



## OPEN ACCESS

## EDITED BY

Pierre Echaubard,  
SOAS University of London, United Kingdom

## REVIEWED BY

Andrew W. Bartlow,  
Los Alamos National Laboratory (DOE),  
United States  
Sara Savic,  
Scientific Veterinary Institute Novi Sad, Serbia

## \*CORRESPONDENCE

Ted J. Lawrence  
✉ TJL222@cornell.edu

RECEIVED 21 June 2023

ACCEPTED 16 October 2023

PUBLISHED 02 November 2023

## CITATION

Lawrence TJ, Takenaka BP, Garg A, Tao D,  
Deem SL, Fèvre EM, Gluecks I, Sagan V and  
Shacham E (2023) A global examination of  
ecological niche modeling to predict emerging  
infectious diseases: a systematic review.  
*Front. Public Health* 11:1244084.  
doi: 10.3389/fpubh.2023.1244084

## COPYRIGHT

© 2023 Lawrence, Takenaka, Garg, Tao, Deem,  
Fèvre, Gluecks, Sagan and Shacham. 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.

# A global examination of ecological niche modeling to predict emerging infectious diseases: a systematic review

Ted J. Lawrence <sup>1\*</sup>, Bryce P. Takenaka <sup>2</sup>, Aastha Garg <sup>2</sup>,  
Donghua Tao <sup>3</sup>, Sharon L. Deem <sup>4</sup>, Eric M. Fèvre <sup>5,6</sup>,  
Ilona Gluecks <sup>5</sup>, Vasit Sagan <sup>1,7</sup> and Enbal Shacham <sup>1,2</sup>

<sup>1</sup>Taylor Geospatial Institute, St. Louis, MO, United States, <sup>2</sup>College for Public Health and Social Justice, Saint Louis University, St. Louis, MO, United States, <sup>3</sup>Medical Center Library, Saint Louis University, St. Louis, MO, United States, <sup>4</sup>Institute for Conservation Medicine, Saint Louis Zoo, St. Louis, MO, United States, <sup>5</sup>International Livestock Research Institute, Nairobi, Kenya, <sup>6</sup>Institute of Infection, Veterinary and Ecological Sciences, University of Liverpool, Liverpool, United Kingdom, <sup>7</sup>Department of Earth and Atmospheric Sciences, Saint Louis University, St. Louis, MO, United States

**Introduction:** As emerging infectious diseases (EIDs) increase, examining the underlying social and environmental conditions that drive EIDs is urgently needed. Ecological niche modeling (ENM) is increasingly employed to predict disease emergence based on the spatial distribution of biotic conditions and interactions, abiotic conditions, and the mobility or dispersal of vector-host species, as well as social factors that modify the host species' spatial distribution. Still, ENM applied to EIDs is relatively new with varying algorithms and data types. We conducted a systematic review (PROSPERO: CRD42021251968) with the research question: *What is the state of the science and practice of estimating ecological niches via ENM to predict the emergence and spread of vector-borne and/or zoonotic diseases?*

**Methods:** We searched five research databases and eight widely recognized One Health journals between 1995 and 2020. We screened 383 articles at the abstract level (included if study involved vector-borne or zoonotic disease and applied ENM) and 237 articles at the full-text level (included if study described ENM features and modeling processes). Our objectives were to: (1) describe the growth and distribution of studies across the types of infectious diseases, scientific fields, and geographic regions; (2) evaluate the likely effectiveness of the studies to represent ecological niches based on the biotic, abiotic, and mobility framework; (3) explain some potential pitfalls of ENM algorithms and techniques; and (4) provide specific recommendation for future studies on the analysis of ecological niches to predict EIDs.

**Results:** We show that 99% of studies included mobility factors, 90% modeled abiotic factors with more than half in tropical climate zones, 54% modeled biotic conditions and interactions. Of the 121 studies, 7% include only biotic and mobility factors, 45% include only abiotic and mobility factors, and 45% fully integrated the biotic, abiotic, and mobility data. Only 13% of studies included modifying social factors such as land use. A majority of studies (77%) used well-recognized ENM algorithms (MaxEnt and GARP) and model selection procedures. Most studies (90%) reported model validation procedures, but only 7% reported uncertainty analysis.

**Discussion:** Our findings bolster ENM to predict EIDs that can help inform the prevention of outbreaks and future epidemics.

**Systematic review registration:** <https://www.crd.york.ac.uk/prospero/>, identifier (CRD42021251968).

KEYWORDS

disease early detection, ecological and environmental phenomena, epidemiology, One Health, outbreaks, public health

## 1. Introduction

Amid an increasing global trend in emerging infectious diseases (EIDs) and outbreaks (1, 2), the need to better understand underlying drivers of EIDs has never been greater (3). The social and environmental conditions and their complex interactions that have been commonly implicated as driving EIDs include: climate and land-use change; human-wildlife interactions; shifting human demographics and behaviors including livelihood patterns; and poverty and inequality (1, 4–6). However, modeling the fundamental mechanisms that drive EIDs is one of the most complex and challenging scientific problems as these factors are intricately intertwined and difficult to disentangle (7–10). Thus, there has been hesitance in traditional epidemiology to model these outcomes as the predictive accuracy is questionable (8).

Ecological niche modeling (ENM) is an approach that expands beyond traditional spatial epidemiology. Specifically, ENM is intended to determine geographic distribution and spatial relationships of environmental conditions, such as climate and land cover with disease occurrence, and thereby represent the ecological niche of the disease (11, 12). ENM enables the spatial prediction and mapping of geographic regions (beyond the locations analyzed) that are at risk to EIDs, so that public health measures can be implemented to prevent future outbreaks and epidemics. The main elements to ENM include occurrence data, environmental data (abiotic and/or biotic variables), and correlative or classification algorithms that link environmental conditions and disease occurrence (11). Ecological niches comprise environmental spaces that include sets of ecological variables, such as precipitation, land cover, and soil characteristics as well as human social factors that modify the behavior and spatial distribution of a host species, such as land use and types of land tenure that influence habitat modifications. These factors shape the potential occurrence of pathogens; and these niches translate into geographic distributions according to the combined effects of the spatial structure of the biotic conditions and interactions, abiotic conditions, and mobility or dispersal capacity of vector-host species (Figure 1) (12, 13).

Elucidating ecological niches in the ENM process is primarily based on algorithms that can be divided into three categories: presence-absence; presence-background; and presence-only (11). Presence is defined as the number of observed occurrences or presence sites of each individual from a host species. Presence-absence algorithms compare environmental conditions where a pathogen is present versus where it is absent. The presence-absence approach can be highly robust, but absence data may be limited due to actual absence, lack of data or untested data, and many datasets may only contain locations of presence. Such limited datasets of absences are due to incomplete sampling efforts needed to be confident of absence, which is often difficult to achieve and estimates of species range

changes based on survey data (14). Presence-background algorithms involve the same algorithms used for presence-absence models, but include *background samples*, which are usually based on random sampling of locations taken where the presence or absence of a species is unknown in the study area (15–17). The sampled background points are then used as pseudo-absences (18). The background points are compared to the observed presence data to help differentiate the environmental conditions under which a species can occur or not (15). The presence-background algorithms are sometimes referred to as enhanced presence-only algorithms and are intended to give a more accurate species distribution prediction than simple presence-only algorithms (19). Presence-only algorithms focus solely on the environmental values linked to each occurrence record for calibration (i.e., no direct information about absences is used) (20). Each of these model types enable the spatial prediction of EIDs, while the choice of approach is largely dependent on scale of analysis and data availability.

