production, for instance, 47% of the cultivated areas in Vietnam were situated in the VMD, and they outputted 54% of the total production; in terms of aquaculture, 2/3 of the Vietnamese production originated from the VMD. According to Young et al. (2002), these performances, which have roughly tripled in the last 30 years in all sectors, have fueled the boom of the Vietnamese exports of agricultural products (especially rice, shrimps, and fruits). This spectacular rise is due to a number of factors: a better economic environment (thanks to reforms more favorable to the private sector), the adoption of modern techniques (fertilizers, mechanical harvesting, progresses in aquaculture), yield improvements, improved irrigation and drainage, among others.

This growth has logically been accompanied by a deep transformation of the agricultural land-use. However, other factors, like the sea level rise, the general urbanization of the country or the progression of soil salinity (Smajgl et al., 2015), have also played a role in this transformation, and it is not trivial to sort out its different causes. In a country like Vietnam, this difficulty raises some concerns because agricultural land-use is traditionally strictly planned under the control, and following the national circulars, of the Ministry of Natural Resources and Environment (MONRE, 2009; VNA, 2013). Plans are produced every 10 years and readjusted at mid-term using a land-use inventory in order to rectify divergences with the reality. This actually results in two five-years long plans, detailed down to the level of provinces, that both recommend a given distribution in terms of land-use and cultivation types, but also schedule national and provincial investments (irrigation infrastructures, protection against flooding or salinity intrusion, transportation infrastructures, and so on) based on this distribution. In an ideal situation, where every province would follow the plan, there would not be any difference between the recommended distribution of land-use and its forecast. However, during the period covered by the latest plan (from 2000 to 2010), the planned—and then expected—distribution has been systematically offset, sometimes by an important margin, from the reality of land-use as measured by remote-sensing techniques. In **Figure 1**, for instance, it is easy to see that land-use has had a trend to shift from rice to shrimps. The surface dedicated to rice crops has strongly decreased, while the one dedicated to shrimp aquaculture has increased.

At the levels of provinces, the same shift can be observed. For instance, in the province of Ben Tre (see **Figure 2**), a total cultivated area of 175,824 ha was planned for 2010, but it eventually reached 179,671 ha. This average value (which gathers all kinds of agricultural activities) might be seen as a not so significant change at the macro-level, but profound divergences can be unveiled when studying the numbers in more details and especially the spatial inscription of the changes. For example, the rice area expanded to 38,000 ha, compared to a planned 30,000 ha (+ 27%); the surface devoted to aquaculture, which was supposed to reach 39,200 ha, only reached 30,289 ha (−23%); and the forests, expected to cover 350 ha (PCBT,

2011), remained at a low 1.30 ha. At the scale of villages, the divergences are sometimes even more spectacular, as entire villages have shifted to one type of production to another without any incentives from the plans—and without, of course, following their recommendations. To understand the importance of these changes, we consider in this paper a specific case study, which will be the focus of the remaining sections. This case study comprises five villages situated in the middle of the Thanh Phu district (Ben Tre province). They have been carefully chosen as they exhibit a variety of land cover characteristics while remaining geographically close to each other, at least close enough to reasonably allow us to consider that the farmers living in these villages share common "cultural traits" and traditions. As shown on **Figure 3**), four of these villages are situated in areas invaded by brackish water, while the remaining one benefits from fresh water due to its inclusion in a dike-protected area; the three villages to the west are bordering the river, the easternmost is located near a mangrove and the remaining one has no direct access to the river.

On **Figure 4**, we show the results of a study we conducted on these five villages in order to assess the shift of land-use between, on one hand, the two projections for year 2010 of the plans produced in 2000 and 2005 and, on the other hand, the actual land-use map in 2010 (PCBT, 2011). Changes are measured using a Fuzzy Kappa indicator, a variant of Kappa (Cohen, 1960) that provides a measure close to how humans compare maps. The darkness of the areas on the two right-hand maps is proportional to the change in land-use. It is easy to see that, while the average changes for the whole province may not be spectacular, they translate into local changes that mark complete shifts from one type of production to another. With respect to this, the plan published in 2000 is completely wrong in its projections (almost all parcels have changed) and the rectified plan published in 2005, while correct for the most part, completely misses the shifts in two villages and along the canals.

This finding would not be a concern if these projections were not also used to plan long-term investments. When a plan projects an extension of rice areas in a particular place, this usually is accompanied by a planning of irrigation infrastructures, dikes, or other facilities supposed to foster the cultivation of rice. But what happens, for instance, when the farmers living in this particular place take the decision to massively switch from rice cultivation to shrimp aquaculture? Would the investments be canceled, or, even worse, maintained and then play a counter-productive role? Land-use planning, as it goes in Vietnam, and despite its inherent interest, is not precisely the most appropriate way to react to such changes in a timely manner. This implies two possible views on how to better connect investments with the reality of land-use. The first option is to completely abandon the use of a global plan as it is designed now, and opt for a more decentralized and reactive land-use planning strategy, which would allow districts or even villages to become responsible for their own local land-use allocation strategy (or the best part of it). Considering and discussing the potential strengths (better and more timely adaptation) and drawbacks (possible lack of coordination between adjacent areas) of this option clearly goes beyond the scope of this article and would be, in any case, ultimately, a political decision. A second option would be, therefore, to investigate how the predictions/recommendations

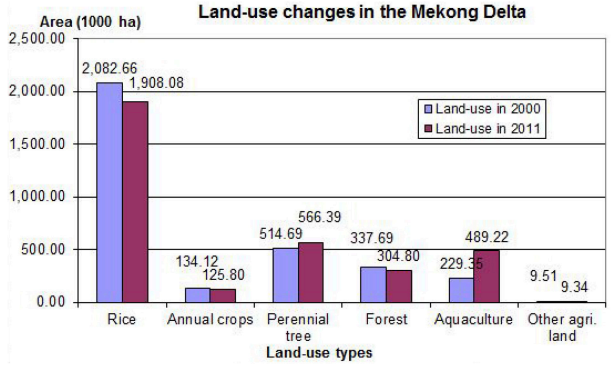


FIGURE 1 | Land-use area in the Mekong Delta in 2000 and in 2011. Source: Vietnamese General Statistics Office (VGSO, 2000) and Ministry of Natural Resources and Environment (MONRE, 2012) of Vietnam.

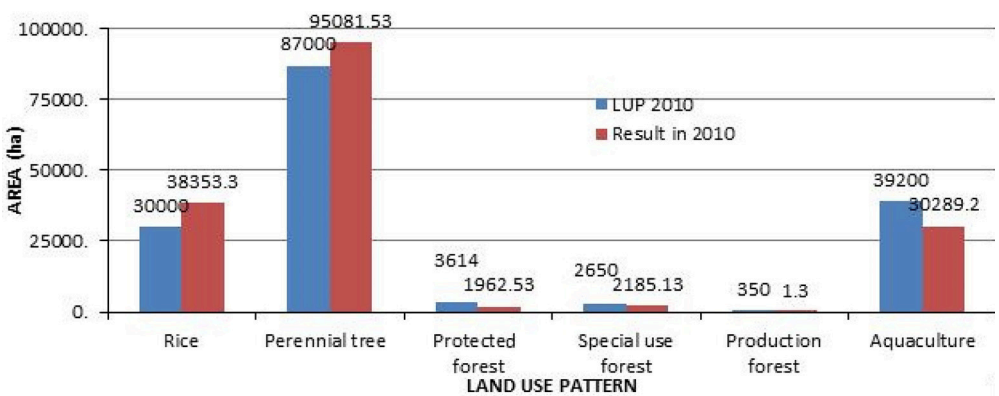


FIGURE 2 | Land-use planning assessment of Ben Tre province in 2010. Source: Department of Environmental and Natural resources of Ben Tre province, Vietnam (PCBT, 2011).

of the plans could be made more accurate, so as to support planners in designing more realistic plans. This is the option we take in this article.

One major difficulty in this option is that we do not have access to the methodology used by planners to produce their plans. Their predictions, moreover, possess an ambiguous status, as they serve both as forecasts (what they expect the land-use to become in 10 years) and recommendations (what they want the land-use to become in 10 years). As such, they are probably designed using a mixture of science and politics: models based on past performances certainly play a role, together with data collection and aggregation (from villages to provinces), but it is difficult to

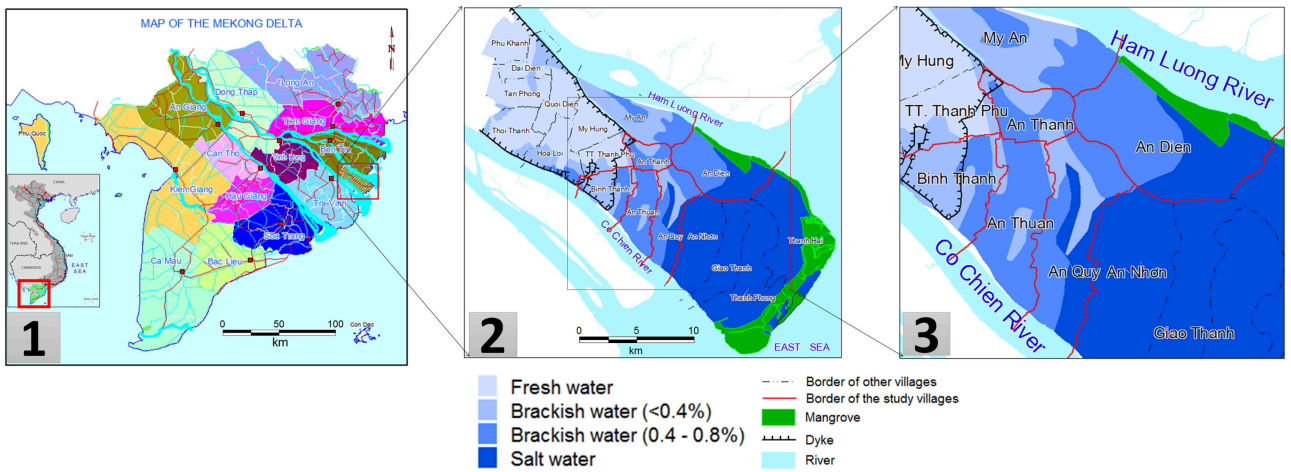


FIGURE 3 | Maps of the five villages of Thanh Phu district in the Mekong Delta. Source: Combined from the source of Land resource department, Can Tho University, Vietnam (Vo and Le, 2006) and the Department of Environmental and Natural resources of Ben Tre province, Vietnam (PCBT, 2011).

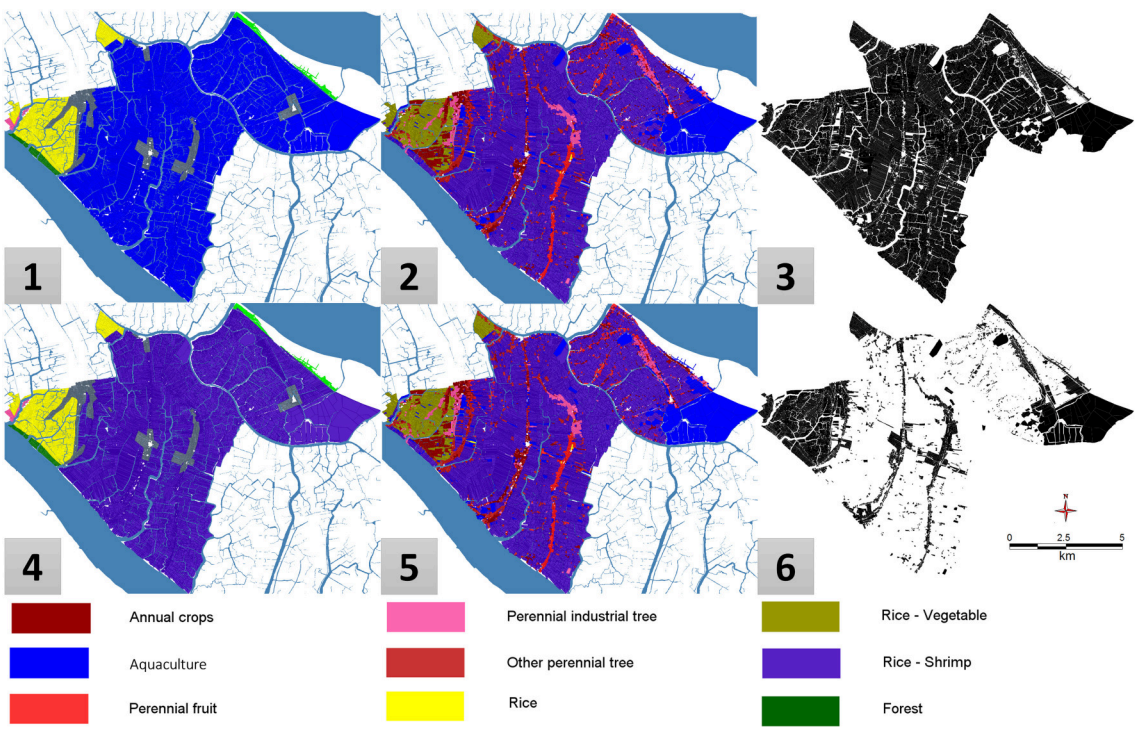


FIGURE 4 | Comparison of land-use planning map and land-use map in 2010 of 5 villages of Thanh Phu district, Ben Tre province. (1) Land-use planning map 2010 (planned in 2001), (4) Land-use planning map 2010 (modified in 2005), (2), (5) Land-use map in 2010, (3) Fuzzy Kappa map between 1 and 2, (6) Fuzzy Kappa map between 4 and 5.

sort out what belongs to political or economic wishes and what is the fruit of a scientific approach to build accurate projections. In that respect, it is appropriate to start by the beginning and try to understand—or at least draw serious hypotheses about—the causes behind the current trends in land-use change in the VMD.

As pointed by Lambin and Geist (2007) and JunJie (2008), the dynamics of land-use change at a regional scale results from the interactions of various actors and factors at different scales, among them institutional policies, individual farming choices, land-cover and environmental changes, economic conditions, social dynamics, just to name a few. Understanding its evolution, for example, in our case, to better support future planning efforts, requires to use models that can represent the individual contributions of each of these actors, and of course a description of how these models interact, based on a more general theory of the interplay of the different causes.

1.2. Proposal: Design of an Integrated Hybrid Model

We address this question through the design of an integrated model of land-use change in the case study presented above (**Figure 3**), which relies on the central hypothesis that the main force driving land-use change is the individual choices made by farmers at their local level. Farmers are the actors who decide (or not), given their knowledge, preferences or interactions, to switch from one culture to another. In the model, the shifts observed at more global levels (village, district, province, and region) are then a consequence of the aggregation—and interactions–of individual models of farmers' decisions.

Given the multiplicity of factors that can be represented in such a model, we need to rely on a very flexible modeling architecture in order to easily combine models and test their different combinations. We need also to avoid reinventing the wheel: the reuse (and comparison) of legacy models of specific factors (economic models, for instance), when they are available, is clearly a necessity, which in turn implies that we use a framework capable of supporting and coupling models expressed using multiple modeling paradigms. These two reasons have led us, among different choices, to rely on the multi-agent based "co-modeling" architecture of Huynh et al. (2014) and to implement it in the open-source GAMA platform (GAMA, 2007), which possesses the advantage of providing, out of the box, a first-class support for spatial data and spatial analysis, as well as a componential architecture that supports the concurrent use of multiple formalisms within a model (Grignard et al., 2013). The methodology we followed is discussed in this paper and represents one of the two outcomes of this research, which we hope will be seen as generic enough to be reusable by modelers in other contexts.

The second outcome is the integrated model itself. As stated above, the central component of this model is a model of farmers, provided with a sophisticated mechanism of decision-making that can be influenced, at different degrees, by their perception of the contexts in which they act or interact with other actors. Three of these "contexts" are described in the paper: the economic context, represented by a model of market prices of the agricultural products; the environmental context, represented by several models of the changes observed or planned in the farmers' physical environment (e.g., natural or man-made land-cover changes, rise of salinity, and so on); and finally the social context in which the farmers interact with other human actors (e.g., neighbors, family members, opinion leaders, policy makers), represented by a model of the dynamics of their social network.

The sub-model of farmers uses an agent-based BDI formalism (Rao and Georgeff, 1991) to describe their decision-making process. This sub-model is then coupled with other sub-models, each of them built using a suitable formalism for carrying out a realistic description of the context they represent. As a matter of fact, the dynamics present in these sub-models obey to different logical processes, operate at different scales and rely on different datasets, so they need to be designed with specific modeling techniques. In its most stable form, the one discussed in this paper, the integrated model makes use of a GIS-based model of parcels, soils, and rivers; a cellular automaton is then used to account for the evolution of land-cover changes and the diffusion of salinity; an aggregated mathematical model represents the fluctuation of prices on the regional and national markets; and, finally, a graph-based social network is used to represent familial networks of influence.

The validation of the resulting model makes full use of the "co-modeling" architecture to produce meaningful outcomes. On the methodological side, we show how to conduct experiments that use different combinations of these sub-models, exchange them with alternative models, or explicitly modify their relationships and interactions, thereby providing modelers with an extraordinary flexibility for conducting controlled experiments and testing various hypotheses. On the applicative side, we show how certain combinations of causes (represented by corresponding combinations of sub-models) can provide new insights on the land-use processes at play in the VMD and discuss how these insights could be reused to improve land-use planning in the future.

2. INTEGRATED MODELS OF LAND-USE CHANGES

2.1. State of the Art and Applications of Land-Use Change Models

Land-use and land-cover changes (LUCC) models have a long and rich history in spatial modeling (Parker et al., 2002). They can be broadly classified into two, not necessarily exclusive, categories: in the first one, we find models we might describe as "descriptive," whose primary concern is not to represent realistic mechanisms but to faithfully reproduce high-level dynamics of land-use change (Lambin, 1997); these models are essentially used for monitoring or prediction. They habitually rely on a discretization of the space, identifying "parcels" or "patches," and aggregating the influence of several drivers into high-level rules (equations in mathematical models Serneels and Lambin, 2001, transition rules in Cellular Automata models Zhao and Peng, 2012; Subedi et al., 2013), transition functions or matrices in Markov Chain models (Kemeny and Snell, 1983) to

describe how these patches change state over time. The second category groups "explicative" models, seeking to generate realistic dynamics of land-use change based on a detailed and faithful representation of the possible drivers of the changes; these models are more explicitly targeting decision-support, providing for example a support to "what-if" experiments (Trickett and Trafton, 2007). In the last 10 years, thanks to the generalization of developments in agent-based modeling (ABM, Michel et al., 2009), the design of large-scale, data-driven, individual-based models has even made possible the modeling of individual land-use change decisions and the simulation of their collective effects at higher geographical scales (Parker et al., 2003, Valbuena et al., 2010, Mena et al., 2011, Bakker et al., 2015), albeit with simple behavioral models. It is only recently that more sophisticated representations of the stakeholders' behaviors have emerged, again based on ABM, for example with the use of the the cognitive formalism known as BDI (Belief Desire Intention; Taillandier and Therond, 2011, Truong et al., 2016), which allows to model human actors as agents with beliefs concerning the world (i.e., the context outside the agent but accessible by it through its perceptual mechanisms), desires for their long term activities (i.e., goals), and intentions (i.e., the short-term actions that need be applied to fulfill their desires).

These two categories of LUCC models have long remained somehow separated, firstly because they had different objectives and, secondly, because they relied on different modeling paradigms. But their objectives are, actually, quite convergent: explaining large-scale changes in land-use without taking into account human activities, and especially their variability over time, casts doubt on the ability of the first category of models to produce realistic predictive models; conversely, the "environment" of the agents cannot be considered solely as a product of their activity. Especially in countries (like Vietnam) that are threatened by climate change, land-cover changes as well as other stressors (economy, innovations) need to be taken into account—and the first category of models can become essential in that respect, in conjunction, of course, with models of the second category. These reasons, and others, have led to the emergence of a new type of models, known in the literature as "hybrid models" (Parrott, 2011), which basically combine different sub-models into one to produce richer insights, at the price, however, of an increased complexity: a complexity in the design of these combinations of models; and a complexity in their exploration. LUDAS (Le et al., 2008), built in NetLogo, or Aporia (Murray-Rust et al., 2014), built on top of the Repast Semaphore platform (North et al., 2006), are two good examples of this trend, and underline both the potentialities of this new approach to modeling, but also its drawbacks, which are summarized in the three following points:

- Lack of "necessary complexity": despite their goal, most of the hybrid LUCC models (Zhao and Peng, 2012, Subedi et al., 2013) remain quite simple in that they do not treat the different dynamics equally. When the environmental factors are represented with great detail, the behavior of stakeholders remains simple (e.g., Lambin and Geist, 2007). And when this behavior is modeled using advanced mechanisms, like the BDI architecture (Taillandier and Therond, 2011, Truong et al., 2016), it is the environment that lacks a proper representation. Of course, everyone prefers to deal with simple models, more tractable from a simulation point of view, but it is their own complexity that, sometimes, provide LUCC models their heuristic power in terms of decision-support (Edmonds and Moss, 2005).
- Lack of genericity: until now, despite the similarity between the objects, processes or actors that can be found across different LUCC case studies, when a model is developed for one case study, it usually remains specific to it. In particular, no real effort is made to generalize and share methodological outcomes (architectures, sub-models, patterns), because they rely on assumptions that cannot be easily translated to other contexts: Aporia (Murray-Rust et al., 2014), for instance, is dedicated to European farmers and their environment, while LUDAS (Le et al., 2008) remains restricted to highlands and mountainous areas in Vietnam.
- Lack of flexibility: finally, with the notable exception of Aporia (which partially supports the change of sub-models), most of the existing hybrid LUCC models are designed as a static composition of carefully chosen (or written) sub-models. This does not allow to consider sub-models as possible parameters of experiments, something that can be necessary to explore different configurations or scenarios. In our case, given the variety of factors identified, explaining LUCC in the Mekong Delta with an integrated model requires that we explore several causes, some of them represented not only by parameters, but by entire sub-models or specific combinations of them. The underlying software architecture need then to provide a high degree of modularity and flexibility, in order to easily add, remove or change sub-models, but also to change their way of interacting, exchanging information and contributing to the overall outcome.

2.2. Models Coupling in Socio-Environmental Modeling

A handful of research works have already addressed the problem of coupling multidisciplinary models in the general domain of socio-environmental modeling. We can classify these proposals in two extreme groups: on one hand, the ones that propose a strong coupling between a given set of well identified sub-models, supporting complete interactions between them; on the other hand, a weak coupling which mainly relies on an interface supporting data exchanges between multiple models described by their input and output parameters. Both groups have their advantages and drawbacks: while the former can produce really integrated solutions, it lacks flexibility in that sub-models cannot easily be replaced; and the latter, more flexible in theory, imposes an interface which also limits the types of sub-models that can be used.

An interesting example of weak coupling is the work of Nicolai et al. (2011), who have coupled UrbanSim with MATSim in order to produce an integrated model of urban mobility. UrbanSim contains information on residential locations, workplaces and

urban development, while MATSim provides access to large-scale models of land-use, transportation network and economic dynamics. The two models are synchronized through the exchange of specific data (mobility needs from UrbanSim to MATSim, accessibility indicators in the reverse way). The literature provides other examples, like Yez et al. (2008), with an integrated model of a marine environment composed of local ecosystems of pelagic species, which adds an explicit translation between models at different scales, or Steiner et al. (2009), which couples a community land model with the regional climate model of the West African monsoon, and of course others. Simple exchanges of data is in general not sufficient in weak coupling approaches because the sub-models, which can be legacy models that cannot be modified, can operate at different scales of space and time and can also have different objectives (Rajeevan and Nanjudiah, 2009). There is a need, then, to provide a form of translation (Moreira et al., 2009 calls it a "coupler") that takes into account the peculiarities of each model, often linked to the formalism in which it is described (e.g., agent-based modeling, discrete event, continuous equations). Numerous works have therefore addressed the problem of combining or coupling models described using different paradigms, like for example Rochette et al. (2012) on the coupling of hydrodynamic continuous models and individual-based models, Quesnel (2005) on the coupling of physical and social models, Rousseaux et al. (2012) on the coupling of continuous and discrete formalisms in ecological modeling or Nguyen et al. (2008) on the coupling between agent-based models and equation-based models through the use of intermediate graph-based representations.

But even complex exchanges of data or parameters between models cannot solve the problems raised by the necessity, sometimes, to control how the different sub-models are executed or simulated within an integrated model. In that case, a stronger coupling, which involves some kind of functional control in addition to the exchange of data, must be used (Huang et al., 2008). Strong coupling requires relying on some kind of operational architecture or framework, which can provide a way to express the control over sub-models. This framework can either use an existing modeling paradigm (like Li et al., 2013, which proposes an "agent-centered" approach in which different modeling formalisms can be translated to individual or agent-based models) or make use of specialized software architectures dedicated to the functional coupling of models, like the High-Level Architecture (HLA; Dahmann and Morse, 1998; Dahmann et al., 1998; Hill, 2002; Kim, 2005; SISO, 2010), the Discrete Event Systems (DEVS Zeigler et al., 1997; Hild, 2000; Vangheluwe, 2000), of the Functional Mock-up Interface (Blochwitz et al., 2012; Bertsch et al., 2014). HLA is a standard and a software framework that has been originally designed as an infrastructure dedicated to the synchronization of simulators whenever they exchange data. The principle of HLA is to consider each simulator as a federate in a Federation, itself defined by three core elements: an object model which describes the simulators and the federation in terms of input and outputs, an interface specification within the RTI (Runtime Infrastructure), and a set of control rules (which more or less represent the behavior of the integrated model resulting from the coupling of the simulators). Despite its numerous advantages in the handling of completely heterogeneous simulators, the main problem of HLA remains its complexity for non-computer scientists, which makes it out of reach of most modelers. DEVS is initially a formalism proposed to model discrete event systems. Its interest for models coupling lies in its recursive definition: a model described in DEVS (with its set of inputs, outputs, states, etc.) can be considered as either "atomic" or "coupled," in which case it is described with additional features like the models it couples (which can themselves be atomic or coupled), a translation function and the influences between these models. DEVS is particularly elegant in its concepts, and very well adapted for building composite models, but relies on an internal description of models that is not easy to provide for stochastic models or complex legacy models. Finally, FMI is an industrial standard for co-simulation, where each sub-model or simulator is wrapped and exposed to the others using a functional interface which specifies how it can be accessed or manipulated.

Despite being operational, HLA, DEVS, and FMI are almost completely absent from the literature on hybrid modeling (Parrott, 2011) for decision-making in environmental or ecological issues. The main reason is that these solutions address the software engineering side of the problem of coupling and require modelers to learn and master quite complex techniques. In the worst cases, sub-models even need to be completely rewritten or redesigned to adapt to the interface required by the coupling infrastructure. Furthermore, these techniques use formalisms and languages that are different from the ones commonly used to build socio-environmental models (like agent-based models, cellular automata, or mathematical models), which is actually a source of confusion for many researchers. That is the reason why most of the existing integrated models make use of *ad-hoc* techniques, designed for the purpose of providing an infrastructure for one specific set of sub-models and unfortunately almost impossible to reuse in other contexts. For example, in Nicolai et al. (2011), the connection between UrbanSim et MATSim is completely bound to this particular instance of integration and there is no way one can reuse it, adapt it or extend it, for instance with an environmental model.

We propose in this paper another way of seeing the problem of coupling, detailed in Section 4.1. It is conceptually close to the notion of "multi-model ecologies" promoted by Bollinger et al. (2015) and comes with a complete agent-based implementation in the GAMA platform. Beyond the particular integrated model we present in the next section, this proposal aims at being as generic as possible and constitutes both the basis and one of the outcomes of our research.

3. INTEGRATED LAND-USE CHANGE MODEL

Our proposal is organized to fulfill the requirements expressed in the Introduction and address the limitations of the existing land-use change models exposed in Section 2.1. Following Lambin and Geist (2007) and JunJie (2008), for whom the farmers' land-use change decisions are influenced by a complex combination of

social, economic and environmental factors, the integrated model we have designed relies on four main hypotheses (and some associated assumptions) that individually shape the sub-models used to represent these factors and their interactions.

3.1. Hypotheses Regarding the Dynamics of Land-Use Change

When analyzing the map on **Figure 4**, one can easily see that the changes in land-use are localized in specific sub-regions (exactly, specific villages) of the district. However, the environmental contexts of these villages are completely different: for instance, one of these villages is located in an area under the protection of dikes (to prevent salt intrusion), while the others are not. The same village borders the main branch of the river, while the others either border canals and roads, or do not seem to follow any specific pattern. In all these cases, it is clear that the farmers do not put the same exact rules into practice. What does it mean in terms of behavior? A first hypothesis would be that farmers from different villages have completely different decision-making behaviors. A second, more realistic, is that farmers, whilst equipped with the same decision-making behaviors, and influenced by the same core factors, including their income, for instance, dynamically adapt the priority of these behaviors with respect to their environmental—or social—contexts, thereby taking into account a rather large number of criteria. This observation is in line with those of Lambin and Geist (2007) and Beratan (2007). This brings us to formulate our first hypothesis:

- Hypothesis H1: Farmers use a complex decision-making process based on multiple criteria, which are adjusted based on their perceptions or individual beliefs. These processes, performed independently, drive the global land-use change process.

More precisely, Ahnstrm et al. (2009) points out that economic factors, such as the price of products, costs of production and benefits expected are the most important factors influencing the decision of farmers. Farmers usually tend to produce products that are supposed to provide them with the highest income in the future. **Figure 5** shows the evolution of average prices of the four main products (rice, vegetables, shrimp, coconut) in Ben Tre and An Giang provinces from 2005 to 2010. These products correspond to different land-use types (Rice, Rice-Vegetables, Rice-Shrimp, Shrimp, Annual crops, Industrial perennial and Fruit perennial). This leads us to establish our second hypothesis:

- Hypothesis H2: Among the various factors driving land-use changes, the market price of products and the income they expect from them play a pivotal role in the decision of farmers.

Farmers, when deciding which type of production they will choose, need of course to take the suitability of their parcel i.e., the type and quality of the soil and the quality of water available, into account. Smajgl et al. (2015) points out that environmental factors, like saltwater intrusion, implicit effects of infrastructure planning, etc., have a negative impact on this suitability, especially in coastal regions, and will then even indirectly lead to changes in land-use. This provides us with a third hypothesis:

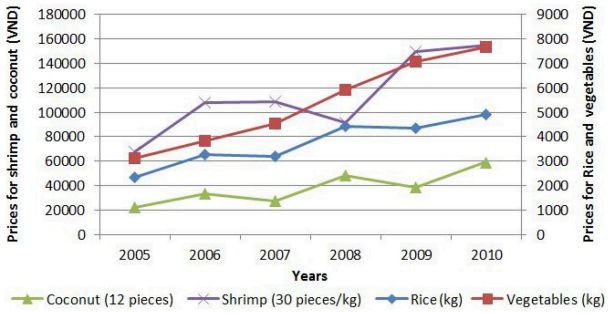


FIGURE 5 | Market prices of the most popular products in the Mekong Delta from 2005 to 2010. Source: Combined from various sources (BTCA, 2015; DARD, 2015; VASEP, 2015; VFA, 2015) (1USD ≈ 21,840VND).

- Hypothesis H3: Environmental characteristics, such as saltwater intrusion, soil type, etc., change the suitability of parcels with respect to some agricultural productions and directly or indirectly influence the strategy of farmers.

Finally, the decision of farmers are influenced by the decision of their neighbors (Case, 1992) or their family. In Vietnam, for example, when farmers succeed with a new land-use type, they usually gain a lot of attention from their neighbors or through their social relationships; moreover, a number of associations of farmers exist in every village, where they are encouraged to exchange their experiences and techniques of cultivation with different networks. This brings us to formulate a fourth hypothesis:

- Hypothesis H4: Social interactions, between neighbors, within families or more extended social networks, exert an influence on the decisions of farmers, either because of the exchanges of information or because of the display of successful experiments.

3.2. Sub-Models Organization

The four hypotheses presented above rely on quite heterogeneous data and studies. Actually, testing them all at once in one single model would certainly result in a complex construction difficult to design, maintain, adapt and experiment. But we can exploit their independence and consider each of them to be the subject of a much simpler (sub-)model, in order to test them individually, but also to test their respective influences in various patterns of composition. To this aim, we propose in Section 4 a novel approach to models coupling that provides modelers with more flexibility than existing approaches (it allows them to freely and dynamically add, remove, or substitute models).

Concretely, the four hypotheses are translated into five independent sub-models, each of them representing the factors or part of the factors present in each hypothesis, so that they could be tested in isolation or in interaction with each other. **Figure 6** summarizes this decomposition. We provide more details on each sub-model in the following sections.

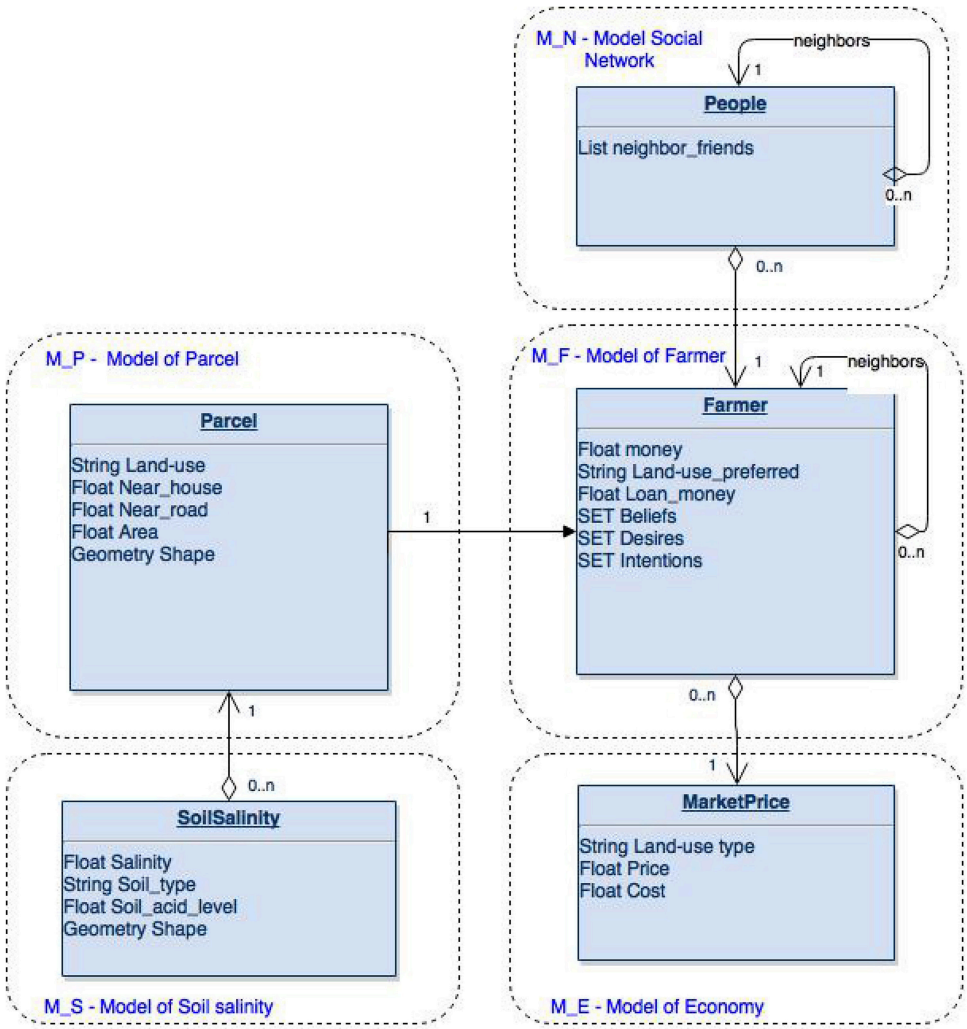


FIGURE 6 | Conceptual view of the five sub-models that constitute our integrated model.

3.3. M_S: Sub-Model of Soil Salinity

3.3.1. Presentation

As discussed in Section 3.1, environmental factors influence the decisions of farmers and the salinity of the soil is one of the most important factors farmers take into account when they need to assess the suitability of parcels for different types of cultures. We then designed a first sub-model exclusively dedicated to the reproduction of the dynamics of soil salinity from 2005 to 2010. As shown in **Figure 7**, the input of this first sub-model are the salinity maps of 2005 Vo and Le (2006) and 2010, available thanks to an efficient regional network, the GIS data on dikes and dike-protected areas for the year 2010 (PCBT, 2011), the GIS data on parcels and their land-use and the GIS data on rivers and canals.

3.3.2. Model Design

This sub-model relies on a discretization of the environment in 18,400 parcels, obtained from the land-use map, where each parcel is linked with its immediate neighbors in a radius of 100 m and is provided with a set of attributes, among them its salinity [classified into 4 levels (less than 0.4%; 0.4–0.8%; 0.8–1.2%; greater than 1.2%)] whether or not it is in dike-protected area, and whether or not it is bordering a river (obtained by overlapping the rivers and canals maps). Initial salinity levels in Thanh Phu district, in 2005, are computed after (Vo and Le, 2006).

The dynamics of the model is voluntarily kept simple and deterministic: at each iteration (1 year) it reevaluates, like in a cellular automaton, the level of salinity of each parcel. Parcels considered as protected by dikes do not change. Parcels bordering rivers see their salinity automatically rise up to 1.2%. And salinity is diffused in the remaining parcels using the following function:

$$\text{salinity(x)} = \left(\frac{salinity(x) + \sum salinity(y | distance(x, y) <= 100)}{1 + |salinity(y | distance(x, y) <= 100)|} \right)$$
$$\text{with x, y} \in \text{Parcels}$$

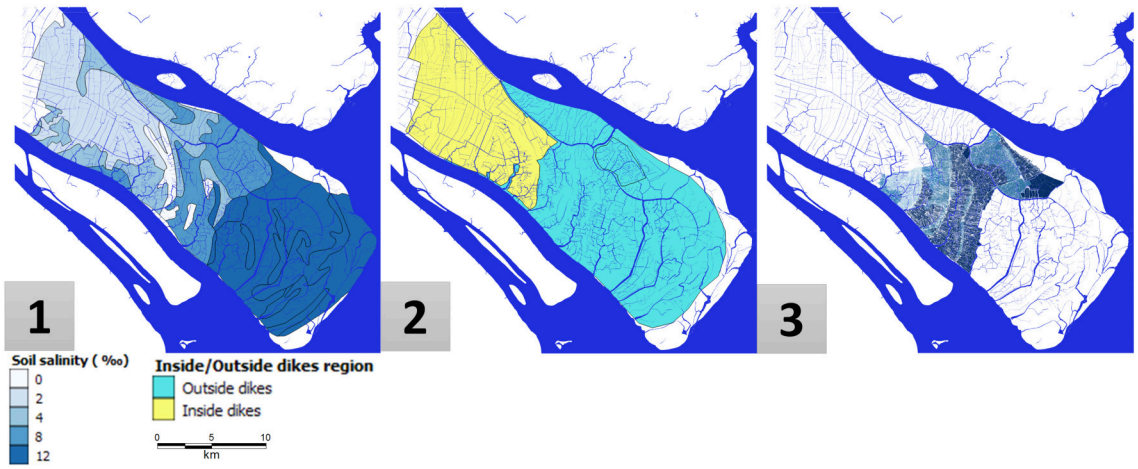


FIGURE 7 | Data for the soil salinity model: (1) Soil salinity map in 2005, (2) Regions protected by dikes in 2010, (3) Soil salinity map in 2010.

One major limitation of this model is that we did not consider the various flooding episodes that occurred during these 5 years, principally to keep it as generic as possible, and also because of the lack of accurate data on these episodes. Taking them into account would probably require the use of a stochastic component, which could in any case be added later if necessary. In its deterministic form, the model has nevertheless been successfully validated by comparing the map obtained by simulation and the actual map of salinity for the year 2010 using the fuzzy-kappa indicator (see Section 5.1.1), with a score of 62.24%, and the absolute deviation percentage (24.28%). Both have been considered as satisfying with respect to its limitations.

3.4. M_P: Sub-Model of Parcels

3.4.1. Presentation

The environment in which land-use changes are simulated is represented by a set of parcel agents, initialized after a land-use map at the level of villages (**Figure 8**). By combining this map with a soil map and a flooding map, each parcel agent is provided with a given land-use and other attributes such as its soil type, its level of salinity, and the extent and depth of flooding episodes on it. The two main purposes of this sub-model are, on one hand, to provide other sub-models with a unified way of accessing and modifying these attributes and, on the other hand, to compute a synthetic indicator called "land suitability." Land suitability represents the compatibility of a given parcel with the different land-use types. It can take 4 values (S1: Highly suitable; S2: Moderately suitable; S3: Marginally suitable ; N: Not suitable) (Soil Resources, Management and Conservation Service, 1981). Based on the type of soil and the level of salinity, we defined (with the help of domain experts) a suitability matrix for each of the 8 land-use types considered in the model (e.g., Rice, Rice-Vegetables, Rice-Shrimp, Shrimp, Annual crops, Industrial perennial, Fruit perennial and Other perennial).

3.4.2. Model Design

In the current instantiation of the integrated model, this sub-model is not provided with any internal dynamics. Instead, it is supposed that some attributes can be manipulated by external models (e.g., M_S for the level of salinity) and that the type of the soil remains unchanged. Each year, each of the parcels then simply computes and updates its land suitability matrix.

3.5. M_E: Sub-Model of Economic Factors

3.5.1. Presentation

In H2 (see Section 3.1), we make the hypothesis that economic factors like regional market prices and costs of production strongly influence the decision of farmers regarding the land-use to adopt. To verify this, we then designed a simple economic model to represent the evolution of these two factors. The data concerning market prices has been collected from 2005 to 2010 (averaged every year) from regional sources. However, the costs of production within the corresponding period could not be obtained so easily; we then used the costs in 2010 (evaluated in Nguyen et al., 2014) and extrapolated them from 2005 to 2010 using the regression equations depicted in **Figure 9** (Judge et al., 1988). Two main assumptions are made in this sub-model: (1) the investment cost and selling price of each agricultural product are the same for all farmers; (2) external incentives (such as the opening of new markets) and perturbations (such as economic crises) are not taken into account.

3.5.2. Model Design

The main components of the model are the 5 regression equations below, where "x" represents the time in year from 1 to 5 (i.e., from 2005 to 2010) and the parameters have been computed after the values for 2010. The costs are expressed in the Vietnamese currency, Dongs, per square meter (VND/m^2). Such a model allows to easily compute the expected benefit of a given production, by subtracting its cost from its selling price,

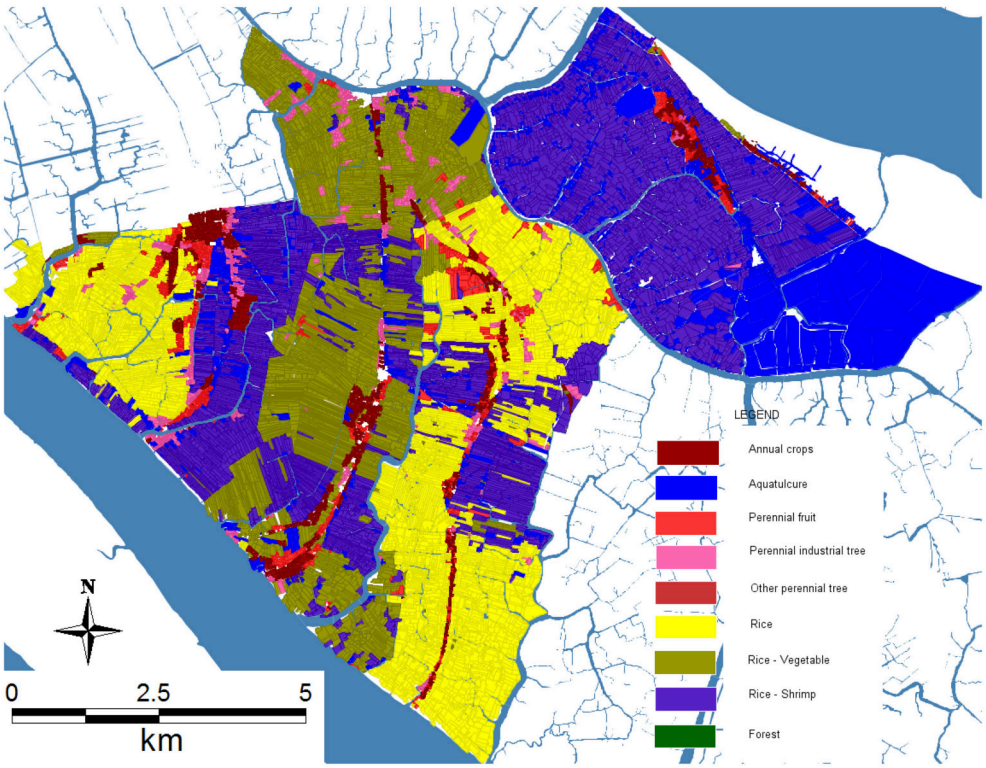


FIGURE 8 | Land-use map of five villages (An Thanh, Binh Thanh, An Thuan, An Quy, An Nhon, An Dien) of Thanh Phu district in 2005.

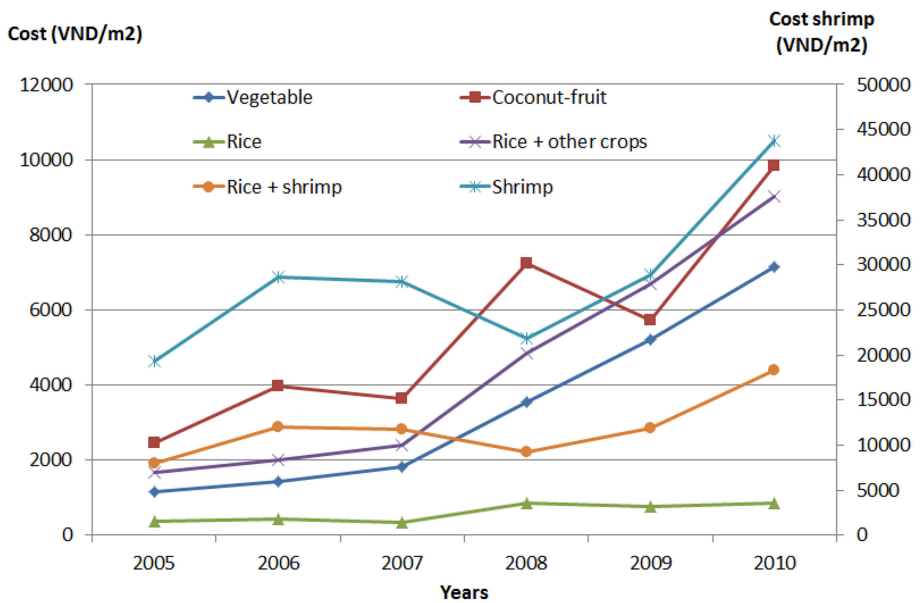


FIGURE 9 | Production costs of the most popular products in the Mekong Delta from 2005 to 2010. Source: Computed from the prices from 2005 to 2010 and the production costs in 2010 (1USD ≈ 21,840VND).

multiplied by the surface of the parcel on which it is cultivated.

$$cost\_vegetable = (1226.4x - 917.55) \quad (1)$$

$$cost\_coconut\_fruit = (1304.5x + 910.91) \quad (2)$$

$$cost\_rice = (-17.71x^3 + 189.3x^2 - 471.95x + 688.9) \quad (3)$$

$$cost\_rice\_other = (1519.1x - 880.31) \quad (4)$$

$$cost\_shrimp = (1345.2x^3 - 13094x^2 + 38752x - 7459.8) \quad (5)$$

$$cost\_rice\_shrimp = (137.74x^3 - 1345.6x^2 + 3998.2x - 865.13) \quad (6)$$

3.6. M_F: Sub-Model of Farmers Behavior

3.6.1. Presentation

In order to validate the hypothesis that farmers use complex decision-making processes based on a multitude of factors (discussed in Section 3.1), we designed an agent-based model that represents farmers (or households, as we do not distinguish individuals from their immediate familial environment) as agents provided with behaviors expressed in the BDI formalism (Caillou et al., 2015). In Truong et al. (2016), the authors have compared three formalisms (i.e., BDI, multi-criteria and probabilistic) for representing complex individual decision-making processes in agent-based models and arrived to the conclusion that BDI is the most relevant when dealing with heterogeneous factors such as the ones we consider in our integrated model. Basically, BDI is a cognitive architecture that relates beliefs (i.e., how the agents perceive their world), desires (i.e., what their goals are) and intentions (i.e., what plans can be undertaken to reach these goals).

The data used to define these behavioral components is based on a survey of 25 households who have changed at least two times their land-use from 1997 to 2014. All come from the village of Binh Thanh, Thanh Phu district, Ben Tre province. The pie chart on **Figure 10** depicts the results of this survey, and we can see that there is a diversity of motivations and expectations that emerge from them. Looking for a higher profit is important, but following what the neighbors do or adapting to changes in land suitability are also frequently cited. This sub-models relies on two main assumptions: (1) these is a 1:1 relationship between farmers and parcels (one farmer only exploits one parcel); (2) the productivity of farmers remains constant and is not affected by technical progress or population growth. While these assumptions are mostly correct for the period of our study, they would need to be reevaluated for longer periods.

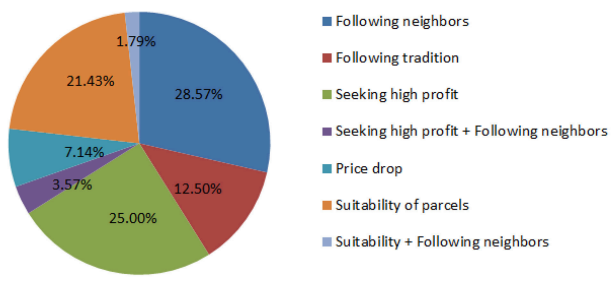


FIGURE 10 | Survey results about the main motivations behind land-use change in Binh Thanh Village, Thanh Phu district, Ben Tre province, Vietnam.

3.6.2. Model Design

Farmer agents are defined as in **Figure 11**, with their attributes, potential actions and potential plans. Each agent is also provided with specific properties, called predicates, that can be either true or false. These predicates, which usually represent a combination of attributes and perceptions of the agent, are used to track which belief, desire or intention should be made active.

- **Beliefs**: Beliefs are a set of predicates, which represent the perception farmers have of their environment and themselves.

  The beliefs base is updated by a function called update_beliefs() at initialization and after each step of the simulations. The beliefs of farmers are defined in **Table 1**, together with the conjunction of conditions to meet to make each belief become true. The three first beliefs are computed based on the available amount of money of farmers. The others are based on the perception of farmers agent regarding the prices of products, the costs of production and the land suitability of their parcel.
- **Desires**: Desires represent a set of objectives that farmers would like to achieve based on their beliefs. Each desire has a priority, which is used to choose the next intention of farmers. A desire can be achieved only when the predicate of the desire is added into the beliefs base or is destroyed by the agent (i.e., when the agent believes that the desire has been realized). Based on our survey data (5), each farmer can have different desires which are not necessarily exclusive. The desires of farmers are listed in **Table 2**. These desires are set or added by an action called *update_desires()*, which allows farmers to update their desires based on their beliefs base.

- Farmers *try_not_to_change*.

  Every farmer can have this desire (it is their initial one), notably after they have changed to their preferred land-use. During simulations, this desire is added to the base when farmers believe that they do not have enough money to change, or that their income is greater than the average income, or that they use their land like their neighbors, or, finally, that the price of their product has not changed.
- Farmers want to *minimize_risks* (environmental and financial ones).

  This desire is designed for farmers who believe that they belong to an average category, that the price of their products is increasing, and that they do not have enough money to select another land-use type. It is also influenced by the belief regarding the suitability of their parcel. If it appears to be (or to have become) not suitable for their current land-use, they will load the desire to change to another one, even if it does not provide the highest income.
- Farmers want to *earn_the_highest_possible_income*.

  Normally, the land uses that require high investments will provide higher incomes. This desire is used by standard

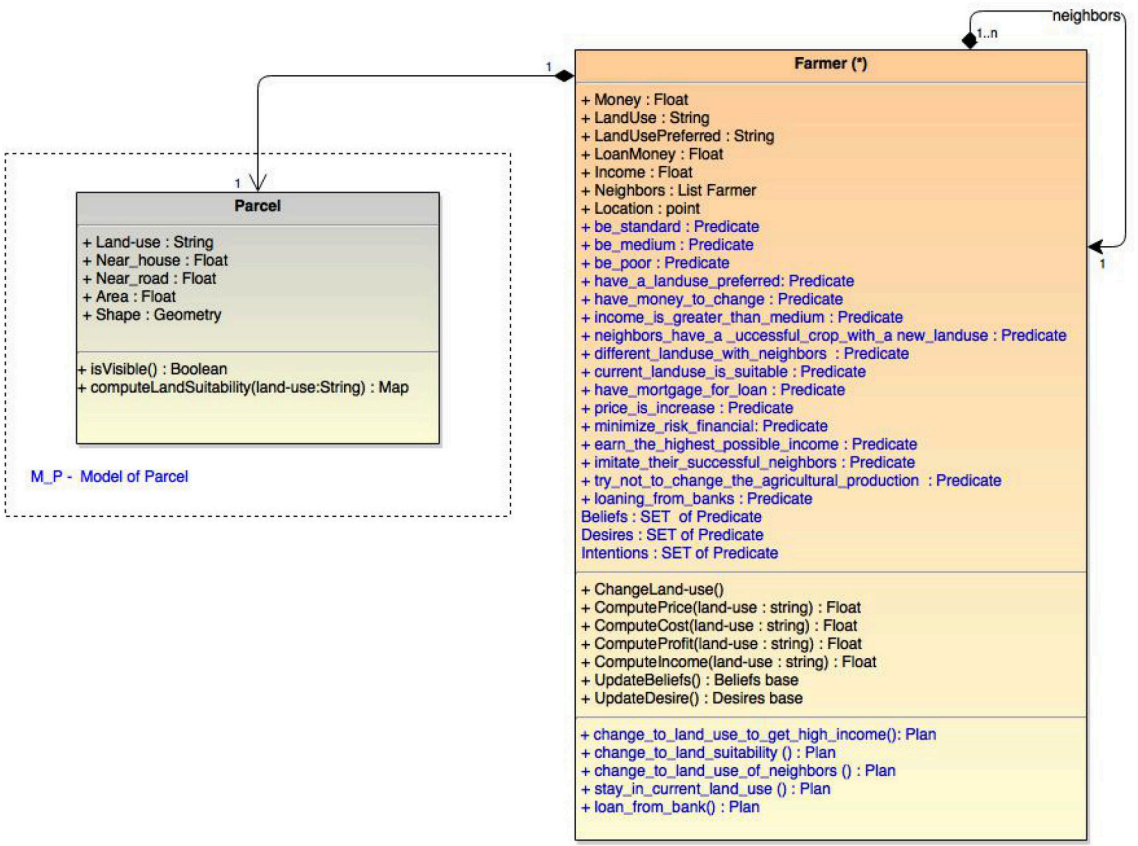


FIGURE 11 | UML representation of farmers.

farmers who believe they have enough money to shift to a land-use type providing the highest possible profit.

- Farmers want to *imitate_their_successful_neighbors*.

  This desire is setup for farmers who believe that they are poor. If they perceive that their neighbors have changed their land use, and that they have been successful in doing so, then they have the desire to change to the same land-use; this desire will also trigger the desire to loan money from banks if the shifting cost is higher than their money.
- Farmers want to *loan_money_from_banks* in order to shift to a new land-use type.

  After farmers select a land-use type different from the one they currently undertake, and if they believe they do not have enough money to change, they add a temporary desire to loan money from a bank, which is provided with the highest priority until they can effectively change.

- **Intentions**: An intention (i.e., what the agent intends to do) is defined for every desire with the goal of fulfilling it as soon as possible. Theoretically, when an agent has two or more simultaneous desires, the corresponding intentions are executed simultaneously. However, because of technical limitations of the BDI architecture we used, intentions are organized sequentially. For example, if a farmer has three simultaneous desires, the three corresponding intentions are added into a first-in-last-out stack of intentions, and the last one is set as the current intention while the others are put on hold. The current intention will determine the selected plan. Plans are a set of actions, which can be executed over the course of several steps of simulation until its post-conditions become true or the related intention is removed from the base of intentions. The relationships amongst Beliefs, Desires and Intentions are showed in **Table 2**. For example, when farmers intend to change their land-use to *earn_the_highest_possible_income* but do not *have_money_to_change*, they desire to *loan_money_from_banks*, which triggers the corresponding plan (*loan_from_banks*) and postpones *change_to_highest_income*.

  The set of plans defined for farmers is:

- "change_to_land_suitability()": This plan is executed when "minimize_risk" is set as the current intention. In this plan, farmer agents select a new land-use type based on the criteria of highest suitability, so as to minimize potential risks, even if the expected income is not the highest.
- "change_to_highest_income()": This plan is executed when farmers have the intention to "earn_the_highest_possible_income." In this plan, they

select a land-use type that is expected to provide them with the highest income, even if it is risky (if the land is not really suitable, for instance).
- "change_to_land_use_of_neighbors()": This plan is executed when farmers have the intention to "imitate_their_successful_neighbors." In this plan, agents undertake the necessary actions to change to the land-use type chosen by their immediate (geographical) neighbors who have earned a higher income than them the year before.

TABLE 1 | Beliefs base of farmers and conditions for the update_beliefs function.

Belief	Condition to set
be_standard	money = 2* mediumIncome
be_medium	money = mediumIncome
be_poor	money = 0.5* mediumIncome
have_ a_preferred_landuse	landUsePreferred ≥ ""
have_money_to_change	money - costs(landUsePreferred) >0
loaning_from_banks	loanMoney >0
neighbors_have_a_successful_ crop_with_a_new_landUse	averageIncome(neighbors, landUse) ≥ income(landUse)
different_landUse_with_neighbors	landUse ≠ neighbors.landUse
have_mortgage_loan	loanMoney = 0 and mortgaged = True
current_landUse_is_not_suitable_with_ the_parcel	computeSuitability(landUse) >2
income_is_greater_than_medium_income	income >mean(income)
price_is_increase	if exist price(landUseTypes(i), step) >price(landUseTypes(i), step-1)

- "loan_from_banks()" : In this plan, farmers request a loan from banks when they have planned to change their land use but do not have enough money to invest in this shift. This request is not automatically fulfilled; as it is the case in reality, this intention is added sequentially when farmers execute the intentions "imitate_their_successful_neighbors" and "earn_the_highest_possible_income." The result of the loan request is decided by the use of a probabilistic function, which takes into account the beliefs of farmers and the compatibility of the new land-use with that projected in the provincial plan (see **Table 2**).
- "keep_current_land_use()" : This plan is executed when farmers have the intention to "try_not_to_change." It does not change anything.

In addition to the actions defined in these different plans, each farmer agent undertakes a number of mandatory actions every simulation step, such as paying back loans to the bank or computing their income.

3.7. M_N: Sub-Model of Farmers Relationships

3.7.1. Presentation

In Hamill and Gilbert (2009), the author supposes the existence of a network in which farmers can be influenced by and can influence their "neighbors." This concept of "neighborhood" can take many forms, from topological or geographical relationships, which rely on the proximity between farmers, to familial or social-economic ones, in which, for instance, the level of income would be used as a filter. A first assumption is made here by considering that the familial network is superseded by the proximity network since in Vietnam, especially in rural areas, it is common that members of the same family live next to each other.

TABLE 2 | Relationships between the Beliefs, Desires and Intentions of farmers.

Beliefs	Desires and corresponding intentions	Plans
not have_money_to_change income_is_greater_than_medium_income not different_landUse_with_neighbors not price_is_increasing	try_not_to_change _production	stay_in_current_land_use()
be_medium price_is_increasing not current_landUse_is_suitable	minimize_risks	change_to_land_suitability()
be_standard price_is_increasing	earn_the_highest_possible_income	change_to_highest_income() and loan_from_banks() based on beliefs
be_poor neighbors_have_a_successful_crop _with_a_new_landUse	imitate_their_successful_neighbors	change_to_land_use_of_neighbors() and loan_from_banks() based on beliefs
not have_money_to_change have_mortgage_for_loan not loaning_from_banks	loan_money_from_banks	loan_from_banks()

A second assumption is that the exchanges of influence take place between farmers that belong to the same "social level" (or income group).

Statistical population data used at the provincial level (VGSOBT, 2010) distinguishes between 3 different profiles of farmers, essentially based on their level of income: (1) P1: rich and standard farmers, (2) P2: average farmers, (3) P3: poor and nearly poor farmers. We reuse this classification and couple it with the proximity network in order to produce an "influence network" for each farmer.

3.7.2. Model Design

This network is recomputed at every iteration of the simulation (as farmers may change their income) and its main purpose is to serve as a "social topology" for farmers, i.e., to modify the way they compute their set of neighbors. In the absence of this sub-model, the neighbors of a farmer are the farmers located in a radius of 100 m around it. When this sub-model is used, the neighbors become the farmers located in the same radius and belonging to the same profile.

4. INTEGRATION OF THE SUB-MODELS

4.1. Models as Agents

To couple these different models, and to address the limitations pointed out in Section 2.2, we introduce in this section the concept of "co-modeling." This concept is relying on a multi-agent approach to knowledge and control decentralization, but also borrows from DEVS its recursive design of models.

Basically, a "co-model" is an agent-based model in which some agents represent other models (called, in that case, "micro-models"). As in any "regular" ABM, these agents can be added or removed dynamically from their "macro-model." They have their own attributes, life cycle, operations, collaborations, conflict resolution mechanisms, etc. and the only difference with regular agents is that they wrap one or several instances of other models (which can, themselves, be co-models).

The interest for modelers (at least for modelers already using the ABM paradigm) is that they don't have to learn a new set of concepts for coupling sub-models. If they know how to write an agent-based model, they are normally able to write a co-model. One difference, though, is that a co-model is intended to capture and represent a particular collaboration between these "micro-models," which can be based on a collaboration scheme of experts, for instance, or on any other way of organizing their contributions. Conversely, any regular ABM can be viewed as a very specific implementations of co-models, where agents only wrap models of individuals.

This proposal, implemented in the GAMA simulation platform (but which could be implemented in any other platform as long as its meta-model supports multi-level definitions and recursivity within agents), does not solve the numerous problems (of relating the inputs and outputs, harmonizing the scales, and so on) raised by coupling heterogeneous models. Its main purpose is to introduce a natural way of writing elegant solutions to these problems, where the relationships between models are viewed as interactions between artificial agents. The whole literature on multi-agent systems (Michel et al., 2009) can then be used to imagine and implement specific collaboration protocols between micro-models, which can consist in exchanges of data, of course, but also more sophisticated interactions (control of one model by another, self-organization of micro-models, hierarchical organization of micro-models, exchange of semantic information, etc.). Furthermore, considering models as agents allows to envision new ways of designing integrated models, for instance through an interactive participatory process with the modeler (Guyot et al., 2005).

This approach to models coupling is still experimental for the moment, and it needs to be validated especially in its advanced uses, for example dynamically coupling/uncoupling models during simulations. What is interesting, however, is that it supports modelers in designing flexible experiments, with the possibility to add, remove or change micro-models at runtime. In our case, this possibility will be extensively used for testing, individually and collectively, the different hypotheses.

As specified in Section 2.2, current land-use change integrated models are not flexible enough when it comes to switch between different micro-models. By considering a model as an agent, and an integrated model as a multi-agent system, our approach tries to bring this flexibility to modelers. In addition, GAMA offers several software design capabilities that support very powerful features: because models are agents, they can inherit from other agent definitions, even abstract ones that only define interfaces for getting and setting parameters or manipulating the model. This allows to define abstract integrated models (like frameworks use to be defined in an object-oriented approach) in which micro-models, which only need to respect the interfaces, are instantiated at runtime, whatever their actual contents or definition. Thanks to its multi-formalism approach, GAMA also allows these models to be written in whatever formalism is available as a plug-in to the platform. Finally, because of the possibility to call external programs or libraries during simulations, a model in GAMA can perfectly be a simple wrapper around a legacy model written in another language or on another platform, provided this model can be launched and monitored during execution by an external process.

In **Figure 6** we present the conceptual diagram that describes the wrapping and the coupling of the models listed above: M_F, M_E, M_P, M_S, M_N respectively represent the models of farmers, parcels, environment, economy and social relationships. M_F is coupled with the other models by the way of data exchanges: it requires inputs like product prices (called LU price in M_E), current parcel suitability (LU suitability in M_P) or the decision of neighbors (LU type of M_N). Conversely, it makes the new land use type chosen by the farmers available for other models (it is for example used by M_E and M_N). The two models M_S and M_P are also coupled in that they exchange the salinity attribute.

4.2. Implementation Using GAMA/GAML

The integrated co-model is implemented in the GAML language (Grignard et al., 2013). Its workflow is presented on **Figure 12**. The micro-models are first imported as "wrapping" agents, which need to be instantiated and initialized like any other agent. Each step of the co-model simulation corresponds to 1 year, during which micro-models are executed depending on their temporal

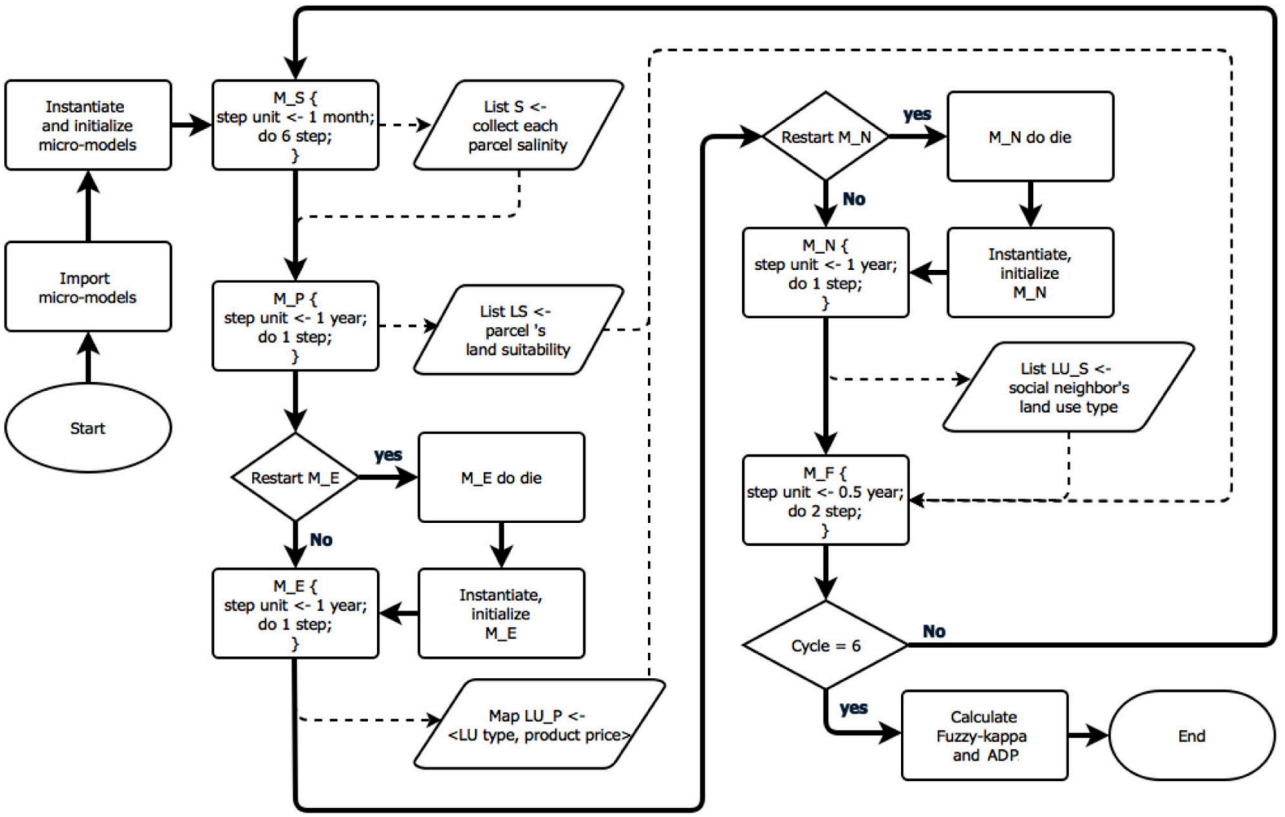


FIGURE 12 | Co-model workflow implemented in GAMA. Bold arrows are the flow of processes. Dotted lines represent the flows of data.

scale (from 1 month to 1 year), their inputs and outputs being taken in charge by the co-model in a completely transparent way.

The full GAML code used to implement the co-model can be found in the Annex. In the following, we highlight some of its parts, mainly to show that manipulating micro-models is not different than manipulating agents. The life cycle of the co-model begins with the importation of the micro-models, which are provided with an alias identifier:

```
import <model_file.gaml> as <micro-model identifier>
```

Micro-models can then be instantiated using the regular "create" statement that specifies the name of the experiment (in GAML, a model is a specification of a simulation, and an experiment is a particular way of instantiating this simulation) and the number of instances of this experiment to create.

```
create <micro-model identifier>.
 <experiment name> [number: number];
```

As for any regular agents, access to the attributes and actions of micro-models use the "ask" statement

```
ask <micro-model identifier>.<experiment name> { do action;}
```

The synchronization of the temporal scales of micro-models is realized by asking their instances to run one or several times, for example, we ask model M_S do 6 simulation cycles.

```
ask M_S.simEnv {loop times:6 {do _step_;}}
```

Finally, a micro-model can be dynamically destroyed (e.g., "ask micro_model do die;") and re-created during the course of the co-model, which allows to run it in its virgin state every time it is needed.

5. EXPERIMENTS

5.1. Input Data

In order to validate our integrated model, we chose a scenario with initial data for the agents in the sub-models are resumed in **Table 3**.

5.1.1. Indicators and Comparison

The main outcome of each simulation of the integrated model is a map of land-use in 2010. To assess its validity, two indicators are used for comparing it to the land-use map observed in 2010: Absolute Deviation Percentages (ADP), which measures the global absolute difference between the maps, and Fuzzy Kappa (FKappa, Visser and de Nijs, 2006), used to measure their similarity based on local correlations, as defined in Section

TABLE 3 | Initialization.

