Biodiversity Informatics, 19, 2025, pp. 61-84

ECOLOGICAL NICHE MODELING APPLICATIONS TO INFECTIOUS DISEASES

Shariful Islam[1,2], Mariana Castaneda-Guzman[1], Diego Soler-Tovar[3], and Luis E. Escobar[1,2,4*]

[1]Department of Fish and Wildlife Conservation, Virginia Tech, Blacksburg, VA, USA
[2]Global Change Center, Virginia Tech, Blacksburg, VA, USA
[3]Faculty of Agricultural Sciences, Universidad de La Salle, Bogotá, Colombia
[4]Center for Emerging Zoonotic and Arthropod-Borne Pathogens, Virginia Tech, Blacksburg, VA, USA

Abstract. Ecological niche modeling (ENM) is a widely used analytical approach for predicting species distributions and has been applied to study the spatial epidemiology of infectious diseases. Nevertheless, research evaluating the key components and assumptions of ENM in disease systems remains limited, raising concerns about its robustness, reproducibility, and transparency. To address this limitation, we conducted a systematic review and evaluated articles on ENM applications to infectious diseases between 2020 and 2022. We reviewed 78 articles to extract information following a standard protocol for reporting ENM analysis and summarized the information for each component (e.g., study subject, location, duration). The spatial extent of study areas varied from village to global scales, temporal duration ranged from 1 to 101 years, and the organismal levels ranged from individuals (57.7%) to populations (33.3%). Less frequently reported components included temporal autocorrelation tests (2.66%), algorithmic uncertainty (28.21%), temporal resolution (35.90%), background data selection (44.87%), coordinate reference system (41.02%), model performance from validation data (46.15%), and model averaging (20.51%). Our findings highlight a lack of consistency and transparency in disease ecology and disease biogeography studies, which may lead to misleading ENM applications in spatial epidemiology. Researchers and reviewers applying ENM to disease systems should clearly report key modeling components to ensure biologically sound outputs. This article identified trends and gaps in reporting ENM protocols for mapping disease transmission risk.

Keywords: Biogeography, Disease, Ecological Niche, Epidemiology, Reproducibility, Spatial

Introduction

Emerging infectious diseases are increasing in frequency and represent a major threat to global public health and global economies (Dobson et al. 2020). Human-animal interaction is considered an important driver of disease emergence along with climate change, biodiversity loss, and socio-economic conditions (Keesing and Ostfeld 2024). Nevertheless, our ability to understand and accurately forecast disease emergence and spread remains limited (Escobar and Craft 2016). One primary reason for inaccurate disease forecasting is the limited standards in data availability, model parameterization, and the inherent stochasticity (i.e., randomness or unpredictability) of disease transmission (Escobar 2020; Peterson 2014).

Disease ecology and biogeography can be used to understand why diseases emerge in some areas but not others (Peterson 2008). The "ecological niche", defined as the set of favorable conditions (biotic and abiotic) allowing an organism to persist and disperse in the long term, is a key concept in disease ecology and disease biogeography (Escobar 2020; Peterson 2014). The ecological niche of pathogens, vectors, hosts, or disease transmission events is complex considering host-pathogen relationships. In some cases, the distribution of a pathogen is shaped by the distribution of its host. Nevertheless, there is also evidence that some pathogens do not follow the niches of their hosts and instead exhibit their own distinct ecological niche (Maher et al. 2010; Astorga et al. 2018). To estimate or character-

---

* *Corresponding author: Address: 1015 Life Science Circle, Blacksburg, Virginia, 24061, United States. escobar1@vt.edu.*
ORCID: https://orcid.org/0000-0001-5735-2750

ize the ecological niche of an organism causing an infectious disease, researchers commonly use ecological niche modeling (ENM). For example, Romero-Alvarez et al. (2023) assessed the potential spread of *Oropouche virus* in humans, and Escobar et al. (2017) identified areas of transmission risk for *Heterosporis*, a fish disease caused by the microsporidian parasite *Heterosporis sutherlandae*. A complete ENM framework consists of five major sections (Overview/Conceptualization, Data, Model Fitting, Assessment, and Prediction), each containing components essential for building robust, transparent, and reproducible models (Zurell et al. 2020).

With increasing access to ENM training materials, tools, and data, the implementation of ENM in the study of diseases is becoming more common (Frans and Liu 2024). For instance, a comprehensive open-access online course facilitates the learning and development of ENM for biodiversity and disease (Peterson et al. 2019; Peterson and Ingenloff 2016). As the use of ENM in disease ecology and biogeography continues to expand (Escobar and Morand 2021), it is increasingly important to assess whether studies are adhering to best practices and reporting important modeling components transparently. Nevertheless, researchers may overlook key modeling components (e.g., study objects, location, and duration) and assumptions during ENM implementations (Escobar 2020), which can lead to an untrustworthy model and the misinterpretation of infectious diseases systems (Araújo et al. 2019; Peterson 2014; Soley-Guardia et al. 2024). For instance, a study conducted by Brito-Hoyos et al. (2013) aiming to map the risk of bat-borne rabies outbreaks in Colombia employed a climate-based ENM. The authors, however, skipped critical steps of ENM (Escobar and Peterson 2013). One of them was the use of all available occurrence points for model calibration without considering spatial clustering and sampling bias. Additionally, Brito-Hoyos et al. (2013) used 15 climatic variables for their model building without considering the multicollinearity among the variables and spatial lags. Later, Escobar and Peterson (2013) revised the Brito-Hoyos et al. (2013) predictions and found that key biogeographic principles were missing, resulting in a misinformation of rabies transmission risk in Colombia. Similarly, studies have identified artifacts in four bioclimatic variables (Bio8, Bio9, Bio18, and Bio19), evidenced by unusual spatial anomalies and inconsistencies among adjacent pixels (Booth 2022; Escobar et al. 2014). Therefore, it has been recommended to assess the discontinuities among the four problematic variables for the specific study areas before final inclusion in an ENM. Nevertheless, Olivera et al. (2021) still included all the available bioclimatic variables in their ENM without quality check, which may fail to identify actual suitable areas. In a systematic review, Feng et al. (2019a) found a lack of critical modeling information (e.g., source of occurrence data, spatial extent, source of environmental data) in ENM applications to biodiversity.

