



OPEN ACCESS

EDITED BY

Lars Hendrik Wegner,
Foshan University, China

REVIEWED BY

Renata Retkute,
University of Cambridge, United Kingdom

*CORRESPONDENCE

Michiel Stock
✉ michiel.stock@ugent.be

RECEIVED 22 September 2023

ACCEPTED 07 December 2023

PUBLISHED 16 January 2024

CITATION

Stock M, Pieters O, De Swaef T and wyffels F (2024) Plant science in the age of simulation intelligence. *Front. Plant Sci.* 14:1299208. doi: 10.3389/fpls.2023.1299208

COPYRIGHT

© 2024 Stock, Pieters, De Swaef and wyffels. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Plant science in the age of simulation intelligence

Michiel Stock^{1*}, Olivier Pieters^{2,3}, Tom De Swaef³ and Francis wyffels²

¹KERMIT and Biobix, Department of Data Analysis and Mathematical Modelling, Ghent University, Ghent, Belgium, ²IDLAB-AIRO, Ghent University, imec, Ghent, Belgium, ³Plant Sciences Unit, Flanders Research Institute for Agriculture, Fisheries and Food, Melle, Belgium

Historically, plant and crop sciences have been quantitative fields that intensively use measurements and modeling. Traditionally, researchers choose between two dominant modeling approaches: mechanistic plant growth models or data-driven, statistical methodologies. At the intersection of both paradigms, a novel approach referred to as “*simulation intelligence*”, has emerged as a powerful tool for comprehending and controlling complex systems, including plants and crops. This work explores the transformative potential for the plant science community of the nine simulation intelligence motifs, from understanding molecular plant processes to optimizing greenhouse control. Many of these concepts, such as surrogate models and agent-based modeling, have gained prominence in plant and crop sciences. In contrast, some motifs, such as open-ended optimization or program synthesis, still need to be explored further. The motifs of simulation intelligence can potentially revolutionize breeding and precision farming towards more sustainable food production.

KEYWORDS

simulation intelligence, digital agriculture, phenotyping, quantified plant, digital twin, modeling, scientific computing, artificial intelligence

1 Introduction

As computational capabilities have grown, modeling has become a specialized area of agricultural sciences. These models can accurately simulate the performance of plants, crops, and greenhouses under various conditions and have therefore been used extensively for generating scientific hypotheses, informing and accelerating breeding programs, optimizing crop management and providing policy recommendations (Silva and Giller, 2020). However, models remain highly species-specific, complex, and difficult to calibrate due to many interlinked parameters (Wallach et al., 2021). Current developments and increased availability of phenotyping data provide an extensive source of data for model development and calibration, required for the extension of model applicability to novel or ‘forgotten’ crops and to studying impacts of pests and micronutrients. (Silva and Giller, 2020). However, more modern and powerful

modeling paradigms are needed to address these issues and to infer large sets of parameters from phenotyping data. This paper introduces some fundamental concepts of the emerging field of *simulation intelligence* (SI) to plant science. SI is the merger of scientific computing and artificial intelligence (Lavin et al., 2021). Specifically for plant science, this will result in a new field that combines novel phenotyping approaches with modeling.

Phenotyping is quantifying (a subset of) plant traits that result from the interaction between plant genetics and the environmental conditions to which plants are exposed (Walter et al., 2015). Due to internal regulatory mechanisms, these phenotypic responses are situated at multiple organizational scales (cell, tissue, organ, plant, field) and across timescales. Biotic and abiotic external drivers also influence these mechanisms. While phenotyping ideally involves capturing the entire state of a plant in space and time, only partial observations are practically feasible, leading to the need for a wide range of sensory devices that capture part of the phenotype. A more holistic and dynamic view of phenotyping is necessary to overcome challenges involved in current approaches, including improving the temporal resolution and broadening results from specific studies to more diverse conditions (Das Choudhury et al., 2018).

Most models consist of mathematical equations for predicting plant behavior, morphology and growth as a function of environment, genetics, and management. Plant models describe and connect plant processes typically studied in isolation and consequently predict integrated responses. As such, models are often used for hypothesis development and improved understanding of plant processes, but also as decision support tools for breeding (e.g., genomic prediction), crop management (e.g., irrigation scheduling) and policy-making (e.g., climate change scenarios) (Peng et al., 2020).

Depending on their objective, models vary at the level with which processes are included (black-box ↔ mechanistic axis) and at the scale they operate in terms of space (molecule ↔ ecosystem axis) and time (second ↔ century axis). Models are considered ‘process-based’ or ‘mechanistic’ when parameters have a biophysical interpretation and their equations explicitly describe processes (e.g., photosynthesis, water transport). They operate at a different spatial and (often also) temporal scale. At each spatial level, there is an extra level of abstraction, but, interestingly, the scale does not necessarily determine whether a model is more or less mechanistic, as plants tend to adapt to their environment in an integrated way (Tardieu et al., 2020). On the other end of the spectrum, entirely data-driven models based on machine learning algorithms often lack interpretable parameters. The latter group of models is vital in breeding [e.g., genomic prediction (Korte and Farlow, 2013; Hickey et al., 2017; De Meyer et al., 2023)], but also in greenhouse climate control (Hemming et al., 2020). Consequently, models often only operate on a single point in the tempo-spatial domain, limiting their use beyond their initial conceptualization. Nevertheless, there are efforts to connect modeling scales from the molecular level up to the crop system (Peng et al., 2020).

Recently, surrogate plant models have become popular (Corrales et al., 2022; Cheng et al., 2023; Zhang et al., 2023), because these allow for creating a “digital twin” of the plant systems for decision support and control. A surrogate data-driven model is trained to mimic the mechanistic model’s output

accurately. When properly trained, such a surrogate can be several magnitudes faster than the original model while behaving nearly identically (Gherman et al., 2023). Apart from the complete replacement of mechanistic models by data-driven models, these can also be combined. For example, Zhang et al. (2023) demonstrate how these can be coupled in series (e.g., crop model simulation outputs as input to a machine learning model), in parallel (e.g., data assimilation in crop models) or via modules (e.g., part of a crop model is replaced by a machine learning module). This is a stepping stone towards SI, leading to cross-pollination between scientific computing, artificial intelligence, plant modeling and phenotyping.

2 Combining scientific computing with artificial intelligence

Complex biological systems require powerful tools to study them. On the one hand, many of these systems require substantial domain knowledge, often as conservation laws and reaction mechanisms, for which traditional mechanistic modeling and simulation paradigms are well suited. This is known as “scientific computing” and relies mainly on ordinary differential equations, partial differential equations, agent-based models and their ilk. On the other hand, many mechanisms are yet to be elucidated while, at the same time, a plethora of multimodal data is available. This motivates using a data-driven approach (referred to as “artificial intelligence” or “machine learning”). Recent advances blur the lines between traditional methodologies, and so-called *scientific machine learning* combines both, for example, in neural ordinary differential equations (Chen et al., 2018; Innes et al., 2019; Rackauckas et al., 2021), where the solvers are treated as differentiable programs that can fit data to learn unknown dynamics of the problems.