In the early 2000s, professionals concerned with multi-host pathogens began applying ENM to disease transmission systems (21). Since then, ENM applied to EIDs has been increasingly employed to

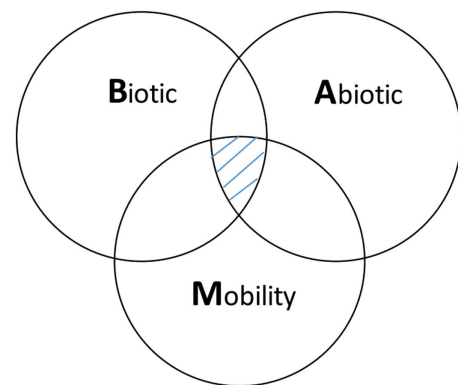


FIGURE 1

Biotic-abiotic-mobility (BAM) framework. The BAM framework is a simplified representation of a species' geographic distribution determined by the intersection of suitable biotic conditions and interactions, abiotic conditions, and the species' mobility, which is shown in the pinstriped center of the diagram. The abiotic component refers to the physical environment (e.g., temperature, precipitation, soil type) that limits the presence of a host-pathogen. The biotic component refers to the biological environment that shapes the distribution of a host-pathogen. Biotic factors include the biological components present in the environment (i.e., biotic conditions), such as vegetation and land cover and the biotic interactions between hosts and vectors, such as vector, host density and distribution that can promote or limit a pathogen's occurrence. Mobility or dispersal capacity of vector-host species refers to the areas that are accessible to a vector-host species due to limited dispersal abilities or biogeographical barriers.

model diverse pathogens with a focus on emerging and re-emerging vector-borne and zoonotic diseases (22). However, ENM applied to EIDs has been criticized due to the limited understanding of users regarding the potential data and algorithm limitations of this modeling framework (11). This is especially true for the inclusion of all BAM factors (Figure 1), which is considered the most rigorous approach and the users understanding of their biological meaning, assumptions, and available data in ENM (22). Moreover, few practical guidelines exist on the proper application of both BAM and ENM techniques (23) and especially applied to EIDs. Due to the growing, yet relatively new and limited number of studies using this approach, a better understanding about the application of ENM to EIDs is needed.

We conducted a global examination (via systematic review) of ENM to predict EIDs as a response to the need to better understand such applications. Our research question was: *What is the state of the science and practice of estimating ecological niches via ENM to predict the emergence and spread of vector-borne and/or zoonotic diseases?* We focused our research question on vector-borne and zoonotic diseases because the majority of emerging infectious diseases are zoonotic (24) and vector-borne diseases are major causes of mortality and morbidity globally (5, 25). Our objectives for this study were to: (1) describe the growth and distribution of studies across the types of infectious diseases, scientific fields, and geographic regions; (2) evaluate the likely effectiveness of the studies to represent ecological niches based on the BAM framework; (3) explain some potential pitfalls of ENM algorithms and techniques used in those studies; and (4) provide specific recommendations for future studies on the analysis of ecological niches to predict EIDs. Our intent is not to provide a practical step-by-step guide to the application of ENM, but rather to provide an overview of the science and practice of ENM that can bolster collaborations between non-experts and ENM experts on interdisciplinary teams to predict infectious disease outbreaks and epidemics. Our findings and recommendations can inform professionals concerned with multi-host pathogens that focus on the interconnections between the health of humans, animals, and their shared environments to help prevent outbreaks and future epidemics (26).

## 2. Methods

### 2.1. Literature search

Our systematic review followed PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-analyses) guidelines (27) and was registered as a protocol with PROSPERO (registration number: CRD42021251968). The bibliographic search of scientific articles was performed using Scopus, ISI Web of Science, PubMed, Medline, and Google Scholar databases. Also, we searched widely recognized One Health journals that publish articles on the linkages between environmental factors and EIDs, which included: *EcoHealth*; *One Health*; *International Journal of One Health*; *One Health Outlook*; *Infection Ecology and Epidemiology*; *The Lancet Planetary Health*; *Challenges*; and *Frontiers in Public Health – Planetary Health*. Our search terms were based on the concepts of modeling geospatial relationships between environmental factors and EIDs, and the scientific fields that conduct such research (Appendix 1). We did a pairwise search involving each individual search term from the 1st

concept area of geospatial modeling of environmental conditions coupled with each individual search term from the 2nd concept area of emerging infectious disease and scientific fields, which resulted in 40 searches in each of the five databases and in each of the seven journals for a total of 480 searches. Our search included articles published between 1995 and 2020 with no geographic or language limitations. Articles were included if they contained the specific terms shown in Table 1 that are associated with the concepts of geospatial modeling of environmental conditions and EIDs and scientific fields. Gray literature and articles reporting ENMs of species only and not involving an emerging infectious disease were excluded from our search. We managed citation, abstract, and article retrieval using Zotero (28).

### 2.2. Article filtering

Our search strategy resulted in a total 758 potentially relevant papers, and 383 articles after duplicates were removed (Figure 2) (29). Two reviewers screened all retrieved articles at the abstract and full-text levels. Articles were excluded ( $n = 146$ ) at the abstract level if they: (i) did not examine a vector-borne or zoonotic disease; and (ii) were purely conceptual (i.e., review, perspective, and commentary) and had not included ENM. Thus, only articles that reported applications of ENMs of vector-borne or zoonotic infectious diseases, or those whose content was unclear based on reading the abstract alone were retained for the second step of the review. Next, the full-text articles were assessed, and articles were excluded ( $n = 116$ ) if they: (i) did not describe key features of the ENM (i.e., describe and provide a rationale for data, algorithm, and application of the algorithm to data); and (ii) if they did not clearly explain or adequately describe the data, algorithm, application of the algorithm to the data, and the critical decisions necessary in the ENM construction process. Therefore, full-text articles were assessed to distinguish between studies that involved a click-and-run approach where researchers did not carefully consider and justify the data, algorithm, and application of the algorithm to the data (i.e., model construction process) versus studies that involved a robust-and-thorough approach that included some of the critical decisions necessary to develop a comprehensive, replicable ecological niche model that accounts for the natural history of the organism, the quality of the data, and the goal of the model (30). We based our full-text review criteria on Escobar (11) to differentiate between studies that did and did not adequately explain the ENM process and to identify studies that properly applied ENM to EIDs. Article exclusion rates for both reviewers at the abstract level were 63 and 55% and at the full-text level were 62 and 80%. The reviewers disagreed on 10% of articles at the abstract level and 33% at the full-text level. Disagreement between reviewers on potentially excluded articles were resolved by a blinded third reviewer and majority vote. In total, 121 publications were retained for analysis.

### 2.3. Data compilation, analysis and synthesis

Based on the studies that we identified as properly applying ENM to EID, we compiled and categorized data from the articles into a database according to the table in Appendix 2. First, we assessed

**TABLE 1** ENM features by studies in the systematic review of ecological niche modeling applied to vector-borne and zoonotic infectious diseases emergence.

Step	Prominent ENM features	Description	% Of studies
1	<b>Data approach</b>	Presence-absence	2%
		Presence-background	84%
		Presence-only	14%
2	<b>Predictor selection</b>	Environmental	
		Vegetation	36%
		Land cover types	43%
		Climate (temperature, precipitation)	89%
	Soil	21%	
Human social	Primarily population density and land use	13%	
3	<b>Algorithm applied</b>		
	Machine learning	Maxent	55%
		GARP	22%
	Statistical algorithm	Generalized linear models, boosted regression trees	3%
Mixed Methods	Maxent or GARP and boosted regression trees	20%	
4	<b>Model selection</b>		
	Jackknife	Commonly used with maxent algorithm	38%
	Best Subset	Commonly used with GARP algorithm	23%
	Other	Chi-squared, bootstrap, etc. used with statistical algorithms	39%
5	<b>Model validation</b>	Area under the curve (AUC-ROC)	56%
		Partial area under the curve (pAUC)	13%
		Other	31%
6	<b>Uncertainty estimation</b>	Sensitivity analysis	7%