Sub-model	Agent	Main attributes, Input data	Initial value	Purpose	Source
M_P	Parcel	Land-use type	Land-use type in 2005 {Rice, Rice-Vegetables, Rice-Shrimp, Shrimp, Annual crops, Industrial perennial, Fruit perennial and Other perennial}	Initial state of parcels	PCBT, 2011
		Land suitability matrix	Suitability {S1, S2, S3, N} for the land-use types based on soil type and salinity of parcels	The impact of environmental factors is represented by this indicator	Values evaluated in Soil Resources, Management and Conservation Service, 1981
M_F	Farmer	Money 30% of poor: 0–3000, 40% of medium:3000–6000, 20% of standard: 6000–9000, 10% of rich: 9000+	Initial yearly revenue of farmers (in USD)		
		Beliefs base	Set by update_beliefs() function. Classifying profile of farmers based on their revenue	Initializing the beliefs base of farmers	
		Desires and corresponding Intentions	Set by update_desires() function	Initializing the desires base of farmers	
M_S	Soil	Soil_name	{Anthrosol, Gleyic Arenosol, Mollic Arenosol, Mollic Fluvisol, Molli-Salic Fluvisol, Sali Thionic Gleysol, Salic Anthrosol, Salic Fluvisol, Salic Gleysol, Salic Histosol, Thionic Fluvisol}	Static data, combined with the salinity to compute the land suitability matrix of parcels	
	Salinity cell	Salinity	salinity ={1,2,3,4} with 1: 0-4/1000; 2: 4-8/1000; 3: 8-12/1000; 4: 12/1000+	Dynamic data, combined with the soil type to compute the land suitability matrix of parcels	salinity map in 2005 of Vo and Le (2006)
	River branch	Salinity	near-sea branch = 12/1000; others = 0/1000	Provides salinity for rivers in the model M_S	
M_N	Social node	Neighbors	Neighbors = {Nodes ‖ distance(self,node) $\leqslant$ 100m and money(self) = money(node)}	As we do not have real survey data of familial relationships, we use a proximity network and the income level as a proxy	
M_E		Regression equations of prices and costs for the land-use types	Prices and costs of products are based on these equations	Provide price and cost of products used by farmers	costs in 2010 Nguyen et al. (2014)

1.1. These indicators are already implemented in the GAMA platform.

$$ADP(\%) = 100\frac{\sum_{i=1}^{n}|\widehat{X}_i - X_i|}{\sum_{i=1}^{n}\widehat{X}_i} \quad (7)$$

with: $\widehat{X}_i$ the observed quantity of parcels with land-use i and X_i the simulated quantity of parcels with land-use i.

An experiment is defined as a specific combination of micro-models. Because of the possible stochasticity of some of these combinations, 100 simulations are launched for every experiment. Their ADP and Fuzzy Kappa are computed, and we use a one-way ANOVA with the assumption of equal variances and a 95% confidence interval (Bewick et al., 2004) to produce their average values.

These two average values are then compared to the ones obtained for, respectively, a hypothetical scenario characterized by an absence of changes (i.e., where the map obtained in 2010 is supposed to be the same as the one observed in 2005), a hypothetical random model (i.e., a model where farmers choose their land-use randomly every year) and, finally, the map used by the planners in 2005 to estimate the situation in 2010 (see **Figure 4**).

The random model experiments output an average FKappa of 8% and an ADP of 126%, displaying a very low correlation with reality, which is not really a surprise given the randomness of the farmers' behaviors. We can nevertheless note that these random simulations give on average a better result than the projected map established in 2000 by the planners (which FKappa and ADP respectively amount to −0.007% and 186%) !

In the "no changes" scenario, the comparison between the initial situation (2005) and the final one (2010) provides a FKappa with a value of 24.4% and an ADP of 94%, which acknowledges the fact that, actually, numerous changes occurred during these 5 years. The surprise is that this low correlation is still beating the one computed for the projected map revised in 2005, with an FKappa of 17,5% and an ADP of 59%. Although these numbers represent an improvement over both the map for year 200 and the random map, they represent an extremely low correlation with the observed land-use, moreover affected by a strong variance.

5.1.2. Validation of the Co-Model

In order to verify the relevance of the hypotheses discussed in Section 3.1, three experiments have been designed and will be presented here. They are summarized in **Table 4**.

- The first experiment supposes that the farmer decision is based on the sole criterion of the products markets prices. It is therefore a combination of M_P (basic model of parcels, without salinity, which means that the land suitability does not evolve), M_F (model of farmers) and M_E (economic model). Note that the BDI architecture is a bit underused in this case, as only one desire will be triggered in the absence of other predicates.

- The second experiment consists in the coupling of the M_S (salinity) micro-model with the previous ones. This aims at feeding the beliefs, desires and intentions of the agents with dynamic values of land suitability in addition to the market prices. The BDI architecture is then really put to test in this second experiment.

- Finally, the third experiment consists in adding M_N (neighbors) to the previous micro-models, which fuels beliefs and desires with new predicates about the neighbors attitudes.

5.2. Experimental Results

The results of the three experiments are displayed in **Figure 13**. For the first experiment, FKappa is measured at 39.4% and ADP at 43.22%. While definitely more accurate than the maps used by planners, the comparison between simulated and real land-use maps reveals a high variance that prevents them to be really exploitable. The assessment of these results can only conclude that, while the economic factors certainly play a role in explaining land-use changes dynamics, they cannot explain the totality of the changes observed in reality.

By coupling the salinity model to the other micro-models, and therefore adding the criterion of land suitability to the beliefs of farmers, the second experiment produces better results. FKappa increases from 39.42% to almost 43.00%, while the ADP sharply decreases from 43.22% to 31.47%. The dynamic combination of the three first hypotheses (use of the BDI architecture (H1), influence of market prices (H2) and land suitability (H3)) produces a land-use change dynamics that explains nearly half of the observed changes, which is a definite improvement over the maps produced by planners.

The results of the third experiment reveal an ADP of 22% (meaning a global accuracy of 78% in terms of surfaces devoted to each land-use) and a FKappa of 47.92%. The very simple social model M_N, coupled with the three previous ones, appears to improve again the accuracy of the integrated model over its previous incarnations. Although this model represents only one basic way to implement hypothesis H4, it underlines the importance of social factors in the decision-making of farmers, and paves the way for the design of more complex social models (fed with more accurate data on farmers' associations, familial networks, and farmers' local interactions).

5.3. Assessment of the Co-Modeling Architecture

These three experiments have allowed us to test the agent-based co-modeling architecture used to build the different integrated models. By considering each micro-model as an agent, co-modeling definitely eased the experimental process depicted above, allowing us to add or remove micro-models with a minimum of efforts (a few code changes, and only in the co-model, not in any of the micro-models). Moreover, we have been able to manipulate, without any particular constraints, models defined in different formalisms and make them communicate and exchange data and control in a completely transparent way.

This experimental flexibility will be precious to progressively refine and test either the integrated model or the existing micro-models. Adding, removing, changing and verifying hypotheses translates, in the co-modeling architecture, into adding, removing, changing and running micro-models. The only changes required concern the "glue" between micro-models (i.e., the code inside the co-model itself), which allows, if the initial code is sufficiently modular, to create hierarchies of co-models, from the simplest ones to more sophisticated ones.

5.4. Assessment and Discussion

The use of a model based on the decision-making of farmers for better predicting land-use changes in the Mekong Delta was the first aim of this research and the three experiments

TABLE 4 | Description of experiments.

Experiment	Factors	Sub-models	Duration	Repetition	Indicators
1	Economic	M_F, M_E	5 years	100 times	FKappa, ADP
2	Economic, environmental	M_F, M_E, M_P M_S	5 years 60 months	100 times	FKappa, ADP
3	Economic, environmental, social	M_F, M_E, M_P, M_N M_S	5 years 60 months	100 times	FKappa, ADP

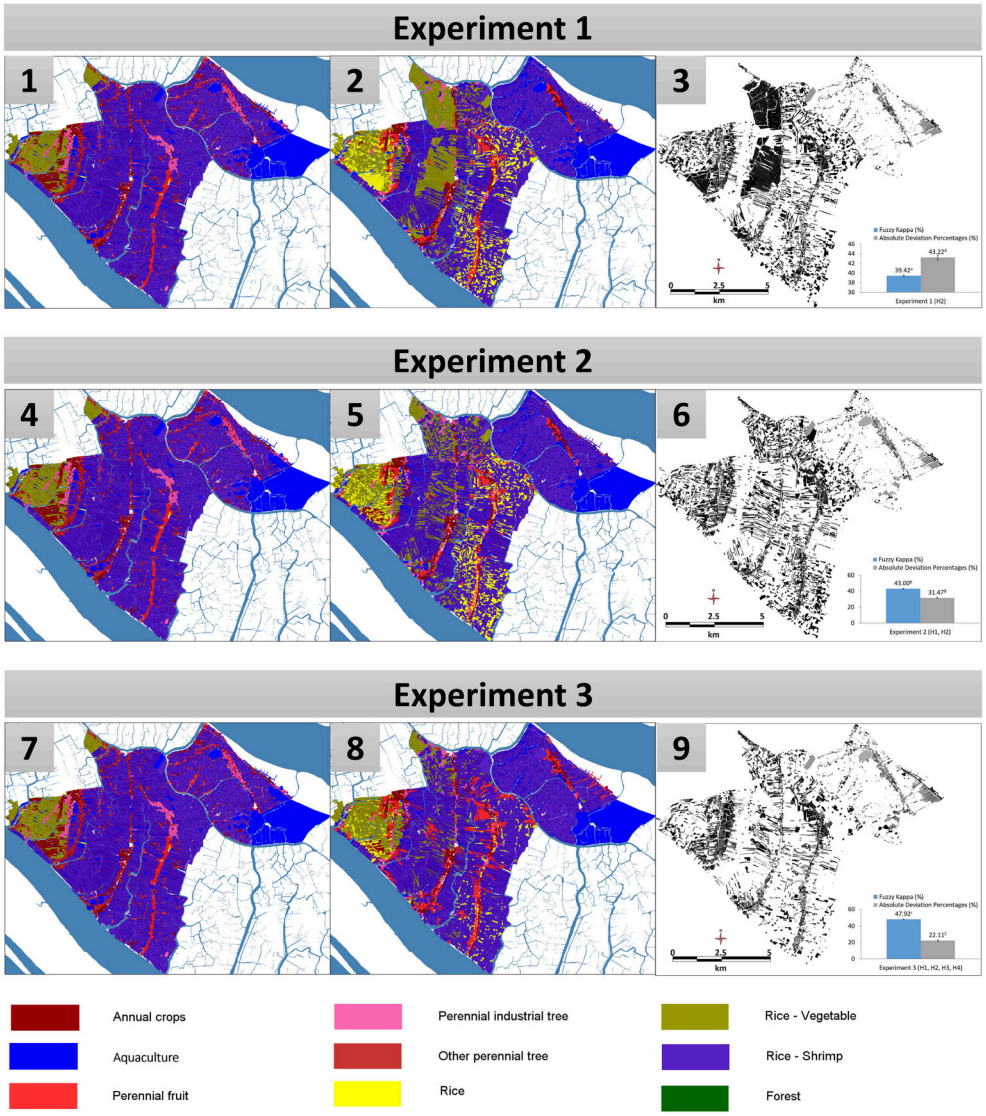


FIGURE 13 | Experiment 1: Comparison between the simulated and observed land-use maps in 2010. Experiment 2: Comparison between the simulated and observed land-use maps in 2010. Experiment 3: Comparison between the simulated and observed land-use maps in 2010. (1), (4), (7) are the results of simulations. (2), (5), (8) are the observed land-use maps in 2010. (3), (6), (9) are the FKappa maps, which show the differences between the two other ones.

described above are an indication that this approach is clearly promising. When we look at **Figure 14**, it is easy to see that the projections used by planners, in 2000 and 2005, do not even show a remote correlation with the state of land-use in 2010. However, starting with the same input data (the land-use map of 2005) and progressively providing farmer agents with knowledge about their economic, environmental and social contexts (with appropriate decision-making processes and behaviors), we have been able to reach nearly 50% of similarity between the map produced by simulation and the map of 2010. Given the extreme simplicity of two of the micro-models employed (the economic and social micro-models), this already represents an interesting achievement, and our plan is to push this further and contact provincial planning authorities so that this model, or a particular instantiation of it, becomes the de facto standard for projecting future land-use in the Mekong Delta. However, it is still, today, in its preliminary stage and we need to improve it if we want it to become really usable by land-use planners.

The first improvement concerns enhancing the realism of the economic and social models, M_E and M_N. The former, which is presently based on the assumption that market prices will follow the dynamics observed in the years 2005-2010, should at least take into account: (1) the global evolution of agricultural trade in South-East Asia, because Vietnam is a net exporter of various products (rice, shrimps, etc.); (2) the local balance between demand and offer, since, besides the price of rice, which is guaranteed by the Vietnamese government, all other products are dependent on the market. Moreover, the market also

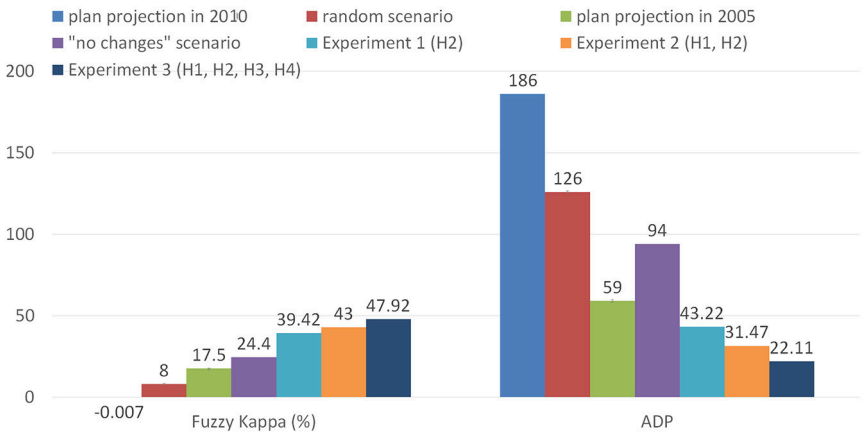


FIGURE 14 | Comparison of the FKappa and ADP indicators of the three experiments.

influences the amount of the investment necessary for shifting from one land-use to another: the prices of fertilizers, seeds and facilities are fluctuating depending on the demand but also on the government's incentive policies. All of these factors need to be taken into account in a more realistic economic model. Regarding the social model, the network of influence goes actually well beyond the immediate neighbors, as is the case in the current model. Farmers' trade unions, extended family networks, political organizations, TV broadcasts all contribute to how innovation or experiments can spread to one village to another. In Vietnam's rural areas, moreover, people often bypass traditional lending institutions when it comes to invest and rely instead on groups of people, often families or a combination of family and trusted friends and neighbors, to lend them money through a mixed mechanism of money pool and lottery (called "hui," translated by "tontine"). A more accurate social model would need to take into account these different networks and how they add up to facilitate the diffusion of new agricultural trends and practices. Crucial data is however missing in the latter case, as very few studies have been undertaken to understand how these networks are constituted and their real influence on farmers. This is where a model like the one presented in this paper is also important, as it points social scientists toward the necessity to undertake specific surveys or fieldwork.

Another area of improvement concerns the addition of a micro-model that would represent the role of the provincial authorities themselves. In our current model, institutions are not taken into account, but they nevertheless play a crucial role in reality, by promoting the plan (through social or political channels), regulating the access to loans or parcels, providing training on agricultural techniques or undertaking large infrastructure investments (like sluice gates and dikes building, irrigation facilities, but also transport infrastructure). Neglecting this direct or indirect role in land-use change, especially in a country like Vietnam where the state is so present, is of course not possible in a complete integrated model. An interesting side effect of incorporating institutions in the model would also be to engage planners in the design of the model by allowing them, maybe in participatory setups, to study the impacts of different policies or investment scenarios.

6. CONCLUSION

In this article, we have proposed a solution for improving land-use change planning, through the design of a hybrid integrated model that captures environmental, social and economic dynamics of this complex phenomenon. The two main outcomes of our work are, on one hand, the model itself, which have proved, despite the simplicity of some sub-models, to be already more accurate than the projections used by planners in the Vietnamese Mekong Delta, and, on the other hand, a flexible agent-based approach to the coupling of heterogeneous models, implemented in an existing simulation platform.

The necessity to use an integrated model has been established from a careful review of the context and existing research, which showed that land-use change should be understood as a consequence of the interplay of numerous factors and actors, including of course farmers. We then proposed a breakdown into several sub-models, each of them representing one of these factors or conjunction of factors, and aiming at verifying one hypothesis. The most central of them, the models of farmers, is implemented using a cognitive AI architecture, BDI, which allows to represent them as complex actors whose beliefs help them to arbitrate between several, sometimes contradictory, desires.

All these sub-models are composed using a new coupling infrastructure called "co-modeling," where they are represented as agents belonging to a higher-level agent-based model (called macro-model); this infrastructure, implemented on the GAMA modeling and simulation platform, provides the flexibility missing in most of the existing proposals, either in the field of land-use change modeling (where the majority of models is specific to a given case study), or in the more general field of multi-modeling, where the price to pay for building integrated models is generally higher in terms of technological investment.

This flexibility has been tested when designing the experiments presented in this article: adding, removing or

switching models, even dynamically, has allowed us to present a progressive assessment of the different hypotheses on which our approach to land-use change was based. Though we did not test all the possible combinations of micro-models and did not propose alternate versions of them, this possibility exists and will be exploited in the future versions of our integrated model. In its current state, the integrated model already offers an interesting combination of modeling techniques and formalisms: a BDI agent-based model of farmers, a cellular automaton based model of diffusion of salinity, a dynamic GIS data based model of parcels, an economic equation-based model and a graph-based social network model. All of them, for the sake of the article, have been programmed in the GAML language (which natively offers the support of multiple formalisms), but reusing legacy models just involves wrapping their invocation in the definition of an agent behavior.

There are three main perspectives of this research that we plan to explore in the near future. The first one will consist in refining the current integrated model, by progressively completing it with more micro-models targeting the other factors identified in land-use change (e.g., an institutional model, local climatic models, models of urban growth) or enhancing the realism of the existing micro-models (by conducting more focused surveys and fieldwork with farmers). The second one will consist in expanding the case study so as to address the problem of land-use change in non-coastal regions of the Mekong Delta, where the environmental conditions are quite different. This perspective will allow us to test one very important feature of the co-modeling architecture, namely the possibility to use different versions of the same micro-model depending on some conditions (for instance, using different environmental micro-models depending on the location of the parcels). Finally, the third perspective will consist in abstracting, as much as we can, the macro-model and the model of farmers decision-making, so as to propose a generic framework that could be instantiated in other geographical contexts. These three perspectives will be pursued throughout 2016 as part of a cooperation between the IRD and Can Tho University.

AUTHOR CONTRIBUTIONS

The three authors have worked together, each in his own specialty, to design and describe the research presented in the paper. AD is the lead designer of the GAMA platform, on which the experiments have been performed, and one of the researchers behind the idea of the "co-modeling" architecture. He has mainly worked on the general outline, introductory and conclusive parts of the paper. NH is the main developer of the "co-modeling" architecture and responsible for its implementation in the GAMA platform. He has also designed some sub-models, among which the salinity model, and helped designing complex experiments. QT is the developer of the BDI decision-making architecture used for farmer agents and the main designer of the integrated model. He has also been responsible for collecting the social, spatial and environmental data used in the model and analysing and validating the outputs of simulations.

ACKNOWLEDGMENTS

The authors would like to thank Prof. Patrick Taillandier (IDEES, France), Prof. Benoit Gaudou (IRIT, France), Prof. Jean-Daniel Zucker (IRD, France), Prof. Philippe Caillou (LRI, France), Prof. Huynh Xuan Hiep (CICT, Vietnam), Prof. Nguyen Hieu Trung (CENRES, Vietnam), and Prof. Vo Quang Minh (CENRES, Vietnam) for their contributions to the various stages of this research.

REFERENCES

Ahnstrm, J., Hckert, J., Berge, H. L., Francis, C. A., Skelton, P., and Hallgren, L. (2009). Farmers and nature conservation: what is known about attitudes, context factors and actions affecting conservation? *Renew. Agric. Food Syst.* 24, 38. doi: 10.1017/S1742170508002391

Bakker, M. M., Alam, S. J., van Dijk, J., and Rounsevell, M. D. A. (2015). Land-use change arising from rural land exchange: an agent-based simulation model. *Landsc. Ecol.* 30, 273–286. doi: 10.1007/s10980-014-0116-x

Beratan, K. K. (2007). A cognition-based view of decision processes in complex social ecological systems. *Ecol. Soc.* 12:27.

Bertsch, C., Ahle, E., and Schulmeister, U. (2014). "The functional mockup interface - seen from an industrial perspective," in *Linköping Electronic Conference Proceedings*, 2014 (Lund), 27–33. doi: 10.3384/ecp1409627

Bewick, V., Cheek, L., and Ball, J. (2004). Statistics review 9: one-way analysis of variance. *Crit. Care* 8, 130–136. doi: 10.1186/cc2836

Blochwitz, T., Otter, M., Akesson, J. R., Arnold, M., Clau, C., and Elmqvist, H., et al. (2012). "The functional mockup interface for tool independent exchange of simulation models," in *Proceedings of the 9th International MODELICA Conference* (Munich), 173–184.

Bollinger, L. A., Nikoli, I., Davis, C. B., and Dijkema, G. P. (2015). Multimodel ecologies: cultivating model ecosystems in industrial ecology. *J. Indust. Ecol.* 19, 252–263. doi: 10.1111/jiec.12253

BTCA (2015). *Weekly Newsletters of Coconut Prices*. Ben Tre: Ben Tre Coconut Association. Retrieved from: http://hiephoiduabentre.com.vn/index.php?Module=Content&Action=blogCategory&id=108&Itemid=207

Caillou, P., Gaudou, B., Grignard, A., Truong, Q. C., and Taillandier, P. (2015). "A simple-to-use BDI architecture for agent-based modeling and simulation," in *The Eleventh Conference of the European Social Simulation Association* (Groningen).

Case, A. (1992). Neighborhood influence and technological change. *Regional Sci. Urban Econ.* 22, 491–508. doi: 10.1016/0166-0462(92)90041-X

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educ. Psychol. Meas.* 20, 37–46. doi: 10.1177/001316446002000104

Dahmann, J., Fujimoto, R., and Weatherly, R. (1998). "The DoD high level architecture: an update," in *Simulation Conference Proceedings, 1998, Winter*, Vol. 1 (Washington, DC: IEEE), 797–804. doi: 10.1109/wsc.1998.745066

Dahmann, J. S., and Morse, K. L. (1998). "High level architecture for simulation: an update," in *Proceedings of the Second International Workshop on Distributed Interactive Simulation and Real-Time Applications*, DIS-RT '98 (Washington, DC: IEEE Computer Society), 32. doi: 10.1109/disrta.1998.694563

DARD (2015). *Weekly Newsletters of Agricultural Prices*. An Giang: Department of Agriculture. Retrieved from: http://sonongnghiep.angiang.gov.vn/wps/portal/

Edmonds, B., and Moss, S. (2005). "From KISS to KIDS - an "anti-simplistic" modelling approach," in *Multi-Agent and Multi-Agent-Based Simulation, volume 3415 of Lecture Notes in Computer Science*, eds P. Davidsson, B. Logan, and K. Takadama (Berlin; Heidelberg: Springer), 130–144.

GAMA (2007). *GAMA Platform - GAMA is an Agent-Based, Spatially Explicit, Modeling and Simulation Platform*. Available online at: http://gama-platform.org

Grignard, A., Taillandier, P., Gaudou, B., Vo, D. A., Huynh, N. Q., and Drogoul, A. (2013). "GAMA 1.6: advancing the art of complex Agent-Based Modeling and Simulation," in *PRIMA 2013: Principles and Practice of Multi-Agent Systems, Number 8291 in Lecture Notes in Computer Science*, eds G. Boella, E. Elkind, B. T. R. Savarimuthu, F. Dignum, and M. K. Purvis (Berlin; Heidelberg: Springer), 117–131.

Guyot, P., Drogoul, A., and Lemaître, C. (2005). "Using emergence in participatory simulations to design multi-agent systems," in *Proceedings of the Fourth International Joint Conference on Autonomous Agents and Multiagent Systems, AAMAS '05* (New York, NY: ACM), 199–203. doi: 10.1145/1082473.1082503

Hamill, L., and Gilbert, N. (2009). Social circles: a simple structure for agent-based social network models. *J. Artif. Soc. Soc.* 12:3.

Hild, D. R. (2000). *Discrete Event System Specification (Devs) Distributed Object Computing (Doc) Modeling and Simulation*. PhD thesis, The University of Arizona.

Hill, D. R. C. (2002). Theory of modelling and simulation: integrating discrete event and continuous complex dynamic systems: second edition by B. P. Zeigler, H. Praehofer, T. G. Kim, Academic Press, San Diego, CA, 2000. *Int. J. Robust Nonlin. Control* 12, 91–92. doi: 10.1002/rnc.610

Huang, H., Wang, L., Zhang, X., Luo, Y., and Zhao, L. (2008). "Coupling multi-agent model and GIS to simulate pine wood nematode disease spread in ZheJiang province," in *Proceedings of Spie the International Society for Optical Engineering, International Conference, 16th Geoinformatics* (Guangzhou). doi: 10.1117/12.812557

Huynh, N. Q., Huynh, H. X., Drogoul, A., and Cambier, C. (2014). "Co-modeling: an agent-based approach to support the coupling of heterogeneous models," in *Nature of Computation and Communication, Number 144 in Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering, International Conference, ICTCC 2014*, eds P. C. Vinh, E. Vassev, and M. Hinchey (Ho Chi Minh City: Springer), 156–170.

Judge, G. G., Hill, R. C., Griffiths, W., Lutkepohl, H., and Lee, T.-C. (1988). *Introduction to the Theory and Practice of Econometrics*. Singapore: Wiley.

JunJie, W. (2008). Land use changes: economic, social, and environmental impacts. *Choices* 23, 6–10. Available online at: http://purl.umn.edu/94681

Kemeny, J. G., and Snell, J. L. (1983). *Finite Markov Chains: With a New Appendix "Generalization of a Fundamental Matrix"*. New York, NY: Springer.

Kim, K. T. G. (2005). "Proposal of high level architecture extension," in *Artificial Intelligence and Simulation, number 3397 in Lecture Notes in Computer Science*, ed T. G. Kim (Berlin; Heidelberg: Springer), 128–137.

Lambin, E., and Geist, H. J. (2007). *Causes of Land-Use and Land-Cover Change*. Retrieved from: http://www.eoearth.org/view/article/150964

Lambin, E. F. (1997). Modelling and monitoring land-cover change processes in tropical regions. *Prog. Phys. Geogr.* 21, 375–393. doi: 10.1177/030913339702100303

Le, Q. B., Park, S. J., Vlek, P. L. G., and Cremers, A. B. (2008). Land-use dynamic simulator (LUDAS): a multi-agent system model for simulating spatio-temporal dynamics of coupled humanlandscape system. I. Structure and theoretical specification. *Ecol. Inform.* 3, 135–153. doi: 10.1016/j.ecoinf.2008.04.003

Li, X., Wang, W., Lei, Y., and Li, Q. (2013). "An agent-centered multi-formalism modeling framework for SoS behavior," in *Proceedings of the 2nd International Conference on Computer Science and Electronics Engineering (ICCSEE 2013), Advances in Intelligent Systems Research Series* (Paris: Atlantis Press), 624–629. doi: 10.2991/iccsee.2013.159

Mena, C. F., Walsh, S. J., Frizzelle, B. G., Xiaozheng, Y., and Malanson, G. P. (2011). Land use change on household farms in the Ecuadorian Amazon: design and implementation of an agent-based model. *Appl. Geogr.* 31, 210–222. doi: 10.1016/j.apgeog.2010.04.005

Michel, F., Ferber, J., and Drogoul, A. (2009). "Multi-agent systems and simulation: a survey from the agents community's perspective," in *Multi-Agent Systems: Simulation and Applications, Computational Analysis, Synthesis, and Design of Dynamic Systems*, ed A. U. Danny Weyns (CRC Press; Taylor & Francis), 47.

MONRE (2009). *Detailing the Establishment, Regulation and Evaluation Planning, Land-Use Planning*. Circular of Ministry of Natural Resources and Environment, No 19/2009/TT-BTNMT, 2009, Hanoi.

MONRE (2012). *Decision on Publishing Statistical Data of Land-Use Change in 2011*. Decision of Minister, Ministry of Natural Resources and Environment of Vietnam, No 1482/QD-BTNMT, 2012, Hanoi.

Moreira, E., Costa, S., Aguiar, A. P., Cmara, G., and Carneiro, T. (2009). Dynamical coupling of multiscale land change models. *Landsc. Ecol.* 24, 1183–1194. doi: 10.1007/s10980-009-9397-x

Murray-Rust, D., Robinson, D. T., Guillem, E., Karali, E., and Rounsevell, M. (2014). An open framework for agent based modelling of agricultural land use change. *Environ. Model. Softw.* 61, 19–38. doi: 10.1016/j.envsoft.2014.06.027

Nguyen, M. T. H., Vo, M. Q., and Vo, G. T. (2014). Suitable land use zoning in line with climate change scenarios for three coastal districts in Ben Tre province. *Vietnam Soil Sci.* 44, 68–74.

Nguyen, N. D., Drogoul, A., and Auger, P. (2008). "Methodological steps and issues when deriving individual based-models from equation-based models: a case study in population dynamics," in *Proceedings of the Intelligent Agents and Multi-Agent Systems, 11th Pacific Rim International Conference on Multi-Agents, PRIMA 2008, Hanoi, December 15–16, 2008* (Hanoi), 295–306. doi: 10.1007/978-3-540-89674-6_33

Nicolai, T. W., Wang, L., Nagel, K., and Waddell, P. (2011). "Coupling an urban simulation model with a travel modela first sensitivity test," in *Proceedings of the Computers in Urban Planning and Urban Management (CUPUM '11)* (Lake Louise, AB), 11–07.

North, M. J., Collier, N. T., and Vos, J. R. (2006). Experiences creating three implementations of the repast agent modeling toolkit. *ACM Trans. Model. Comput. Simul.* 16, 1–25. doi: 10.1145/1122012.1122013

Parker, D. C., Berger, T., and Manson, S. M. (2002). "Agent-based models of land-use and land-cover change," in *Report and Review of an International Workshop* (Irvine, CA).

Parker, D. C., Manson, S. M., Janssen, M. A., Hoffman, M. J., and P., D. (2003). Multi-agent systems for the simulation of land-use and land- cover change: a review. *Ann. Assoc. Am. Geograph.* 93, 316–340. doi: 10.1111/1467-8306.9302004

Parrott, L. (2011). Hybrid modelling of complex ecological systems for decision support: recent successes and future perspectives. *Ecol. Inform.* 6, 44–49. doi: 10.1016/j.ecoinf.2010.07.001

PCBT (2011). *Report of Land-Use Zoning in 2020 and the Land-Use Plan for 5 Early Years 2011–2015 of Ben Tre Province*. Ben Tre: People's Committee of Ben Tre Province.

Quesnel, G. V. D. (2005). "Coupling of physical models and social models: multi-modeling and simulation with VLE," in *Joint Conference on Multi-Agent Modelling for Environmental Management (CABM-HEMA-SMAGET05)* (Bourg Saint Maurice), 21–25.

Rajeevan, M., and Nanjudiah, R. (2009). Coupled model simulations of twentieth century climate of the indian summer monsoon. *Curr. Trends Sci.* 537–567.

Rao, A., and Georgeff, M. (1991). "Modeling rational agents within a BDI-architecture," in *Proceedings of the 2nd International Conference on Principles of Knowledge Representation and Reasoning* (San Mateo, CA), 473–484.

Rochette, S., Huret, M., Rivot, E., and Le Pape, O. (2012). Coupling hydrodynamic and individual-based models to simulate long-term larval supply to coastal nursery areas. *Fish. Oceanogr.* 21, 229–242. doi: 10.1111/j.1365-2419.2012.00621.x

Rousseaux, F., Bocher, E., Gourlay, A., and Petit, G. (2012). "Toward a coupling between GIS and agent simulation: USM, an OrbisGIS extension to model urban evolution at a large scale," in *OGRS 2012 Proceedings* (Switzerland), 206–214.

Serneels, S., and Lambin, E. F. (2001). Proximate causes of land-use change in Narok District, Kenya: a spatial statistical model. *Agric. Ecosyst. Environ.* 85, 65–81. doi: 10.1016/S0167-8809(01)00188-8

SISO, S. (2010). *IEEE Standard for Modeling and Simulation (M&S) High Level Architecture (HLA) – Framework and Rules. IEEE Std 1516-2010 (Revision of IEEE Std 1516-2000)* (New York, NY), 1–38.

Smajgl, A., Toan, T. Q., Nhan, D. K., Ward, J., Trung, N. H., Tri, L. Q., et al. (2015). Responding to rising sea levels in the Mekong Delta. *Nat. Clim. Change* 5, 167–174. doi: 10.1038/nclimate2469

Soil Resources, Management and Conservation Service (ed.). (1981). "A framework for land evalution," in *Number 32 FAO Soils Bulletin, Vol. 2, Print Edition* (Rome: FAO).

Steiner, A. L., Pal, J. S., Rauscher, S. A., Bell, J. L., Diffenbaugh, N. S., Boone, A., et al. (2009). Land surface coupling in regional climate simulations of the west african monsoon. *Clim. Dyn.* 33, 869–892. doi: 10.1007/s00382-009-0543-6

Subedi, P., Subedi, K., and Thapa, B. (2013). "Application of a hybrid cellular automaton markov (CA-markov) model in land-use change prediction: a case study of saddle creek drainage Basin, Florida. *Appl. Ecol. Environ. Sci.* 1, 126–132. doi: 10.12691/aees-1-6-5

Taillandier, P., and Therond, O. (2011). "Use of the belief theory to formalize agent decision making processes: application to cropping plan decision making," in *European Simulation and Modelling Conference* (Guimarães), 138–142.

Trickett, S. B., and Trafton, J. G. (2007). What if: the use of conceptual simulations in scientific reasoning. *Cogn. Sci.* 31, 843–875. doi: 10.1080/03640210701530771

Truong, Q. C., Taillandier, P., Gaudou, B., Vo, M. Q., Nguyen, T. H., and Drogoul, A. (2016). "Exploring agent architectures for farmer behavior in land-use change. A case study in coastal area of the Vietnamese Mekong Delta," in *Multi-Agent Based Simulation XVI, Lecture Notes in Computer Science*, eds B. Gaudou and J. S. Sichman (Istanbul: Springer International Publishing), 146–158. doi: 10.1007/978-3-319-31447-1_10

Valbuena, D., Verburg, P. H., Bregt, A. K., and Ligtenberg, A. (2010). An agent-based approach to model land-use change at a regional scale. *Landsc. Ecol.* 25, 185–199. doi: 10.1007/s10980-009-9380-6

Vangheluwe, H. (2000). "DEVS as a common denominator for multi-formalism hybrid systems modelling," in *IEEE International Symposium on Computer-Aided Control System Design*, ed A. Varga (Anchorage, AL: IEEE Computer Society Press), 129–134.

VASEP (2015). *Price of Aquaculture Products in Soc Trang Province from 2005 to 2010*. Vietnam Association of Seafood Exporters and Producers. Retrieved from: http://vasep.com.vn/119/Thong-ke-thuy-san/Gia-trong-nuoc.htm

VFA (2015). *Prices of Domestic Rice from 2007 to 2010*. Ho Chi Minh City: Vietnam Food Association. Retrived from: http://www.vietfood.org.vn/vn/default.aspx?c=101

VGSO (2000). *Report of Census for Land-Use Area in 2000*. Vietnamese General Statistics Office. Retrived from: http://www.gso.gov.vn/Modules/Doc_Download.aspx?DocID=659

VGSOBT (2010). *Statistical Yearbook of Ben Tre Province 2010*. Ben Tre: Vietnamese General Statistics Office.

Visser, H., and de Nijs, T. (2006). The map comparison kit. *Environ. Model. Softw.* 21, 346–358. doi: 10.1016/j.envsoft.2004.11.013

VNA (2013). *Land on Law, Chap. 4*. Hanoi: Vietnamese National Assembly. Law No. 45/2013/QH13.

Vo, Q. M., and Le, Q. T. (2006). "Soils of the Mekong delta, based on WRB (World Reference Based) - FAO system, at the scale 1/250.000," in *Collection of the research scientific of College of Agriculture and Applied Biology, Can Tho University*, 147–156.

Yez, E., Hormazbal, S., Silva, C., Montecinos, A., Barbieri, M. A., Valdenegro, A., et al. (2008). Coupling between the environment and the pelagic resources exploited off northern chile: ecosystem indicators and a conceptual model. *Lat. Am. J. Aquat. Res.* 36, 159–181. doi: 10.3856/vol36-issue2-fulltext-3

Young, K. B., Wailes, E. J., Cramer, G. L., and Khiem, N. T. (2002). *Vietnam's Rice Economy: Developments and Prospects. Research report, Arkansas Agricultural Experiment Station, Arkansas*. Available online at: http://arkansasagnews.uark.edu/968.pdf

Zeigler, B., Moon, Y., Kim, D., and Ball, G. (1997). The DEVS environment for high-performance modeling and simulation. *IEEE Comput. Sci. Eng.* 4, 61–71. doi: 10.1109/99.615432

Zhao, L., and Peng, Z.-R. (2012). LandSys: an agent-based Cellular Automata model of land use change developed for transportation analysis. *J. Transp. Geogr.* 25, 35–49. doi: 10.1016/j.jtrangeo.2012.07.006

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Environmental Science

ORIGINAL RESEARCH
published: 20 November 2015
doi: 10.3389/fenvs.2015.00074

Hybrid Bottom-up/Top-down Energy and Economy Outlooks: A Review of IMACLIM-S Experiments

Frédéric Ghersi *

Centre National de la Recherche Scientifique, CIRED, Nogent-sur-Marne, France

In this paper we survey the research undertaken at the *Centre International de Recherche sur l'Environnement et le Développement* (CIRED) on the combination of the IMACLIM-S macroeconomic model with "bottom-up" energy modeling, with a view to associate the strengths and circumvent the limitations of both approaches to energy-economy-environment (E3) prospective modeling. We start by presenting the two methodological avenues of coupling IMACLIM-S with detailed energy systems models pursued at CIRED since the late 1990s': (1) the calibration of the behavioral functions of IMACLIM-S that represent the producers' and consumers' trade-offs between inputs or consumptions, on a large set of bottom-up modeling results; (2) the coupling of IMACLIM-S to some bottom-up model through the iterative exchange of some of each model's outputs as the other model's inputs until convergence of the exchanged data, comprising the main macroeconomic drivers and energy systems variables. In the following section, we turn to numerical application and address the prerequisite of harmonizing national accounts, energy balance, and energy price data to produce consistent hybrid input-output matrices as a basis of scenario exploration. We highlight how this data treatment step reveals the discrepancies and biases induced by sticking to the conventional modeling usage of uniform pricing of homogeneous goods. IMACLIM-S rather calibrates agent-specific margins, which we introduce and comment upon. In a further section we sum up the results of 4 IMACLIM-S experiments, insisting upon the value-added of hybrid modeling. These varied experiments regard international climate policy burden sharing; the more general numerical consequences of shifting from a biased standard CGE model perspective to the hybrid IMACLIM approach; the macroeconomic consequences of a strong development of electric mobility in the European Union; and the resilience of public debts to energy shocks. In a last section we offer some conclusions and thoughts on a continued research agenda.

Keywords: hybrid energy economy modeling, induced production frontier, technical change modeling, hybrid energy economy accounts, climate and energy policy modeling

OPEN ACCESS

Edited by:
Christian E. Vincenot, Kyoto University, Japan

Reviewed by:
Fei Wang, Institute of Soil and Water Conservation, CAS, China
Luis Gomez, University of Las Palmas de Gran Canaria, Spain

***Correspondence:**
Frédéric Ghersi
ghersi@centre-cired.fr

Specialty section:
This article was submitted to Environmental Informatics, a section of the journal Frontiers in Environmental Science

Received: *03 August 2015*
Accepted: *06 November 2015*
Published: *20 November 2015*

Citation:
Ghersi F (2015) Hybrid Bottom-up/Top-down Energy and Economy Outlooks: A Review of IMACLIM-S Experiments. Front. Environ. Sci. 3:74. doi: 10.3389/fenvs.2015.00074

INTRODUCTION

Bottom-up (BU) and top-down (TD) models of the energy systems have been opposed since at least Grubb et al. (1993), echoing even older distinctions dating back to the oil crises aftermath, before the revival of energy modeling prompted by the climate affair. The 2nd IPCC report highlights what is arguably the peak of the competition between the two modeling chapels (Hourcade and Shukla, 2001), after which both communities start considering bridging the gap between them through the development of hybrid models. We specifically discussed the need for such hybrid approaches in Hourcade et al. (2006). This need stems from recognizing that standard macroeconomic models, whatever their paradigm (optimal control, computable general equilibrium, econometric estimation, or the recently developing dynamic stochastic general equilibrium approach) simulate rather aggregate energy supplies and demands through mathematical functions that are (1) quite plain (for most of them continuous and infinitely differentiable), for a sheer analytical convenience that comes at the cost of maladaptation to observed or expected flexibilities of energy systems; (2) devised as first order proxies valid only in the vicinity of some initial state of the economy; (3) applied in the framework of growth theories unfit to represent technical change biased toward a drastic reduction of the energy intensity of growth. Standard macroeconomic models are thus structurally incapable of representing the major and complex technical evolutions of the energy systems, the now so-called energy transition, that the conjunction of conventional energy scarcity and the climate conundrum seems bound to induce in coming years.

Conversely, detailed energy system models offer the scope and detail necessary to capture (and aggregate at need) the complex expected evolutions of the manifold energy production techniques and end-uses;[1] but they lack the integrative macroeconomic framework necessary to evaluate the total economic costs attached to these evolutions—be it only a comprehensive description of investment markets—and thus cannot possibly account for the feedbacks between the macroeconomic constraints and the energy systems shifts.

This recognition led to develop the IMACLIM-S model as an alternative to the standard computable general equilibrium (CGE) model, focusing on providing macroeconomic consistency to the partial equilibrium results of bottom-up models. Beyond the much explored hybrid modeling strategies that link some disaggregated energy supply module to a standard CGE model, IMACLIM-S indeed constrains not only energy supply but also the energy demands of all economic agents (however aggregated) based on the explicit representation of energy systems specific to bottom-up approaches. It has done so following two quite distinct methodological avenues, which we successively present in Section Two Coupling Avenues below. In Section Data Harmonization Requirements Building Hybrid Energy/economy Accounting Tables we insist on the data harmonization necessary to a sound application of both modeling approaches of Section Two Coupling Avenues. In Section Hybrid Modeling Results we sum up various results of applied work based on either one of the two modeling methods, insisting upon the value-added of BU and TD coupling. In Section Conclusion and Further Research Agenda we draw some conclusions and reflect upon a continued research agenda.

TWO COUPLING AVENUES

CIRED's years of research on hybrid modeling build on a profound dissatisfaction with the extension, in the wake of the 1973 oil crisis, of Solow's neoclassical growth model (Solow, 1957) to energy questions[2]. That extension simultaneously meant:

- Applying Solow's "wrinkle" to energy, i.e., interpreting the observed energy cost shares (either aggregate or indeed sectoral) as the instantaneous results of cost minimization facing a vector of input prices.
- Generalizing the resulting production function(s) to decades of prospective outlook by simply submitting it to exogenous technical change—at first, to neutral (uniform) factor productivity improvements, as do Hudson and Jorgenson (1974); in the second wave of models spurred by the climate affair from the 1990s on, to biased productivity improvements including a specific "autonomous energy efficiency improvement" (AEEI)[3].

Under such a modeling framework of energy-economy interactions, a change in energy prices at any point in time, whatever its intensity, only displaces the KLEM input mix along some pre-determined, exogenous isoquant. This cannot be a proper framework, especially in the shorter temporal terms, to assess the dramatic shifts expected from highly ambitious climate policy action, or for that matter from potential geopolitical instabilities, under the inescapable constraint of (mostly) strongly inert energy supply and end-use equipment stocks[4].

This critique has two main implications. The first is that any relevant model of energy-economy interactions must work in an induced technical change framework, in the tradition of Kennedy (1964)—[5], i.e., in technical terms, one that guarantees

[1]Notwithstanding, bottom-up models have been criticized for their too-naive representation of supply or demand technologies investment decision (Sutherland, 1991)—the source of a perceived "efficiency gap" between the available techniques and those really implemented (Jaffe and Stavins, 1994), quite unsatisfactorily explained by extremely high private discount rates. The CIMS model provides some answers to this issue (Jaccard, 2009), which is however beyond our scope here.

[2]Berndt and Wood are among the first to estimate a Capital, Labor, Energy, Materials, or KLEM production function of the aggregate United States economy (research leading to Berndt and Wood, 1975). As early as 1974, Jorgenson publishes a multisectoral KLEM model of the United States, which he applies to a prospective outlook up to the year 2000 (Hudson and Jorgenson, 1974).

[3]Löschel (2002) reviews technical change in economic models of climate or energy policy of the 1990s.

[4]Equipment stocks are understood in the broadest sense here. Consider notably the impact of urban forms, which evolve over centuries, on the share of individual vs. collective housing (with obvious consequences on heating requirements) or on the substitutability of public to private transport.

[5]Kennedy explicitly builds on Hicks: "A change in the relative prices of the factors of production is itself a spur to invention and to inventions of a particular kind—directed at economizing the use of a factor which has become relatively expensive"

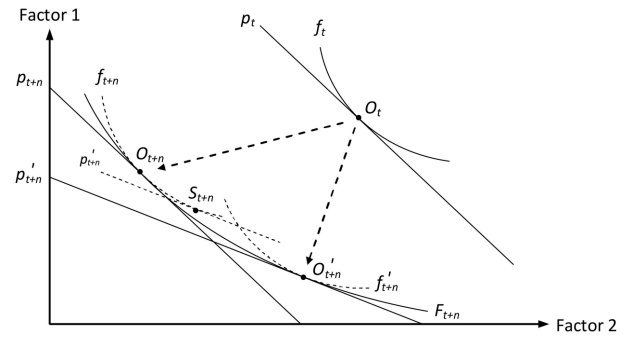


FIGURE 1 | Induced technical change as a dynamic production frontier.

the path-dependency of substitution isoquants. To illustrate this following Ruttan (2002) (**Figure 1**), let us consider some time t perturbation that shifts the $t+n$ price vector of some economy from p_{t+n} to p. Standard comparative statics searches the $t+n$ impact of the perturbation on the single pre-determined production function f_{t+n}, thus settling on S_{t+n}; the point is rather to acknowledge that the perturbation induces technical change leading to a specific f production function, which induces an O'_{t+n} optimum. It is therefore not any unique f_{t+n} function but the "dynamic production frontier" F_{t+n} enveloping all production functions reachable from year t that should structure economic analysis[6].

The second implication is that any macroeconomic model of energy-economy interactions is bound to turn to energy systems expertise if it is to describe with any relevance the evolution of the techniques that supply energy and convert it into service (heat, light, motion, information and communication), particularly in the short term. This has the unfortunate precondition of requiring modeling "capital" in a way closer to the physical capital actually mobilized in the processes of energy supply and demand, rather than as the ambiguous residual of value-added, labor costs subtracted, of the standard neoclassical approach.

Both implications ultimately echo the Cambridge controversy on capital, i.e., the unsettled neoclassical confusion of the economic productivity of investments for the technical efficiency of equipment[7]. Solow's own improvement of his initial model by the embodiment of technical progress in successive capital vintages (Solow, 1959) only partly addressed this critique: it introduced path-dependency but did not address the conceptual gap between flexible, continuous investment and discrete physical capital. Manne (1977) also contributed by focusing on a bottom-up representation of energy supply, thereby explicitly modeling physical capital; but he resorted to an aggregate KLE production function to project energy demand. For the same latter reason, the many subsequent CGE models that substituted some bottom-up representation of factor combinations (in varying shapes and hues) to the standard cost-minimizing energy producing sectors only partly improved on the initial KLEM model as well[8]. In the two following subsections we present our own two methods to try and overcome both identified shortcomings of the KLEM abstraction.

Calibration of Reduced Forms of Bottom-up Behaviors[9]

Our first two attempts at improving on the KLEM extension to Solow's model were literal applications of the "dynamic production frontier" interpretation of technical change. One early construction was applied work on the international climate negotiations covering 2030 projections of 14 major economies (Ghersi et al., 2003, cf. Section Reconciling the Equity and Efficiency of International Burden-sharing Climate Agreements below). A few years later, we published our improved method with illustrative projections of the global economy, also to 2030 (Ghersi and Hourcade, 2006, cf. Section The Continued Fable of the Elephant and Rabbit below). In both endeavors our bottom-up source was the POLES model of energy systems[10]. Focusing on climate policy, we purposely used POLES runs for a range of carbon prices broad enough to capture the "asymptotic" behavior of energy systems, i.e., the expected floor to energy intensities resulting from equipment inertia at the retained temporal horizon. To reveal our envelopes, we interpreted the results of POLES policy runs as the partial price derivatives of the static trade-off functions generated by each sequence of relative prices (**Figure 2**)[11]. We did so under the notable assumption that all prices not explicitly modeled in the partial equilibrium framework of POLES, including the capital and labor prices, are constant[12].

The complex practical implementation of the method and the extended geographical scope of our first experiment prompted us to focus on two-sector economies distinguishing only two goods, one energy good vs. one composite aggregate of all non-energy productions, while retaining an input-output framework easier to connect to BU expertise than value-added abstractions. In such a framework and for each economy modeled we thus required 3 "production frontiers": one for energy production, one for non-energy production and one for households' trade-off between energy and composite goods. For the sake of concision, in what follows we only report the stabilized method thoroughly explored in Ghersi and Hourcade (2006) (**Figure 2**)[13]. We also do not detail the prerequisite of constructing macroeconomic tables compatible with the POLES no-policy projection to 2030

(Hicks, 1932, p. 124). The work of Kennedy spurred much literature up to Magat (1979), which includes a thorough review.

[6]**Figure 1** is taken from Ghersi and Hourcade (2006), where it is further commented upon.

[7]Cohen and Harcourt (2003) provide a thorough review from a viewpoint similar to our own.

[8]See e.g., McFarland et al. (2004), Sue Wing (2006), Laitner and Hanson (2006), Schumacher and Sands (2007), Böhringer and Rutherford (2008), Fujimori et al. (2014), Cai et al. (2015).

[9]This section draws on Ghersi and Hourcade (2006).

[10]The POLES model is jointly developed by CNRS, IPTS, and ENERDATA, cf. http://edden.upmf-grenoble.fr/spip.php?article1471.

[11]**Figure 2** proposes scenario range alternatives to the carbon tax issue. These remain to be explored in future work.

[12]In the implicitly "constant" (deflated) currency of POLES.

[13]The method applied in Ghersi et al. (2003) differs in its more complex treatment of capital. We succinctly explain how in Section Reconciling the Equity and Efficiency of International Burden-sharing Climate Agreements below.

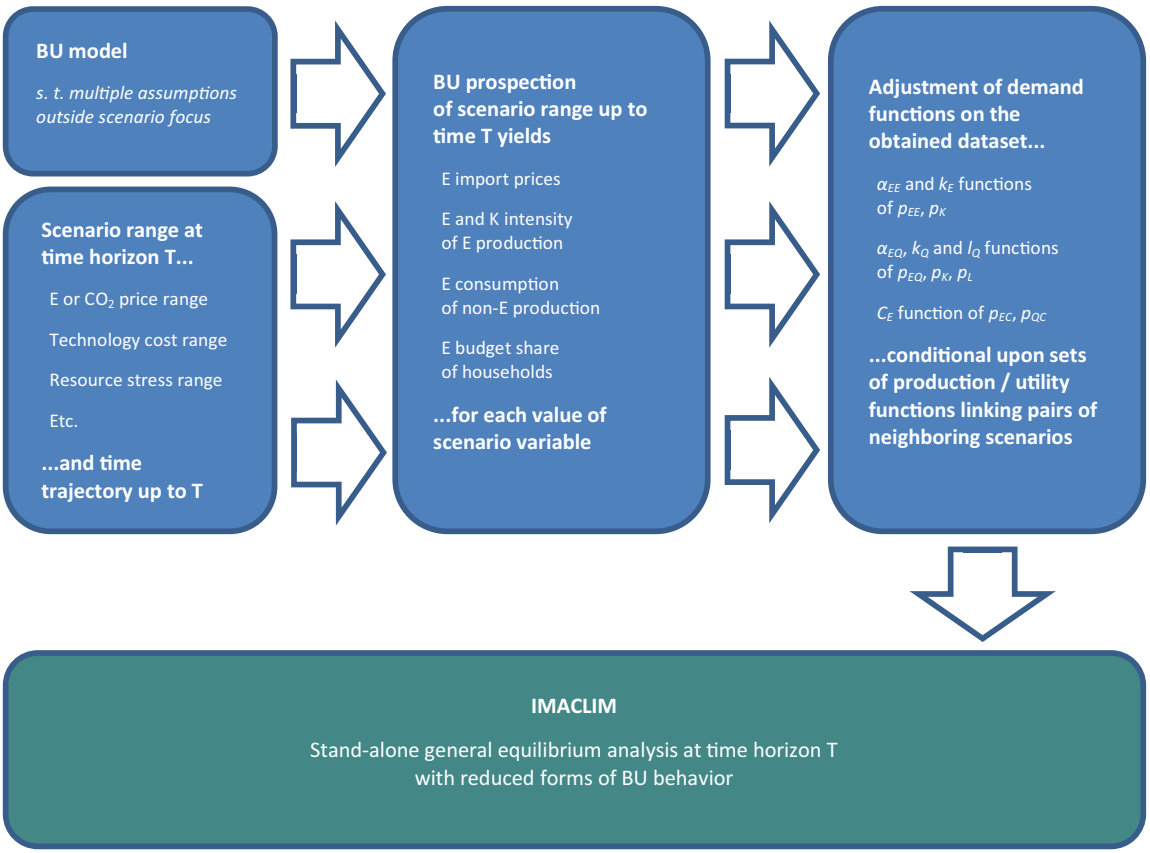


FIGURE 2 | Calibration of reduced forms of bottom-up behaviors.

to rather focus on the method of construction of the envelopes *per se* at the 2030 horizon[14].

Concerning energy production, for the sake of simplicity we consider that energy prices only alter the energy and capital intensities of the energy good but do not impact its labor and material intensities[15]. From POLES runs we draw sets of matching variations of factor intensities and prices for a range of carbon prices. Assuming constant non-energy prices allows interpreting as capital intensity shifts the variations of the monetary capital stock *per* physical unit of energy produced. We use the resulting dataset to calibrate the energy and capital intensities of energy production, α_{EE} and k_E, as functions of the ratio of their prices p_{EE} and p_K (**Figure 2**)[16].

As already hinted above, a non-negligible difficulty regards identifying actual capital expenses in the cost structure of the energy sector of national accounts. The remainder of value-added (VA) net of labor costs encompasses not only capital depreciation, but also elements as heterogeneous as interest payments, rents (on land, water, mineral, and fossil resources) and a mark-up reflecting static returns to scale and/or market characteristics. Using this remainder as an index of productive equipment is therefore questionable, all the more so as the capital costs of energy supply play a key role in policy assessment. We overcome this difficulty by distinguishing, in the non-labor VA, genuine equipment expenditures, which we calibrate on total gross fixed capital formation (GFCF) data net of housing investment, and the corresponding interest payments, which we estimate thanks to a limited set of exogenous assumptions: an average capital lifespan and a real interest rate[17].

Contrary to energy production, l_Q the labor content of composite (non-energy) production has a paramount influence on cost assessment, especially under imperfect labor market assumptions. We must therefore reveal a set of functions f to

[14]In Ghersi and Hourcade (2006) we detail our method of projection of use and resource tables under constraint of energy system information.

[15]The non-energy variable costs of energy production reported by POLES provide an estimate of the sum of material and labor costs. Because the labor content of energy production is low, its response to relative prices shifts can be neglected at a macroeconomic level. Changes in the non-energy intermediate consumption of new techniques may be more significant, but are not reported by the state-of-the-art of BU models—if they were, they could easily find their way into our methodology.

[16]In Ghersi and Hourcade (2006) we resort to the least-square adjustment of an arctangent specification, purposely selected to allow reproducing asymptotes to substitution possibilities.

[17]Interest payments are a percentage of equipment expenditures that is easily computed by setting an average lifespan of capital and a constant rate of growth of equipment expenditures together with a constant real interest rate over this lifespan—the two rates being assumed equal on a stabilized growth path.

compute the labor content l_Q, and as a matter of fact the capital content k_Q, necessary to the calibration of the envelope of these functions. We do so assuming that:

- All policy-induced *t+n* economies are on a steady equilibrium path, guaranteeing to each function *f* the first-order conditions of relative marginal productivities equating relative prices (for any two production factors).
- For a given output and around a given energy price p_{EQ}, the price elasticity of energy demand is derived from POLES considering a marginal increase of p_{EQ}.

For a selected functional form, there is a single *f* making these assumptions compatible with the no-policy prices and factor-demands vectors. The same mathematical property holds successively for every pair of equilibria separated by a marginal increase of the energy or carbon price.

In Ghersi and Hourcade (2006) we assume, considering their widespread use in the E3 modeling community[18], that CES functions of capital *K*, labor *L* and energy *E* approximate each real *f* at the neighborhood of the corresponding equilibrium. We calibrate CES_0 the CES prevailing at the (K_0, L_0, E_0) point of the no-policy projection by imposing (1) the linear homogeneity condition, (2) the first-order conditions at the no-policy equilibrium, and (3) the energy demand E_1 resulting from a marginally higher energy price p_{EQ1} under constant other prices and output, as computed by POLES. CES_0 then yields the optimal K_1 and L_1 induced by the marginally higher price regime. We iterate this method on the newly defined (K_1, L_1, E_1) equilibrium, considering the impact of a further marginal energy price increase, as again computed by POLES. This allows the successive identification of equilibrium (K_i, L_i) compatible with POLES information on (p_{EQi}, E_i) couples over the whole spectrum of analysis. Note that, even though we assume a CES function at the neighborhood of each equilibrium, the resulting envelope has no reason to exhibit a constant elasticity of substitution, unless in the implausible case of a constant price elasticity of *E* over the range of policies explored. We use the resulting set of prices (p_{EQ}, p_L, p_K) and factor demands (α_{EQ}, l_Q, k_Q) to adjust functional forms of conditional demands of the three factors (**Figure 2**).

Turning to household behavior, POLES does not systematically report on the proper arguments of utility functions, i.e., energy services (heating, lighting, passenger-kilometers, etc.) whose variations may differ from those of energy consumptions *per se* thanks to efficiency gains. Our method consequently focuses on Marshallian demands without revealing the underlying set of utility functions. To calibrate an envelope of the Marshallian energy demand C_E, we first translate in budget share terms the changes in household energy demand computed by POLES, assuming that POLES implicitly considers constant total household expenditures[19]; we then adjust Marshallian demand functions by linking variations of this share to shifts of the energy and composite price, p_{EC} to p_{QC} ratio—again, considering that POLES implicitly considers constant non-energy prices.

Last but not least, we derive the impact of carbon constraints on total factor productivity in the composite sector[20] from a comparative-static analysis of an endogenous growth mechanism: we impact all factor intensities with a Hicks-neutral technical progress coefficient that is a function of cumulated investments. Assuming all *t+n* projections on a steady equilibrium path justifies using variations of the *t+n* equipment expenditures as a proxy of those of cumulated investment[21]. Under this specification, the crowding-out effect of mobilizing more resources in the production and consumption of energy is not accounted for through the allocation of a fixed capital stock or GFCF. Rather, firms finance their investments (equipment expenditures augmented by interest payments) under the double constraint of market balances—investment goods are produced by the composite sector—and of the ability of households' purchasing power to sustain the resulting price increases. Cumulated investments and the induced productivity of the composite sector consequently align.

Coupling Through Iterative Convergence of Linking Variables Trajectories

The ambitious "reduced form" method of Section Calibration of Reduced Forms of Bottom-up Behaviors suffers from two major drawbacks. First, its extension to a multi-sectoral framework raises conceptual issues regarding the form that the trade-off functions united by envelopes could take if further disaggregation of inputs or consumption goods were considered. Secondly, it is conditional upon the carbon pricing trajectory surmised in the bottom-up source. There is no question that simple options like a constant vs. a linearly increasing vs. an exponentially increasing tax yield contrasted "response surfaces" at the retained horizon, which implies that the calibration of reduced forms must be performed again for each specific pricing trajectory. What is more, the correspondence of BU and TD pricing trajectories is not as straightforward as it appears: the question of the reference against which the tax should be measured, i.e., the price deflator that must be applied to it to reveal its real time profile, is complex[22].

[18]E.g., in models as G-Cubed, MS-MRT, SGM, EPPA. Cf. respectively McKibbin and Wilcoxen (1995), Bernstein et al. (1999), Fisher-Vanden et al. (1993), Babiker et al. (2001).

[19]Note that the assumptions of constant expenditures, constant composite consumption and constant composite price are incompatible with variations of the energy expenditures. Given necessarily constant non-energy prices, we prefer to consider a constant income (more compatible with the fixed GDP assumption of POLES) rather than a constant consumption of the composite good.

[20]Because energy models increasingly account for the impacts of learning-by-doing and R&D efforts on the costs of energy technologies, the envelope of energy production functions is assumed to embody such effects and is therefore not subject to productivity adjustments.

[21]We calibrate the specification to allow a doubling of cumulated investment triggering a 20% cost decrease, based on 1978–2000 time-series for France and OECD. Sensitivity analyses demonstrate that variations of the elasticity of total factor productivity to real investment do not qualitatively affect the model.

[22]In a TD framework the consumer and GDP price indexes come to mind as deflators of respectively the consumer's and producer's carbon taxes. Because the two indexes generally do not match, one uniform tax of the BU framework translates into two distinct taxes in the TD framework—which does not simplify policy analysis or the interpretation of modeling results. Besides, both deflators are endogenous to the TD framework. Ultimately, the carbon pricing trajectory

For these reasons we developed a second avenue of coupling IMACLIM-S to bottom-up modeling. Rather than aiming at a standalone version of IMACLIM-S embarking bottom-up expertise (**Figure 2** above), this alternative method consists in a systematic joint running of IMACLIM-S and the linked bottom-up model (**Figure 3**). Consistency between the two models builds on an iterative exchange of the largest possible set of shared variables or parameters, up to convergence of all elements of this set[23]. More precisely, the method consists in:

- Harmonizing all the exogenous parameters common to both models, i.e., typically demography, but also possibly international energy prices, if these are exogenous to the bottom-up model as they are in the open-economy single-region versions of IMACLIM-S[24].
- Forcing, in the input-output framework of IMACLIM-S, bottom-up simulation results on (1) the international price of energy commodities; (2) the energy consumptions of households and the energy intensities of productions; (3) the capital intensities of energy supplies and, as far as possible, energy end-uses; (4) the volume of energy exports to foreign markets, together with that of domestic output of primary fossil energies, disaggregation permitting.
- Running IMACLIM-S under constraint of these exogenous data to compute various economic indicators (such as GDP, output at various aggregation levels but also possibly relative price variations on non-energy markets), which are sent back to the bottom-up model to serve as drivers of energy demand in a renewed simulation.
- Iterating the exchange of inputs and outputs between the two models up to convergence, in both modeling systems, of the exchanged data.

sustaining any modeling output of IMACLIM-S with reduced forms is not easily pinpointed.

[23]The approach is similar to that of Dai et al. (2015) but with data exchange extended to prices; or to that of Fortes et al. (2014) or Labriet et al. (2015), although energy prices, rather than exogenously taken from the bottom-up source, remain largely endogenous variables of IMACLIM-S (cf. below). The idea of coupling economic and energy models *via* iteration of linking variables exchanges up to convergence dates back at least to Hoffman and Jorgenson (1976). It was picked up throughout the years to present times (cf., e.g., Messner and Schrattenholzer, 2000; Schäfer and Jacoby, 2006; Martinsen, 2011) where it appears to flourish (cf. the above 3 recent papers but also Igos et al., 2015).

[24]Contrary to the endogenous primary resources markets of the global multiregional version of IMACLIM, IMACLIM-R (Sassi et al., 2010).

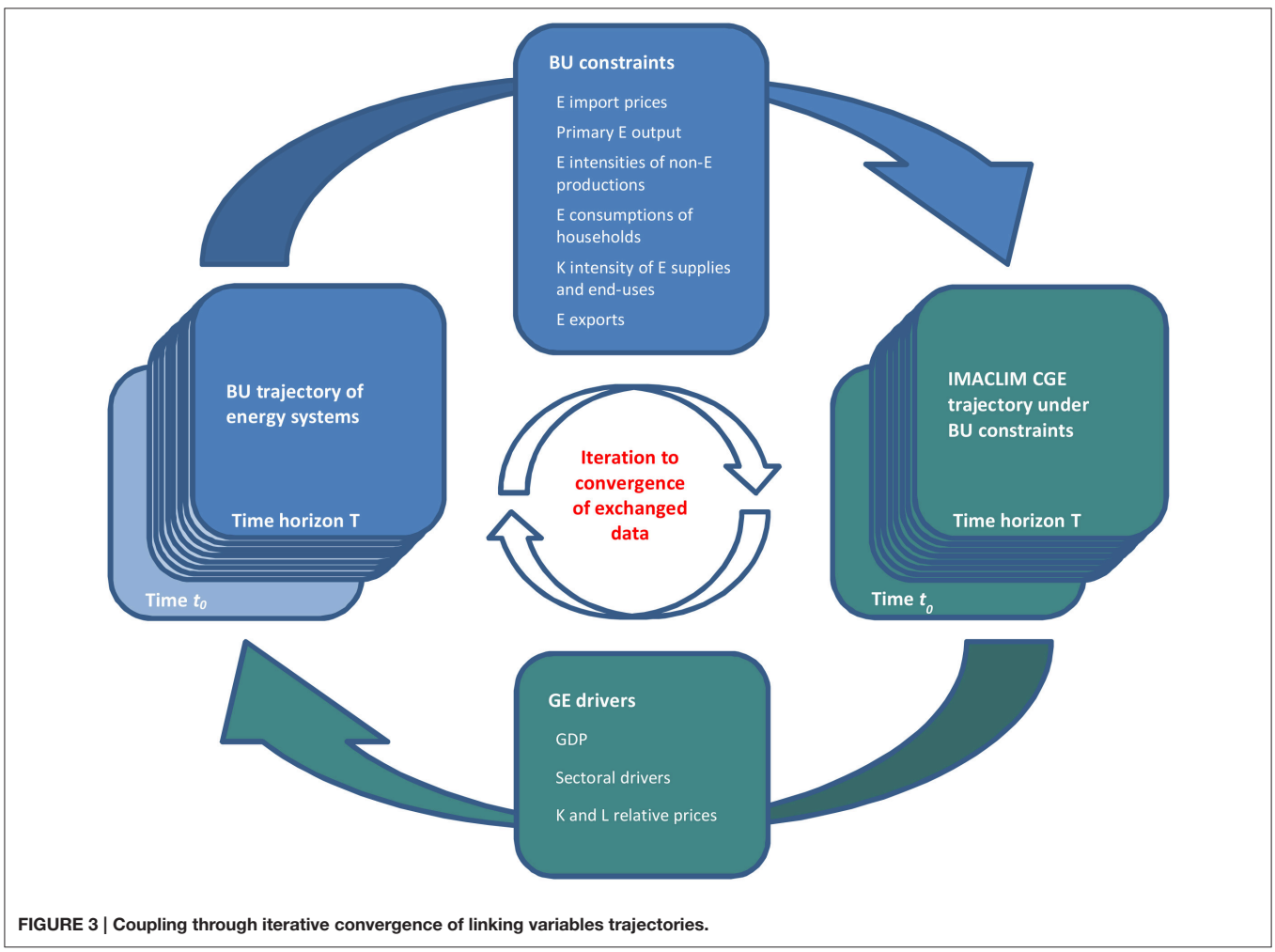


FIGURE 3 | Coupling through iterative convergence of linking variables trajectories.

The outcome of this procedure is a consolidated set of variables consistent with both the macroeconomic framework of IMACLIM-S and the detailed energy systems modeling of the linked bottom-up model. The method naturally calls for a dynamic recursive implementation of IMACLIM-S rather than for the static comparative approach of Section Calibration of Reduced Forms of Bottom-up Behaviors: it would be pointless to feed back a deviation of GDP and other demand drivers at some unique time horizon of the initial bottom-up trajectory, while maintaining the initial trajectory up to that horizon. The exchanged data is thus in fact a set of time trajectories of the various linking variables from some common initial year (the later of the two models' starting years) to some common prospective horizon (**Figure 3**). Looking back at the interpretive framework of **Figure 1**, any envelope F_{t+n} now remains implicit and it is the whole path of O optima, from O_t to O_{t+n}, that the iterative convergence reveals—conditionally upon a parameterized context (regarding policies, technologies, resources, etc.) that changes from one run to the other.

However, contrary to the reduced form alternative the iterative method does not benefit from multiple bottom-up runs that could allow settling non-energy inputs trade-offs. One central question is again that of end-use capital, i.e., the supplemental investment in machinery, home equipment, housing insulation and transportation vehicles at the source of energy efficiency. This energy efficiency investment is increasingly explicit in bottom-up models and could eventually be fed to IMACLIM-S alongside energy demand data and energy supply capital data. Notwithstanding, the question of labor demand dynamics in non-energy sectors, together with the broader impact of relative price variations on trade-offs disconnected from energy matters, would remain unsettled. In applied work, we have thus so far resorted to conventional behavioral functions, as indeed the CES function, to settle all trade-offs but these imported from the linked bottom-up model (cf. Sections An Outlook on the Macroeconomic Impacts of Electric Vehicles Penetration in EU28 and Sovereign Risk and Energy: the RISKERGY Program below).

Concerning the behavior of producers, we thus assume that the output Y of all goods (whatever our level of disaggregation) is a function of inputs of the primary factors K and L, aggregate energy E and the composite non-energy good M[25], which combine in a nested structure echoing recent literature[26]. At the bottom of the structure (**Figure 4**), capital K and labor L trade off to produce a KL aggregate—the capital intensity of energy production(s) is however simultaneously subject to specific productivity variations, to mirror the energy supply investment dynamics reported by the bottom-up model (cf. below). At the second tier of the input structure the KL and E aggregates combine into a KLE aggregate—which allows inferring KL intensity from the E intensity imported from BU modeling[27]. At the third tier of the input structure, the KLE and $\alpha_M\ Y$ aggregates combine into output Y. Note that such treatment of non-energy input substitutions contradicts our induced innovation take on technical progress. We settled on it temporarily for sheer convenience, considering its widespread and well referenced use, but amending it is part of our further research agenda (cf. Section Conclusion and Further Research Agenda below). Nonetheless, the induced production frontier interpretation still holds as far as substitution between the energy bundle and value-added is concerned, thanks to the natural path-dependency of the bottom-up description of energy systems.

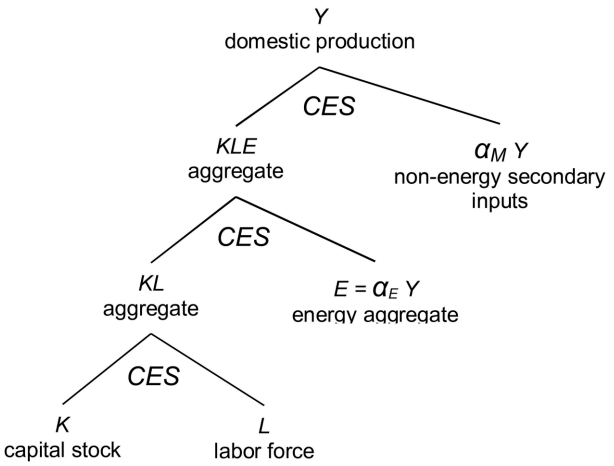


FIGURE 4 | Nested production structure in IMACLIM-S versions iterating to convergence with BU models.

Turning to households, we directly import the volume(s) of energy consumption(s) from BU modeling and devote the remainder of the consumption budget to the composite good. In disaggregate versions, similar to production we must surmise some behavioral function to settle the competition between components of the composite good—although we will see below (Section An Outlook on the Macroeconomic Impacts of Electric Vehicles Penetration in EU28) that more than energy consumptions can be inferred from bottom-up modeling.