ENMs of infectious diseases is inherently more complex than traditional applications to biodiversity (Escobar and Craft 2016). Understanding disease ecology, particularly transmission dynamics involving hosts, vectors, and pathogens, is often complicated by small sample sizes, sampling biases, and limited geographic and environmental data (Escobar 2020). Moreover, failing to incorporate key biological components in ENM for disease systems may hinder the prediction of disease distribution and transmission risk (Escobar and Craft 2016; Peterson, 2006). Given the growing role of ENM in the understanding of the ecology of disease systems and informing spatial epidemiology, it is essential to evaluate how researchers incorporate fundamental components and assumptions to ensure robust and biologically meaningful model results. A standardized protocol may promote consistency and best practices in model development (Zurell et al. 2020).

Previous studies have shown that the transparency and methodological rigor of individual-based or agent-based models have improved following the introduction of the Overview, Design Concepts, and Details (ODD) protocol (Grimm et al. 2010). Similarly, the adoption of the Darwin Core Standard has had a positive influence on data sharing practices (Wieczorek et al. 2012). Considering the importance of having a standard checklist for ecological niche studies, and to support and guide key steps in ENM, Zurell et al. (2020) adapted the ODD protocol into the Overview/Conceptualization, Data, Model Fitting, Assessment, and Prediction (ODMAP) framework. The ODMAP protocol complements and integrates the Range Model Metadata Standard (RMMS) by providing a step-by-step guide along with a comprehensive checklist of components required to build an ecological niche model (Merow et al. 2019, Zurell et al. 2020; Fitzpatrick et al. 2021). The ODMAP protocol also includes a Shiny web application that offers guidance for developing standardized ENM documentation using the RMMS dictionary and allows users to download the ODMAP table (Zurell et al. 2020; Fitzpatrick et al. 2021). As an updated and adaptable metadata framework, ODMAP is well-suited for documenting ENM components and is designed to automatically accommodate future updates to the dictionary (Merow et al. 2019, Zurell et al. 2020; Fitzpatrick et al. 2021). Zurell et al. (2020) demonstrated the utility of their ODMAP pro-

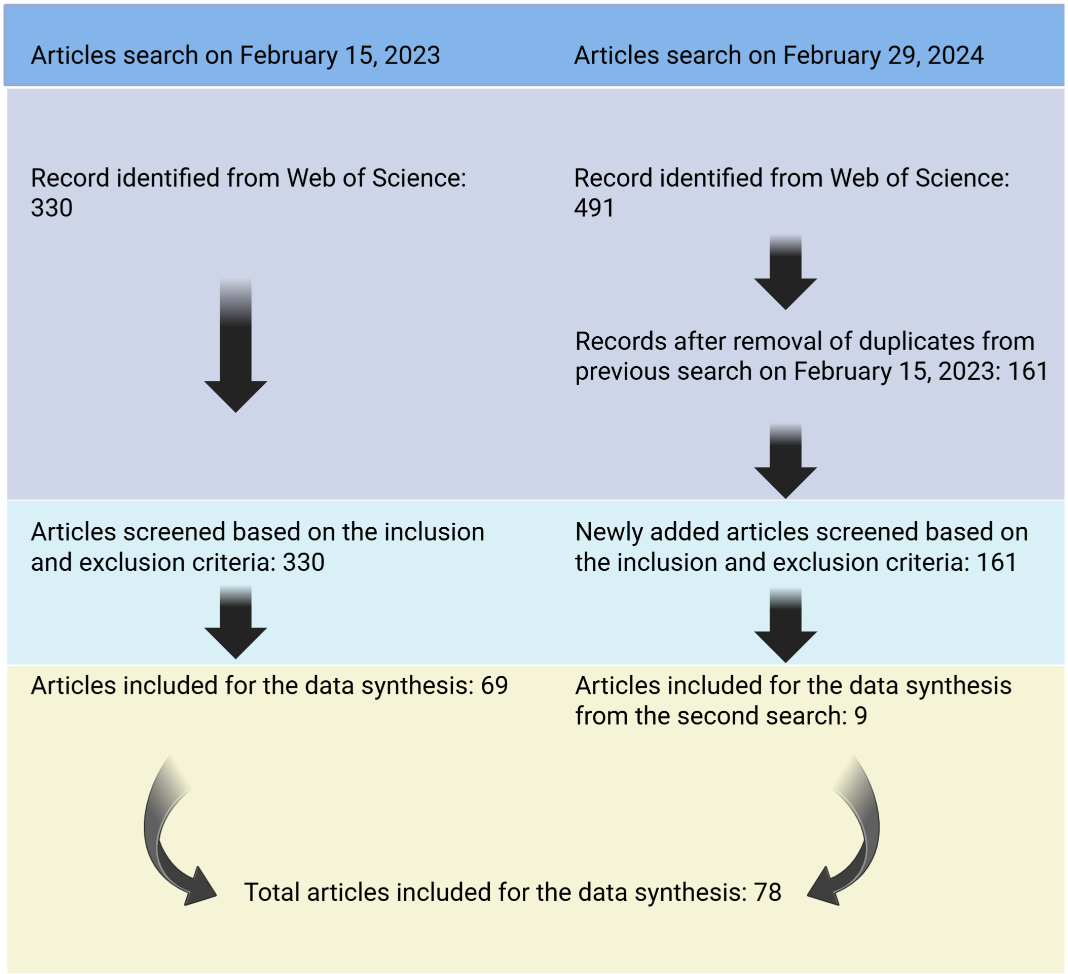


Figure 1: Flow diagram of systematic article search and selection for the data synthesis. From top to bottom, different background color indicates the article search on Web of Science, articles identification, article screening following inclusion and exclusion criteria, and final list of articles selected to include and data synthesis.

tocol through nine case studies. It has been expected that ODMAP will have a similar impact like ODD protocol in biodiversity informatics (Zurell et al. 2020; Fitzpatrick et al. 2021).