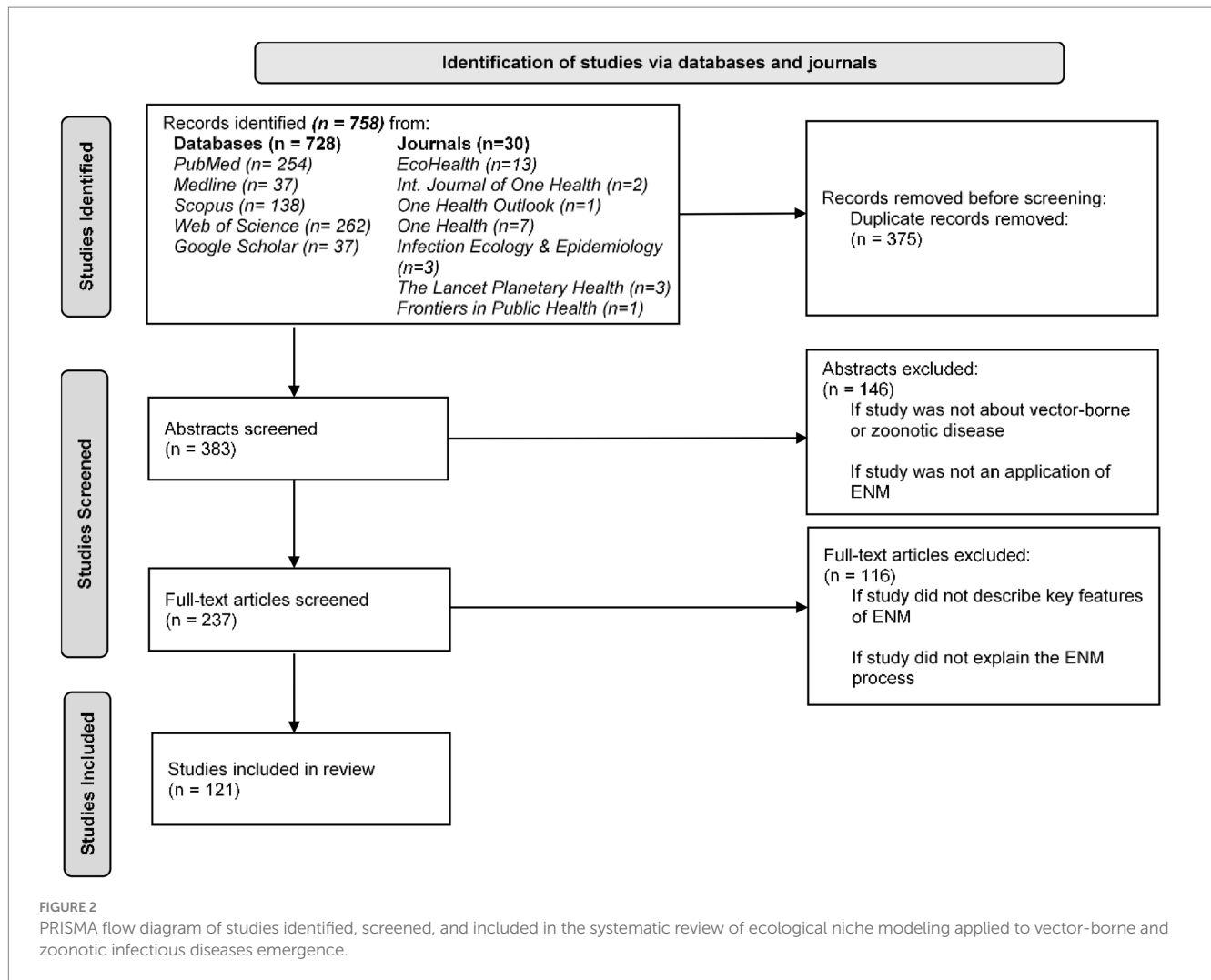
publications over time, journal field, geographic distribution of study location by country, and disease characteristics and emergence factors with the studies. Next, we geolocated and mapped the studies by country and according to each study's selection of social and environmental predictor variables, particularly climate zones, land cover type, and land-use biomes. We then assessed the studies according to integration of biotic, abiotic, and mobility factors. In particular, we evaluated the effectiveness of the studies to represent ecological niches based on the BAM framework (13, 21–23, 31). The evaluation comprises the identification of biotic, abiotic, and mobility factors included in the study, and specifically the combination of variables included (e.g., abiotic and mobility or biotic and mobility), which may represent potential versus actual ecological niches and sink or dead-end host populations versus source host populations, and the extent to which the geographic regions of the biotic, abiotic, and mobility factors overlap. Additionally, the ENM features were extracted from the articles and collated according to the criteria shown in Appendix 3. We divided the methodological framework into six steps according to methodological approach (presence-absence, presence-background, and presence-only), data selection, choice and application of algorithm, model selection, model validation, and uncertainty estimation. We describe the ENM features according to the percent of articles that exhibit specific features, number of articles with the most common features, and a description of the most common features included and excluded. We integrated the collated information into a stepwise ENM process and provide a descriptive

summary of the essential details that need to be reported in published articles that apply ENM to EIDs.

### 3. Results

The number of studies that applied ENM to predict vector-borne and zoonotic diseases ( $n = 121$ ) has steadily increased primarily over the last decade of this review, with only 2 or 3 studies per year in each year 2000–2009 (Figure 3A). Most studies were conducted in the fields of epidemiology ( $n = 31$ ), biomedicine ( $n = 29$ ), and public health ( $n = 25$ ), whereas less than half as many were conducted in the fields of ecology and environment ( $n = 21$ ) and veterinary medicine ( $n = 15$ ; Figure 3B). The application of ENM to EIDs has been somewhat unevenly distributed geographically across North America ( $n = 27$ ), South America ( $n = 29$ ), Africa ( $n = 22$ ), Asia ( $n = 22$ ), Australia ( $n = 3$ ), Europe ( $n = 10$ ), as shown in Figure 4. Calculating the number of studies relative to the area of the regions showed that North America, South America, Africa, and Europe had 2–4 times more studies per million km<sup>2</sup> than Australia and Asia. A majority of the studies (62%) were conducted in developing countries, while far fewer studies (33%) were conducted in developed countries (Country status was based on the UN World Economic Situation and Prospects Report, 2021). In terms of climate zones, the largest percent of studies (33%) were located in the tropical climate zone compared to 22% in the subtropics, 21% in warm temperate, 19% in cool temperate, and



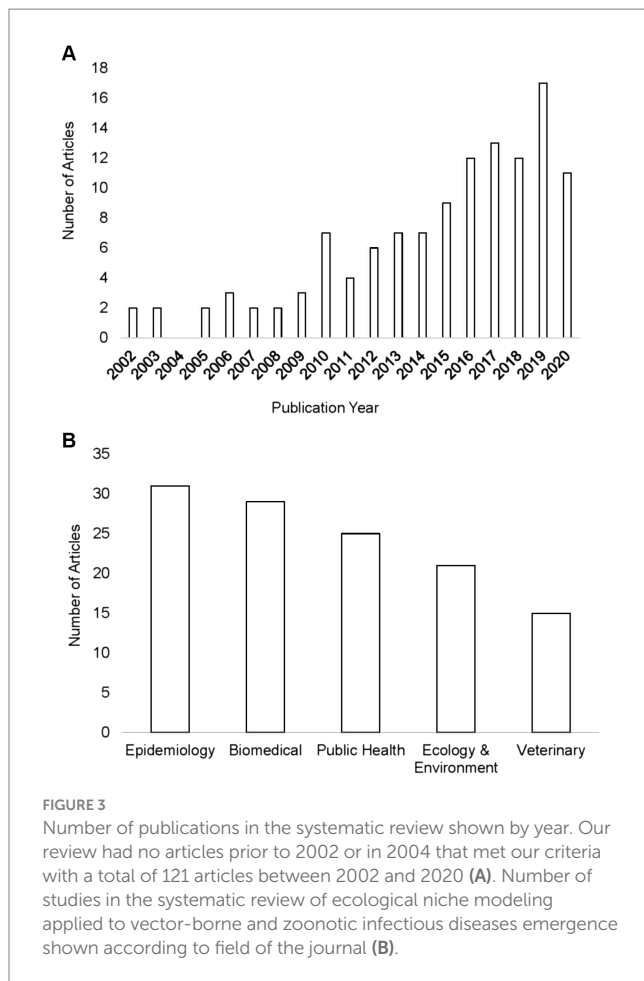


4% in polar climate zones ( $n = 108$ ; [Figure 5](#)). A similar distribution exists across global land cover regions, but there were far fewer studies that consider the biotic conditions, such as vegetation or land cover ( $n = 65$ ; [Figure 6](#)).

Very few studies accounted for the anthropogenic factors (e.g., land use and resource overexploitation) that modify the spatial distribution of EIDs ( $n = 14$ ; [Figure 7](#)). Of the studies that included social modify factors, all included human population data and to a lesser extent (roughly 1/3) included land-use data. The studies primarily used abiotic factors (e.g., physical environment; 90%;  $n = 109$ ) and mobility (e.g., species' dispersal ability; 99%;  $n = 120$ ) in the application of ENM to predict vector-borne and zoonotic diseases ([Figure 8](#)). Only 54% of studies included biotic conditions and interactions ( $n = 65$ ). Additionally, few studies (7%,  $n = 9$ ) included only biotic and mobility factors, while 45% of studies modeled only abiotic and mobility factors ( $n = 54$ ) and similarly (45%;  $n = 55$ ) studies included all (biotic, abiotic, mobility) factors consistent with the BAM framework. The BAM framework was mostly applied to mosquitoes followed by sandflies and livestock as vectors or reservoirs, and Malaria, Leishmaniasis, West Nile virus, and Dengue as diseases, as shown in [Figure 8](#).