Regarding energy trade, which we never properly addressed in the reduced form alternative, one difficulty is that most bottom-up approaches only compute net imports as the difference between domestic output and consumption. We consequently have to disaggregate an evolution of gross exports and imports from the reported evolution of net trade[28]. We then choose to force the resulting exports volumes and let imports balance

[25] In disaggregated version the composite M is an aggregate of non-energy goods, whose trade-off we must again settle resorting to some behavioral assumption. In the applied work reported Section An Outlook on the Macroeconomic Impacts of Electric Vehicles Penetration in EU28 we fall back on the Leontief assumption of fixed intensities.

[26] Van der Werf (2008) and Okagawa and Ban (2008) econometrically establish the superiority of the chosen structure over other possible choices (i.e., substituting K to E then the KE aggregate to L). They also are precious sources of estimation of the elasticities of substitution of CES functions at each tier of the structure.

[27] In disaggregated version, this E intensity is a vector of energy intensities reflecting some energy mix.

[28] In both the applied works of Section An Outlook on the Macroeconomic Impacts of Electric Vehicles Penetration in EU28 and Sovereign Risk and Energy: The RISKERGY Program we do so by assuming an inverse evolution of imports and exports, which we apply to gross imports and exports volumes of the calibration year as reported by IEA balances.

demand. Working on major oil and gas producers (cf. Section Sovereign Risk and Energy: The RISKERGY Program) also incites us to control the price of energy exports, to reflect BU-modeled rents variations on international markets. In our approach to production, prices build up from costs—although we still consider, as we did in the reduced forms method, a fix mark-up stemming from the differentiation of capital depreciation and profits in value-added. To account for the state of international energy markets as depicted by the linked BU model, we simply adjust a specific margin (rent) on energy exports.

Similar to export prices, we want the production prices of energy to mirror variations computed in the technology-explicit framework of the linked BU model. These variations mainly reflect the capital intensification of energy supply resulting from increasingly costly conventional resources, the higher costs of the 3rd (and 4th) nuclear generation, stronger environmental constraints, the penetration of renewables, the capital requirements of "smart" distribution networks, etc. They could also reflect increased "operation and maintenance" (O&M) costs induced by shifts of the technology mix (or mixes, depending on the disaggregation of energy in IMACLIM-S), which should mean a change of the labor intensity of energy production—although the split between labor costs, materials, and indeed possible genuine capital expenses in the O&M aggregate attached to BU technologies is generally unknown.

Considering the minor share of O&M expenses in total energy costs, in applied work so far we translate all energy costs variations in capital intensity shifts. To do so we develop an original additional iterative method[29]: (1) along the trajectory computed by IMACLIM-S we reveal an approximation of BU energy costs by combining base year prices and intensities of non-energy factors (capital, labor and materials) with current year energy price(s) and intensity (intensities); (2) we compare the resulting energy cost trajectory with that computed by the bottom-up model (3) we compute the capital intensity change that allows both trajectories to match, all other things equal; (4) we exogenously adjust the capital productivity of energy production to reflect this trajectory of capital intensity changes; (5) we iterate until our IMACLIM-S approximation of BU energy costs evolves as the genuine BU trajectory.

Despite these extensive developments, consistency between the BU model and IMACLIM-S remains imperfect because the BU model settles production and end-use technology competition under the assumption of constant relative non-energy prices—which biases the relative prices of energy vectors—rather than the relative prices computed in the macroeconomic framework of IMACLIM-S. This is a shortcoming of soft linking approaches stressed by Bauer et al. (2008)—although for a slightly different soft linking experiment. It could however be addressed by an extension of the set of linking variables, as we further discuss when outlining our future research agenda (cf. Section Conclusion and Further Research Agenda below).

DATA HARMONIZATION REQUIREMENTS BUILDING HYBRID ENERGY/ECONOMY ACCOUNTING TABLES

Somewhat surprisingly, the first experiments of linking IMACLIM-S to bottom-up analysis did not mechanically lead us to compare the aggregate energy volumes and prices common to both modeling systems. The reason was our use, in early versions of IMACLIM-S, of the standard CGE model assumption of normalized production prices: without loss of generality we could set the "producer" (net of trade and transport margins and of sales taxes) prices of all goods and services disaggregated by IMACLIM-S to 1 at our base year[30], thus forbidding any comparison of the consecutive selling prices and sold volumes with corresponding energy data. As a consequence, we resorted to "base 1" trajectories to translate bottom-up results into the framework of IMACLIM-S.

However, despite the mask of normalization—and indeed that of differing monetary units—, massive discrepancies rapidly caught our attention. A first one regarded energy expenses, which varied between IMACLIM-S and the connected bottom-up model by more than acceptable statistical discrepancies either in their total or for those economic agents similarly aggregated in both systems. Reasons for such discrepancies are that the energy expenses of national accounts build on data that (1) are, for many years but a few reference years, the products of surmised energy intensity gains and output volume indexes, without guarantee of matching the explicit energy consumptions reported in energy balances (2) are constructed as expenditures of branches of activities, which are in turn corrected to yield expenditures of products produced by branches[31]; (3) undergo some statistical treatment to add up to an equilibrium of uses and resources.

A second, more subtle discrepancy regarded the distribution of volumes of energy consumptions among economic agents, which substantially changed the consequences of carbon or energy policies focused on some of them—e.g., targeting households emissions would mean targeting a quite different percentage of total carbon emissions in IMACLIM-S than in the linked bottom-up model. In IMACLIM-S this distribution of volumes across agents is univocally induced by the assumption of a unique producer price, which implies selling prices differentiated by taxes only. This strongly contradicts energy price data in many countries where the average energy price of firms is quite below that of households, and masks indeed a wide

[29]This iteration is embedded in the larger iteration of linking variables exchanges between models, i.e., it is performed for each run of IMACLIM-S under updated constraints from the linked BU model. In the case of the applied work of Section An Outlook on the Macroeconomic Impacts of Electric Vehicles Penetration in EU28 convergence of cost trajectories at a 10^{-3} precision requires in the order of 5 iterations of capital productivity changes.

[30]The year at which the model is calibrated to match national accounts statistics.

[31]Surprisingly enough, the French national statistics institute (INSEE) indicates that "product transfers" impact but a few sectors, mainly related to agriculture (INSEE, 2012). The 4 energy sectors of the 88-sector INSEE IO table are indeed barely concerned by product transfers corrections, and in a positive way only (which means that some firms mainly involved in non-energy activities have a minor share of their activity on energy markets). This implies that energy firms do not sell anything but energy products strictly speaking, an arguably questionable statistics.

variety of prices faced by producing sectors, depending on the average size of the firms they aggregate.

The threat of strong biases to our policy analysis thus prompted us to envisage basing our modeling on hybrid matrixes reconciling national accounts with energy-specific data. Over the course of our applications of IMACLIM-S we developed two distinct methods to that effect, which the two following subsections detail, while a third subsection addresses the important connected question of modeling agent-specific energy prices.

Building from IEA Energy Balances and Energy Price Data[32]

The most extensive energy/economy data hybridizing method developed for IMACLIM models builds on full-fledged energy balances and energy price data, typically that available from the International Energy Agency (IEA). Its starting point consists in reorganizing the disaggregated energy balance (in million tons-of-oil equivalent, Mtoe) into an input-output format compatible with that of national accounts. This is a much more time-consuming and data-intensive procedure than it appears, as it entails, for dozens of energy products[33].

- Correcting, when working on an IEA region that aggregates different countries, the reported imports and exports. For both accounts the IEA indeed only sums up the data of each country within the region, without subtracting intra-regional trade. For, e.g., the European Union this is an absolutely necessary step.
- Absorbing statistical discrepancies, by e.g., a homogeneous adjustment of all uses in one direction and of all resources in the other, to bridge the gap between the two totals.
- Reallocating international bunkers and internal transport fuel consumptions to exports and domestic consumption. The IEA treats energy consumptions from a geographical perspective, whereas input-output data aggregate the economic accounts of resident businesses—two quite orthogonal perspectives. This is indeed one of the biggest data treatment challenges, considering how difficult it is to obtain the data required to perform it. The stakes are however high, with the energy intensity of transports recognized as one of the main deadlocks of energy transition.
- Vertically integrating (1) the energy consumptions motivated by electricity auto-production—which only appear as primary energy consumptions in national accounts; (2) the product transfers from the refining industry: IEA balances detail how refineries recycle some of their outputs as inputs; this is irrelevant from a national accounts perspective, which only record what refineries actually buy on markets, and eventually sell back to markets.
- Absorbing stock variations by adjustment of resources or uses (depending on their signs), similarly to our usual practice on national accounts—the alternative being to aggregate stock variations to investment, but this has the undesired consequence of immobilizing energy flows into productive capital.
- Distributing among productive sectors (including those producing transportation services) and households the energy consumptions of transports, which are isolated as such in energy balances. This is the second most delicate operation, which suffers from too-rare statistics on business vehicle fleets and particularly the attached fuel consumptions.

One particular difficulty regards the energy consumptions of energy sectors. The input-output tables of many economies report auto-consumptions of the electricity or the natural gas sector that imply volume consumptions flagrantly above those reported in energy balances[34]. These auto-consumptions are dominantly commodity trade on liberalized markets, and only quite marginally genuine consumptions. In applied work so far we have hesitated between (1) stripping down auto-consumptions to genuine energy consumptions and consequently moving the value-added of commodity trading to non-energy sectors when re-balancing our hybrid national accounts, cf. below; or (2) acknowledging energy trade, but this implies breaking the link to energy balances and falling back on national accounting data only[35].

The second step of our hybridizing procedure consists in complementing the obtained IO table of energy flows with a table of applying energy prices (in currency *per* volume, e.g., Euros *per* ton-of-oil equivalent, €/toe). This is again a delicate step, as IEA databases provide energy price statistics with limited disaggregation and must therefore be augmented with other data sources, either national or international[36]. Notwithstanding, the precise average price applying to some "cell" of the constructed IO table of volume consumptions often remains unavailable. However, the market prices of primary energy products can be found and applied uniformly across (mostly industrial) agents with some confidence; statistics for gas and electricity usually report prices dependent on the volume of consumption, which offer ranges for assumptions—or which can confirm, in last resort, the average price resulting from crossing national accounts expenses and the IO table of energy flows. There is no denying that this step leaves room for judgment, but at the very least it forces to make explicit, educated choices on the prices of the energy expenditures of all economic agents.

[32] This section draws on yet unpublished research co-authored with Emmanuel Combet, Julien Lefévre, and Gaëlle Le Treut at CIRED.

[33] In their most disaggregated format IEA energy balances detail 71 resources and uses for 60 energy products.

[34] For example, the 2010 IO table of France by INSEE reports a 45 billion Euro auto-consumption of the gas and electricity production and distribution sector. This is incompatible with the *ca.* 8 million ton-of-oil equivalent (toe) consumptions of said industries reported by the IEA energy balance–which cannot be valued at more than €1000 *per* toe on average.

[35] In our applied work on the European Union (cf. Section An Outlook on the Macroeconomic Impacts of Electric Vehicles Penetration in EU28), sectoral disaggregation has prompted us to move commodity trade from the energy sectors to the composite sector (which includes trading activities). Conversely, in the sectorally aggregated but geographically extensive framework of our RISKERGY work (cf. Section Sovereign Risk and Energy: the RISKERGY Program) we have decided to rather adjust the auto-consumption of our aggregate energy sectors *prorata* the adjustment of other energy uses by the hybridizing process (cf. Section Building from BU Model Variables at Base Year below).

[36] CIRED makes frequent use of the ENERDATA database at http://www.enerdata.net.

The third step of our hybridizing method is to substitute the disaggregated energy expenses obtained by the term-by-term multiplication of the volumes and prices tables, to that pre-existing in the system of national accounts. We consequently adjust other components of the system to maintain the accounting identities, under the purposely set constraint of not modifying any of the cross-sectoral totals of uses or resources in the economy—which notably implies that we maintain the total value-added of domestic production. We do so (1) on the uses side, for the intermediate consumptions of sectors, household consumption and exports[37], by compensating the difference between the recomputed energy expenses and the original statistics through an adjustment of the expense on the most aggregated non-energy good—a composite remainder of not specifically described economic activities, usually encompassing all service activities in E3 models; (2) for the energy sectors resources, by adjusting all non-energy expenses (including value-added components) *pro rata* the induced adjustment of total energy expenses; (3) for the intermediate consumption of sectors, labor and capital costs, input and product taxes and imports, by compensating the difference between the recomputed resources of the energy sectors and the original statistics through an adjustment of the resource of the same most aggregated non-energy good.

Building from BU Model Variables at Base Year

As underlined in the preceding section, the construction of detailed prices and volumes tables of energy consumptions compatible with national accounts is both time-consuming and data-intensive. At the time of writing, CIRED has thus only treated France, Brazil, South Africa and the European Union (the 28-Member State aggregate) at such level of precision. In the course of recent applied work (cf. Section Sovereign Risk and Energy: The RISKERGY Program), faced with the necessity to articulate an aggregated IMACLIM-S to a large number of countries, we consequently came up with an alternative, much simpler method of hybridizing IMACLIM-S with BU data.

This method bypasses the tedious data collection and treatment effort aimed at recomposing a BU-consistent table of energy expenses by constructing this table directly from aggregate variables of the linked bottom-up model. We implemented it to reconcile the base years of aggregate 2-sector models initially calibrated on the GTAP database with data for the corresponding year of, again, the POLES model[38]. The procedure is similar but not identical to that of Section Building from IEA Energy Balances and Energy Price Data above.

- A first step similar to Section Building from IEA Energy Balances and Energy Price Data is to substitute to GTAP energy consumptions of the composite (non-energy) sector and of households the corresponding price x volume statistics that can be aggregated from POLES data for the relevant year.
- A second, quite specific step is to adjust the energy expenses of the energy sector *prorata* the adjustment of the sum of other uses induced by the first step. This rough procedure is prompted by the quite specific issue of commodity trading, which we introduced Section Building from IEA Energy Balances and Energy Price Data above (cf. also footnote 35).
- A third and a fourth step, both identically shared with the approach of Section Building from IEA Energy Balances and Energy Price Data, consist in adjusting all non-energy resources of the energy sector *prorata* the adjustment of its energy expenses induced by the second step, and in compensating all induced changes of the uses and resources of the energy sector by adjusting the corresponding uses and resources of the non-energy sector, thereby guaranteeing conservation of all cross-sectoral accounting totals: intermediate, household and public consumption, investment and exports on the side of uses; intermediate consumption again, labor costs, capital costs, input and product taxes and imports on the side of resources.

Note that this alternate method is all the more relevant as the linked bottom-up model has a description of energy markets compatible with that of IMACLIM-S. It is the case of the POLES model, which is closer to an economist's view of energy matters than, e.g., models of the MARKAL family, such as the TIMES PanEU model with which we worked on the EU28 economy (cf. Section An Outlook on the Macroeconomic Impacts of Electric Vehicles Penetration in EU28 below). Also, one shortcoming of the method is that the sectoral disaggregation of the resulting hybrid matrix is limited to that explicitly available in the linked bottom-up model—even at the highest possible 2-sector aggregation of energy vs. non-energy goods the question of households' automotive fuel expenses is not easily settled[39].

Calibration of Agent-specific Prices Through Specific Margins

Nothing forbids applying the standard uniform (normalized) pricing rule of CGE models to the hybrid accounts resulting from the methods of either Section Building from IEA Energy Balances and Energy Price Data or Section Building from BU Model Variables at Base Year. As we already stressed, this has the joint consequences of limiting, for each good in the model, price differences across consumers to tax specificities; and of biasing the distribution of energy volume consumptions. The only advantage of hybridizing data would therefore be to have pinpointed the economic

[37]National accounts treat the energy consumptions of public administrations as these of an aggregate public service sector, whose uses focus on the final consumption of public administrations. For this reason, the direct final energy consumptions of public administrations are nil. Understandably, gross fixed capital formation does not consume (immobilize) any energy flow either, at least notwithstanding stock variations (cf. above).

[38]In the 8.1 version we used, GTAP details harmonized 57-sector input-output tables for 134 countries or regions covering the globe, cf. https://www.gtap.agecon.purdue.edu/databases/v8/default.asp (accessed July 2015).

[39]In the applied work of Section Sovereign Risk and Energy: The RISKERGY Program we crudely attribute to households all fuel consumptions of light duty vehicles. A non-negligible part of such consumptions are paid for by firms, though. Conversely, some fuel consumptions attached to light trucks but also to boats and even airplanes (although statistically negligible) are paid for by households.

value of the energy sectors and their actual cost shares in productions. It might appear counterproductive to thus discard all the collected information on the real distribution of consumption volumes and on agent-specific prices. However, this is indeed what modelers do when exploiting the hybrid matrixes derived from the widely used Global Trade Analysis Program (GTAP) database in standard CGE approaches. In our instance, having produced the hybrid tables ourselves naturally lead us to fully exploit them by departing from standard CGE modeling practice and introducing agent-specific prices.

The question of how to model agent-specific prices should of course be linked to the reasons why prices faced by different economic agents for an identical volume of some energy good actually vary. Two main reasons prevail. One reason is that aggregation masks the heterogeneity of energy goods. This is obvious in models where energy is one single good, within which the volume mix of natural gas, electricity, petroleum products, etc. may vary substantially from one economic agent to the other. Because the prices of all vectors *per* energy unit are not aligned (for interesting practical reasons beyond our scope here) the average price of the energy consumption of agents varies too. But this holds too in models where energy is more disaggregated. Not mentioning primary forms, E3 models typically distinguish coal products vs. natural gas vs. petroleum products vs. electricity. The plural in coal and petroleum products betrays product heterogeneity that is generally echoed in pricing—cf. simply the contrasted prices of diesel and gasoline fuels in most countries.

A second reason for agent-specific energy prices is that strictly identical energy goods, as typically a kWh of electricity or a cubic meter of natural gas, face distribution costs that sharply increase from centralized (e.g., large firms) to decentralized (e.g., small commercial or residential customers) consumptions. These increasing costs are a complex blend of equipment costs, maintenance costs and even harder-to-assess specific costs attached to varying contractual commitments. It is doubtful that any data outside undisclosed corporate data could allow a meaningful distribution of these extra costs on the cost structure of energy production.

Faced with this lack of information, in our applied work tackling this issue (that reported Sections An Outlook on the Macroeconomic Impacts of Electric Vehicles Penetration in EU28 and Sovereign Risk and Energy: The RISKERGY Program below) we decide to introduce a set of "specific margins" aggregating, for each economic user category, deviations from the average basic price emerging from the cost structure. By construction the aggregate margins compensate and thus do not alter the balance of each energy sector (or the energy aggregate) resulting from the hybridizing processes of either Section Building from IEA Energy Balances and Energy Price Data or Section Building from BU Model Variables at Base Year above. Let us underline at last that, to the best of our knowledge, no other modeling team has thus tackled the issue of agent-specific pricing.

HYBRID MODELING RESULTS

In this last section we briefly present 4 numerical applications of the hybrid modeling methods outlined in preceding sections[40]. Two of the 4 applications concern the calibration of reduced forms on results of the POLES model, but date back to times when we had not developed data harmonization yet (**Table 1**). Two later applications build on harmonized data but diverge in their geographical vs. sectoral disaggregation, as well as on the coupled bottom-up model (**Table 1**).

Reconciling the Equity and Efficiency of International Burden-sharing Climate Agreements

Our first attempt at coupling IMACLIM-S to bottom-up modeling is applied work on the equity and efficiency of international climate burden-sharing agreements (Ghersi et al., 2003). The aim of the paper is to clarify the equity-efficiency debate around post-2012 international greenhouse gases (GHG) emissions control. To this end, it explores the 2030 consequences of two contrasting GHG quota allocation rules: a "Soft Landing" rule extending the Kyoto protocol approach to burden sharing and a "Contraction and Convergence" rule progressively departing from the Soft Landing burden sharing to arrive at a global common *per capita* emissions endowment in 2050. The assessment of both rules leans on runs of 14 IMACLIM-S models embarking reduced forms of POLES modeling outputs for 14 countries or regions covering the globe.

The reduced forms calibration method is a similar but not identical predecessor to the stabilized approach of Section Calibration of Reduced Forms of Bottom-up Behaviors above. The 14 regional models aggregate the two energy and non-energy products of Section Calibration of Reduced Forms of Bottom-up Behaviors. Calibration on POLES results thus concerns the energy and fixed capital consumption intensities of energy production, the energy vs. labor trade-off in the production of the composite good—the capital intensity of the composite good is the remainder of total fixed capital consumption, which is itself a constant share of GDP; this is where the method diverges most from that of Section Calibration of Reduced Forms of Bottom-up Behaviors—and the energy consumption of households, constrained as a linear expenditure system, i.e., a Cobb-Douglas with constant budget shares beyond some basic need (floor consumption) of both goods. It also regards the carbon intensity of all these consumptions, modeled as a sheer function of the carbon price. The data on which each regional calibration is performed is a set of 60 POLES runs that explore the consequences of carbon pricing linearly increasing from 0 in year 2000 to between 10 and 600 US dollars *per* metric ton of carbon in 2030.

[40]All numerical applications reported here, boiling down to the resolution of either one single set (at time horizon T) or a succession of sets (from time 0 to time horizon T) of simultaneous non-linear equations, are successfully conducted in Microsoft Excel with the help of the Excel solver. Depending on sectoral disaggregation, the number of solver variables varies between 12 and 67 (with many more variables functions of these).

TABLE 1 | The settings of 4 hybrid modeling experiments involving IMACLIM-S.

Section	BU source	Coupling type	Hybrid data	Scope
Reconciling the equity and efficiency of international burden-sharing climate agreements	POLES model	Reduced forms calibration	No	14 2-sector economies
The continued fable of the Elephant and Rabbit	POLES model	Reduced forms calibration	No	World, 2 sectors
An outlook on the macroeconomic impacts of electric vehicles penetration in EU28	TIMES PanEU model	Convergence through iteration	Yes	EU28, 12 sectors
Sovereign risk and energy: the RISKERGY program	POLES model	Convergence through iteration	Yes	All POLES countries, 2 sectors

The determinant impact of coupling on analysis is caught at a glance when considering the variety of policy responses reported by POLES across economies (**Table 2**). The (non-energy) firms and households of different economies exhibit flexibilities that are contrasted both in their magnitudes and "curvatures"—betrayed by how far from a six-fold increase of energy or carbon savings the six-fold increase of carbon pricing triggers. Notwithstanding the ability of CES functions to capture the said curvatures (which we comment upon Section The Continued Fable of the Elephant and Rabbit below), it is quite unlikely that any conventional CGE approach could have built on enough region-specific calibration data to embark this wealth of information on the differentiated elasticities of future regional energy systems[41].

Accounting for this diversity allows proper elicitation of the winners and losers of the two, somewhat polar policy courses explored—although one important conclusion of the paper is that assessments differ whether considering aggregate GDP (a commonly used measure of efficiency) or induced carbon pricing (an arguable measure of implicit distributive impacts). The analysis is further enriched by considering free vs. auctioned allocation of emission allowances, together with the existence or not of an international market for permits, which also impact on the comparison of allowance rules.

The main methodological shortcoming of this first hybrid modeling endeavor is that it does not build on hybrid energy-economy tables: we disaggregate the 1995 GTAP tables that form our core input-output data using the standard assumption of normalized prices; as a consequence we distort the distribution of energy consumptions across agents (cf. Section Data Harmonization Requirements Building Hybrid Energy/economy Accounting Tables above). Ultimately, by calibrating reduced forms on distorted data we misuse the information extracted from POLES runs.

The Continued Fable of the Elephant and Rabbit

In 2006 we publish the central methodological piece of our hybrid modeling endeavor (Ghersi and Hourcade, 2006), which details the method surveyed in Section Calibration of Reduced Forms of Bottom-up Behaviors above and illustrates its significance at a global scale in 2030, with the purpose of contesting that the "elephant and rabbit stew" of Hogan and Manne (1977) should necessarily taste so much like the former, massively larger animal. The linked bottom-up model is again the POLES model and the runs on which we calibrate global reduced forms are those runs which already fed our 2003 publication (cf. Section Reconciling the Equity and Efficiency of International Burden-sharing Climate above).

Ironically, the response surfaces extracted from POLES for the composite sector and households' energy consumptions turn out quite compatible with the standard CES functions to which we compare our envelope approach: the late 1990s version of POLES that delivered these responses is still only bottom-up from an energy supply point of view—indeed we demonstrate the inability of CES functions to accommodate its responses concerning the energy and capital intensities of energy production—while energy demands largely remain governed by econometric specifications well replicated by CES functions. For this reason we also calibrate our reduced forms on an ALTER dataset, which is in fact a modified POLES dataset with increased curvatures and explicit absolute floors to the 3 energy intensities we work on. We refer to these floors as "technical asymptotes."

For the sake of argument, our year-2030 policy runs explore a wide range of carbon taxes in the simplified setting of perfect markets, including that of labor, and under the assumption of lump-sum recycling of the carbon tax proceeds to households. Besides the original IMACLIM-S model resorting to reduced forms according to Section Calibration of Reduced Forms of Bottom-up Behaviors, we implement a standard static CES model "CES Kfix" considering an exogenous stock of capital (following, e.g., Böhringer, 1998) and a "CES Kvar" variant of this model, which (1) sorts out the various components of value-added net of labor costs and values its depreciation part as the composite good; (2) treats depreciation as an index of capital demand, which must be balanced by investment.

Even in the controlled environment of perfect markets and lump-sum recycling, IMACLIM-S modeling results for the 2030 global economy turn out to diverge from those of the CES Kfix or CES Kvar approaches. What is more, the sign and magnitude of divergences varies whether GDP, households' non-energy consumption or the marginal cost of decarbonization are considered, as well as whether the POLES or ALTER calibration

[41] Löschel (2002) stresses how most CGE modelers set key elasticity of substitution parameters judgmentally rather than based on genuine economic analysis. The later rare econometric efforts of Van der Werf (2008) and Okagawa and Ban (2008) offer limited regional coverage that falls short of concerning the 14 global regions of our hybrid modeling.

TABLE 2 | Response to carbon pricing of 8 economies in 2030 according to the POLES model.

2030 policy response w.r.t. baseline	Africa (%)	Brazil (%)	China (%)	EU15 (%)	FSU (%)	India (%)	Japan (%)	USA (%)
Firms' E consumption, $100/tC	−11	−9	−21	−12	−18	−17	−4	−21
Firms' E consumption, $600/tC	−42	−31	−53	−35	−41	−49	−16	−47
Firm's carbon emissions, $100/tC	−15	−12	−26	−16	−21	−22	−8	−30
Firms' carbon emissions, $600/tC	−53	−40	−64	−45	−47	−61	−28	−62
Households' E consumption, $100/tC	−7	−3	−15	−4	−6	−11	−2	−9
Households' E consumption, $600/tC	−25	−11	−40	−17	−20	−32	−11	−28
Households' carbon emissions, $100/tC	−11	−7	−20	−4	−8	−13	−5	−8
Households' carbon emissions, $600/tC	−39	−25	−57	−23	−28	−44	−20	−34

This table does not appear in Ghersi et al. (2003) but was extracted from archived data. For the sake of concision it only reports on 8 of the 14 modeled economies. Carbon pricing is expressed in constant (2005) US dollar per ton of carbon ($/tC). "E consumption" stands for energy consumption in volume.

data is used. For limited ranges of decarbonization and for one or the other indicator the Kvar or Kfix approaches can deliver estimates reasonably close to the IMACLIM-S computation; however they never simultaneously match on the 3 retained indicators and often diverge by up to 100%, thereby justifying the effort invested in a proper modeling of induced production frontiers.

An Outlook on the Macroeconomic Impacts of Electric Vehicles Penetration in EU28

In the framework of a recent EV-STEP research program we developed the coupling of IMACLIM-S with the TIMES PanEU model of the University of Stuttgart, to explore the macroeconomic consequences of accelerated penetration of electric vehicles in the EU28 economy. We worked on a geographically aggregated EU28 but disaggregated 12 productions: 5 energy vectors, 2 types of vehicles (conventional vehicles and electric cars), 1 electric equipment good, which encompasses the batteries of electric vehicles; 3 transportation services (air, water, and land) and 1 composite good, which aggregates all remaining economic products and services[42].

We implemented the iteration-to-convergence coupling method of Section Coupling Through Iterative Convergence of Linking Variables Trajectories above with a high level of detail, eventually forcing in IMACLIM-S 39 different trajectories either directly taken from TIMES or inferred from its results. These trajectories cover international energy prices, energy exports, primary energy outputs, energy intensities and households' energy demands, but also the capital intensity of the electricity production, the transport services intensity of the composite good (based on ton-kilometer data from TIMES), the transport services demand of households (based on passenger-kilometer data from TIMES), the "electric vehicle intensity" of GDP (uniformly applied to all production sectors) or the actual vehicle demand of households, and indeed the assumed trajectory of battery costs, which we apply to the electric equipment intensity of the electric car production.

Despite the care taken to ground IMACLIM-S on hybrid economy/energy accounts following Section Building from IEA Energy Balances and Energy Price Data above, the data imports from TIMES into IMACLIM-S systematically take the form of trends developing from 2007 on, to be applied to IMACLIM-S variables, rather than of absolute values. The main reason for this is that TIMES PanEU is calibrated on 2005 data, and thus does not match the statistics of further years, including the 2007 data on which IMACLIM-S is calibrated. Another reason is that some non-energy variables, as vehicle sales or transport activities, are not described in IMACLIM-S with the same level of detail as they are in TIMES. A last reason is that we lacked time, in the limited course of a first collaborative, to scan the extensive content of TIMES PanEU for more focused information that we might have used in IMACLIM-S in a more direct way.

We implemented IMACLIM-S for 4 scenarios developed by TIMES PanEU (**Table 3**) as well as for an optimistic variant on electric cars (EC) exports for the 3 scenarios where EC sales pick up: REF, EU- and EU. Because iterating to convergence was out of the scope of this first research collaborative with the University of Stuttgart, we had to content ourselves with calibrating IMACLIM-S growth under the REF- scenario (the closest to a baseline) on the GDP trajectory common to all 4 TIMES PanEU scenarios. To do so we adjusted the labor productivity gains and export trends of IMACLIM-S to have it replicate the TIMES PanEU GDP trajectory under constraint of its REF- forecast of energy systems. The subsequent runs of the REF, EU- EU scenarios and their EC-exporting variants were developed with unchanged growth drivers, but also without convergence. As a consequence the results reported below are only preliminary results, which at least overestimate those scenario impacts that are imported as variations of total volumes[43].

[42]In Ghersi (2014) we thoroughly describe the construction of a hybrid energy-economy EU28 matrix and lay out the coupling method, although not in its ultimate refinements. In Ghersi (2015) we publish scenario results together with appendixes describing complementary data mining and treatment, ultimate amendments to the model and the complete detail of the data imports from the TIMES PanEU model.

[43]I.e., variations of households' energy consumptions and of energy exports. Impacts on firms' consumptions are imported as intensities, which minimizes divergence between the first iteration and the ultimate converged equilibrium. Note that the volume output of primary energies is probably exogenous to TIMES scenarios, which also reduces divergence between the first and the converged iteration.

TABLE 3 | Overview of 4 scenarios explored by TIMES PanEU.

Scenario	REF-	REF	EU-	EU
GHG abatement target	EU-ETS emissions: −21% in 2020 compared to 2005 then −1.74% *per annum*		Total emissions −20% in 2020, −80% in 2050 compared to 2005	
Renewables	60% of electricity consumption by 2050			
Electric mobility	No target	National targets	No target	National targets
Biofuels	No target	National targets	No target	National targets

Source: Markus Blesl, IER, University of Stuttgart.

Under our pessimistic assumptions on the competitiveness of the EU electric car (EC) industry[44], EC penetration negatively impacts growth for both the mitigation contexts we explore (**Table 3** above), but only slightly so (**Figure 5**, left panel): in the REF context, the GDP loss of the REF scenario over its REF- counterpart is highest in 2030 at 0.17%. In the EU context, the GDP loss is highest in 2025 at 0.25%. Detailed modeling results and sensitivity analysis reveal that the main cause of the GDP loss is thus not the forced higher penetration of durably more expensive electric cars, which is strongest in 2050 in the REF context and in 2045 in the EU context, but rather the impact of increased electric mobility on energy and especially electricity prices, which culminates at dates corresponding with peaking GDP losses: the REF scenario induces a maximum 1.4% increase in electricity prices over the REF- scenario in 2030, whereas the EU scenario induces a maximum 3.6% increase in electricity prices over the EU- scenario in 2025.

EC production in optimistic (hereafter v2) scenarios strongly exceed EC productions in their original (hereafter v1) counterparts[45]. The impact on GDP is of course not as straightforward as the multiplication of extra exported units by export price, because EC production competes for inputs with other sectors. One exception is however the model's imperfect labor market, where EC production finds unemployed resources to put to use. This is one important mechanism through which the v2 assumptions favorably impact growth. Another major mechanism is international trade: v2 assumptions guarantee substantial exports (and lesser imports) regardless of terms-of-trade competition[46]; they thus lift some pressure off the trade balance, which allows EU prices appreciating *vis-à-vis* international prices and thus diminishing the brunt of imports.

As a consequence, in 2050 the GDP of REF v2 ends up 0.2% above that of REF v1; that of EU- v2, 0.6% above EU-v1; and that of EU v2, 0.8% above EU v1 (cf. right vs. left panel of **Figure 5**). In the case of REF, the v2 variant indeed induces a GDP trajectory significantly above that of the REF- v1 scenario: under our modeling assumptions and in the context of moderate climate policy objectives, an optimistic development of international demand for EU-produced electric cars does more than compensate the increased electricity investment needs prompted by EC penetration. Besides, in the ambitious framework of a "factor 5" mitigation objective, favorable assumptions on the competitiveness of the emerging EU electric car industry turn the moderate costs of increased EC penetration (maximum loss of 0.25% GDP in 2025 between the EU- v1 and EU v1 scenarios) into moderate benefits (maximum benefit of 0.10% GDP in 2050 from the EU- v2 to the EU v2 scenario).

The main strength of this coupling exercise is how the aggregate EU28 macroeconomic framework of IMACLIM-S can lean on the 28-country detail of energy systems of TIMES PanEU. Particularly, the impact of electric vehicle penetration on electric systems is assessed at a level of detail that warrants proper modeling of actual regional supply constraints. It is hard to imagine that any CGE framework extended to BU electricity supply could perform in comparable manner.

However, one important limitation of the coupling exercise is the paradigm gap between the simulation, recursive framework of IMACLIM-S, where economic agents take "myopic" decisions only concerned with the state of the economy at the current year, and the optimization framework of TIMES PanEU, where some benevolent planner perfectly anticipates the model's multiple trajectories up to its temporal horizon and accordingly adapts all investment decisions to minimize the total discounted cost of the energy system[47]. Mixing the two approaches could be acceptable if TIMES PanEU modeled energy supply alone, for which centralized energy planning is still the norm in many countries—with the obvious limitation that no central planner has infallible knowledge of the future of energy prices, energy demands and energy technology developments. But TIMES PanEU extends to all end-uses and thus models the investment decision of every single household regarding, e.g., its vehicle or heating equipment. Our coupling experiment must therefore not be interpreted beyond what it really is: a somewhat patchwork outlook on possible energy/economy futures combining macroeconomic simulation and energy system optimization. It was nonetheless a rich experiment of extending to as many variables as possible the connection of IMACLIM-S to a BU model, which can serve in future coupling with simulation models as POLES.

Sovereign Risk and Energy: The RISKERGY Program

Our latest hybrid modeling endeavor is part of the RISKERGY research program, which aims at developing a method of sovereign risk rating that pays due attention to energy issues.

[44] One industry that does not export and that, at base year, imports half its domestic sales from abroad, whereas conventional vehicles are only for 8% of them imported goods.

[45] Optimistic export scenarios consider (1) vehicle exports climbing from 10 to 200% of vehicle sales to households between 2008 and 2020, then remaining at 200% of household sales up to 2050; (2) an aggregate vehicle import share at base year identical to that of conventional vehicles, i.e., of 8% rather than 50%.

[46] To give an order of magnitude: 10 million vehicles at 25 thousand Euros amounts to €250 billion exports; the REF v2, EU- v2, and EU v2 scenarios respectively reach 7, 25, and 27 million units exports. In the REF- v1 scenario, EU exports start from *ca*. €1,700 billion in 2007 and climb to €5,600 billion in 2050.

[47] Schäfer and Jacoby (2006), Martinsen (2011) (although not iterating to convergence) or Labriet et al. (2015) have similarly coupled optimisation models of the MARKAL family with recursive TD models.

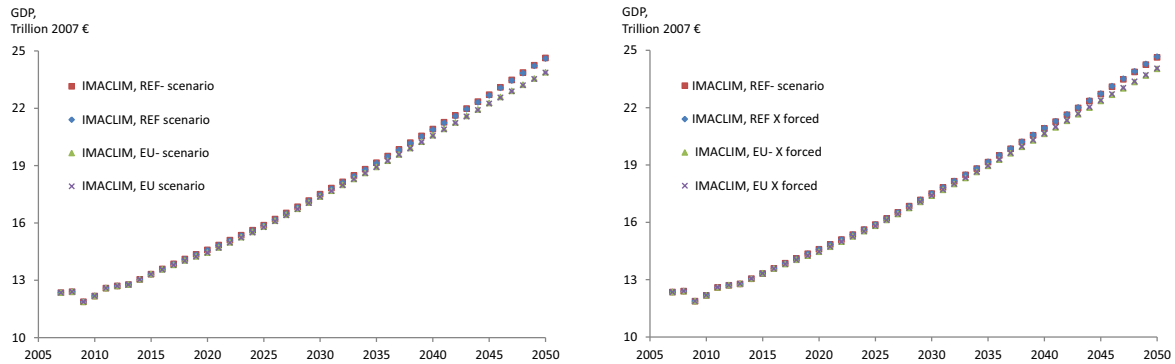


FIGURE 5 | EU28 growth trajectories under 4 scenarios and 3 forced exports variants. Computed from year-to-year real variations according to a chained Fischer GDP price index.

The bottom-up model to which we couple IMACLIM-S is again the POLES simulation model but in its most recent version. The requirements of the program prompt us to favor the dynamic framework of the iteration-to-convergence method of Section Coupling Through Iterative Convergence of Linking Variables Trajectories. The purpose is indeed to test the resilience of economies to and the impact on public debt accumulation of short-term shocks along the growth trajectory to 2020.

One key constraint on our research is that our modeling must apply indistinctly to the entire set of countries individualized in POLES, i.e., to close to 50 economies with widely different economic and energy characteristics. As a consequence we have so far worked on aggregate 2-sector economies, although we might disaggregate the energy good in the 4–5 usual vectors in future work. We have also been careful to systematically resort to international datasets covering a large number of countries (at the World Bank, the IMF, the IEA, etc.) to feed our modeling framework with parameters as recent statistical trajectories of GDP growth, labor supply, labor productivity, unemployment but also the trade balance and prevailing exchange rates.

We calibrate our multiple IMACLIM-S country models at the 2007 base year on hybrid matrixes resulting from the combination of the extensive GTAP database (cf. footnote 38 above) with POLES data, following the synthetic method presented Section Building from BU Model Variables at Base Year above. In this instance, shortcutting the alternate extensive method of Section Building from IEA Energy Balances and Energy Price Data appears particularly relevant, considering the daunting task of extending this method to several dozen countries, many of which suffering from poor statistical apparatus. As we already stressed this comes at the cost of some approximations, particularly in the split between households' and firms' transport fuel consumptions.

Working on short term sovereign debt accumulation requires unusual alignment on available statistics. The Solow growth engine we mobilize, despite the path-dependency imposed by importing energy system constraints from POLES, but also statistics on investment, capital utilization and exchange rates, would "miss" the massive 2008 global economic crisis if it were to run on long-term trends of labor productivity improvements only. For this reason, from our 2007 base year up to 2013 and for all country models we perform an original targeting of 3 core macroeconomic statistics (GDP, unemployment, and the trade balance) by adjusting a set of shocks on capital and labor productivities, the negotiation power of workers (which impacts endogenous equilibrium unemployment via a wage curve) and international trade. Beyond 2013 we phase out the crisis perturbations at a constant rate up to 2020 rather than abruptly canceling them, which would induce a somewhat artificial break of macroeconomic trends.

We have so far produced only preliminary runs of the coupled model on the 10 richest world economies, for a baseline as well as for a scenario of oil supply shock, which envisages a 15-million barrel drop of the daily exports of Persian Gulf countries in 2016. POLES and IMACLIM-S converge at 1% (IMACLIM-S computes a GDP impact of updated POLES data below 1%) after 3 iterations only, but for the quite wrong reason that we still have to extend the data imports from POLES to IMACLIM-S to elements other than international energy prices, energy consumptions and intensities—first and foremost to the capital intensity of energy supply, which our applied work on the EU identifies as one major channel of impact of energy systems on growth (cf. Section An Outlook on the Macroeconomic Impacts of Electric Vehicles Penetration in EU28 above).

Although at this early stage of development our coupled model behaves in interpretable manner for all 10 contrasted economies, in the baseline scenario as well as in the oil shock scenario. Compare, e.g., how France and Russia fare facing the 2016 oil shock supply (**Figure 6**), which translates into a fivefold increase of the international oil price and a consecutive three- to fourfold increase of international gas prices (depending on regional markets). France, highly dependent on oil imports, marks a 3.2% recession in 2016, from which it only partially recovers from 2017 on when the restriction on oil supply lifts and prices fall back to baseline levels or indeed slightly below. Conversely Russia, as major oil and gas exporter, strongly benefits from the price hike induced by the shock. The impact is masked in 2016 because of degraded conditions in Russia's trading

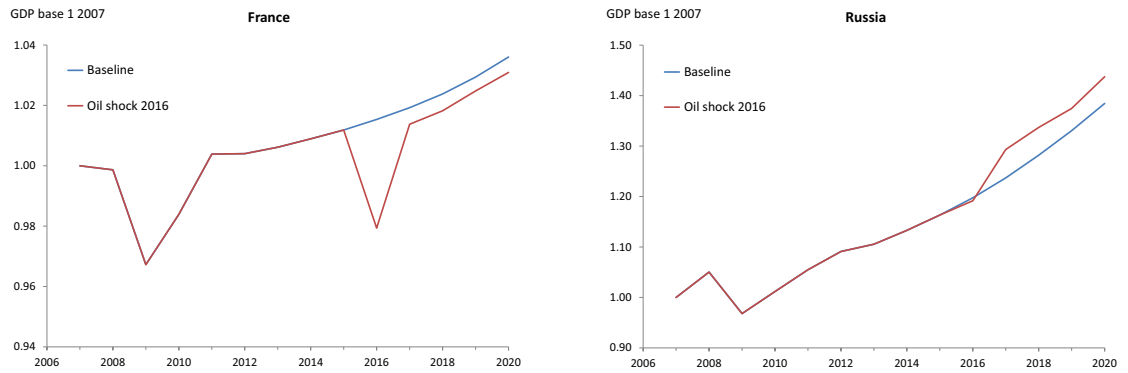


FIGURE 6 | GDP trajectories of France and Russia under baseline and oil shock scenario, IMACLIM-S and POLES coupled modeling.

partners, but is manifest in later years, as the country reaps the profits of its 2016 investment boost[48].

Similar to Section Reconciling the Equity and Efficiency of International Burden-sharing Climate Agreements but all the more striking in a geographically disaggregated setting, the main force of the coupling is how it allows extending the analysis to a large number of countries with properly characterized energy systems. Such coverage at such level of precision is unquestionably out of reach of standard (non-hybrid) approaches, whatever their macroeconomic paradigm, be it only for the simple reason that the available literature is quite far from providing the country-specific parameterization that should sustain it.

CONCLUSION AND FURTHER RESEARCH AGENDA

It is our hope that this survey demonstrates significant advances in our ability to propose hybrid modeling solutions to the energy and climate modeling challenge. Over the last 15 years we have outlined and have started implementing two operational methods of coupling extensive bottom-up modeling to our IMACLIM-S macroeconomic framework. We have also addressed data harmonization issues to guarantee the relevance of model dialogue, which has incidentally led us to model agent-specific prices, a significant improvement over other TD modeling endeavors. Our numerical applications are few but growing, and should shortly benefit from the contributions of other authors on the Brazilian and South African economies.

Some unsettled issues naturally require further research. Regarding data harmonization and the production of hybrid base year accounts, we still have to settle on a more satisfactory treatment of energy auto-consumptions. The issue depends on whether we are hybridizing in link with disaggregated energy balances or with the coupled BU model. The high level of disaggregation of IEA balances allows proper identification of the genuine energy inputs (for consumption or transformation) of all energy sectors—notwithstanding whether these end up aggregated or not. The decision is then easier made to strip down auto-consumptions to genuine marketed inputs, or to identify, by comparing energy with national accounting data, putative commodity trade, which can then be given specific modeling. Hybridizing base year data by importing energy expenses and volumes from the coupled BU model offers less choice because of a lasting simplification of BU models: these rarely distinguish between crude oil and petroleum products, which forbids pinpointing the crude oil expenses of the refining industry—one major auto-consumption of the aggregate energy sector of hybrid accounts built on BU data. However, the refining data are readily available from even aggregated IEA balance sheets and could be exploited in future work.

Our coupling methods also call for further developments. The calibration of reduced forms is conceptually stabilized, but only in the case of an aggregate and "closed" (global) economy. Questions remain as to how to extend the method to a multisectoral setting. Similar to the two-sector setting but with increased degrees of freedom, a prerequisite is to settle on some conceptual form for the instantaneous trade-offs between inputs or consumptions. The implication of such choice on modeling results remains to be assessed. Another particularly open field of investigation regards households' consumptions: in the wake of the research on aggregate macroeconomic models boiling down to some aggregate production function, E3 modelers have devoted much attention to the producer's side of the energy conservation problem and very little to the consumer's side of it, as testifies the poor treatment of households energy trade-offs in current models[49].

[48] Results are for the polar case of fixed exchange rates, which implies that the boost in trade balance causes households' consumption to shrink relative to investment and public consumptions (with these fixed shares of GDP and consumption its remainder). In further research we will envisage some flexible representation of terms-of-trade.

[49] E.g., the DART (Klepper et al., 2003), EPPA (Paltsev et al., 2005), PACE (Böhringer, 2002), or the PHOENIX model (Sue Wing et al., 2011) still resort to nested CES functions to represent consumer demand. The nesting commonly isolates one aggregate energy bundle from all other consumptions, thus leveling off both the substitution possibilities among energy goods on one side and those between the energy bundle and the bundle of all non-energy goods on the other side. Cai et al. (2015) circumvent some of these limitations by resorting to CRESH functions rather than CES functions in their GTEM-C model. Schäfer and Jacoby (2005) or Karplus et al. (2013), both implementing the EPPA model of the MIT, are rare examples of an elaborate treatment of household demand (concerning private car).

The iteration-to-convergence method is also quite advanced, although it has developed in a more pragmatic way. However, our provisory treatment of trade-offs other than those involving energy is too standard and we must at the very least move to a modeling of capital vintages with putty-clay characteristics, lest the benefits of importing explicit energy systems inertias be lost in a too lax treatment of substitution possibilities for non-energy factors. The feedback from IMACLIM-S to the coupled BU model can also surely be expanded to more than GDP or other activity indicators. As we hint at the end of Section Coupling Through Iterative Convergence of Linking Variables Trajectories, IMACLIM-S computes endogenous relative prices for all factors of production, which, if transmitted to the capital and operation & maintenance costs identified by BU models for all technologies, could turn out to alter the "merit order" of technologies.

REFERENCES

Bauer, N., Edenhofer, O., and Kypreos, S. (2008). Linking energy system and macroeconomic growth models. *Comput. Managem. Sci.* 5, 95–117. doi: 10.1007/s10287-007-0042-3

Babiker, M. H., Reilly, J. M., Mayer, M., Eckaus, R. S., Sue Wing, I., and Hyman, R. C. (2001). *The MIT Emissions Prediction and Policy Analysis (EPPA) Model: Emissions, Sensitivities and Comparison of Results*. Joint Program on the Science and Policy of Global Change report #71, Cambridge, MA: Massachusetts Institute of Technology. Available online at: http://web.mit.edu/globalchange/www/MITJPSPGC_Rpt71.pdf (accessed July 2015)

Berndt, E. R., and Wood, D. O. (1975). Technology, prices, and the derived demand for energy. *Rev. Econ. Stat.* 57, 259–268. doi: 10.2307/1923910

Bernstein, P. M., Montgomery, W. D., and Rutherford, T. F. (1999). Global impacts of the Kyoto agreement: results from the MS-MRT model. *Res. Energy Econ.* 21, 375–413. doi: 10.1016/S0928-7655(99)00009-3

Böhringer, C. (1998). The synthesis of bottom-up and top-down in energy policy modeling. *Energy Econ.* 20, 233–248. doi: 10.1016/S0140-9883(97)00015-7

Böhringer, C. (2002). Climate politics from Kyoto to Bonn: from little to nothing? *Energy J.* 23, 51–71. doi: 10.5547/ISSN0195-6574-EJ-Vol23-No2-2

Böhringer, C., and Rutherford, T. F. (2008). Combining bottom-up and top-down. *Energy Econ.* 30, 574–596. doi: 10.1016/j.eneco.2007.03.004

Cai, Y., Newth, D., Finnigan, J., and Gunasekera, D. (2015). A hybrid energy-economy model for global integrated assessment of climate change, carbon mitigation and energy transformation. *Appl. Energy* 148, 381–395. doi: 10.1016/j.apenergy.2015.03.106

Cohen, A. J., and Harcourt, G. C. (2003). Whatever happened to the cambridge capital theory controversies? *J. Econ. Perspect.* 17, 199–214. doi: 10.1257/089533003321165010

Dai, H., Mischke, P., Xie, X., Xie, Y., and Masui, T. (2015). Closing the gap? Top-down versus bottom-up projections of China's regional energy use and CO_2 emissions. *Appl. Energy*. doi: 10.1016/j.apenergy.2015.06.069. Available online at: http://www.sciencedirect.com/science/article/pii/S0306261915008272. [Epub ahead of print].

Fisher-Vanden, K., Edmonds, J. A., Pitcher, H. M., Barns, D., Baron, R., Kim, S., et al. (1993). *The Second Generation Model of Energy Use, the Economy, and Greenhouse Gas Emissions*. Washington, DC: Pacific Northwest Laboratory. 17. Available online at: http://sedac.ciesin.org/mva/FV1993/FV1993.html (accessed July 2015).

Fortes, P., Pereira, R., Pereira, A., and Seixas, J. (2014). Integrated technological-economic modeling platform for energy and climate policy analysis. *Energy* 73, 716–730. doi: 10.1016/j.energy.2014.06.075

Fujimori, S., Masui, T., and Matsuoka, Y. (2014). Development of a global computable general equilibrium model coupled with detailed energy end-use technology. *Appl. Energy* 128, 296–306. doi: 10.1016/j.apenergy.2014.04.074

Ghersi, F. (2014). *Macroeconomic Modeling of Electrified Mobility Systems in 2030 European Union*. CIRED Working Paper 55-2014. Paris: CIRED. 32. Available online at: http://www.centre-cired.fr/IMG/pdf/CIREDWP-201455.pdf

Ghersi, F. (2015). *Macroeconomic Modeling of Electrified Mobility Systems in 2030 European Union. Scenario Results*. CIRED Working Paper 58-2015. Paris: CIRED. 32. Available online at: http://www.centre-cired.fr/IMG/pdf/ciredwp-201458.pdf

Ghersi, F., and Hourcade, J.-C. (2006). Macroeconomic consistency issues in E3 modeling: the continued fable of the elephant and the rabbit. *Energy J.* S2, 39–62. doi: 10.5547/issn0195-6574-ej-volsi2006-nosi2-3

Ghersi, F., Hourcade, J.-C., and Criqui, P. (2003). Viable responses to the equity-responsibility dilemma: a consequentialist view. *Clim. Policy* 3S1, S115–S133. doi: 10.1016/j.clipol.2003.10.011

Grubb, M., Edmonds, J., ten Brink, P., and Morrison, M. (1993). The costs of limiting fossil-fuel CO_2 emissions: a survey and analysis. *Ann. Rev. Energy Environ.* 18, 397–478. doi: 10.1146/annurev.eg.18.110193.002145

Hicks, J. R. (1932). *The Theory of Wages*. London: MacMillan.

Hoffman, K. C., and Jorgenson, D. (1976). Economic and technological models for evaluation of energy policy. *Bell J. Econ. Managem. Sci.* 8, 444–466. doi: 10.2307/3003296

Hogan, W. W., and Manne, A. S. (1977). "Energy-economy interactions: the fable of the elephant and the rabbit?" in *Modeling Energy-Economy Interactions: Five Approaches*, ed C. J. Hitch (Washington, DC: Resources for the Future), 247–277.

Hourcade, J.-C., Jaccard, M. K., Bataille, C., and Ghersi, F. (2006). Hybrid modeling: new answers to old challenges. *Energy J.* S2, 1–12. doi: 10.5547/ISSN0195-6574-EJ-VolSI2006-NoSI2-1

Hourcade, J.-C., and Shukla, P. (2001). "Global, regional and national costs and ancillary benefits of mitigation," in *Climate Change 2001: Mitigation,* eds B. Metz, O. Davidson, R. Swart, and J. Pan (Cambridge: Cambridge University Press), 499–559.

Hudson, E. A., and Jorgenson, D. W. (1974). U. S. energy policy and economic growth, 1975-2000. *Bell J. Econ. Managem. Sci.* 5, 461–514. doi: 10.2307/3003118

Igos, E., Rugani, B., Rege, S., Benetto, E., Drouet, L., and Zachary, D. S. (2015). Combination of equilibrium models and hybrid life cycle-input–output analysis to predict the environmental impacts of energy policy scenarios. *Appl. Energy* 145, 234–245. doi: 10.1016/j.apenergy.2015.02.007

INSEE. (2012). *Méthodologie des Comptes Trimestriels. INSEE Méthodes 126*. Paris: Institut National de la Statistique et des Études Économiques. 143.

Jaccard, M. K. (2009). "Combining top down and bottom up in energy economy models," in *International Handbook on the Economics of Energy*, eds J. Evans and L. C. Hunt (Cheltenham: Edward Elgar), 311–331.

Jaffe, A. B., and Stavins, R. N. (1994). The energy-efficiency gap. What does it mean? *Energy Pol.* 22, 804–810. doi: 10.1016/0301-4215(94)90138-4

Karplus, V. J., Paltsev, S., Babiker, M., and Reilly, J. M. (2013). Applying engineering and fleet detail to represent passenger vehicle transport in a computable general equilibrium model. *Econ. Model.* 30, 295–305. doi: 10.1016/j.econmod.2012.08.019

Kennedy, C. (1964). Induced bias in innovation and the theory of distribution. *Econ. J.* 74, 541–547. doi: 10.2307/2228295

Klepper, G., Peterson, S., and Springer, K. (2003). DART97: *A Description of the Multi-regional, Multi-sectoral Trade Model for the Analysis of Climate Policies*. Kiel Working Papers 1149, Kiel Institut für Weltwirtschaft, Kiel. 40. Available online at: http://www.ifw-members.ifw-kiel.de/publications/dart97-a-description-of-the-multi-regional-multi-sectoral-trade-model-for-the-analysis-of-climate-policies [accessed December 2014].

Labriet, M., Drouet, L., Vielle, M., Loulou, R., Kanudia, A., and Haurie, A. (2015). *Assessment of the Effectiveness of Global Climate Policies Using Coupled Bottom-up and Top-down Models*. Nota di Lavoro 23.2015. Milan: Fondazione ENI Enrico Mattei. 22.

Laitner, J. A. S., and Hanson, D. A. (2006). Modeling detailed energy-efficiency technologies and technology policies within a CGE framework. *Energy J.* S2, 151–169. doi: 10.5547/issn0195-6574-ej-volsi2006-nosi2-8

Löschel, A. (2002). Technological change in economic models of environmental policy: a survey. *Ecol. Econ.* 43, 105–126. doi: 10.1016/S0921-8009(02)00209-4

Magat, W. A. (1979). Technological advance with depletion of innovation possibilities - implications for the dynamics of factor shares. *Econ. J.* 89, 614–623. doi: 10.2307/2231871

Manne, A. S. (1977). *ETA-MACRO: A Model of Energy-Economy Interactions*. NASA STI/Recon Technical Report, 78, 26612.

Martinsen, T. (2011). Introducing technology learning for energy technologies in a national CGE model through soft links to global and national energy models. *Energy Pol.* 39, 3327–3336. doi: 10.1016/j.enpol.2011.03.025

McFarland, J. R., Reilly, J. M., and Herzog, H. J. (2004). Representing energy technologies in top-down economic models using bottom-up information. *Energy Econ.* 26, 685–707. doi: 10.1016/j.eneco.2004.04.026

McKibbin, W., and Wilcoxen, J. (1995). The theoretical and empirical structure of the G-Cubed model. Brookings discussion paper in International Economics, 118.

Messner, S., and Schrattenholzer, L. (2000). MESSAGE–MACRO: linking an energy supply model with a macroeconomic module and solving it iteratively. *Energy* 25, 267–282. doi: 10.1016/S0360-5442(99)00063-8

Okagawa, A., and Ban, K. (2008). Estimation of substitution elasticities for CGE models. Discussion Papers in Economics and Business 08-16, Graduate School of Economics and Osaka School of International Public Policy (OSIPP), Osaka: Osaka University. Available online at: http://www2.econ.osaka-u.ac.jp/library/global/dp/0816.pdf (accessed July 2015)

Paltsev, S., Reilly, J. M., Jacoby, H. D., Eckaus, R. S., McFarland, J., Sarofim, M., et al. (2005). The MIT Emissions Prediction and Policy Analysis (EPPA) model: Version 4. MIT Joint Program on the Science and Policy of Global Change, Report 125, Cambridge, USA. 78. Available online at: http://globalchange.mit.edu/files/document/MITJPSPGC_Rpt125.pdf [accessed December 2014].

Ruttan, V. (2002). "Sources of technical change: Induced innovation, evolutionary theory, and path dependence," in *Technological Change and the Environment*, eds A. Grübler, N. Nakicenovic, and W. D. Nordhaus (Washington, DC: Resources for the Future), 9–39.

Sassi, O., Crassous, R., Hourcade, J.-C., Gitz, V., Waisman, H., and Guivarch, C. (2010). IMACLIM-R: a modeling framework to simulate sustainable development pathways. *Inter. J. Glob. Environ. Iss.*10, 5–24. doi: 10.1504/IJGENVI.2010.030566

Schäfer, A., and Jacoby, H. D. (2005). Technology detail in a multisector CGE model: transport under climate policy. *Energy Econ.* 27, 1–24. doi: 10.1016/j.eneco.2004.10.005

Schäfer, A., and Jacoby, H. D. (2006). Experiments with a hybrid CGE-MARKAL model. *Energy J.* S2, 171–177. doi: 10.5547/issn0195-6574-ej-volsi2006-nosi2-9

Schumacher, K., and Sands, R. D. (2007). Where are the industrial technologies in energy–economy models? An innovative CGE approach for steel production in Germany. *Energy Econ.* 29, 799–825. doi: 10.1016/j.eneco.2006.12.007

Solow, R. M. (1957). Technical change and the aggregate production function. *Rev. Econ. Stat.* 39, 312–320. doi: 10.2307/1926047

Solow, R. M. (1959). "Investment and technical progress," in *Mathematical Methods in the Social Sciences*, eds K. J. Arrow, S. Karlin, and P. Suppes (Stanford, CA: Stanford University Press), 89–104.

Sue Wing, I. (2006). The synthesis of bottom-up and top-down approaches to climate policy modeling: electric power technologies and the cost of limiting US CO2 emissions. *Energy Pol.* 34, 3847–3869. doi: 10.1016/j.enpol.2005.08.027

Sue Wing, I., Daenzer, K., Fisher-Vanden, K., and Calvin, K. (2011). *Phoenix Model Documentation*. Maryland: Joint Global Change Research Institute, University of Maryland. 84. Available online at: Http://www.globalchange.umd.edu/data/models/phx_documentation_august_2011.pdf [accessed December 2014].

Sutherland, R. (1991). Market barriers to energy efficiency investments. *Energy J.* 12, 15–34. doi: 10.5547/ISSN0195-6574-EJ-Vol12-No3-3

Van der Werf, E. (2008). Production functions for climate policy modeling: an empirical analysis. *Energy Econ.* 30, 2964–2979. doi: 10.1016/j.eneco.2008.05.008

Conflict of Interest Statement: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Environmental Science

METHODS
published: 19 June 2015
doi: 10.3389/fenvs.2015.00043

Quantifying tracer dynamics in moving fluids: a combined Eulerian-Lagrangian approach

Fanny Chenillat[1], Bruno Blanke[2], Nicolas Grima[2], Peter J. S. Franks[1], Xavier Capet[3] and Pascal Rivière[4]*

[1] Integrative Oceanography Division, Scripps Institution of Oceanography, University of California, San Diego, La Jolla, CA, USA, [2] Laboratoire de Physique des Océans, UMR 6523 Centre National de la Recherche Scientifique-Ifremer-IRD-UBO, Brest, France, [3] Laboratoire d'Océanographie et du Climat, / IPSL, UMR 7159 Centre National de la Recherche Scientifique-IRD-MNHN-UPMC, Paris, France, [4] Laboratoire des Sciences de l'Environnement Marin, UMR 6539 Centre National de la Recherche Scientifique-Ifremer-IRD-UBO, Plouzané, France

OPEN ACCESS

Edited by:
Christian E. Vincenot, Kyoto University, Japan

Reviewed by:
Ana María Durán-Quesada, University of Costa Rica, Costa Rica
Jinbo Wang, Scripps Institution of Oceanography, USA

***Correspondence:**
Fanny Chenillat, Integrative Oceanography Division, Scripps Institution of Oceanography, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093, USA
fchenillat@ucsd.edu

Specialty section:
This article was submitted to Environmental Informatics, a section of the journal Frontiers in Environmental Science

Received: *22 April 2015*
Accepted: *29 May 2015*
Published: *19 June 2015*

Citation:
Chenillat F, Blanke B, Grima N, Franks PJS, Capet X and Rivière P (2015) Quantifying tracer dynamics in moving fluids: a combined Eulerian-Lagrangian approach. Front. Environ. Sci. 3:43. doi: 10.3389/fenvs.2015.00043

Eulerian models coupling physics and biology provide a powerful tool for the study of marine systems, complementing and synthesizing *in situ* observations and *in vitro* experiments. With the monotonic improvements in computing resources, models can now resolve increasingly complex biophysical interactions. Quantifying complex mechanisms of interaction produces massive amounts of numerical data that often require specialized tools for analysis. Here we present an Eulerian-Lagrangian approach to analyzing tracer dynamics in moving fluids. As an example of its utility, we apply this tool to quantifying plankton dynamics in oceanic mesoscale coherent structures. In contrast to Eulerian frameworks, Lagrangian approaches are particularly useful for revealing physical pathways, and the dynamics and distributions of tracers along these trajectories. Using a well-referenced Lagrangian tool, we develop a method to assess the variability of biogeochemical properties (computed using an Eulerian model) along particle trajectories. We discuss the limitations of this new method, given the different biogeochemical and physical timescales at work in the Eulerian framework. We also use Lagrangian trajectories to track coherent structures such as eddies, and we analyze the dynamics of the local ecosystem using two techniques: (i) estimating biogeochemical properties along trajectories, and (ii) averaging biogeochemical properties over dynamic regions (e.g., the eddy core) defined by ensembles of similar trajectories. This hybrid approach, combining Eulerian and Lagrangian model analyses, enhances the quantification and understanding of the complex planktonic ecosystem responses to environmental forcings; it can be easily applied to the dynamics of any tracer in a moving fluid.

Keywords: biophysical interactions, plankton dynamics, ecosystem functioning, Lagrangian trajectories, ocean dynamics, mesoscale processes, eddy tracking, coherent structures

Introduction

The mechanisms of transport in fluids are important to various fields such as physics and chemistry (e.g., Ban and Gilbert, 1975; Bird et al., 2007 and references within), as well as medicine and biology (e.g., Lih, 1975). Though transport occurs at various scales (Bird et al., 2007 and references within), the underlying processes are similar among different systems. Transport is often affected by turbulence, from blood flow to atmospheric and oceanic currents (Dewan, 2011). Turbulent transport remains challenging to study because of its non-linearity and episodicity. Here we present a combined Eulerian-Lagrangian approach that allows us to follow and quantify tracer dynamics in a turbulent fluid. As an example of its utility, we apply this method to the study of biological-physical interactions in a marine planktonic system.

Biological and physical processes interact strongly in the ocean, leading to an inhomogeneous distribution of marine life. Plankton, the lower trophic levels of this ecosystem, are strongly influenced by their environment. Planktonic ecosystems are composed of small organisms, phytoplankton and zooplankton, with a typical body size ranging from microns to millimeters, and, by definition, drift with the ocean currents. Phytoplankton comprise the first trophic level of the marine food web, using light and nutrients to grow. The second trophic level is composed of zooplankton feeding on phytoplankton. Thus, planktonic ecosystems are strongly structured by light and nutrient availability, but also by temperature, mixing and other physical processes.

Biological-physical (biophysical) interactions occur over a wide range of timescales (Stommel, 1963; Haury et al., 1978). At small scales (millimeters to few meters), biophysical interactions are rapid (minutes to hours) and involve phytoplanktonic physiological processes and predation. For example, small-scale turbulence affects nutrient uptake by phytoplankton cells, which in turn influences the size structure and species composition of the phytoplankton community (e.g., Kiørboe, 1993; Estrada and Berdalet, 1997). At the global scale (several thousands of kilometers), biophysical interactions determine biogeographic provinces, defined by climatological conditions (Longhurst, 1995). The California Upwelling Coastal Province represents one of these biogeographic regions. In the California Upwelling System, seasonal equatorward winds blow along the coast, leading to upwelling of cold, salty, nutrient-rich water in the nearshore region. This nutrient-rich water supports intense biological activity of planktonic ecosystems and enhanced fish production. Between the small scale and the global scale, the mesoscale (10 km to ∼200 km) is characterized by two-dimensional vortex features known as eddies. These eddies last from a week to several months, and are often associated with enhanced biological activity compared to the surrounding oligotrophic waters (i.e., low nutrient concentration and limited biological activity) (e.g., Haury et al., 1978; The Ring Group, 1981).

Though biophysical interactions in the sea have been extensively investigated through *in situ* observations, *in vitro* experiments, and *in silico* analyses, most mechanisms remain unclear. In particular, the mechanisms associated with planktonic ecosystem variability in eddies and upwelling systems are still in debate. Oceanic Global Circulation Models (OGCMs) provide a key tool for studying such complex mechanisms. Given the monotonic improvements in computing resources, OGCMs are able to resolve increasingly complex biophysical interactions. These models produce massive amounts of numerical data that often require specialized tools for the identification, extraction, and quantification of features and dynamics. Here, we discuss how the combination of Lagrangian and Eulerian methods can help us understand the mechanisms driving planktonic dynamics in complex physical systems. In Section Modeling Biological-physical Interactions, we briefly describe these methods, and in Section Description of the Combined Eulerian-Lagrangian Approach, we explain why their combined use presents a useful and powerful approach. We illustrate this combined approach through case studies of eastern boundary upwelling systems (EBUS) in the last Section, Combined Eulerian-Lagrangian Analyses in EBUS.

Modeling Biological-physical Interactions

Biogeochemical models are able to describe the ecosystem with simple sets of biological dynamics translated into mathematical functions. Because plankton can be considered as a continuous property (present in large quantities in most oceanic systems), they can be mathematically modeled with box model structures. Such modeling consists in computing the transfers of nutrients (the biological fluxes) between the organic and inorganic components of the ecosystem (the biomass). These models can be very simple, e.g., with only Nutrient-Phytoplankton-Zooplankton (NPZ, e.g., Franks, 2002), or include complex biogeochemical cycling models with several *N*-, *P*-, or *Z*-compartments, additional detritus compartments (particulate or dissolved matter), and a variety of interconnections.

To study spatial patterns of plankton and assess complex biophysical interactions, these ecosystem models have been coupled with three-dimensional ocean circulation models. OGCMs represent a particularly powerful tool for studying the complex, non-linear biophysical interactions of planktonic ecosystems. In an Eulerian framework, the evolution of a tracer concentration can be described at each fixed point in space by the advection-diffusion equation:

$$\frac{\partial C}{\partial \mathrm{t}} + \vec{u} \cdot \nabla C = \mathit{Diffusion} + \mathit{Sources} - \mathit{Sinks} \qquad (1)$$

where C is the tracer concentration, the $\partial/\partial \mathrm{t}$ operator is the local rate of change of the tracer, $\vec{u}$ is the mean velocity field, and ∇C is the partial spatial derivative of C. The *Sources* and *Sinks* terms represent the biological gains and losses driven by biological fluxes (e.g., nutrient uptake for growth, predation by zooplankton grazing, or loss by natural mortality, respiration, excretion, egestion, etc.). Eulerian model analyses usually focus on biological budgets within selected geographical boxes (e.g., Chai et al., 2002; Gruber et al., 2011; Chenillat et al., 2013). Within these stationary boxes, the biological concentration C and its

spatial gradients move with the flow. These variations of C due to advection can make the direct understanding of biophysical processes difficult.

An Eulerian framework can limit the possibilities for analysis: the ocean has contrasting dynamical regimes (LaCasce, 2008) including eddies and fronts. These small-scale features rapidly impact the environment. In an Eulerian system, computations are made at fixed points, on a two-dimensional or three-dimensional grid (e.g., **Figure 1A**). At a fixed point, changes in the physical and biological properties can be driven by the passage of these contrasting dynamical regimes. From a biological point of view, however, one is usually interested in the dynamics within such moving features. Thus, using only an Eulerian framework limits our ability to understand planktonic dynamics in relation to their environment.

On the other hand, a Lagrangian approach consists of following a water parcel or a living individual as it moves with the flow. In this 'particle-tracking' method, the biological changes are analyzed along discrete trajectories of passive particles that move with the flow. Compared with the biological concentration equation in an Eulerian framework (Equation 1), in the Lagrangian frame the advection term disappears and the biomass concentration evolution equation along a moving fluid parcel becomes:

$$\frac{dC}{dt} = Diffusion + Sources - Sinks \qquad (2)$$

For some biological applications, the particle is considered as a biotic component or a living individual, and can move relative to the water (e.g., the krill). An Individual Based Model (IBM) will accommodate these biological traits. Each individual advected by the current is constrained by its own physiological or behavioral traits (see Batchelder et al., 2002 for complete details) added to Equation (2). In the simplest case, a swimming behavior can be added to account for active migration of zooplankton (Carr et al., 2008) or fish eggs and larvae (Lett et al., 2008; Blanke et al., 2012). This active migration can be strong and reach up to several thousands of meters in the water column for some zooplanktonic taxa like Euphausiids (Brinton, 1967). Including behavior in Equation (2) helps to diagnose how individual and environmental processes combine to influence the transport of active organisms. On the other hand, understanding how the environment regulates nutrient input variability and controls phytoplankton growth does not typically require the addition of swimming behavior, and Equation (2) may be used just as is. In this case the particle is neutrally buoyant, representing an infinitesimal volume of water with movements controlled by the ambient currents.