The advances in scientific computing and machine learning and their use in studying complex, dynamical multi-scale systems gave rise to a more generalized view: the new field of *simulation intelligence* (SI). Lavin et al. (2021) outlined nine vital, interconnected computing technology motifs, visually represented in Figure 1:

1. **Multi-scale and multi-physics modelling** (Karniadakis et al., 2021): integrating different types of simulators;
2. **Surrogate modelling and emulation** (Purcell and Neubauer, 2023): replacing a complex model or system with a different one;
3. **Simulation-based inference** (Cranmer et al., 2020): using the simulator to infer parameters or states;
4. **Causal modelling and inference** (Schölkopf et al., 2021): including or identifying causal concepts within the model;
5. **Agent-based modelling** (Zhang and DeAngelis, 2020): simulating a system as a collection of semiautonomous agents;
6. **Probabilistic programming** (Schoot et al., 2021): interpreting code as a stochastic program;
7. **Differentiable programming** (Baydin et al., 2018): computing and using derivatives and gradients of computer code and simulators;

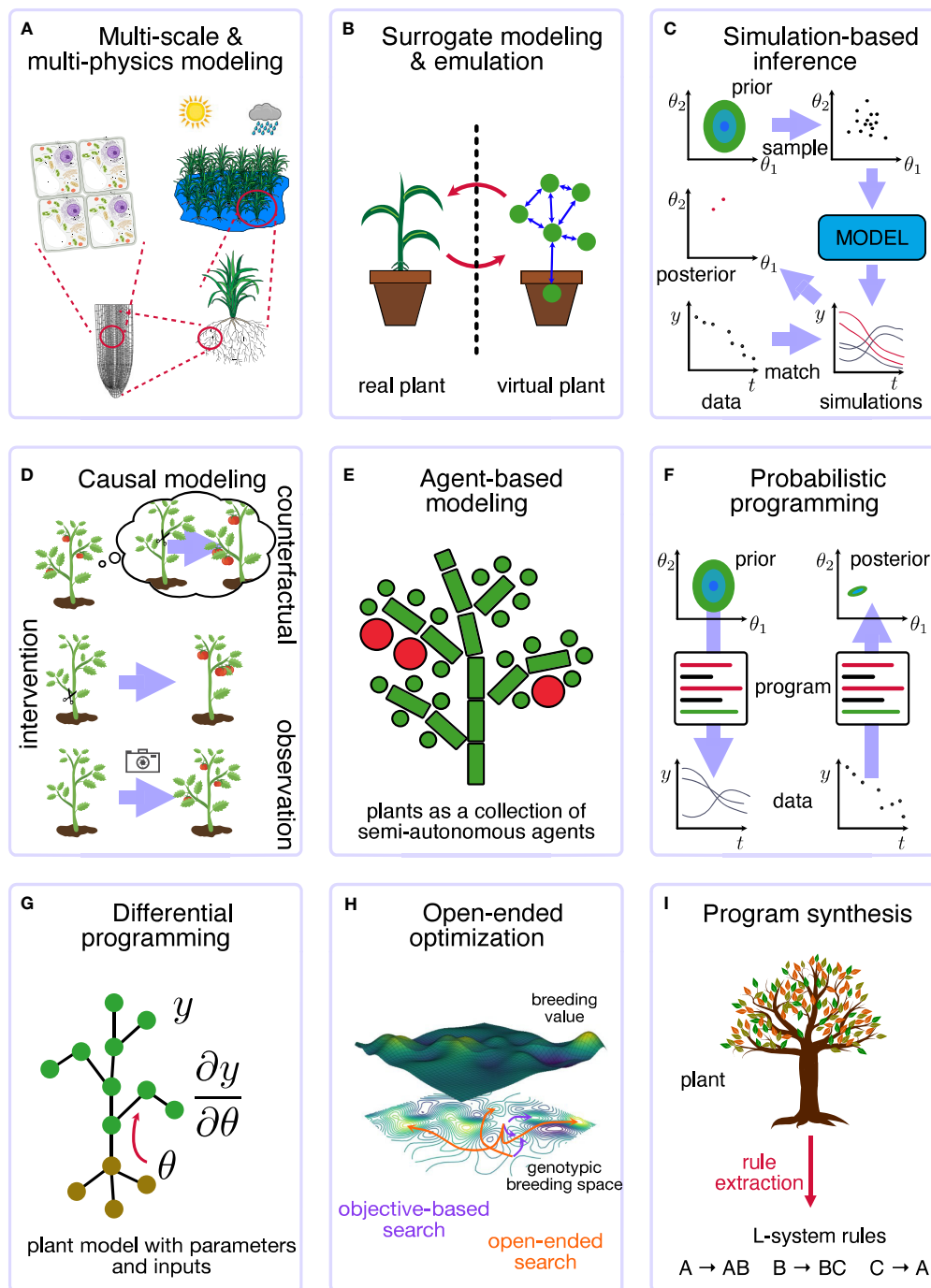


FIGURE 1

Overview of the different SI motifs for plants. (A) **Multi-scale and multi-physics modeling** considers the different scales (from cell to ecosystem or field) and physical processes (radiation, hydraulics, fluidics). (B) **Surrogate modeling and emulation** considers a virtual digital twin of the plant system. (C) **Simulation-based inference**, such as approximate Bayesian computing, allows a simulator to infer parameters or states from data. (D) **Causal modeling and inference** takes into account the different levels of causal reasoning that are possible. (E) **Agent-based modeling** simulates a system as a collection of semiautonomous agents. (F) **Probabilistic programming** allows for general computing with stochastic components and performs general inferences about parameters and states. (G) **Differential programming** computing and simulation with gradients and derivatives. (H) **Open-ended optimization** aims at finding continuous improvements and adaptations, for example, in plant breeding. (I) **Program synthesis** automatically generates the code to solve a problem, e.g., extract the L-system to describe a plant.

8. **Open-ended optimization** (Stock and Gorochowski, 2023): trying to find continuous improvements;
9. **Program synthesis** (David and Kroening, 2017): automatically discovering the code to solve a problem.

We included a reference for each motif that covers this specific topic more in-depth.

This technology stack has enormous potential to advance fields such as material science, agriculture, chemistry, medicine, climate,

and synthetic biology. Simulation intelligence can also significantly impact plant sciences. By combining modeling and phenotyping, one can uncover mechanisms underlying plant dynamics. For example, functional-structural plant modeling aims to develop holistic plant growth and function models, harmonizing architecture with (eco)physiology. The Quantitative Plant¹ initiative collects plant phenotyping data sets, plant analysis tools and models. The availability of plant phenotypic data repositories and plant biophysical models are prerequisites for SI.¹

Concretely, SI aims to handle the following challenges in using computational modeling for real-world problems:

- **Inverse problem solving**, where one wants to use a model to infer hidden states or parameters from observations or measurements. For example, in root phenotyping, researchers use electrical resistance tomography measurements to infer root properties non-invasively (Whalley et al., 2017).
- **Uncertainty reasoning**, which relates to the inherent uncertainty of dealing with biological systems, both epistemic (i.e., incomplete knowledge of the processes) and aleatoric (i.e., the irreducible noise, for example, due to biological stochasticity) (Hüllermeier and Waegeman, 2021). Quantifying uncertainty is of great importance for plant breeding and precision crop management (Asseng et al., 2013; Tao et al., 2018; Folberth et al., 2019; Nelson et al., 2019; Hernández and López, 2020; Dokoohaki et al., 2021), especially when dealing with a changing climate.
- **Human-machine teaming** relates to the interaction between the model or machine intelligence and the breeder, farmer or other users. This includes intelligent dashboards and ways for users to query the simulator for decision-making and inject data, observations and results into the model. Bridging the gap between models and users is a significant challenge in digital agriculture (Antle et al., 2017; Slob et al., 2023; Zhang et al., 2023).

3 Nine simulation intelligence motifs for plant science

This section discusses the nine SI motifs outlined in Lavin et al. (2021) and discusses how they can relate to plant science. We speculate about the opportunities they could present in digital agriculture when we find only a few plant-related examples. Due to the broad scope of these topics, we have to be concise. Our primary aim is to inform the quantitative plant scientist of available SI tools. We refer to the work of Lavin et al. (2021) or specific overview papers for an extensive introduction.

3.1 Multi-scale and multi-physics modeling

Plant processes occur on different scales (Figure 1A). These can be spatial, from the molecular processes in the cell (micrometers) to the ecosystem (up to kilometers), or temporal, the processes of interest range from nanoseconds (e.g., photosynthesis) to months or years (e.g., growth). For example, plants generate and use various rhythms and oscillations at all scales and organization levels (Damineli et al., 2022). Plant modelers are aware of the fractal complexity of plant modeling, where lower-level processes can be abstracted away in so-called *meta-mechanisms* (Tardieu et al., 2020). Meta-mechanisms are, e.g., response curves of plant traits to environmental conditions, which can be characterized in a high-throughput fashion. Such meta-mechanisms are largely determined by physical trade-offs that limit evolution (Kempes et al., 2019). Meta-mechanisms can be tailored to specific plant species or cultivars, an open challenge in plant modeling (Silva and Giller, 2020).