A large majority of studies involve a presence-background approach ( $n = 102$ ) with few studies using presence-only ( $n = 17$ ) and

presence-absence ( $n = 2$ ) approaches, as shown in [Table 1](#). Most studies focus on environmental predictors rather than modifying social factors. Specifically, climate variables are most often used (89%;  $n = 108$ ) as predictors of disease emergence followed by land cover (43%;  $n = 52$ ), vegetation cover (36%;  $n = 44$ ), and soil characteristics (21%;  $n = 25$ ). However, host species density alone was included in nearly all studies (96%;  $n = 116$ ). Few studies (13%;  $n = 16$ ) include human social variables as modifying factors in ENM to predict vector-borne and zoonotic diseases. The studies often utilized machine learning (88%;  $n = 106$ ). Particularly, the MaxEnt algorithm was used in 64% of studies ( $n = 77$ ). The GARP algorithm (Genetic Algorithm for Rule-set Production) was applied in 25% of the studies ( $n = 30$ ). Most studies (74%;  $n = 90$ ) reported model selection procedures. The Jackknife procedure, where each variable is excluded in turn to test model performance and a model is created with the remaining variables, was commonly used with MaxEnt. The Best Subset procedure (all possible models are compared using a specified data-driven set of predictors) was commonly used with GARP for model selection. Most studies (90%;  $n = 109$ ) reported model validation procedures with most studies applying the Area Under the Curve (AUC) of the Receiver Operator Characteristic (ROC), which provides an aggregate measure of performance across all possible classification thresholds. The next most common validation procedures were a



binomial probability approach and the true skill statistic. Only 7% of studies ( $n = 8$ ) reported uncertainty analysis.

## 4. Discussion

### 4.1. Growth of ENM applied to EIDs and the distribution of studies

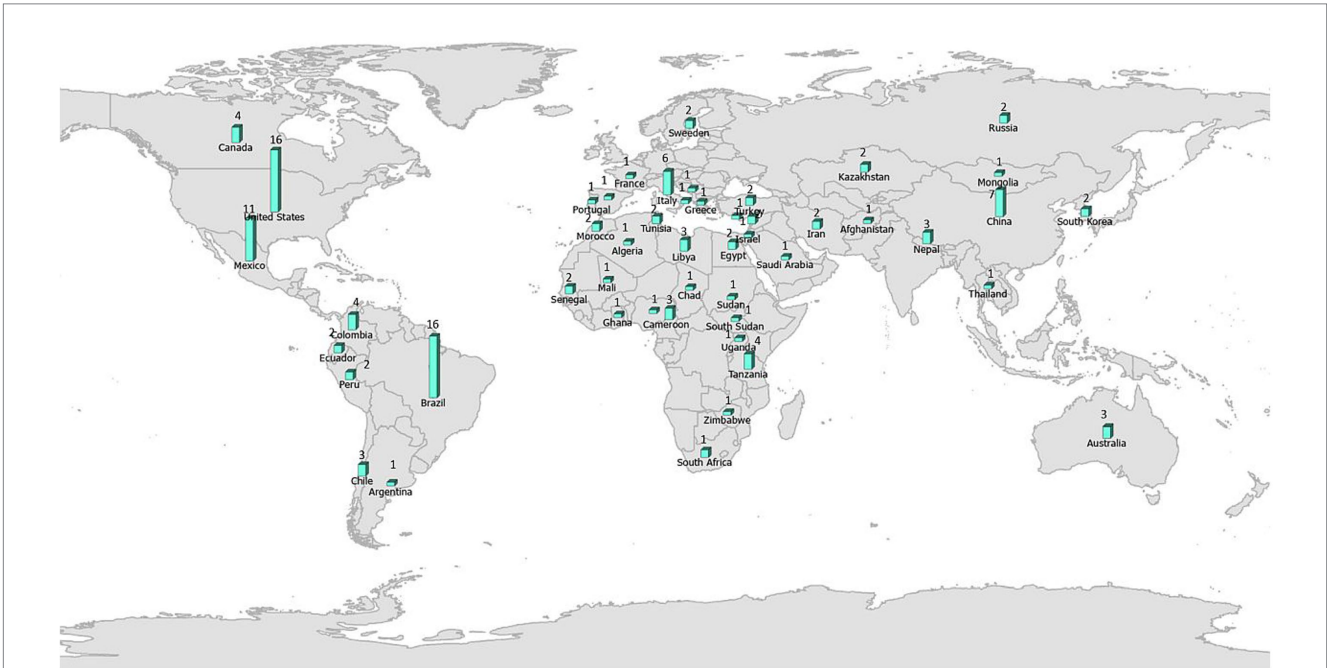
The uneven distribution of ENM to predict vector-borne and zoonotic diseases across journal fields in the past 20 years indicates that an increased engagement with the environmental and veterinary sciences is needed. Such engagement can result in greater insights into data selection, such as biotic factors contributing to emergence (26). Further, there was uneven geographic distribution of studies throughout the world. Most studies were in developing countries (62%), which is likely because most studied diseases are found in higher rates and with more negative impacts in developing regions. Also, most studies (33%) were within tropical climate zones where many infectious diseases emerge, which reflects the need to conduct studies in locations that are most vulnerable to emergence, and which likely comprises less-developed tropical countries with fewer resources and health infrastructure to respond to outbreaks. Nevertheless, continuing the growth and utility of spatial data and analyses will enable better predictions of EIDs and much-needed interventions earlier.

### 4.2. Likely effectiveness of reviewed studies to represent ecological niches

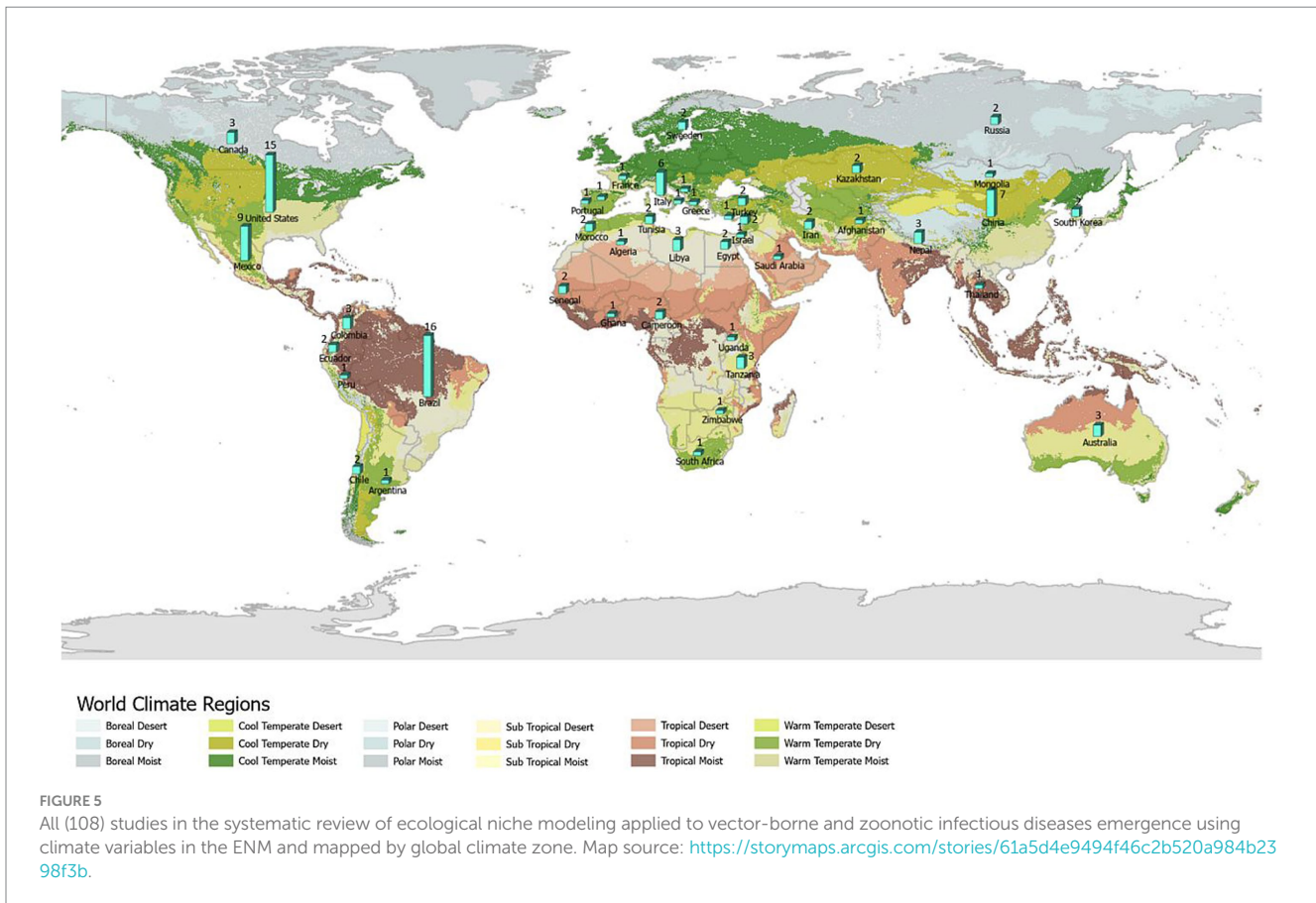
Studies that focus solely on abiotic and mobility factors (45%) comprise regions in geographic space where conditions allow growth rates of host-species populations to be positive, but the physiological conditions limit the ability of a host species to persist in that area (13, 32). As a result, such studies likely have investigated host-species populations where the infectious disease was (at least temporarily) not found or was potentially transient in dead-end hosts, which limit the persistence of an infectious disease (22). Essentially, the combination of only abiotic and mobility factors can represent a sink of host-species populations where host species are unable to persist (13). Similarly, the studies that focused solely on biotic and mobility factors (7%) also represent potential sink populations where an infectious disease is (at least temporarily) not found or potentially transient in dead-end hosts (22). For example, predation and competition between species may restrict a host-species distributions and the infectious disease may predispose the host species to increased predation. Stable host-species populations will be found primarily in regions where the ENM includes biotic, abiotic and mobility factors (13). Such a region has the appropriate set of biotic and abiotic factors for the host species to be present and persist, is accessible to the host species, and is equivalent to the geographic distribution of the host species (32).