A fully Lagrangian approach (i.e., solving the original dynamical and biogeochemical equations following Lagrangian parcels) is rarely practical: the strain and vorticity of the flow rapidly produce complex spatial distributions of the particles that form the calculation nodes. Understanding such output usually requires remapping the Lagrangian particle locations and properties to a regular grid. However, this approach has numerous drawbacks (e.g., Bailey et al., 2010).

A more fruitful approach is to use the Eulerian advective flows to create Lagrangian trajectories. Such approaches have been used in biogeochemical and ecological studies to characterize planktonic niches (Lehahn et al., 2007; D'Ovidio et al., 2010), to study how zooplankton transport is affected by mixing (Qiu et al., 2010), and to follow planktonic changes along trajectories (Abbott et al., 1990). Significantly, no such studies have aimed to identify how dynamical regional features, like mesoscale eddies or submesoscale filaments, drive the local ecosystem dynamics.

Lagrangian methods have been extensively used not only in numerical studies but also directly at sea. For example, *in situ* drifters were used to describe lateral advection and eddy dispersion (Davis, 1983, 1991), and the main circulation of the ocean (e.g., Owen, 1991; Richardson, 1993) or to study air-sea CO_2 fluxes (Merlivat et al., 2014), and biological processes along their path (Landry et al., 2009).

Numerical Lagrangian approaches have some distinct advantages over field deployments. First, numerical experiments can provide stronger statistical power due to the large numbers

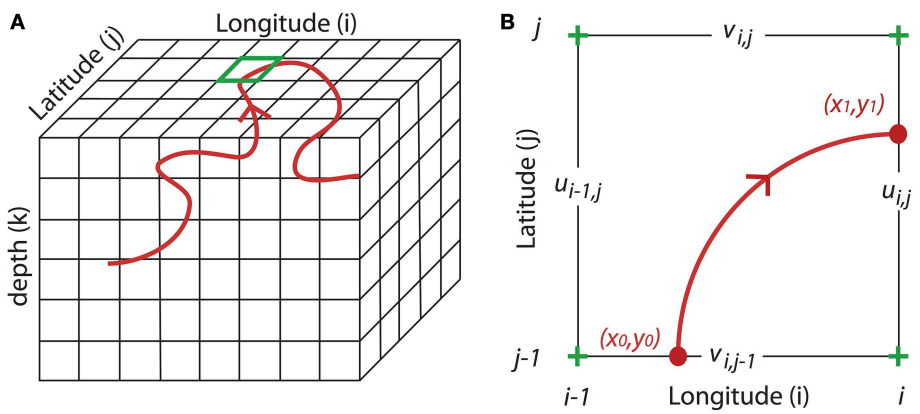


FIGURE 1 | Principles of the combined Eulerian-Lagrangian approach. (A) Schematic of a random Lagrangian trajectory (red line) driven by the Eulerian current (u,v,w) in a three-dimensional grid model. **(B)** Details of the green box in **(A)**, showing a Lagrangian computation in the horizontal plane ($w = 0$) (Döös et al., 2013).

of drifters that can be deployed. Second, particles can be tracked backward in time to reconstruct the history of a water parcel. And third, in models the drifters can track vertical motions, moving in three dimensions, unlike most *in situ* drifters that follow only the horizontal components of the current. Present-day computational resources make it possible to obtain sufficient spatial and temporal resolution to resolve meso- and submesoscale features using Lagrangian methods. It is this aspect that we highlight with our combined Eulerian-Lagrangian approach. In particular, we apply Lagrangian analyses to the study of coherent eddies—an ideal testbed (Beron-Vera et al., 2008; D'Ovidio et al., 2013), as Lagrangian particles can quantify the coherent nature of the particular flow features they are trapped in. This combined Eulerian-Lagrangian approach allows tracking of the evolution of all the biological properties associated with a moving water parcel, in both coherent and transient structures. The combined approach is introduced in the following section (Description of the Combined Eulerian-Lagrangian Approach). The method has been applied to identify the biological transformations following a moving water parcel. More precisely, analyses have been performed in EBUS to study ecosystem dynamics along mean pathways, along individual trajectories, and in complex dynamical features such as mesoscale eddies (in Section Combined Eulerian-Lagrangian Analyses in EBUS).

Description of the Combined Eulerian-Lagrangian Approach

Principles of the Combined Method in Physical Models

The Lagrangian trajectories are computed from time- and space-dependent Eulerian velocity fields, which are simulated by an OGCM such as the Regional Ocean Model System (ROMS) (Shchepetkin and McWilliams, 2005), Massachusetts Institute of Technology general circulation model (MITgcm) (Hill and Marshall, 1995) or Nucleus for European Modeling of the Ocean (NEMO) (Madec, 2008), among others. These numerical models solve the primitive equations, i.e., the Navier-Stokes equations, under the assumptions of Boussinesq and hydrostaticity, in an Eulerian grid framework—that is, at fixed points in space (**Figure 1A**).

The Lagrangian trajectories can be computed online, i.e., concurrently with the Eulerian run, or offline using archived Eulerian velocity fields. The latter strategy, initiated by Döös (1995) and Blanke and Raynaud (1997), is often preferable because it enables the calculation of more trajectories, without the need for strict assumptions about the distribution of the initial particle locations. Multiple offline computations require much less CPU time than online calculations because the OGCM does not need to be run every time. Finally, unlike online calculations, offline computations allow tracking the particle backward in time. However, offline computations are highly dependent on the archiving frequency of the OGCM results. This frequency needs to be settled depending on the grid resolution and scientific goals.

The Lagrangian computation is based on the trajectory equation, solved at successive positions:

$$\frac{d\vec{x}_p}{dt} = \vec{v}\left(\vec{x}_p, t\right) \tag{3}$$

where $\vec{x}_p$ is the location of particle p, $\vec{v}$ is the local velocity field with (u, v) components in a two-dimensional field [or (u, v, w) components in a three-dimensional field] (**Figure 1B**). Several Lagrangian tools have been developed, for example Ariane (Blanke and Raynaud, 1997), Tracmass (Döös et al., 2013), or Roff (Carr et al., 2008). We focus here on Ariane for which the assumption of mass conservation, i.e., the non-divergence of the velocity field, enables the robust tracking of particles in a three-dimensional velocity field.

Two types of Lagrangian analyses can be run with a tool such as Ariane: *qualitative* analyses allow the description of full trajectory details whereas *quantitative* analyses give average pathways within a region of interest. The former consists of the computation of individual trajectories from initial positions of particles that can be deployed anytime and anywhere within the Eulerian grid, over the whole domain or over selected smaller sub-domains (**Figure 1A**). At each time step of the Lagrangian integration, the position of the particles and the associated physical properties (position, salinity, temperature and depths) are recorded. Ideally, the number of particles does not exceed several thousand so they can be treated individually; this must be optimized depending on the study and computational hardware. Quantitative experiments are a generalization of qualitative experiments and consist of deploying up to several millions of particles along a control section in a predefined subdomain, during a given period (usually several weeks to several years). The experiment ends when most particles have exited the subdomain. Additional diagnostics can be performed along each trajectory computation, by testing a criterion at each time step. This criterion can be the time of integration or a spatial constraint defined by a threshold of physical properties. When the criterion is met, the particle's trajectory is stopped. The movement of a water mass, defined by the particles that compose it, can then be studied using the trajectories of these particles (Blanke and Raynaud, 1997; Blanke et al., 1999), and the physical properties can be recovered along the trajectories. From a physical point of view, quantitative experiments help determine the Lagrangian stream function and mean circulation in a specific oceanic region. In both types of experiments, the particles can be followed forward or backward in time. The latter technique is useful for tracing the origins of a given water mass.

In the real ocean, small-scale movements associated with turbulent diffusion can alter the particle motions. Lagrangian methods can account for this diffusion using a random walk algorithm that adds a stochastic component to the deterministic part of the flow. This random walk approach has also been used to simulate the biological behavior of individual organisms (e.g., Sakai, 1973) or organism aggregation (Yamazaki and Haury, 1993). However, this approach introduces a supplementary degree of complexity. In "pure advective methods" as in Ariane, the advection term is dominant and this small-scale

turbulent diffusion is ignored. In this case, only the deterministic velocity components from the OGCM are used to calculate the trajectories. This is a common approach (e.g., Chen et al., 2003; Pous et al., 2010; Auger et al., 2015) and we recognize its limitation. However, OGCMs include such turbulent mixing as a subgridscale diffusivity, thus its signature is present in the physical and biogeochemical properties interpolated along the Lagrangian trajectories. Moreover, if diffusion were included in the movements using a random walk algorithm, one should be very reserved about the interpolation of backward integrations to trace the history of water parcels because of the irreversibility of the diffusion process. Note that the computed trajectories might vary depending on the archiving frequency; with a higher frequency, trajectories more accurately follow fine-scale structures. Accounting for fine-scale structures in the Eulerian framework, the variations of physical tracers along the corresponding Lagrangian trajectories are thus related to the lateral and vertical turbulent mixing processes that are deliberately ignored in the calculation of the particle motions.

Extension of the Combined Method for Analysis of Plankton Dynamics

Our present focus on the first trophic level, exclusively composed of passive planktonic organisms whose dynamics are mainly constrained by advection, allows us to use the Lagrangian approach as a "pure advective model". Such an approach has been used for decades and, as detailed in Section Modeling Biological-physical Interactions, numerous trajectory models have been developed, often focusing on the transport of planktonic organisms such as zooplankton (Qiu et al., 2010), fish eggs or fish larvae (Lett et al., 2007; Pous et al., 2010; Blanke et al., 2012) and more recently on jellyfish (Berline et al., 2013) or young turtles (Gaspar et al., 2012).

Note that the temporal and spatial scales associated with physical and biogeochemical/ecological processes are different: most physical processes involved in these transport studies vary over weeks to months, while biological processes can fluctuate at scales less than a day. These differences in scale imply that the Lagrangian tools used until now with relatively coarse resolution OGCMs were not well suited for planktonic marine ecology—and in particular the study of the first trophic level. However, recent advances in numerical methods allowing higher temporal and spatial resolution of the OCGMs have enabled the modeling of physical processes at scales that match those of the biological processes.

The diffusion terms in Equation (2) are neglected in our Lagrangian approach under the assumption that they are much smaller than the biological terms. This simplifies the interpretation of biological variations along the trajectories: the variations experienced by biogeochemical tracers are driven only by biological processes. Diagnostics of a given tracer, including biological stock and fluxes, can be performed along the Lagrangian trajectories $\vec{x}_p$: quantitative experiments help to assess planktonic dynamics along the main circulation pathways at regional scales (applied in Section Biological Transformations along Regional-scale Pathways), while qualitative experiments help to assess the biological transformations along specific trajectories (in Section Biological Transformations along Individual Trajectories).

Combine Method to Follow Mesoscale Features and associated Biological Processes

The basic principles through which mesoscale eddies act on oceanic tracers are well understood (Bleck et al., 1988; Gent and McWilliams, 1990; Lee et al., 1997). However, their roles in ecosystem dynamics have not been fully described and explained despite continued interest in the effects of mesoscale turbulence on biogeochemical activity (Gower et al., 1980; Jenkins, 1988; McGillicuddy et al., 1998, 1999, 2001; Lévy, 2008; Klein and Lapeyre, 2009) and numerous available observations (e.g., Nencioli et al., 2008). In particular, because the overall effect of mesoscale turbulence is a mixture of advective and diffusive processes, it is difficult to delimit the boundaries of eddies and assess the associated biophysical dynamics. We used the Lagrangian approach to follow numerical particles trapped in eddies that were coherent and quasi-axisymmetric, and analyze the biological processes occurring in these features.

Multiple techniques have been used to detect and delimit eddy boundaries; the most common can be merged into two categories. The first method is based on physical fields such as the Sea Surface Height (SSH) anomaly (Henson and Thomas, 2008; Chaigneau et al., 2009; Chelton et al., 2011), the vorticity (McWilliams et al., 1999), the velocity gradient tensor or Okubo-Weiss (OW) parameter (Isern-Fontanet et al., 2003, 2004; Morrow et al., 2004; Chelton et al., 2007; Sangrà et al., 2009). The second method detects the curvature or the shape of instantaneous streamlines, assuming that eddies evolve as quasi-circular flow patterns. These geometrical approaches include the winding-angle method (Sadarjoen and Post, 2000), vector geometry (Nencioli et al., 2010), wavelets (Luo and Jameson, 2002; Doglioli et al., 2007; Sangrà et al., 2009) or Lagrangian coherent structures (Beron-Vera et al., 2008). The SSH anomaly and OW criteria remain the most popular approaches. There is no acknowledged best method and both categories have their limitations. For example, a physical-based method often requires the definition of a threshold to determine the boundaries of the eddy: small variations of this threshold can significantly change its geometry (Souza et al., 2011). Disadvantages of the geometric method include the possible inclusion of filaments surrounding the eddy (with the wavelet methods) and substantial computational costs (with the winding-angle method). Sometimes, these techniques are combined to allow for better eddy detection, tracking and delimitation (e.g., Nencioli et al., 2010).

Here, we propose a method to define boundaries within mesoscale structures through qualitative experiments using the Lagrangian approach by defining specific areas based on particle positions and properties. Particles that are trapped within eddies and transported together can be used to define the extent of relevant areas in relation to the moving coherent structure. These areas—dynamical boxes—are used to compute diagnostics of the biological and physical processes calculated from the Eulerian model. The details of this combined Eulerian-Lagrangian method are developed in Section Biological Transformations in Dynamical Boxes.

Combined Eulerian-Lagrangian Analyses in EBUS

An Introduction to EBUS

In this section, we present various applications of (1) the extended Lagrangian approach to get biogeochemical variations along trajectories and (2) the Eulerian-Lagrangian approach to study planktonic ecosystem dynamics. The examples presented here focus mainly on two of the four major EBUS: the Benguela Current upwelling system located along the southwest coast of Africa (**Figure 2A**, and Section Biological Transformations along Regional-scale Pathways, and Section Biological Transformations along Individual Trajectories: A Case Study and the Importance of the Eulerian Output Frequency), and the California Current upwelling system along the U.S. west coast (**Figure 2B**, and Section Biological Transformations along Individual Trajectories: A Case study to Assess Biology along Trajectories, and Section Biological Transformations in Dynamical Boxes).

EBUS provide numerous unresolved questions concerning biophysical interactions. EBUS are characterized by strong alongshore seasonal winds that create a depression of the sea surface along the coast. This depression is compensated by an intense upwelling of deep, cold, salty and nutrient-rich water. This fertilization leads to sustained biological activity of the planktonic system and higher trophic levels. Enhanced biological activity, concentrated at the coast or nearshore, contrasts with offshore oligotrophic (i.e., nutrient-poor, low-productivity) conditions. The resulting cross-shore gradient is influenced by export of coastal material to the offshore region. This export is characterized by high variability induced by wind-driven Ekman transport and mesoscale eddy activity (Combes et al., 2013), as well as submesoscale processes including downward subduction in filaments (Ramp et al., 1991; Lathuilière et al., 2010).

EBUS regimes show pronounced lateral gradients in their physical and biogeochemical ocean components. Therefore, they represent excellent systems for exploring the role of physical processes on planktonic ecosystem dynamics using the combined Eulerian-Lagrangian approach. We introduce quantitative experiments to study large-scale processes (in Section Biological Transformations along Regional-scale Pathways), then we present qualitative experiments to follow biological transformations along specific trajectories (in Section Biological Transformations along Individual Trajectories), and within dynamical boxes (in Section Biological Transformations in Dynamical Boxes).

Biological Transformations Along Regional-scale Pathways

We address the use of the combined Eulerian-Lagrangian approach in a biophysical interaction study first using quantitative experiments. Here, we aim to trace biogeochemical properties in relation to the dominant circulation. The ROMS hydrodynamic model was configured for the Benguela Current upwelling system with 8 km horizontal resolution. The ecosystem model is a simple NPZD model that was previously used in this region (Koné et al., 2005). The Eulerian model output consists of two calendar years, archived daily. To study the across-shore transport of biogeochemical components and evolution of the properties of nearshore water masses after leaving the coastal region, Lagrangian experiments were performed over a sub-domain that covers the continental slope (**Figure 2A**). Lagrangian particles were deployed along a section close to the coastline (the control section) over 360 days. They were intercepted along an offshore section parallel to the control section after a maximum duration of 360 days (**Figure 3**).

The average cross-shore pathways taken by the nearshore particles are given by the Lagrangian stream function (**Figure 3**).

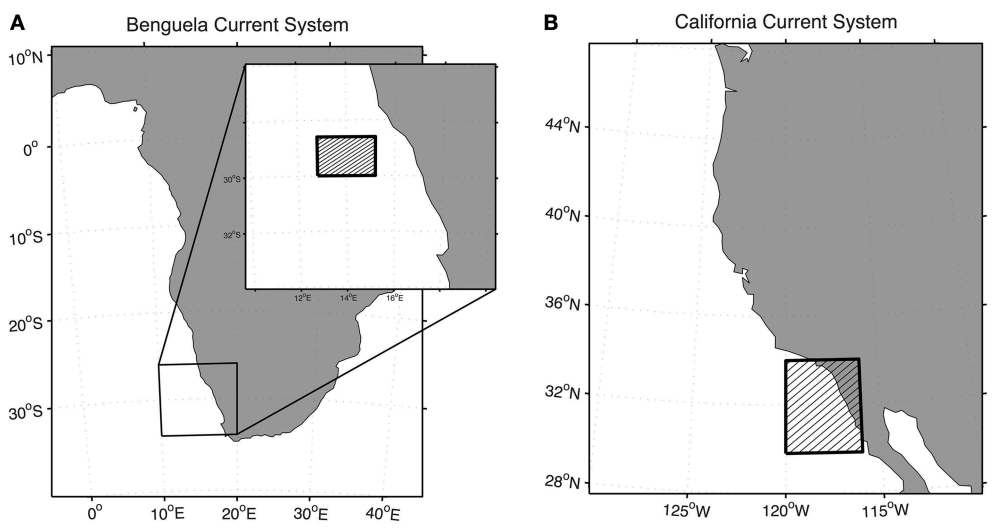


FIGURE 2 | Regions of study. (A) The Benguela Current upwelling system located along the southwest coast of Africa and **(B)** the California Current upwelling system along the U.S. west coast. For each region, the studies focused on the shaded area.

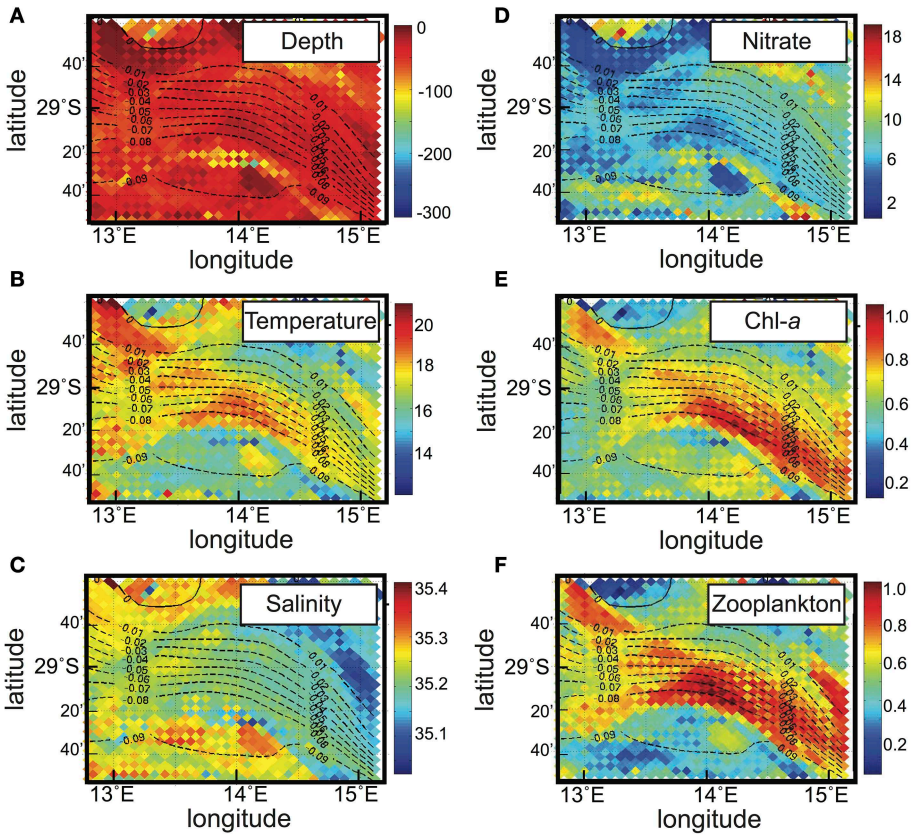


FIGURE 3 | Results from a quantitative experiment in the Benguela Current upwelling system (Figure 2A). Vertical averages of **(A)** depth (m), **(B)** temperature (°C), **(C)** salinity, **(D)** nitrate concentration (mmolN m^{-3}), **(E)** chlorophyll-a concentration (mg Chl-a m^{-3}) and **(F)** zooplankton concentration (mmolN m^{-3}), calculated along trajectories of the particles that traveled offshore, from the right to the left in this longitudinal section. The corresponding mean Lagrangian stream function is superimposed (black dashed contours) with a 0.01 Sv interval.

The associated transport reaches ∼0.1 Sv. This dominant connection occurs in the surface layers (i.e., the upper 100 m), with a significant cross-shore transport of fresh and warm water compared to the surrounding water masses (**Figures 3A–C**), which is associated with many eddies traveling offshore (not shown). Because the dominant pathway occurs within the euphotic layer, i.e., the illuminated part of the water column where primary production—the basis of biological activity—occurs, interesting results concerning biological tracers might be expected. Moreover eddies associated with this main pathway are likely to trap and transport biological material from the coast (Chenillat et al., submitted ms. a,b). Indeed, compared to surrounding waters, along this pathway both the nutrient and planktonic concentrations seem to be affected: (i) the nutrient concentration is lower, because it is likely consumed by the phytoplankton; (ii) both chlorophyll-a (a proxy for phytoplankton concentration) and zooplankton biomasses are higher (**Figures 3D–F**). Thus, this Eulerian-Lagrangian approach has allowed us to explain and quantify the changes of nearshore planktonic organisms in relation to the physical cross-shore transport, at a regional scale. This cross-shore transport seems here to be driven by mesoscale activity; unfortunately, these analyses are insufficient to clarify the effect of this mesoscale activity on ecosystem dynamics.

In the context of biophysical interaction studies, this approach has many potential applications. For example, in regions with strong biogeochemical or physical gradients, quantitative experiments allow assessment of the transport, and the redistribution and associated changes of biological components. Large gradients of biogeochemical and physical components may develop along strong currents, along the coastline or in the cross-shore direction, or between two or more basins. Eulerian-Lagrangian diagnostics can help us understand (i) how physical properties maintain these large gradients, (ii) the exchanges among oceanic biogeographical provinces, and (iii) at finer scale, the effects of runoff on nearshore and coastal ecosystems.

A consideration of time scales is essential in such offline studies: the frequency of the available Eulerian outputs can significantly affect the results. The influence of this frequency on the biogeochemical inferences from the Lagrangian experiments is tested in the following section (Biological Transformations along Individual Trajectories).

Biological Transformations along Individual Trajectories

A Case Study and the Importance of the Eulerian Output Frequency

We performed qualitative experiments with the Lagrangian approach to find the values of biogeochemical tracers along particle paths. The trajectory of one particular particle is shown in **Figure 4A**, with the same Eulerian output used in Section Biological Transformations along Regional-scale Pathways. This particle was selected from among many other particles initially released along the nearshore section at around 30 m depth (see **Figure 3**). The trajectory shows several loops as the particle moves offshore (**Figure 4A**), indicating that it was trapped in an eddy. Within the eddy, the particle was upwelled to the surface, experiencing an increase in temperature of ∼3°C and changes in the local biogeochemistry (not shown). This experiment clearly shows that a qualitative experiment—calculated from daily Eulerian velocity output—enables tracking of particles trapped in a mesoscale eddy. However, along-trajectory variations of biogeochemical properties could be sensitive to the frequency of the Eulerian output, leading to uncertainties on their interpretation. We thus set up specific diagnostics to study the impact of the frequency of the Eulerian output on the Lagrangian biogeochemical results.

The more frequent the Eulerian output and the higher the grid resolution, the more accurate the resolution of small-scale features in the Lagrangian trajectories. Moreover, the ecosystem processes occur at high frequency, i.e., at scales shorter than a day, which is modeled through a suitable parameterization in the Eulerian framework. Because data storage is an issue for modelers, it is essential to estimate the optimal output frequency: it must be fast enough to capture biological transformations, but coarse enough to generate a manageable amount of data. The Eulerian output frequency influences the results obtained in the offline combined Eulerian-Lagrangian approach, as shown below through a study that compared several frequencies for the same Eulerian model.

The purpose of this analysis is to compare different archiving intervals: hourly, daily and every 5 days. We chose the hourly output as a reference for the Lagrangian computations because it is fully compatible with the time scales of the biological processes and is close to the best achievable given the Eulerian model time step, usually of around 500–1000 s. The daily and 5-day archiving outputs are obtained by degrading the hourly output with a simple arithmetic average. Our diagnostics were performed on a population of particles initialized around the selected particle (**Figure 4A**) using a seeding strategy that consisted of duplicating the initial position in its closest neighborhood in time (t_0) and space (x_0, y_0, z_0) (as illustrated in **Figure 4B**). A total of 21^3 (9261) particles were released. Only a few trajectories are represented in **Figure 4C** for the three different output intervals. This shows that despite the use of the same initial conditions, some trajectories and associated properties differ depending on the interval. These differences can be highlighted by comparing the tracer values when the particles reach the offshore section (**Figure 5**). We chose to explore a physically conservative tracer—salinity, and a non-conservative biological tracer—chlorophyll-a. The results obtained at hourly intervals are used as the reference

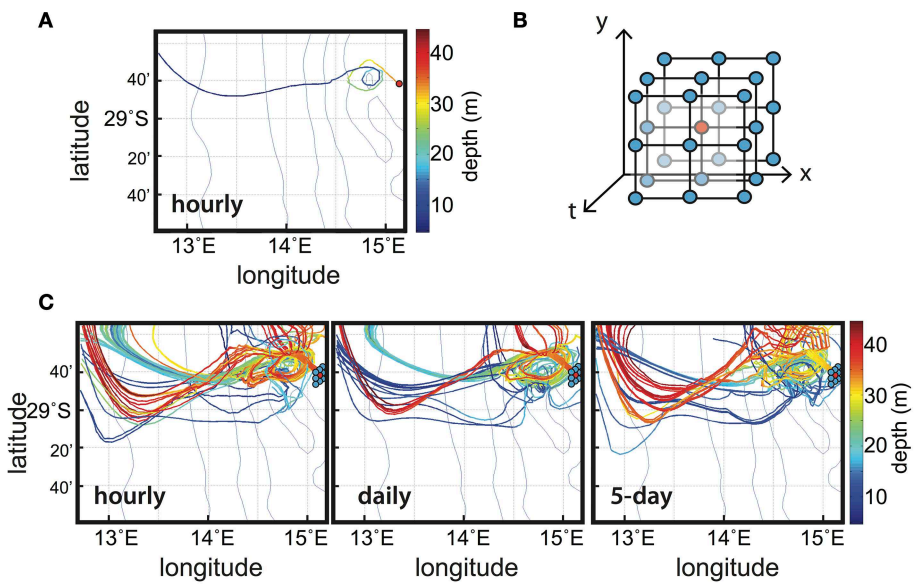


FIGURE 4 | Results from qualitative experiments in the Benguela Current upwelling system (Figure 2A). (A) Trajectory of one reference particle that was temporarily trapped in a mesoscale eddy, calculated with hourly Eulerian outputs; its initial position is shown by the red dot. **(B)** Seeding strategy (2-dimensional in space and 1-dimensional in time) in the immediate neighborhood of the initial position (the red dot). **(C)** Trajectories of the particles initialized around the reference initial position for several sampling frequencies of the Eulerian model output: hourly (left), daily (middle), and 5-day (right). In **(A,C)** the color bar shows depth (in m) along trajectories.

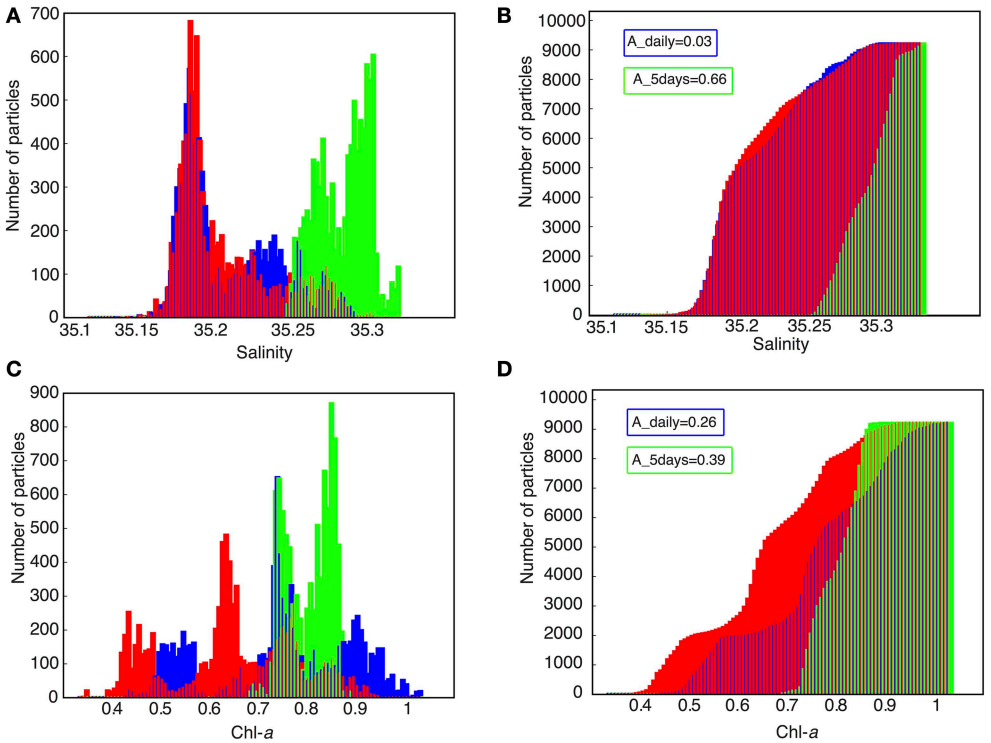


FIGURE 5 | Effect of various Eulerian archiving intervals on Lagrangian qualitative experiments. Histograms **(A–C)** and cumulative histograms **(B–D)** of some properties of the particles initialized in the immediate neighborhood of the reference particle (see **Figure 3**), after trajectory calculations using hourly (red), daily (blue), and 5-day (green) subsampling frequencies of the Eulerian model output. The properties are salinity **(A,B)** and chlorophyll-a concentration **(C,D)**.

(**Figures 5A,C**). The more the results at daily intervals and 5-day intervals differ from the reference, the less accurate they are. Cumulative histograms (**Figures 5B,D**) are useful for comparing the individual distributions to evaluate the effects of sampling interval. The following index estimates the differences between the test and reference distributions by integrating the square of the area of the difference between the corresponding cumulative histograms:

$$A = \frac{\sum_{i=1}^{N}(Tref\ (i) - T(i))^2 \Delta T}{\sum_{i=1}^{N}(Tref\ (i))^2} \quad (4)$$

where T is the physical or biogeochemical property tested (i.e., from either the daily or 5-day experiment), *Tref* is the value of reference (calculated from the hourly experiment), N is the total number of classes of the histogram, and ΔT is the class size of the histogram. ΔT is chosen to account for a total of 100 classes, and thus A represents a percentage of area difference. The 5-day output histograms are strikingly different from the hourly output histograms for both physical ($A \sim 66\%$) and biogeochemical tracers ($\sim 39\%$). This shows that 5-day output is far too coarse to reproduce accurate biophysical interactions. At the daily scale, the sensitivity to the Eulerian sampling is more marked for biogeochemical tracers ($A \sim 26\%$) than for physical tracers ($\sim 3\%$). This sensitivity to hourly vs. daily archive resolution is explained by the rapidity of biological processes: the induced non-linearities must be sampled at a suitable frequency from the Eulerian archive to ensure the quality of the Lagrangian diagnostics. It is important to note that biophysical processes can occur at time scales shorter than a day or a hour. For example, internal gravity waves can rapidly impact vertical nutrient fluxes (Wang et al., 2007). Eulerian regional models are just starting to resolve such super-inertial dynamics (e.g., Alford et al., 2015). To explicitly account for these processes in the Lagrangian model, one would need output at intervals of a few minutes. At present this would place a heavy burden on data storage, though it will become increasingly feasible in the future.

Combined Eulerian-Lagrangian methods clearly need to be used with caution when analyzing biogeochemical fields from Eulerian models. The optimal archive strategy is likely the highest frequency that allows sampling of the fastest non-linear biogeochemical processes while maintaining reasonable data storage costs. In the present study, with an eddy-resolving model of $\sim$5 km horizontal resolution, we chose a daily archive to conduct the biophysical analyses. Note that storage constraints may be somewhat lessened by aggregating horizontal grid point information as explored for offline Eulerian tracer experiments by Lévy et al. (2012).

A Case Study to Assess Biology along Trajectories

Qualitative experiments allow not only assessment of properties along individual trajectories but also visualization of the complete

story of parcels of water. We present here a study that follows water trapped in a mesoscale eddy in the California Current system (Chenillat et al., submitted ms. a,b). In this numerical experiment, ROMS was coupled with an NPZD model (with several nutrient, phytoplankton, and zooplankton boxes) to study one particular cyclonic (counterclockwise) eddy evolving in the Southern California Bight (SCB) to understand how dynamical and biological processes were able to maintain biological activity within such a mesoscale feature. They chose an eddy formed at the coast that trapped and transported some nearshore water and its associated coastal upwelling ecosystem offshore.

The combined Eulerian-Lagrangian approach (**Figure 6**) was implemented as follows: a 2-year-long simulation was run with the Eulerian model (the first and the second year are noted Y1 and Y2, respectively) with daily output. A cyclonic eddy was chosen from this simulation based on a local minimum of the SSH (see **Figure 7C**) and locally enhanced biological activity. Over 7 months, this eddy traveled from the coast to the offshore region (**Figure 7**). From the daily archive of this 2-year-long simulation, several sets of qualitative Lagrangian experiments were performed to follow the eddy, as detailed in Chenillat et al. (submitted ms. a) (**Figure 6**). In a pre-experiment ∼500 particles were released along the coastline (0–200 m depth), monthly from September Y1 to March Y2 (on the first day of each month). The particles were followed until 30 March Y2 when they had spread over the SCB. Only particles within the eddy (see **Figure 7C**) were selected for the main Lagrangian experiments. These particles were tracked backward (see **Figures 7A–B**) and forward (**Figure 7D**) in time to identify the source of the particles, and their ultimate fate.

The gathering of physical information along the trajectories allows quantification of the kinematics of this eddy and an accurate spatial definition of the eddy's edge and core (Chenillat et al., submitted ms. a). Biological properties along these trajectories show different characteristics depending on the position of the particles within the eddy: biological activity was higher in the core than at the edge, and both regions were higher than the regional average (not shown).

Though this approach gives some insights concerning the properties in the eddy, it does not reveal the dynamics and biophysical interactions that structure the ecosystem. Additional diagnostics are thus needed and will be developed in Section Biological Transformations in Dynamical Boxes.

This combined qualitative Eulerian-Lagrangian approach to study biophysical interactions was also used by Auger et al. (2015), who released particles nearshore in a region of high cadmium (Cd) concentration, to diagnose the dispersion plume of Cd-rich water in an EBUS. They aimed to identify the source of Cd (natural or anthropogenic) that was bio-accumulating, estimated from Cd concentrations and primary production, along cross-shore trajectories.

More generally, this method can be used in marine systems to identify the origin or fate of any parcel of water, and understand any biological processes at work in the Eulerian ecosystem model. All modeled biological concentrations and fluxes can be assessed along Lagrangian trajectories, providing information about ecosystem functioning as in Auger et al. (2015). Derived properties like euphotic depth or nitracline depth can also be estimated along trajectories to define a water mass origin from biological point of view (D'Ovidio, pers. comm.). This method could provide information about water mass origins with the tracking of freshly advected nutrients from below the euphotic layer—controlling the phytoplanktonic new production—vs. "old" water that has traveled horizontally, controlling phytoplanktonic regenerated production.

Biological Transformations in Dynamical Boxes

In this last section, we highlight the role of mesoscale eddies in ecosystem dynamics by defining dynamical boxes. Mesoscale features move over time; the water linked to them can be followed using region boundaries that shift with the features: dynamical boxes. This method can help quantify and understand ecosystem evolution in such features, and assess ecosystem response to local physically driven forcing.

The dynamical box concept enhances the Lagrangian approach by providing a tool to analyze groups of particles trapped within coherent structures (see Section Biological Transformations along Individual Trajectories). The Lagrangian approach used in Chenillat et al. (submitted ms. a) showed that the cyclonic eddy was able to retain nearshore water for several months while moving offshore (**Figure 7**). This led to the

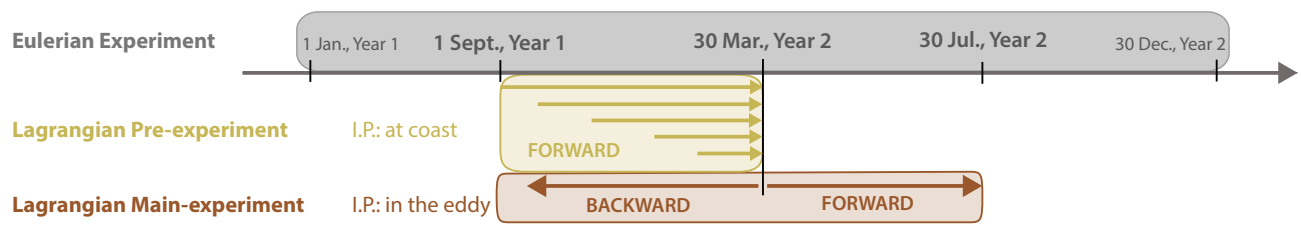


FIGURE 6 | Schematic view of combined Eulerian-Lagrangian qualitative experiments to track a specific eddy. The gray box represents the time axis of the 2-year long (year 1 and year 2, noted Y1 and Y2, respectively) Eulerian experiment (ROMS/NPZD). The yellow and brown boxes represent two complementary Lagrangian experiments based on the Eulerian outputs. The Lagrangian pre-experiment (yellow box) consisted of releasing ∼500 particles along the coastline (0–200 m depth) and monthly (from 1 September Y1 to 1 March Y2), and following them until 30 March Y2. The Lagrangian main experiment (brown box) consisted of selecting the particles present in an eddy on 30 March Y2 (see **Figure 7D**), and tracking them both backward and forward in time (I.P., Initial Positions).

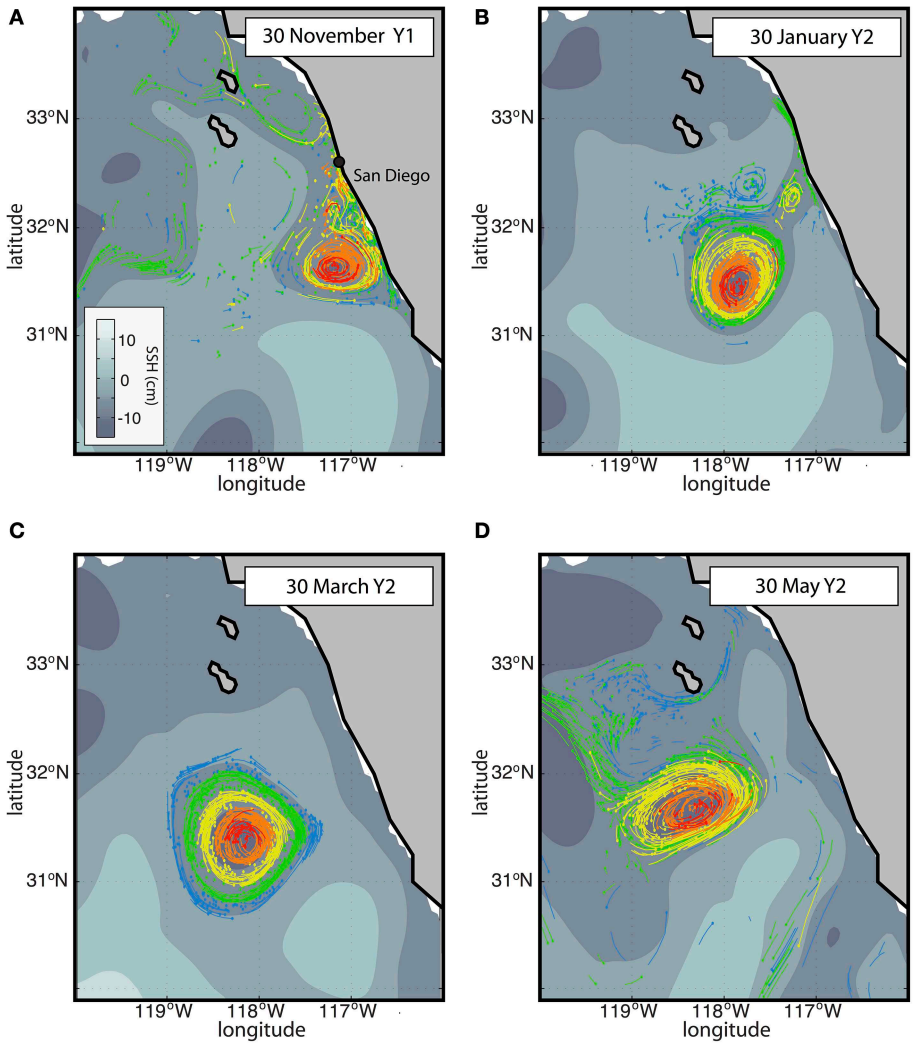


FIGURE 7 | Time evolution of the model SSH (Eulerian outputs, in gray scale) and Lagrangian particle positions (colored dots) in an eddy of the California Current upwelling system (Figure 2B). The color code refers to different pools that were defined based on their position in the eddy on 30 March Y2, from the center (red and orange) to the edge (green and blue). The core region is defined by the red and orange particles, experiencing solid body rotation during the eddy lifetime (Chenillat et al., submitted ms. a). The trajectories during the 3 days prior to each date are drawn with thin colored lines (to keep the figure clear, only half of them are represented). The particles on display are those trapped within the eddy on 30 March Y2 **(C)**, tracked either backward **(A,B)** or forward **(D)** in time (see **Figure 6**).

conclusion that there was little horizontal exchange within the eddy. Nevertheless, because the particles only represent a portion of the water trapped within the eddy, they do not capture all the effects of eddies on ecosystem dynamics or the possible vertical exchanges between the euphotic zone and deeper waters: the particles trapped in the eddy tracked coastal water whereas local vertical exchanges involved non-coastal eddy waters that were not tracked in the study.

To get a complete picture of the biological responses in the eddy core, we must assess and understand ecosystem evolution in response to physically driven vertical nutrient input (Falkowski et al., 1991; McGillicuddy and Robinson, 1997). Using dynamical boxes defined according to the Lagrangian particle kinematics, Chenillat et al. (submitted ms. a) characterized the eddy core region as those particles experiencing solid body rotation during the eddy lifetime (**Figure 8A**). This core region changed spatially and moved over time, precluding the use of a stationary analysis region. Within this dynamical box defined by the Lagrangian particles, Eulerian diagnostics of biophysical processes showed that the core area remained relatively constant (Chenillat et al., submitted ms. a,b).

The Eulerian diagnostics in these dynamic boxes require horizontal averaging of the Eulerian output over the box region at each time step. Vertical exchanges such as loss of material by biological sinking (output from the eddy) or upwelling of nutrient by vertical advection (input) can then be quantified. In these dynamical boxes, any biological

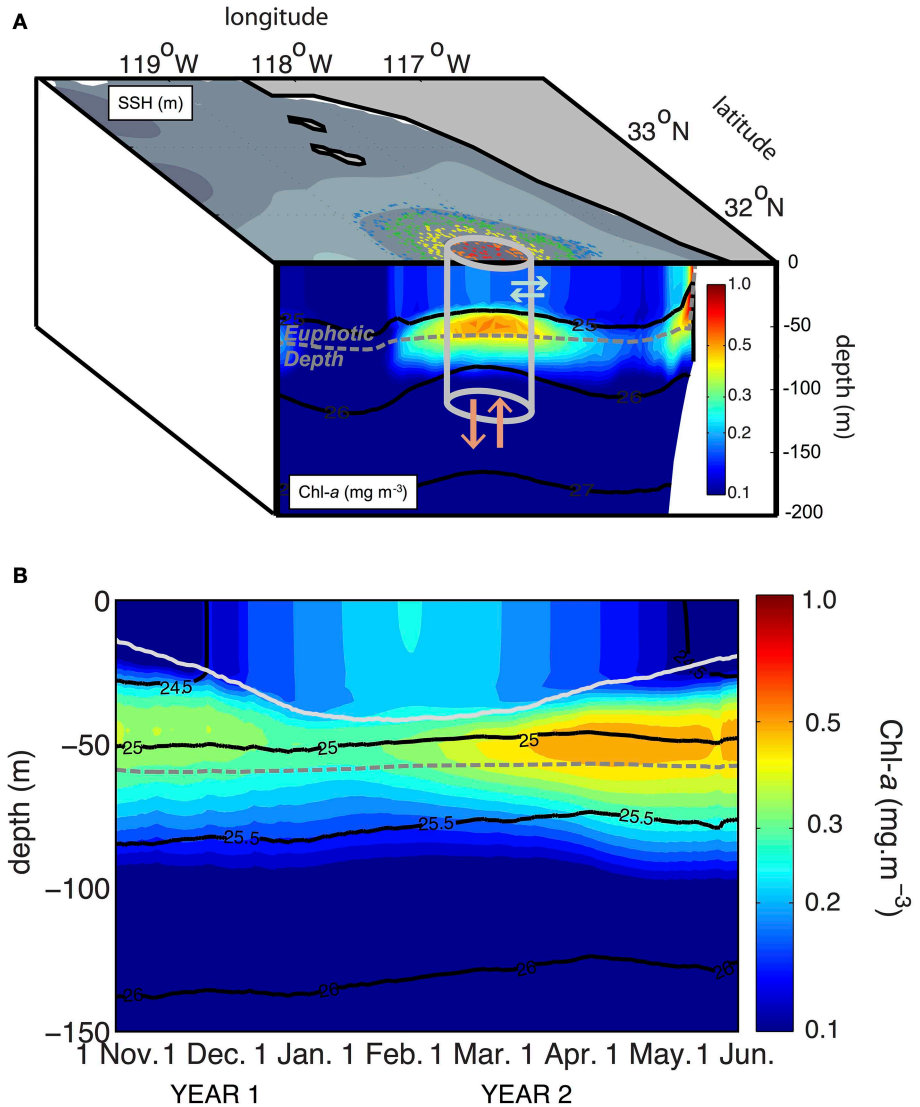


FIGURE 8 | Biological response in an eddy of the California Current upwelling system. (A) Schematic of the dynamical box (the eddy core) defined by the red and orange pools of particles (see **Figure 7C**). The eddy core contains an enhanced subsurface chlorophyll-a maximum. Within this box, at each time step, vertical exchanges (orange arrows) for example across the euphotic depth can be assessed. **(B)** Time evolution of the total chlorophyll-a concentration averaged horizontally over the eddy core dynamical box. Superimposed on the chlorophyll-a concentration (color scale in **A,B**) are isolines of density (black) and euphotic depth (dashed gray line). In **(B)**, the depth of the mixed layer is represented by the light gray line.

component represented in the ecosystem model can be diagnosed with high confidence due to the weak lateral exchanges within the box (eddy) or with the surrounding waters (**Figure 8A**). These components include biological concentrations (nutrients, phytoplankton, zooplankton), biological fluxes (e.g., nutrient uptake for growth, grazing, remineralization) or advective fluxes of biological components (i.e., any tracer concentrations advected by the vertical or horizontal flow). More precisely, within the dynamical box and at each time step, horizontal averages can be applied to calculate the vertical structure of any physical or biological tracer in the water column (**Figure 8B**). At each time step, we can also estimate the euphotic depth and evaluate any vertical exchange across this boundary, such as the nutrient input driven by advective fluxes or sinking of material (Chenillat et al., submitted ms. b).

Dynamical boxes were also defined for the waters directly surrounding the eddy. Analyses of the properties of particles in this box showed that biological activity was lower than in the eddy core, that vertical input of nutrients to the euphotic layer were low, and that horizontal exchanges were not negligible. This dynamical box method is an excellent tool for comparing processes among regions; in this case they revealed the significance of vertical nutrient input for maintaining biological

activity in the eddy core compared to surrounding waters. The method can be generalized to study vertical exchanges within any coherent structure. Another interesting and fruitful application of the Eulerian-Lagrangian method combined with dynamical boxes would be to compare the dependence of the biology in eddies on the eddy properties such as their site of generation, their time of formation, their intensity, their persistence, and whether they are cyclonic or anticyclonic (see Morales et al., 2012 and references within).

Conclusions

We have developed a combined Eulerian-Lagrangian approach to study planktonic dynamics in moving fluids from numerical studies. Lagrangian approaches are particularly useful for revealing physical pathways, and the dynamics and distributions of tracers along these trajectories. A similar combined approach has been used previously in oceanography to study biological processes in response to ocean dynamics (e.g., D'Ovidio et al., 2010; Blanke et al., 2012; Berline et al., 2013); our approach provides new perspectives: it allows quantification of biogeochemical tracer dynamics both along particle trajectories, and within coherent structures—such as eddies—that are characteristic of a turbulent flow. Because physical and biological processes often occur at different frequencies, combined Eulerian-Lagrangian methods must be used with caution when analyzing biogeochemical fields from Eulerian models: the Eulerian output frequency will influence the results obtained from the offline combined Eulerian-Lagrangian approach. We showed that the optimal archive strategy balances the highest frequency that allows sampling of the fastest non-linear biogeochemical processes while maintaining reasonable data storage costs: it must be fast enough to capture biological transformations, but coarse enough to generate a manageable amount of data.

The application of this approach was focused on the planktonic ecosystem response in eastern boundary upwelling systems, where ocean dynamics drive export of coastal material offshore through intense cross-shore transport events in coherent structures. This method can be widely used to quantify any tracer dynamics in moving fluids. The code for the combined Eulerian-Lagrangian method is available upon request (Ariane-Tracer); we will use it in future work to assess the origin and age of water forming a frontal structure. In general, this method can be used to quantify any tracer dynamics, biotic or abiotic, evolving in any moving fluid or in a non-homogenous environment.

Acknowledgments

Support for this study has been provided by CNES (Centre National d'Études Spatiales), as part of the Ifesta-Up project funded by the TOSCA (Terre, Ocean, Surfaces Continentales, Atmosphère) program and by National Science Foundation (NSF) funding (OCE-10-26607) to the CCE-LTER site.

References

Abbott, M. R., Brink, K. H., Booth, C. R., Blasco, D., Codispoti, L. A., Niiler, P. P., et al. (1990). Topography san syntactic foam. *J. Geophys. Res.* 95, 9393–9409. doi: 10.1029/JC095iC06p09393

Alford, M. H., Peacock, T., MacKinnon, J. A., Nash, J. D., Buijsman, M. C., Centuroni, L. R., et al. (2015). The formation and fate of internal waves in the South China Sea. *Nature* 521, 65–69. doi: 10.1038/nature14399

Auger, P.-A., Machu, E., Gorgues, T., Grima, N., and Waeles, M. (2015). Comparative study of potential transfer of natural and anthropogenic cadmium to plankton communities in the North-West African upwelling. *Sci. Total Environ.* 505, 870–888. doi: 10.1016/j.scitotenv.2014.10.045

Bailey, D., Berndt, M., Kucharik, M., and Shashkov, M. (2010). Reduced-dissipation remapping of velocity in staggered arbitrary Lagrangian–Eulerian methods. *J. Comput. Appl. Math.* 233, 3148–3156. doi: 10.1016/j.cam.2009.09.008

Ban, V. S., and Gilbert, S. L. (1975). The chemistry and transport phenomena of chemical vapor deposition of silicon from SiCl 4. *J. Cryst. Growth*. 31, 284–289. doi: 10.1016/0022-0248(75)90142-6

Batchelder, H. P., Edwards, C. A., and Powell, T. M. (2002). Individual-based models of copepod populations in coastal upwelling regions: implications of physiologically and environmentally influenced diel vertical migration on demographic success and nearshore retention. *Prog. Oceanogr.* 53, 307–333. doi: 10.1016/S0079-6611(02)00035-6

Berline, L., Zakardjian, B., Molcard, A., Ourmières, Y., and Guihou, K. (2013). Modeling jellyfish *Pelagia Noctiluca* transport and stranding in the ligurian Sea. *Mar. Pollut. Bull.* 70, 90–99. doi: 10.1016/j.marpolbul.2013.02.016

Beron-Vera, F. J., Olascoaga, M. J., and Goni, G. J. (2008). Oceanic mesoscale eddies as revealed by Lagrangian coherent structures. *Geophys. Res. Lett.* 35, L12603. doi: 10.1029/2008gl033957

Bird, R. B., Stewart, W. E., and Lightfoot, E. N. (2007). *Transport Phenomena, 2nd Edn.* New York, NY: John Wilwe & Sons, Inc.

Blanke, B., Arhan, M., Madec, G., and Roche, S. (1999). Warm water paths in the equatorial Atlantic as diagnosed with a general circulation model. *J. Phys. Oceanogr.* 29, 2753–2768.

Blanke, B., Bonhommeau, S., Grima, N., and Drillet, Y. (2012). Sensitivity of advective transfer times across the North Atlantic Ocean to the temporal and spatial resolution of model velocity data: implication for European eel larval transport. *Dyn. Atmos. Oceans.* 55–56, 22–44. doi: 10.1016/j.dynatmoce.2012.04.003

Blanke, B., and Raynaud, S. (1997). Kinematics of the pacific equatorial undercurrent: an eulerian and lagrangian approach from GCM results. *Oceanogr. J. Physical.* 27, 1038–1053.

Bleck, R., Onken, R., and Woods, J. D. (1988). A two-dimensional model of mesoscale frontogenesis in the ocean. *Q. J. R. Meteorol. Soc.* 114, 347–371. doi: 10.1002/qj.49711448005

Brinton, E. (1967). Distributional atlas of *Euphausiacea* (*Crustacea*) in the California current region. *Calif. Coop. Oceanic Fish. Invest.* 5:275.

Carr, S. D., Capet, X., McWilliams, J. C., Pennington, J. T., and Chavez, F. P. (2008). The influence of diel vertical migration on zooplankton transport and recruitment in an upwelling region: estimates from a coupled behavioral-physical model. *Fish. Oceanogr.* 17, 1–15. doi: 10.1111/j.1365-2419.2007.00447.x

Chai, F., Dugdale, R. C., Peng, T., Wilkerson, F. P., and Barber, R. T. (2002). One-dimensional ecosystem model of the equatorial pacific upwelling system. Part I: model development and silicon and nitrogen cycle. *Deep Sea Res. Part II.* 49, 2713–2745. doi: 10.1016/S0967-0645(02)00055-3

Chaigneau, A., Eldin, G., and Dewitte, B. (2009). Eddy activity in the four major upwelling systems from satellite altimetry (1992–2007). *Prog. Oceanogr.* 83, 117–123. doi: 10.1016/j.pocean.2009.07.012

Chelton, D. B., Schlax, M. G., and Samelson, R. M. (2011). Global observations of nonlinear mesoscale eddies. *Prog. Oceanogr.* 91, 167–216. doi: 10.1016/j.pocean.2011.01.002

Chelton, D. B., Schlax, M. G., Samelson, R. M., and de Szoeke, R. A. (2007). Global observations of large oceanic eddies. *Geophys. Res. Lett.* 34, L15606. doi: 10.1029/2007gl030812

Chen, C., Xu, Q., Beardsley, R. C., and Franks, P. J. S. (2003). Model study of the cross-frontal water exchange on Georges Bank: a three-dimensional Lagrangian experiment. *J. Geophys. Res.* 108, 3142. doi: 10.1029/2000JC000390

Chenillat, F., Rivière, P., Capet, X., Franks, P. J. S., and Blanke, B. (2013). California coastal upwelling onset variability: cross-shore and bottom-up propagation in the planktonic ecosystem. *PLoS ONE* 8:e62281. doi: 10.1371/journal.pone.0062281

Combes, V., Chenillat, F., Di Lorenzo, E., Rivière, P., Ohman, M. D., and Bograd, S. J. (2013). Cross-shore transport variability in the California current: ekman upwelling vs. eddy dynamics. *Prog. Oceanogr.* 109, 78–89. doi: 10.1016/j.pocean.2012.10.001

Davis, R. E. (1983). Oceanic property transport, lagrangian particle statistics, and their prediction. *J. Mar. Res.* 41, 163–194. doi: 10.1357/002224083788223018

Davis, R. E. (1991). Observing the general circulation with floats. *Deep Sea Res. Part A.* 38, S531–S571. doi: 10.1016/S0198-0149(12)80023-9

Dewan, A. (2011). *Tackling Turbulent Flows in Engineering*. Berlin; Heidelberg: Springer-Verlag.

Doglioli, A. M., Blanke, B., Speich, S., and Lapeyre, G. (2007). Tracking coherent structures in a regional ocean model with wavelet analysis: application to cape basin eddies. *J. Geophys. Res.* 112, C05043. doi: 10.1029/2006JC003952

Döös, K. (1995). Interocean exchange of water masses. *J. Geophys. Res.* 100, 13499–13514. doi: 10.1029/95JC00337

Döös, K., Kjellsson, J., and Jönsson, B. (2013). "TRACMASS—a lagrangian trajectory model," in *Preventive Methods for Coastal Protection, Towards the Use of Ocean Dynamics for Pollution Control* (Berlin; Heidelberg: Springer), 225–249.

D'Ovidio, F., De Monte, S., Alvain, S., Dandonneau, Y., and Lévy, M. (2010). Fluid dynamical niches of phytoplankton types. *Proc. Natl. Acad. Sci. U.S.A.* 107, 57–59. doi: 10.1073/pnas.1004620107

D'Ovidio, F., De Monte, S., Della Penna, A., Cotté, C., and Guinet, C. (2013). Ecological implications of eddy retention in the open ocean: a lagrangian approach. *J. Phys. A Math. Theor.* 46, 1–21. doi: 10.1088/1751-8113/46/25/254023

Estrada, M., and Berdalet, E. (1997). Phytoplankton in a turbulent world. *Sci. Mar.* 61, 125–140.

Falkowski, P. G., Ziemann, D., Kolber, Z., and Bienfang, P. K. (1991). Role of eddy pumping in enhancing primary production in the ocean. *Nature* 352, 55–58. doi: 10.1038/352055a0

Franks, P. J. S. (2002). NPZ models of plankton dynamics: their construction, coupling to physics, and application. *J. Oceanogr.* 58, 379–387. doi: 10.1023/A:1015874028196

Gaspar, P., Benson, S. R., Dutton, P. H., Réveillère, A., Jacob, G., Meetoo, C., et al. (2012). Oceanic dispersal of juvenile leatherback turtles: going beyond passive drift modeling. *Mar. Ecol. Prog. Ser.* 457, 265–284. doi: 10.3354/meps09689

Gent, P. R., and McWilliams, J. C. (1990). Isopycnal mixing in ocean models. *J. Oceanogr.* 20, 150–155.

Gower, J. F. R., Denman, K. L., and Holyer, R. J. (1980). Phytoplankton patchiness indicates fluctuation spectrum of mesoscale oceanic structure. *Nature* 288, 157–159. doi: 10.1038/288157a0

Gruber, N., Lachkar, Z., Frenzel, H., Marchesiello, P., Munnich, M., McWilliams, J. C., et al. (2011). Eddy-induced reduction of biological production in eastern boundary upwelling systems. *Nat. Geosci.* 4, 787–792. doi: 10.1038/ngeo1273

Haury, L. R., McGowan, J. A., and Wiebe, P. H. (1978). "Patterns and processes in the time-space scales of plankton distributions," in *Spatial Pattern in Plankton Communities*, ed J. H. Steele (New York, NY: Springer), 277–327.

Henson, S. A., and Thomas, A. C. (2008). A census of oceanic anticyclonic eddies in the Gulf of Alaska. *Deep Sea Res. Part I* 55, 163–176. doi: 10.1016/j.dsr.2007.11.005

Hill, C., and Marshall, J. (1995). "Application of a parallel navier-stokes model to ocean circulation in parallel computational fluid dynamics," in *Proceedings of Parallel Computational Fluid Dynamics: Implementations and Results Using Parallel Computers* (New York, NY: Elsevier Science B.V), 545–552.

Isern-Fontanet, J., Font, J., García-Ladona, E., Emelianov, M., Millot, C., and Taupier-Letage, I. (2004). Spatial structure of anticyclonic eddies in the algerian basin (Mediterranean Sea) analyzed using the okubo–weiss parameter. *Deep Sea Res. Part II* 51, 3009–3028. doi: 10.1016/j.dsr2.2004.09.013

Isern-Fontanet, J., García-Ladona, E., and Font, J. (2003). Identification of marine eddies from altimetric maps. *J. Atmos. Oceanic Technol.* 1995, 772–778. doi: 10.1175/1520-0426(2003)20<772:IOMEFA>2.0.CO;2

Jenkins, W. J. (1988). Nitrate flux into the euphotic zone near Bermuda. *Nature* 331, 521–523. doi: 10.1038/331521a0

Kiørboe, T. (1993). Turbulence, phytoplankton cell size, and the structure of pelagic food webs. *Adv. Mar. Biol.* 29, 1–72.

Klein, P., and Lapeyre, G. (2009). The oceanic vertical pump induced by mesoscale and submesoscale turbulence. *Ann. Rev. Mar. Sci.* 1, 351–375. doi: 10.1146/annurev.marine.010908.163704

Koné, V., Machu, E., Penven, P., Andersen, V., Garçon, V., Fréon, P., et al. (2005). Modeling the primary and secondary productions of the southern benguela upwelling system: a comparative study through two biogeochemical models. *Glob. Biogeochem. Cycles* 19:GB4021. doi: 10.1029/2004GB002427

LaCasce, J. H. (2008). Statistics from lagrangian observations. *Prog. Oceanogr.* 77, 1–29. doi: 10.1016/j.pocean.2008.02.002

Landry, M. R., Ohman, M. D., Goericke, R., Stukel, M. R., and Tsyrklevich, K. (2009). Lagrangian studies of phytoplankton growth and grazing relationships in a coastal upwelling ecosystem off Southern California. *Prog. Oceanogr.* 83, 208–216. doi: 10.1016/j.pocean.2009.07.026

Lathuilière, C., Echevin, V., Lévy, M., and Madec, G. (2010). On the role of the mesoscale circulation on an idealized coastal upwelling ecosystem. *J. Geophys. Res.* 115, C09018. doi: 10.1029/2009JC005827

Lee, M. M., Marshall, D. P., and Williams, R. G. (1997). On the eddy transfer of tracers: advective or diffusive? *J. Mar. Res.* 55, 483–505. doi: 10.1357/0022240973224346

Lehahn, Y., D'Ovidio, F., Lévy, M., and Heifetz, E. (2007). Stirring of the Northeast Atlantic spring bloom: a lagrangian analysis based on multisatellite data. *J. Geophys. Res.* 112, C08005. doi: 10.1029/2006JC003927

Lett, C., Penven, P., Ayón, P., and Fréon, P. (2007). Enrichment, concentration and retention processes in relation to anchovy (*Engraulis Ringens*) eggs and larvae distributions in the northern humboldt upwelling ecosystem. *J. Mar. Syst.* 64, 189–200. doi: 10.1016/j.jmarsys.2006.03.012

Lett, C., Verley, P., Mullon, C., Parada, C., Brochier, T., Penven, P., et al. (2008). A lagrangian tool for modelling ichthyoplankton dynamics. *Environ. Model. Softw.* 23, 1210–1214. doi: 10.1016/j.envsoft.2008.02.005

Lévy, M. (2008). "The modulation of biological production by oceanic mesoscale turbulence," in *Transport and Mixing in Geophysical Flows* (Berlin; Heidelberg: Springer), 219–261.

Lévy, M., Resplandy, L., Klein, P., Capet, X., Iovino, D., and Ethé, C. (2012). Grid degradation of submesoscale resolving ocean models: benefits for offline passive tracer transport. *Ocean Model.* 48, 1–9. doi: 10.1016/j.ocemod.2012.02.004

Lih, M. M. S. (1975). *Transport Phenomena in Medicine and Biology (Biomedical Engineering and Health Systems), 1st Edn*. New York, NY: John Wiley & Sons Inc.

Longhurst, A. R. (1995). Seasonal cycles of pelagic production and consumption. *Prog. Oceanogr.* 36, 77–167. doi: 10.1016/0079-6611(95)00015-1

Luo, J., and Jameson, L. (2002). A wavelet-based technique for identifying, labeling, and tracking of ocean eddies. *J. Atmos. Oceanic Technol.* 19, 381–390. doi: 10.1175/1520-0426-19.3.381

Madec, G. (2008). *NEMO Ocean Engine*. France, Institut Pierre-Simon Laplace (IPSL), 300. (Note du Pole de Modélisation 27)

McGillicuddy, D. J., Johnson, R., Siegel, D. A., Michaels, A. F., Bates, N. R., and Knap, A. H. (1999). Mesoscale variations of biogeochemical properties in the Sargasso Sea. *J. Geophys. Res.* 104, 381–394. doi: 10.1029/1999jc900021

McGillicuddy, D. J., Kosnyrev, V. K., Ryan, J. P., and Yoder, J. A. (2001). Covariation of mesoscale ocean color and sea-surface temperature patterns in the Sargasso Sea. *Deep Sea Res. Part II.* 48, 1823–1836. doi: 10.1016/S0967-0645(00)00164-8

McGillicuddy, D. J., and Robinson, A. R. (1997). Eddy-induced nutrient supply and new production in the sargasso sea. *Deep Sea Res. Part I.* 44, 1427–1450. doi: 10.1016/S0967-0637(97)00024-1

McGillicuddy, D. J., Robinson, A. R., Siegel, D. A., Jannasch, H. W., Johnson, R., Dickey, T. D., et al. (1998). Influence ofmesoscale eddies on newproduction in the Sargasso Sea. *Nature* 394, 263–266. doi: 10.1038/28367

McWilliams, J. C., Weiss, J. B., and Yavneh, I. (1999). The vortices of homogeneous geostrophic turbulence. *J. Fluid Mech.* 401, 1–26. doi: 10.1017/S0022112099006382

Merlivat, L., Boutin, J., and Antoine, D. (2014). Roles of biological and physical processes in driving seasonal air–sea CO_2 flux in the southern ocean: new insights from CARIOCA pCO_2. *J. Mar. Syst.* 147, 9–20. doi: 10.1016/j.jmarsys.2014.04.015

Morales, C. E., Hormazabal, S., Correa-Ramirez, M., Pizarro, O., Silva, N., Fernandez, C., et al. (2012). Mesoscale variability and nutrient–phytoplankton distributions off central-southern chile during the upwelling season: the influence of mesoscale eddies. *Prog. Oceanogr.* 104, 17–29. doi: 10.1016/j.pocean.2012.04.015

Morrow, R., Birol, F., Griffin, D., and Sudre, J. (2004). Divergent pathways of cyclonic and anti-cyclonic ocean eddies. *Geophys. Res. Lett.* 31, L24311. doi: 10.1029/2004gl020974

Nencioli, F., Dong, C., Dickey, T., Washburn, L., and McWilliams, J. C. (2010). A vector geometry–based eddy detection algorithm and its application to a high-resolution numerical model product and high-frequency radar surface velocities in the Southern California bight. *J. Atmos. Oceanic Technol.* 27, 564–579. doi: 10.1175/2009JTECHO725.1

Nencioli, F., Kuwahara, V. S., Dickey, T. D., Rii, Y. M., and Bidigare, R. R. (2008). Physical dynamics and biological implications of a mesoscale eddy in the lee of Hawai'i: cyclone opal observations during E-Flux III. *Deep Sea Res. Part II* 55, 1252–1274. doi: 10.1016/j.dsr2.2008.02.003

Owen, W. B. (1991). A statistical description of the mean circulation and eddy variability in the northwestern Atlantic using SOFAR floats. *Prog. Oceanogr.* 28, 257–303. doi: 10.1016/0079-6611(91)90010-J

Pous, S., Feunteun, E., and Ellien, C. (2010). Investigation of tropical eel spawning area in the South-Western Indian Ocean: influence of the oceanic circulation. *Prog. Oceanogr.* 86, 396–413. doi: 10.1016/j.pocean.2010.06.002

Qiu, Z. F., Doglioli, A. M., Hu, Z. Y., Marsaleix, P., and Carlotti, F. (2010). The influence of hydrodynamic processes on zooplankton transport and distributions in the North Western Mediterranean: estimates from a lagrangian model. *Ecol. Modell.* 221, 2816–2827. doi: 10.1016/j.ecolmodel.2010.07.025

Ramp, S. R., Jessen, P. F., Brink, K. H., Niiler, P. P., Daggett, F. L., and Best, J. S. (1991). The physical structure of cold filaments near point Arena, California, during June 1987. *J. Geophys. Res.* 96, 14859–14883. doi: 10.1029/91JC01141

Richardson, P. L. (1993). A census of eddies observed in North Atlantic SOFAR float data. *Prog. Oceanogr.* 31, 1–50. doi: 10.1016/0079-6611(93)90022-6

Sadarjoen, I. A., and Post, F. H. (2000). Detection, quantification, and tracking of vortices using streamline geometry. *Comput. Graph.* 24, 333–341. doi: 10.1016/S0097-8493(00)00029-7

Sakai, S. (1973). A model for group structure and its behavior. *Biophys. Jpn.* 13, 82–90. doi: 10.2142/biophys.13.82

Sangrà, P., Pascual, A., Rodríguez-Santana, A., Machín, F., Mason, E., McWilliams, J. C., et al. (2009). The canary eddy corridor: a major pathway for long-lived eddies in the subtropical North Atlantic. *Deep Sea Res. Part I* 56, 2100–2114. doi: 10.1016/j.dsr.2009.08.008

Shchepetkin, A. F., and McWilliams, J. C. (2005). The regional oceanic modeling system (ROMS): a split-explicit, free-surface, topography-following-coordinate oceanic model. *Ocean Model.* 9, 347–404. doi: 10.1016/j.ocemod.2004.08.002

Souza, J. M. A. C., de Boyer Montégut, C., and Le Traon, P. Y. (2011). Comparison between three implementations of automatic identification algorithms for the quantification and characterization of mesoscale eddies in the South Atlantic Ocean. *Ocean Sci.* 7, 317–334. doi: 10.5194/os-7-317-2011

Stommel, H. (1963). Varieties of oceanographic experience. *Science* 139, 572–576. doi: 10.1126/science.139.3555.572

The Ring Group. (1981). Gulf stream cold-core rings: their physics, chemistry, and biology. *Science* 212, 1091–1100. doi: 10.1126/science.212.4499.1091

Wang, Y. H., Dai, C. F., and Chen, Y. Y. (2007). Physical and ecological processes of internal waves on an isolated reef ecosystem in the South China Sea. *Geophys. Res. Lett.* 34, L18609. doi: 10.1029/2007GL030658

Yamazaki, H., and Haury, L. R. (1993). A new lagrangian model to study animal aggregation. *Ecol. Modell.* 69, 99–111. doi: 10.1016/0304-3800(93)90051-S

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Environmental Science

HYPOTHESIS AND THEORY
published: 20 August 2015
doi: 10.3389/fenvs.2015.00058

Adaptive submodel selection in hybrid models

Randall Gray[1]* *and Simon Wotherspoon*[2]

[1] *Mathematics, University of Tasmania, Hobart, TAS, Australia,* [2] *Institute of Marine and Antarctic Studies, University of Tasmania, Hobart, TAS, Australia*

Hybrid modeling seeks to address problems associated with the representation of complex systems using "single-paradigm" models: where traditional models may represent an entire system as a cellular automaton, for example, the set of submodels within a hybrid model may mix representations as diverse as individual-based models of organisms, Markov chain models, fluid dynamics models of regional ocean currents, and coupled population dynamics models. In this context, hybrid modelers try to choose the best representations for each component of a model in order to maximize the utility of the model as a whole. Even with the flexibility afforded by the hybrid approach, the set of models constituting the whole system and the dynamics associated with interacting models may be most efficient only in parts of the global state space of the system. The immediate consequence of this possibility is that we should consider adaptive hybrid models whose submodels may change their representation based on their own state and the states of the other submodels within the system. This paper uses a simple example model of an artificial ecosystem to explore a hybrid model which may change the form of its component submodels in response to their local conditions and internal state relative to some putative optimization choices. The example demonstrates the assessment and actions of a "monitor" agent which adjusts the mix of submodels as the model run progresses. A simple mathematical structure is also described and used as the basis for a submodel selection strategy, and alternative approaches are briefly discussed.