The ODMAP protocol is designed to assist beginners in ENM, provides a structured workflow for advanced modelers, and serves as an efficient tool for editors and reviewers to evaluate model-based studies (Zurell et a. 2020). The aim of this study was to assess recent ENM applications to disease systems using the ODMAP checklist.

To our knowledge, no studies have been conducted to evaluate the quality standards of the implementation of ENM in disease ecology and disease biogeography. Based on key ENM requirements, data limitations, and the complexities of disease transmission systems, we hypothesize that most articles fail to account for critical components during the modeling of infectious disease.

Materials And Methods

Search strategy

We conducted a systematic literature review on Web of Sciences following the PRISMA framework (Moher et al. 2009) (Figure 1, date 2/15/2023). We recovered articles published between 2020 and 2022 using the terms "ecological niche modeling" OR "species distribution modeling" AND "disease". We selected this period since it corresponds to the release of a disease-specific article published with recommendations on ENM applications to describe and forecast disease (Escobar 2020). The study period also provides an overview of the early 2020s as a proxy of contemporary methods, including the period of explosion of COVID-19 articles among disciplines (Riccaboni and Verginer 2022). To avoid missing any articles, we repeated the search following the same terms on February 29, 2024.

Table 1: **Inclusion criteria considered selecting and removing the articles during article selection step**.

Inclusion criteria	Exclusion criteria
Articles written in English	Articles written in a language other than English
Articles published between 2020 to 2022	Removed all the articles outside the study period
Research articles	Removed review articles, commentaries, rebuttals, opinion letters or editorial notes
Study used ENM to understand the ecology and distribution of a pathogen, or vector, or reservoir or definite host of an infectious disease	Study fails to define a component of the disease system, non-infectious disease (e.g., cancer, mental health)
Full text article is available	Full text article is not available

Study selection

All articles retrieved were downloaded and revised using detailed inclusion criteria (Table 1). We screened each article to assess its eligibility for this study, including language, publication period, research article, and application of ENM to understand the disease systems. Articles retained were read in full and data were extracted and standardized in terms of units to allow comparisons.

Data collection and preparation

Following Zurell et al. (2020), we assessed and extracted data from each article following five standard sections to gather data or model parameterization, including an overview, data, modeling, assessment, and prediction (ODMAP) (Supplementary Material). For each section model components or parameters are required for robust ENM calibration and validation. Besides, we collected the most recent impact factor available for each journal in which the selected articles were published, using information from the respective journal websites.

All data were entered and processed using MS Excel (Microsoft Office ProfessionalPlus 2021, Microsoft Corporation, Washington) and saved as .csv file. Pieces of information were continuous or binary (yes or no) for some parameters. If the answer was "yes," we recorded the corresponding details. For example, if the question was whether the author mentioned the statistical algorithm used for ENM building, and the answer was "yes," we noted the algorithm employed in the study. Similarly, if the question was whether the author mentioned the temporal duration of the study, and the answer was "yes," we recorded the duration as a quantitative value.

To standardize the geographical extent of the studies, we categorized them as regional if the study area included more than one country, or state-level if articles specified a sub-area within a country. Some articles reported spatial resolution in arcminutes. We converted the resolution from arcminutes to meters using conversion factors provided by the United States Geological Survey (USGS 2011).

After extracting information from all selected articles, we saved the dataset as a .csv file and imported it into R statistic software version 4.4.1 for the descriptive analysis (R Core Team 2024). We used the ggplot2 package for visualization (Wickham 2016). For components with discrete or categorical data (e.g. Yes/No), we used bar plot, while for continuing data, we used box plots. For categorical components with only one observation, we excluded that category from the bar plot. We used VOSviewer (www.vosviewer.com) version 1.6.20, to conduct keyword network analysis and visualization. We considered the minimum of 2 occurrences of keywords as a threshold for the keyword network in VOSviewer.

Results

We collected a total of 491 published articles between 2020 and 2022. After screening articles following inclusion and exclusion criteria, 78 articles were selected for review and evaluated for data collection. Results were categorized into five sections, including Overview, Data, Modelling, Assessment, and Prediction (ODMAP) following protocol by (Zurell et al. 2020). The journals in which the selected articles were published had impact factors ranging from 0.6 to 14.8 (n=75; mean=3.2) (Figure S1). Based on keywords network analysis, the application of ENM in the field of infectious disease research appears to be increasing. The larger node "ecological niche model" and "species distribution model" highlight its recurrent and central themes in the analysis of the reviewed articles (Figure 2).

Overview

For the overview section there were twenty-two components, only eight components were explained by all 78 articles (36.4%; n=22) (Figure 3). For instance, only a small subset of articles provided information about code availability (9.0%; n=7), spatial extent (i.e., longitude, latitude) (26.9%; n=21), temporal resolution (35.9%; n= 8), data availability (36.4%; n=44), and model assembling (64.1%; n=50).