### 4.3. Potential pitfalls of ENM techniques in reviewed studies

The popularity of the MaxEnt (Maximum Entropy) algorithm and the Genetic Algorithm for Rule-set Prediction (GARP) is apparent in our systematic review. Despite the heavy reliance on machine learning techniques, other statistical algorithms were used in a few studies of our review to predict EIDs, such as generalized linear models and boosted regression trees. Notably, boosted regression trees were used in conjunction with MaxEnt or GARP in a sizable number (20%) of studies in our review. Still, the application of the MaxEnt or GARP algorithms can present different challenges (21, 33, 34). In particular, the predictive performance may vary depending on the modeled host species' ecological characteristics as the different properties of the data and/or of the species may affect the accuracy of predictive maps (35). When applying ENM to infectious diseases users should use multiple metrics to assess how the data and analyses influence predictive performance. Further, model selection procedures chosen in the studies were often associated with the algorithm applied to the data and differ in important ways. When applying ENM it is important to understand that the Jackknife test is primarily used in the model development process to assess the contribution of each variable to the model (36, 37). The jackknife test runs the model (1) once with all variables, (2) dropping out each variable in turn, and (3) with a single variable at a time (36). The importance of a variable is determined based on having a large training gain when the variable is used alone in the model and a subsequent decrease in training gain when removed from the model (38). Alternatively, the Best Subset procedure assess multiple possible models and it presents the best candidates, but it is up to the user to compare and choose one model (39). Further, the best model is not always apparent, and judgment is required.



**FIGURE 4**  
All (121) studies in the systematic review of ecological niche modeling applied to vector-borne and zoonotic infectious diseases emergence mapped by country. Map source: <https://www.arcgis.com/home/item.html?id=deb60dd7744048cd9ba4fe203881fd12>.

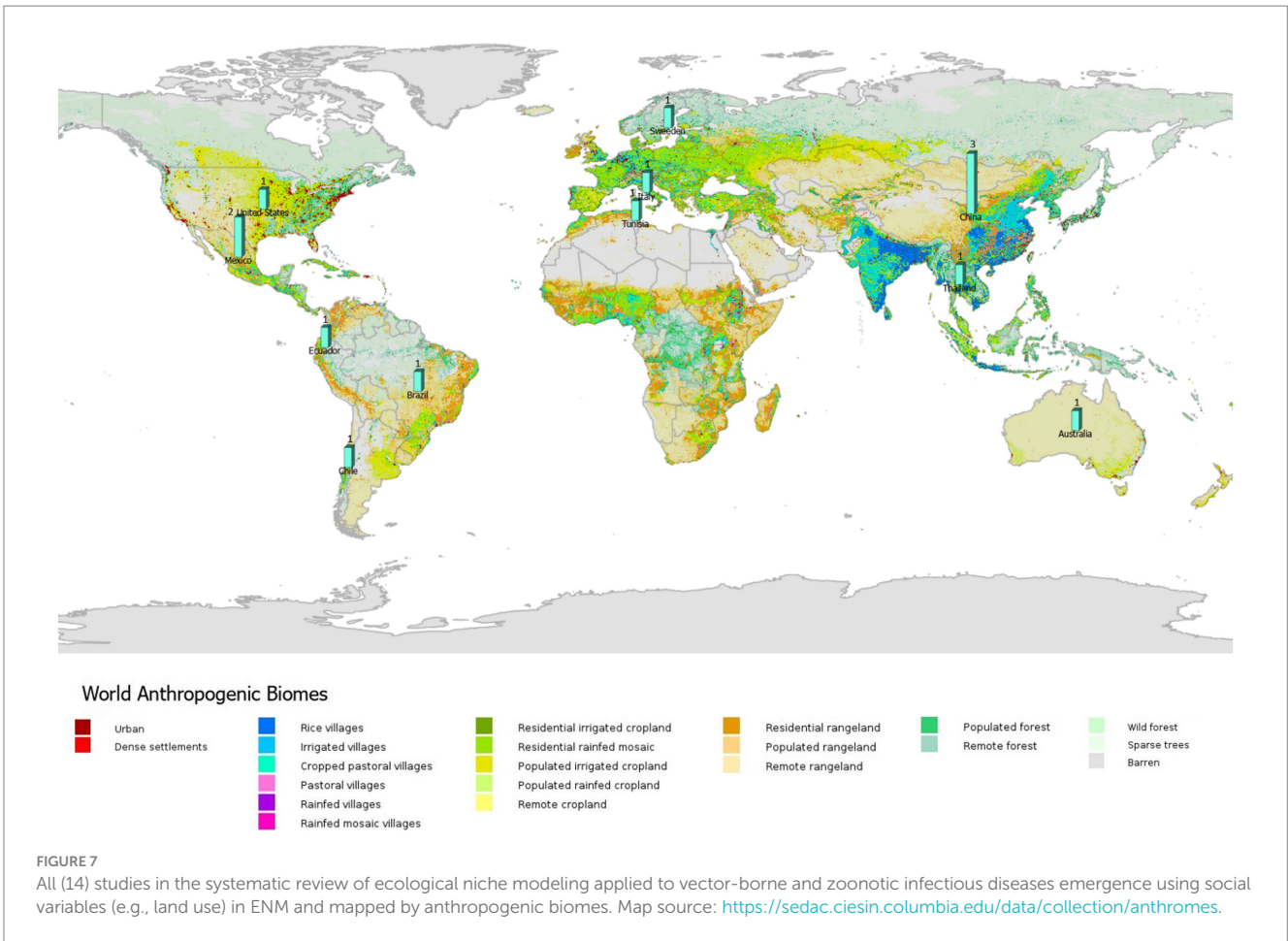
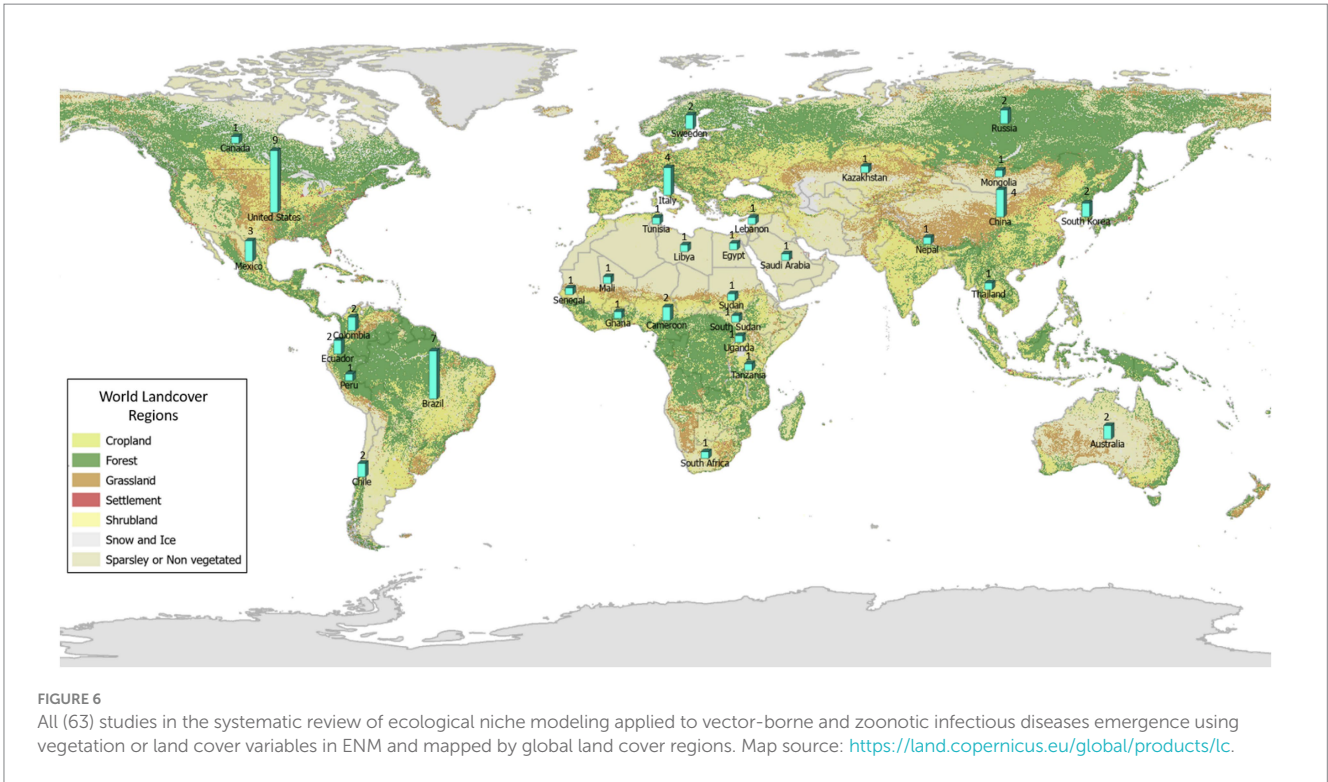