Keywords: hybrid modeling, adaptive models, environmental modeling, cross-paradigm modeling, agent-based modeling, adaptive agents

OPEN ACCESS

Edited by:
Christian E. Vincenot, Kyoto University, Japan

Reviewed by:
Guennady Ougolnitsky, Southern Federal University, Russia
William John Chivers, The University of Newcastle, Australia

***Correspondence:**
Randall Gray, Mathematics, University of Tasmania, Sandy Bay Campus, Churchill Avenue, Sandy Bay, Hobart, TAS 7005, Australia
Randall.Gray@limnal.net

Specialty section:
This article was submitted to Environmental Informatics, a section of the journal Frontiers in Environmental Science

Received: *10 May 2015*
Accepted: *03 August 2015*
Published: *20 August 2015*

Citation:
Gray R and Wotherspoon S (2015) Adaptive submodel selection in hybrid models. Front. Environ. Sci. 3:58. doi: 10.3389/fenvs.2015.00058

1. Introduction

The case has been made for developing systems with submodels that change their representation according to their state. Vincenot et al. (2011) identify reference cases describing the major ways system dynamics models (*SD*) and individual-based models (*IB*) can be coupled. Their final case, *SD–IB* model swapping, is exemplified in the models described by Bobashev et al. (2007) and Gray and Wotherspoon (2012). These papers argue that we can improve on conventional hybrid models, in terms of efficiency, fidelity, model clarity, or execution speed by using an approach that allows the submodels themselves to change during a simulation. The last two papers implement simple models which demonstrate the approach, with correspondingly simple mechanisms to control transitions between different submodels.

Some authors argue that the explicit coupling of *SD* models and *IB* models may provide greater clarity and resolution in modeling (Fulton, 2010; Vincenot et al., 2011): parts of a model that are most clearly the result of aggregate processes are likely to be better suited

to modeling with a *SD* approach. In contrast, the parts of a system where individuals have a significant influence on their neighbors (Botkin et al., 1972) are better suited to an *IB* approach. This argument is closely tied to the notion of model fidelity. Following DelSole and Shukla (2010), we take *fidelity* to be the degree to which a model's trajectory is compatible with real trajectories. If our immediate goal is to maximize the utility of the set of submodels within a model as it runs, this must include the fidelity of the system in the decision process.

Measuring or estimating execution speed and numerical error are comparatively straight-forward, but determining model fidelity is not. Models with a high degree of fidelity should produce results which are consistent with observed data from real instances of the system they model across both a wide range of starting conditions and under the influence of *ad hoc* perturbations, such as fires through a forested domain. Model fidelity is addressed by DelSole and Shukla (2010) in the context of seasonal forecasting models. They explore the relationship between fidelity and skill using an information-theoretic approach. They describe *skill* loosely as the ability to reproduce actual trajectories, and they describe *fidelity* as measuring the difference between the distribution of model results and the distribution of real world results. They highlight the attractiveness of mutual information and relative entropy as measures (or at least indices) of skill and fidelity, but they observe that in their domain, climate modeling, the necessary probability distributions are unknown.

The issues of fidelity and the attendant cost/benefit balance are central to the discussion in Bailey and Kemple (1992). This paper assesses the costs and benefits of three different upgrades to an existing model designed to help determine the best mix of types of radios used in a military context; their objective is to prioritize implementation of the refinements of their model. The fundamental issues they address are substantially the same as issues that influence dynamic model selection.

The paper by Yip and Marlin (2004) compares three models used for real-time optimization of a boiler network: simple linear extrapolation from the system's current state, quadratic prediction with the coefficients based on historical data and updated at every step, and a detailed process model that corresponds closely with the physical elements of the modeled system. Their conclusion correlates the fidelity of the model with its ability to control the real-time optimization of the system. They explicitly note that there are real costs associated with the increased fidelity. These costs include model development and the need for more expensive sensors. They note that increasing fidelity in the model enabled the system to adapt to changing fuel more efficiently, and that when there were frequent changes in fuel characteristics the simpler models performed poorly.

The projects described in Little et al. (2006) and Fulton et al. (2011) both used hybrid models as a means of decreasing the run-time, and increasing the fidelity of the modeled contaminant uptake in simulated organisms. This was accomplished by mixing individual-based submodels and regional population-based systems models. Gray and Wotherspoon (2012) explicitly used changes in the representation of agents to improve the execution speed of a contamination tracking model, without losing the fidelity of the individual based uptake model. In this paper, we will develop a more general strategy which may be appropriate for more complex systems.

2. Model Organization

For clarity, we will take the term *niche* to refer to something in the model which could be modeled in several ways: a "porpoise" niche could be filled by many instances of an individual-based model, models of pods, or a regional *SD* model of the porpoises. This is essentially the same as the term *component* in Vincenot et al. (2011). The motivation for departing from this convention arose from confusion resulting from inadvertently using *component* both in a technical and non-technical sense. The close analogy between the nature of a niche in an ecosystem and the nature of a component as discussed in Vincenot et al. (2011) suggested the choice of *niche*.

Each of the alternative ways of representing a niche can be viewed as a *submodel*, and the word *representation* will be used to reflect a particular choice of submodel within a niche. An explicit instance of a submodel (such as a specific pod or an *SD* model) will be referred to as an *agent*. The *configuration* of the model at any moment consists of the particular set of submodels which fill the niches that comprise the model as a whole. For an adaptive hybrid model, there may be a large number of possible configurations and the selection of a "best" configuration is a complex matter.

Each agent running in a model must necessarily have data which can serve to characterize it for these assessments. This data would typically be some subset of its state variables, but the data alone may not be enough to base an assessment on: there may also be extrinsic data which play a role in a particular submodel's or agent's activity and impinges on its suitability. Then, the characterization of an agent—its *state vector*—is an amalgam of its own state and the state of other niches it interacts with, and it can be regarded as a point in the state space which the submodel is defined over.

A corresponding set of data characterizes a niche in the model; here, it is typically some appropriate aggregation of agent-level state variables (a biomass-by-size distribution, for example), relative rankings of the suitability of agents and alternative submodels, and indications of what extrinsic support all of the various alternatives require. This niche-level state vector provides the data needed for optimizing the configuration globally, and for managing the configuration when niche-wide effects become significant, for example, for an incipient epidemic.

Thus, there are three distinct levels of organization which may influence the considerations regarding the current configuration, and inform any decision about what may need to change, namely

1. Agent-level data need to be examined to determine how well-suited each agent is to its current state and the context provided by the agents it interacts with,
2. A niche-level assessment which compares the utility of each of its current agents within a niche with their alternative submodels, and

3. A model-wide assessment which determines whether there are cross-agent conflicts or unmet needs arising from a particular configuration.

The state vectors which form the domains of submodels and niches are loci in appropriate state spaces and can be encoded as an elements in appropriate vector spaces. The mathematical tools to manipulate these state vectors can then be applied to calculate the distances between two states, the similarity of loci which represent models or niches, or to identify trends or clusters.

2.1. Implications of Changing Configurations

At a basic level, hybrid models are designed to represent entities or processes in the real world in a way which brings more clarity, efficiency, or fidelity that may be possible with more traditional approaches. Adaptive hybrid models, implicitly acknowledge that the appropriate representation may change through time. An important consequence is that when a submodel in a niche changes, it may trigger changes in representation elsewhere in the model.

We might consider an example where an *SD* submodel which represents the prey for an *SD* based predator changes to *IB* submodels. It seems reasonable to expect the representation of the predator might follow suit. This may change the spatial resolution, the fineness of the "quantities" represented, and possibly the time steps associated with the predators and prey. Disparities in either of the first two are simple enough to deal with: modelers routinely use interpolation as a means of removing inappropriate edges, or generating subscale data, for example. Changes in an agent's time step can have a dramatic causal influence on the subsequent simulation.

Chivers (2009) discusses how individual-based models are sensitive to when state variables are updated. In his discussion, the issue arises as a result of when the probability of a predator–prey interaction is calculated relative to when the prey are removed from the system, though similar effects are also likely to occur in other contexts. The temporal sensitivity of submodels' interactions needs careful examination in order to construct submodels that proceed through time coherently and interact correctly.

Multi-agent models must have strategies to manage the agents as they step from the start of the simulation to its end. The simplest method is to make everything within the model use the shortest time step required. This is computationally inefficient in a heterogeneous model.

A better approach is the technique of variable speed splitting, such as in Walters and Martell (2004) and many others. (**Figure 1**) This approach allows models to step through time in different intervals by dividing the largest interval required into smaller steps that are more appropriate for the submodels with naturally shorter time scales. While models with uniform time steps are a trivial example of this approach, variable speed splitting is almost as simple and much more efficient. This technique can keep the subjective times of a set of agents moderately consistent, but *ad hoc* stepping changes would still seem to be awkward or difficult.

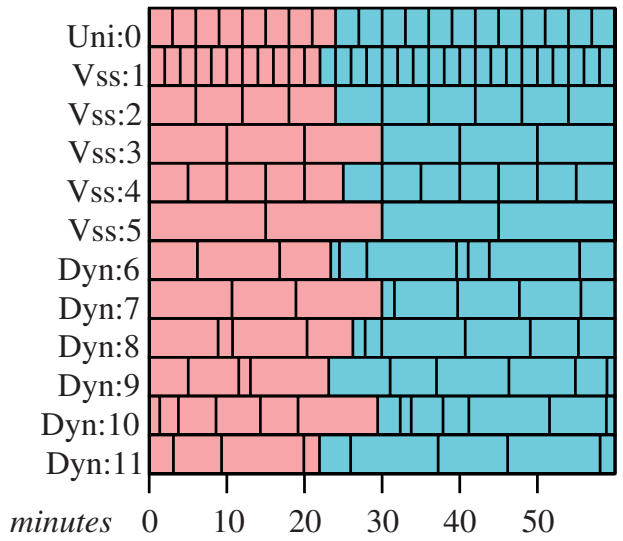


FIGURE 1 | Time scheduling strategies. Red boxes represent time steps that have already passed, blue boxes represents scheduled time steps that have not yet been run. "Uni:" and "Vss:" submodels are members of a uniform or variable speed splitting submodels and require uniform time steps, and "Dyn:" submodels have adaptive time steps.

Both of these strategies may be subject to artifacts arising from the sequence in which agents are given their time step. The general class of model errors of the sort described in Chivers (2009) arise as a consequence of structure of the processing across the set of agents in a simulation. *IB* models which process agents species-by-species will be particularly vulnerable to these sorts of artifacts, since there will be an implicit advantage or disadvantage to being early in the list. Similarly, advantage or disadvantage can arise when there is a change in representation, perhaps from an *SD* submodel to an *IB* submodel; a shorter time step in this situation may introduce a great many small time steps which agents may exploit. This kind of problem can be overcome by introducing a randomizing process within each time step. Early versions of the variable speed splitting model in Lyne et al. (1994) suffered from predator–prey artifacts arising from a naïve introduction of predators and prey into the list of agents, and such randomizing was introduced to minimize the effects. In situations where the time steps of the interacting agents differ, implementing a randomization strategy may require a significant increase in the complexity of the system to accommodate irregular stepping through the lists of agents, or a significant change in the basic structure of the model.

Gray et al. (2006) and Fulton et al. (2009) describe models that have a well-developed approach to coordinating agents using adaptive time steps. In these models agents may set their own time steps to intervals that are suitable for their current activity or role. This strategy can readily incorporate submodels with uniform time steps, or collections that employ a variable speed splitting strategy. When agents interact, they either explicitly become synchronous before interaction occurs by setting their time steps appropriately and waiting, or they

implicitly acknowledge that there is a temporal mismatch (**Figure 1**).

While some agents should be given execution priority (such as an agent which models ocean currents), most agents will have their execution order within a time step randomized, effectively preventing a large class of execution order dependent artifacts. The associated overhead in the most recent work, (Gray et al., 2006; Gray and Wotherspoon, 2012), is marginally higher than one would expect from single-stepping or variable speed stepping systems, but the advantages arising from the ability to ensure synchrony and change time steps in response to environmental stimulus outweigh the small computational overhead. This last approach seems likely to be the most appropriate for a general hybrid model that supports swapping models.

General adaptive hybrid models must have a mechanism for scheduling each agent's execution which keeps the cohort of agents roughly synchronous, and it should able to handle changes in an agent's time step when the agent changes its representation; where possible, agents should also be designed so that they may run at other time steps as well as their own preferred time step so they can become synchronous and interact at the appropriate temporal scale with other agents.

2.2. Systematically Adjusting the Model Configuration

A model's configuration should only change when there is an overall benefit in the efficiency or fidelity of the system. A straightforward way of determining this is to have a monitoring routine that runs periodically, polling the agents, and ranking likely configurations according to their relative benefit or cost. This means that each submodel would need a way to provide, to the monitor, a measure of its current suitability, and to indicate what it needs from other niches.

The last step in Algorithm 1 is deliberately vague.

Algorithm 1 illustrates a possible assessment pass for a monitor, though how appropriate it may be is an open question. Configuration ranking for the example model will be cast in terms of evaluating an objective function based on elements of the vector space of tree elements described in the Supplementary Material.

A monitor may have large number of potential candidate configurations, but we would like to keep the actual number quite low. The example model described below has a global domain associated with a particular representation, along with local domains (subregions of the global domain) which are associated with finer scale representations of the modeled entities. The set of potential candidate trees could be quite large; in practice we reduce the number by casting the candidate trees in a more general way—including trees representing particularly good representations and particularly poor representations: the first to steer the configuration toward good choices, and the second to drive it away from poor choices. We can use the hierarchical organization (whole-model, niche, submodel, agent) to help limit our search space, as well as the geographic context of the agents (whole-domain, local cell, immediate-locus).

The sets of candidate trees which are associated with particular configurations will need to be crafted carefully as a part of the model design. These trees reflect the modelers understanding of the strengths and weaknesses of each of the submodels (or sets of different submodels) which may be employed.

Algorithm 1 | Basic processing pass for the monitor

```
for all niches do
    for all submodels in the niche do
        for all agents in the submodel do
            generate agent state vector
            generate the submodel state vector
                note extrinsic requirements
        end for
    end for

    generate niche state vector
end for

Run niche-level assessment
Flag any whole of model issues
for all candidate configurations do
    Deprecate untenable configuration
    Adjust for unavoidable extrinsic
        requirements
end for

Select best indicated configuration
```

Exactly how a monitoring routine is integrated into the model framework is a subjective choice best left to the team implementing the models, but one very attractive option is to implement the monitor as an agent in the system. This would allow the monitor to assess its own performance and the needs of other agents with respect to its own suitability with the option of swapping itself for a monitor which implements some alternative strategy.

3. The Example Model

The purpose of the example model described below, is to provide a context for a discussion of the dynamics associated with a hypothetical simulation using this model. The ends of the spectrum between *SD* models and *IB* models are represented, and the environment is unrealistically simple in order to keep us from being swamped by detail.

The model consists of a spatially explicit environment that is partitioned into nine cells (**Figure 2**). The biotic elements consist of plants, fruit, seeds, herbivores, and carnivores. The herbivores feed on the plants and their fruit; and carnivores prey upon juvenile herbivores. The plants and herbivores are interdependent: fruit is the sole diet for juvenile herbivores and the plants need juvenile herbivores to make the seeds viable by eating the fruit.

The representations are equation-based *SD* models of the interactions between the plants and animals and *IB* models for plants and animals. The *SD* submodels model the biomass with respect to size, for plants and animals, or simply numeric

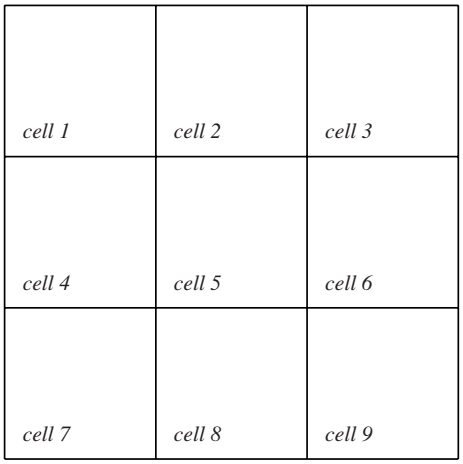


FIGURE 2 | The model domain is divided into nine cells. An *SD* agent is associated with each of these cells and with the domain as a whole. Any *IB* agents which are created during the simulation will be associated with one cell at any given time.

quantities for fruit and seeds, and they can operate at either the global or cell-sized scale. Modeling biomass in this way makes it possible to minimize the loss of fidelity incurred by swapping from *IB* agents to *SD* agents and visa-versa, since we preserve more of the essential nature of the populations. A more detailed description of the *SD* agents is presented in the Supplementary Material.

Fruit and Seeds

Fruit and seeds are treated somewhat differently to the rest of the niches. They exist principally as numbers of entities that are updated as a result of the activities of other, more explicit *SD* or *IB* models. There are explicit routines that deal with uniquely "fruit" and "seed" processing to handle spoilage and germination, respectively.

For fruit and seeds we have the following relationships

$$dN_{\mathsf{F}}(t) = Production - Spoilage - FruitEaten$$

and

$$dN_{\mathsf{S}}(t) = s * FruitEaten - \left(1 - \frac{N_{\mathsf{P}}(t)}{K_{\mathsf{P}}}\right) Germ.$$

where $N_{\mathsf{P}}(t)$ is the biomass of plants at time t, and K_{P} is the carrying capacity of the pertinent domain (either global or cell-based). The processing for fruit is quite simple and consists only of applying "spoilage"; no reference to other agents in the system is required, and only the number of fruit is adjusted as a result (Algorithm 2). Seed models will adjust their "seed count" as well as the biomass distribution for plants in their time step, according to the level of germination. Germination is probabilistic as is the size of the plant a germinated seed becomes in its pass, though the distribution of possibly sizes is quite restrained (Algorithm 3).

Algorithm 2 | Basic processing pass for fruit

$N_{\mathsf{F}} \leftarrow N_{\mathsf{F}} - (Spoilage_{\mathsf{F}} \cdot N_{\mathsf{F}})$

Algorithm 3 | Basic processing pass for seeds

$NewTreeCount \leftarrow Germination \cdot SeedCount$
$SeedCount \leftarrow SeedCount - (NewTreeCount + Spoilage_{\mathsf{S}} \cdot SeedCount)$
generate *NewTreeCount* new plant agents and introduce them into the system

SD Representations

Each of the niches has an integral equation expressing the change in biomass for a given size; an animal's equation is of the form[1]

$$dN_{\mathsf{A}}(t,x) = Growth\,\&\,starve + Repr - PredMort - NatMort.$$

We do not include migration terms in the *SD* models, since that will be addressed by the *IB* forms. The assumption is that the *SD* representation is most appropriate when population levels are moderately high, and there is adequate food; under these conditions, we will assume that the net migration associated with a domain will be close to zero.

Plants are represented by similar equations, namely

$$dN_{\mathsf{P}}(t,x) = \left(1 - \frac{N_{\mathsf{P}}(t)}{K_{\mathsf{P}}}\right)[Growth + Germ] - PredMort - NatMort$$

where $N_{\mathsf{P}}(t,x)$ is the biomass of plants of size x at time t.

The important state variables for the *SD* are, for each domain, the biomass-by-size distributions for plants, herbivores, and carnivores, and the raw numbers of fruit and viable seeds.

The system of equations described in the Supplementary Material is evaluated using a fourth order Runge–Kutta algorithm; the numbers of fruit and seeds, and both the global and cell-based biomass distributions for plants and animals are updated at the end of the calculation. The model will adjust the values in the global and cell-based models to allow data from models running with better resolution (usually more localized models) (Algorithm 4) to take precedence.

Most of the important parameters and many of the functions associated with the life history of the modeled entities are not specified. This way we may consider possible trajectories without being tied to a particular conception or parameterization of the system.

IB Representations

Individual-based representations for plants, herbivores, and carnivores follow the pattern in Little et al. (2006); fruit and seeds are only modeled in the *SD* representation, though their numbers are modified by the activities of the herbivores irrespective of how those herbivores are represented.

[1] See the Supplementary Material for a more detailed set of equations.

Algorithm 4 | Basic processing pass for the *SD* models

```
for all agents in this domain do
    Incorporate quantities that are
        controlled in other agents
    Run Runge-Kutta4
    Update only quantities that are
        controlled by this agent
end for
```

3.1. *IB* Plants

Plants maintain a reference to their cell, their location, a mass, and a peak mass. If a plant's mass drops below a certain proportion ($\mathsf{P}_{M\Omega}$) of its peak mass, it dies—this provides a means for the herbivores to drive the plant population to local extinction.

We will suppose that plants grow according to a sigmoidal function with some reasonable asymptote and intermediate sharpness; fruiting occurs probabilistically as in the *SD* representation.

The plant agent goes through the steps in Algorithm 5 in each of its time steps. In the algorithm, $\Gamma_{\mathsf{P}}(\delta t, mass)$ is an analog of the probability of a plant growing from one size to another from the *SD* representation, P_{Mature} is the parameter that indicates the mass a plant must be before it fruits, P_{Fruits} is the probability of a mature plant fruiting, and P_{ρ} is the amount of fruit relative to the fruiting area. The routine ADDFRUIT updates the models representing fruit in the domain.

Algorithm 5 | Basic processing pass for plants

```
if (Mass ≥ P_Mature) ∧ (P_Fruits ≥ rnd_0,1) then
    AddFruit(P_ρ Mass^(2/3))
end if
if (Mass ≤ P_MΩ PkMass) ∨ (Ω_indP < rnd_0,1) then
    Die
else
    Mass ← Γ_P(δt, Mass)
    if Mass > PkMass then
        PkMass ← Mass
    end if
end if
```

3.2. *IB* Animals

Like the plants, animals maintain a reference to their cell, their location, and a mass. They also maintain several variables that are associated with foraging or predation, namely the amount of time until they need to eat (*Sated*), and the amount of time they have been hungry (*Hungry*).

Animals will grow while they do not need to eat and will only forage when they are hungry. Reproduction happens in a purely probabilistic way once the animal is large enough, and the young are not cared for by the parents.

Animal movement is constrained so that they will tend to stay within their nominated home cell, only migrating (changing their home cell to an adjacent cell) when food becomes scarce or if the population exceeds some nominated value and causes crowding.

The analogs of the mechanisms for growth and starvation in the *SD* representation are quite different to those of the *IB* version. In the *SD* models, starvation and growth occur as a result of the relative population levels of the consumer and the consumed rather than the local availability of food.

There are no real programmatic differences between the *IB* representations of herbivores and carnivores; their differences lie in their choices of food and the way their "time-to-eat" variable is initially managed. Individual-based, new-born carnivores begin with a long time till they need to eat. This reflects a reliance on some unmodeled foodstuff until they are large enough to prey on the juvenile herbivores. In contrast, the juvenile herbivores must begin eating fruit immediately, and only switch to foraging on plants when they are larger (but before they can reproduce). For both species, if the amount of time they have been hungry exceeds a particular value, H_{Ω} or C_{Ω}, the individual dies.

So, if we take A to represent either carnivores (C) or herbivores (H) below, then the processing pass for an animal is shown in Algorithm 6, where A_{moveT} is the amount of time an animal can be hungry before it migrates, A_{Ω} is the amount of time it takes for the animal to starve, $\mathsf{A}_{EatLimit}$ is the most the animal can eat as a proportion of its mass, $\mathsf{A}_{RepSize}$ is the minimum size an animal may breed at and A_{RepP} is the probability of reproducing. The routines $\text{PREYPRESENT}_{\mathsf{H}}$ and EAT_{H} have different cases for juvenile and adult herbivores, since juveniles prey upon fruit, and the seeds from the fruit they eat need to be accounted for in the appropriate places. There is a similar issue with juvenile carnivores. Their *preylist* will always be set to a value that indicates that they may eat as much as they like, and the corresponding call to EAT_{C} will handle this value appropriately.

3.3. The Monitor and Model Dynamics

The following may be typical of the types of situations that could or should cause changes in the configuration:

- *Low population*—If, in an *SD* representation, the number of individuals filling a niche (either explicitly taken from a distribution, or estimated using a mean and a biomass) drops below a nominated value, then the biomass in that niche should be converted to *IB* agents representing those individuals. This type of change is motivated by the observation that at low population levels the assumption that we can treat the population as having uniform access to resources (or be uniformly available to predators) breaks down;
- *High population*—If a niche in a cell is represented by *IB* agents and the number of individuals exceeds a (higher) nominated value, the biomass those agents represent should be subsumed by the distribution in the local *SD* submodel. The change in representation is attractive here for two reasons: an equation-based representation will be much faster, and *SD* submodels are arguably simpler to calibrate;
- *Starvation risk*—If the mean amount of time an animal in a cell spends *hungry* in a cell exceeds half of A_{ω} (or some other nominated time), the prey biomass must convert to *IB* agents if it isn't already so (bearing in mind that this isn't pertinent for fruit). This mean is calculated by averaging the means of each animal in the cell. If this is triggered, it indicates that the

Algorithm 6 | Basic processing pass for herbivores and carnivores

if $(\Omega_{\text{ind}\mathsf{A}} > \text{rnd}_{0,1}) \vee (Hungry \geq \mathsf{A}_{\Omega})$ **then**
  DIE
end if
PreyList ← PREYPRESENT$_{\mathsf{A}}$(*Locus*, *Mass*)
if *Sated* ≥ 0 **then**
  Mass ← *Mass* + GROWTH$_{\mathsf{A}}$(*mass*, δt)
else if (*Hungry* ≥ 0) ∧ (len(*PreyList*) > 0) **then**
  Sated ← EAT(*PreyList*, $\mathsf{A}_{EatLimit}$, *mass*)
  Hungry ← 0
  ForageCt ← 0
else if (*Hungry* ≥ 0) ∧ *len*(*PreyList*) = 0) **then**
  FORAGE
  ForageCt ← *ForageCt* + 1
else if (*Hungry* ≥ A_{moveT}) ∨ CROWDED$_{\mathsf{A}}$ **then**
  MIGRATE$_{\mathsf{A}}$(*Locus*)
else
  if (*mass* ≥ $\mathsf{A}_{RepSize}$) ∧ (A_{RepP} ≥ $\text{rnd}_{0,1}$) **then** REPRODUCE$_{\mathsf{A}}$(*Locus*)
  end if
end if

biomass of the prey species is sparse enough that homogeneity assumption is unlikely to hold;

- *Relative biomass*—If the biomass available for predation is represented in a local *SD* agent and its density drops below some proportion of the minimum required to support the predators in the domain, the prey species should convert its biomass into *IB* agents and, if the predator is represented by a *SD* agent, it should also convert to an *IB* form. If the biomasses are such that the effective predation rate is unsustainable, the mixing assumption is unlikely to hold.

The pertinent data for conditions will be periodically reported to the monitor through a set of status trees. The trees are able to represent single entities, nested entities, and aggregates equally well, and can preserve structural information which may also be used in the comparison of these trees. One of the basic elements we can easily incorporate into a submodel's status tree is the agent's own assessment of its competence relative to its state-vector and its local conditions. This measure of "self-confidence" can probably be maintained at little computational cost for most agents, and may be the most significant component in a monitor's assessment. The *high* and *low* population level conditions can clearly be determined by the agent in question; it can set its level of self-confidence upward or downward as appropriate. *Starvation* can also be encoded in the relevant node of an agent's status tree, but since starvation alone may not indicate a problem with the way the entity is represented, it probably wouldn't reduce the value for its confidence.

A starvation trigger may usually arise as a natural consequence of the population dynamics, but it may also occur when there is a mismatch in representations which has not been adequately addressed in the design stage. The final condition based on the relative biomasses is one which properly lies in the realm of the monitor—it would be quite inefficient for each of the candidate animals to be querying their prey for available biomass, summing the result, and then noting the need for change.

The monitor will primarily use the confidence values associated with agents and their niches, and the distance from trees which describe the state of the model or its set of submodels to trees which describe "known good" configurations. With data obtained directly from the agents in the system and from alternative representations it generates status trees,

- $\check{\boldsymbol{\tau}}^{\Sigma}_{sn}$, is a candidate status tree tied to a specific configuration. The serial number, *sn*, ties it to a configuration with that serial number,
- $\check{\boldsymbol{\tau}}^{\Sigma}_{d}$, is a candidate tree which represents the current state of a domain,
  $\boldsymbol{\tau}^{\Sigma}_{t}$, an aggregate tree for the whole domain at time *t*,
  $\boldsymbol{\tau}^{\Sigma}_{SD(n),t}$, aggregate trees for each cell, $n \in \{1, \ldots, 9\}$,
  $\boldsymbol{\tau}_{\text{R}(i),t}$, specific status trees for each agent,
  $\boldsymbol{\tau}_{R,t}$, specific status trees for a representation *R* for each representation associated with a niche,
  and
  $\hat{\boldsymbol{\tau}}_{\text{R}(i),t}$, candidate trees for all possible representations of each agent *i*,

at the beginning of each of its steps. The model may have a mix of *SD* and *IB* representations, and some of the trees will have to incorporate data from many agents ($\boldsymbol{\tau}^{\Sigma}_{t}$, any of the $\hat{\boldsymbol{\tau}}_{\text{R}(i),t}$, and $\boldsymbol{\tau}_{R,t}$, for example). A candidate tree is a status tree which represents an alternative submodel in a niche, and candidate trees are generated for specific agents and for each niche. When the monitor begins to generate status or candidate trees for a given agent, it first looks to see if it has generated an appropriate tree already. If it finds one, it incorporates or adjusts the tree appropriately; perhaps by incorporating the agent's biomass and size into the tree's data. We will also denote the configuration of a domain (global or local) with $\check{\boldsymbol{\tau}}^{\Sigma}_{c}$ where *c* identifies the domain in question.

The monitor assesses the trees by calculating aggregate values of particular attributes, comparing the trees' divergences from allegedly ideal configurations, and by looking how uniform groups are – groups of individuals that are all very similar are good candidates for simpler representations.

We can calculate the average confidence value from any of these trees by evaluating

$$\frac{\langle \overline{\text{mask}}(\boldsymbol{\tau}, \textit{confidence}, 0) \rangle}{\text{supp}(\overline{\text{mask}}(\boldsymbol{\tau}, \textit{confidence}, 0))},$$

for example. The trees and functions to manipulate them are described in the Appendix(Supplementary Material).

Now let us consider what a simulation might look like. **Figure 3** provides an overview of the configuration of the system as our hypothetical simulation runs. The model begins with eleven agents (not counting the monitor). The monitor runs its first step generating the status trees: $\boldsymbol{\tau}^{\Sigma}_{0}$, which characterizes the model in aggregate, $\boldsymbol{\tau}^{\Sigma}_{SD(0),0}, \ldots, \boldsymbol{\tau}^{\Sigma}_{SD(9),0}$, which record the aggregate state of the 10 *SD* submodels, the aggregate status tree for the *IB* agent, $\boldsymbol{\tau}^{\Sigma}_{IB(0),[}9]$, status trees for the *SD* submodels:

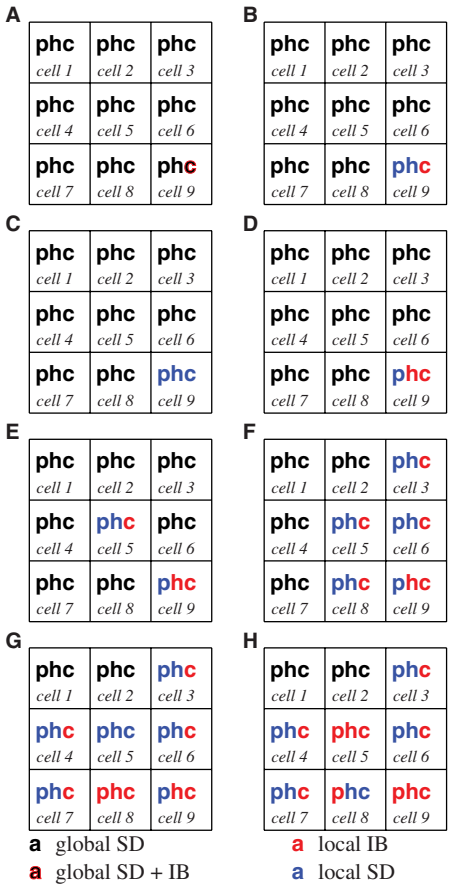


FIGURE 3 | The color of the p,h, and c indicate an agent's current representation within a cell at various points in the description of a simulation. In each, a black symbol indicates that the biomass of plants (p), herbivores (h), or carnivores (c) is modeled with the global *SD* agent, a blue symbol indicates that the biomass is modeled with a cell's *SD* agent, and red indicates that an *IB* model is being used. Symbols composed of two colors indicate that more than one representation is currently controlling portions of the relevant biomass. The initial configuration of the model is represented in **(A)**. Each of the subsequent panels **(B–H)** represent snapshots of the changing configuration during the simulation.

$\boldsymbol{\tau}_{\mathrm{SD}(0),0}$–$\boldsymbol{\tau}_{\mathrm{SD}(10),0}$, the status tree for the lone carnivore, $\boldsymbol{\tau}_{\mathrm{IB}(11),0}$, followed by the trees which represent alternative agents: $\hat{\boldsymbol{\tau}}_{\mathrm{SD}(0),0}$–$\hat{\boldsymbol{\tau}}_{\mathrm{SD}(10),0}$ and $\hat{\boldsymbol{\tau}}_{\mathrm{IB}(11),0}$. As mentioned earlier, there is only the single tree for agent 11 (the carnivore) since its alternative representation is embodied in $\hat{\boldsymbol{\tau}}_{\mathrm{SD}(10),0}$. During the simulation a simulated fire will occur.

The first steps which must be taken before ranking of potential configurations is to find the sets of candidate trees which best approximate the current configuration at both the global and cell levels. We do this by calculating a similarity index or a distance which indicates how close each of the candidate trees are to the configuration of each of the domains. There are many ways we could do this: for an index which only considers structural similarity we might use something like the simple function

$$\mathrm{ssim}(c, \boldsymbol{\tau}_d) = \frac{\mathrm{overlap}(c, \boldsymbol{\tau}_d)}{\max(\|c\|_{\mathsf{T}}, \|\tau_d\|_{\mathsf{T}})},$$

but for a more comprehensive treatment which factors values which are incorporated into the candidate and status trees we might apply the $\Delta(,)$ or dist functions described in the Supplementary Material. The dist function is a well-defined distance over the vector space of trees, while the $\Delta(,)$ function is an index of similarity that incorporates structural characteristics as well as the numerical distance between compatible subtrees. To refine such an analysis we could apply mask and mask to select only the relevant parts of the candidate and status trees.

So to assess the configuration of a domain, we would use our chosen measure to construct a set of the results of applying an optimization function, *opt*, to each of the candidate trees and their similarity to the current configuration. So if $\boldsymbol{S}$ is the set of all serial numbers for candidates, $\check{\boldsymbol{\tau}}_d^{\Sigma}$ is the status tree fo the current domain, and, we calculate

$$(C) = \{(\delta(\check{\boldsymbol{\tau}}_d^{\Sigma}, \check{\boldsymbol{\tau}}_i^{\Sigma}), i) : \forall i \in S\},$$

and this is used to generate

$$C^* = \{(\mathrm{opt}(\check{\boldsymbol{\tau}}_i^{\Sigma}), c, \check{\boldsymbol{\tau}}_i^{\Sigma}, i) : \forall (c, i) \in C\}$$

where δ stands for our chosen measure of similarity.

The elements in C^* are then assessed by the monitor, and the best permissible candidate is selected. If there is only a small improvement on the current configuration, $\check{\boldsymbol{\tau}}_d^{\Sigma}$, the monitor will leave the configuration as it is; otherwise, the monitor would then manage the creation of new agents to replace less optimal representations and manage the exchange of state data.

So the early phase of our simulation might begin like so:

1. Both of the aggregate trees $\boldsymbol{\tau}_0^{\Sigma}$ and $\boldsymbol{\tau}_{SD(9),0}^{\Sigma}$ indicate that there is an *IB* agent in their domain and that their *SD* representation does not perform well for the indicated biomass. Both the status and candidate trees for agent 11, $\boldsymbol{\tau}_{11(0),,}\boldsymbol{\tau}_{IB(11),0}$ and $\hat{\boldsymbol{\tau}}_{\mathrm{IB}(11),0}$, indicate that it is confident that it can represent the biomass, and that there are no immediate unmet requirements from other agents. (**Figure 3A**).
2. The monitor assesses the trees against a prepared set of configurations: each of the alternative configurations (including the current configuration) is compared to a set of prepared, "efficient" configurations. The configuration of cell 9, $\check{\boldsymbol{\tau}}_9^{\Sigma}$, notes global *SD* representations for plants and herbivores. This configuration is ranked lower than the alternative which has an individual based model for carnivores and a local *SD* submodel for the other entities in the cell. The monitor makes this change in configuration, and informs the global *SD* agent that it is no longer controlling the biomasses in cell 9 (**Figure 3B**).
3. The model may run for some time without any change in configuration. Both the herbivores and carnivores breed. The increased execution speed between **C** and **D** in **Figure 4** is a result of a change in representation: the number of carnivores, recorded in $\boldsymbol{\tau}_{SD(C),t_4}^{\Sigma}$, reaches a point that prompts the monitor to convert them to an *SD* form (**Figure 3C**).

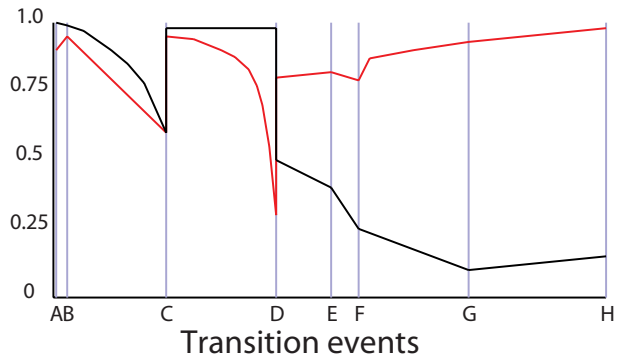


FIGURE 4 | Normalized indexes of execution speed (black) and fidelity (red) the against configuration changes through time associated with Figure 3.

4. The biomass of carnivores has increased significantly by the time the model reaches **D** in **Figure 4**, and they are now eating all the young herbivores; as a result the carnivore population is now prey-limited, and the *Relative biomass* condition is triggered. Both the carnivore and herbivore populations are converted to *IB* representations. Notice that dynamics in the fidelity in **Figure 4** around **D** arise from the collapse of the carnivore's prey, followed by the increase in fidelity after the representation change at **D** (**Figure 3D**).
5. A carnivore, agent 43, has been hungry ($(Hungry \geq \mathsf{A}_{moveT})$) and has migrated to the cell 5 (noted in $\boldsymbol{\tau}_{\mathrm{IB}(43),t_5}$). As occurred in cell 9 at step 2, the monitor converts plants and herbivores in cell 5 to a local *SD* representation, with *IB* carnivores (**Figure 3E**).
6. A lot of activity has occurred in this monitor interval: a *Starvation risk* is triggered in cell 9 because too many of the carnivores are hungry (many of the $\boldsymbol{\tau}_{\mathrm{IB}(n),t_7}$ trees indicate that the elapsed time without eating is greater than *Hungry*). There has been more migration to cells 5,6 and 8 from cell 9 (more of the $\boldsymbol{\tau}_{\mathrm{IB}(n),t_7}$ trees indicate residence in new cells), and a chance migration has introduced a carnivore into cell 3 from cell 5. Cells 3,6 and 8 are converted to local *SD* and *IB* representations as happened in step 3 (**Figure 3F**).
7. The population of carnivores in cell 9 crashes as a result of migration and the scarcity of prey, (reported in $\check{\boldsymbol{\tau}}_9^{\Sigma}$) The *IB* juvenile herbivores are patchy and harder to find, so only a few carnivores are getting enough to eat. There will be many $\boldsymbol{\tau}_{\mathrm{IB}(n),t_{10}}$ which indicate hunger or death due to starvation. The monitor cleans up the dead agents. There are chance migrations from cell 5 into cells 4 and 7 (in $\boldsymbol{\tau}_{SD(4),0}^{\Sigma}$ and $\boldsymbol{\tau}_{SD(7),0}^{\Sigma}$). A fire begins in cell 8, moving through cell 5: biomass loss in all niches causes all niches to shift to *IB* representations (**Figure 3G**).
8. Juvenile herbivores are reappearing in cell 9, but the available plant biomass (recorded in $\boldsymbol{\tau}_{\mathrm{SD}(9),t_{11}}$) has dropped due to reduced germination rates, triggering the *Relative biomass* condition in cell 9 causing the plants to convert to an *IB* representation. The fire in cell 8 has killed all animal biomass in the cell; they *do not* return to the global *SD* representation because their status trees diverge by too much. Instead, they convert to local *SD* representations (which represent zero biomass quite efficiently). Plants remain as *IB* agents The fire spreads to cell 5. **Figure 4** shows a modest increase in execution speed between **G** and **H** due to the population losses associate with the fire (**Figure 3H**).

- . . . the simulation continues.

4. Discussion

Adaptive hybrid models can be constructed so that each submodel is aware of its other representations and is able to change form as appropriate (Gray and Wotherspoon, 2012). This approach requires each model to have a reasonably close coupling with its alternative representations, and the burden of instrumenting (and maintaining) the necessary code quickly becomes untenable in complex models. Worse, it removes the possibility of more subtle configuration management that can accept poor performance in one part of a system in exchange for much better performance elsewhere. It seems that a guiding principle should be that in an adaptive hybrid model, each representation should know only as much about the rest of the model as it must know, and no more. The facility for a submodel to delve into the workings of other submodels, or the workings of the model as a whole, decreases the clarity that hybrid modeling makes possible, and opens avenues for unwanted, unanticipated behavior.

The major argument in favor of closely integrated representations for submodels is that it makes common (or at least similar) state variables easy to maintain across representations, even in the face of many representation changes. It is an attractive arguement, but the long term consequence is an ever growing burden of code maintenance.

Constructing hybrid models isn't significantly more complex than constructing traditional models. Adaptive hybrid models of the sort described in this paper will require a more significant investment in the design of a monitoring routine, and in the crafting of appropriate sets of candidate configurations. The transition dynamics such a model will exhibit depend on the sets of candidate configurations, and it seems likely that a combination of analysis and experimentation may be the most effective way to develop a set of useful configurations. The hybrid models associated with (Lyne et al., 1994; Little et al., 2006; Fulton et al., 2009) were built by extending the repertoire of ways of representing elements of the ecosystem or the anthropic components rather than wholescale redesign and replacement.

We can imagine an ideal adaptive hybrid model, where any state information which must be passed on is accompanied by an appropriate, opaque parcel of code to perform the maintenance. As long as the monitor knows what information each of these maintenance interfaces needs, they can be updated each time the monitor interrogates the agent which has control of the state data. This is a readily attainable ideal: many programming languages support first class functions with closures, and these features are precisely what we need to address this problem. *Scheme, Python, ML, Common Lisp, Lua, Haskell*, and *Scala* all have first order

functions with closures and, hence, the capacity to build model systems with this capability.

The state vectors and their supporting maintenance procedures can be treated as data and passed in lists associated with the status trees. If a monitor decides to swap representations, the accumulated lists of maintenance functions may be passed on to the new representation. A new representation inherits a maintenance list with variables that are part of its native state, it can claim them as its own and continue almost as though it had been running the whole time. In this way, a new representation doesn't need to know anything about its near kin, only that it must be able to run these black-box functions that come from other submodels, and to pass them on when required.

It may seem that this concentrates the global domain knowledge in the monitor, but this is not really the case. The monitor knows how to blindly query agents for state data and to the data in maintenance procedures. The monitor also knows how to recognize and rank characterizations of the states of the submodels or niches and to use those data to select a configuration.

The domain knowledge is encapsulated in the sets of targets the monitor matches the current configuration against, and in the heuristic triggers (such as *Starvation risk*) associated with a submodel or niche.

The essential problems any monitor is likely to deal with are problems of set selection (recognition, pattern matching…) and optimisation. These are common tasks: web searches, voice recognition, and route planning have become ingrained parts of modern society. Like route planning, the monitor needs to be able to reassess the "optimal" strategy as an ongoing process.

There are many options to choose from to rank configurations. A few of the likely candidates include

- Using an objective function to evaluate each of the possible configurations,
- Selecting a configuration based on decision trees,
- Using neural nets to match model states and direct us to an appropriate configuration,
- Using Bayesian networks to determine the most likely candidate,

and

- Using support vector machines to select the target/configuration pairs.

In writing this paper, one of the vexing difficulties has been finding a suitable mathematical representation which would allow comparisons between configurations, submodel states and the states of niches. We need proxies that describe models and configurations of models in a way that we may readily understand, manipulate and reason about, and being able to deal with submodels which are, in themselves, adaptive hybrid models, seems to be a naturally desirable trait. The vector space of trees described in the Appendix (Supplementary Material) has some nice properties, and may be directly useful with many of the options above: it forms a commutative ring (without necessarily having a unit), and would naturally inherit the body of techniques which only require the properties of such a ring.

5. Conclusion

There are still some major obstacles to developing a fully fledged adaptive hybrid model which is generic enough to tackle instances as varied as marine ecosystem modeling and urban planning. Foremost is a relative lack of real examples. The simulation of the hypothetical model[2] has tried to expose the character of an adaptive hybrid model which uses a monitor to manage the configuration of the system. There are parts of the description of the example system which are conspicuous by their absence; this is largely because they lie in almost wholly uncharted water. As a modeling community, we need to develop a wide range of approaches to how a model may assess the relative merits of a set of configurations. Many of the mechanisms we need for adaptive hybrid models already exist, but are found in domain specific models, and in wholly different domains, such as search engines and GPS navigation.

Establishing a suitable mathematical representation for model configurations which gives us access to well-developed techniques for set selection, pattern recognition and component analysis would seem to be almost as urgent as adaptive hybrid examples of real systems.

Acknowledgments

The authors would like to thank the two reviewers whose comments have improved the paper immeasurably. Thanks also go to a patient and understanding editor at Frontiers, and to Dr. Tony Smith, who gave up a weekend to work a scientifically and grammatically fine toothed comb through the paper. The responsibility for any mistakes, awkward sentences, or places where it just does not make sense now rests completely with the lead author.

Supplementary Material

The Supplementary Material for this article can be found online at: http://journal.frontiersin.org/article/10.3389/fenvs.2015.00058

[2]The model described in this paper is currently under development and will be made freely available when it has been completed.

References

Bailey, M. P., and Kemple, W. G. (1992). "The scientific method of choosing model fidelity," in *Proceedings of the 24th Conference on Winter* (New York, NY: ACM), 791–797.

Bobashev, G. V., Goedecke, D. M., Yu, F., and Epstein, J. M. (2007). "A hybrid epidemic model: combining the advantages of agent-based and equation-based approaches," in *Simulation Conference, 2007 Winter* (Miami, FL: IEEE), 1532–1537.

Botkin, D. B., Janak, J. F., and Wallis, J. R. (1972). Some ecological consequences of a computer model of forest growth. *J. Ecol.* 60, 849–872. doi: 10.2307/2258570

Chivers, W. J. (2009). *Generalised, Parsimonious, Individual-based Computer Models of Ecological Systems*. Newcastle, NSW: University of Newcastle.

DelSole, T., and Shukla, J. (2010). Model fidelity versus skill in seasonal forecasting. *J. Clim.* 23, 4794–4806. doi: 10.1175/2010JCLI3164.1

Fulton, E. (2010). Approaches to end-to-end ecosystem models. *J. Mar. Syst.* 81, 171–183. doi: 10.1016/j.jmarsys.2009.12.012

Fulton, E., Gray, R., Sporcic, M., Scott, R., and Hepburn, M. (2009). "Challenges of crossing scales and drivers in modelling marine systems," in *18th World IMACS Congress and MODSIM09 International Congress on Modelling and Simulation* (Cairns, QLD), 2108–2114.

Fulton, E., Gray, R., Sporcic, M., Scott, R., Little, L., Hepburn, M., et al. (2011). *Ningaloo Collaboration Cluster: Adaptive Futures for Ningaloo*. Technical Report 5.3, Ningaloo Collaboration Cluster, Hobart, TAS.

Gray, R., Fulton, E., Little, L., and Scott, R. (2006). *Operating Model Specification Within an Agent Based Framework. North West Shelf Joint Environmental Management Study Technical Report*. Number 16 in CSIRO-CMAR NWSJEMS Technical Reports. CSIRO, Hobart, TAS. doi: 10.1016/j.envsoft.2011.08.012

Gray, R., and Wotherspoon, S. (2012). Increasing model efficiency by dynamically changing model representations. *Environ. Model. Softw.* 30, 115–122. doi: 10.1016/j.envsoft.2011.08.012

Little, L., Fulton, E., Gray, R., Hayes, D., Scott, R., McDonald, A., et al. (2006). *Management Strategy Evaluation Results and Discussion for the North West Shelf*. Final Report 14, CSIRO Australia, Hobart, TAS.

Lyne, V., Gray, R., Sainsbury, K., and Scott, R. (1994). *Integrated Biophysical Model Investigations*. Final report, CSIRO Australia, Division of Fisheries, Hobart, Tasmania.

Vincenot, C. E., Giannino, F., Rietkerk, M., Moriya, K., and Mazzoleni, S. (2011). Theoretical considerations on the combined use of system dynamics and individual-based modeling in ecology. *Ecol. Modell.* 222, 210–218. doi: 10.1016/j.ecolmodel.2010.09.029

Walters, C. J., and Martell, S. J. (2004). *Fisheries Ecology and Management*. New Jersey: Princeton University Press.

Yip, W., and Marlin, T. E. (2004). The effect of model fidelity on real-time optimization performance. *Comput. Chem. Eng.* 28, 267–280. doi: 10.1016/S0098-1354(03)00164-9

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Environmental Science

METHODS
published: 10 June 2015
doi: 10.3389/fenvs.2015.00042

Bayesian inference for stochastic individual-based models of ecological systems: a pest control simulation study

Francesca Parise[1†], *John Lygeros*[1] *and Jakob Ruess*[2*†]

[1] *Automatic Control Laboratory, Department of Information Technology and Electrical Engineering, ETH Zürich, Zurich, Switzerland,* [2] *Institute of Science and Technology Austria, Klosterneuburg, Austria*

Mathematical models are of fundamental importance in the understanding of complex population dynamics. For instance, they can be used to predict the population evolution starting from different initial conditions or to test how a system responds to external perturbations. For this analysis to be meaningful in real applications, however, it is of paramount importance to choose an appropriate model structure and to infer the model parameters from measured data. While many parameter inference methods are available for models based on deterministic ordinary differential equations, the same does not hold for more detailed individual-based models. Here we consider, in particular, stochastic models in which the time evolution of the species abundances is described by a continuous-time Markov chain. These models are governed by a master equation that is typically difficult to solve. Consequently, traditional inference methods that rely on iterative evaluation of parameter likelihoods are computationally intractable. The aim of this paper is to present recent advances in parameter inference for continuous-time Markov chain models, based on a moment closure approximation of the parameter likelihood, and to investigate how these results can help in understanding, and ultimately controlling, complex systems in ecology. Specifically, we illustrate through an agricultural pest case study how parameters of a stochastic individual-based model can be identified from measured data and how the resulting model can be used to solve an optimal control problem in a stochastic setting. In particular, we show how the matter of determining the optimal combination of two different pest control methods can be formulated as a chance constrained optimization problem where the control action is modeled as a state reset, leading to a hybrid system formulation.

Keywords: stochastic population dynamics, moment equations, Bayesian parameter inference, optimal control, agricultural pests

OPEN ACCESS

Edited by:
Christian E. Vincenot, Kyoto University, Japan

Reviewed by:
Marc Thilo Figge, Hans Knöll Institute, Germany
Jean-Christophe Augustin, Ecole Nationale Vétérinaire d'Alfort, France

***Correspondence:**
Jakob Ruess, Institute of Science and Technology Austria, Am Campus 1, Klosterneuburg 3400, Austria
jruess@ist.ac.at

[†] *These authors have contributed equally to this work.*

Specialty section:
This article was submitted to Environmental Informatics, a section of the journal Frontiers in Environmental Science

Received: *09 April 2015*
Accepted: *22 May 2015*
Published: *10 June 2015*

Citation:
Parise F, Lygeros J and Ruess J (2015) Bayesian inference for stochastic individual-based models of ecological systems: a pest control simulation study. Front. Environ. Sci. 3:42. doi: 10.3389/fenvs.2015.00042

1. Introduction

The use of mathematical models in population ecology and epidemiology has a long history (Murray, 2002). Among the wide range of available models, two major categories can be distinguished: *population-level* and *individual-based* models (Black and McKane, 2012). Population-level models (PLMs) implicitly assume an infinite population

size and provide a phenomenological description of the overall population behavior. The main advantage of PLMs is that they involve ordinary differential equations that can be analyzed using dynamical systems theory; their main drawback is that they neglect any effect that may be caused by finite population sizes or by the inherently random nature of interactions between individuals. Individual-based models (IBMs), on the other hand, are discrete models that represent an ecological system as a collection of a finite number of individuals that are modeled explicitly (DeAngelis and Mooij, 2005; Railsback and Grimm, 2012). Depending on how many details are included for each individual, IBMs can be further distinguished into agent-based models, which provide great detail but are limited to algorithmic and numerical analysis, and stochastic process models that typically distinguish a limited number of different types of individuals but are more amenable to analytical investigation (Black and McKane, 2012). Among IBMs we restrict our attention to stochastic process models where the possible interactions of species occur with a probability that is proportional to the number of individuals present in the system. Mathematically, these models can be represented by continuous-time discrete-state Markov chains (CTMCs).

Discrete stochastic models naturally arise in many applications both in ecology (Marion et al., 1998; Ovaskainen and Cornell, 2006; Ovaskainen and Meerson, 2010) and in epidemiology (Isham, 1991; Nåsell, 2002). While stochastic IBMs and CTMCs in particular, provide an intuitive description of these systems, their analysis tends to be difficult. Consequently, the use of these models has for a long time remained limited to systems containing at most a handful of different interacting species. Recent years have seen, on the one hand, an immense increase in computational resources, and on the other hand, a surge of studies in which IBMs have been used to study biological systems at the cellular level. Specifically, in these applications interacting molecules play the role of interacting individuals (Balazsi et al., 2011; Goutsias and Jenkinson, 2013; Neuert et al., 2013). These developments have stimulated research on new methods for analyzing IBMs (see, e.g., Munsky and Khammash, 2006; Wolf et al., 2010; Ruess et al., 2011, 2013), which opened the path for using larger and more complicated models that are more suitable to represent the complex systems encountered in applications. As a consequence of these new modeling capabilities, problems that were for a long time solvable only using PLMs can now be analyzed (and hence receive renewed interest) in the context of IBMs. In the case study of this paper, for example, we consider the use of IBMs to address optimal control problems for pest management. Many studies on optimal control problems in ecology were performed using PLMs during the 1970's (see Wickwire, 1977 and references therein). In these pioneering works, the control actions were usually included by changing the model parameters (e.g., birth or death rates) according to the control action or by developing a new model for the controlled system, in which the control is included as a continuous input. In the case study presented here, we adopt a different modeling approach, similar to the idea presented in Jaquette (1970) for a simple birth-death process. Specifically, we assume that the control operator can reset the state of the system (for example reduce the number of pest individuals by applying pesticide) at certain given intervention times, while the dynamics of the system between the reset times follow the uncontrolled IBM. The resulting model is thus a stochastic hybrid system with controlled state reset (Branicky et al., 1998; Bensoussan and Menaldi, 2000).

In the 1970's, the main difficulty in translating the control theoretical studies into real world applications was the lack of efficient and accurate procedures to estimate the required model parameters (Wickwire, 1977). While for PLMs these difficulties have been to a large extent overcome, for the more detailed IBMs the problem of reverse engineering parameter values from measured data is still a major challenge (Poovathingal and Gunawan, 2010; Stumpf, 2014). Most of the available approaches, in fact, require iterative evaluation of parameter likelihoods, which are usually not available analytically and computationally very expensive. To circumvent this problem, likelihood-free approaches for parameter inference, such as approximate Bayesian computation, have found applications in ecology (McKinley et al., 2009; Lagarrigues et al., 2014) and other fields (Toni et al., 2009).

In this paper, we suggest a different approach for parameter inference based on an expression that allows for fast (but approximate) evaluation of the likelihood. To obtain this expression, we move from a full description of the stochastic model to ordinary differential equations that describe the time evolution of only means and (co)variances (and possibly higher order moments) of the interacting species. We then use moment closure techniques to approximate the solution of these equations and show that the results can be used to approximate the likelihood. Moment closure methods have a long history in population biology (Whittle, 1957; Nåsell, 2003; Krishnarajah et al., 2005; Singh and Hespanha, 2006; Hespanha, 2008), but only recently first attempts have been made to use these methods for parameter inference (Kügler, 2012; Zechner et al., 2012; Milner et al., 2013). The approach we take here is motivated by recent developments in the modeling of biochemical reaction networks (Zechner et al., 2012) and differs from typical methods in ecology (Ross et al., 2009; Gillespie and Golightly, 2010) in that it is intended to be used with data that are obtained from many independent observations of a system, instead of being tailored to a single observation of an ecological system over time. Accordingly, it is most suitable for applications in which many replicates of the same experiment can be performed, as typically done in microcosm and mesocosm experiments (Srivastava et al., 2004; Hekstra and Leibler, 2012; Altermatt et al., 2014), but only one measurement per replica is feasible. This might be the case when measurements are costly or time-consuming, resulting in a trade-off between the number of measurements and the number of replicates (e.g., Carrara et al., 2014), or if they perturb the system so that further measurements of the same replicate are not possible or meaningful. This is the typical scenario for destructive measurements as, for instance, in the study of bacterial colonization dynamics where measurements may require sequencing the bacterial genomes of the whole community (Cordero et al., 2012). At the subcellular scale, this type of data is naturally obtained when the experimental

replicates are individual cells and the intracellular dynamics of chemical species are measured (e.g., Zechner et al., 2012).

2. Materials and Methods

2.1. Individual-Based Modeling

Consider an ecological system comprising m species, $X_1, \ldots, X_m$ that can interact, in a given habitat, according to K different types of interactions

$$\nu'_{1k}X_1 + \ldots + \nu'_{mk}X_m \xrightarrow{\theta_k} \nu''_{1k}X_1 + \ldots + \nu''_{mk}X_m,\ k = 1, \ldots, K. \tag{1}$$

The expression on the left hand side of the arrow denotes the amount ν'_{ik} of individuals for each species X_i needed for interaction k to happen. The expression on the right hand side describes the result of the interaction. In other words, the effect of interaction k is to update the number of individuals of each species X_i by the net amount $\nu_{ik} := \nu''_{ik} - \nu'_{ik}$. Let $X(t) = \left[X_1(t) \cdots X_m(t)\right]^\top$ denote the amount of individuals of each species present in the system at time t. In the following, we assume that each interaction is a stochastic event whose probability to occur depends on the probability that the required amounts of individuals meet in some location of the habitat and on a parameter θ_k that determines the probability that the individuals successfully interact when they meet. Since the interactions are stochastic events, $X(t)$ is a stochastic process that takes values $x = [x_1 \cdots x_m]^\top \in \mathbb{N}_0^m$. Under the assumption of random movement of the individuals, the probability that a given interaction takes place in the infinitesimal time interval $[t, t+dt]$, given the current population state $X(t) = x$, can be determined by the law of mass action as

$$a_k(x, \theta)dt := \theta_k h_k(x)dt, \quad \text{where } h_k(x) = \prod_{i=1}^{m} \binom{x_i}{\nu'_{ik}},$$
$$k = 1, \ldots, K \quad \text{and } \theta = \{\theta_1, \ldots, \theta_K\}.$$

Note that we assumed that the habitat is homogeneous, that is, the probabilities $a_k(x, \theta)dt$ do not depend on the spatial location, but only on the total amount x of individuals present in the system.

2.2. Data Description

Since the model introduced in the previous section is stochastic, for a given initial population $X(0) = x_0$, many different evolutions of the system, corresponding to different realizations of the stochastic process $X(t)$, are possible. In the following, we assume that we can monitor many of these different replicates of the system, but only one measurement per replica can be taken (possibly at different times). Therefore, the collected data consists of several measurements of the number of individuals of one species[1] X_j at different time points, each coming from a different replicate. This means that we assume that the collected data contains information about the dynamics but not about the correlation of the species abundance between different time points. Let $t_1, \ldots, t_S$ denote the measurement times and suppose that for each measurement time we have measured n different replicates[2]. The data set is then of the form $D = \left\{X_j^1(t_s), \ldots, X_j^n(t_s)\right\}_{s=1}^{S}$, where all the measurements $X_j^i(t_s),\ i = 1, \ldots, n,\ s = 1, \ldots, S$ are statistically independent.

A feature of such data, which is at the same time a strength and a serious complication, is that the measurements described above correspond to observations from $n \cdot S$ (supposedly) identical replicates of the system. In a realistic situation, these replicates might be performed in different days or come from slightly different ambient conditions. Not taking into account this source of variability can have deleterious effects on model predictions and accordingly also on any strategy for optimal interventions and population control. Consequently, the variability observed in different replicates is an asset that should be used in order to identify a model that is not tailored to one particular experimental condition but can describe all of them. To allow this type of flexibility, we assume that in different repetitions of an experiment some of the parameters $\theta_k,\ k = 1, \ldots, K$ may be slightly different. Without loss of generality, let us denote by $k = 1, \ldots, r$ the interactions for which the rate θ_k varies between different repetitions of the experiment, and by $k = r+1, \ldots, K$ the remaining ones that are the same for all repetitions. We describe the experimental variability by assuming that the success of the first r interactions is given by $\theta_k \cdot Z_k,\ k = 1, \ldots, r$, where $Z = [Z_1 \cdots Z_r]^\top$ is a random vector with unknown distribution P_Z. Moreover, we assume that the marginal means of P_Z are all equal to one, so that $\theta_k,\ k = 1, \ldots, r$ are the average interaction success rates over different replicates of the experiment. This gives rise to a model akin to mixed-effects models in the statistics literature (Lavielle, 2014).

2.3. The Conditional Master Equation and Moment Dynamics

Under the assumption of a homogeneous environment, for fixed success parameters θ_k, the time evolution of the number of individuals $X(t)$, as described in Section 2.1, follows a continuous-time Markov chain. Consequently, the time evolution of its probability distribution $p(x, t) := P\left[X(t) = x\right]$, can be described by a master equation (Black and McKane, 2012). The same theory holds in the case of random success rates $\theta_k \cdot Z_k$, $k = 1 \ldots, r$, if we fix a specific realization z of the random vector Z. Mathematically, the conditional process $X(t) \mid Z = z$ is Markovian and can be described by the master equation

$$\dot{p}(x, t|z) = -p(x, t|z) \sum_{k=1}^{K} a_k(x, \theta, z) + \sum_{k=1}^{K} p(x - \nu_k, t|z) a_k(x - \nu_k, \theta, z), \tag{2}$$

where[3] $p(x, t|z) := P\left[X(t) = x | Z = z\right]$. Typically, Equation (2) cannot be solved explicitly and for complex systems also

[1] Extension to multiple measured species is straightforward.

[2] The assumption that n is the same for each time point is only for notational convenience and is by no means necessary.

[3] For ease of notation we omit the dependence of this probability on the initial condition.

numerical approaches might fail. However, if one is not interested in the whole probability distribution, Equation (2) can be used to derive evolution equations for some of the moments of the distribution, as mean and variance. Specifically, denoting by $\tilde{\psi}$ the l-dimensional vector of the (uncentered) moments up to some order L of the joint process $\widetilde{X}(t) = [X(t)^\top Z^\top]^\top$ and by $\bar{\tilde{\psi}}$ the vector containing the higher order moments of $\widetilde{X}(t)$, one obtains

$$\frac{d}{dt}\tilde{\psi}(t) = \tilde{A}(\theta)\tilde{\psi}(t) + \tilde{B}(\theta)\bar{\tilde{\psi}}(t), \tag{3}$$

where $\tilde{A}(\theta) \in \mathbb{R}^{l\times l}$ and $\tilde{B}(\theta) \in \mathbb{R}^{l\times\infty}$ are matrices defined by the reaction network and by the parameters θ [a detailed derivation is given in Zechner et al. (2012)]. For mass action kinetics with at most pairwise interactions of the species, $\bar{\tilde{\psi}}$ contains moments of order at most $L + 2$, so that $\tilde{B}(\theta)$ is a finite dimensional matrix (Ruess and Lygeros, 2015). Nonetheless, the system in Equation (3) is not solvable because it depends on the unknown quantities $\bar{\tilde{\psi}}(t)$ (which act as an external input, that is, they are not part of the state vector). To overcome this issue, one can use moment closure techniques to approximate the unknown higher order moments $\bar{\tilde{\psi}}(t)$ by non-linear functions of the lower order moments, that is, $\bar{\tilde{\psi}}(t) \cong \tilde{f}(\tilde{\psi}(t))$. As a consequence, the right hand side of Equation (3) can be approximated with an expression that depends only on the state variables leading to the solvable closed system

$$\frac{d}{dt}\tilde{\nu}(t) = \tilde{A}(\theta)\tilde{\nu}(t) + \tilde{B}(\theta)\tilde{f}(\tilde{\nu}(t)). \tag{4}$$

Note that in Equation (4) we used a different symbol for the state vector to stress the fact that $\tilde{\nu}(t)$ are approximations of the true moments $\tilde{\psi}(t)$, since they are obtained as solution of the approximated dynamics.

Given that the marginal moments corresponding to Z of the joint process $\widetilde{X}(t)$ are constant, we can write Equation (4) as

$$\frac{d}{dt}\nu(t) = A(\theta)\nu(t) + C(\theta)\mu^Z + B(\theta)f(\nu(t), \mu^Z), \tag{5}$$

where $\nu(t)$ are the moments of $\widetilde{X}(t)$ excluding the marginal moments of Z, μ^Z are the moments of Z up to order L and the matrices $A(\theta)$, $B(\theta)$, $C(\theta)$ are sub-matrices of $\tilde{A}(\theta)$, $\tilde{B}(\theta)$. This form of the equations is convenient because we can now regard the moments μ^Z as additional parameters and the system of moment equations for $\nu(t)$ as being parameterized by $\gamma = \{\theta, \mu^Z\}$. Consequently, for a specified parameter vector γ, the system of Equation (5) can be solved numerically allowing one to compute any desired moment of $X(t)$ up to order L.

2.4. Bayesian Inference with Population Data

In real applications, usually neither the rates θ nor the variability μ^Z between repetitions of the experiments are known. Hence, to obtain a model that is useful in practice, we need to estimate the parameters γ from measured data. This task can be posed as a Bayesian parameter inference problem, where any available knowledge about γ can be specified as an *a priori* parameter distribution $p(\gamma)$. The result of the Bayesian inference procedure is a parameter posterior distribution $p(\gamma|d)$ that reflects the updated belief about γ, given the observed realization d of the data D. According to Bayes' rule, this posterior distribution can be obtained as

$$p(\gamma|d) = \frac{p(d|\gamma)\cdot p(\gamma)}{p(d)},$$

where $p(d|\gamma)$ is the likelihood of γ for the observed realization d, $p(\gamma)$ is the prior distribution, and $p(d) = \int p(d|\gamma)p(\gamma)d\gamma$ is the marginal likelihood of the data. Since computing the posterior distribution analytically is usually impossible, Monte Carlo schemes are typically used to draw samples $\gamma_1, \ldots, \gamma_M$, from $p(\gamma|d)$, allowing one to construct an empirical estimate of the posterior distribution. The iterative evaluation of the likelihood $p(d|\gamma)$, needed in these schemes, can however be computationally very expensive or even impossible for complex high dimensional systems. For the data considered in this paper, for example, evaluating the parameter likelihood requires computing the distribution of the measured species at all the measurement time points. Specifically, the likelihood is given by

$$p(d|\gamma) = \prod_{s=1}^{S}\prod_{i=1}^{n} P\left[X_j^i(t_s) = x_j^i(t_s)|\gamma\right] = \prod_{s=1}^{S}\prod_{i=1}^{n} p_j(x_j^i(t_s), t_s|\gamma),$$

where $p_j(\cdot, t|\gamma)$ is the distribution of species X_j at time t given that γ are the model parameters, and $x_j^i(t_s)$, $i = 1, \ldots, n$, $s = 1, \ldots, S$ are the measured abundances of species X_j. The factorization of the joint distribution over time points and samples stems from the assumption that all the measurements are statistically independent. It is evident that evaluating this likelihood requires computing $p_j(\cdot, t|\gamma)$. This cannot be done (except in some special cases) without first computing the entire joint distribution of $\widetilde{X}(t)$ at all the measurement time points, hence solving Equation (2). For these reasons, exact Bayesian inference is very difficult, if not impossible.