Plant models involve various kinds of physical models, going from molecular and metabolic processes (Farquhar et al., 1980), hydraulic functioning (De Swaef et al., 2022), to soil and atmospheric physics (Liu et al., 2020b). Modern (functional-structural) plant modeling involves advanced physics simulation such as ray tracing to assess radiation (De Visser et al., 2014; Bailey, 2018; Retkute et al., 2018) and computational fluid dynamics (Bartzanas et al., 2013; Jiao et al., 2020). The latter are often computationally demanding and might require appropriate tools, such as surrogate modeling (discussed later), to make them feasible for, e.g., greenhouse control.

Physics-informed machine learning can be a powerful aid in incorporating the different scales and physics (Karniadakis et al., 2021). Here, data-driven models are fitted not only to match their training data but also to adhere to known physical laws and are ideally suited to integrate data into different physical processes. This has shown great success in biomedical modeling, for example, in modeling blood flow in an intracranial aneurysm (Raissi et al., 2020). In crop science, Cavanagh et al. (2021) used physics-informed deep learning to study the morphological changes induced by Asian soybean rust. Similar directions for holistic plant modeling will undoubtedly be fruitful.

3.2 Surrogate modeling and emulation

A *surrogate model* is a model that replaces an often expensive computation or process (Figure 1B). In scientific computing, expensive simulations, such as computational fluid dynamics, are often replaced by relatively inexpensive methods of training and deploying machine learning surrogates, such as Gaussian processes or artificial neural networks. For example, Cheng et al. (2023) used a data-driven surrogate model in combination with a multi-objective genetic algorithm to reduce irrigation and nitrogen fertilization by 44% and 37%, respectively.

Surrogate models play a pivotal role in developing of *digital twins* – real-time synchronized virtual representations of products, processes, or environments. These dynamic digital counterparts

1 <https://quantitative-plant.org/>(accessed on March 2023).

facilitate a bidirectional flow of information, leveraging real-world data while influencing management and decision-making processes. Positioned at the forefront of digital agriculture and smart farming (Verdouw et al., 2021; Slob et al., 2022; Purcell and Neubauer, 2023), digital twins seamlessly integrate with the principles of Industry 4.0 tailored for agricultural contexts.

Digital twins exhibit versatility, being applied to emulate plants, greenhouses, or entire supply chains. Their primary utility is improving cost-efficiency, such as reducing water and fertilizer consumption and elevating prediction accuracy (Ariessen-Verschuur et al., 2022). Verdouw et al. (2021) categorizes digital twins based on their relationship to virtual objects—whether it pertains to an imaginary entity (e.g., a yet-to-grow cultivar), an existing object, future states for predictive analysis, or a historical object. Additionally, digital twins serve distinct purposes, being employed for both monitoring and prescription.

In recent years, digital twins have demonstrated significant successes in agriculture. Examples include emulating various wheat development stages to predict yield (Skobelev et al., 2020), optimizing yield and minimizing energy requirements in underground hydroponic farms (Jans-Singh et al., 2020), exploring virtual replicas of greenhouses through immersive VR experiences (Slob et al., 2023), monitoring the health and quality of individual plants in orchards (Moghadam et al., 2020), and fine-tuning greenhouse control systems (Chaux et al., 2021). Using simulations to govern systems and explore hypothetical interventions aligns closely with causal reasoning, as discussed in Section 3.4.

3.3 Simulation-based inference

In plant modeling, knowledge of the processes of interest is often encoded in process-based models. Typically, given the initial conditions and the parameter values, these models can simulate data that can be compared with quantitative measurements such as biomass growth, development stage, or transpiration. Generating this data is called forward modeling. However, when one observes the data, one would often infer the likely hidden states or parameter values, a process that is a much more challenging inverse problem (Figure 1C). The field of simulation-based inference deals with developing inference methods for highly intricate simulators, i.e., to extract the parameters of a mechanistic model algorithmically (Cranmer et al., 2020). Simulation-based inference is often called *likelihood-free inference* – as contrasted with classical statistical estimation problems. The likelihood function implicitly defined by the simulator is often not tractable, making this a challenging endeavor. Inverse problems are usually solved using a Bayesian perspective, where the parameters or states have associated prior distributions. The simulator acts as an implicit likelihood function, linking the model with the data and parameters.

Approximate Bayesian Computation (ABC) is a widely utilized approach for simulation-based inference (Marin et al., 2012; Romero-Cuellar and Francés, 2023). In ABC, the simulator generates synthetic data by sampling parameters from a prior distribution or proposal distribution and using these parameters to perform a simulation.

These synthetic datasets are characterized by summary statistics, such as the total biomass, used to compare the simulated data with collected observations. Parameter values producing synthetic data with summary statistics closely aligned with those of the actual data, often measured using Euclidean distance, are retained. These selected values provide approximate samples of the posterior distribution. ABC's most commonly used variant operates similarly to rejection sampling, and its sampling properties are well understood. However, the conventional ABC method becomes inefficient, especially when dealing with large parameter spaces. Notably, ABC has been successfully applied in plant science to merge crop growth models with whole genome data (Technow et al., 2015), to infer root architecture (Ziegler et al., 2019) and to characterize the morphodynamic progression of Asian soybean rust (Cavanagh et al., 2021). The progression of machine learning and SI techniques, including probabilistic and differentiable programming (see Section 3.5), has significantly influenced simulation-based inference. For instance, in a study by Monti et al. (2023), the parameters of an agent-based model were learned directly from data by redefining it as a probabilistic program.

3.4 Causal modeling and inference

Data-driven modeling has achieved remarkable success across various scientific and technological domains. However, purely statistical models often fail to uncover the underlying causal mechanisms behind the observed data. As an illustration, consider a simple linear regression model predicting yield based on nutrient inputs. This model might erroneously suggest that fertilization decreases yield, neglecting the confounding effect of poorer soils where fertilizers are commonly applied.

The significance of understanding causality has been underscored by Judea Pearl in his work, including “The Book of Why” (Pearl and Mackenzie, 2018). Pearl introduced a hierarchy of causal reasoning that data-driven models can accomplish, comprising:

1. **Observations:** Detecting associations in data, such as estimating tree biomass from their diameter at breast height.
2. **Interventions:** Predicting outcomes resulting from active manipulations of the system, like projecting the effects of flower pruning on fruit production.
3. **Counterfactuals:** Imagining potential scenarios where conditions or interventions differed, as in assessing whether larger fruits would have resulted from more extensive flower removal.

Pearl's mathematical insights reveal that some models are inherently limited in performing higher-level causal reasoning. Thus, plant scientists who aim to predict and manage must exercise caution when employing data-driven models from observational data because the causal link between the predictors is not exploited by default. For example, a data-driven model might conclude that watering harms a plant's water status, as irrigation and water stress are correlated. This scenario highlights the

limitations of relying solely on observational data, representing the first level of causal reasoning in Judea Pearl's hierarchy, where associations in data are detected without considering active manipulations or counterfactual scenarios. The limitations contrast with many mechanistic models, which can often be used directly for interventions and counterfactuals. The crux lies in developing models incorporating the structural relationships between variables of interest, advocating for mechanistic and hybrid models. The evolving field of causal machine learning continues to gain prominence (Schölkopf et al., 2021) (Figure 1D).

3.5 Agent-based modelling

Agent-based models (ABMs) depict complex systems as interconnected, autonomous agents (e.g., organs or whole plants) interacting from the bottom up (Figure 1E). These models frequently encompass stochastic elements and can replicate macro-level processes stemming from micro-level interactions. Consequently, ABMs are a natural fit for elucidating multi-scale phenomena. In ecology and plant science, ABMs are widely employed (DeAngelis and Mooij, 2005; McLane et al., 2011; Zhang and DeAngelis, 2020), offering insights into growth, carbon allocation, reproduction, and more. These models can portray individual plants within functional-structural plant models (FSPMs) or capture entire plant communities, such as field ecosystems.