**FIGURE 5**  
All (108) studies in the systematic review of ecological niche modeling applied to vector-borne and zoonotic infectious diseases emergence using climate variables in the ENM and mapped by global climate zone. Map source: <https://storymaps.arcgis.com/stories/61a5d4e9494f46c2b520a984b2398f3b>.

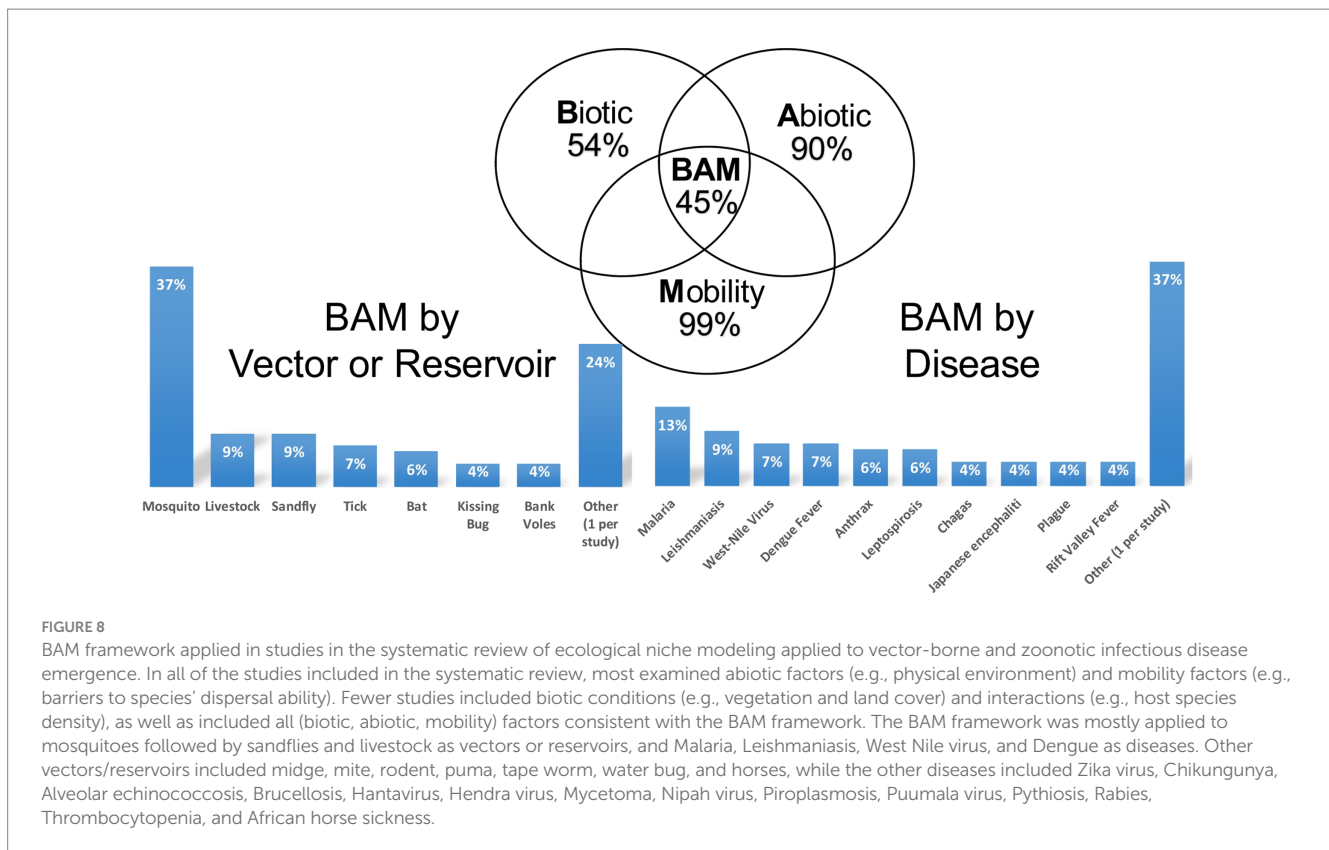
Model performance evaluated in studies that were included in our review primarily used the area under the curve (AUC) of the receiver operating characteristic (ROC). However, using such an approach has

been criticized (40, 41). Particularly, Lobo et al. (42) recommend not using AUC for five reasons: (1) it ignores the predicted probability values and the goodness-of-fit of the model; (2) it summarizes the test









performance over regions of the ROC space in which one would rarely operate; (3) it weights omission and commission errors equally; (4) it does not give information about the spatial distribution of model errors; and, most importantly, (5) the total extent to which models are carried out highly influences the rate of well-predicted absences and the AUC scores. To account for the potential limitations of the AUC in 15 of the reviewed articles published between 2015 and 2020, the overall accuracy of the models was assessed using the partial AUC (pAUC) procedure (43), which allows the user to set bounds on the types of predictions that are to be considered when certain portions of that space are not directly relevant to applications of interest. Overall, pAUC seems recently considered as more practically relevant than AUC (44).

Few studies conducted uncertainty analysis. When ENMs are transferred in space and time, it is important to understand the sources and location of uncertainty in their predictions (45). Uncertainties associated with ENMs have increased partially because of the increased number of environmental datasets available (46). Measuring uncertainty is crucial when applying ENM to predict EIDs as uncertainty can arise from several sources (45). Some studies in our review did test whether uncertainties were related to the choice of methods. Still, more uncertainty analyses are needed in future studies.

#### 4.4. Recommendations for future studies

Despite the importance of anthropogenic pressures to EIDs few studies (roughly 10%) that we assessed included such variables in ENM. Further, all of the studies primarily used human population growth or density as a predictor that modifies the emergence and spread of infectious diseases. Far more attention must be given to a

wide range of human social factors that modify the behavior and presence of a host species in future ENM studies. Land use is another important variable that was used in only a few of those small number of studies that included social (modifying) factors. Such factors that modify or influence the behavior and presence of a host species is particularly important to include in ENM when modeling domesticated vs. wild host species, such as livestock. Additionally, other modeling approaches, such as agent-based modeling and metapopulation models are used in epidemiology but overlooked in ENM and could be utilized to contribute to the inclusion of social factors. Agent-based models require detailed data entry for each individual in the population of interest resulting in a comprehensive description of the human behavior linked to an epidemic, while structured metapopulation models take into account geographic area census data and interpopulation mobility patterns (47). Overall, the lack of social factors (e.g., land use, land tenure, human population density) that can modify disease patterns reflects an overly simplistic view of pathogens, and disconnects pathogens from social and ecological contexts, whereas a holistic perspective that incorporates social as well as physical, chemical, and biological dimensions of our planet's systems is more realistic (7).