A naive idea to overcome these issues would be to approximate $p_j(\cdot, t|\gamma)$ by using the moments computed according to Equation (5) together with the assumption that $p_j(\cdot, t|\gamma)$ belongs to a certain family (e.g., that it is a Gaussian distribution). Such assumptions are, however, in general not satisfied and, as detailed in the following, they are not really necessary. By using as data the first L moments of the measured samples only, it is in fact possible to derive a different likelihood function that is correct for any distribution $p_j(\cdot, t|\gamma)$, in the limit of $n \to \infty$. Specifically, set $L = 2$ and let $\hat{\mu}_1(t_s)$ and $\hat{\mu}_2(t_s)$ be sample mean and variance of the random samples $D(t_s) = \left\{X_j^1(t_s), \ldots, X_j^n(t_s)\right\}$, which represent the measured species abundances at time t_s, $s = 1, \ldots, S$. Furthermore, denote by $\mu_1(t)$ the mean and by $\mu_i(t)$, $i = 2, \ldots 4$, the centered moments up to order four of $p_j(\cdot, t|\gamma)$. By the central limit theorem, the probability density function $p_{\hat{\mu}}(\cdot|\gamma)$ of $\hat{\mu}: = \left[\hat{\mu}(t_1)^\top \cdots \hat{\mu}(t_S)^\top\right]^\top$, where $\hat{\mu}(t_s): = \left[\hat{\mu}_1(t_s)\ \hat{\mu}_2(t_s)\right]^\top$ is

the vector of sample moments up to order 2 at time t_s, is for n large enough given by

$$\begin{aligned} p_{\hat{\mu}}(\cdot|\gamma) &= \prod_{s=1}^{S} p_{\hat{\mu}(t_s)}(\cdot|\gamma), \quad \text{where} \\ p_{\hat{\mu}(t_s)}(\cdot|\gamma) &= \mathcal{N}(M(t_s), \Sigma(t_s)) \quad \text{and} \\ M(t_s) &= \begin{bmatrix} \mu_1(t_s) \\ \mu_2(t_s) \end{bmatrix} \quad \text{and} \\ \Sigma(t_s) &= \frac{1}{n} \begin{bmatrix} \mu_2(t_s) & \mu_3(t_s) \\ \mu_3(t_s) & \mu_4(t_s) - \frac{n-3}{n-1}\left(\mu_2(t_s)\right)^2 \end{bmatrix}. \end{aligned} \tag{6}$$

Equation (6) allows one to evaluate the likelihood for the collection of sample moments $\hat{\mu}$ up to order two, given the first four moments of $p_j(\cdot, t|\gamma)$. These moments can be obtained efficiently by numerically solving Equation (5). Consequently, using this approach it becomes feasible to draw samples from $p(\gamma|\hat{\mu})$ using a Markov chain Monte Carlo scheme. The downside is that, by transitioning from the full data to the sample moments up to order $L = 2$, all the information about the parameters γ that might have been provided by higher order statistics of the data is discarded (Ruess and Lygeros, 2013). However, formulas for the likelihood of sample moments up to any desired order L can be obtained in exactly the same way, and thus, this approach is not limited to sample means and variances only. Evaluating the likelihood for sample moments up to order L requires computing moments of $p_j(\cdot, t|\gamma)$ up to order $2L$, which becomes computationally expensive for large L. The choice of how many sample moments to include in the parameter inference is therefore a trade-off between computational cost and neglected information.

3. Case Study: Optimal Pest Control

3.1. The Model

As case study we consider the problem of modeling, and eventually controlling, the evolution of an agricultural pest. To this end, we consider an extension of the model of cotton aphids proposed in Matis et al. (2007) and Gillespie and Golightly (2010). Specifically, we introduce an additional immigration term to the original model and we include a recovery process of the habitat. In more detail, our model consists of a discrete state stochastic process $N(t)$ that describes the size of the current pest population and a discrete state stochastic process $C(t)$ that is used as an indicator of how much the environment has been deteriorated, up to time t, by the infestation. In the following, we assume that these two processes are updated according to the occurrence of the following stochastic events:

$$\begin{aligned} \emptyset &\xrightarrow{\alpha} N + C \\ N &\xrightarrow{\lambda Z} 2N + C \\ N + C &\xrightarrow{\eta} C \\ C &\xrightarrow{r} \emptyset. \end{aligned} \tag{7}$$

Specifically, we suppose that new pest individuals arise in the system due to immigration, with rate α, or birth events, with a rate that is proportional to the current population size. To capture variability between different replicates, stemming for instance from different ambient conditions, we assume that the birth rate is given by $\lambda N(t)Z$, where Z is a one-dimensional random variable distributed according to a log-normal distribution P_Z with mean one and unknown variance. The death rate of the pest is given by $\eta N(t)C(t)$, i.e., it depends on the current population size, but also on the damage to the environment.

Furthermore, we assume that the process describing the state of the environment, $C(t)$, is increased by one unit whenever a new pest individual is added in the system (either via immigration or due to a birth event). Since pest individuals deteriorate the environment for a time period that may exceed their own life span, we assume that the death of pest individuals leaves $C(t)$ unchanged. However, we model the fact that the environment may eventually recover by assuming that $C(t)$ decreases with rate $rC(t)$.

This model induces a conditional master equation, see Equation (2), with state $x = [n \quad c]^\top$ and parameters $\theta = \{\alpha, \lambda, \eta, r\}$, in which $a_k(x, \theta, z)$ and ν_k, $k = 1, \ldots, 4$ are given by

$$\begin{aligned} a_1(x, \theta, z) &= \alpha, & \nu_1 &= \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \\ a_2(x, \theta, z) &= \lambda \cdot n \cdot z, & \nu_2 &= \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \\ a_3(x, \theta, z) &= \eta \cdot n \cdot c, & \nu_3 &= \begin{bmatrix} -1 \\ 0 \end{bmatrix}, \\ a_4(x, \theta, z) &= r \cdot c, & \nu_4 &= \begin{bmatrix} 0 \\ -1 \end{bmatrix}. \end{aligned}$$

From this master equation we can derive moment equations and use moment closure to obtain a closed system in the form of Equation (5). In the following, we use equations for the moments up to order four and a fifth-order derivative matching closure, as described in Singh and Hespanha (2006). Solving the resulting approximate systems, for given parameter values θ and given first two moments of Z, enables us to approximately compute the moments of $X(t) = [N(t)\, C(t)]^\top$ up to order four.

3.2. Inference Results

Since we assumed that Z has a log-normal distribution with mean one, it is sufficient to include only the variance of Z as an unknown parameter, so that $\gamma = \{\theta, \mathrm{Var}[Z]\}$. For the *in silico* case study, we assume that $N(0) = C(0) = 0$, that is initially no pests are present in the system, and that the true values of the parameters are given by

$$\begin{aligned} \alpha = 0.03, \quad \lambda = 0.012, \quad \eta = 0.25 \cdot 10^{-4}, \\ r = 0.003 \quad \text{and} \quad \mathrm{Var}[Z] = 0.05, \end{aligned} \tag{8}$$

where we used hours as time units. These parameters produce pest outbreaks that are on the timescale of realistic profiles

of aphids pest infestations. As hypothetical data set $D = \{N^1(t_s), \ldots, N^n(t_s)\}_{s=1}^{S}$, we consider a case study in which $n = 100$ different replicates of the system are measured once a week ($t_s = 7 \cdot 24 \cdot s$ hours) for a total of $S = 5$ weeks. We note that the chosen measurements times are consistent with the dynamics generated by the parameters given in Equation (8). If, as in real scenarios, the parameters and hence the timescale of the systems dynamics are not known one may use a sequential experiment design approach to produce informative datasets. To generate the dataset considered in this *in silico* case study we used the stochastic simulation algorithm (SSA), described in Gillespie (1976), with randomly drawn values z from P_Z. The sample means and variances of the data are denoted by $\hat{\mu} = \left[\hat{\mu}(t_1)^\top \cdots \hat{\mu}(t_5)^\top\right]^\top$ where $\hat{\mu}(t_s) := \left[\hat{\mu}_1(t_s)\, \hat{\mu}_2(t_s)\right]^\top$, $s = 1, \ldots, 5$. In principle, this gives us all the ingredients to evaluate the likelihood using Equation (6). However, the fact that we approximate the moments μ_i, $i = 1, \ldots, 4$ using moment closure means that we only have an approximation of the true covariance matrices $\Sigma(t_s)$, $s = 1, \ldots, 5$. Since these approximations are not guaranteed to be positive semi-definite, a further step may be required in which the approximated symmetric matrices are projected onto the cone of positive semi-definite matrices. Another possibility, which we follow here, is to construct empirical estimates $\hat{\Sigma}(t_s)$ of $\Sigma(t_s)$, $s = 1, \ldots, 5$, from the measured data and use these in Equation (6). This procedure is reasonable whenever sufficient data is available to estimate moments up to fourth order to acceptable precision.

We assume that no prior information about γ is available and accordingly choose flat prior distributions for all the parameters. To draw samples from the posterior distribution $p(\gamma|\hat{\mu})$ we used a Metropolis-Hastings Markov chain Monte Carlo algorithm

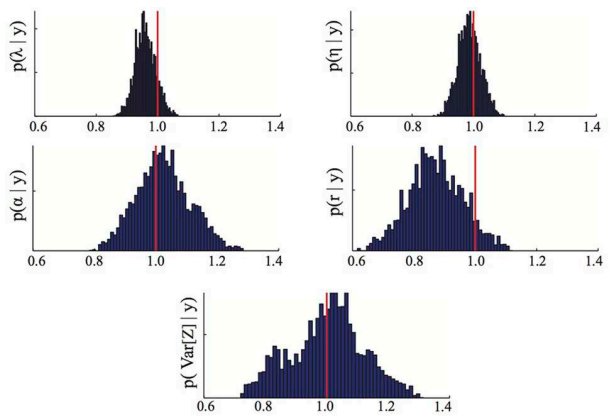


FIGURE 1 | Parameter posterior distribution. The panels show different marginals of the posterior distribution computed using the Bayesian inference MCMC approach described in Section 3.2. For each panel, the x-axis has been rescaled to show ratios of inferred to true parameter value. The red line highlights the 1/1 ratio, which corresponds to perfectly inferred parameters. The maximum a posteriori estimates $\hat{\gamma}_{MAP}$ are those maximizing the posterior distribution. The parameters are λ = birth, η = death, α = immigration, r = recovery rate.

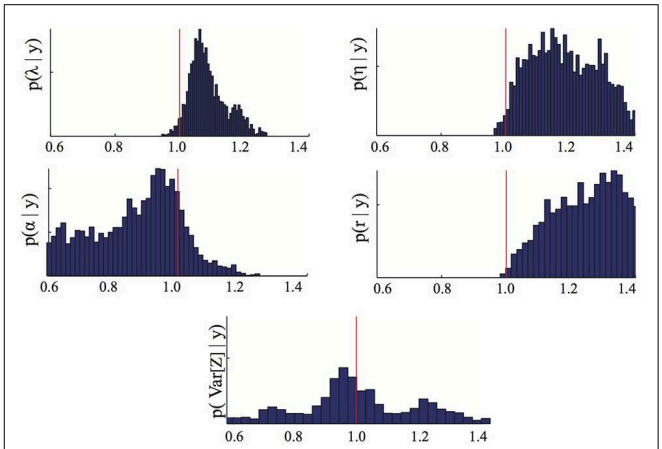


FIGURE 3 | Parameter posterior distribution. The panels show different marginals of the posterior distribution computed using the Bayesian inference MCMC approach if only the means of the dataset D are used. For each panel, the x-axis has been rescaled to show ratios of inferred to true parameter value. The red line highlights the 1/1 ratio, which corresponds to perfectly inferred parameters. The parameters are λ = birth, η = death, α = immigration, r = recovery rate.

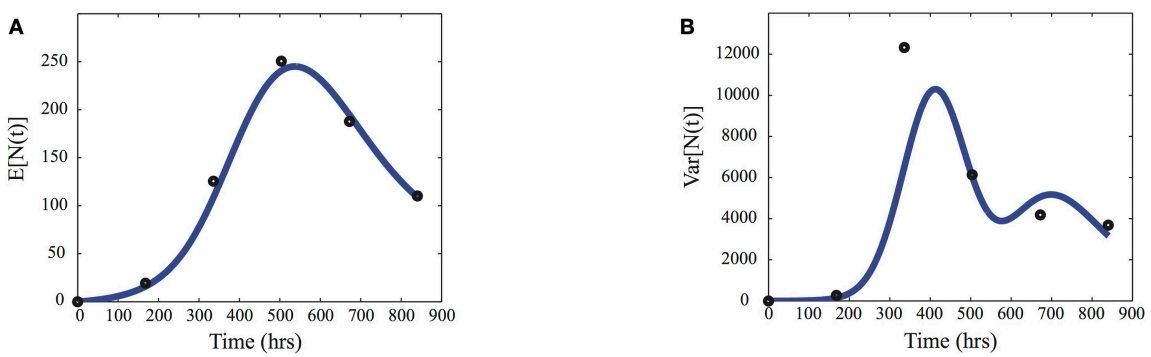


FIGURE 2 | Inference data. Comparison of the sample means **(A)** and variances **(B)** of the pest population $N(t)$ used to infer the parameters (dots) and the predictions given by the moment Equation (5) using the maximum a posteriori estimates $\hat{\gamma}_{MAP}$ given in Equation (9).

with randomly chosen initial parameter guesses, log-normal proposal distributions and a chain length of 10000 (for more information on this algorithm see, for instance, Zechner et al., 2012). The first 3000 iterations of the chain were discarded as a burn-in period and an empirical estimate of the posterior distribution was obtained from the remaining 7000 iterations of the chain. The results are shown in **Figure 1**. It can be seen that the posterior distributions are relatively tight with mode close to the true parameters. The small deviations between posterior mode and true parameter value visible in some of the panels stem from a combination of approximation error due to moment closure and errors coming from the fact that the moments used as inference data were estimated from a finite ($n = 100$) number of replicates and are hence affected by the noise described in Equation (6). The obtained maximum a posteriori estimates $\hat{\gamma}_{MAP}$ are

$$\hat{\alpha} = 0.0307, \quad \hat{\lambda} = 0.0115, \quad \hat{\eta} = 0.247 \cdot 10^{-4}, \\ \hat{r} = 0.0026 \quad \text{and} \quad \hat{\mathrm{V}}\mathrm{ar}\,[Z] = 0.0501. \tag{9}$$

In **Figure 2**, the mean and variance, computed from the model using Equation (5) and $\gamma = \hat{\gamma}_{MAP}$, are compared to the sample means and variances $\hat{\mu}(t_s)$ of the considered data set.

To assess the advantages of the proposed stochastic approach with respect to more standard methods based on measurements of the average species density only, we repeated the previously described inference process using only the means $\{\hat{\mu}_1(t_s)\}_{s=1}^{S}$ as data. By comparing the parameter posterior distribution obtained in this case (**Figure 3**) with the one obtained using as inference data also the variance (**Figure 1**) one can immediately see that higher order statistics may contain valuable information regarding the parameters. As a consequence, the proposed approach could help solving identifiability problems of standard inference approaches based on deterministic models.

The previous results were obtained assuming a dataset that contains $n = 100$ replicates of the system for each measurement time. As rigorously encoded in the mathematical description of the noise given in Equation (6), the variance of the estimates is inversely proportional to the number of samples n. Consequently, if the number of replicates is very low, the variance of the data used in the inference may be large resulting in inconclusive (i.e., very spread) parameter posterior distributions. This is due to the fact that many different parameters give rise to model predictions that are consistent with the high level of noise given by Equation (6). To test the performance of the proposed approach for different levels of noise, we performed a case study in which we estimated the real parameter values given in Equation (8) using mean and variances estimated from different numbers of replicates. In particular, we considered four different scenarios with $n = 10$, $n = 25$, $n = 50$ and $n = 100$ replicates and simulated 10 different datasets for each scenario to reduce the influence of the particular realization of the data in the results. We then performed the parameter inference for all scenarios and all datasets (i.e., 40 times) and computed the relative error $\left(\text{e.g., } 100 \cdot \frac{|\hat{\lambda}-\lambda|}{\lambda}\right)$ of the MAP estimates with respect to the real ones. The results are reported in **Table 1**. It can be seen that the precision of the estimates becomes larger as the number of replicates increases. In particular, $n = 10$ replicates lead to very imprecise results, whereas the estimates obtained from both $n = 50$ and $n = 100$ replicates attain reasonable precision. The scenario with $n = 25$ replicates provides an intermediate case and may or may not be sufficiently accurate depending on what errors are tolerable for the application.

TABLE 1 | Average error of the MAP estimates.

	λ	η	α	r	Var[Z]
$n = 10$	20.35	20.26	44.94	49.54	64.46
$n = 25$	5.82	9.61	15.82	22.01	29.65
$n = 50$	4.50	3.20	9.72	12.05	11.25
$n = 100$	3.32	2.72	8.62	7.28	12.97

The table reports the relative error (in per cent) with respect to the true parameters averaged over 10 different realizations of the dataset for different numbers n of replicates.

3.3. Optimal Control

The identified model can be used to derive optimal control strategies for pest control. Specifically, we suppose in the following that we can influence the system in Equation (7) by means of two different control strategies: pesticides and release of sterile pest individuals. While pesticides are probably the most used strategy for pest control, they present some disadvantages, as for example progressive reduction of efficiency, negative impact on beneficial insect populations (as pest natural enemies) or chemical residues in crops and in the ecosystem (Rafikov and Balthazar, 2005). For these reasons, the use of complementary or alternative biological control approaches has been suggested (Bhattacharyya and Bhattacharya, 2006; Greenman and Norman, 2007; Vreysen et al., 2007). The release of sterile insects, in particular, is a biological control method that aims at reducing the pest population size by introducing in the ecosystem sterile insects, usually male, that compete with the wild type for reproduction (Dyck et al., 2005). The desired effect is thus achieved as a result of the fact that females mating with sterile males will have no offspring. In other words sterile releases reduce the number of pest individuals available for reproduction and hence the birth rate. This approach has been successfully employed, for example, to eradicate screwworm flies, melon flies, the codling moth and pink bollworm among others, see Barclay and Li (1991) and references therein. Pesticide and sterile release are very different strategies also from an economical perspective: while the cost of pesticide is proportional to the area that has to be treated, the cost of steriles depends on the amount released. We notice that, in order to prevent a given fraction of the population from reproduction, the amount of released steriles should be roughly proportional to the number of pest individuals present in the system. Finally, the two approaches differ in the effect that they have on the ecosystem. In the following, we model the effect of pesticide as an instantaneous state reset of the pest population to $N(t^+) = (1 - u_p(t))N(t^-)$, where $u_p(t) \in [0, 1]$ is the percentage of the field treated with pesticide at time t. Note that since $X(t)$ is a stochastic process, this reset influences all the moments and cross-moments of $X(t)$ involving the random

process $N(t)$. For example, for the ith moment of $N(t)$, that is $\mu_i^N(t) := \mathbb{E}\left[N^i(t)\right]$, we get $\mu_i^N(t^+) = (1-u_p(t))^i \mu_i^N(t^-)$ in Equation (5). To model the effect of the release of sterile individuals we need to include them as a new species $S(t)$. In particular, we model the interaction between healthy and sterile individuals by

$$\begin{aligned} N + S &\xrightarrow{\kappa_1} B \\ B &\xrightarrow{\kappa_2} N + S \\ B + C &\xrightarrow{\eta} C \\ S + C &\xrightarrow{\eta} C. \end{aligned} \tag{10}$$

Note that we assume that the steriles have the same death rate as the healthy individuals, but they cannot reproduce. The interaction between the two species is modeled by assuming that each sterile individual can prevent one healthy individual from reproducing for a random time period during which both the individuals can die. Accordingly, the introduction of steriles effectively reduces the birth rate of the population. This is captured by $B(t)$, which quantifies how many of the healthy individuals cannot reproduce at a certain time. If we denote by $u_s(t) \in [0, 1]$ the percentage of steriles introduced at time t, with respect to an assumed maximal number of steriles $\bar{S}$ that can be introduced, we can again model this control action as a state reset of the extended model. Specifically, let $Q(t) := u_s(t)\bar{S}$ be the deterministic amount of steriles added at time t, then $S(t^+) = S(t^-) + Q(t)$ and $C(t^+) = C(t^-) + Q(t)$. Note that $C(t)$ is updated as well since we assume that sterile individuals are also damaging the field. Again, the state reset action leads to an update of the moment equations: for example, for the moments involving only $S(t)$, we get $\mu_i^S(t^+) = E\left[S(t^+)^i\right] = E\left[(S(t^-)+Q)^i\right] = \sum_{h=0}^{i} \binom{i}{h} \cdot \mu_h^S(t^-) \cdot Q^{i-h}$. Overall, the effect of the two control actions is to reset the state of the extended stochastic system, leading to a hybrid model.

One of the main problems in pest management is to determine what combinations of the available treatments are most effective for maintaining the infesting population below a given economic threshold, as described in Barclay and Li (1991), (or eventually eradicating the invasion) while minimizing the economic cost. Specifically, consider a given time horizon $\mathcal{T} = [0, T]$ and suppose that control actions can be taken at discrete time intervals $t_h = h\Delta_{T_{ac}}$ where $h = 1, \ldots, H$ with $H := \lfloor T/\Delta_{T_{ac}} \rfloor$. The optimal pest management problem can be stated as the following optimization problem

$$\begin{aligned} \min_{\substack{u_p(t_h), u_s(t_h) \\ h=1,\ldots,H}} \quad & \sum_{h=1}^{H} \left[\rho^P u_p(t_h) A + \rho^S u_s(t_h) \bar{S}\right] + \rho^C \mu_1^C(T) \\ \text{s.t.} \quad & P\left[N(t) > \xi\right] \le \delta, \quad \forall t \in [0, T], \end{aligned} \tag{11}$$

where $\rho^P > 0$ models the cost of pesticide per area (possibly including a disincentive to penalize the use of pesticide with respect to biological control), A is the total area of the habitat, $\rho^S > 0$ is the cost per sterile and $\rho^C > 0$ is a factor that translates the expected value of the process C at the end of the period into an economic cost due to damage to the field and hence decrement in productivity. The parameter ξ represents the economic threshold below which the pest should be contained. Note that since the considered model is stochastic it is not possible to guarantee that, for a given control strategy, all the realizations of the process will be below the economic threshold. We can however impose that the constraint should be satisfied with a given probability $1 - \delta$, that is, the constraint should be satisfied in $100(1-\delta)\%$ of the realizations. In the following we assume that the average population starts below the economic threshold ξ. If this was not the case, for example due to on-going infestations, one could substitute the constraint in Equation (11) with a time-varying decreasing threshold $\xi(t)$, which is higher at the beginning (to guarantee the feasibility of the optimization problem) and eventually reaches the desired threshold ξ.

The problem in Equation (11) is a chance constrained optimal control problem and is in general very difficult to solve (Prékopa, 1995). Some possible approaches are based on sampling techniques (Vapnik and Chervonenkis, 1971; Tempo et al., 2012; Grammatico et al., in press) or convex relaxations (see Nemirovski and Shapiro, 2006 and references therein). Here, we decided to solve a simplified version of the problem in Equation (11) by assuming that the distribution of the pest $N(t)$ is approximately Gaussian. If this is the case, we can rewrite the constraint $P\left[N(t) > \xi\right] \le \delta$ in terms of mean and variance of the stochastic process $N(t)$ as follows

$$\begin{aligned} \min_{\substack{u_p(t_h), u_s(t_h) \\ h=1,\ldots,H}} \quad & \sum_{h=1}^{H} \left[\rho^P u_p(t_h) A + \rho^S u_s(t_h) \bar{S}\right] + \rho^C \mu_1^C(T) \\ \text{s.t.} \quad & \mu_1^N(t) + \Phi^{-1}(1-\delta) \cdot \sigma^N(t) \le \xi, \quad \forall t \in [0, T], \end{aligned} \tag{12}$$

where $\Phi(\cdot)$ is the cumulative distribution function of a normalized Gaussian random variable[4] and $\sigma^N(t) := \sqrt{\mu_2^N(t) - \mu_1^N(t)^2}$ is the standard deviation of the process $N(t)$ (Boyd and Vandenberghe, 2004, p. 157; Nemirovski and Shapiro, 2006). Note that the problem in Equation (12) depends on the moments of the stochastic processes $N(t)$ and $C(t)$ only. Hence it can be solved using the approximated moment equations derived in Equation (5). Given the fact that we modeled the control actions as state resets, the controlled system in Equation (5) can be thought of as a deterministic continuous-time system with discrete-time controlled jumps, leading to a hybrid optimal control problem (Branicky et al., 1998; Bensoussan and Menaldi, 2000; Shahid Shaikh and Caines, 2007).

We assume in the following that $\rho^C = \rho^P = 1$, $\rho^S = 4$, $\bar{S} = 200$, and $A = 100$. **Figure 4** reports the result of the optimization problem, solved using the function *fmincon* of Matlab, for $\delta = 0.1$, a horizon of 6 weeks, $\Delta_{T_{ac}} = 1$ week and deterministic initial state $N(0) = C(0) = 10$. The lowest possible economic threshold that guarantees feasibility of the problem in

[4] If no information is known about the distribution, Chebyshev's inequality can be used to enforce the constraint $P\left[N(t) > \xi\right] \le \delta$, leading to the more restrictive bound $\mu_1^N(t) + \sqrt{\frac{1}{\delta}} \cdot \sigma^N(t) \le \xi$.

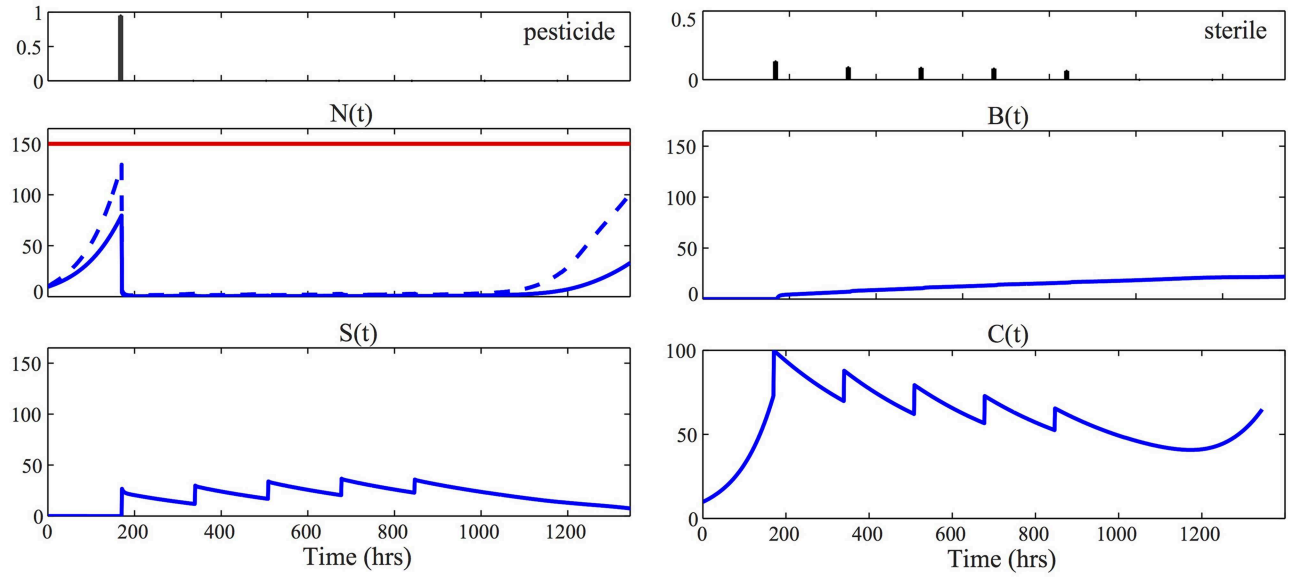


FIGURE 4 | Optimal control strategy. The two panels at the top visualize the optimal control strategy $u_p^\star, u_s^\star$, solution of Equation (12). For all the other panels, the solid blue line represents the time evolution of the mean of the corresponding process, [respectively, $N(t), S(t), B(t)$, and $C(t)$], according to the optimal control laws. To obtain these results we solved the approximated moment Equations (5) using the identified parameters $\hat{\gamma}_{MAP}$, given in Equation (9), together with $\kappa_1 = \kappa_2 = 0.01$. In the plot of the healthy population $N(t)$, the dashed line represents the quantity $\mu_1^N(t) + \Phi^{-1}(0.9) \cdot \sigma^N(t)$. The fact that this line is below the economic threshold $\xi = 150$ (red line) guarantees that the constraint of the problem in Equation (12) is satisfied. Consequently the stochastic realizations of the pest population are contained below the economic threshold, that is $N(t) < \xi$, with probability $100(1 - \delta) = 90\%$.

Equation (12), given the chosen initial condition and the fact that the first intervention is after 1 week, is 130. For the results of **Figure 4**, we fixed $\xi = 150$. To find the global minimum of the non-convex problem in Equation (12) we restarted the optimization from 10 different random initial control vectors and then selected the strategy with minimum cost. For solving the moment equations in Equation (12), we used the identified parameters $\hat{\gamma}_{MAP}$ given in Equation (9) and we set the unknown parameters $\kappa_1 = \kappa_2 = 0.01$. The resulting optimal control laws, $u_p^\star, u_s^\star$, consist of applying pesticide to almost all the crop field at the first possible control time $t_1 = 168$ hrs and subsequently controlling the population with consecutive sterile releases. This result is consistent with the analysis reported in Barclay and Li (1991), where it was shown that releasing steriles is economically preferable when the infesting population is low, while if the population is high, pesticide has to be preferred.

In order to test the performance of the derived control law, we simulated the behavior of the model in Equations (7) and (10), according to the real parameters given in Equation (8), using the stochastic simulation algorithm. Specifically, we performed 500 simulations and reported in **Figure 5** the median (blue line) and the probability distribution of $N(t)$ and $C(t)$. **Figure 5A** illustrates the behavior of the system if no control action is taken. We see that in this case the pest population exceeds the economical threshold $\xi = 150$. In **Figure 5B**, on the other hand, we see that the computed optimal control laws $u_p^\star, u_s^\star$ successfully regulate all the possible realizations, maintaining the population well-under the given threshold. This result is obtained by applying pesticide at time $t_1 = 168$ h and then using only sterile release. To show that a single application of pesticide at the beginning of the horizon would not suffice to control the population, we show in **Figure 5C** the behavior of the system if only the optimal control law $u_p^\star$ is applied. From the result it appears that the use of steriles is fundamental to complement the effect obtained by applying $u_p^\star$.

4. Discussion

One of the major future challenges for ecologists is to find strategies for organizing human interactions with the environment in a long-term sustainable way. We believe that dynamical models, inferred from measured data, together with optimal control theory have the potential to be of substantial help in achieving this task. While it is usually straightforward to incorporate the effect of human actions in a model and to formulate related optimal control problems, solving these problems may be a challenge. This has however not prevented the advancement of control theory across all engineering disciplines. So why is it that optimal control has not found more applications in ecology? A reason for this is the intrinsic complexity of ecological systems. Contrary to engineering disciplines, where the systems usually have known structure and parameters, ecological systems have been shaped by evolution in ways that we do not yet fully understand. Therefore, before we can attempt to control an ecological system with the help of a mathematical model, we first need to identify an appropriate model and its parameters from measured field data. This task is complicated by the fact that ecological systems are inherently driven by

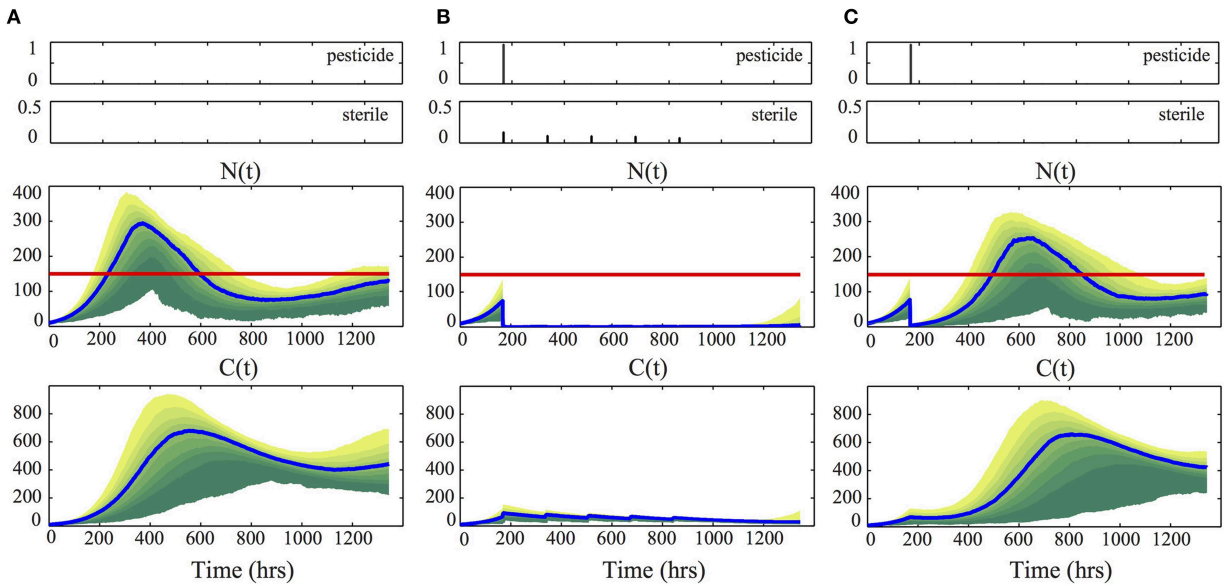


FIGURE 5 | Control performances. For each column, the performance of the control strategy shown at the top is illustrated. Specifically, the middle and bottom panels visualize the probability distribution of $N(t)$ and $C(t)$, obtained from 500 stochastic simulations, using the real parameters given in Equation (8) and $\kappa_1 = \kappa_2 = 0.01$. The blue line denotes the median of the distribution (i.e., 50% of the realizations are below this line), the nested colored regions represent the cumulative distribution of $N(t)$ and $C(t)$, with steps of 10% (i.e., 10% of the realizations are inside the dark green region and 90% are inside the yellow one). The red line represent the economical threshold $\xi = 150$. **(A)** refers to no control action, **(B)** to the optimal strategy $u_p^\star, u_s^\star$, and **(C)** to the use of $u_p^\star$, only.

random interactions between individuals that take place in spatially structured habitats and may be influenced by different environmental conditions. To be applicable in practice, a model should take into account all these factors.

In this paper, we took a first step in this direction by proposing an approach for dealing with stochasticity and varying environmental conditions, neglecting the spatial aspect of the problem. Our approach requires data from many different replicates of the system. In the current paper we used a simulated dataset; as future work it is important to test this method on real data. These may not be straightforward to obtain in some applications, as for the pest control application considered here. However, one could envision grouping a habitat into many small and clearly separated patches. Each patch could then provide a replicate of the system. Another possibility would be to "zoom out" and regard the model not as a model for one specific habitat (i.e., one specific crop), but as a model for all habitats of this kind (i.e., all cotton crops), for instance in an entire country. Based on the identified model, we formulated an optimal control problem in which two different control strategies can be used to reduce the pest population, both resulting in a state reset. This lead to a hybrid system in which the control operator can reset the state at certain given intervention times, while the dynamics of the system between the reset times are given by a continuous-time stochastic process.

All of the results of this paper rely on the possibility to obtain good approximations of the moments of the species abundances, that is, on the existence of an adequate moment closure method for the studied system. In some applications, it can happen that all the available methods do not perform adequately. To address this issue, further work is required on the development of new moment closure methods. From a hybrid systems perspective, an appealing approach is to group the interacting species into highly and lowly abundant species and to model them using continuous deterministic and discrete stochastic dynamics, respectively. Methods to analyze such hybrid models have been developed recently (Jahnke, 2011), but their use for parameter inference or optimal control has so far not been documented.

Author Contributions

FP and JR designed and performed the research. All authors wrote the paper.

Funding

The research leading to these results has received funding from the People Programme (Marie Curie Actions) of the European Union's Seventh Framework Programme (FP7/2007-2013) under REA grant agreement No. [291734] and from SystemsX under the project SignalX.

Acknowledgments

The authors would like to acknowledge contributions from Baptiste Mottet who performed preliminary analysis regarding parameter inference for the considered case study in a student project (Mottet, 2014/2015).

References

Altermatt, F., Fronhofer, E. A., Garnier, A., Giometto, A., Hammes, F., Klecka, J., et al. (2014). Big answers from small worlds: a user's guide for protist microcosms as a model system in ecology and evolution. *Methods Ecol. Evol.* 6, 218–231. doi: 10.1111/2041-210X.12312

Balazsi, G., van Oudenaarden, A., and Collins, J. (2011). Cellular decision making and biological noise: from microbes to mammals. *Cell* 144, 910–925. doi: 10.1016/j.cell.2011.01.030

Barclay, H. J., and Li, C. (1991). Combining methods of pest control: minimizing cost during the control program. *Theor. Popul. Biol.* 40, 105–123.

Bensoussan, A., and Menaldi, J. L. (2000). Stochastic hybrid control. *J. Math. Anal. Appl.* 249, 261–288. doi: 10.1006/jmaa.2000.7102

Bhattacharyya, S., and Bhattacharya, D. (2006). Pest control through viral disease: mathematical modeling and analysis. *J. Theor. Biol.* 238, 177–197. doi: 10.1016/j.jtbi.2005.05.019

Black, A., and McKane, A. (2012). Stochastic formulation of ecological models and their applications. *Trends Ecol. Evol.* 27, 337–345. doi: 10.1016/j.tree.2012.01.014

Boyd, S., and Vandenberghe, L. (2004). *Convex Optimization*. Cambridge, UK: Cambridge University Press.

Branicky, M. S., Borkar, V. S., and Mitter, S. K. (1998). A unified framework for hybrid control: model and optimal control theory. *IEEE Trans. Automat. Control* 43, 31–45.

Carrara, F., Giometto, A., Seymour, M., Rinaldo, A., and Altermatt, F. (2014). Experimental evidence for strong stabilizing forces at high functional diversity of aquatic microbial communities. *Ecology* 96, 1340–1350. doi: 10.1890/14-1324.1

Cordero, O., Ventouras, L., DeLong, E., and Polz, M. (2012). Public good dynamics drive evolution of iron acquisition strategies in natural bacterioplankton populations. *Proc. Natl. Acad. Sci. U.S.A.* 109, 20059–20064. doi: 10.1073/pnas.1213344109

DeAngelis, D. L., and Mooij, W. M. (2005). Individual-based modeling of ecological and evolutionary processes. *Annu. Rev. Ecol. Evol. Syst.* 36, 147–168. doi: 10.1146/annurev.ecolsys.36.102003.152644

Dyck, V. A., Hendrichs, J., and Robinson, A. S. (2005). *Sterile Insect Technique*. Dordrecht: Springer.

Gillespie, C., and Golightly, A. (2010). Bayesian inference for generalized stochastic population growth models with application to aphids. *J. R. Stat. Soc. C* 59, 341–357. doi: 10.1111/j.1467-9876.2009.00696.x

Gillespie, D. (1976). A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *J. Comput. Phys.* 22, 403–434. doi: 10.1016/0021-9991(76)90041-3

Goutsias, J., and Jenkinson, G. (2013). Markovian dynamics on complex reaction networks. *Phys. Rep.* 529, 199–264. doi: 10.1016/j.physrep.2013.03.004

Grammatico, S., Zhang, X., Margellos, K., Goulart, P., and Lygeros, J. (in press). A scenario approach for non-convex control design. *IEEE Trans. Automat. Control*. doi: 10.1109/TAC.2015.2433591

Greenman, J. V., and Norman, R. A. (2007). Environmental forcing, invasion and control of ecological and epidemiological systems. *J. Theor. Biol.* 247, 492–506. doi: 10.1016/j.jtbi.2007.03.031

Hekstra, D. R., and Leibler, S. (2012). Contingency and statistical laws in replicate microbial closed ecosystems. *Cell* 149, 1164–1173. doi: 10.1016/j.cell.2012.03.040

Hespanha, J. (2008). "Moment closure for biochemical networks," in *3rd International Symposium on Communications, Control and Signal Processing* (St Julians), 142–147.

Isham, V. (1991). Assessing the variability of stochastic epidemics. *Math. Biosci.* 107, 209–224. doi: 10.1016/0025-5564(91)90005-4

Jahnke, T. (2011). On reduced models for the chemical master equation. *Multiscale Model. Simul.* 9, 1646–1676. doi: 10.1137/110821500

Jaquette, D. (1970). A stochastic model for the optimal control of epidemics and pest populations. *Math. Biosci.* 8, 343–354.

Krishnarajah, I., Cook, A., Marion, G., and Gibson, G. (2005). Novel moment closure approximations in stochastic epidemics. *Theor. Popul. Biol.* 67, 855–873. doi: 10.1016/j.bulm.2004.11.002

Kügler, P. (2012). Moment fitting for parameter inference in repeatedly and partially observed stochastic biological models. *PLOS ONE* 7:e43001. doi: 10.1371/journal.pone.0043001

Lagarrigues, G., Jabot, F., Lafond, V., and Courbaud, B. (2014). Approximate Bayesian computation to recalibrate individual-based models with population data: illustration with a forest simulation model. *Ecol. Modell.* 306, 278–286. doi: 10.1016/j.ecolmodel.2014.09.023

Lavielle, M. (2014). *Mixed Effects Models for the Population Approach: Models, Tasks, Methods and Tools*. Boca Raton, FL: Chapman & Hall; CRC.

Marion, G., Renshaw, E., and Gibson, G. (1998). Stochastic effects in a model of nematode infection in ruminants. *Math. Med. Biol.* 15, 97–116. doi: 10.1093/imammb/15.2.97

Matis, J., Kiffe, T., Matis, T., and Stevenson, D. (2007). Stochastic modeling of aphid population growth with nonlinear power-law dynamics. *Math. Biosci.* 208, 469–494. doi: 10.1016/j.mbs.2006.11.004

McKinley, T., Cook, A., and Deardon, R. (2009). Inference in epidemic models without likelihoods. *Int. J. Biostat.* 5, 24. doi: 10.2202/1557-4679.1171

Milner, P., Gillespie, C., and Wilkinson, D. (2013). Moment closure based parameter inference of stochastic kinetic models. *Stat. Comput.* 23, 287–295. doi: 10.1007/s11222-011-9310-8

Mottet, B. Q. (2014/2015). *Application of Moment-Based Methods for Parameter Inference to Nonlinear Stochastic Population Growth Model*. Technical Report, Semester Project, ETH Zurich. Supervisors: F. Parise, J. Ruess, and J. Lygeros.

Munsky, B., and Khammash, M. (2006). The finite state projection algorithm for the solution of the chemical master equation. *J. Chem. Phys.* 124:044104. doi: 10.1063/1.2145882

Murray, J. D. (2002). "Mathematical biology I. An introduction," in *Interdisciplinary Applied Mathematics*, Vol. 17, eds S. S. Antman, J. E. Marsden, L. Sirovich, and S. Wiggins (New York, NY: Springer), 1–537.

Nemirovski, A., and Shapiro, A. (2006). Convex approximations of chance constrained programs. *SIAM J. Optimiz.* 17, 969–996. doi: 10.1137/050622328

Neuert, G., Munsky, B., Tan, R., Teytelman, L., Khammash, M., and van Oudenaarden, A. (2013). Systematic identification of signal-activated stochastic gene regulation. *Science* 339, 584–587. doi: 10.1126/science.1231456

Nåsell, I. (2002). Stochastic models of some endemic infections. *Math. Biosci.* 179, 1–19. doi: 10.1016/S0025-5564(02)00098-6

Nåsell, I. (2003). Moment closure and the stochastic logistic model. *Theor. Popul. Biol.* 63, 159–168. doi: 10.1016/S0040-5809(02)00060-6

Ovaskainen, O., and Cornell, S. (2006). Space and stochasticity in population dynamics. *Proc. Natl. Acad. Sci. U.S.A.* 103, 12781–12786. doi: 10.1073/pnas.0603994103

Ovaskainen, O., and Meerson, B. (2010). Stochastic models of population extinction. *Trends Ecol. Evol.* 25, 643–652. doi: 10.1016/j.tree.2010.07.009

Poovathingal, S., and Gunawan, R. (2010). Global parameter estimation methods for stochastic biochemical systems. *BMC Bioinformatics* 11:414. doi: 10.1186/1471-2105-11-414

Prékopa, A. (1995). *Stochastic Programming*. Dordrecht: Kluwer Academic Publishers.

Rafikov, M., and Balthazar, J. M. (2005). Optimal pest control problem in population dynamics. *Comput. Appl. Math.* 24, 65–81. doi: 10.1590/S1807-03022005000100004

Railsback, S., and Grimm, V. (2012). "Individual-based ecology," in *Encyclopedia of Theoretical Ecology*, eds A. Hastings and L. J. Gross (Berkeley: University of California Press), 365–370.

Ross, J., Pagendam, D., and Pollett, P. (2009). On parameter estimation in population models II: multi-dimensional processes and transient dynamics. *Theor. Popul. Biol.* 75, 123–132. doi: 10.1016/j.tpb.2008.12.002

Ruess, J., and Lygeros, J. (2013). "Identifying stochastic biochemical networks from single-cell population experiments: a comparison of approaches based on the Fisher information," in *IEEE 52nd Annual Conference on Decision and Control (CDC)* (Florence), 2703–2708.

Ruess, J., and Lygeros, J. (2015). Moment-based methods for parameter inference and experiment design for stochastic biochemical reaction networks. *ACM Trans. Model. Comput. Simul.* 25:8. doi: 10.1145/2688906

Ruess, J., Milias-Argeitis, A., and Lygeros, J. (2013). Designing experiments to understand the variability in biochemical reaction networks. *J. R. Soc. Interface* 10:20130588. doi: 10.1098/rsif.2013.0588

Ruess, J., Milias-Argeitis, A., Summers, S., and Lygeros, J. (2011). Moment estimation for chemically reacting systems by extended Kalman filtering. *J. Chem. Phys.* 135:165102. doi: 10.1063/1.3654135

Shahid Shaikh, M., and Caines, P. E. (2007). On the hybrid optimal control problem: theory and algorithms. *IEEE Trans. Automat. Control* 52, 1587–1603. doi: 10.1109/TAC.2007.904451

Singh, A., and Hespanha, J. (2006). "Lognormal moment closures for biochemical reactions," in *45th IEEE Conference on Decision and Control* (San Diego, CA), 2063–2068. doi: 10.1109/CDC.2006.376994

Srivastava, D. S., Kolasa, J., Bengtsson, J., Gonzalez, A., Lawler, S. P., Miller, T. E. et al. (2004). Are natural microcosms useful model systems for ecology? *Trends Ecol. Evol.* 19, 379–384. doi: 10.1016/j.tree.2004.04.010

Stumpf, M. (2014). Approximate Bayesian inference for complex ecosystems. *F1000Prime Rep.* 6:60. doi: 10.12703/P6-60

Tempo, R., Calafiore, G., and Dabbene, F. (2012). *Randomized Algorithms for Analysis and Control of Uncertain Systems: With Applications*. London: Springer Science & Business Media.

Toni, T., Welch, D., Strelkowa, V., Ipsen, A., and Stumpf, M. (2009). Approximate Bayesian computation scheme for parameter inference and model selection in dynamical systems. *J. R. Soc. Interface* 6, 187–202. doi: 10.1098/rsif.2008.0172

Vapnik, V. N., and Chervonenkis, A. Y. (1971). On the uniform convergence of relative frequencies of events to their probabilities. *Theory Probab. Appl.* 16, 264–280.

Vreysen, M. J., Robinson, A., and Hendrichs, J. (2007). *Area-wide Control of Insect Pests: From Research to Field Implementation*. Dordrecht: Springer Science & Business Media.

Whittle, P. (1957). On the use of the normal approximation in the treatment of stochastic processes. *J. R. Stat. Soc. B* 19, 268–281.

Wickwire, K. (1977). Mathematical models for the control of pests and infectious diseases: a survey. *Theor. Popul. Biol.* 11, 182–238.

Wolf, V., Goel, R., Mateescu, M., and Henzinger, T. (2010). Solving the chemical master equation using sliding windows. *BMC Syst. Biol.* 4:42. doi: 10.1186/1752-0509-4-42

Zechner, C., Ruess, J., Krenn, P., Pelet, S., Peter, M., Lygeros, J., et al. (2012). Moment-based inference predicts bimodality in transient gene expression. *Proc. Natl. Acad. Sci. U.S.A.* 109, 8340–8345. doi: 10.1073/pnas.1200161109

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Environmental Science

OPINION
published: 30 April 2015
doi: 10.3389/fenvs.2015.00033

Machine learning components in deterministic models: hybrid synergy in the age of data

Evan B. Goldstein[1]* *and Giovanni Coco*[2]

[1] Department of Geological Sciences, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA, [2] School of Environment, University of Auckland, Auckland, New Zealand

Keywords: hybrid model, machine learning, data driven modeling, hybrid modeling, model theory

"When the information changes, I change my mind. What do you do sir?"
–paraphrased from Paul Samuelson (1970)[1].

Introduction

Physics-based numerical models designed to study processes on the surface of the Earth are commonly built with conservation laws. Yet conservations laws are based on a treatable subset of all physical and biological processes operating in the system of interest, and empirical relationships are always required to fully describe systems (and provide mathematical "closure"). In contrast to conservation laws, empirical expressions used in model construction are inductive, and based on observations of phenomena.

Any model that requires empirical expressions is also subject to revision as the empirical parameterization is refined: all empiricism is open to revision by corroborating or conflicting data. As more data becomes available, and more degrees of freedom are explored, it becomes harder to incorporate all available data into a single optimal empirical predictor. We argue in this contribution that empirical parameterizations for numerical models should be constructed using machine learning techniques because these techniques are built to operate on large, high dimensional datasets. Machine learning, in the context of this paper, defines a suite of algorithms used to develop predictive relationships (correlations) using a set of input data. Examples of commonly used machine learning techniques in the Earth sciences are artificial neural networks (e.g., Maier and Dandy, 2000; Pape et al., 2007; van Maanen et al., 2010), regression trees (e.g., Snelder et al., 2009; Oehler et al., 2012), Bayesian networks (e.g., Aguilera et al., 2011; Yates and Le Cozannet, 2012), and evolutionary algorithms (e.g., Knaapen and Hulscher, 2002; Ruessink, 2005; Goldstein et al., 2013). Machine learning techniques offer insight and are high-performance, reproducible, and scalable. The inclusion of powerful inductive techniques like those provided by machine learning offers the opportunity to enhance the predictions obtained from deductive approaches. Additionally the use of machine learning often leads to further insight. The use of empiricisms built from machine learning in a physics-based model results in a "hybrid" model.

Combining the strengths of inductive (data-driven) and deductive (physics-based) approaches in a single hybrid model has been suggested as a valuable step forward in model development because of increases in accuracy (e.g., Babovic et al., 2001; Hall, 2004) and speed (Krasnopolsky and Fox-Rabinovitz, 2006). An additional benefit is the direct coupling of models to data,

[1] This quote (and a discussion of its history) is available as a long-lived URL through the 'Internet Archive': https://web.archive.org/web/20150305151453/http://quoteinvestigator.com/2011/07/22/keynes-change-mind/comment-page-1/

OPEN ACCESS

Edited by:
Christian E. Vincenot, Kyoto University, Japan

Reviewed by:
Mikhail Kanevski, University of Lausanne, Switzerland
Nathaniel K. Newlands, Science and Technology, Government of Canada, Canada

***Correspondence:**
Evan B. Goldstein, evan.goldstein@unc.edu

Specialty section:
This article was submitted to Environmental Informatics, a section of the journal Frontiers in Environmental Science

Received: *05 March 2015*
Paper pending published: *29 March 2015*
Accepted: *16 April 2015*
Published: *30 April 2015*

Citation:
Goldstein EB and Coco G (2015) Machine learning components in deterministic models: hybrid synergy in the age of data. Front. Environ. Sci. 3:33. doi: 10.3389/fenvs.2015.00033

especially valuable for exploratory models (Murray, 2003). Several studies have already used this approach, embedding machine learning components directly in models (e.g., Jain and Srinivasulu, 2004; Corzo et al., 2009; Goldstein et al., 2014; Limber et al., 2014) or during model calibration (Knaapen and Hulscher, 2002, 2003; Ruessink, 2005). The focus of this article is on machine learning components that are directly embedded within physics-based models. We believe the further use of this "hybrid" approach will result in more accurate model components (i.e., individual empirical predictors) and more accurate models. We do not intend to cast machine learning solely as a data fitting procedure. There is a rich set of problems where machine learning is applicable, but a clear use is in optimizing the fit of empirical predictors in nonlinear multivariate datasets.

Five main points highlight the advantages of this hybrid approach: These benefits are not without costs, mostly in the form of extra work/time and problems associated with less-rigorous usage of machine learning. As a convention we refer to machine learning model components as "parameterizations" and the hybrid model as the "model."

Machine Learning Highlights "Theory-Less" or "Data-Less" Model Components

Building machine learning predictors is usually motivated by a lack of theory or a perceived inadequacy in theory. Parameterizations that do not have an accurate or well-developed "theory" might negatively reflect upon model predictions because they are incorrect, or poorly parameterized. As an example, Goldstein et al. (2013) built a new predictor for bedforms generated under wave action motivated by the fact that no predictors were explicitly tested in conditions of large grain size variation and strong wave conditions. This predictor was eventually used in a larger "hybrid" numerical model with success (Goldstein et al., 2014).

Machine learning predictors also highlight heuristic or theoretical elements of a numerical model that do not have sufficient data to test. Both types of problems (lacking theory and lacking data) can motivate future research, specifically theory creation and targeted data collection.

Machine Learning Can Be Used to Gain New Theoretical Insight

Machine learning techniques on a dataset may provide theoretical insight. Crutchfield (2014) has termed this process "artificial science," where theoretical insight is derived directly from data. New machine learning techniques suggest that this is possible (Schmidt and Lipson, 2009). In this way developing predictors might provide theoretical insight into model or system behavior, at the very least giving new hypotheses to test. The wave ripple predictor developed by Goldstein et al. (2013) provides a new, inductively derived predictor for ripples under various grain sizes and forcing conditions. This new mathematical relationship describes an observed relationship (between grain size, wave forcing, and ripple size) derived from observations. This new relationship (an inductive statement) provides a testable hypothesis.

The Possibility for Emulation

Beyond detecting a new theory, hypothesis, or physical relationship, machine learning may provide a more parsimonious empirical relation than existed previously (e.g., Tinoco et al., 2015). This could help to speed up model runtime, a main goal of previous hybrid model work where the use of artificial neural networks allowed for emulation of entire components of a global climate model based on physical processes with no accuracy loss (e.g., Krasnopolsky and Fox-Rabinovitz, 2006). The computational gain associated with the use of a hybrid model cascades into a series of additional advantages including the possibility of simulating more scenarios, decreasing grid size or exploring finer-scale parameterizations.

Machine Learning Outperforms Arbitrary Curve Fitting

Multidimensional empirical parameterization are often built by assembling the data, collapsing the data to a two- dimensional plane, and fitting a user–defined function through the data cloud. Several steps require "user input," and may be arbitrary. First, the collapse of the multidimensional data onto a 2D plane may require developing nondimensional groups. Though the number of nondimensional groups is mandated by the well-known Buckingham's Pi theorem, the actual makeup of each group is not, and is often guided by utility, physical reasoning or user intuition (e.g., Bridgman, 1922). Second, a user defined curve must be selected to fit to the data. This curve may not be the most optimal basis function to fit to the data.

Both of these ambiguities are avoided in machine learning because: (1) user input can be the raw data parameters, or all possible nondimensional parameter groupings (e.g., Tinoco et al., 2015) and (2) machine learning often does not require the use of a set basis function, or the basis function is sufficiently flexible to allow the approximation of any arbitrary function. These benefits suggest that machine learning is a powerful set of tools for developing parameterizations when data is high-dimensional, noisy, and nonlinear: these techniques outperform traditional curve fitting for their ability to truly provide an optimal curve to be fit to data.

Machine Learning is Reproducible and Scalable

Machine learning is inherently reproducible if the methodology is clearly described and the data is open and available. Reproducibility relies strongly on researchers to provide the exact data used as training data (to "teach" the learner) and those used to test the model. Any specific initialization is also required.

Because machine learning techniques are repeatable, they are also scalable. As new data is collected, it can be integrated into the machine learning routine to develop a new, more optimal

predictor. Not all new data is equally relevant, and Bowden et al. (2012) present a technique to determine if the new data extends the range of the predictor.

Caveats and Open Problems

Several issues remain when using machine learning. Predictors can become overfit if too much data is shown to the learner, or the optimization routine is performed without bounds. Proper time to halt an optimization, and other relevant methods to avoid overfitting, are topics of active research (e.g., Maier and Dandy, 2000; Schmidt and Lipson, 2009; O'Neill et al., 2010; Tuite et al., 2011). Users should invest energy and time to mine the literature for these techniques.

Often data used to train the model is not selected optimally. We have previously advocated a deliberate sampling strategy to select data from the entire range of phase space available (e.g., Goldstein and Coco, 2014; Tinoco et al., 2015), as have others (e.g., Bowden et al., 2002; May et al., 2010; Wu et al., 2013). Predictors tend to be less overfit and more optimal when sampling is a considered process. This step adds extra work (especially to thoroughly document the process for repeatability), but we believe it is needed to develop the most optimal predictor.

Operators of machine learning algorithms should be experts in the data being examined. Nonphysical predictors can often appear as a result of regular usage of machine learning, data and/or computational errors. These erroneous results must be understood and manually discarded. We acknowledge that this adds a level of subjectivity to the analysis, but this subjectivity is also present in traditional empirical techniques (e.g., why did a researcher choose to fit the data using one function vs. another?). As a result, thorough examination of the physical correctness of the predictor should be performed, and expert knowledge should be exercised before machine learning results are accepted as correct and inserted into a hybrid model.

When combining machine learning components with physics-based model components users should be wary of the general structure of the predictor, and the potential for competing or mismatched nonlinearities in model components. We have personally encountered the mismatch between machine learning derived and theoretical components in a numerical model (Goldstein et al., 2014). This mismatch initially restricted our ability to understand sensitivity over a broad range of parameter values. We stress that it is always critical to understand and investigate how model components will interact.

Conclusion

Models constructed to study Earth surface processes are often intended to study large-scale, long-term phenomena. Little data may exist to parameterize long time-scale processes. However, ample data often exists for smaller space- and time- scale processes. Earth surface models should leverage all available data to build empirical parameterizations by adopting a hybrid approach.

Machine learning tools represent our best ability to process empirical data in a reproducible way. Best practices (explicit mentions of data selection and learner initialization) allows for reproducible results. The process of developing these predictors also explicitly highlights known gaps in knowledge. We believe these benefits should motivate the widespread adoption of hybrid models that combine machine learning approaches with physics-based models.

Acknowledgments

We thank R. O. Tinoco, A.B. Murray, P. W. Limber, K. Ells, and Z. Li for many enjoyable and insightful discussions, the editors for proposing this "research topic," and two reviewers for their constructive comments.

References

Aguilera, P. A., Fernández, A., Fernández, R., Rumí, R., and Salmerón, A. (2011). Bayesian networks in environmental modelling. *Environ. Model. Softw.* 26, 1376–1388. doi: 10.1016/j.envsoft.2011.06.004

Babovic, V., Canizares, R., Jensen, H. R., and Klinting, A. (2001). Neural networks as routine for error updating of numerical models. *J. Hydraul. Eng. ASCE* 127, 181–193. doi: 10.1061/(ASCE)0733-9429(2001)127:3(181)

Bowden, G. J., Maier, H. R., and Dandy, G. C. (2002). Optimal division of data for neural network models in water resources applications. *Water Resour. Res.* 38, 1010. doi: 10.1029/2001WR000266

Bowden, G. J., Maier, H. R., and Dandy, G. C. (2012). Real-time deployment of artificial neural network forecasting models: understanding the range of applicability. *Water Resour. Res.* 48, W10549. doi: 10.1029/2012WR011984

Bridgman, P. W. (1922). *Dimensional Analysis*. New Haven: Yale University Press.

Corzo, G. A., Solomatine, D. P., Hidayat, de Wit, M., Werner, M., Uhlenbrook, S., et al. (2009). Combining semi-distributed process-based and data-driven models in flow simulation: a case study of the Meuse river basin. *Hydrol. Earth Syst. Sci.* 13, 1619–1634. doi: 10.5194/hess-13-1619-2009

Crutchfield, J. P. (2014). The dreams of theory. *WIREs Comput. Stat.* 6, 75–79. doi: 10.1002/wics.1290

Goldstein, E. B., and Coco, G. (2014). A machine learning approach for the prediction of settling velocity. *Water Resour. Res.* 50, 3595–3601. doi: 10.1002/2013WR015116

Goldstein, E. B., Coco, G., and Murray, A. B. (2013). Prediction of wave ripple characteristics using genetic programming. *Cont. Shelf Res.* 71, 1–15, doi: 10.1016/j.csr.2013.09.020

Goldstein, E. B., Coco, G., Murray, A. B., and Green, M. O. (2014). Data driven components in a model of inner shelf sorted bedforms: a new hybrid model. *Earth Surf. Dynam.* 2, 67–82, doi: 10.5194/esurf-2-67-2014

Hall, J. W. (2004). Comment on 'Of data and models'. *J. Hydroinform.* 6, 75–77.

Jain, A., and Srinivasulu, S. (2004). Development of effective and efficient rainfall-runoff models using integration of deterministic, real-coded genetic algorithms and artificial neural network techniques. *Water Resour. Res.* 40, W04302. doi: 10.1029/2003WR002355

Knaapen, M. A. F., and Hulscher, S. J. M. H. (2002). Regeneration of sand waves after dredging *Coast. Eng.* 46, 277–289. doi: 10.1016/S0378-3839(02)00090-X

Knaapen, M. A. F., and Hulscher, S. J. M. H. (2003). Use of a genetic algorithm to improve predictions of alternate bar dynamics. *Water Resour. Res.* 39, 1231. doi:10.1029/2002WR001793

Krasnopolsky, V. M., and Fox-Rabinovitz, M. S. (2006). A new synergetic paradigm in environmental numerical modeling: hybrid models combining deterministic and machine learning components. *Ecol. Model.* 191, 5–18. doi: 10.1016/j.ecolmodel.2005.08.009

Limber, P. W., Murray, A. B., Adams, P. N., and Goldstein, E. B. (2014). Unraveling the dynamics that scale cross-shore headland amplitude on rocky coastlines, Part 1: model development. *JGR Earth Surf.* 119, 854–873. doi: 10.1002/2013JF 002950

Maier, H. R., and Dandy, G. C. (2000). Neural networks for the prediction and forecasting of water resources variables: A review of modelling issues and applications. *Environ. Model. Softw.* 15, 101–124. doi: 10.1016/S1364-8152(99) 00007-9

May, R. J., Maier, H. R., and Dandy, G. C. (2010). Data splitting for artificial neural networks using SOM-based stratified sampling. *Neural Netw.* 23, 283–294. doi: 10.1016/j.neunet.2009.11.009

Murray, A. B. (2003). "Contrasting the goals, strategies, and predictions associated with simplified numerical models and detailed simulations", in *Prediction in Geomorphology*, eds R. M. Iverson and P. R. Wilcock (Washington, DC: AGU, AGU Geophysical Monograph 135), 151–165.

O'Neill, M., Vanneschi, L., Gustafson, S., and Banzhaf, W. (2010). Open issues in genetic programming. *Genet. Program. Evol. M.* 11, 339–363. doi: 10.1007/s10710-010-9113-2

Oehler, F., Coco, G., Green, M. O., and Bryan, K. R. (2012). A data-driven approach to predict suspended-sediment reference concentration under non-breaking waves. *Cont. Shelf Res.* 46, 96–106. doi: 10.1016/j.csr.2011.01.015

Pape, L., Ruessink, B. G., Wiering, M. A., and Turner, I. L. (2007). Re- current neural network modeling of nearshore sandbar behavior, *Neural Netw.* 20, 509–518. doi: 10.1016/j.neunet.2007.04.007

Ruessink, B. G. (2005). Calibration of nearshore process models: application of a hybrid genetic algorithm. *J. Hydroinform.* 7, 135– 149.

Schmidt, M., and Lipson, H. (2009). Distilling free-form natural laws from experimental data. *Science* 324, 81–85. doi: 10.1126/science.1165893

Snelder, T. H., Lamouroux, N., Leathwick, J. R., Pella, H., Sauquet, E., and Shankar, U. (2009). Predictive mapping of the natural flow regimes of France. *J. Hydrol.* 373, 57–67. doi: 10.1016/j.jhydrol.2009.04.011

Tinoco, R. O., Goldstein, E. B., and Coco, G. (2015). A data-driven approach to develop physically sound predictors: application to depth-averaged velocities on flows through submerged arrays of rigid cylinders. *Water Resour. Res.* 51, 1247–1263. doi: 10.1002/2014WR016380

Tuite, C., Agapitos, A., O'Neill, M., and Brabazon, A. (2011). "Tackling Overfitting in Evolutionary-Driven Financial Model Induction," in *Natural Computing in Computational Finance*, eds A. Brabazon, M. O'Neill, and D. Maringer (Springer, Heidelberg), 141–161.

van Maanen, B., Coco, G., Bryan, K. R., and Ruessink, B. G. (2010). The use of artificial neural networks to analyze and predict alongshore sediment transport. *Nonlinear Process. Geophys.* 17, 395–404. doi: 10.5194/npg-17-395-2010

Wu, W., May, R. J., Maier, H. R., and Dandy, G. G. (2013). A benchmarking approach for comparing data splitting methods for modeling water resources parameters using artificial neural networks. *Water Resour. Res.* 49, 7598–7614. doi: 10.1002/2012WR012713

Yates, M. L., and Le Cozannet, G. (2012). Brief communication 'Evaluating European coastal evolution using Bayesian networks', *Nat. Hazards Earth Syst. Sci.* 12, 1173–1177. doi: 10.5194/nhess-12-1173-2012

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Environmental Science

ORIGINAL RESEARCH
published: 09 July 2015
doi: 10.3389/fenvs.2015.00047

Parametric linear hybrid automata for complex environmental systems modeling

Samar H. K. Tareen[1], *Jamil Ahmad*[2] *and Olivier Roux*[3]*

[1] *Maastricht Centre for Systems Biology (MaCSBio), Maastricht University, Maastricht, Netherlands,* [2] *Department of Computational Sciences, Research Centre for Modeling and Simulation, National University of Sciences and Technology, Islamabad, Pakistan,* [3] *MeForBio, Institut de Recherche en Communications et Cybernétique de Nantes, Ecole Centrale de Nantes, France*

OPEN ACCESS

Edited by:
Christian E. Vincenot, Kyoto University, Japan

Reviewed by:
Ovidiu Radulescu, Université de Montpellier 2, France
Saumitra Mukherjee, Jawaharlal Nehru University, India

***Correspondence:**
Olivier Roux, MeForBio, Institut de Recherche en Communications et Cybernétique de Nantes, Ecole Centrale de Nantes, 1 rue de la Noë - B.P.92101, Nantes, 44321, France
olivier.roux@irccyn.ec-nantes.fr

Specialty section:
This article was submitted to Environmental Informatics, a section of the journal Frontiers in Environmental Science

Received: *16 March 2015*
Accepted: *26 June 2015*
Published: *09 July 2015*

Citation:
Tareen SHK, Ahmad J and Roux O (2015) Parametric linear hybrid automata for complex environmental systems modeling. Front. Environ. Sci. 3:47. doi: 10.3389/fenvs.2015.00047

Environmental systems, whether they be weather patterns or predator–prey relationships, are dependent on a number different variables, each directly or indirectly affecting the system at large. Since not all of these factors are known, these systems take on non-linear dynamics, making it difficult to accurately predict meaningful behavioral trends far into the future. However, such dynamics do not warrant complete ignorance of different efforts to understand and model close approximations of these systems. Toward this end, we have applied a logical modeling approach to model and analyze the behavioral trends and systematic trajectories that these systems exhibit without delving into their quantification. This approach, formalized by René Thomas for discrete logical modeling of Biological Regulatory Networks (BRNs) and further extended in our previous studies as *parametric biological linear hybrid automata (Bio-LHA)*, has been previously employed for the analyses of different molecular regulatory interactions occurring across various cells and microbial species. As relationships between different interacting components of a system can be simplified as positive or negative influences, we can employ the Bio-LHA framework to represent different components of the environmental system as positive or negative feedbacks. In the present study, we highlight the benefits of hybrid (discrete and continuous combined) modeling which lead to refinements among the fore-casted behaviors in order to find out which ones are actually possible. We have taken two case studies: an interaction of three microbial species in a freshwater pond, and a more complex atmospheric system, to show the applications of the Bio-LHA methodology for the timed hybrid modeling of environmental systems. Results show that the approach using the Bio-LHA is a viable method for behavioral modeling of complex environmental systems by finding timing constraints while keeping the complexity of the model at a minimum.

Keywords: hybrid modeling, environmental systems, linear hybrid automata, René Thomas formalism, network analysis, SMBioNet, GenoTech, HyTech

1. Introduction

1.1. Environmental Systems

Environmental systems are a broad category of systems which interact or have an impact on the environment. These systems can range from different chemical processes found in nature to large scale interactions between different biotic and abiotic factors, as represented in the modeling studies of Lake Balaton (Somlydy, 1982; Ttrai et al., 2000). In contrast to more traditional systems, as studied in other physical sciences, environmental systems are tightly integrated with each other, forming different sub-domains within each system (Hanrahan, 2010). As such, they are difficult to study in isolation since their behavior changes with respect to other environmental factors working in conjunction with each other (Hanrahan, 2010).

1.2. Modeling of Environmental Systems

Because of their broad spectrum, different approaches have been employed in the modeling and analysis of environmental systems over the years. The approaches can be grouped into stochastic and probabilistic (Refsgaard et al., 2007; Gottschalk et al., 2010; Sun et al., 2014), statistical (Refsgaard et al., 2007; Uusitalo, 2007; Berie, 2014), differential (Casulli and Zanolli, 2002; Coulthard et al., 2005; Hanrahan, 2010), or a combinations of these methodologies. In recent years, different artificial intelligence (AI) approaches have also been applied, such as machine learning (Wiley et al., 2003; Kanevski et al., 2004) and agent based modeling (Sengupta and Bennett, 2003; Crooks et al., 2008). However, certain aspects of the system elude these approaches, such as the larger view of the behaviors of the system–statistical methods do not represent the dynamics of the system, probabilistic approaches lack deterministic predictability, and differential and AI models suffer from high levels of complexity when modeling realistic parameters.

1.3. Our Contribution

In this paper, we present a hybrid modeling approach using linear hybrid automata for the behavioral modeling of the system. We have already applied this approach in the domain of systems biology, particularly in the modeling of biological regulatory networks (Ahmad et al., 2012; Aslam et al., 2014). The advantage of this approach is that it allows the modeling of large regulatory systems, assisting in the inference of the dynamics of the system, without having to deal with the exact rates or parameters governing the said system. To demonstrate our approach, we have applied our framework on two case studies, a microbial system in a freshwater pond, and a slightly more complex atmospheric system (adapted from Seppelt, 2007), explained below. For the sake of readability, the given examples only deal with a small number of components, but the presented methods are applicable on larger systems as well.

1.4. Case Study 1: Microbial Population

Consider the microbial populations of three particular microbes in a freshwater pond. The first of these microbes, labeled $M1$, generates a certain product which acts as a nutrient for the second microbe, labeled $M2$. However, the microbe $M2$ produces a toxin which is harmful to the first microbe, $M1$. Apart from the toxin, the microbe $M2$ also produces a nutrient for the third microbe, $M3$. The third microbe acts as a dominant predator instead, and generates a toxin which is harmful to both the first microbe, $M1$, and the second microbe $M2$. However, in order to sustain itself, the third microbe can also produce its own nutrients, but only when it is present in sufficient numbers to form colonies.

1.5. Case Study 2: Atmospheric System

Consider a slightly simplified version of the atmosphere, particularly pertaining to the interconnectedness of the temperature of the planet and the water cycle. Whenever the sun shines, it increases the temperature, and makes evaporation possible (provided that there is a source of water available). As the temperature increases and the water evaporates, clouds begin to form which can produce two effects: (i) the sun gets blocked, lowering the temperature; and (ii) precipitation is produced which resupplies the water sources for future evaporation. However, persistent high temperatures can increase the air temperature, blocking condensation, and cloud formation. Likewise, persistent evaporation without precipitation can drain the available water source(s).