Interestingly, ABMs precisely capture individual plant behaviors due to plants' modular structure, comprising elements like roots, leaves, stems, and branches. Each module functions autonomously, gathering, producing, or distributing resources for the overall plant's advantage. Remarkably, plants lack a central controlling entity, resembling a decentralized "swarm intelligence" (Baluška et al., 2010a; Oborny, 2019; van Schijndel et al., 2022). For instance, a plant's root tips exhibit both sensory and command center roles, independently deciding growth directions and even forming symbiotic relationships with mycorrhizal fungi (Baluška et al., 2010b). Some liken this to a "solid" brain, where individual units are fixed. In contrast, others argue for "liquid" brain aspects (van Schijndel et al., 2022), like in plants with vegetative propagation, like strawberries, exploring diverse niches to optimize their niche.

3.6 Probabilistic programming

The language of probability theory is an effective way to describe biological systems, given their inherent stochastic nature (Figure 1F). Specifically, Bayesian statistics is a consistent framework to update the scientist's prior beliefs (encoded in prior distributions) with measurements and observations (encoded in the likelihood) into the so-called posterior distribution (Schoot et al., 2021). In plant science, Bayesian reasoning is applied in, for instance, plant pathology and epidemics (Mila and Carriquiry, 2004), modeling life stage events (Humplík et al., 2020), and predicting maize yield (Lacasa et al., 2020). Though powerful,

Bayesian and probabilistic methods can be complex in practice because conditioning a distribution (e.g., computing the posterior) requires normalization, often involving computing intractable integrals or sums. Probabilistic programming is a relatively new, general approach to making probabilistic methods more accessible in the scientific community.

A *probabilistic programming language* (PPL) allows one to write, in principle, arbitrary complex stochastic programs from which the scientist can make inferences by sampling. Hence, a universal PPL provides two constructs: i) a way to sample from the stochastic program and ii) a way to condition during inference. For example, one can write a program to simulate flowering vines and then constrain regions where they are present. This allows one to sample vines that grow in a specific shape, such as a letter (Ritchie et al., 2016). PPLs have shown success throughout the biological sciences, for example, in inferring phylogeny (Ronquist et al., 2020), protein structure alignment (Moreta et al., 2019) and inferring signaling pathways (Merrell and Gitter, 2020). There are a plethora of PPLs available, many interfacing with scientifically popular programming languages for sciences, for example, Stan (Stan Development Team, 2023), Pyro (Bingham et al., 2019), or Turing (Holt and Cordy, 1988).

3.7 Differentiable programming

While probabilistic programming facilitates generic computations involving probability distributions, *differentiable programming* (Izzo et al., 2016; Innes et al., 2019) extends computation by enabling differentiation of arbitrary computer programs (Figure 1G). This empowers the fine-tuning of program behavior using gradient-based optimization techniques. This achievement relies on *automatic differentiation* (Baydin et al., 2018) – numerically computing (exact) derivatives by directly manipulating the computational graph – a foundational concept in deep learning. Differentiable programming has exerted a profound scientific influence, acting as a cornerstone for nearly all deep learning research over the past decade and diverse domains beyond deep learning. These domains encompass ordinary differential equations (Chen et al., 2018; Rackauckas et al., 2021; Núñez et al., 2023), scientific machine learning, robotics Degraeve et al. (2019), physics, protein science (Ingraham et al., 2019; AlQuraishi and Sorger, 2021), combinatorial optimization (Liu et al., 2020a), and geosciences (Shen et al., 2023). The utility of differentiable programming extends to harmonizing process-based and data-driven models. Within plant sciences, differentiable plant models offer an avenue to assess sensitivity directly, calibrate parameters using gradients, apply probabilistic programming techniques for uncertainty quantification, and gain control over conditions for optimizing growth. Concrete achievements in plant sciences include the creation of 3D digital twin leaf models from image data (Li et al., 2022) and solving inverse problems related to photosynthesis (Aboelyazeed et al., 2023). We also propose that advancements in differentiable ray tracing (Li et al., 2018), computational fluid dynamics (Bezgin et al., 2023), and physics engines (de Avila Belbute-Peres et al., 2018; Degraeve et al., 2019) hold substantial promise for enhancing plant simulations.

3.8 Open-ended optimization

Open-ended systems possess the ability to achieve limitless improvement and continuously generate novelty (Stanley and Lehman, 2015; Banzhaf et al., 2016; Stanley et al., 2017) (Figure 1H). In such systems, the focus primarily lies on creating novelty rather than being driven by a specific objective function (Stanley and Lehman, 2015). Open-endedness is a characteristic observed in various complex systems, including natural evolution and technological innovation. Its principles have been explored for diverse applications such as designing new computer architectures (Ackley and Small, 2014), software development (Fix et al., 2021), artificial neural networks (Guttenberg et al., 2018), and novel cancer treatment strategies (Balaz et al., 2021).

Our other work delves into how open-endedness and quality-diversity algorithms can contribute to biotechnology and synthetic biology (Stock and Goroehowski, 2023). Expanding this perspective, we propose that open-endedness can significantly impact plant breeding, a critical aspect in ensuring global food security (Lenaerts et al., 2019). Conventional breeding approaches often prioritize incorporating positive traits into populations, potentially at the expense of diversity (Louwaars, 2018). A noteworthy exception occurred in the 1970s when Zelder, a breeding company, intentionally bred wheat varieties to enhance diversity as a defense against yellow rust (Groenewegen, 1977). Embracing the open-ended optimization viewpoint, one could design breeding schemes capable of continually generating new cultivars with novel and desirable traits. Insights from the field of quality-diversity optimization (Pugh et al., 2016), which focuses on generating new variants that combine functionality and diversity, have the potential to revolutionize breeding strategies for developing crops and cultivars suited to a dynamically changing world. In silico evolution experiments can help to understand the allometric relations observed in plants due to environmental conditions (Eloy et al., 2017). As such, they might help to design new cultivars.

3.9 Program synthesis

Program synthesis automates software creation to tackle specific problems (David and Kroening, 2017) (Figure 1I). Here, the focus shifts towards generating optimized code by capturing users' intentions. A notable application of this concept is evident in the recently released ChatGPT, where users use natural language queries to program or create computer code. This synthesis technique plays a pivotal role in simplifying intricate mathematical system descriptions.

Program synthesis offers avenues for extracting insights from biological experiments in mathematical modeling. For instance, Koks et al. (2013) automatically generated biological models from mutation experiments and recommended new experiments to differentiate between potential models. This approach holds promise for analyzing high-throughput CRISPR-Cas-based knockout experiments, offering valuable insights for plant breeding (Van Huffel et al., 2022). Symbolic regression uses genetic programming to automatically discover a white-box

model of one system (Angelis et al., 2023; Cranmer, 2023) – ideal to find the earlier-discussed meta-mechanisms. The DreamCoder system can uncover simple programs that generate example datasets (Ellis et al., 2023). These programs encompass diverse forms, such as regular expressions, graphics, symbolic equations, and physical laws. These techniques aid in discovering allometric laws and meta-mechanisms to support model building. They also facilitate the automatic extraction of rules for L-systems, enabling the creation of virtual plants based on a limited set of examples. In summary, custom computer algebra systems (Ma et al., 2022) and language compilers can streamline equations and code, resulting in concise and numerically stable plant models.

4 Discussion and outlook

The sections above discussed SI and its (potential) impact on the plant sciences. Here, we will give a more holistic point of view of why SI can be instrumental in discovering new and improving current practices in plant sciences. SI provides a holistic, top-down look at plant science and a systemic approach for leveraging fragmented phenotypic data and ecophysiological knowledge contained in process-based models.

Process-based plant models (including FSPMs) are continuously under development and updated with relevant knowledge. These are applied for decision support and (climate) scenario analysis, but also for answering scientific, plant-physiological questions, often in combination with plant phenotypic data. High-throughput plant phenotyping data of crop performance and development is now also being applied in crop breeding (Araus and Cairns, 2014; Gill et al., 2022). Still, the impact of these phenotypic data often needs to be more specific to the objectives of the experiments wherein these were collected. Therefore, SI concepts can facilitate the connection between phenotypic data and ecophysiological plant models and, as such, broaden the use of these phenotypic data and expand knowledge of the processes they rely on.