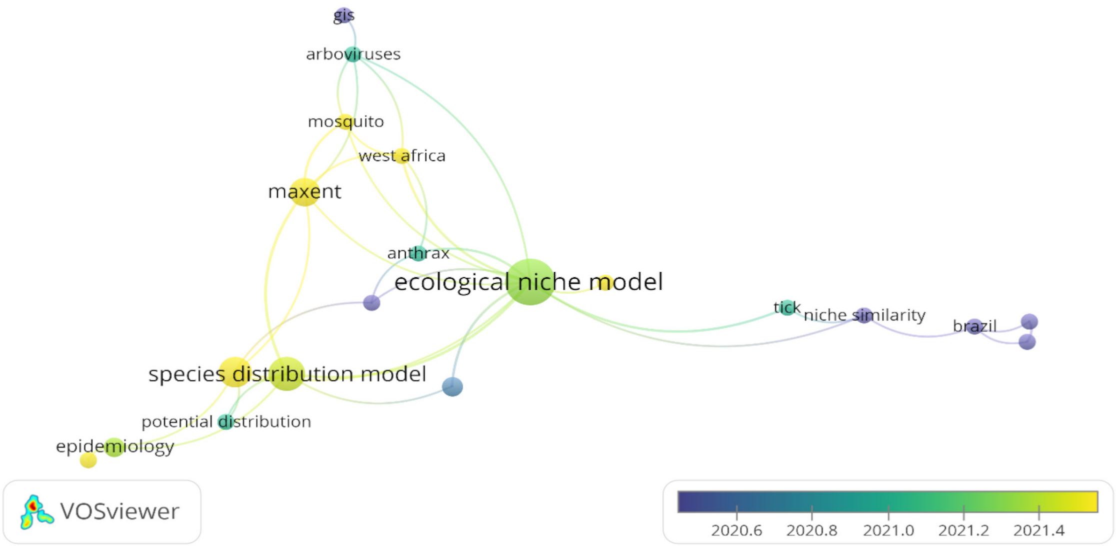


Figure 2: Keyword network analysis for the selected articles (n=78). We used default settings in VOSviewer and considered minimum 2 occurrences of keywords as a threshold. The node size indicates its frequency in the analysis of the reviewed articles. Color gradient indicates the temporal trends of use of different keywords in reviewed articles along with the publication years.

The spatial resolution of data ranged from 6 to 17,000 meters, with a median pixel size of 1000 meters (Figure S2A). Articles varied in spatial extent, from village to district, state, country, multi-country (regional), and global extents. Most of the studies were conducted in the United States (Figure S2B). The temporal extent of the articles varied between one and 101 years, with a median of 11 years (Figure S2C). Articles used different temporal resolutions (time durations between two sequential data collections), from hourly to annual datasets. Most articles used data that were collected monthly, seasonally or annually (Figure S2D).

Researchers applied various statistical methods to implement ENM theory and examine the ecological niches of their disease study subjects. Sixty articles (76.9%) used only MaxEnt as a modeling technique. Other articles also used MaxEnt in combination with other modeling techniques, such as generalized linear models, random forest, boosted regression trees, and generalized additive models (Table S1). Considering subject organisms, most articles aimed to reconstruct the ecological niche of mosquitoes (20.5%; n=16) and ticks (14.1%; n=11). Only a limited number of articles used ENM to understand the ecological niches of pathogens in tandem with their hosts (Figure S2E). Out of 78 articles, six used biotic (e.g., human footprint) in tandem with abiotic (e.g., temperature) variables to study infectious disease systems.

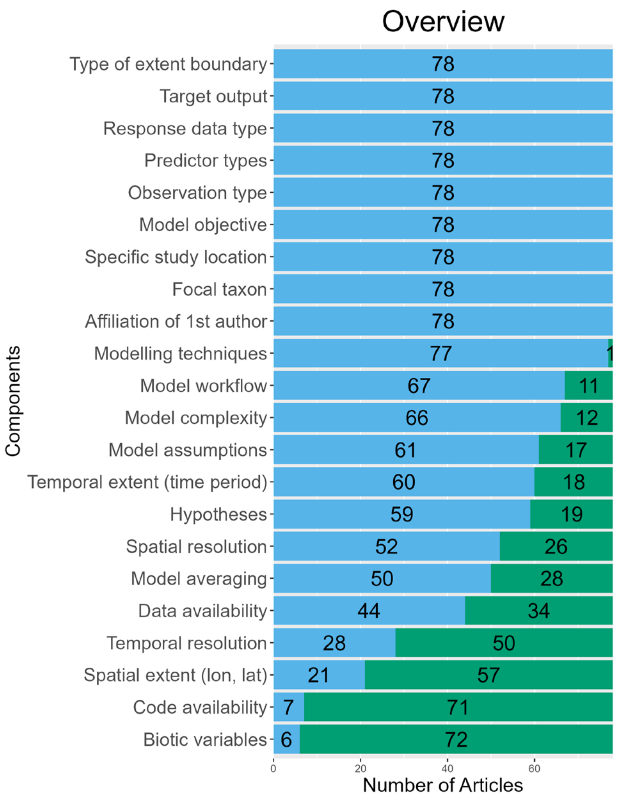


Figure 3: Articles containing or missing information about different components of the Overview section. The Overview section contained key information about analysis. Cyan-blue: Number of articles with Overview information. Green-cyan: Number of articles missing Overview information. The most complete components in the overview section were types of extent boundary, target output, response data type, predictor types, observation type, model objective, specific study location, focal taxon, and affiliation of first author. While the least reported components were biotic variables, code availability, spatial extent (lon, lat), temporal resolution, and data availability.

Data

Four of the 32 components used for modeling were present, including ecological level, taxon names, predictor variables and spatial extent, representing 12.5% (n=32) of the articles reviewed. A limited number of articles provided information on background data (44.9%; n=35), sample design (48.7%; n=38), addressing errors and biases (61.5%; n=48), temporal resolution (24.4%; n=19), and extent of transfer data (30.8%; n=24).

Articles examined pathogens and their hosts at different levels of biological organization. The majority of the research was done at the individual level (57.7%; n=45), followed by population (33.3%; n=26) and community (9.0%; n=7) levels (Figure S3.1A). Articles used a variety of sampling designs, such as field surveys (n=8), random sampling (n=11), and data from online databases (n=7), but most (n=41) of the 78 articles reviewed did not disclose details regarding their sampling methodology (Figure S3.1C). The sample sizes used in ENM research varied from 7 (*Bacillus cereus*) to 1,394,279 (tuberculosis) occurrence records, with a median of 233 occurrences (Figure S3.1B).

Most articles mentioned data partitioning procedures during model evaluation but did not mention percentage or number of observations or occurrences records used for model training, and testing. Data partitioning information for model validation was explained in 54 articles (69.2%), and training data in 59 (75.6%) articles (Figure 4).