1.6. Plan of the Paper

We begin with the Methodology Section where we describe the work flow of our approach, the formalisms and frameworks used, and its step by step application on Case Study 1. Following it is the Results Section in which we apply the given methodology on Case Study 2, ending with the Discussion Section where we discuss the applicability of our approach, its advantages, and its disadvantages. A list of glossary items containing technical terms and their short descriptions are also provided at the end of this article (**Table A1**).

2. Methodology

Our methodology focuses on building an initial discrete model of the system in question, using the Kinetic Modeling formalism set forth by Kauffman in the late 1960s (Kauffman, 1969) and then by Thomas in the late 1970s (Thomas, 1978, 1979, 1998; Thomas et al., 1995) for biological regulatory networks. The basic idea of the approach was that natural phenomena are often observable when there is a switch from a relatively stable mode to another different mode.

This approach appeals for discrete modeling with states and transitions where time elapses in the states and transitions between these states are instantaneous. The important data being the time spent in one state whatever the evolution in this state actually is, abstraction can then be done in order to take into account lengthening actions besides almost infinitely fast switches. This stands for the hybrid feature. Thus, a discrete model allows us to observe the dynamics of the system, in the form of a state graph, using arbitrary discrete values to represent different levels of activities each entity or object can exhibit. Once the discrete dynamics of the system are obtained, *Parametric Biological Linear Hybrid Automata* (Bio-LHA, Ahmad et al., 2006; Ahmad, 2009) can be constructed on targeted trajectories to

isolate conjuncted parametric delay constraints governing these trajectories (Ahmad et al., 2007, 2008; Ahmad and Roux, 2010). These delays assist us in inferring different temporal constraints a system must follow to generate the specified trajectory, and which constraints can be targeted to destabilize the said trajectory. The whole process is represented in **Figure 1** and can be broken into the following steps:

- The pathways pertaining to the system are extracted from the literature and abstracted to the form of a discrete regulatory network. These pathways can comprise of simple linear processes to complex and dense interactions such as feedbacks, feed forward loops, branching processes etc.
- The regulatory network then undergoes model checking to generate sets of logical parameters which satisfy particular system specific observations extracted from the literature.
- The parameters are loaded into the regulatory network to generate discrete dynamics of the system (also called discrete state space), represented as a directed state graph.
- Network analysis, in particular the shortest path betweenness centrality calculation, is conducted for each state of the state graph to isolate specific trajectories comprising of states satisfying particular centrality constraints.
- The isolated trajectories are then converted to respective Bio-LHAs which are then used to find the delay constraints in the form of path constraints (for acyclic trajectories), invariance kernels (for cyclic trajectories), or convergence domains (for asymptotic trajectories).
- These constraints are then refined to form pairwise relations between the respective parameters governing the activation or inhibition delay of the involved entities.

In order to properly understand the requirements and application of the approach, we detail the mathematical definitions and their application on Case Study 1 (Microbial population) in the following subsections.

2.1. Discrete Modeling

The formal definitions of the René Thomas kinetic logic formalism, adapted from Thomas (1979), Thomas and d'Ari (1990), Ahmad et al. (2012), and Paracha et al. (2014), are provided below:

Definition 1. [Regulatory Network]. *A directed graph* $G = (V, E)$ *is a regulatory network (RN) when,*

- V*, with a typical element* v*, is the set of vertexes,*
- E*, with a typical element* $e = (v_m, v_n)$*, is the set of edges directed from a source vertex* v_m *to the target vertex* v_n*,*
- $G^+(v)$ *represents the set of the targets of the vertex* v*, while* $G^-(v)$ *represents the set of its sources,*
- *each edge is labeled by the pair* $(j_{v_m,v_n}, \eta_{v_m,v_n})$ *such that* j_{v_m,v_n} *is the positive integer representing the concentration level required for the interaction, and* $\eta_{v_m,v_n} \in \{+, -\}$ *shows the type of interaction with '+' being activation and '−' being inhibition,*
- $\forall v_n \in G^+(v_m)$*, each* $j_{v_m,v_n} \in \{1, 2, \ldots, max_{v_m}\}$ *where* max_{v_m} *is less than or equal to the number of vertexes in* $G^+(v_m)$*,*
- *each* $v \in V$ *has a set* $Z_v = \{0, 1, \ldots, max_v\}$ *representing its discrete abstracted concentration levels.*

Figure 2 shows the regulatory network of Case Study 1. Here the set $V = \{M1, M2, M3\}$ is the set of vertexes, representing the three entities (discretized populations of microbes), and $E = \{(M1, M2), (M2, M1), (M2, M3), (M3, M2), (M3, M1), (M3, M3)\}$ is the set of edges, each labeled with a positive integer and sign. For example, the edge $e = (M2, M3)$ has $j_{M2,M3} = 1$ and $\eta_{M2,M3} =$"+." The integer represents the level of the source entity required to perform the action, whereas

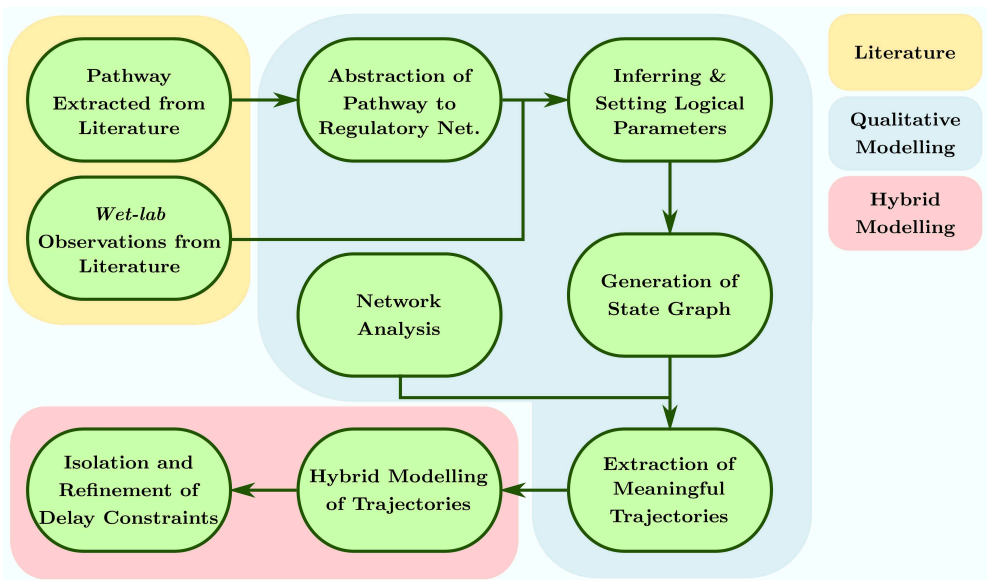


FIGURE 1 | The work-flow of the methodology applied in this study.

the sign shows the type of action: "+" for activation and "−" for inhibition. Thus, the source entity for "+" is the activator, and for "−" the inhibitor. Furthermore, continuing the example, the vertex $M3$ has $max_{M3} \leq 3$ as $G^{+}(M3) = \{M1, M2, M3\}$. Likewise, $G^{-}(M3) = \{M2, M3\}$.

Components of a RN are autonomous processes, with each having different values (discrete abstracted concentration levels) and variables depending upon the interactions with the other components in the dynamic system. At each moment, the whole system is in a given state represented by the tuple formed by the values of each component. In order to understand the dynamics of the discrete model, the states of the regulatory network and the resources and logical parameters of each entity in the respective state are formally defined as,

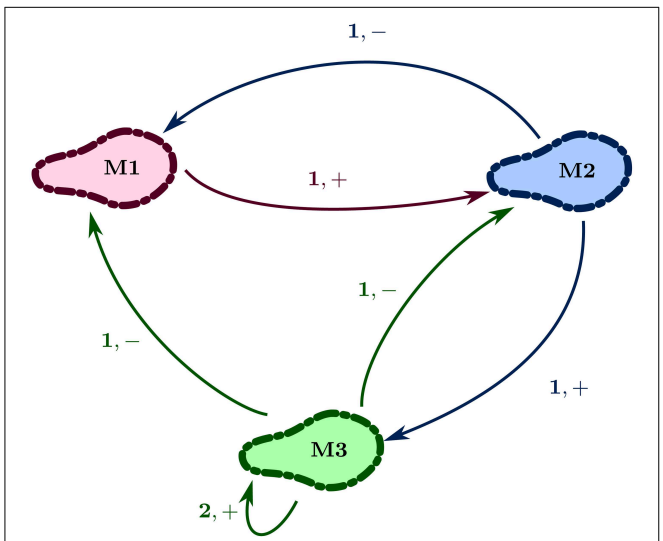


FIGURE 2 | Regulatory network of Case Study 1: Microbial population.

Definition 2. [State]. *A state of an RN is an ordered tuple of the discrete levels of each entity represented as* $(s_{x_v})_{\forall v \in V} \in S$ *where,*

- $S = \Pi_{v \in V} Z_v$ *is the set of all states, and*
- $x_v \in Z_v$ *represents the discrete level of the entity* $v \in V$.

Thus, a state labeled "102," representing the state (1, 0, 2) for Case Study 1 RN (**Figure 2**), would represent the discrete levels as $M1 = 1, M2 = 0, M3 = 2$, in that order.

The resources of one component stand for the set of the components (possibly including itself) that are positively acting on it at the moment. This is again dynamic since it depends on the current value of each of the components, as well as their interactions with the given component. Formally,

Definition 3. [Resources]. *For a given state of an RN, the resource set* $Q_{x_{v_n}}$ *for an entity* $v_n \in V$ *at concentration level* x_{v_n} *is defined as* $Q_{x_{v_n}} = \{v_m \in G^{-}(v_n) | (x_{v_m} \geq j_{v_m,v_n} \wedge \eta_{v_m,v_n} = \text{`+'}) \otimes (x_{v_m} < j_{v_m,v_n} \wedge \eta_{v_m,v_n} = \text{`−'})\}$.

For the state (1, 0, 2), the resource set $Q_{1_{M1}}$ is $\{M2\}$ as $M2$ cannot inhibit it below its respective discrete level "1" while $M3$ is inhibiting it above its respective level "2," $Q_{0_{M2}}$ is $\{M1\}$ as $M1$ is activating it at level "1," and $Q_{2_{M3}}$ is $\{M3\}$ as $M3$ is a self-activator at level "2."

Definition 4. [Logical Parameters]. *The set of logical parameters of an RN "G" is defined as* $K(G) = \{K_{v_m}(Q_{x_{v_m}}) \in Z_{v_m} \; \forall v_m \in V\}$.

The meaning of the logical parameters is that it defines the discrete level that the entity will evolve toward, given the set of resources for the entity available. So, for example, if the parameter

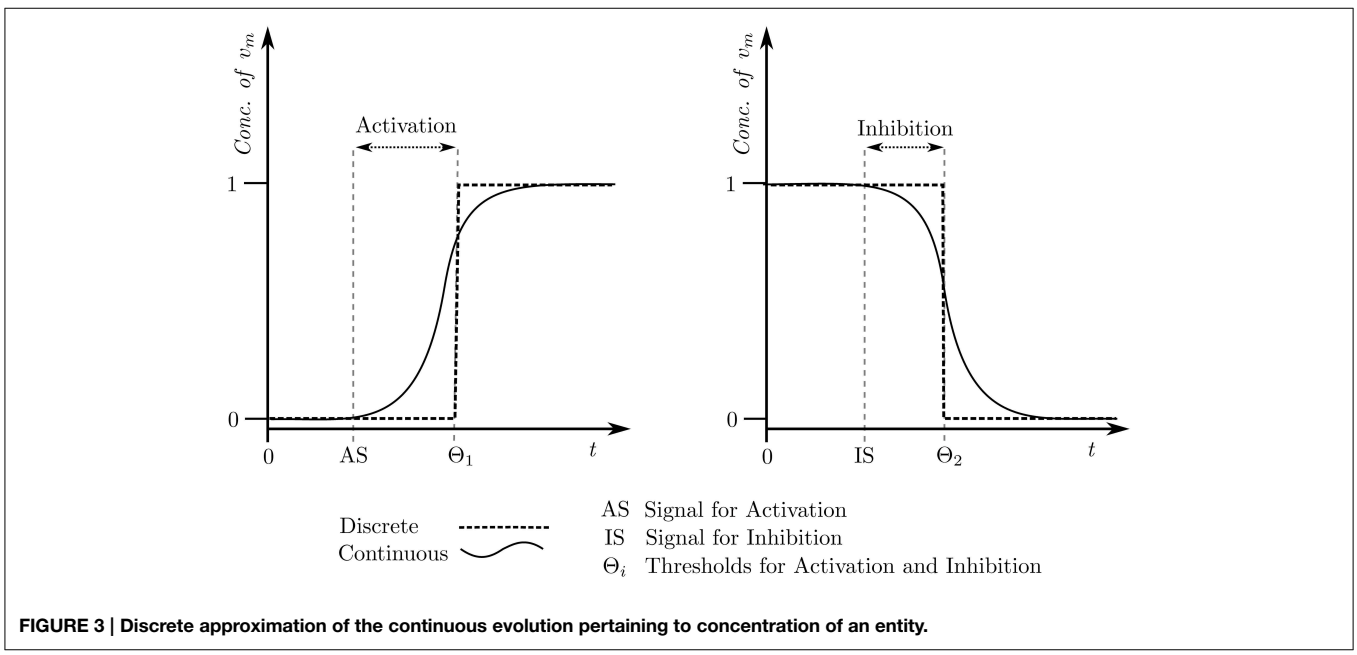


FIGURE 3 | Discrete approximation of the continuous evolution pertaining to concentration of an entity.

$K_{M2}\{M1\}$ is set to "0" (meaning that in the presence of both $M1$ and $M3$, the microbe $M2$ will still decline in population) then the level $M2$ will evolve toward "0" through a step function, similar to the those shown in **Figure 3**.

The parameter K_{v_m} governs the evolution of the entity v_m via the evolution operator "↱" (Bernot et al., 2004) against the rule:

$$x_{v_m} \Rsh K_{v_m}(Q_{x_{v_m}}) = \begin{cases} x_{v_m} + 1 & \text{if } x_{v_m} < K_{v_m}(Q_{x_{v_m}}); \\ x_{v_m} & \text{if } x_{v_m} = K_{v_m}(Q_{x_{v_m}}); \\ x_{v_m} - 1 & \text{if } x_{v_m} > K_{v_m}(Q_{x_{v_m}}). \end{cases}$$

Table 1 shows all the logical parameters of Case Study 1. As a general rule, when only the activators of an entity v are present then the respective entity approaches its max_v, whereas when only the inhibitors are present then it approaches "0." Both of these cases are termed as trivial parameters. Once selected, the logical parameters are then plugged into the respective RN, generating a new directed graph reflecting the discrete dynamics of the system, formally defined as:

Definition 5. [State Graph]. *The state graph of an RN is a directed graph* $SG = (S, T)$ *where,*

- S *is the set of states as defined in Definition 2, and*
- $T \subseteq S \times S$ *is the set of unlabeled edges representing transitions between states such that for each ordered transition* $s \rightarrow s'$, $\exists\ v_m \in V$ *such that* $s_{x_{v_m}} \neq s'_{x_{v_m}}$, $s'_{x_{v_m}} = s_{x_{v_m}} \Rsh K_{v_m}(Q_{x_{v_m}})$, *and* $\forall\ v_n \in V \setminus \{v_m\},\ s_{x_{v_n}} = s'_{x_{v_n}}$.

In summary, each state of the state graph will differ from the other states in at least one entity level. The transitions between the states, on the other hand, will be defined by the logical parameters and available resources of the entities, and will have a target state differing from the source state in exactly one entity level. **Figure 4** shows the state graph of Case Study 1 using the logical parameter set given in the "Selected" column in **Table 1**. As an example, the state transition $(1, 0, 0) \rightarrow (1, 1, 0)$ differs in the level of $M2$ only, and is only possible because of the logical parameter $K_{M2}\{M1, M3\} = 1$. Using the state graph individual behaviors of the system can be studied in the form of trajectories, formally defined as:

Definition 6. [Trajectory]. *A trajectory is defined as a successive series of transitions in a state graph.*

Cyclic trajectories, representing oscillatory behavior, always end up in the starting state $s_i \in S$, whereas acyclic trajectories end at any state other than the starting ($s_n \neq s_i$). A divergent trajectory shares transitions up to a certain extent with another trajectory, and then takes different transitions to different states. The length of the trajectory equals the total number of transitions in the said trajectory. The state graph can also be used to explain deadlocked states as states which do not have any transitions to other states. In practice, such states usually represent configurations of the system which are not favorable in terms of behavior. In **Figure 4**, the state (0, 0, 2) is the deadlock and represents the domination of the microbe $M3$ over the other two species.

TABLE 1 | Case Study 1: Parameter table.

Parameters	Resources	Values		
		Allowed	Generated	Selected
K_{M1}	{}	0	0	0
	{$M2$}	0,1	0	0
	{$M3$}	0,1	0,1	0
	{$M2, M3$}	1	1	1
K_{M2}	{}	0	0	0
	{$M1$}	0,1	0,1	0
	{$M3$}	0,1	0,1	0
	{$M1, M3$}	1	1	1
K_{M3}	{}	0	0	0
	{$M2$}	0,1,2	2	2
	{$M3$}	2	2	2
	{$M2, M3$}	2	2	2

The logical parameters are listed under "Parameters," with the respective resource sets under "Resources," and parameter values under "Values." Allowed Values are the ones that are possible according to logical rules (e.g., the more resources you have, the greater parameter you get). Generated values are the ones that have been generated by the SMBioNet software (see Section 2.2.1) according to some expected properties of the dynamics. Selected Values are the ones we have chosen for our example of Case study 1.

2.2. Optimization of Discrete Modeling

2.2.1. Model Checking

As mentioned earlier in this section, the Hybrid Modeling targets specific trajectories of the RN, present in its state graph. However, the state graph contains many trajectories, and is dependent on a specific set of logical parameters. **Table 1** shows different values that are allowed for each of the logical parameter, with each possible set generating different state graphs and behaviors. Bernot et al. (2004) pioneered an application of automated model checking, specifically Computation Tree Logic (CTL), to select particular sets of logical parameters for discrete modeling. Model checking is an exhaustive automated computational technique which is used to verify different properties in a given system (Clarke et al., 1999). In the SMBioNet software (Bernot et al., 2004; Khalis et al., 2009), the system is provided in the form of an RN, and the observations or known behaviors of the system are encoded in CTL. The software then checks all logical parameters to find and select the ones which can generate the encoded observations.

For Case Study 1, two observations were selected for the screening of logical parameters: from an arbitrary starting state (1, 0, 0), (i) there should exist at least one trajectory which will return to this state to form a cycle, and (ii) there should be at least one trajectory which will end up in the deadlock. Using the general rule of logical parameters described in Section 2.1 after Definition 4 ("Allowed" column in **Table 1**), coupled with the above observations in CTL, SMBioNet generated eight possible logical parameter sets, shown under the "Generated" column in **Table 1**. A single set is then selected from these generated values, given under the "Selected" column, which is then plugged

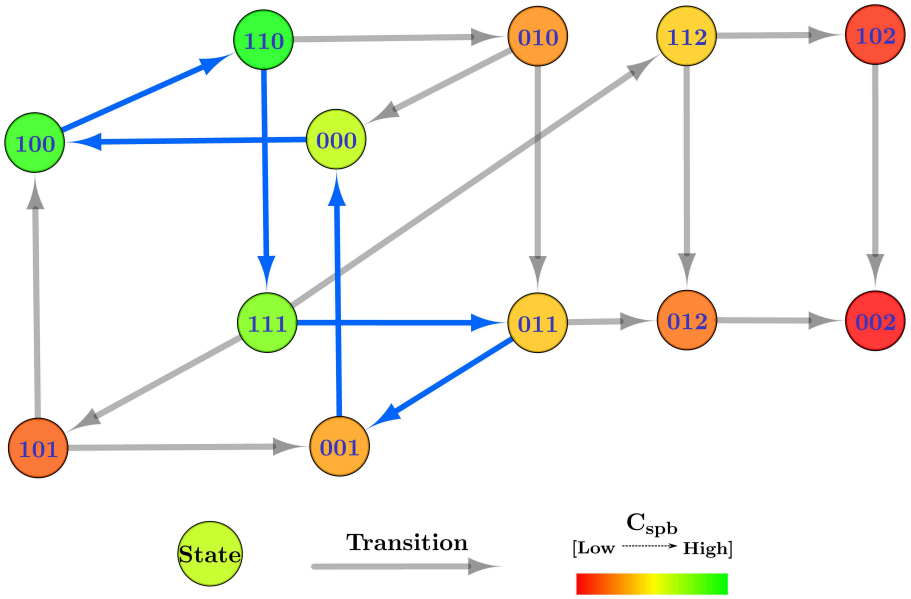


FIGURE 4 | Centrality applied state graph of Case Study 1: Microbial population. The cycle selected for hybrid modeling is shown via blue state transitions.

into the RN. The source file used for the generation of these sets is provided in the Supplementary Materials file. Details of SMBioNet and CTL can be studied in-depth in the excellent review article by Khalis et al. (2009).

2.2.2. Network Analysis

A state graph of a particular system may contain any number of trajectories, depending on the number of entities and the state transitions, which makes the behavior of the system non-deterministic between the said trajectories. As such, selecting specific trajectories for hybrid modeling becomes difficult, especially in the absence of available observations. To assist in the selection of trajectories, we employ a network analysis technique, shortest path betweenness centrality analysis, to isolate trajectories. Adapted from Juncker and Schreiber (2008),

Definition 7. [Shortest Path Betweenness Centrality]. *The shortest path betweenness centrality (C_{spb}), also defined as betweenness centrality, measures the occurrence of the a particular vertex $v \in V$ in all the shortest paths between other vertexes. Mathematically, it is computed as* $C_{spb} = \sum_{p \neq q \neq r \in S} \frac{\sigma_{pr}(q)}{\sigma_{pr}}$ *where,*

- σ_{pr} *is the total number of shortest trajectories from the state p to r, and*
- $\sigma_{pr}(q)$ *is the total number of shortest trajectories from p to r which pass through q.*

In short, C_{spb} measures the frequency of the occurrences of each state in the trajectories between other states, and as a fraction, only ranges from "0" to "1." This allows for a relative measure of the states amongst themselves, with states having C_{spb} of at or near "1" occurring more frequently in the dynamics and trajectories of the system, than states which have their C_{spb} at or near "0." The significance of this measure also increases with the fact that C_{spb} of deadlock states automatically becomes zero because all trajectories get stuck in the deadlock state and do not pass through it to reach any other state. This reflects the tendency of oscillating systems to avoid deadlocked states. In comparison to other centrality analyses available in the field of network analysis, C_{spb} is the only measure which is able to cater to both of these properties simultaneously. For example, closeness centrality measure (Juncker and Schreiber, 2008) caters to lower numerical values for deadlocked states, but does not cater to the relevance of the states themselves in terms of cyclic trajectories, thus not solving non-determinism. Likewise, the eccentricity centrality will represent the overall reachability of a particular state from other states, but will not contribute toward the solution of non-determinism in terms of selecting preferred trajectories in the system.

The table containing the C_{spb} of each state of the state graph for Case Study 1 is provided in the Supplementary Materials file, and **Figure 4** is color coded to reflect higher measures of C_{spb} with green, and lower levels with red. For this case study, a cyclic trajectory comprising of the states with the highest centralities was selected for hybrid modeling, as this would theoretically reflect the most favorable and most probable behavior of the system based on the frequent occurrences of the constituent states. We begin with the state with the highest C_{spb}, in this case state (1, 1, 0), and select the state with the highest centrality from its target states, in this case (1, 1, 1). We then select the state with the highest centrality from the targets of (1, 1, 1), which is (0, 1, 1). Continuing onwards, the trajectory builds toward the state (0, 0, 1), (0, 0, 0), (1, 0, 0), and finally back to the starting state (1, 1, 0), completing a cycle with the highest cumulative C_{spb}. This cycle, (1, 1, 0) → (1, 1, 1) → (0, 1, 1) → (0, 0, 1) →

$(0, 0, 0) \rightarrow (1, 0, 0) \rightarrow (1, 1, 0)$, is represented in **Figure 4** via blue transitions between the states.

2.3. Hybrid Modeling

As mentioned earlier in this section, the hybrid model is constructed on particular trajectories of the state graph. In the approach we use the *Parametric Biological Linear Hybrid Automata* (Bio-LHA) to convert the selected trajectory into a hybrid model by merging the constructed discrete system with time constraints. The formal definition of Bio-LHA, adapted from Ahmad et al. (2007); Ahmad and Roux (2010) is given below.

Let $C^{=}(X, P)$, $C^{\leq}(X, P)$, and $C^{\geq}(X, P)$ be the set of constraints using only $=$, $\leq$, and $\geq$, respectively. Here, X and P are the sets of real valued variables and parameters, respectively.

Definition 8. [Parametric Bio Linear Hybrid Automaton (Bio-LHA)]. *A parametric bio linear hybrid automaton* $\mathbb{B}$ *is a tuple* $(L, l_0, X, P, E, Inv, Dif)$ *where,*

- L *is a finite set of locations,*
- $l_0 \in L$ *is the initial location,*
- P *is a finite set of parameters (delays),*
- X *is a finite set of real-valued variable (clocks),*
- $E \subseteq L \times C^{=}(X, P) \times 2^{X} \times L$ *is a finite set of edges with typical element* $e = (l, g, R, l') \in E$ *representing an edge from* l *to* l' *with guard* g *and the reset set* $R \subseteq X$*. The guard is the timing condition required for the edge to be used for transition, with the clocks used in* $g \in R$*,*
- $Inv : L \rightarrow C^{\leq}(X, P) \cup C^{\geq}(X, P)$ *assigns an invariant to any location,*
- $Dif : L \times X \rightarrow \{-1, 0, 1\}$ *maps each pair* (l, h) *to an evolution rate.*

The selected cyclic trajectory (cycle) was converted into the Bio-LHA shown in **Figure 5**. The locations represent the states of the system, with state $(0, 0, 0)$ being the arbitrary initial point, prominent via the diagonal arrow. Every entity v of the system has two time parameters each, the activation delay d_v^+, and the inhibition delay d_v^-. These delay parameters respectively represent the time it takes for the entity to activate and inhibit, as graphically shown in **Figure 6**. Following the delay parameters are finite real-valued variables $h_v \in X$, which represent individual clocks unique to each entity of the system and used to measure the time for each entity. Each location has an invariant, or the timing restrictions that allow the system to remain within it. The Bio-LHA has to transition to another location before the invariant of the current location is falsified, else the constructed Bio-LHA is erroneous. Likewise, each location transition also has timing restrictions, known as guards, which allow the firing of the transition only when they are true. Lastly, each clock variable has a specific rate in each location: $\dot{h}_v = 0$ for when the respective entity of the clock is not evolving, $\dot{h}_v = 1$ when the respective entity is being activated, and $\dot{h}_v = -1$ when the respective entity is being inhibited. The rates throughout the state graph can be fixed based on the difference of discrete levels of each entity of the state either with its successors, or the successors of its successors. The later method is termed "anticipation" and accurately reflects the interactions between biological entities, but can be forgone for non-biological environmental systems. Thus, for Case Study 1, the rates were based on anticipation.

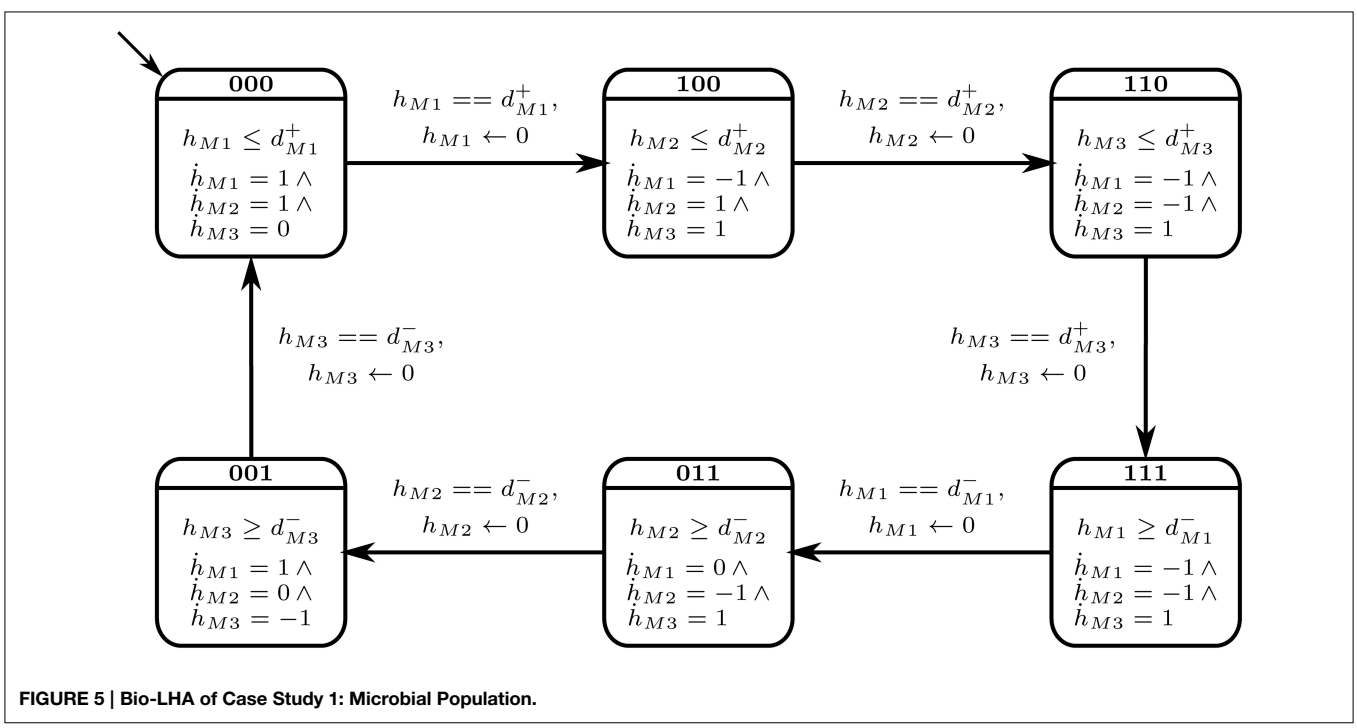


FIGURE 5 | Bio-LHA of Case Study 1: Microbial Population.

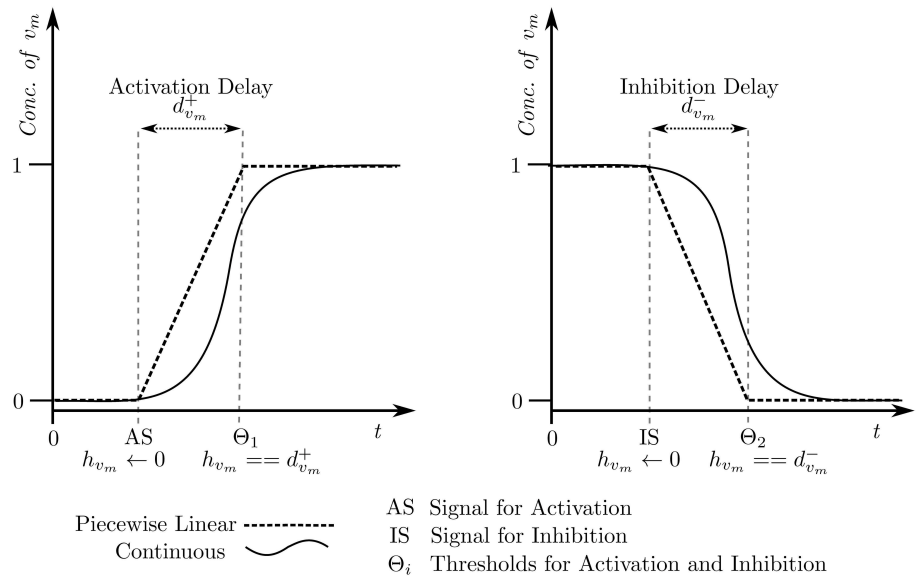


FIGURE 6 | Piecewise affine linear approximation of the continuous evolution of the entity, forming the basis of the hybrid model.

Collectively, the semantics of the of the Bio-LHA are represented as a transition system, adapted from Ahmad et al. (2007) and defined below:

Definition 9. [Semantics of Bio-LHA]. *Let γ be a valuation for the parameters P and ν represents the values of clocks in a location. The $(\mathbb{T}, \gamma)$-semantics of a parametric Bio-LHA $\mathbb{B}$ is defined as a timed transition system $_B = (\mathbb{S}, s_0, \mathbb{T}, \rightarrow)$ where: (i) $\mathbb{S} = \{(\ell, \nu) \mid \ell \in L$ and $\nu \models Inv(\ell)\}$, (ii) s_0 is the initial state, and (iii) the relation $\rightarrow \subseteq \mathbb{S} \times \mathbb{T} \times \mathbb{S}$ is defined for $t \in \mathbb{T}$ as:*

- *discrete transitions: $(\ell, \nu) \xrightarrow{0} (\ell', \nu')$ iff $\exists(\ell, g, R, \ell') \in E$ such that $g(\nu) = true$, $\nu'(h) = 0$ if $h \in R$ and $\nu'(h) = \nu(h)$ if $h \notin R$.*
- *continuous transitions: For $t \in \mathbb{T}^*$, $(\ell, \nu) \xrightarrow{t} (\ell', \nu')$ iff $\ell' = \ell$, $\nu'(h) = \nu(h) + Dif(\ell, h) \times t$, and for every $t' \in [0, t]$, $(\nu(h) + Dif(\ell, h) \times t') \models Inv(\ell)$, where $\models$ represents satisfaction operator.*

In short, the discrete transitions show the transitioning between locations, and is instantaneous. On the other hand, continuous transitions take place within the same location and allows the time to elapse. In **Figure 5**, the Bio-LHA will remain in the initial location $l_0 = (0, 0, 0)$ as long as the clock variable h_{M1} remains less than or equal to d^+_{M1}, the activation delay of $M1$. The rate for that clock is $\dot{h}_{M1} = 1$, given below the invariant. Once the clock variable equals d^+_{M1}, the continuous transitions can no longer occur. At the same time the guard on the transition $h_{M1} == d^+_{M1}$ becomes true and allows a discrete transition from $(0, 0, 0)$ to $(1, 0, 0)$, resetting the clock h_{M1} to "0" in the process, provided that the invariant of the next location $(1, 0, 0)$ is not falsified.

The construction of the Bio-LHA further allows the temporal (time based) analysis of the modeled trajectory. The locations themselves represent independent temporal zones, and collectively represent the temporal state space containing discrete transitions between the temporal zones. In this paper, we analyze the invariance kernel of the hybrid models, based on the temporal state space. The invariance kernel represents the entity activation and inhibition timing constraints governing the modeled cyclic oscillating trajectory. Thus, viability is the core of this representation, and can be defined as (adapted from Aubin, 1991; Ahmad et al., 2007; Ahmad, 2009).

Definition 10. [Viability]. *Suppose that all trajectories emanating from a particular initial state remain bounded within some constraints, making the trajectories always staying in one sub-domain: namely the viability domain. Such trajectories are called viable trajectories.*

The invariance kernel itself is formally defined as (Adapting from Ahmad et al., 2007; Ahmad and Roux, 2010),

Definition 11. [Invariance Kernel]. *Let $\phi(t) \in S \forall t \geq 0$ be a viable trajectory in the temporal state space S. The largest subset $K(S)$ is the invariance kernel if a trajectory starting at point p is viable in K, $\forall p \in K$.*

The algorithm given in our previous study (Ahmad et al., 2007), was utilized to find the invariance kernel for this Bio-LHA. However, the invariance kernel did not converge, indicating that the modeled trajectory may be asymptotic. For such trajectories, the convergence domain analysis (Ahmad and Roux, 2010) provides an over approximation of the delay constraints, within which the trajectory will converge asymptotically. Formally,

Definition 12. [Convergence Domain]. *The subset $K(S)$ is called the convergence domain if $\forall p \in K$, the trajectory starting at point p converges in an asymptotic manner.*

TABLE 2 | The relation matrix of Case Study 1.

	Relation Matrix					
	d^+_{M1}	$\lvert d^-_{M1}\rvert$	d^+_{M2}	$\lvert d^-_{M2}\rvert$	d^+_{M3}	$\lvert d^-_{M3}\rvert$
d^+_{M1}	$=$					
$\lvert d^-_{M1}\rvert$	$\geq, \leq$	$=$				
d^+_{M2}	$\geq, \leq$	$\geq, \leq$	$=$			
$\lvert d^-_{M2}\rvert$	$\geq, \leq$	$\geq, \leq$	$\geq, \leq$	$=$		
d^+_{M3}	$\geq, \leq$	$\leq$	$\geq, \leq$	$\geq, \leq$	$=$	
$\lvert d^-_{M3}\rvert$	$\leq$	$\geq, \leq$	$\geq, \leq$	$\geq, \leq$	$\geq, \leq$	$=$

Refined pairwise relations between the parameters of the modeled cycle in the form of a matrix. The matrix is read row first, column second.

The convergence domain analysis for Case Study 1 showed 12 conjuncted delay constraints, provided as a table in the Supplementary Materials file. Further analysis of the delay constraints, using a linear constraint solver, allowed us to generate pairwise delay constraint relations between the delay parameters which are given as a matrix in **Table 2**. The matrix is read row first, column second, from which we can see that the delay relations $|d^-_{M3}| \leq d^+_{M1}$ and $d^+_{M3} \leq |d^-_{M1}|$ are the only significant relations because they represent the constraints which force the system to remain asymptotically within the given cycle. Closer inspection shows that if the constraint $|d^-_{M3}| \leq d^+_{M1}$ is violated, then the microbe $M3$ will not be inhibited before the activation of microbe $M1$, understandable as $M3$ inhibits $M1$, but it will also nudge the system toward the dominance of $M3$ over $M1$ and $M2$. In terms of ecological balance, this represents how a species can become invasive, displacing other species from the habitat and disrupting other processes dependent on the displaced species. Likewise, if the second constraint $d^+_{M3} \leq |d^-_{M1}|$ is violated, then $M2$ will be able to inhibit $M1$ and will in-turn be inhibited before $M3$ becomes active, diverging into a completely different cycle from the one modeled as the Bio-LHA which can represent the replacement of a particular type of ecological balance with another.

2.4. Software

The software SMBioNet (Bernot et al., 2004; Khalis et al., 2009) was used to isolate the logical parameters for both case studies. These parameters, together with the respective regulatory networks, were then constructed in the tool GenoTech (Ahmad et al., 2012), to generate the state graph of the trajectories. Cytoscape (Shannon et al., 2003) was used to conduct centrality analysis of the state graph to isolate particular trajectories. Finally, HyTech (Henzinger et al., 1997) was used to construct the Bio-LHA and analyze the delay constraints of the isolated trajectories.

3. Results

In this section we apply the procedure detailed in the previous section on the slightly larger and more complex Case Study 2, described in Section 1.5.

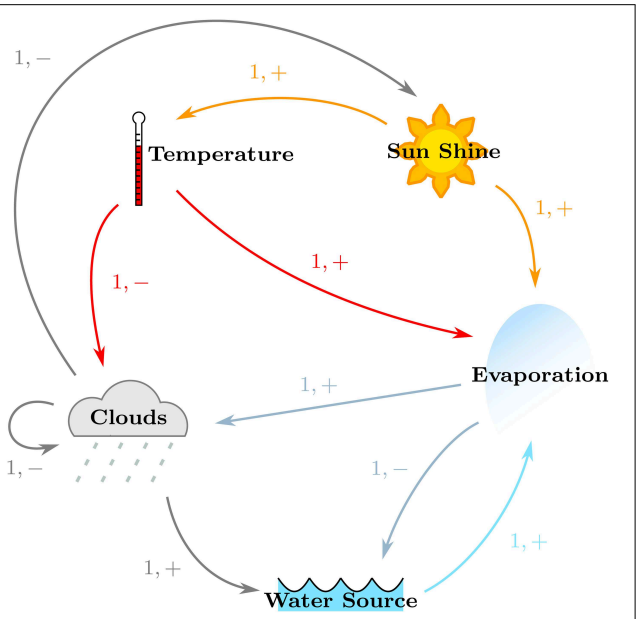


FIGURE 7 | Regulatory network of Case Study 2: Atmospheric system.

3.1. Regulatory Network and Logical Parameters

The regulatory network was constructed for Case Study 2, with particular focus on satisfying the given specifications. Its regulatory network, shown in **Figure 7**, was constructed in the same manner, and shows the positive effects of Sun Shine on the Temperature and Evaporation, the positive effects of the Earth Temperature on Evaporation, the positive effect of Evaporation on Cloud production, the positive effect of Clouds on Water via precipitations, the contribution of the Water source toward Evaporation, as well as the negative effects of Clouds on Sun Shine and themselves (because of precipitation), that of Temperature on Clouds, and of Evaporation on the Water source. For ease in analysis, the names of the entities were shortened to "Temp," "Sun," "Evap," and "Water" for "Temperature," "Sun Shine," "Evaporation," and "Water Source," respectively. Thus, the formal description of Case Study 2 yields the sets:

- $V = \{Temp, Sun, Evap, Clouds, Water\}$, and
- $E = \{(Sun, Temp), (Sun, Evap), (Temp, Evap), (Temp, Clouds), (Evap, Clouds), (Evap, Water), (Clouds, Clouds), (Clouds, Sun), (Clouds, Water), (Water, Evap)\}$

After the construction of the regulatory network, the logical parameters were generated using SMBioNet. The observations encoded in CTL checked that from an arbitrary starting state of Sun and Water being available ($Temp = 0$, $Sun = 1$, $Evap = 0$, $Clouds = 0$, $Water = 1$), to contain: (i) at least one trajectory where Water is persistently available, and (ii) one cyclic or acyclic trajectory where Water is no longer available after some transitions. Apart from restricting the trivial logical parameters, those parameters for Evap which did not have Water as a resource, and those parameters for Clouds which did not have Evap as resource, were restricted to "0" value in order to reflect the dependence of the entities on their respective resources. A

TABLE 3 | Case Study 1: Parameter table.

Parameters	Resources	Values		
		Allowed	Generated	Selected
K_{Temp}	{}	0	0	0
	{*Sun*}	1	1	1
K_{Sun}	{}	0	0	0
	{*Clouds*}	1	1	1
K_{Evap}	{}	0	0	0
	{*Sun*}	0	0	0
	{*Temp*}	0	0	0
	{*Water*}	0	0	0
	{*Sun*, *Temp*}	0	0	0
	{*Temp*, *Water*}	0,1	0,1	1
	{*Sun*, *Water*}	0,1	0,1	1
	{*Temp*, *Sun*, *Water*}	1	1	1
K_{Clouds}	{}	0	0	0
	{*Clouds*}	0	0	0
	{*Evap*}	0,1	0,1	0
	{*Temp*}	0	0	0
	{*Clouds*, *Evap*}	0,1	0,1	1
	{*Evap*, *Temp*}	0,1	0,1	0
	{*Clouds*, *Temp*}	0	0	0
	{*Clouds*, *Evap*, *Temp*}	1	1	1
K_{Water}	{}	0	0	0
	{*Clouds*}	0,1	0,1	1
	{*Evap*}	0,1	0	0
	{*Clouds*, *Evap*}	1	1	1

The logical parameters are listed under "Parameters," with the respective resource sets under "Resources," and parameter values under "Values."

total of 32 sets were generated which satisfied the CTL property, from which a single set given under the "Selected" column of **Table 3** was plugged into the Case Study 2 RN.

3.2. State Graphs and Centrality Analysis

After plugging in the logical parameters, the state graphs of Case Study 2 was generated, shown in **Figure 8**. The state graph contains 566 cyclic trajectories (cycles) and a single deadlock state (1, 1, 0, 0, 0), for the entity order ⟨*Temp*, *Sun*, *Evap*, *Clouds*, *Water*⟩. The deadlock state, in particular, shows a unique configuration where the whole system has turned arid with persistent high temperatures and sun shine.

Due to the large number of generated cyclic trajectories, a filtering method was applied in addition to C_{spb}, by selecting cycles of length 10 only from the total cycles, leaving behind 152 cycles. The reason for the selected length is that it allows all five entities of the system to oscillate between their boolean values, but only once throughout the cycle. To further ease the cycle selection, all 152 cycles were sorted in descending order based on their mean C_{spb}, instead of using cumulative C_{spb} for selection as done in Case Study 1 (Section 2.2.2). From the sorted list, the first cycle which had all five entities oscillating was

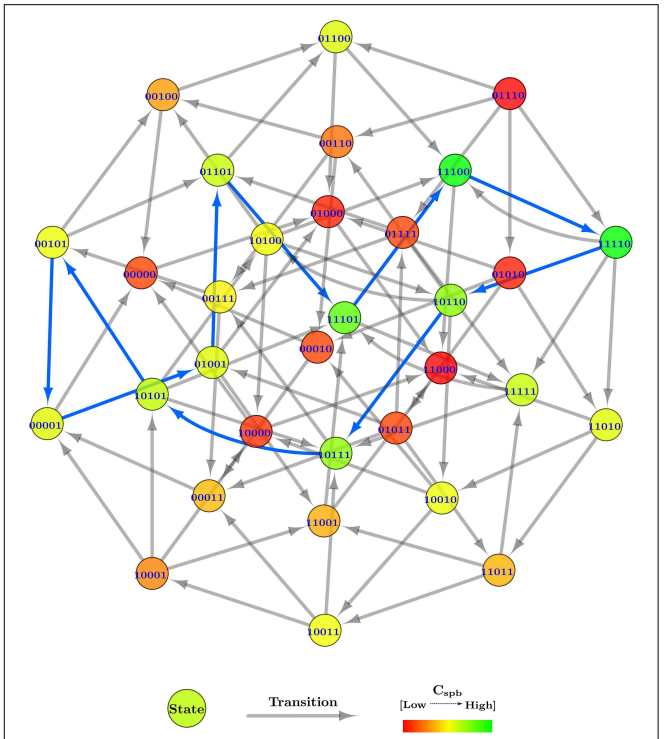


FIGURE 8 | Centrality applied state graph of Case Study 2: Atmospheric system. The cycle selected for hybrid modeling is shown via blue state transitions.

selected, i.e., (0, 1, 0, 0, 1) → (0, 1, 1, 0, 1) → (1, 1, 1, 0, 1) → (1, 1, 1, 0, 0) → (1, 1, 1, 1, 0) → (1, 0, 1, 1, 0) → (1, 0, 1, 1, 1) → (1, 0, 1, 0, 1) → (0, 0, 1, 0, 1) → (0, 0, 0, 0, 1) → (0, 1, 0, 0, 1), having the mean $C_{spb} = 0.145$ approximately. The cycle is shown via blue transitions in **Figure 8**, and starts in the initial state with the sun shining and a water source available, collectively inducing evaporation in the second state. Meanwhile, the temperature also increases as indicated in the third state, while the water source is drained due to persistent evaporation in the fourth state. Afterwards, clouds are formed in the fifth state due to the evaporated water vapors, blocking out the sun in the sixth state, and rejuvenating the water source in the seventh state via precipitation. The clouds disseminate in the eighth state because of the collective effect of precipitation and high temperatures. However, due to persistent blockage of the sun in the previous states, the temperature also goes down in the ninth state, effectively stopping evaporation in the 10th state. Given that the clouds have been disseminating for a while now, the sun begins to shine again, bringing the whole trajectory to its initial state.

3.3. Hybrid Modeling and Delay Constraints

The selected cyclic trajectory was then converted to a Bio-LHA, as shown in **Figure 9**. The rates for each entity were set based on the immediate successors of the state in the state graph, unlike Case Study 1 where anticipation was used to set the rates. The primary reason is that in biological systems, a small concentration of a source entity can start affecting its target

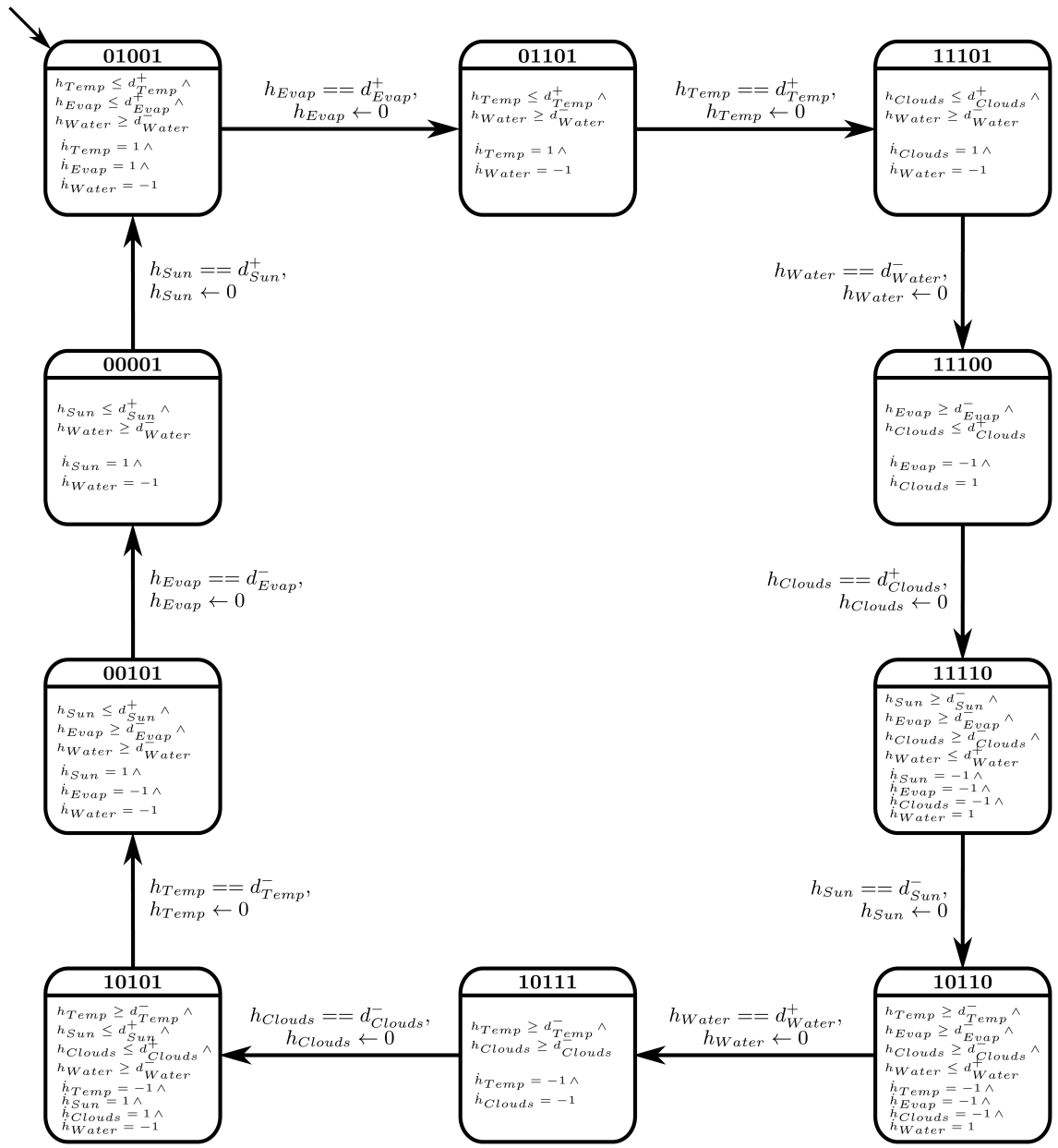


FIGURE 9 | Bio-LHA of Case Study 2: Atmospheric System. To improve readability, only the clock variables with non-zero rates are shown in each location.

entities, thus initiating their activation or inhibition early (before achieving the threshold). In larger environmental systems, on the other hand, a small concentration of the source entity has negligible effects on its target entities, thus we forgo the concept of anticipation for Case Study 2. Once set, the algorithm cited in Section 2.3 was used to find the invariance kernel of the cycle, converging upon a single set of delay constraints. The seven non-trivial conjuncted constraints are provided in the Supplementary Materials file.

The constraints were then fed to the linear constraint solver to generate the pairwise relation matrix given in **Table 4**, revealing nine significant constraint relations. These significant relations force the trajectory to remain within the cycle, and violation of these relations will diverge the system toward other cyclic or acyclic trajectories. Thus, the violation of the constraint $d^+_{Evap} \leq d^+_{Temp}$ will diverge the system toward states where the activation of Temperature occurs before Evaporation, possibly restricting or limiting the production of Clouds due to persistent high temperature. Likewise, the inhibition delay of the Water entity ($|d^-_{Water}|$) strictly takes more duration than activation of Temperature, activation of Sun Shine, and activation of Evaporation. Violation of the first relation will allow the

TABLE 4 | The relation matrix of Case Study 2.

	Relation Matrix									
	d^+_{Temp}	$\lvert d^-_{Temp}\rvert$	d^+_{Sun}	$\lvert d^-_{Sun}\rvert$	d^+_{Evap}	$\lvert d^-_{Evap}\rvert$	d^+_{Clouds}	$\lvert d^-_{Clouds}\rvert$	d^+_{Water}	$\lvert d^-_{Water}\rvert$
d^+_{Temp}	=									
$\lvert d^-_{Temp}\rvert$	$\geq, \leq$	=								
d^+_{Sun}	$\geq, \leq$	$\geq, \leq$	=							
$\lvert d^-_{Sun}\rvert$	$\geq, \leq$	$\geq, \leq$	$\geq, \leq$	=						
d^+_{Evap}	$\leq$	$\geq, \leq$	$\geq, \leq$	$\geq, \leq$	=					
$\lvert d^-_{Evap}\rvert$	$\geq, \leq$	$\geq, \leq$	$\geq, \leq$	$\geq$	$\geq, \leq$	=				
d^+_{Clouds}	$\geq, \leq$	$\geq, \leq$	$\geq, \leq$	$\geq, \leq$	$\geq, \leq$	$\geq, \leq$	=			
$\lvert d^-_{Clouds}\rvert$	$\geq, \leq$	$\geq, \leq$	$\geq, \leq$	$\geq$	$\geq, \leq$	$\geq, \leq$	$\geq, \leq$	=		
d^+_{Water}	$\geq, \leq$	$\geq, \leq$	$\geq, \leq$	$\geq$	$\geq, \leq$	$\leq$	$\geq, \leq$	$\leq$	=	
$\lvert d^-_{Water}\rvert$	$>$	$\geq, \leq$	$>$	$\geq, \leq$	$>$	$\geq, \leq$	$\geq, \leq$	$\geq, \leq$	$\geq, \leq$	=

Refined pairwise relations between the parameters of the modeled cycle in the form of a matrix (read row first, column second). The order of entities in each state of the cycle is ⟨Temp, Sun, Evap, Clouds, Water⟩ for "Temperature," "Sun Shine," "Evaporation," "Clouds," and "Water Source."

Water Source to be drained before the ultimate activation of Temperature, nudging the trajectory closer to the deadlock state. Then, in the relation $d^+_{Water} \leq |d^-_{Clouds}|$, it is clear that the Water Source should be replenished before the dissemination of the Clouds, the violation of which will also diverge the trajectory toward the deadlock state. Recall that the deadlock state is (1, 1, 0, 0, 0).

4. Discussion

Although the application of *Parametric Biological Linear Hybrid Automata* is not new (Ahmad et al., 2009, 2012; Aslam et al., 2014; Paracha et al., 2014 to name a few), in this paper we present its application to a new set of problems, namely the modeling of environmental systems. We successfully model two case studies representing different systems from a broad spectrum of mechanisms encompassed by environmental systems, which shows the versatility of the modeling framework toward its applicability. In the first case study, our framework modeled the arbitrary population levels between different microbes in a freshwater pond, showing how hybrid modeling can be used to study different behavioral tendencies of the populations represented by population balance (cyclic trajectories) and overpopulation (deadlocked trajectory), while removing the complexity of dealing with actual population numbers. In the second case study, the modeling framework was applied on a slightly more complex atmospheric system, providing us with the constrained dynamics that the system exhibits in a particular trajectory, while also pinpointing particular constrained relations between parameters which are essential to keep a trajectory stable and non-divergent.

In the analysis of the two case studies, the hybrid framework and application methodology was able to answer all objections pertaining to other methodologies raised in Section 1.2, namely that this methodology (i) presents the complete dynamics of the system in the form of state graphs, (ii) is able to deterministically predict different behaviors via delay constraints, (iii) whilst keeping the complexity of the system manageable. Another possibility is the conversion of the hybrid to a continuous model via Timed Hybrid Petri Nets (David and Alla, 2010) where these delay constraints can be applied using arbitrary numeric values to timed or continuous transitions. The advantage of this conversion is that precise perturbations (disturbances) and *what-if* scenarios can be constructed to experiment and analyze the system, resulting in added insights of the behavioral dynamics of the system when faced with different stimuli.

That being said, the approach has its limitations, primarily the number of entities being modeled. Although theoretically any number of entities can be modeled, practically the complexity and state space of the system increase exponentially with the number of entities, leading to state space explosion. This limitation puts a major emphasis on effective abstraction of the system in order to preserve the behaviors of the system whilst representing as few entities as possible. Although useful, there are times when a large number of entities need to be represented for different purposes (Hanrahan, 2010), leaving abstraction a bitter-sweet process. In Conclusion, the application of *Parametric Biological Linear Hybrid Automata* allows the modeling of complex environmental systems for the analysis of behavioral dynamics of these system.

Author Contributions

Study design: ST, JA, OR; Experiments: ST; Analyses: ST, JA; Write-up and review: ST, JA, OR.

Acknowledgments

The authors acknowledge the valuable input of the respected reviewers for the improvement of quality of this research article.

Supplementary Material

The Supplementary Material for this article can be found online at: http://journal.frontiersin.org/article/10.3389/fenvs.2015.00047

Supplemental Data

A supplementary materials file is provided which contains the SMBioNet and HyTech source codes of both case studies, along with the respective convergence domain and invariance kernel tables.

References

Ahmad, J., and Roux, O. (2010). Invariance kernel of biological regulatory networks. *Int. J. Data Mining Bioinformatics* 4, 553–570. doi: 10.1504/IJDMB.2010.035900

Ahmad, J., Richard, A., Bernot, G., Comet, J.-P., and Roux, O. (2006). "Delays in biological regulatory networks (brn)," in *Computational Science ICCS 2006*, Vol. 3992 *of Lecture Notes in Computer Science* eds V. Alexandrov, G. van Albada, P. Sloot, and J. Dongarra (Berlin; Heidelberg: Springer), 887–894.

Ahmad, J., Bernot, G., Comet, J.-P., Lime, D., and Roux, O. (2007). Hybrid modelling and dynamical analysis of gene regulatory networks with delays. *Complexus* 3, 231–251. doi: 10.1159/000110010

Ahmad, J., Roux, O., Bernot, G., and Comet, J.-P. (2008). Analysing formal models of genetic regulatory networks with delays. *Int. J. Bioinformatics Res. Appl.* 4, 240–262. doi: 10.1504/IJBRA.2008.019573

Ahmad, J., Bourdon, J., Eveillard, D., Fromentin, J., Roux, O., and Sinoquet, C. (2009). Temporal constraints of a gene regulatory network: refining a qualitative simulation. *Bio Syst.* 98, 149–159. doi: 10.1016/j.biosystems.2009.05.002

Ahmad, J., Niazi, U., Mansoor, S., Siddique, U., and Bibby, J. (2012). Formal modeling and analysis of the mal-associated biological regulatory network: insight into cerebral malaria. *PLoS ONE* 7:e33532. doi: 10.1371/journal.pone.0033532

Ahmad, J. (2009). *Modélisation Hybride et Analyse des Dynamiques des Réseaux de Régulations Biologiques en Tenant Compte des Délais.* Ph.D. thesis, Ecole Centrale de Nantes.

Aslam, B., Ahmad, J., Ali, A., Paracha, R., Tareen, S., Niazi, U., et al. (2014). On the modelling and analysis of the regulatory network of dengue virus pathogenesis and clearance. *Comput. Biol. Chem.* 53, 277–291. doi: 10.1016/j.compbiolchem.2014.10.003

Aubin, J. P. (1991). *Viability Theory*. Cambridge, MA: Birkhauser Boston Inc.

Berie, H. (2014). Suitability analysis for jatropha curcas production in ethiopia - a spatial modeling approach. *Environ. Syst. Res.* 3, 25. doi: 10.1186/s40068-014-0025-7

Bernot, G., Comet, J.-P., Richard, A., and Guespin, J. (2004). Application of formal methods to biological regulatory networks: extending thomas asynchronous logical approach with temporal logic. *J. Theor. Biol.* 229, 339347. doi: 10.1016/j.jtbi.2004.04.003

Casulli, V., and Zanolli, P. (2002). Semi-implicit numerical modeling of nonhydrostatic free-surface flows for environmental problems. *Math. Comput. Model.* 36, 1131–1149. doi: 10.1016/S0895-7177(02)00264-9

Clarke, E. M. Jr., Grumberg, O., and Peled, D. A. (1999). *Model Checking*. Cambridge, MA: MIT Press.

Coulthard, T., Lewin, J., and Macklin, M. (2005). Modelling differential catchment response to environmental change. *Geomorphology* 69, 222–241. doi: 10.1016/j.geomorph.2005.01.008

Crooks, A., Castle, C., and Batty, M. (2008). Key challenges in agent-based modelling for geo-spatial simulation. *Comput. Environ. Urban Syst.* 32, 417–430. doi: 10.1016/j.compenvurbsys.2008.09.004

David, R., and Alla, H. (2010). *Discrete, Continuous, and Hybrid Petri Nets*. Berlin; Heidelberg: Springer.

Gottschalk, F., Sonderer, T., Scholz, R. W., and Nowack, B. A. (2010). Possibilities and limitations of modeling environmental exposure to engineered nanomaterials by probabilistic material flow analysis. *Environ. Toxicol. Chem.* 29, 1036–1048. doi: 10.1002/etc.135

Hanrahan, G. (2010). *Modelling of Pollutants in Complex Environmental Systems*. Hertfordshire: ILM Publications.

Henzinger, T. A., Ho, P.-H., and Wong-Toi, H. (1997). "Hytech: a model checker for hybrid systems," in *Computer Aided Verification*, ed O. Grumberg (Heidelberg: Springer), 460–463.

Juncker, B. H., and Schreiber, F. (2008). *Analysis of Biological Networks*. Hoboken, NJ: John Wiley & Sons, Inc.

Kanevski, M., Parkin, R., Pozdnukhov, A., Timonin, V., Maignan, M., Demyanov, V., et al. (2004). Environmental data mining and modeling based on machine learning algorithms and geostatistics. *Environ. Model. Softw.* 19, 845–855. doi: 10.1016/j.envsoft.2003.03.004

Kauffman, S. (1969). Metabolic stability and epigenesis in randomly constructed genetic nets. *J. Theor. Biol.* 22, 437–467. doi: 10.1016/0022-5193(69)90015-0

Khalis, Z., Comet, J.-P., Richard, A., and Bernot, G. (2009). The SMBioNet method for discovering models of gene regulatory networks. *Genes Genomes Genomics* 3, 15–22.

Paracha, R., Ahmad, J., Ali, A., Hussain, R., Niazi, U., Tareen, S., et al. (2014). Formal modelling of toll like receptor 4 and jak/stat signalling pathways: insight into the roles of socs-1, interferon-β and proinflammatory cytokines in sepsis. *PLoS ONE* 9:e108466. doi: 10.1371/journal.pone.0108466

Refsgaard, J. C., van der Sluijs, J. P., Hjberg, A. L., and Vanrolleghem, P. A. (2007). Uncertainty in the environmental modelling process a framework and guidance. *Environ. Model. Softw.* 22, 1543–1556. doi: 10.1016/j.envsoft.2007.02.004

Sengupta, R. R., and Bennett, D. A. (2003). Agent-based modelling environment for spatial decision support. *Int. J. Geogr. Inf. Sci.* 17, 157–180. doi: 10.1080/713811747

Seppelt, R. (2007). *Computer-Based Environmental Management*. Wiley-VCH Verlag GmbH & Co. KGaA.

Shannon, P., Markiel, A., Ozier, O., Baliga, N., Wang, J., Ramage, D., et al. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 13, 2498–2504. doi: 10.1101/gr.1239303

Somlydy, L. (1982). Modelling a complex environmental system: the lake balaton study. *Math. Model.* 3, 481–502. doi: 10.1016/0270-0255(82)90044-6

Sun, T. Y., Gottschalk, F., Hungerbhler, K., and Nowack, B. (2014). Comprehensive probabilistic modelling of environmental emissions of engineered nanomaterials. *Environ. Pollut.* 185, 69–76. doi: 10.1016/j.envpol.2013.10.004

Ttrai, I., Mtys, K., Korponai, J., Paulovits, G., and Pomogyi, P. (2000). The role of the kis-balaton water protection system in the control of water quality of lake balaton. *Ecol. Eng.* 16, 73–78. doi: 10.1016/S0925-8574(00)00091-4

Thomas, R. (1979). Kinetic logic: a boolean approach to the analysis of complex regulatory systems. *Lect. Notes Biomath.* 29, 507.

Thomas, R. (1998). Laws for the dynamics of regulatory networks. *Int. J. Dev. Biol.* 42, 479–485.

Thomas, R., and d'Ari, R. (1990). *Biological Feedback*. Boca Raton, FL: CRC Press.

Thomas, R., Thieffry, D., and Kaufman, M. (1995). Dynamical behaviour of biological regulatory networks-i. biological role of feedback loops and practical use of the concept of the loop-characteristic state. *Bull. Math. Biol.* 57, 247–276.

Thomas, R. (1978). Logical analysis of systems comprising feedback loops. *J. Theor. Biol.* 73, 631–656.

Uusitalo, L. (2007). Advantages and challenges of bayesian networks in environmental modelling. *Ecol. Model.* 203, 312–318. doi: 10.1016/j.ecolmodel.2006.11.033

Wiley, E., McNyset, K. M., Peterson, A. T., Robins, C. R., and Stewart, A. M. (2003). Niche modeling and geographic range predictions in the marine environment using a machine-learning algorithm. *Oceanography* 16, 120–127. doi: 10.5670/oceanog.2003.42

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Appendix

TABLE A1 | Glossary of technical terms.

Terms	Descriptions
Kinetic model	An expansion of the boolean model to incorporate discrete values above "1."
Regulatory Network (RN)	A network of interconnected components which regulate the system.
State	A particular configuration of the regulatory network representing an instance of the system.
Resources	A set of components of the network which collectively influence another component.
Logical parameters	Discretized parameters which govern how the system evolves or progresses.
Bio-LHA	Parametric biological linear hybrid automata
State graph	A graph showing possible behaviors of the system by connecting different states together.
Trajectory	A successive series of connected states of the state graph, representing a particular behavior of the system.
Computation Tree Logic (CTL)	A logical language used in model checking and verification of systems.
Selection of Models of Biological Networks (SMBioNet)	A software used to apply CTL model checking to the RN.
Shortest path betweenness centrality (C_{spb})	A relative measure used to sort states of the state graph based on their frequent occurrences in trajectories.
System dynamics	Refers to the abstract number and diversity of behaviors that can exhibited by the system.
Automaton	Plural: Automata. Refers to a construct which can self operate.
Parametric biological linear hybrid automata (Bio-LHA)	A class of automata which simulates hybrid systems using affine linear dynamics.
Hybrid Technology (HyTech)	A software used to construct and simulate hybrid automata including Bio-LHA.

frontiers in Environmental Science

METHODS
published: 28 July 2015
doi: 10.3389/fenvs.2015.00053

Modeling and simulation of multi-scale environmental systems with Generalized Hybrid Petri Nets

Mostafa Herajy[1]* *and Monika Heiner*[2]

[1] Department of Mathematics and Computer Science, Faculty of Science, Port Said University, Port Said, Egypt, [2] Computer Science Institute, Brandenburg University of Technology Cottbus-Senftenberg, Cottbus, Germany

Predicting and studying the dynamics and properties of environmental systems necessitates the construction and simulation of mathematical models entailing different levels of complexities. Such type of computational experiments often require the combination of discrete and continuous variables as well as processes operating at different time scales. Furthermore, the iterative steps of constructing and analyzing environmental models might involve researchers with different background. Hybrid Petri nets may contribute in overcoming such challenges as they facilitate the implementation of systems integrating discrete and continuous dynamics. Additionally, the visual depiction of model components will inevitably help to bridge the gap between scientists with distinct expertise working on the same problem. Thus, modeling environmental systems with hybrid Petri nets enables the construction of complex processes while keeping the models comprehensible for researchers working on the same project with significantly divergent educational background. In this paper we propose the utilization of a special class of hybrid Petri nets, Generalized Hybrid Petri Nets ($\mathcal{GHPN}$), to model and simulate environmental systems exposing processes interacting at different time-scales. $\mathcal{GHPN}$ integrate stochastic and deterministic semantics as well as some other types of special basic events. To this end, we present a case study illustrating the use of $\mathcal{GHPN}$ in constructing and simulating multi-timescale environmental scenarios.

Keywords: modeling and simulation, Hybrid Petri Nets, multi-scale environmental systems, Chagas disease, Triatoma infestans

OPEN ACCESS

Edited by:
Christian E. Vincenot, Kyoto University, Japan

Reviewed by:
Guennady Ougolnitsky, Southern Federal University, Russia
Luis Gomez, University of Las Palmas de Gran Canaria, Spain

***Correspondence:**
Mostafa Herajy, Department of Mathematics and Computer Science, Faculty of Science, Port Said University, 23 December St., Port Said 42521, Egypt
mherajy@sci.psu.edu.eg

Specialty section:
This article was submitted to Environmental Informatics, a section of the journal Frontiers in Environmental Science

Received: *08 May 2015*
Accepted: *13 July 2015*
Published: *28 July 2015*

Citation:
Herajy M and Heiner M (2015) Modeling and simulation of multi-scale environmental systems with Generalized Hybrid Petri Nets. Front. Environ. Sci. 3:53. doi: 10.3389/fenvs.2015.00053

Introduction

The process of constructing and analyzing environmental systems is increasingly becoming a complex procedure (Seppelt et al., 2009; Uusitalo et al., 2015). On the one hand, it can require the amalgamation of different simulation techniques to accurately and efficiently find a solution to the problem under consideration (see e.g., Gillet, 2008; Gregorio et al., 1999). On the other hand, complex environmental systems require the collection and analysis of various data and information that cannot be tackled by researchers coming from just one area of expertise (Seppelt et al., 2009).

While the ordinary differential equations (ODEs) approach is widely used to construct and simulate many problems in the environmental domain, certain classes of such problems cannot be adequately addressed using this approach alone. For instance in Khoury et al. (2013) construct a simple, but elegant ODEs model to study food and population dynamics in honey bee colonies. However, such a continuous approach cannot capture the effect of seasonal variations

on many parameters. Contrary, in Schmickl and Crailsheim (2007) and Russell et al. (2013) a discrete simulation of recurrence and difference equations has been deployed to emulate the discrete changes in bee population taking into account seasonal variations. Nevertheless, certain scenarios necessitate the interplay of different simulation strategies to efficiently and accurately simulate a given problem. For example, the ODE approach can be used to efficiently execute model components with fast dynamics where elucidated discrete simulation does not affect the result, while stochastic simulation has to be used to discretely model components whose individual and random occurrence plays a key role for the overall result. Such type of models possess more than one time scale. Therefore, it requires hybrid simulation to successfully reproduce their dynamics. This results in two or more simulation regimes which have to work simultaneously to solve a given problem. However, these regimes are not isolated, instead, they closely interact and influence the dynamics of each other (Herajy and Heiner, 2012).