We identify three key prerequisites to embrace SI's philosophy in plant sciences fully. (1) The development of cheap sensor technology for environmental monitoring and plant phenotyping enables continuous monitoring of larger populations in real-time. (2) Open datasets and code, following the FAIR principles (Wilkinson et al., 2016), which is finding its way into plant sciences (Saint Cast et al., 2022) but is already much more prevalent in other scientific domains (Scheffler et al., 2022). (3) Interdisciplinary collaborations bridge potential knowledge gaps and enable cross-disciplinary approaches to succeed faster.

The increased amount of open data, along with AI, allows the processing of larger amounts of combined data and models and opens up new or improved applications in plant breeding or greenhouse control, as seen in other domains, e.g. Degraeve et al. (2022) who leveraged simulation and experimental data to learn a closed-loop controller a tokamak reactor. Recent research explores similar hybrid approaches to control plant and crop systems (Kang and Wang, 2017; Ifrim et al., 2021; Mahmood et al., 2023). Simulation and scientific computing is centered around creating mechanistic computational models to simulate real-world phenomena, while

machine learning focuses on leveraging learning algorithms to extract knowledge and insights from scientific data. Both approaches have strengths and can be combined to enhance scientific understanding and decision-making.

Throughout our literature survey, we identified the SI motifs embedded in numerous plant-related projects. This finding aligns with the core objectives of SI, which are geared towards addressing the issues inherent in modeling complex systems:

- Solving inverse problems;
- Integrating mechanistic knowledge with data;
- Navigating uncertainty, and;
- Fostering effective communication between the model and the user.

In applied domains, such as plant and crop modeling, advancements often trail the cutting edge of computational methodologies. Consequently, it is unsurprising that relatively mature SI motifs, such as surrogate models, agent-based modeling, and differentiable programming, showcase the highest prevalence in plant-related examples. In contrast, motifs in their infancy, such as probabilistic programming, open-ended optimization, and program synthesis, exhibit fewer concrete applications in plant and crop science. These emerging SI topics can advance plant and agricultural sciences toward a more sustainable future.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material. Further inquiries can be directed to the corresponding author.

References

- Abuelyzeed, D., Xu, C., Hoffman, F. M., Liu, J., Jones, A. W., Rackauckas, C., et al. (2023). A differentiable, physics-informed ecosystem modeling and learning framework for large-scale inverse problems: Demonstration with photosynthesis simulations. *Biogeosciences* 20, 2671–2692. doi: 10.5194/bg-20-2671-2023
- Ackley, D., and Small, T. (2014). “Indefinitely scalable computing = artificial life engineering,” in *ALIFE 14: The Fourteenth International Conference on the Synthesis and Simulation of Living Systems* (Manhattan, New York: MIT Press), 606–613. doi: 10.1162/978-0-262-32621-6-ch098
- AlQuraishi, M., and Sorger, P. K. (2021). Differentiable biology: Using deep learning for biophysics-based and data-driven modeling of molecular mechanisms. *Nat. Methods* 18 (10), 1169–1180. doi: 10.1038/s41592-021-01283-4
- Angelis, D., Sofos, F., and Karakasidis, T. E. (2023). Artificial intelligence in physical sciences: Symbolic regression trends and perspectives. *Arch. Comput. Methods Eng.* 30, 3845–3865. doi: 10.1007/s11831-023-09922-z
- Antle, J. M., Jones, J. W., and Rosenzweig, C. (2017). Next generation agricultural system models and knowledge products: Synthesis and strategy. *Agric. Syst.* 155, 179–185. doi: 10.1016/j.agsy.2017.05.006
- Araus, J. L., and Cairns, J. E. (2014). Field high-throughput phenotyping: The new crop breeding frontier. *Trends Plant Sci.* 19, 52–61. doi: 10.1016/j.tplants.2013.09.008
- Ariesen-Verschuur, N., Verdouw, C., Tekinerdogan, B., Ariesen-Verschuur, N., Verdouw, C., and Tekinerdogan, B. (2022). Digital Twins in greenhouse horticulture: A review. doi: 10.1016/J.COMPAG.2022.107183
- Asseng, S., Ewert, F., Rosenzweig, C., Jones, J. W., Hatfield, J. L., Ruane, A. C., et al. (2013). Uncertainty in simulating wheat yields under climate change. *Nat. Climate Change* 3, 827–832. doi: 10.1038/nclimate1916
- Bailey, B. N. (2018). A reverse ray-tracing method for modelling the net radiative flux in leaf-resolving plant canopy simulations. *Ecol. Model.* 368, 233–245. doi: 10.1016/j.ecolmodel.2017.11.022
- Balaz, I., Petrić, T., Kovacevic, M., Tsompanas, M.-A., and Stillman, N. (2021). Harnessing adaptive novelty for automated generation of cancer treatments. *Biosystems* 199, 104290. doi: 10.1016/j.biosystems.2020.104290
- Baluška, F., Lev-Yadun, S., and Mancuso, S. (2010a). Swarm intelligence in plant roots. *Trends Ecol. Evol.* 25, 682–683. doi: 10.1016/j.tree.2010.09.003
- Baluška, F., Mancuso, S., Volkmann, D., and Barlow, P. W. (2010b). Root apex transition zone: A signalling–response nexus in the root. *Trends Plant Sci.* 15, 402–408. doi: 10.1016/j.tplants.2010.04.007
- Banzhaf, W., Baumgaertner, B., Beslon, G., Doursat, R., Foster, J. A., McMullin, B., et al. (2016). Defining and simulating open-ended novelty: Requirements, guidelines, and challenges. *Theory Biosci.* 135, 131–161. doi: 10.1007/s12064-016-0229-7
- Bartzanas, T., Kacira, M., Zhu, H., Karmakar, S., Tamimi, E., Katsoulas, N., et al. (2013). Computational fluid dynamics applications to improve crop production systems. *Comput. Electron. Agric.* 93, 151–167. doi: 10.1016/j.compag.2012.05.012
- Baydin, A. G., Pearlmutter, B. A., Radul, A. A., and Siskind, J. M. (2018). Automatic differentiation in machine learning: A survey. *J. Mach. Learn. Res.* 18, 1–43.
- Bezgin, D. A., Buhendwa, A. B., and Adams, N. A. (2023). JAX-Fluids: A fully-differentiable high-order computational fluid dynamics solver for compressible two-phase flows. *Comput. Phys. Commun.* 282, 108527. doi: 10.1016/j.cpc.2022.108527
- Bingham, E., Chen, J. P., Jankowiak, M., Obermeyer, F., Pradhan, N., Karaletsos, T., et al. (2019). Pyro: deep universal probabilistic programming. *J. Mach. Learn. Res.* 20, 1–6.

Author contributions

MS: Writing – original draft, Writing – review & editing. OP: Writing – original draft, Writing – review & editing. TD: Writing – original draft, Writing – review & editing. FW: Writing – original draft, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work was partially funded by Ghent University grant number BOF-GOA-01G01923.

Conflict of interest

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.