Articles employed either geographic coordinate reference systems (n=12) or projected coordinate systems (n=16) to spatially identify occurrence records. Only 32 (42.2%) articles provided details regarding the coordinate reference system (Figure S3.2A). Of the 78 articles, 74 used bioclimatic variables, either alone or in tandem with other variables. In addition to bioclimatic variables, articles used sociodemographic, topography, elevation, Normalized Difference Vegetation Index, and population density data as ENM predictor variables (Figure S3.2B). Variable data were mostly derived (N=57) from the WorldClim database, either by itself or in conjunction with other data sources (Figure S3.2C; Table S2). Articles also downloaded bioclimatic data from MERRAclim (n=3), CHELSA (n=4), and other sources (Figure S3.2C; Table S2).

Almost all articles (96.2%; n=75) mentioned model transference. A limited number of articles provided information about the spatial resolution and extent of variables for model transference. Similarly, a small number of articles mentioned the temporal resolution (24.4%; n=19) and temporal extent (30.8%; n=24) of their study. Only 38.5% of articles (n=30) provided information about environmental scenarios used in model transferring (Figure 4).

Modelling

There were twelve distinct components in the modeling section, but only one of them (8.3%, n=12) was explained in 100% of articles (Figure 4). Information about

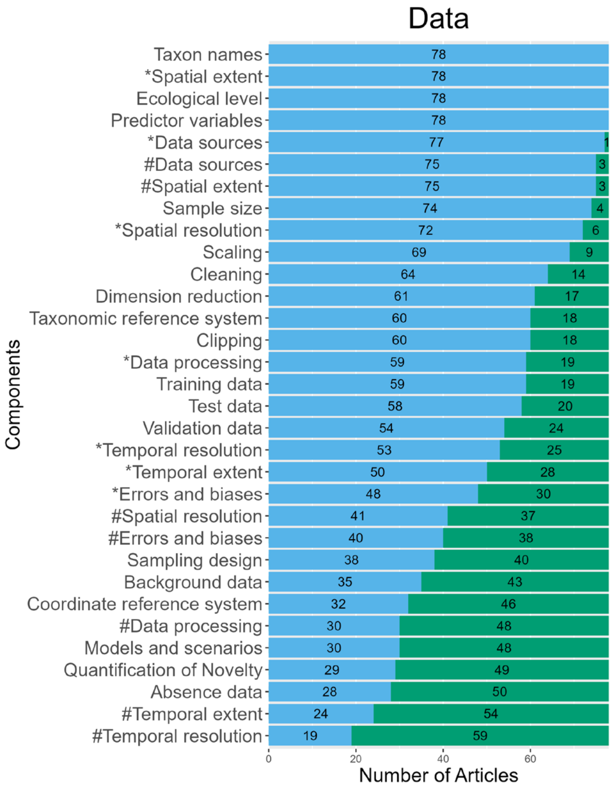


Figure 4: Data section is divided into four different subsections (Biodiversity data, Predictor variables, Data partitioning and Transfer data). This sections summaries the information about species and predictor variables, and data processing. Cyan-blue: Number of articles with Data information. Green-cyan: Number of articles missing Data information. The most complete components in the data section were taxon names, spatial extent (predictor), ecological level, predictor variables, and data sources. While the least reported components were temporal resolution (transfer), temporal extent (transfer), absence data, quantification of novelty, and models and scenarios.
* is for predictor, # is for transfer.

nested data was provided in 2.7% (n=2), temporal autocorrelation in 2.7% (n=2), model averaging in 20.5% (n=16), model ensemble in 39.7% (n=31), and model setup for extrapolation 46.2% (n=36).

Most articles (n=53) did not provide information about the pre-selection of the predictor variables. There were 15 articles merely referring to "correlation analysis" without explaining the precise methodology used to assess variable redundancy. Studies used seven different methods or techniques for the pre-selection of the variables (Figure S4A). For variables, the most commonly used methods were principal component analysis (n=13), Pearson correlation analysis (n=12), Spearman correlation (n=5), and variance inflation factor (n=4) (Figure S4A). Articles used eight different techniques to identify the variable of importance (Figure S4B). For identifying variable importance, studies implemented the Jackknife test (n=20), correlation score (n=6), principal component analysis (n=6),