Biotic conditions and interactions are closely linked to landscape structure (48). Specifically, biotic factors, such as species richness, distribution, population density and the structure of host communities and reservoirs are essential in the transmission of infectious or parasitic agents (49, 50). Population and community structure are further associated with habitat variables, such as the amount of habitat available for species in a single land cover patch or the total amount of habitat available for species across all the different land cover patches, which can influence the type, frequency, and intensity of biotic interactions (48). Generally, habitat is delineated

and included in ENM through vegetation and land-cover imagery (51), which were applied in most (95%) of studies that modeled all BAM components. In turn, changes to habitat structure (e.g., fragmentation) affects the prevalence of infectious diseases (50). However, our systematic review revealed a lack of studies overall that include biotic factors in ENM. The lack of inclusion may be because biotic conditions and interactions are hypothesized to affect distributions only locally, but there is growing evidence that biotic interactions may have a larger role in shaping broad-scale distributions (22). Thus, more effort is needed to test and include biotic variables in ENM, such as the number of different habitat types, the prevalence of edge habitat, and the configuration of habitat. Habitat edges in particular can influence biotic interactions through forming a barrier to species movement and dispersal, may have differential influences on species' populations and change the intensity of outcomes of biotic interactions, and the proximity to edge and the associated influence of surrounding matrix can also affect the outcome of biotic interactions (48).

Future studies should consider the full BAM approach in developing ENM of EIDS. Furthermore, the spatial overlap of the BAM components must be carefully considered as an ENM study can result in estimating the potential rather than the actual host-species niche (13). Our review indicates most studies tend to focus on abiotic conditions and mobility factors. Even when biotic factors are included together with abiotic and mobility factors, the ENM can still represent the geographic space of primarily the abiotic and mobility factors. In such a case, the biotic interactions restrict the presence of host-species populations to a small spatial extent of geographic space of the abiotic and mobility factors (21). Additionally, the ENM can represent the geographic space of primarily the biotic and abiotic factors. Even if regions of biotic and abiotic largely overlap, the interactions can be weak, diffuse and non-specific if the accessible region due to mobility factors overlap with the regions of biotic and abiotic in only a relatively small region (13). In such a situation the potential niche of the host-species population will generally be larger than the actual distribution of the host-species population's actual niche. Overall, unless the three sets of factors affecting geographic distributions of species overlap almost entirely, most ENM algorithms estimate the potential rather than the actual niche. These considerations are foremost if ENMs are to have biological meaning and reality (52).

## 5. Conclusion

The use of ENM to predict EIDs is growing, and our review can help guide this approach. We clarified the socio-environmental factors that are commonly implicated as drivers of vector-borne and zoonotic diseases, social and environmental settings of study locations, the variables modeled, and the principal elements of ENM and the modeling process to predict EIDs. Rigorous modeling of underlying drivers of EIDs can better inform epidemiological and One Health professionals to help prevent outbreaks and future epidemics.

## Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

## Author contributions

TL conceptualized the systematic review, was the third reviewer to settle disagreements on articles through majority voting during the abstract and full-text screening, conducted the data analyses, verified the underlying data, and was the primary author of the article. BT conducted article searches in the research databases, abstract and full-text screening, verified the underlying data, and managed the systematic review database and workflow processes. AG conducted article searches in the research databases, abstract, and full-text screening. DT provided guidance and validation of the systematic review methods. SD, EF, IG, VS, and ES had full access to the data, contributed equally to the interpretation of data, and critically revising the manuscript for important intellectual content. All authors contributed to the article and approved the submitted version.

## Funding

This work was supported in part by the Taylor Geospatial Institute and a seed research grant from the Living Earth Collaborative at Washington University in St. Louis. Furthermore, this research was also supported in part by the United Kingdom Biotechnology and Biological Sciences Research Council, the Department for International Development, the Economic and Social Research Council, the Medical Research Council, the Natural Environment Research Council, and the Defense Science and Technology Laboratory, under the Zoonoses and Emerging Livestock Systems (ZELS) programme, grant reference BB/L019019/1. This study also received support from the CGIAR One Health initiative "Protecting Human Health Through a One Health Approach," which was supported by contributors to the CGIAR Trust Fund (<https://www.cgiar.org/funders/>).

## Acknowledgments

We thank the University of Liverpool's Open Access team for support of the CC-BY open access license for this article.

## 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.

## Supplementary material

The Supplementary material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpubh.2023.1244084/full#supplementary-material>