Publisher’s note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

- Cavanagh, H., Mosbach, A., Scalliet, G., Lind, R., and Endres, R. G. (2021). Physics-informed deep learning characterizes morphodynamics of Asian soybean rust disease. *Nat. Commun.* 12, 6424. doi: 10.1038/s41467-021-26577-1
- Chaux, J. D., Sanchez-Londono, D., and Barbieri, G. (2021). A digital twin architecture to optimize productivity within controlled environment agriculture. *Appl. Sci.* 11, 8875. doi: 10.3390/AP11198875
- Chen, R. T. Q., Rubanova, Y., Bettencourt, J., and Duvenaud, D. (2018). "Neural ordinary differential equations," in *Proceedings of the 32nd Conference on Advances in Neural Information Processing Systems* (Montreal, Canada), 1–13. doi: 10.1007/978-3-662-55774-73
- Cheng, D., Yao, Y., Liu, R., Li, X., Guan, B., and Yu, F. (2023). Precision agriculture management based on a surrogate model assisted multiobjective algorithmic framework. *Sci. Rep.* 13, 1142. doi: 10.1038/s41598-023-27990-w
- Corrales, D. C., Schoving, C., Raynal, H., Debaeke, P., Journet, E.-P., and Constantin, J. (2022). A surrogate model based on feature selection techniques and regression learners to improve soybean yield prediction in southern France. *Comput. Electron. Agric.* 192, 106578. doi: 10.1016/j.compag.2021.106578
- Cranmer, M. (2023). Interpretable machine learning for science with PySR and SymbolicRegression.jl. doi: 10.48550/arXiv.2305.01582. Dataset.
- Cranmer, K., Brehmer, J., and Louppe, G. (2020). The frontier of simulation-based inference. *Proc. Natl. Acad. Sci.* 117, 30055–30062. doi: 10.1073/pnas.1912789117
- Damineli, D. S. C., Portes, M. T., and Feijó, J. A. (2022). Electrifying rhythms in plant cells. *Curr. Opin. Cell Biol.* 77, 102113. doi: 10.1016/j.cob.2022.102113
- Das Choudhury, S., Bashyam, S., Qiu, Y., Samal, A., and Awada, T. (2018). Holistic and component plant phenotyping using temporal image sequence. *Plant Methods* 14, 35. doi: 10.1186/s13007-018-0303-x
- David, C., and Kroening, D. (2017). Program synthesis: Challenges and opportunities. *Philos. Trans. R. Soc. A: Mathematical Phys. Eng. Sci.* 375, 20150403. doi: 10.1098/rsta.2015.0403
- DeAngelis, D. L., and Mooij, W. M. (2005). Individual-based modeling of ecological and evolutionary processes. *Annu. Rev. Ecology Evolution Systematics* 36, 147–168. doi: 10.1146/annurev
- de Avila Belbutte-Peres, F., Smith, K., Allen, K., Tenenbaum, J., and Kolter, J. Z. (2018). "End-to-end differentiable physics for learning and control," in *Advances in Neural Information Processing Systems*, vol. 31. (Curran Associates, Inc).
- Degrave, J., Felici, F., Buchli, J., Neunert, M., Tracey, B., Carpanese, F., et al. (2022). Magnetic control of tokamak plasmas through deep reinforcement learning. *Nature* 602, 414–419. doi: 10.1038/s41586-021-04301-9
- Degrave, J., Hermans, M., Dambre, J., and Wyffels, F. (2019). A differentiable physics engine for deep learning in robotics. *Front. Neurobotics* 13. doi: 10.3389/fnbot.2019.00006
- De Meyer, S., Cruz, D. F., De Swaef, T., Lootens, P., De Block, J., Bird, K., et al. (2023). Predicting yield of individual field-grown rapeseed plants from rosette-stage leaf gene expression. *PLoS Comput. Biol.* 19, e1011161. doi: 10.1371/journal.pcbi.1011161
- De Swaef, T., Pieters, O., Appeltans, S., Borra-Serrano, I., Coudron, W., Couvreur, V., et al. (2022). On the pivotal role of water potential to model plant physiological processes. *silico Plants* 4, diab038. doi: 10.1093/silicoplants/diab038
- De Visser, P., van der Heijden, G., and Buck-Sorlin, G. (2014). Optimizing illumination in the greenhouse using a 3D model of tomato and a ray tracer. *Front. Plant Sci.* 5. doi: 10.3389/fpls.2014.00048
- Dokoohaki, H., Kivi, M. S., Martinez-Feria, R., Miguez, F. E., and Hoogenboom, G. (2021). A comprehensive uncertainty quantification of large-scale process-based crop modeling frameworks. *Environ. Res. Lett.* 16, 084010. doi: 10.1088/1748-9326/ac0f26
- Ellis, K., Wong, C., Nye, M., Sable-Meyer, M., Cary, L., Morales, L., et al. (2023). DreamCoder: Growing 474 generalizable, interpretable knowledge with wake-sleep Bayesian program learning. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 381, 1–18. 476. doi: 10.1098/rsta.2022.0050
- Eloy, C., Fournier, M., Lacoite, A., and Mouliat, B. (2017). Wind loads and competition for light sculpt trees into self-similar structures. *Nat. Commun.* 8, 1–12. doi: 10.1038/s41467-017-00995-6
- Farquhar, G. D., von Caemmerer, S., and Berry, J. A. (1980). A biochemical model of photosynthetic CO₂ assimilation in leaves of C₃ species. *Planta* 149, 78–90. doi: 10.1007/BF00386231
- Fix, S., Probst, T., Ruggli, O., Hanne, T., and Christen, P. (2021). Open-ended automatic programming through combinatorial evolution. *Intelligent Syst. Design Appl.* 418, 1–12. doi: 10.48550/arXiv.2102.10475
- Folberth, C., Elliott, J., Müller, C., Balković, J., Chrysanthopoulos, J., Izaurralde, R. C., et al. (2019). Parameterization-induced uncertainties and impacts of crop management harmonization in a global gridded crop model ensemble. *PLoS One* 14, e0221862. doi: 10.1371/journal.pone.0221862
- Gherman, I. M., Abdallah, Z. S., Pang, W., Gorochowski, T. E., Grierson, C. S., and Marucci, L. (2023). Bridging the gap between mechanistic biological models and machine learning surrogates. *PLoS Comput. Biol.* 19, e1010988. doi: 10.1371/journal.pcbi.1010988