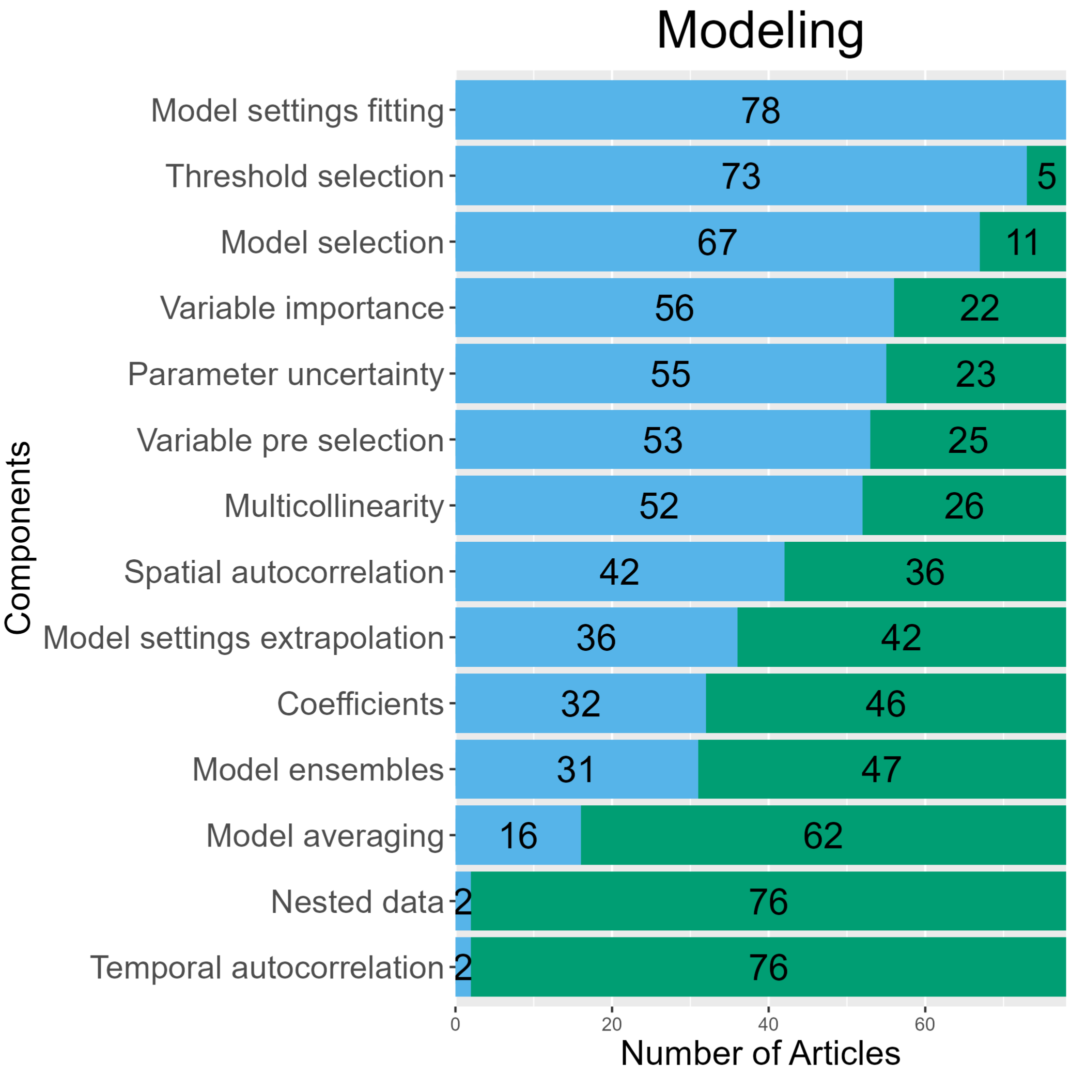


Figure 5: Articles containing or missing information about different components of the Modelling section. The Modelling section contained information about repeatability of the model building. Cyan-blue: Number of articles with Modelling information. Green-cyan: Number of articles missing Modelling information. The most completely reported components in the modelling section were model setting fitting, threshold selection, model selection, variable importance, and parameter uncertainty. While the least reported components were temporal autocorrelation, nested data, model averaging, model ensembles, and coefficients.

permutation importance (n=6), and cross-validation (n=3) most commonly (Figure S4B). Articles mentioned that they checked for multicollinearity among the variables, though most of them did not mention specific methods they had used (n=26). We found that the articles used eight different techniques for multicollinearity check. The most common techniques for multicollinearity check in ENM were principal component analysis (n=7), Pearson correlation (n=6), and variance inflation factor (n=5) (Figure S4C).

Assessment

In the assessment section, five components were included, which are useful to explain the estimated relationship between biodiversity and environmental data (Figure 5). None of the five components of assessment section were explained in all articles (Figure 6). A limited number of articles provided information about expert judgement about ENM assessment (11.5%; n=9), performance on validation data (46.2%; n=36), and response shape of predictor variables (62.8%; n=49).

Prediction

In ENM articles, seven components were considered for prediction. Only the information related to prediction unit was explained in all articles (Figure 7). Only a limited number of articles addressed algorithmic uncertainty (28.2%; n=22), novel environment (38.2%; n=29), and scenario uncertainty (38.5%; n=30).

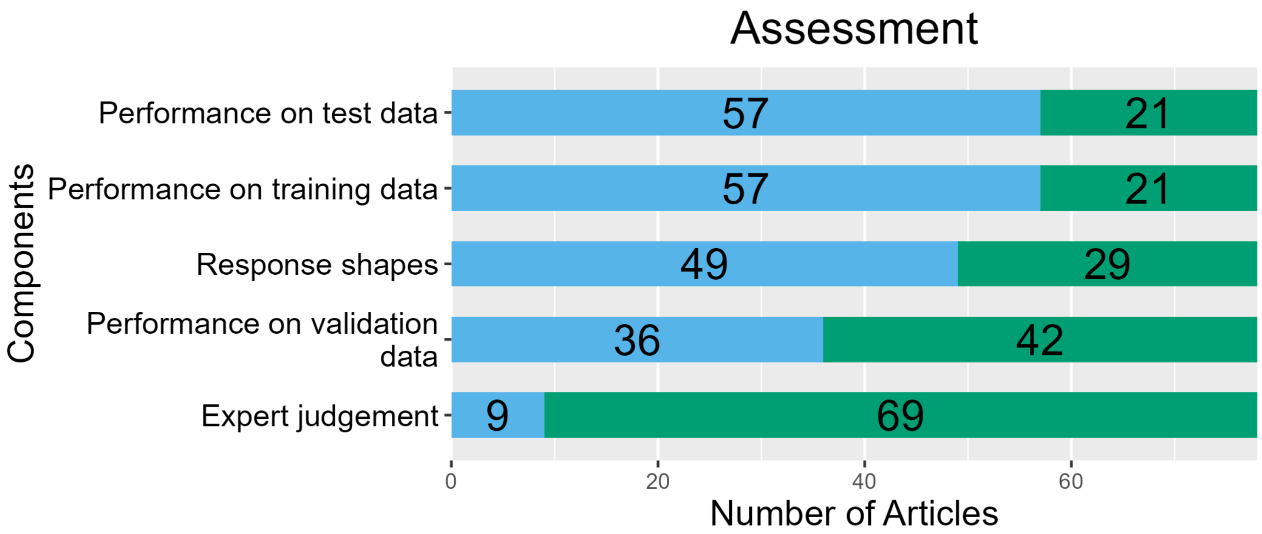


Figure 6: Articles containing or missing information about different components of the Assessment section. The Assessment section contained key information about the performance of the built ENM. Cyan-blue: Number of articles with Assessment information. Green-cyan: Number of articles missing Assessment information. The most complete components of Assessment section were performance of test data and performance on training data. While the least reported components were expert judgement and performance on validation data.