## References

- Jones KE, Patel NG, Levy MA, Storeygard A, Balk D, Gittleman JL, et al. Global trends in emerging infectious diseases. *Nature*. (2008) 451:990–3. doi: 10.1038/nature06536
- Smith KF, Goldberg M, Rosenthal S, Carlson L, Chen J, Chen C, et al. Global rise in human infectious disease outbreaks. *J R Soc Interface*. (2014) 11:20140950. doi: 10.1098/rsif.2014.0950
- Ogden N, AbdelMalik P, Pulliam J. Emerging infectious diseases: prediction and detection. *Can Commun Dis Rep*. (2017) 43:206–11. doi: 10.14745/ccdr.v43i10a03
- Olson SH, Benedum CM, Mekaru SR, Preston ND, Mazet JAK, Joly DO, et al. Drivers of emerging infectious disease events as a framework for digital detection. *Emerg Infect Dis*. (2015) 21:1285–92. doi: 10.3201/eid2108.141156
- Morens DM, Fauci AS. Emerging infectious diseases: threats to human health and global stability. *PLoS Pathog*. (2013) 9:e1003467. doi: 10.1371/journal.ppat.1003467
- Aguirre AA, Tabor GM. Global factors driving emerging infectious diseases. *Ann N Y Acad Sci*. (2008) 1149:1–3. doi: 10.1196/annals.1428.052
- Wilcox BA, Colwell RR. Emerging and reemerging infectious diseases: biocomplexity as an interdisciplinary paradigm. *EcoHealth*. (2005) 2:244. doi: 10.1007/s10393-005-8961-3
- Russell RE, Katz RA, Richgels KLD, Walsh DP, Grant EHC. A framework for Modeling emerging diseases to inform management. *Emerg Infect Dis*. (2017) 23:1–6. doi: 10.3201/eid2301.161452
- Arthur RF, Gurley ES, Salje H, Bloomfield LSP, Jones JH. Contact structure, mobility, environmental impact and behaviour: the importance of social forces to infectious disease dynamics and disease ecology. *Philos Trans R Soc B Biol Sci*. (2017) 372:20160454. doi: 10.1098/rstb.2016.0454
- Metcalf CJE, Lessler J. Opportunities and challenges in modeling emerging infectious diseases. *Science*. (2017) 357:149–52. doi: 10.1126/science.aam8335
- Escobar LE. Ecological niche Modeling: an introduction for veterinarians and epidemiologists. *Front Vet Sci*. (2020) 7:519059. doi: 10.3389/fvets.2020.519059
- Escobar LE, Craft ME. Advances and limitations of disease biogeography using ecological niche modeling. *Front Microbiol*. (2016) 7:1174. doi: 10.3389/fmicb.2016.01174
- Soberon J, Peterson AT. Interpretation of models of fundamental ecological niches and species' distributional areas. *Biodivers Inform*. (2005) 2:2. doi: 10.17161/bi.v2i0.4
- Ashcroft MB, King DH, Raymond B, Turnbull JD, Wasley J, Robinson SA. Moving beyond presence and absence when examining changes in species distributions. *Glob Change Biol*. (2017) 23:2929–40. doi: 10.1111/gcb.13628
- Ward G, Hastie T, Barry S, Elith J, Leathwick JR. Presence-only data and the EM algorithm. *Biometrics*. (2009) 65:554–63. doi: 10.1111/j.1541-0420.2008.01116.x
- Hazen EL, Abrahms B, Brodie S, Carroll G, Welch H, Bograd SJ. Where did they not go? Considerations for generating pseudo-absences for telemetry-based habitat models. *Mov Ecol*. (2021) 9:5. doi: 10.1186/s40462-021-00240-2
- Phillips SJ, Dudík M, Elith J, Graham CH, Lehmann A, Leathwick J, et al. Sample selection bias and presence-only distribution models: implications for background and pseudo-absence data. *Ecol Appl*. (2009) 19:181–97. doi: 10.1890/07-2153.1
- Warton D, Aarts G. Advancing our thinking in presence-only and used-available analysis. *J Anim Ecol*. (2013) 82:1125–34. doi: 10.1111/1365-2656.12071
- Valavi R, Guillera-Arroita G, Lahoz-Monfort JJ, Elith J. Predictive performance of presence-only species distribution models: a benchmark study with reproducible code. *Ecol Monogr*. (2022) 92:e01486. doi: 10.1002/ecm.1486
- Royle JA, Chandler RB, Yackulic C, Nichols JD. Likelihood analysis of species occurrence probability from presence-only data for modelling species distributions. *Methods Ecol Evol*. (2012) 3:545–54. doi: 10.1111/j.2041-210X.2011.00182.x
- Peterson AT. Ecologic niche Modeling and spatial patterns of disease transmission. *Emerg Infect Dis*. (2006) 12:1822–6. doi: 10.3201/eid1212.060373
- Johnson EE, Escobar LE, Zambrana-Torrelío C. An ecological framework for Modeling the geography of disease transmission. *Trends Ecol Evol*. (2019) 34:655–68. doi: 10.1016/j.tree.2019.03.004
- Sillero N, Arenas-Castro S, Enriquez-Urzelai U, Vale CG, Sousa-Guedes D, Martínez-Freiria F, et al. Want to model a species niche? A step-by-step guideline on correlative ecological niche modelling. *Ecol Model*. (2021) 456:109671. doi: 10.1016/j.ecolmodel.2021.109671
- Haider N, Rothman-Ostrow P, Osman AY, Arruda LB, Macfarlane-Berry L, Elton L, et al. COVID-19—zoonosis or emerging infectious disease? *Front Public Health*. (2020) 8:596944. doi: 10.3389/fpubh.2020.596944
- McArthur DB. Emerging infectious diseases. *Nurs Clin North Am*. (2019) 54:297–311. doi: 10.1016/j.cnur.2019.02.006
- Mackenzie JS, Jeggo M. The one health approach—why is it so important? *Trop Med Infect Dis*. (2019) 4:88. doi: 10.3390/tropicalmed4020088
- Moher D, Shamseer L, Clarke M, Ghersi D, Liberati A, Petticrew M, et al. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Syst Rev*. (2015) 4:1. doi: 10.1186/2046-4053-4-1
- Trinosey J, Brahma FA, Gall C. Zotero: a product review. *J Electron Resour Med Libr*. (2009) 6:224–9. doi: 10.1080/15424060903167229
- Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ*. (2021) 372:n71. doi: 10.1136/bmj.n71
- Robinson NM, Nelson WA, Costello MJ, Sutherland JE, Lundquist CJ. A systematic review of marine-based species distribution models (SDMs) with recommendations for best practice. *Front Mar Sci*. (2017) 4:421. doi: 10.3389/fmars.2017.00421
- Peterson AT, Soberon J. Species distribution Modeling and ecological niche Modeling: getting the concepts right. *Nat Conserv*. (2012) 10:102–7. doi: 10.4322/natcon.2012.019
- Peterson AT, Soberon J, Pearson RG, Anderson RP, Martínez-Meyer E, Nakamura M, et al. *Ecological Niches and Geographic Distributions (MPB-49)* Princeton University Press (2011).
- Merow C, Smith MJ, Silander JA Jr. A practical guide to MaxEnt for modeling species' distributions: what it does, and why inputs and settings matter. *Ecography*. (2013) 36:1058–69. doi: 10.1111/j.1600-0587.2013.07872.x
- Stockman AK, Beamer DA, Bond JE. An evaluation of a GARP model as an approach to predicting the spatial distribution of non-vagile invertebrate species. *Divers Distrib*. (2006) 12:81–9. doi: 10.1111/j.1366-9516.2006.00225.x
- Chikerema SM, Gwitira I, Murwira A, Pfukenyi DM, Matope G. Comparison of GARP and Maxent in modelling the geographic distribution of *Bacillus anthracis* in Zimbabwe. *Zimb Vet J*. (2017) 35:7.
- Miller RH, Masuoka P, Klein TA, Kim HC, Somer T, Grieco J. Ecological niche Modeling to estimate the distribution of Japanese Encephalitis virus in Asia. *PLoS Negl Trop Dis*. (2012) 6:e1678. doi: 10.1371/journal.pntd.0001678
- Phillips SJ, Anderson RP, Schapire RE. Maximum entropy modeling of species geographic distributions. *Ecol Model*. (2006) 190:231–59. doi: 10.1016/j.ecolmodel.2005.03.026
- Masuoka P, Klein TA, Kim HC, Claborn DM, Achee N, Andre R, et al. "Modeling and Analysis of Mosquito and Environmental Data to Predict the Risk of Japanese Encephalitis," in *ASPRS Annual Conference, Baltimore, Maryland, USA*. (2009).
- Haase CG, Yang A, McNyset KM, Blackburn JK. GARPTools: R software for data preparation and model evaluation of GARP models. *Ecography*. (2021) 44:1790–6. doi: 10.1111/ecog.05642
- Mughini-Gras L, Mulatti P, Severini F, Boccolini D, Romi R, Bongiorno G, et al. Ecological niche modelling of potential West Nile virus vector Mosquito species and their geographical association with equine epizootics in Italy. *EcoHealth*. (2014) 11:120–32. doi: 10.1007/s10393-013-0878-7
- Fan J, Yuan X, Wang M, Zhu G. Mapping the potential distribution of the schistosomiasis intermediate host *Biomphalaria straminea* in China. *Geospat Health*. (2018) 13:13. doi: 10.4081/gh.2018.723
- Lobo JM, Jiménez-Valverde A, Real R. AUC: a misleading measure of the performance of predictive distribution models. *Glob Ecol Biogeogr*. (2008) 17:145–51. doi: 10.1111/j.1466-8238.2007.00358.x
- Peterson AT, Papeş M, Soberon J. Rethinking receiver operating characteristic analysis applications in ecological niche modeling. *Ecol Model*. (2008) 213:63–72. doi: 10.1016/j.ecolmodel.2007.11.008
- Ma H, Bando AI, Rockette HE, Gur D. On use of partial area under the ROC curve for evaluation of diagnostic performance. *Stat Med*. (2013) 32:3449–58. doi: 10.1002/sim.5777
- Alkhishe A, Cobos ME, Peterson AT, Samy AM. Recognizing sources of uncertainty in disease vector ecological niche models: an example with the tick *Rhipicephalus sanguineus* sensu lato. *Perspect Ecol Conserv*. (2020) 18:91–102. doi: 10.1016/j.pcon.2020.03.002
- Parreira MR, Nabout JC, Tessarolo G, de Souza L-RM, Teresa FB. Disentangling uncertainties from niche modeling in freshwater ecosystems. *Ecol Model*. (2019) 391:1–8. doi: 10.1016/j.ecolmodel.2018.10.024
- Christaki E. New technologies in predicting, preventing and controlling emerging infectious diseases. *Virulence*. (2015) 6:558–65. doi: 10.1080/21505594.2015.1040975
- Zarnetske PL, Baiser B, Strecker A, Record S, Belmaker J, Tuanmu MN. The interplay between landscape structure and biotic interactions. *Curr Landsc Ecol Rep*. (2017) 2:12–29. doi: 10.1007/s40823-017-0021-5
- Bommarito C, Wahl M, Thielges DW, Pansch C, Zucchetto M, Pranovi F. Biotic and abiotic drivers affect parasite richness, prevalence and abundance in *Mytilus galloprovincialis* along the northern Adriatic Sea. *Parasitology*. (2022) 149:15–23. doi: 10.1017/S0031182021001438
- Morand S, Lajaunie C. Outbreaks of vector-borne and zoonotic diseases are associated with changes in Forest cover and oil palm expansion at global scale. *Front Vet Sci*. (2021) 8:8. doi: 10.3389/fvets.2021.661063
- Stuber EF, Gruber LE, Fontaine JJ. A Bayesian method for assessing multi-scale species-habitat relationships. *Landsc Ecol*. (2017) 32:2365–81. doi: 10.1007/s10980-017-0575-y
- Peterson AT, Anamza T. Ecological niches and present and historical geographic distributions of species: a 15-year review of frameworks, results, pitfalls, and promises. *Folia Zool*. (2015) 64:207–17. doi: 10.25225/fozo.v64.i3.a3.2015