- Gill, T., Gill, S. K., Saini, D. K., Chopra, Y., de Koff, J. P., and Sandhu, K. S. (2022). A comprehensive review of high throughput phenotyping and machine learning for plant stress phenotyping. *Phenomics* 2, 156–183. doi: 10.1007/s43657-022-00048-z
- Greenewegen, L. (1977). Multilines as a tool in breeding for reliable yields. *Cereal Res. Commun.* 5, 125–132.
- Guttenberg, N., Virgo, N., and Penn, A. (2018). On the potential for open-endedness in neural networks. *Artif. Life* 25, 145–167. doi: 10.48550/arXiv.1812.04907
- Hemming, S., de Zwart, F., Elings, A., Petropoulou, A., and Righini, I. (2020). Cherry tomato production in intelligent greenhouses—Sensors and AI for control of climate, irrigation, crop yield, and quality. *Sensors* 20, 6430. doi: 10.3390/s20226430
- Hernández, S., and López, J. L. (2020). Uncertainty quantification for plant disease detection using Bayesian deep learning. *Appl. Soft Computing* 96, 106597. doi: 10.1016/j.asoc.2020.106597
- Hickey, J. M., Chiurugwi, T., Mackay, I., and Powell, W. (2017). Genomic prediction unifies animal and plant breeding programs to form platforms for biological discovery. *Nat. Genet.* 49, 1297–1303. doi: 10.1038/ng.3920
- Holt, R. C., and Cordy, J. R. (1988). The Turing programming language. *Commun. ACM* 31, 1410–1423. doi: 10.1145/53580.53581
- Hüllermeier, E., and Waegeman, W. (2021). Aleatoric and epistemic uncertainty in machine learning: An introduction to concepts and methods. *Mach. Learn.* 110, 457–506. doi: 10.1007/s10994-021-05946-3
- Humplik, J. F., Dostál, J., Ugena, L., Spíchal, L., De Diego, N., Vencálek, O., et al. (2020). Bayesian approach for analysis of time-to-event data in plant biology. *Plant Methods* 16, 14. doi: 10.1186/s13007-020-0554-1
- Ifrim, G., Barbu, M., Horincar, G., and Titica, M. (2021). "Data-driven multivariable control of a microalgae growth process," in *2021 25th International Conference on System Theory, Control and Computing (ICSTCC)* (Iasi, Romania), 321–326. doi: 10.1109/ICSTCC52150.2021.9607111
- Ingraham, J., Riesselman, A., Sander, C., and Marks, D. (2019). "Learning protein structure with a differentiable simulator," in *7th International Conference on Learning Representations, ICLR 2019*. 1–24.
- Innes, M., Edelman, A., Fischer, K., Rackauckas, C., Saba, E., Shah, V. B., et al. (2019). A differentiable programming system to bridge machine learning and scientific computing. doi: 10.48550/arXiv.1907.07587
- Izzo, D., Biscani, F., and Mereta, A. (2016). Differentiable genetic programming. doi: 10.48550/arXiv.1611.04766. Dataset.
- Jans-Singh, M., Leeming, K., Choudhary, R., and Girolami, M. (2020). Digital twin of an urban-integrated hydroponic farm. *Data-Centric Eng.* 1, e20. doi: 10.1017/DCE.2020.21
- Jiao, W., Liu, Q., Gao, L., Liu, K., Shi, R., and Ta, N. (2020). Computational fluid dynamics-based simulation of crop canopy temperature and humidity in double-film solar greenhouse. *J. Sensors* 2020, e8874468. doi: 10.1155/2020/8874468
- Kang, M., and Wang, F.-Y. (2017). From parallel plants to smart plants: Intelligent control and management for plant growth. *IEEE/CAA J. Automatica Sin.* 4, 161–166. doi: 10.1109/JAS.2017.7510487
- Karniadakis, G. E., Kevrekidis, I. G., Lu, L., Perdikaris, P., Wang, S., and Yang, L. (2021). Physics-informed machine learning. *Nat. Rev. Phys.* 3, 422–440. doi: 10.1038/s42254-021-00314-5
- Kempes, C. P., Koehl, M. A., and West, G. B. (2019). The scales that limit: The physical boundaries of evolution. *Front. Ecol. Evol.* 7. doi: 10.3389/fevo.2019.00242
- Koksal, A. S., Pu, Y., Srivastava, S., Bodik, R., Fisher, J., and Piterman, N. (2013). "Synthesis of biological models from mutation experiments," in *Proceedings of the 40th Annual ACM SIGPLAN-SIGACT Symposium on Principles of Programming Languages* (New York, NY, USA: Association for Computing Machinery), 469–482, POPL '13. doi: 10.1145/2429069.2429125
- Korte, A., and Farlow, A. (2013). The advantages and limitations of trait analysis with GWAS: A review. *Plant Methods* 9, 29. doi: 10.1186/1746-4811-9-29
- Lacasa, J., Gaspar, A., Hinds, M., Jayasinghe Don, S., Berning, D., and Ciampitti, I. A. (2020). Bayesian approach for maize yield response to plant density from both agronomic and economic viewpoints in North America. *Sci. Rep.* 10, 15948. doi: 10.1038/s41598-020-72693-1
- Lavin, A., Zenil, H., Paige, B., Krakauer, D., Gottschlich, J., Mattson, T., et al. (2021). Simulation intelligence: Towards a new generation of scientific methods. doi: 10.48550/arXiv.2112.03235
- Lenaerts, B., Collard, B. C. Y., and Demont, M. (2019). Review: Improving global food security through accelerated plant breeding. *Plant Sci.* 287, 110207. doi: 10.1016/j.plantsci.2019.110207
- Li, T.-M., Aittala, M., Durand, F., and Lehtinen, J. (2018). Differentiable Monte Carlo ray tracing through edge sampling. *ACM Trans. Graphics* 37, 222:1–222:11. doi: 10.1145/3272127.3275109
- Li, W., Zhu, D., and Wang, Q. (2022). A single view leaf reconstruction method based on the fusion of ResNet and differentiable render in plantgrowthdigitaltwinsystem. *Comput. Electron. Agric.* 193, 106712. doi: 10.1016/j.compag.2022.106712
- Liu, J., Gao, F., and Zhang, J. (2020a). Gumbel-softmax optimization: A simple general framework for combinatorial optimization problems on graphs. *Stud. Comput. Intell.* 881, 879–890. doi: 10.1007/978-3-030-36687-273
- Liu, Y., Kumar, M., Katul, G. G., Feng, X., and Konings, A. G. (2020b). Plant hydraulics accentuates the effect of atmospheric moisture stress on transpiration. *Nat. Climate Change* 10, 691–695. doi: 10.1038/s41558-020-0781-5