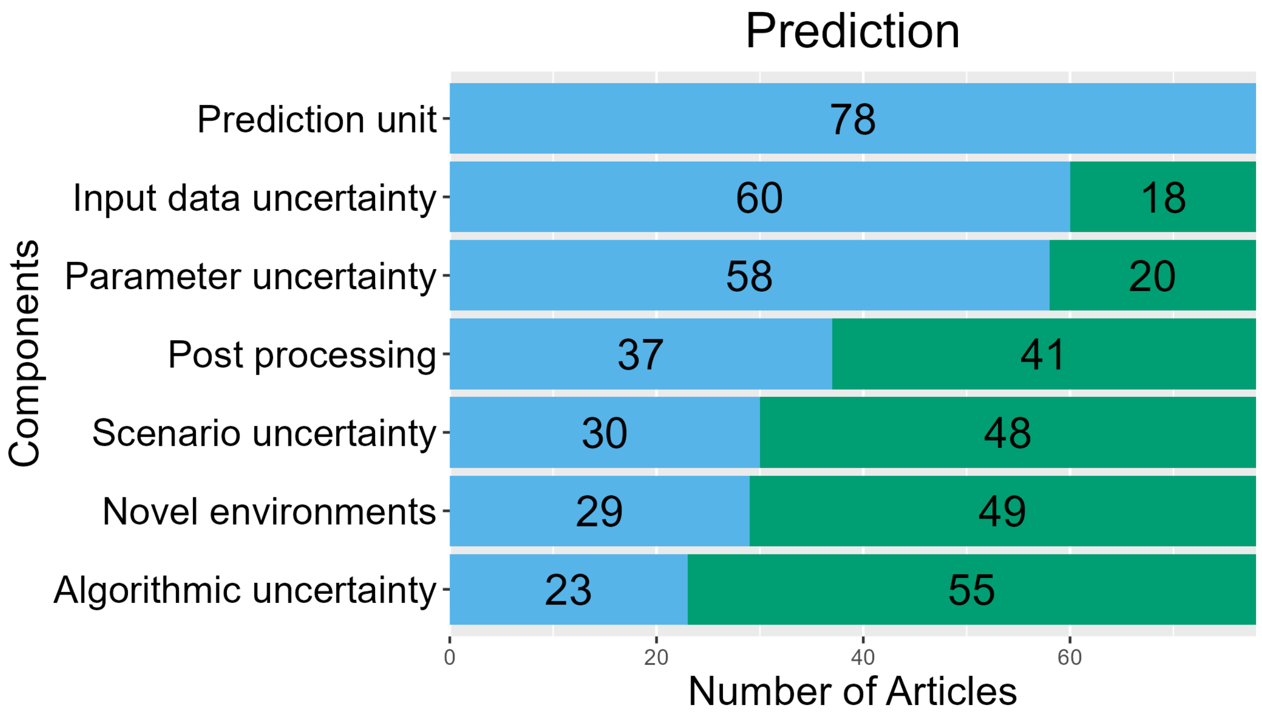


Figure 7: Articles containing or missing information about different components of the Prediction section. The Prediction section contained information about inter or extrapolating of the developed ENM. Cyan-blue: Number of articles with Prediction information. Green-cyan: Number of articles missing Prediction information. The most complete components in the prediction section were prediction unit, input data uncertainty, and parameter uncertainty. While the least reported components were algorithmic uncertainty, novel environment, and scenario uncertainty.

Discussion

We evaluated published articles for consistency and transparency in reporting key modeling and data components in ecological niche modeling (ENM) applications in disease ecology and disease biogeography. The use of ENM in infectious disease research is increasing with the journals where articles are being published varying in impact factor. A previous study reported that species distribution modeling research is published across a wide range of journals, with 12% annual growth (Vasconcelos et al. 2024). The increasing trends of articles publication in diverse journals reflect the interdisciplinary nature and broad applicability of ENM. We found that most articles missed important components during ENM building and were biased towards a certain modeling technique (i.e., MaxEnt), disease system (i.e., vector-borne), and study region (i.e., United States). Our study identified methodological gaps of published articles, which could limit the replicability of studies. Our findings highlight an urgent need to improve ENM protocols and workflows among researchers, reviewers, and editors to as a means to improve rigor and reproducibility.

Our analysis revealed that many infectious disease articles lacked comprehensive reporting of key components essential for the proper implementation of ENM across the ODMAP sections (Feng et al. 2019a; Zurell et al. 2020). For instance, temporal (n=19; 24.4%) and spatial resolution (n=41, 52.6%), which directly influence modeling outputs, were inadequately addressed in most articles (Figure 4). Ecological niches are highly context dependent and vary across different spatial and temporal scales. To understand pathogen transmission dynamics, comprehensive and scale-aware approaches are important for understanding ecology and physiology of the organisms involved in disease transmission (Escobar 2020; Zarzo-Arias et al. 2023).

The absence of a clear description of model parameterization limits the accurate validation and reproducibility of ENM research, posing challenges considering the fields rapid growth and the lack of reproducibility standards (Araújo et al. 2019; Feng et al. 2019a; Hallgren et al. 2019; Kass et al. 2025; Sandel et al 2011; Schmolke et al. 2010; Sillero et al. 2021; Soley-Guardia et al. 2024; Valavi et al. 2022; Veronica and Liu 2024). Previous case studies have similarly highlighted the unstructured and incomplete explanation of components (e.g., Braga et al. 2014; Escobar and Peterson 2013; Escobar et al. 2014; Escobar et al. 2018; Kass et al. 2025; Merow et al. 2013; Peterson and Nakazawa 2008; Xu et al. 2024; Zurell et al. 2020). To address these issues, infectious disease modelers should adhere to a standardized framework that ensures detailed reporting of fundamental modeling components, facilitating both effective communication among modelers and better understanding by readers.