Furthermore, with the increasing demand for interdisciplinary science, modeling complex environmental systems may involve researchers with different scientific and educational background. For instance, researchers from ecosystems, mathematics, and computer science may collaborate in constructing and analyzing a computational experiment. Nonetheless, maintaining the communication in such interdisciplinary teams is one of the key issues in constructing environmental models (Seppelt et al., 2009). Thus, a visual language may be of help to accelerate the communication between team members with diverse professional background. As an example, consider the problem of water resource management. Numerical modeling and simulation play a remarkable role in predicting future water demand as well as managing water quality (Qi and Chang, 2011; Liu et al., 2015). However, the procedure of constructing a realistic and accurate model for this purpose mandates that a team of experts coming from different fields (e.g., environmental science, mathematics, geography, hydrological modeling, and computer science) collaborate closely together.

One of those modeling tools that can contribute in overcoming these challenges are Petri nets. Petri nets (Murata, 1989) are a visual modeling language highly suitable to model concurrent, asynchronous and distributed systems. In addition to their graphical representation, Petri nets enjoy a well established mathematical theory to analyze the constructed model. However, the basic place/transition nets are not very helpful in constructing and executing quantitative models exposing certain level of complexities. Therefore, many extensions have been proposed over the years to overcome these limitations. For instance, continuous Petri nets (Alla and David, 1998) can be used as an alternative technique which exactly corresponds to the ODEs approach (Gilbert and Heiner, 2006; Soliman and Heiner, 2010). Similarly, stochastic Petri nets (Ajmone et al., 1995) provide a graphical tool to permit the stochastic exploration of a constructed model. Nowadays, a variety of Petri nets with different extensions have been used to model various technical and biological systems (e.g., see Reddy et al., 1993; Matsuno et al., 2003; Fujita et al., 2004; Herajy et al., 2013).

Hybrid Petri Nets ($\mathcal{HPN}$) (David and Alla, 2010) are another interesting class of Petri nets. $\mathcal{HPN}$ permit the integration of discrete and continuous variables (places) in addition to discrete and continuous processes (transitions) into one model. In a typical scenario, discrete places serve as signals that control the firing of continuous transitions. $\mathcal{HPN}$ allow the efficient simulation of systems which entail large number of states by approximating them via continuous simulation, while discrete events can be pertained using discrete transitions. In Matsuno et al. (2003), $\mathcal{HPN}$ are adapted to provide a very specific approach dedicated to the simulation of biological systems. In general, hybrid modeling using $\mathcal{HPN}$ is a promising technique since it permits the simulation of more complex systems (see e.g., Tian et al., 2013). Furthermore, models with interacting components working at different scales can be easily executed via $\mathcal{HPN}$. Nevertheless, so far, little attention has been paid to the employment of this approach in the context of modeling environmental systems.

In this paper, we focus on a particular class of hybrid Petri nets, Generalized Hybrid Petri nets ($\mathcal{GHPN}$) (Herajy and Heiner, 2012), as a promising tool for model-based exploration of environmental systems. $\mathcal{GHPN}$ provide various transitions, arcs, and places, which together have the power to substantially facilitate the modeling of different processes in the environmental science. One important aspect of $\mathcal{GHPN}$ is their ability to simulate systems that expose different time scales: fast and slow. The former time scale is continuously simulated, while the latter one is stochastically and individually executed. Furthermore, the interaction between continuous and stochastic dynamics is appropriately captured. We illustrate the use of $\mathcal{GHPN}$ in modeling environmental systems via a case study, the Chagas disease infection cycle. All the discussed features in this paper are implemented in a general platform-independent Petri net editing tool called Snoopy (Heiner et al., 2012) which can be downloaded free of charge for academic use from Snoopy (2015).

The rest of this paper is organized as follows: after this introduction, the different aspects of $\mathcal{GHPN}$ are discussed by presenting a formal definition as well as the different modeling elements of $\mathcal{GHPN}$. Afterwards, the main steps involved in the simulation of $\mathcal{GHPN}$ are briefly summarized. In the Result section, we provide a case study to illustrate the use of $\mathcal{GHPN}$ for modeling environmental systems, namely the simulation of infection transmission of Chagas disease. This example explains the motivation behind most of the $\mathcal{GHPN}$ modeling components. Finally, we conclude with a few remarks concerning the utilization of $\mathcal{GHPN}$ to implement the simulation of multi-scale environmental models.

Methods

In this section we provide an overview of $\mathcal{GHPN}$ including the formal definition as well as their different modeling elements. We concentrate in this part on the use of $\mathcal{GHPN}$ for the modeling of environmental systems. Thus, the semantics of places, transitions and arcs are discussed according to this context.

Generalized Hybrid Petri Nets

Elements

As a Petri net class, the specification of $\mathcal{GHPN}$ involves defining the three main components, namely: places, transitions, and arcs. **Figure 1** illustrates the different entities that can be found in a typical $\mathcal{GHPN}$ model. In the sequel, we briefly discuss the semantics and usage of each constituent.

Places

Places correspond to the model variables. They are further classified into discrete and continuous. On the one hand, discrete places are drawn as single line circles. They are used to represent discrete variables (e.g., the number of trees in a forest, the number of eggs laid by a bee, or a species population). Discrete places can hold nonnegative integer numbers called tokens. On the other hand, continuous places are drawn with shaded line circles and are used to depict continuous variables (e.g., the amount of water in a lake, the concentration of contaminated water, or the number of infected individuals in an epidemic model). Therefore, they can hold nonnegative real values. In certain modeling scenarios, continuous places serve as an approximation of discrete places where the numbers of tokens reach large values. The value assigned to a place is called *place marking*. In $\mathcal{GHPN}$ models, the system state is described at any time point during the simulation as the union of discrete and continuous place marking.

Transitions

Transitions correspond to the basic events. $\mathcal{GHPN}$ employ five transition types for convenient modeling of different types of systems: stochastic, immediate, deterministically time delayed, scheduled, and continuous transitions. The first four transition types are discrete ones. However, they differ from each other by the time delay assigned to them.

Stochastic transitions fire at discrete time steps, however, after random time delays. These random delays are exponentially distributed. Stochastic transitions can represent events that take place at random time steps. During execution, the simulator calculates the time at which the next event will occur, and subsequently it decides the event type (which transition to fire). Theoretically, an effective conflict (David and Alla, 2010) between two stochastic transitions is not possible in a such random firing scheme. Immediate transitions are also fired in a discrete manner, but with zero delays. They fire directly as soon as they are enabled. Similarly, deterministically time delayed transitions are fired after a deterministic time delay. The delay of this transition type could be set to zero. Nevertheless, when an immediate transition and a deterministically delayed transition are concurrently enabled, the immediate one will have higher priority to fire first. Moreover, scheduled transitions are a special type of deterministically time delayed transitions which fire at certain time point(s) previously programmed by the user.

In contrast, continuous transitions fire continuously with respect to time. The firing speeds of continuous transitions are specified by their rates. Besides, the semantics of continuous transitions is represented by a set of ODEs that account for in- and outflow of each place.

Arcs

Arcs model the relation between the model variables and the basic events. Arcs connect places with transitions and maybe vice versa

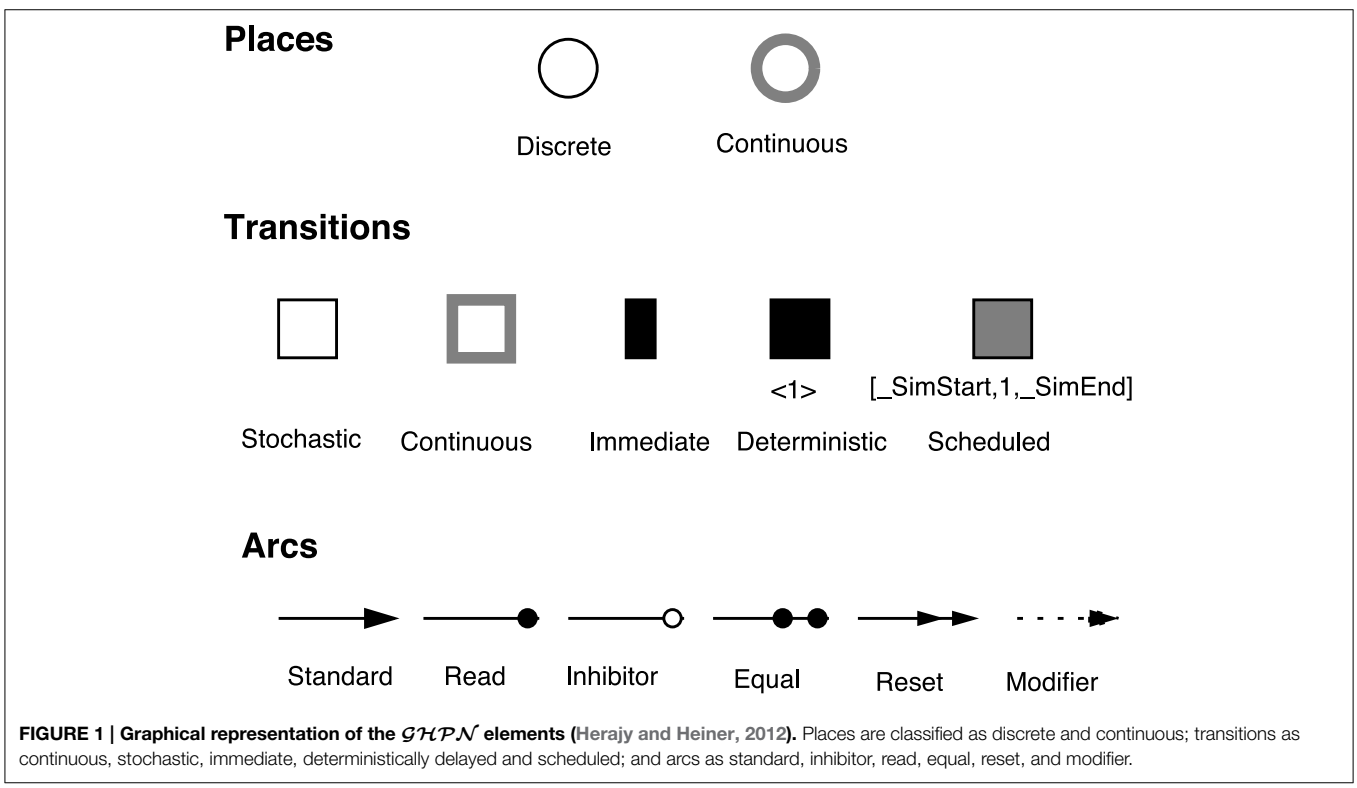


FIGURE 1 | Graphical representation of the $\mathcal{GHPN}$ elements (Herajy and Heiner, 2012). Places are classified as discrete and continuous; transitions as continuous, stochastic, immediate, deterministically delayed and scheduled; and arcs as standard, inhibitor, read, equal, reset, and modifier.

depending on their type. There are six type of arcs in $\mathcal{GHPN}$: standard, read, inhibitor, equal, reset, and modifier arcs.

Standard arcs connect places with transitions and vice versa. They control the enabling of the target transition as well as affecting the preplaces (postplaces) when the target (source) transition fires. Standard arcs can be discrete or continuous. Discrete arcs adopt positive integer values as arc weights, while continuous arcs use positive rational numbers as arc weights. The rules that determine the type of arc weights are illustrated in **Figure 2**.

In contrast, read arcs affect only the enabling of the target transition. A transition connected with a preplace via a read arc is enabled only (with respect to this preplace), if the marking of this preplace is greater than or equal to the corresponding arc weight. Similarly, inhibitor arcs govern the enabling of transitions. However, a transition connected with a preplace using an inhibitor arc is enabled only if the current marking of the preplace is less than the arc weight. Equal arcs enforce more stronger conditions on the enabling of transitions. A transition connected with a preplace via an equal arc is enabled only if the current marking of the preplace is exactly equal to the arc weight.

The other two remaining arcs do not influence the enabling of the connected transitions. For example, reset arcs set the value of the preplace marking to zero when the corresponding transition fires. They are useful to implement certain model semantics. Similarly, modifier arcs do not affect the enabling nor the firing of a transition. They facilitate the use of a preplace in defining a transition rate function while preserving the structure-related constraints of the transitions' rate functions.

Marking-dependent Arc Weights

$\mathcal{GHPN}$ permit arc weights to be specified as an algebraic expression involving place names rather than just a constant. This feature is called marking-dependent arc weights (Valk, 1978; Matsuno et al., 2003; Herajy et al., 2013). Implementing certain model semantics without the help of marking-dependent arc weights may become intricate and even impossible in certain circumstances.

For instance, consider the following model expression that requires to be implemented using Petri nets:

```
IF x>=y AND y<5 THEN
     z:=x+y
 END IF
```

When x, y, and z are continuous variables, it is impossible to represent the above semantics using just arcs with constant weights. However, using marking-dependent arc weights this can be easily modeled as it is depicted in **Figure 3**.

As another interesting example of the usefulness of marking-dependent arcs consider the transformation of the whole population from one age category to another one after an elapsed period of time. For example in Xiang et al. (2013), new-born snails are considered as old snails at the beginning of the year. This process can be intuitively modeled using marking-dependent arc weights where the outgoing and the ingoing arc weights equal the value of the preplace.

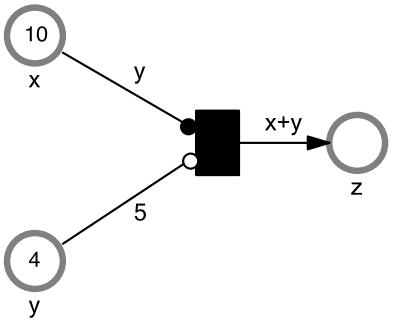


FIGURE 3 | A $\mathcal{GHPN}$ example of using marking-dependent arc weights. The example contains three places x, y, z and one immediate transition t_1. The transition t_1 can fire only when the two conditions: $x \geq y$ (enforced by the read arc) and $y < 5$ (enforced by the inhibitor arc) hold.

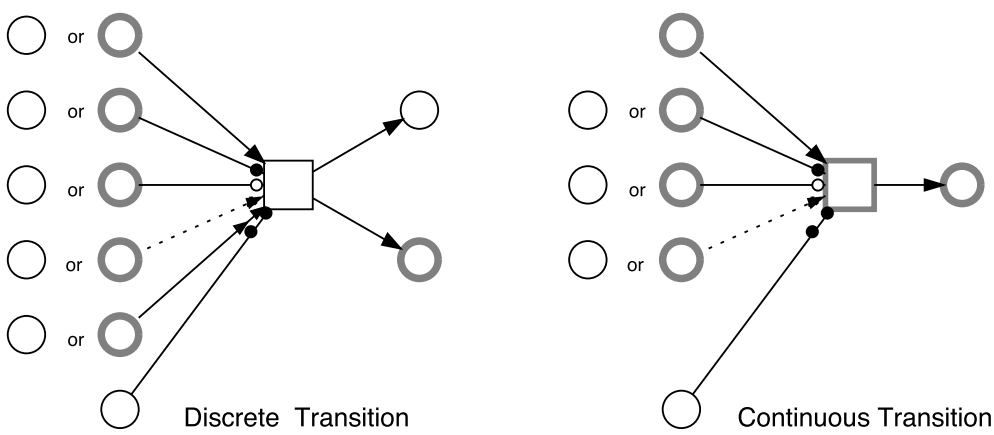


FIGURE 2 | Possible connections between $\mathcal{GHPN}$ elements. The (obvious) restrictions are: discrete places cannot be connected with continuous transitions using standard arcs, continuous places cannot be tested with equal arcs, and continuous transitions cannot use reset arcs.

Obviously, marking-dependent arc weights extend $\mathcal{GHPN}$ and permit the modeling of a larger class of environmental systems that require the corresponding semantics.

Formal Definition

In this section we formally define the syntax of $\mathcal{GHPN}$. The formal semantics including the enabling and firing rules as well as the conflict resolution are given in Herajy and Heiner (2012).

Definition 1 (Generalized Hybrid Petri Nets). *Generalized Hybrid Petri Nets are a 6-tuple* $\mathcal{GHPN} = [P, T, A, F, V, m_0]$, *where P, T are finite, non-empty and disjoint sets. P is the set of places, and T is the set of transitions with:*

- $P = P_{disc} \cup P_{cont}$ *whereby* P_{disc} *is the set of discrete places to which non-negative integer values are assigned, and* P_{cont} *is the set of continuous places to which non-negative real values are assigned.*
- $T = T_D \cup T_{cont}$, $T_D = T_{stoch} \cup T_{im} \cup T_{timed} \cup T_{scheduled}$ *with:*
  1. T_{stoch} *is the set of stochastic transitions, which fire randomly after exponentially distributed waiting time.*
  2. T_{im} *is the set of immediate transitions, which fire with waiting time zero; they have highest priority among all transitions.*
  3. T_{timed} *is the set of deterministically delayed transitions, which fire after a deterministic time delay.*
  4. $T_{scheduled}$ *is the set of scheduled transitions, which fire at predefined time points.*
  5. T_{cont} *is the set of continuous transitions, which fire continuously over time.*
- $A = A_{disc} \cup A_{cont} \cup A_{inhibit} \cup A_{read} \cup A_{equal} \cup A_{reset} \cup A_{modifier}$ *is the set of directed arcs, with:*
  1. $A_{disc} \subseteq ((P \times T) \cup (T \times P))$ *defines the set of discrete arcs.*
  2. $A_{cont} \subseteq ((P_{cont} \times T) \cup (T \times P_{cont}))$ *defines the set of continuous arcs.*
  3. $A_{read} \subseteq (P \times T)$ *defines the set of read arcs.*
  4. $A_{inhibit} \subseteq (P \times T)$ *defines the set of inhibits arcs.*
  5. $A_{equal} \subseteq (P_{disc} \times T)$ *defines the set of equal arcs.*
  6. $A_{reset} \subseteq (P \times T_D)$ *defines the set of reset arcs,*
  7. $A_{modifier} \subseteq (P \times T)$ *defines the set of modifier arcs.*
- *the function F*

$$F: \begin{cases} A_{cont} \rightarrow D_q, \\ A_{disc} \rightarrow D_n, \\ A_{read} \rightarrow D_q, \\ A_{inhibit} \rightarrow D_q, \\ A_{equal} \rightarrow D_n, \\ A_{reset} \rightarrow \{1\}, \\ A_{modifier} \rightarrow \{1\}. \end{cases}$$

assigns a marking-dependent function to each arc, where D_n *and* D_q *are sets of functions defined as follows:*

$$D_n = \{d_n | d_n : \mathbb{N}_0^{|{}^\bullet t_j|} \rightarrow \mathbb{N}, t_j \in T\},$$

$$D_q = \{d_q | d_q : \mathbb{R}_0^{|{}^\bullet t_j|} \rightarrow \mathbb{Q}^+, t_j \in T\}.$$

- *V is a set of functions* $V = \{g, d, w, f\}$ *where :*
  1. $g : T_{stoch} \rightarrow H_s$ *is a function which assigns a stochastic hazard function* h_{s_t} *to each transition* $t_j \in T_{stoch}$, *whereby* $H_s = \{h_{s_t} | h_{s_t} : \mathbb{R}_0^{|{}^\bullet t_j|} \rightarrow \mathbb{R}_0^+, t_j \in T_{stoch}\}$ *is the set of all stochastic hazard functions, and* $g(t_j) = h_{s_t}, \forall t_j \in T_{stoch}$.
  2. $w : T_{im} \rightarrow H_w$ *is a function which assigns a weight function* h_w *to each immediate transition* $t_j \in T_{im}$, *such that* $H_w = \{h_{w_t} | h_{w_t} : \mathbb{R}_0^{|{}^\bullet t_j|} \rightarrow \mathbb{R}_0^+, t_j \in T_{im}\}$ *is the set of all weight functions, and* $w(t_j) = h_{w_t}, \forall t_j \in T_{im}$.
  3. $d : T_{timed} \cup T_{scheduled} \rightarrow \mathbb{R}_0^+$, *is a function which assigns a constant time to each deterministically delayed and scheduled transition representing the (relative or absolute) waiting time.*
  4. $f : T_{cont} \rightarrow H_c$ *is a function which assigns a rate function* h_c *to each continuous transition* $t_j \in T_{cont}$, *such that* $H_c = \{h_{c_t} | h_{c_t} : \mathbb{R}_0^{|{}^\bullet t_j|} \rightarrow \mathbb{R}_0^+, t_j \in T_{cont}\}$ *is the set of all rates functions and* $f(t_j) = h_{c_t}, \forall t_j \in T_{cont}$.
- $m_0 = m_{disc} \cup m_{cont}$ *is the initial marking for both the continuous and discrete places, whereby* $m_{cont} \in \mathbb{R}_0^{|P_{cont}|}$, $m_{disc} \in \mathbb{N}_0^{|P_{disc}|}$.

  Here, $\mathbb{N}_0$ *denotes the set of non-negative integer numbers,* $\mathbb{R}_0$ *denotes the set of non-negative real numbers,* $\mathbb{Q}^+$ *denotes the set of positive rational numbers, and* ${}^\bullet t_j$ *denotes the set of pre-places of a transition* t_j. □

A distinguishing feature of $\mathcal{GHPN}$ compared with other hybrid Petri net classes is its support of the full interplay between stochastic and continuous transitions. Such interplay is implemented by updating and monitoring the rates of stochastic transitions The crucial point for our paper is how stochastic transitions are simulated when mixed with continuous ones. So the next section focuses in particular on the simulation of stochastic transitions, while numerically solving the set of ODEs induced by the continuous transitions (for more details see Herajy and Heiner, 2012). By this way, accurate results are obtained during simulation.

Simulation

The simulation of a $\mathcal{GHPN}$ model has to take into account the different types of $\mathcal{GHPN}$ transitions. Although it is easy to simulate individual transition types when they are isolated, it becomes more challenging to simulate a model combining discrete and continuous transitions. Thus, the most important aspect is how discrete and continuous transitions are interleaved during the simulation, particularly, stochastic and continuous ones.

Continuous transitions are fired continuously. Thus, they necessitate the simultaneous (numerical) solution of a system of ODEs representing the continuous part of a $\mathcal{GHPN}$ model. From this perspective, the simulation of discrete transitions are considered as events which are triggered whenever a discrete transition is enabled and needs to be fired. Therefore, we have different events corresponding to each transition type. When an event occurs, a dispatcher is called to handle the

appropriate actions. The corresponding system of ODEs is generated using (1).

$$\frac{dm(p_i)}{d\tau} = \sum_{t_j \in {}^\bullet p_i} F(t_j, p_i) \cdot v_j(\tau) \cdot read(u, m(p_i)) \cdot inhibit(u, m(p_i)) - \sum_{t_j \in p_i^\bullet} F(p_i, t_j) \cdot v_j(\tau) \cdot read(u, m(p_i)) \cdot inhibit(u, m(p_i)) \quad (1)$$

where $m(p_i)$ represents the marking of the place p_i, $v_j(\tau) = f_j$ is the marking-dependent rate function of the continuous transition t_j, and the functions $read(u, p_i)$, $inhibit(u, p_i)$, which consider the effects of read and inhibitor arcs, respectively, are defined as follows:

For a given transition $t_j \in T^C$,

$$read(u, m(p_i)) = \begin{cases} 1 & \text{if } m(p_i) \geq u \\ 0 & \text{else} \end{cases}$$

with $u = F(p_i, t_j) \wedge (p_i, t_j) \in A_{read}$, and

$$inhibit(u, m(p_i)) = \begin{cases} 1 & \text{if } m(p_i) < u \\ 0 & \text{else} \end{cases}$$

with $u = F(p_i, t_j) \wedge (p_i, t_j) \in A_{inhibit}$.

Furthermore, two issues are of paramount importance concerning this simulation procedure: how an event is detected during the numerical solution of the set of ODEs and how we know that a stochastic transition is enabled and needs to be fired.

Concerning the former issue, a special type of ODE solver should be used that supports a root finding feature (Mao and Petzold, 2002). The occurrence of enabling conditions of discrete transitions are then formulated as a root that can be detected by the ODE solver. As soon as a root is encountered by the ODE solver, the control is transferred to the discrete regime to fire the enabled transition(s). Afterwards, the ODE solver continues the integration using the new system state.

Moreover, stochastic transitions are considered as a special type of discrete events called stochastic events. Stochastic events are detected by introducing a new ODE, described by Equation (2), to the set of ODEs.

$$g(\mathbf{x}) = \int_t^{t+\tau} a_0^s(\mathbf{x})dt - \xi = 0\,, \quad (2)$$

where ξ is a random number exponentially distributed with a unit mean, and $a_0^s(\mathbf{x})$ is the cumulative (the sum) rate of all stochastic transitions.

The newly added ODE monitors the difference between the summation of all the rates of the stochastic transitions and a small, exponentially distributed random number. When Equation (2) equals zero, the continuous simulation is interrupted to call the dispatcher to fire the enabled stochastic transition. Afterwards, the simulation is resumed as previously discussed.

Results

In this section we apply $\mathcal{GHPN}$ to model and simulate a case study from the environmental domain. $\mathcal{GHPN}$ can be used for models which are completely deterministic, completely stochastic, or a combination of them. The chosen example illustrates the use of $\mathcal{GHPN}$ to represent and simulate the dynamics of environmental and ecological systems. We show how stochastic and continuous transitions are used to provide the interplay between a discrete regime representing the environment fluctuations and a deterministic one representing the simulation of large populations. Additionally, deterministically time delayed transitions and immediate transitions proved to be useful in modeling real-life examples.

Modeling the Transmission of Chagas Disease Infection

Background

The Chagas disease has been a major public health concern in Latin America for some decades (Nouvellet et al., 2015). The transmission of Chagas infection among humans involves complex ecological and epidemiological interacting processes (Cohen and Gürtler, 2001; Nouvellet et al., 2015). The Chagas disease is caused by the protozoan Trypanosoma Cruzi (T. Cruzi for short). The main insect vector responsible for the transmission of T. Cruzi is a bug known as Triatoma infestans (Cohen and Gürtler, 2001; Castañera et al., 2003). A vector is an insect that transmits a disease, while the disease transmitted via such an insect is referred to as a vector-borne disease. Vectors are living organisms that can transmit infectious diseases between humans or from animals to humans. Many of these vectors are bloodsucking insects. Within a household, Chagas disease is mainly transmitted to humans via the bitting by infected bugs (Castañera et al., 2003). Bugs acquire infections by the feeding on infected mammals (humans or dogs) (Cohen and Gürtler, 2001). Chickens are another feeding source for Triatoma infestans. However, blood meals taken from chickens do not transmit the infection to bugs. Therefore, chickens can serve as an alternative feeding source to Triatoma infestans such that bitting rates of vectors to humans and infected dogs are minimized (Cohen and Gürtler, 2001). Nevertheless, the four species involved in the Chagas disease cycle are: humans, dogs, chickens, and infected vectors. Besides, the population of Triatoma infestans is oscillating seasonally with the highest population of vectors recorded in warm seasons (spring and summer) (Cohen and Gürtler, 2001).

Mathematical modeling of the transmission of the Chagas disease is an important tool to understand the biological and ecological factors influencing the spread of infections among humans and household animals. To this end, many mathematical models have been constructed (see e.g., Cohen and Gürtler, 2001; Castañera et al., 2003; Coffield et al., 2013; Nouvellet et al., 2015). However, all of these models utilize solely either the deterministic or the stochastic approach. For the former modeling paradigm, authors argue that the population size of the interacting species is large enough so that the ODE approach can be deployed to study the model dynamics (Coffield et al., 2013). In contrast,

the latter models assume that the population size of interacting species is relatively small. Thus, stochastic simulation will be more accurate (Castañera et al., 2003). For instance, in Cohen and Gürtler (2001) the number of humans of a household consists of just five persons divided into different age categories. Each category contains only one human. Nevertheless, modeling the transmission of the Chagas disease can encompass variables interacting at different time scales. For instance, vertebrate species (humans, dogs, and chickens) can be found in scales of tens or hundreds at the very most, because the majority of realistic models operate on the level of small villages. In contrast, vector population is abundant and exists at the scale of thousands. Thus, hybrid modeling is a desirable approach worth being investigated to gain deeper understanding of the Chagas disease transmission.

Model Specification

In this section we use $\mathcal{GHPN}$ to model the transmission of Chagas infection between humans, dogs, and vectors. Our $\mathcal{GHPN}$ model is based on the deterministic one by Coffield et al. (2013) as it accounts for the high-level transmission of infections without considering the detailed stages of nymph bugs.

Coffield et al. (2013) simulated the evolution of the total population of vectors, humans, and dogs involved in the transmission of Chagas disease. The chicken population is considered to be constant. Their population change is not taken into account since they do not acquire infection. **Figure 4** provides a $\mathcal{GHPN}$ representation of the Chagas transmission cycle. To simplify the discussion, we divide the human population into two groups: infected (denoted by the place H_i), and susceptible (denoted by the place H_s). Similarly, we divide

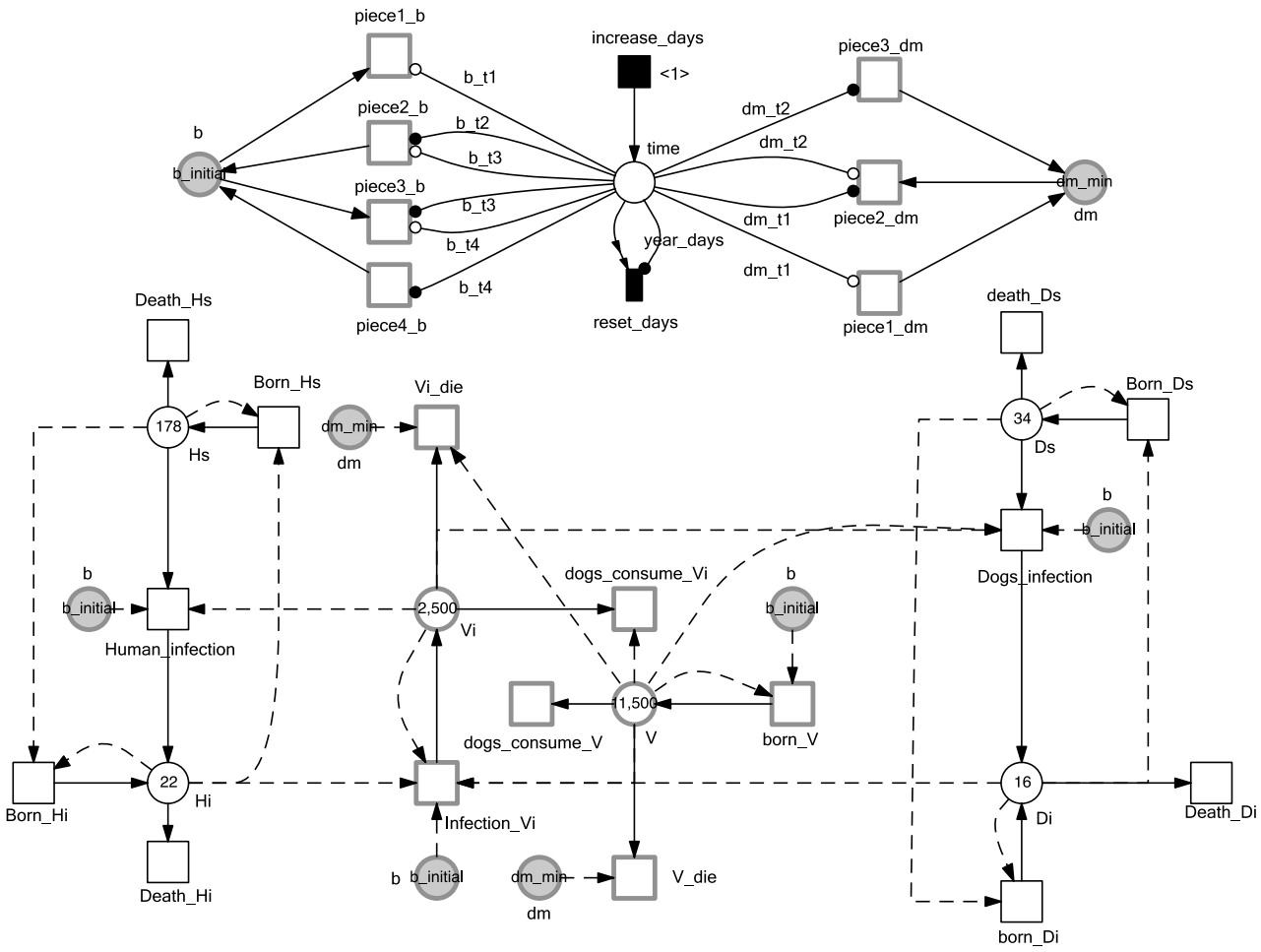


FIGURE 4 | $\mathcal{GHPN}$ model of Chagas disease transmission. Continuous and discrete places are used to model the population of vectors and mammals, while continuous transitions are adopted to represent the processes operating on the model species. The simulation time is monitored by the discrete place *time* which is increased by one time step (one day) when the deterministically delayed transition *increase_days* fires. Stochastic and continuous transitions describe physical processes that operate on the model species. Places given in gray are logical places which help to simplify the connections between model components. Please note the use of modifier arcs to include non-preplaces into the transition rate functions. Modifier arcs make this kind of dependency explicit. Moreover, arc weights and initial markings specified by constants make the model easy to configure using different constant values.

the population of dogs into infected dogs (D_i), and susceptible ones (D_s). In contrast, we consider the total population of vectors (V) and the infected ones (V_i) to minimize the connection among the model components. We adopt the concept of logical places (places represented in gray colors) to keep the connection between model components intelligible. Moreover, the current simulation time (in days) is represented by the discrete place: *time*. The deterministically time delayed transition *increase_days* increases the simulation time by a one-day step. The value of the place *time* is reset after a duration of 365 days. Read and reset arcs as well as the immediate transition *reset_days* implement the reset semantics of the current year as it can be seen in **Figure 4**. Furthermore, continuous and stochastic transitions model the dynamics of the different processes involved in the Chagas disease infection cycle.

In the sequel we elucidate the definition of each transition rate. Moreover, we discuss our motivation of modeling certain processes as stochastic transitions and others as continuous ones by showing the effect of the random firing of stochastic transitions on the overall dynamic results. Similar to Coffield et al. (2013), we consider the total population of humans, dogs, and chickens as being constant during the whole simulation period. The total human population is considered to be roughly constant as the sum of the number of infected humans and the number of susceptible humans does basically not change, if we assume equal rates for birth and death. Likewise, the dog population is also considered to be constant. However, the number of chickens are not divided into infected and susceptible, since chickens cannot be infected. More information is provided in the equations below.

First, the growth of the total vector population is defined by Equation (3) (Coffield et al., 2013).

$$d_h \times (V) \times (1 - \frac{V}{K}) \tag{3}$$

Where d_h is the vector hatching rate, and K specifies the maximum number of bugs that can be supported in a village. Equation (3) is used to define the rate of the transition *born_V*. The hatching rate coefficient d_h is defined in terms of the biting rates b, which is varying seasonally (see below). Please note that we assume that the rate at which vectors hatch at time t is equal to the number of eggs laid at time $t+\tau$.

Furthermore, vectors undergo two types of death: degradation due to natural death and by the oral consumption by dogs (Coffield et al., 2013). These two processes are represented by the two transitions: *dogs_consume_V*, and *V_die*, respectively. The rates of the transitions *dogs_consume_V*, and *V_die* are defined by Equations (4) and (5), respectively.

$$\left(\frac{E \times V}{V + A}\right) \times D \tag{4}$$

$$\left(\frac{d_m}{2} \times \left(1 - \frac{V}{K}\right) + d_m\right) \times V \tag{5}$$

where E is the maximum number of vectors consumed by one dog per day, A is the vector number at which dogs consume at the rate $E/2$ vectors per day, and d_m is the mortality rate coefficient.

The mortality rate coefficient of infected and uninfected vectors is not constant. Instead, it is changing with respect to time according to the current season (Castañera et al., 2003; Coffield et al., 2013). **Figure 5** illustrates the time-dependent mortality rate of Triatoma infestants, while **Figure 6** is a Petri net sub model used to reproduce the piecewise function in **Figure 5**. To model the seasonal variation in mortality rate, we adopt read and inhibitor arcs to define the time period of each piece of the piecewise function. The current value of the mortality rate is represented by the continuous place d_m. Two constant values, *dm_initial*, *dm_max* are used to denote the

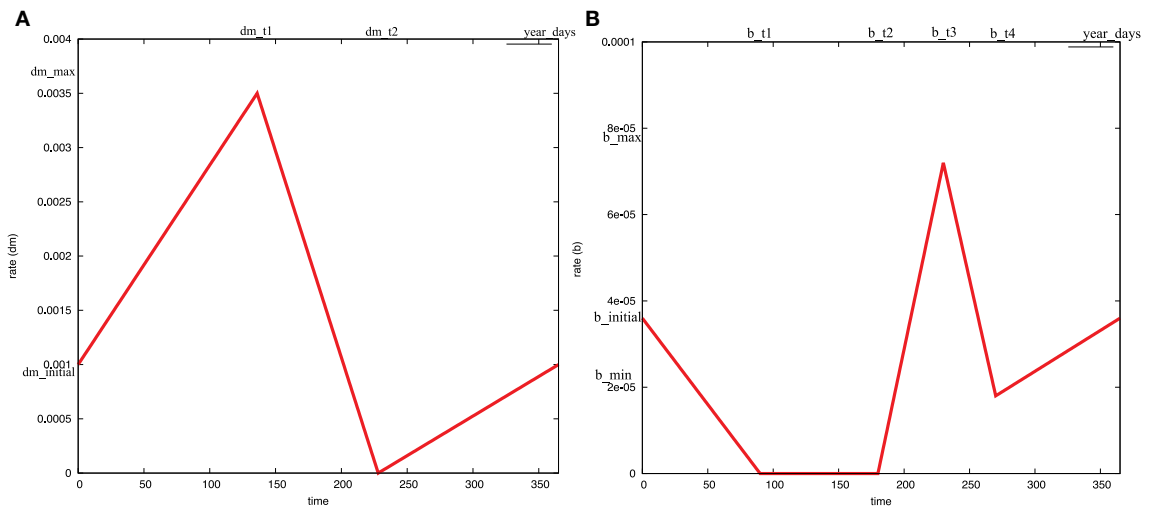


FIGURE 5 | The effects of seasonal variability on: (A) the mortality rate coefficient, and (B) the biting rate coefficient. These curves can be modeled as piecewise linear functions (Coffield et al., 2013). They can be produced using the Petri net submodel in **Figure 6**. The time boundaries where the functions change their behavior from decreasing to increasing or vice versa is shown in the x-axis. The exact time points where the functions change their behavior is illustrated in the upper axis. Similarly, the start and end values of each piece is shown in the y-axis. The exact values of these parameters are given in the Supplementary Material.

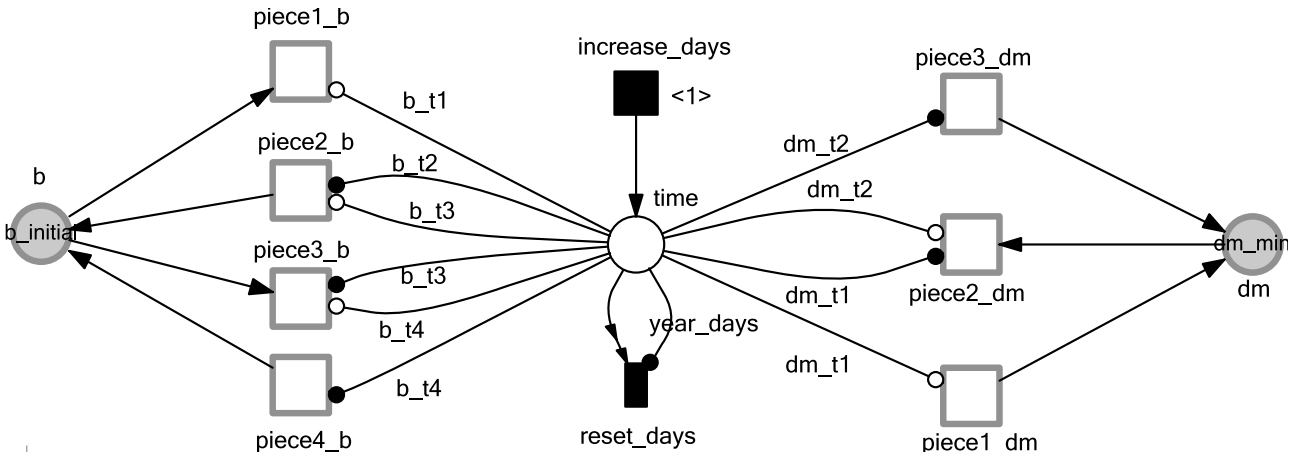


FIGURE 6 | $\mathcal{GHPN}$ submodel to reproduce the seasonal variability on mortality and biting rates. The deterministically time delayed transition, *increase_days*, fires at each time step increasing the current simulation time by one day. The immediate transition *reset_days* resets the current time to zero when it reaches the maximal number of days in a year. Continuous transitions are used to simulate the various pieces of the piecewise linear functions describing the mortality and biting rates. All labels at arcs or places are constants; compare caption of **Figure 4**.

initial and maximum values of *dm*, respectively. The *x*-axis of the piecewise function in **Figure 5** is divided into three intervals $[0, dm\_t_1[$, $[dm\_t_1, dm\_t_2[$, and $[dm\_t_2, year\_days]$. Where *year_days* denotes the number of days per year (in our model we consider each year to consist of 365 days). Read arcs are used to specify the interval's lower value, while inhibitor arcs are used to specify the interval's upper values. Afterwards, each continuous transition *piece*1_*dm*, *piece*2_*dm*, and *piece*3_*dm* get assigned the rates, $(dm\_max - dm\_initial)/dm\_t_1$, $dm\_max/(dm\_t_2 - dm\_t_1)$, and $dm\_inial/(year\_days - dm\_t_2)$, respectively. For the simulation results in this paper we assign the values 0.0003, 0.0017,136, 228, and 365 to *dm_initial*, *dm_max*, dm_t_1, dm_t_2, and *year_days* respectively. A similar procedure is applied to capture the seasonal variation in the biting rate, as it can be noted in **Figures 5**, **6**. The complete list of all constant values is provided in the Supplementary Table 1.

Similarly, the infected vector population can grow, naturally die, or be consumed by dogs. The increase of infected bugs is a result of the transmission of T. Cruzi parasites to some of the uninfected vectors. In our model, this process is represented by the transition $Infection\_V_i$ with a firing rate defined by Equation (6).

$$b \times (V - V_i) \times (P_{hv} \times H_i + P_{dv} \times d_f \times D_i) \tag{6}$$

where P_{hv} is the human to vector infection probability, P_{dv} is the dogs to vector infection probability, and d_f is the human factor of one dog.

Moreover, the natural death of vectors and the loss of vectors due to the consumption by dogs are modeled by the two transitions: *Vi_die* and $dogs\_consume\_V_i$, respectively. The rate of *Vi_die* is defined by Equation (7), similar to the death of the total vectors V, while the rate of *dogs_consume_Vi* is defined by Equation (8).

$$\left(\frac{d_m}{2} \times \left(1 - \frac{V}{K}\right) + d_m\right) \times V_i \tag{7}$$

$$\frac{E \times D \times V_i}{V + A} \tag{8}$$

Now we consider the dynamics of humans and dogs. Susceptible humans (H_s) can be bitten by vectors and become infected (H_i). The infection process is denoted by the transition *Human_infection*. The firing rate of this transition is given by Equation (9)

$$b \times P_{vh} \times H_s \times V_i \tag{9}$$

where P_{vh} is the probability of a susceptible human to be infected. A human infected by Chagas disease unfortunately cannot be recovered in the future. Both susceptible and infected humans can die with rates defined by Equations (10) and (11), respectively.

$$\gamma_{H_s} \times H_s \tag{10}$$

$$\gamma_{H_i} \times H_i \tag{11}$$

where γ_{H_s}, and γ_{H_i} are the mortality rates of susceptible and infected humans, respectively. Equations (10) and (11) define the rates of the transitions: *death_Hs*, and *death_Hi*, respectively.

Under the assumption that the number of humans are constant during the whole simulation period, the growth rate of susceptible and infected humans can be made equal to their corresponding death rate. However, according to Coffield et al. (2013), infection can be transferred from a mother to her fetus. Thus, we can model the growth of susceptible and infected human using Equations (12) and (13), respectively,

$$(1 - T_{hi}) \times (\gamma_{H_i} \times H_i + \gamma_{H_s} \times H_s) \tag{12}$$

$$(T_{hi}) \times (\gamma_{H_i} \times H_i + \gamma_{H_s} \times H_s) \quad (13)$$

where T_{ni} is the congenital transmission probability for infected humans. Equations (12) and (13) imply that we take T_{ni}, the total of died humans (infected and susceptible) as new born infected humans, while the remaining $1 - T_{ni}$ are added to the suspected humans.

Likewise, susceptible dogs can be infected with T. Cruzi parasites. However, an infection is transmitted to dogs either by the bitting by vectors or by the oral consumption of infected vectors by dogs. This process is modeled by the transition *dogs_infection* in **Figure 4**. The transition rate is given by Equation (14).

$$b \times d_f \times P_{vdb} + \frac{P_{vdc} \times E \times D_s \times V_i}{V + A} \quad (14)$$

where P_{vdb} denotes the vector to dog infection probability, and P_{vdc} the vector to dog infection probability via oral consumption. Obviously, the first term of Equation (14) represents the dogs infection via bug biting while the second term represents dog infection via oral consumption. Similar to humans, dogs (susceptible and infected) may die. The death of susceptible and infected dogs is represented by the transitions *death_Ds*, and *born_Di*, respectively. The firing rates of these transitions are given by Equations (15) and (16), respectively.

$$\gamma_{D_s} \times D_s \quad (15)$$

$$\gamma_{D_i} \times D_i \quad (16)$$

Similar to humans, and under the assumption that the overall dog population is constant during the whole simulation period, we set the rate of growth equal to the rate of death. However, new born dogs can be infected if they are born to an infected mother. Thus, the growth rates of susceptible and infected dogs are defined by Equations (17) and (18), respectively.

$$(1 - T_{di}) \times (\gamma_{Di} \times D_i + \gamma_{D_s} \times D_s) \quad (17)$$

$$(T_{di}) \times (\gamma_{Di} \times D_i + \gamma_{D_s} \times D_s) \quad (18)$$

where T_{di} is the congenital transmission probability for infected dogs. The complete model definition is provided in **Figure 4**, while the meaning and rate function of each transition are summarized in the **Tables 1**, **2**.

Model Simulation

The $\mathcal{GHPN}$ model in **Figure 4** is executed using Snoopy's hybrid simulation engine (Herajy and Heiner, 2012; Heiner et al., 2012) to produce the dynamics of the Chagas disease cycle. An initial simulation of this model using the purely deterministic approach reveals that the values of the model transition firing rates are clearly distinguishable. **Figure 7** compares the cumulative firing rates of the model transitions for a simulation period of 30 years (10,950 days). This comparison shows that certain transitions fire very slowly, while others fire very fast. These different timescales can be interpreted as a result of a small population in the preplaces of the corresponding transitions, or they may be due to the relatively small values of the rate coefficients. For a better view of the quantitative differences among the transition rates, we summarize the cumulative firing rates of the net transitions in **Table 3**.

The simulation statistics in **Table 3** show that growth and death of humans and dogs occur infrequently compared with the death and growth of vectors. For instance, the total firing rates of human growth and death is 0.0021%, compared to 32.88% for

TABLE 1 | Detailed specification of the main transitions of the model in Figure 4.

#	Transition name	Type	Rate	Purpose
1	Death_Hs	Stochastic	$\gamma_{Hs} \times H_s$	Death of susceptible humans
2	Born_Hs	Stochastic	$(1 - T_{hi})\ (\gamma_{H_i} \times H_1 + \gamma_{H_s} \times H_s)$	Growth of susceptible humans
3	Death_Hi	Stochastic	$\gamma_{H_i} \times H_i$	Death of infected humans
4	Born_Hi	Stochastic	$(T_{hi})\ (\gamma_{H_i} \times H_1 + \gamma_{H_S} \times H_S)$	Growth of infected humans
5	Human_infection	Stochastic	$b \times P_{vh} \times H_s \times V_i$	Infection of susceptible humans
6	Death_Ds	Stochastic	$\gamma_{Ds} \times D_s$	Death of susceptible dogs
7	Born_Ds	Stochastic	$(1 - T_{di})(\gamma_{Di} \times D_i + \gamma_{D_S} \times D_s)$	Growth of susceptible dogs
8	Death_Di	Stochastic	$\gamma_{D_i} \times D_i$	Death of infected dogs
9	Born_Di	Stochastic	$(T_{di})\ (\gamma_{Di} \times D_i + \gamma_{D_S} \times D_S)$	Growth of infected dogs
10	Dogs_infection	Stochastic	$b \times d_f \times P_{vdb} + \frac{P_{vdc} \times E \times Ds \times V_i}{V+A}$	Infection of susceptible dogs
11	V_die	Continuous	$\left(\frac{d_m}{2} \times \left(1 - \frac{V}{K}\right) + d_m\right) \times V$	Death of total vectors
12	Born_V	Continuous	$d_h \times (V) \times (1 - \frac{V}{K})$	Growth of total vectors
13	Dogs_consume_V	Continuous	$\left(\frac{E \times V}{V+A}\right) \times D$	Dogs oral consumption of total vectors
14	Vi_die	Continuous	$\left(\frac{d_m}{2} \times \left(1 - \frac{V}{K}\right) + d_m\right) \times V_i$	Death of infected vectors
15	Dogs_consume_Vi	Continuous	$\frac{E \times V}{V+A} \times D \times \frac{V_i}{V}$	Dogs oral consumption of infected vector
16	Infection_Vi	Continuous	$b \times (V - V_i)(P_{hv} \times H_i + P_{dv} \times d_f \times D_i)$	Vector infection

TABLE 2 | The specification of the transitions involved in mortality and the biting rate of the submodel in Figure 6.

#	Transition name	Type	Rate	Purpose
1	Piece1_dm	Continuous	$\frac{(dm\_max - dm\_initial)}{dm\_t_1}$	First time period of the vector mortality rate variation
2	Piece2_dm	Continuous	$\frac{dm\_max}{(dm\_t_2 - dm\_t1_)}$	Second time period of the vector mortality rate variation
3	Piece3_dm	Continuous	$\frac{dm\_initial}{(year\_days - dm\_t_2)}$	Third time period of the vector mortality rate variation
4	Piece1_b	Continuous	$\frac{b\_initial}{b\_t_1}$	First time period of the vector biting rate variation
5	Piece2_b	Continuous	$\frac{b\_max}{(b\_t_3 - b\_t_2)}$	Second time period of the vector biting rate variation
6	Piece3_b	Continuous	$\frac{(b\_max - b\_min)}{(b\_t_4 - b\_t_3)}$	Third time period of the vector biting rate variation
7	Piece4_b	Continuous	$\frac{(b\_initial - b\_min)}{(year\_days - b\_t_4)}$	Fourth time period of the vector biting rate variation
8	Increase_days	Deterministic	1	Increase the current time by one day
9	Reset_days	Immediate	1	Reset the current time to zero after the end of the year

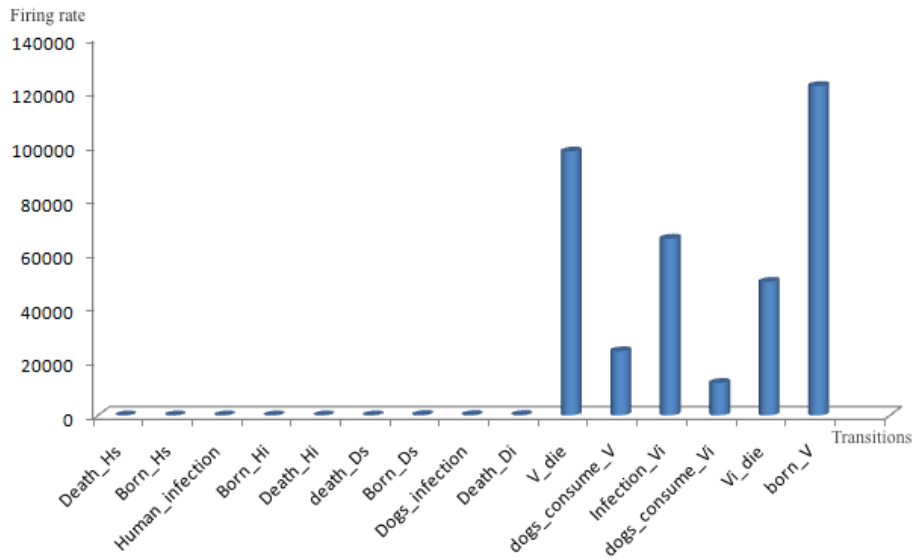


FIGURE 7 | Cumulative firing rates of each transition in the T. Cruzi model during the whole simulation period of 30 years. Transitions representing processes operating on humans and dogs are very slow compared to the other transitions operating on vectors.

TABLE 3 | Comparison of the cumulative transition firing rates (in percentage) of the model in Figure 4.

	Human (%)	Dogs (%)	Vectors (%)	Infected vectors (%)
Growth	0.0021	0.023	32.88	16.33
Death	0.0021	0.023	26.34	13.33
Infection	0.0048	0.0448	17.38	–

The firing rates of processes related to humans or dogs are calculated by taking the average of infected and susceptible species. The simulation time of this experiment is 11,000 days.

the growth rate of vectors. The reason for such a difference is that over a period of 30 years the age of humans and dogs is much larger than the age of bugs. Similarly, the accumulative firing rates of human and dog infections are very low in comparison with the firing rate of vector infections. This is a result of the abundance of the vector population in comparison with the human and dog populations.

Furthermore, the statistics in **Figure 7** and **Table 3** suggest that slow firing processes can be better represented by stochastic transitions, while faster ones should be better modeled via continuous transitions. Therefore, in **Figure 4** all processes related to human and dog populations (e.g., growth, death, and infection) are modeled using stochastic transitions. In contrast, vector-related processes (e.g., vector growth, vector death, and vector infection) are modeled via continuous transitions.

To examine the implication of introducing stochastic transitions to the Chagas disease model, we compare the time course simulation result produced by the purely deterministic approach with the result of the hybrid simulator. **Figures 8**, **9** give the time course simulation results of the population of dogs and infected vectors simulated using both the deterministic and hybrid simulation techniques.

In **Figure 8A**, the population of infected dogs implies the same qualitative conclusions for the deterministic and hybrid results. Both simulation results suggest that the population of infected vectors oscillates with respect to time. However, they differ in the specific quantitative values. Hybrid simulation results imply that the population of infected dogs enter the steady state at a lower value compared with the simulation results produced by the deterministic approach.

To better understand such differences in the quantitative results, we compare the results of the ODE approach with the individual runs of the hybrid simulation. **Figure 8B** presents two single runs of the hybrid simulation. These individual runs show that the population of infected dogs fluctuates as the result of simulating growth, death, and infection processes via stochastic transitions. In fact, modeling such processes in this way is more natural than using the deterministic approach to simulate them. Indeed, growth, death, and infection of dogs and humans are inherently stochastic processes. Moreover, the relatively small population of dogs motivates the use of stochastic transitions to simulate this type of processes.

To examine the influence of such fluctuation on the population of dogs and humans and on the rest of those model components, which remain modeled using the ODE approach, we plot the simulation results of infected vectors for the purely deterministic and the hybrid simulation results. **Figure 9** shows that the population of infected vectors produced through the hybrid simulation technique oscillates at a lower amplitude than the purely deterministic counterpart. This implies that the noise related to the stochastically modeled part also influences the deterministically simulated components.

In summary, although deterministic and hybrid simulation techniques applied to the Chagas disease provide similar qualitative conclusions, the latter technique exhibits more accurate results due to the more realistic representation and simulation of inherently fluctuating natural processes.

Discussion

In this paper we propose the utilization of a special class of hybrid Petri nets, Generalized hybrid Petri nets, for the modeling and simulation of multi-timescale environmental systems. $\mathcal{GHPN}$ provide flexible and rich modeling features to represent and execute the different processes that are frequently encountered during the construction of dynamic models to explore environmental systems. The major advantage of using Petri nets compared with other techniques to represent and simulate environmental models is the graphical depiction of the system components' interactions supporting the communication in a multidisciplinary research team. Hybrid Petri nets extend the

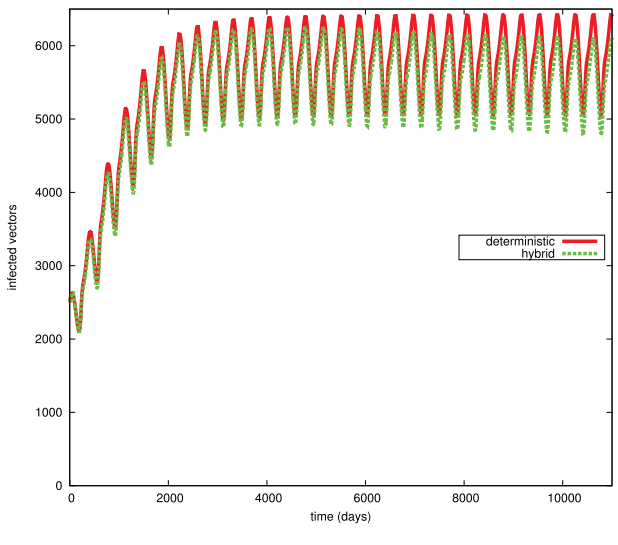


FIGURE 9 | Simulation results for the infected vector population (V_i) in the deterministic and hybrid setting. V_i oscillates in the hybrid setting at slightly lower amplitude than in the purely deterministic setting.

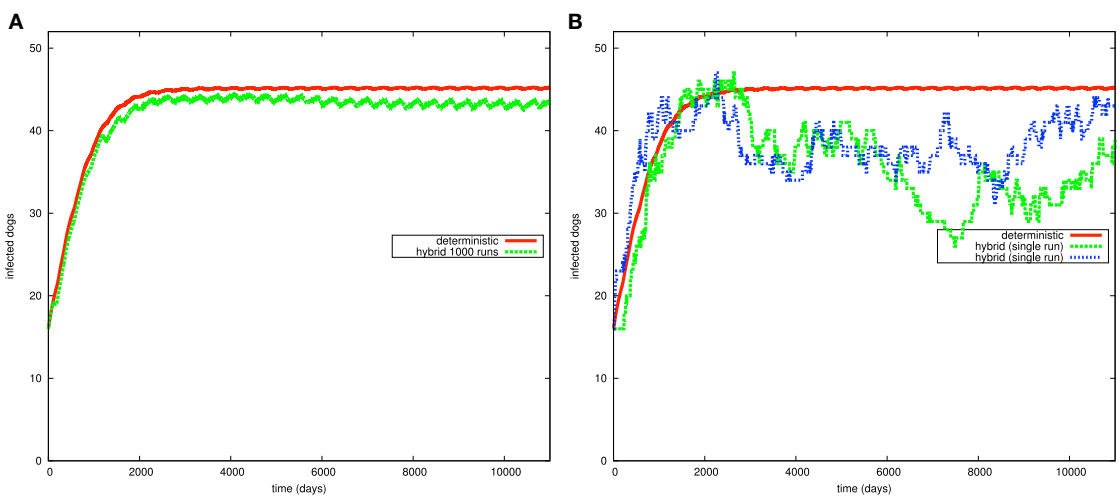


FIGURE 8 | Simulation results of the Chagas model in Figure 4 for the dog population: (A) continuous and average hybrid time course result (1000 runs), (B) deterministic and two single runs of hybrid simulation.

modeling power of standard Petri nets by providing a number of specific elements that can be used to represent physical processes operating at different timescales, which subsequently widens the classes of models that can make use of the Petri net approach and its unifying power.

The case study presented in the Result section explains the motivation behind the different elements of $\mathcal{GHPN}$. For instance, read and inhibitor arcs are used to define boundary conditions for time periods, where vector mortality rates behave in a certain way (increasing or decreasing). Discrete transitions like immediate and deterministically delayed ones can be used to model the duration of time periods. The chosen case study, the Chagas model, involves processes that occur at different scales making the hybrid simulation technique most appropriate to execute such models. In this paper, the processes related to humans and dogs are represented by stochastic transitions. The effects of the other processes could also be investigated by modeling them as stochastic transitions. However, this would increase the simulation runtime for the model. In fact, $\mathcal{GHPN}$ provide a favorable tradeoff between a simulation's accuracy and runtime.

The discussed case study is implemented using the Petri net tool Snoopy (Heiner et al., 2012) which supports the construction and simulation of different Petri net classes including stochastic, continuous, and hybrid Petri nets. Snoopy can be download free of charge for academic use from Snoopy (2015). A $\mathcal{GHPN}$ model constructed with Snoopy can be simulated via a purely deterministic, stochastic, or hybrid simulator. This feature permits to experiment with different simulation techniques using one and the same model. We applied this specific feature to execute the case study in this paper using the deterministic and the hybrid simulator. Besides, a model constructed in Snoopy can be remotely simulated via Snoopy's Simulation and Steering Server (S^4) (Herajy and Heiner, 2014a,b). S^4 provides a further flexible tool to remotely simulate and steer Petri net models constructed using Snoopy. The Snoopy file implementing this model can be downloaded from http://www-dssz.informatik.tu-cottbus.de/DSSZ/Software/Examples. Thus, all our results presented in this paper are reproducible.

In the original model of Coffield et al. (2013), the vector growth rate at time t depends on the hatching rate at a previous time $t - \tau$. The value of τ is approximated to be 20 days (Spagnuolo et al., 2011). This can be simulated as a delayed differential equation with a constant delay. In the discrete world, this delay can be accounted for in the model semantics using a deterministically time delayed transition with a delay of 20 days. However, using continuous transitions to simulate the growth rate of vectors, we need to adjust the semantics of such a transition type to take into account such a delay period while generating and solving the corresponding system of ODEs. This could be added in a future extension of the continuous Petri nets in Snoopy.

Author Contributions

The authors of this paper have equally contributed to the manuscript preparation.

Acknowledgments

This work has been partially funded by the GE-SEED grant (7934) which is administrated by STDF (Science and Technology Development Fund) and DAAD (German Academic Exchange Service).

Supplementary Material

The Supplementary Material for this article can be found online at: http://journal.frontiersin.org/article/10.3389/fenvs.2015.00053

Supplemental Data

Supplemental material (S1) contains the constant coefficients required to specify transition rates of the Chagas disease model.

References

Ajmone, M., Balbo, G., Conte, G., Donatelli, S., and Franceschinis, G. (1995). *Modelling with Generalized Stochastic Petri Nets*. Wiley Series in Parallel Computing, John Wiley and Sons.

Alla, H., and David, R. (1998). Continuous and hybrid Petri nets. *J. Circ. Syst. Comp.* 8, 159–188. doi: 10.1142/S0218126698000079

Castañera, M. B., Aparicio, J. P., and Gürtler, R. E. (2003). A stage-structured stochastic model of the population dynamics of triatoma infestans, the main vector of Chagas disease. *Ecol. Model.* 162, 33–53. doi: 10.1016/S0304-3800(02)00388-5

Coffield, D. J. Jr., Spagnuolo, A. M., Shillor, M., Mema, E., Pell, B., Pruzinsky, A., et al. (2013). A model for Chagas disease with oral and congenital transmission. *PLoS ONE* 8:e67267. doi: 10.1371/journal.pone.0067267

Cohen, J. E., and Gürtler, R. E. (2001). Modeling household transmission of american trypanosomiasis. *Science* 293, 694–698. doi: 10.1126/science.1060638

David, R., and Alla, H. (2010). *Discrete, Continuous, and Hybrid Petri Nets*. Berlin; Heidelberg: Springer.

Fujita, S., Matsui, M., Matsuno, H., and Miyano, S. (2004). Modeling and simulation of fission yeast cell cycle on hybrid functional Petri net. *IEICE Trans. Fundam. Electron. Commun. Comput. Sci.* E87-A, 2919–2927.

Gilbert, D., and Heiner, M. (2006). "From Petri nets to differential equations - an integrative approach for biochemical network analysis," in *Proc. ICATPN 2006, Vol. 4024, of LNCS*, eds S. Donatelli and P. S. Thiagarajan (Berlin; Heidelberg: Springer), 181–200.

Gillet, F. (2008). Modelling vegetation dynamics in heterogeneous pasture-woodland landscapes. *Ecol. Model.* 217, 1–18. doi: 10.1016/j.ecolmodel.2008.05.013

Gregorio, S. D., Serra, R., and Villani, M. (1999). Applying cellular automata to complex environmental problems: the simulation of the bioremediation of contaminated soils. *Theor. Comput. Sci.* 217, 131–156. doi: 10.1016/S0304-3975(98)00154-6

Heiner, M., Herajy, M., Liu, F., Rohr, C., and Schwarick, M. (2012). "Snoopy – a unifying Petri net tool," in *Proc. PETRI NETS 2012, Vol. 7347, of LNCS*, eds S. Haddad and L. Pomello (Berlin; Heidelberg: Springer), 398–407.

Herajy, M., and Heiner, M. (2012). Hybrid representation and simulation of stiff biochemical networks. *J. Nonlin. Anal.* 6, 942–959. doi: 10.1016/j.nahs.2012.05.004

Herajy, M., and Heiner, M. (2014a). Petri net-based collaborative simulation and steering of biochemical reaction networks. *Fundam. Informa.* 129, 49–67. doi: 10.3233/FI-2014-960

Herajy, M., and Heiner, M. (2014b). "A steering server for collaborative simulation of quantitative Petri nets," in *Application and Theory of Petri Nets and Concurrency, Vol. 8489, of Lecture Notes in Computer Science*, eds G. Ciardo and E. Kindler (Switzerland: Springer International Publishing), 374–384.

Herajy, M., Schwarick, M., and Heiner, M. (2013). Hybrid Petri Nets for modelling the eukaryotic cell cycle. *ToPNoC* VIII, 123–141. doi: 10.1007/978-3-642-40465-8/7

Khoury, D. S., Barron, A. B., and Myerscough, M. R. (2013). Modelling food and population dynamics in honey bee colonies. *PLoS ONE* 8:e59084. doi: 10.1371/journal.pone.0059084

Liu, H., Benoit, G., Liu, T., Liu, Y., and Guo, H. (2015). An integrated system dynamics model developed for managing lake water quality at the watershed scale. *J. Environ. Manage.* 155, 11–23. doi: 10.1016/j.jenvman.2015.02.046

Mao, G., and Petzold, L. (2002). Efficient integration over discontinuities for differential-algebraic systems. *Comput. Math. Appl.* 43, 65–79. doi: 10.1016/S0898-1221(01)00272-3

Matsuno, H., Tanaka, Y., Aoshima, H., Doi, A., Matsui, M., and Miyano, S. (2003). Biopathways representation and simulation on hybrid functional Petri net. *In Silico Biol.* 3, 389–404.

Murata, T. (1989). Petri nets: properties, analysis and applications. *Proc. IEEE* 77, 541–580. doi: 10.1109/5.24143

Nouvellet, P., Cucunubá, Z. M., and Gourbière, S. (2015). "Chapter four - ecology, evolution and control of Chagas disease: a century of neglected modelling and a promising future," in *Mathematical Models for Neglected Tropical Diseases: Essential Tools for Control and Elimination, Part A, Vol. 87, of Advances in Parasitology*, eds R. M. Anderson and M. G. Basez (Academic Press), 135–191.

Qi, C., and Chang, N.-B. (2011). System dynamics modeling for municipal water demand estimation in an urban region under uncertain economic impacts. *J. Environ. Manage.* 92, 1628–1641. doi: 10.1016/j.jenvman.2011.01.020

Reddy, V., Mavrovouniotis, M., and Liebman, M. (1993). "Petri net representations in metabolic pathways," in *Proceedings of the 1st International Conference on Intelligent Systems for Molecular Biology*, eds L. Hunter, D. Searls, and J. Shavlik (Bethesda, MD: The AAAI Press), 328–336.

Russell, S., Barron, A. B., and Harris, D. (2013). Dynamic modelling of honey bee (*Apis mellifera*) colony growth and failure. *Ecol. Model.* 265, 158–169. doi: 10.1016/j.ecolmodel.2013.06.005

Schmickl, T., and Crailsheim, K. (2007). Hopomo: a model of honeybee intracolonial population dynamics and resource management. *Ecol. Model.* 204, 219–245. doi: 10.1016/j.ecolmodel.2007.01.001

Seppelt, R., Müller, F., Schröder, B., and Volk, M. (2009). Challenges of simulating complex environmental systems at the landscape scale: a controversial dialogue between two cups of espresso. *Ecol. Model.* 220, 3481–3489. doi: 10.1016/j.ecolmodel.2009.09.009

Snoopy (2015). *Snoopy Website*. Available online at: http://www-dssz.informatik.tu-cottbus.de/snoopy.html[Accessed: 28/3/2015].

Soliman, S., and Heiner, M. (2010). A unique transformation from ordinary differential equations to reaction networks. *PLoS ONE* 5:e14284. doi: 10.1371/journal.pone.0014284

Spagnuolo, A., Shillor, M., and Stryker, G. (2011). A model for Chagas disease with controlled spraying. *J. Biol. Dyn.* 5, 299–317. doi: 10.1080/17513758.2010.505985

Tian, Z., Faure, A., Mori, H., and Matsuno, H. (2013). Identification of key regulators in glycogen utilization in *E. coli* based on the simulations from a hybrid functional Petri net model. *BMC Syst. Biol.* 7:S1. doi: 10.1186/1752-0509-7-S6-S1

Uusitalo, L., Lehikoinen, A., Helle, I., and Myrberg, K. (2015). An overview of methods to evaluate uncertainty of deterministic models in decision support. *Environ. Model. Soft.* 63, 24–31. doi: 10.1016/j.envsoft.2014.09.017

Valk, R. (1978). "Self-modifying nets, a natural extension of Petri nets," in *Proc. of the Fifth Colloquium on Automata, Languages and Programming* (London, UK: Springer-Verlag), 464–476.

Xiang, J., Chen, H., and Ishikawa, H. (2013). A mathematical model for the transmission of schistosoma japonicum in consideration of seasonal water level fluctuations of Poyang Lake in Jiangxi, China . *Parasitol. Int.* 62, 118–126. doi: 10.1016/j.parint.2012.10.004

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Advantages of publishing in Frontiers

OPEN ACCESS
Articles are free to read, for greatest visibility

COLLABORATIVE PEER-REVIEW
Designed to be rigorous – yet also collaborative, fair and constructive

FAST PUBLICATION
Average 85 days from submission to publication (across all journals)

COPYRIGHT TO AUTHORS
No limit to article distribution and re-use

TRANSPARENT
Editors and reviewers acknowledged by name on published articles

SUPPORT
By our Swiss-based editorial team

IMPACT METRICS
Advanced metrics track your article's impact

GLOBAL SPREAD
5'100'000+ monthly article views and downloads

LOOP RESEARCH NETWORK
Our network increases readership for your article

Frontiers
EPFL Innovation Park, Building I • 1015 Lausanne • Switzerland
Tel +41 21 510 17 00 • Fax +41 21 510 17 01 • info@frontiersin.org
www.frontiersin.org

Find us on