- Louwaars, N. P. (2018). Plant breeding and diversity: A troubled relationship? *Euphytica* 214, 114. doi: 10.1007/s10681-018-2192-5
- Ma, Y., Gowda, S., Anantharaman, R., Laughman, C., Shah, V., and Rackauckas, C. (2022). ModelingToolkit: A composable graph transformation system for equation-based modeling. doi: 10.48550/arXiv.2103.05244. Dataset.
- Mahmood, F., Govindan, R., Bernak, A., Yang, D., and Al-Ansari, T. (2023). Data-driven robust model predictive control for greenhouse temperature control and energy utilisation assessment. *Appl. Energy* 343, 121190. doi: 10.1016/j.apenergy.2023.121190
- Marin, J.-M., Pudlo, P., Robert, C. P., and Ryder, R. J. (2012). Approximate Bayesian computational methods. *Stat. Computing* 22, 1167–1180. doi: 10.1007/s11222-011-9288-2
- McLane, A. J., Semeniuk, C., McDermaid, G. J., and Marceau, D. J. (2011). The role of agent-based models in wildlife ecology and management. *Ecol. Model.* 222, 1544–1556. doi: 10.1016/j.ecolmodel.2011.01.020
- Merrell, D., and Gitter, A. (2020). Inferring signaling pathways with probabilistic programming. *Bioinformatics* 36, i822–i830. doi: 10.1093/bioinformatics/btaa861
- Mila, A. L., and Carriquiry, A. L. (2004). Bayesian analysis in plant pathology. *Phytopathology* 94, 1027–1030. doi: 10.1094/PHYTO.2004.94.9.1027
- Moghadam, P., Lowe, T., and Edwards, E. (2020). Digital twin for the future of orchard production systems. *Proceedings* 36, 92. doi: 10.3390/PROCEEDINGS2019036092
- Monti, C., Pangallo, M., De Francisci Morales, G., and Bonchi, F. (2023). On learning agent-based models from data. *Sci. Rep.* 13, 9268. doi: 10.1038/s41598-023-35536-3
- Moreta, L. S., Al-Sibahi, A. S., Theobald, D., Bullock, W., Rommes, B. N., Manoukian, A., et al. (2019). “A probabilistic programming approach to protein structure superposition,” in *2019 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB)*. 1–5. doi: 10.1109/CIBCB.2019.8791469
- Nelson, N., Stubbs, C. J., Larson, R., and Cook, D. D. (2019). Measurement accuracy and uncertainty in plant biomechanics. *J. Exp. Bot.* 70, 3649–3658. doi: 10.1093/jxb/erz279
- Núñez, M., Barreiro, N. L., Barrio, R. A., and Rackauckas, C. (2023). Forecasting virus outbreaks with social media data via neural ordinary differential equations. *Sci. Rep.* 13, 10870. doi: 10.1038/s41598-023-37118-9
- Oborny, B. (2019). The plant body as a network of semi-autonomous agents: A review. *Philos. Trans. R. Soc. B: Biol. Sci.* 374, 20180371. doi: 10.1098/rstb.2018.0371
- Pearl, J., and Mackenzie, D. (2018). *The Book of Why: The New Science of Cause and Effect* (New York City: Basic Books).
- Peng, B., Guan, K., Tang, J., Ainsworth, E. A., Asseng, S., Bernacchi, C. J., et al. (2020). Towards a multiscale crop modelling framework for climate change adaptation assessment. *Nat. Plants* 6, 338–348. doi: 10.1038/s41477-020-0625-3
- Pugh, J. K., Soros, L. B., and Stanley, K. O. (2016). Quality diversity: A new frontier for evolutionary computation. *Front. Robotics AI* 3. doi: 10.3389/frobt.2016.00040
- Purcell, W., and Neubauer, T. (2023). Digital twins in agriculture: A state-of-the-art review. *Smart Agric. Technol.* 3, 100094. doi: 10.1016/j.atech.2022.100094
- Rackauckas, C., Ma, Y., Martensen, J., Warner, C., Zubov, K., Supekar, R., et al. (2021). Universal differential equations for scientific machine learning. doi: 10.48550/arXiv.2001.04385. Dataset.
- Raissi, M., Yazdani, A., and Karniadakis, G. E. (2020). Hidden fluid mechanics: Learning velocity and pressure fields from flow visualizations. *Sci. (New York N.Y.)* 367, 1026–1030. doi: 10.1126/science.aaw4741
- Retkute, R., Townsend, A. J., Murchie, E. H., Jensen, O. E., and Preston, S. P. (2018). Three-dimensional plant architecture and sunlit-shaded patterns: A stochastic model of light dynamics in canopies. *Ann. Bot.* 122, 291–302. doi: 10.1093/aob/mcy067
- Ritchie, D., Thomas, A., Hanrahan, P., and Goodman, N. D. (2016). “Neurally-guided procedural models: Amortized inference for procedural graphics programs using neural networks,” in *Proceedings of the 29th Conference on Neural Information Processing Systems*.
- Romero-Cuellar, J., and Francés, F. (2023). How to assess climate change impact models: Uncertainty analysis of streamflow statistics via approximate Bayesian computation (ABC). *Hydrological Sci. J.* 68, 1611–1626. doi: 10.1080/02626667.2023.2231437
- Ronquist, F., Kudlicka, J., Senderov, V., Borgström, J., Lartillot, N., Lundén, D., et al. (2020). Universal probabilistic programming: A powerful new approach to statistical phylogenetics. *bioRxiv* 4, 244. doi: 10.1101/2020.06.16.154443
- Saint Cast, C., Lobet, G., Cabrera-Bosquet, L., Couvreur, V., Pradal, C., Tardieu, F., et al. (2022). Connecting plant phenotyping and modelling communities: Lessons from science mapping and operational perspectives. *silico Plants* 4, diac005. doi: 10.1093/insilicoplants/diac005
- Scheffler, M., Aeschlimann, M., Albrecht, M., Bereau, T., Bungartz, H.-J., Felsler, C., et al. (2022). FAIR data enabling new horizons for materials research. *Nature* 604, 635–642. doi: 10.1038/s41586-022-04501-x
- Schölkopf, B., Locatello, F., Bauer, S., Ke, N. R., Kalchbrenner, N., Goyal, A., et al. (2021). Towards causal representation learning. *Proc. IEEE* 109, 612–634. doi: 10.1109/JPROC.2021.3058954
- Schoot, R., Depaoli, S., King, R., Kramer, B., Mamp, K., Tadesse, M. G., et al. (2021). Bayesian statistics and modelling. *Nat. Rev. Methods Primers* 1, 1. doi: 10.1038/s43586-020-00001-2
- Shen, C., Appling, A. P., Gentine, P., Bandai, T., Gupta, H., Tartakovsky, A., et al. (2023). Differentiable modelling to unify machine learning and physical models for geosciences. *Nat. Rev. Earth Environ.* 4, 552–567. doi: 10.1038/s43017-023-00450-9
- Silva, J. V., and Giller, K. E. (2020). Grand challenges for the 21st century: What crop models can and can't (yet) do. *J. Agric. Sci.* 158, 794–805. doi: 10.1017/S0021859621000150
- Skobelev, P. O., Mayorov, I. V., Simonova, E. V., Goryanin, O. I., Zhilyaev, A. A., Tabachinskiy, A. S., et al. (2020). Development of models and methods for creating a digital twin of plant within the cyber-physical system for precision farming management. *J. Physics: Conf. Ser.* 1703, 12022. doi: 10.1088/1742-6596/1703/1/012022
- Slob, N., Hurst, W., Slob, N., and Hurst, W. (2022). Digital twins and industry 4.0 technologies for agricultural greenhouses. doi: 10.3390/SMARTCITIES5030059
- Slob, N., Hurst, W., van de Zedde, R., and Tekinerdogan, B. (2023). Virtual reality-based digital twins for greenhouses: A focus on human interaction. *Comput. Electron. Agric.* 208, 107815. doi: 10.1016/j.compag.2023.107815
- Stan Development Team (2023). *Stan Modeling Language Users Guide and Reference Manual*. Available at: <https://mc-stan.org>.
- Stanley, K., and Lehman, J. (2015). *Why Greatness Cannot Be Planned: The Myth of the Objective* (Springer).
- Stanley, K. O., Lehman, J., and Soros, L. (2017). *Open-endedness: The last grand challenge you've never heard of*. Available at: <https://www.oreilly.com/radar/open-endedness-the-last-grand-challenge-you-never-heard-of/>.
- Stock, M., and Gorochoowski, T. (2023). Open-endedness in synthetic biology: a route to continual innovation for biological design. *Sci. Adv.* doi: 10.31219/osf.io/wv5ac
- Tao, F., Rötter, R. P., Palosuo, T., Gregorio Hernández Díaz-Ambrona, C., Mínguez, M. I., Semenov, M. A., et al. (2018). Contribution of crop model structure, parameters and climate projections to uncertainty in climate change impact assessments. *Global Change Biol.* 24, 1291–1307. doi: 10.1111/gcb.14019
- Tardieu, F., Granato, I. S. C., Van Oosterom, E. J., Parent, B., and Hammer, G. L. (2020). Are crop and detailed physiological models equally ‘mechanistic’ for predicting the genetic variability of wholeplant behaviour? The nexus between mechanisms and adaptive strategies. *silico Plants* 2, diaa011. doi: 10.1093/insilicoplants/diaa011
- Technow, F., Messina, C. D., Totir, L. R., and Cooper, M. (2015). Integrating crop growth models with whole genome prediction through approximate bayesian computation. *PLoS One* 10, e0130855. doi: 10.1371/journal.pone.0130855
- Van Huffel, K., Stock, M., Ruttink, T., and De Baets, B. (2022). Covering the combinatorial design space of multiplex CRISPR/cas experiments in plants. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.907095
- van Schijndel, L., Snoek, B. L., and ten Tusscher, K. (2022). Embodiment in distributed information processing: “Solid” plants versus “liquid” ant colonies. *Quantitative Plant Biol.* 3, e27. doi: 10.1017/qpb.2022.22
- Verdouw, C., Tekinerdogan, B., Beulens, A. J. M., and Wolfert, S. (2021). Digital twins in smart farming. *Agric. Syst.* 189, 1–19. doi: 10.1016/j.agsy.2020.103046
- Wallach, D., Palosuo, T., Thorburn, P., Hochman, Z., Gourdain, E., Andrianasolo, F., et al. (2021). The chaos in calibrating crop models: Lessons learned from a multi-model calibration exercise. *Environ. Model. Softw.* 145, 105206. doi: 10.1016/j.envsoft.2021.105206
- Walter, A., Liebisch, F., and Hund, A. (2015). Plant phenotyping: From bean weighing to image analysis. *Plant Methods* 11, 14. doi: 10.1186/s13007-015-0056-8
- Whalley, W., Binley, A., Watts, C., Shanahan, P., Dodd, I., Ober, E., et al. (2017). Methods to estimate changes in soil water for phenotyping root activity in the field. *Plant Soil* 415, 407–422. doi: 10.1007/s11104-016-3161-1
- Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., et al. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Sci. Data* 3, 160018. doi: 10.1038/sdata.2016.18
- Zhang, B., and DeAngelis, D. L. (2020). An overview of agent-based models in plant biology and ecology. *Ann. Bot.* 126, 539–557. doi: 10.1093/aob/mcaa043
- Zhang, N., Zhou, X., Kang, M., Hu, B.-G., Heuvelink, E., and Marcelis, L. F. M. (2023). Machine learning versus crop growth models: An ally, not a rival. *AoB Plants* 15, plac061. doi: 10.1093/aobpla/plac061
- Ziegler, C., Dyson, R. J., and Johnston, I. G. (2019). Model selection and parameter estimation for root architecture models using likelihood-free inference. *J. R. Soc. Interface* 16, 20190293. doi: 10.1098/rsif.2019.0293