Numerous mathematical and statistical algorithms are available to estimate the ecological niches, with some being more frequently used on the assumption that they are the most effective approaches (Araújo et al. 2019; Beery et al. 2021; Blonder et al. 2018; Drake 2015; Elith et al. 2008; Escobar 2020; Kass et al. 2025; Phillips et al. 2006; Qiao et al. 2015; Soley-Guardia et al. 2024; Valavi et al. 2022; Wiens et al. 2009). We found that most articles utilized MaxEnt as the predominant ENM algorithm (Table: S1). Nevertheless, evidences demonstrate that no single modeling algorithm is universally superior (Qiao et al., 2015; Valavi et al., 2022). MaxEnt is the most common algorithm used by researchers modeling ecological niches and spatial distributions (Barker and MacIsaac 2022; Campos et al. 2023; Feng et al. 2019b; Lippi et al. 2023a; Valavi et al. 2022). Previous studies have shown that the performance of ENM varies with changes in species or calibration areas (Elith et al., 2006; Qiao et al., 2019; Valavi et al., 2022) and algorithm employed (Escobar et al. 2018). Strikingly, the selection of modeling algorithms depends on the preferences of the modelers, instead of more scientific approaches, with some modelers preferring default settings and others careful model tuning (Valavi et al. 2022). The preference maybe responds to the modelers' limited understanding of how algorithms work (Joppa et al. 2013). To construct an ENM, the choice of modeling techniques and predictor variables should be informed by the physiology, ecology, and biogeographic history of the organism (Escobar 2020; Peterson et al. 2011). Joppa et al. (2013) recommended that the community should prioritize the understanding of the algorithms over the development of newer, fancier modeling workflows.

We found that most studies used predictor variables from a particular data source (i.e., WorldClim) (Hijmans et al. 2005). Reliance on a specific source suggests that researchers are potentially biased on the use of similar types of readily available predictors, despite differences in study questions, model requirements, and mismatching study periods (Araújo et al. 2019; Booth 2022; Morales-Barbero et al. 2019; Oliver and Morecroft 2014; Regos et al. 2019). Previous studies also found WorldClim data to be the most frequently used (e.g., Barker and MacIsaac 2022; Bobrowski et al. 2021; Booth 2022; Datta et al. 2020; Escobar et al. 20014; Lippi et al. 2023a; Maria et al. 2017; Merkenschlager et al. 2023; Morales-Barbero et al. 2019; Poggio et al. 2018). WorldClim temperature and precipitation data are derived from global weather station averages and interpolated with covariables such as elevation, distance to the coast, and satellite derived data to generate high-resolution bioclimatic variables (e.g., 30 arc seconds, approximately 1 kilometer) (Fick and Hijmans 2017; Hijmans et al. 2005; Waltari et al. 2014; Xu and Hutchinson 2011). Because weather station coverage varies geographically (Hijmans et al. 2005; Waltari et al. 2014), WorldClim interpolations at ~1 km resolution generate 99.99% model-generated, interpolated climatic data (Peterson 2014). In contrast, remote sensing offers consistent, large-scale monitoring of surface temperature and precipitation, including day and night measurements (Adler et al. 2000; Jones et al. 2010; Waltari et al. 2014). A study reported that ENM built with predictor variables from MERRA performed as well as or better than models with WorldClim data (Waltari et al. 2014). As such, the selection of predictor variables for ENM should align with the study question and spatiotemporal scale of the calibration data (Pearson and Dawson 2003; Peterson et al. 2011; Regos et al. 2019). Using multiple predictors, including non-climatic predictors, improves ecological understanding of organisms (Regos et al. 2019). Employing diverse data sources, while considering species behavior and calibration area, helps reduce uncertainty and avoid overreliance in ENM construction (Regos et al. 2019). A next frontier in ENM applications to disease should be the combination of abiotic with biotic predictors, but generation of biotic variables is warranted (Peterson et al. 2019).

We identified significant gaps in published articles regarding clear explanations of occurrence record data collection, cleaning processes, removal of spatial biases, and data partitioning into training/testing datasets. Data quality is a critical factor in building accurate and reliable ENM (Feng et al. 2019a; Escobar 2020; Soley-Guardia et al. 2024; Zurell et al. 2020). Ecological processes vary across spatial scales, and errors in selecting appropriate resolution and extent for occurrence and predictor variables can violate biological and statistical assumptions (Feng et al. 2019a; McGill 2010; Soberón and Nakamura 2009; Soley-Guardia et al. 2024). Biotic and abiotic predictors influence species distributions, but the level of influence vary with spatial resolution of variables, often requiring modification of original resolution of variables for consistency (Soberón and Nakamura 2009; Sunday et al. 2012). As such, researchers should follow standard procedures to describe how spatial and environmental biases were addressed (Boria et al. 2014; Park and Davis 2017), how background data were selected (Barbet-Massin et al. 2012; Phillips et al. 2009), how calibration data were collected and prepared (Anderson and Raza 2010; Escobar 2020; Saupe et al. 2012), and how multicollinearity among predictors were managed (Bucklin et al. 2015; Pliscoff et al. 2014; Synes and Osborne 2011; Zeng et al. 2016).

A higher number of articles utilized ENM to study the ecology and distribution of vector-borne disease caused by mosquitoes and ticks (Figure S2E). Our results align with Lippi et al. (2023b) and Van de Vuurst et al. (2023), who reported a growing focus on modeling vector distribution over the past decade. The higher number of ENM articles may be driven by factors such as the high disease burden posed by vectors, the critical need for risk area identification, policy mandates, and biased allocation of resources, or higher availability of vector datasets.

ENM studies of infectious diseases are also geographically biased. We found a strong focus of ENM applications to disease in the Americas (Figure S2B). Lippi et al. (2023a) and Van de Vuurst et al. (2023) observed a higher number ENM modeling efforts in North America and Europe. Paradoxically, vector-borne disease are predominantly identified in tropical, low-income countries, highlighting the need for integrated, geographically focused studies that account for environmental justice (Rosenberg et al. 2013). The limited number of ENM in regions like southeast Asia may stem from scarce publicly available data, limited research resources, and fewer research effort compared to North and South America, and Africa (Rosenberg et al. 2013). In contrast, the Americas benefit from robust systematic surveillance systems like the National Ecological Observatory Network, which provide long-term, freely available pathogen, vector, and host data for the United States (Springer et al. 2016). Global databases like the Global Biodiversity Information Facility also support parasite, vector, and host research, though their data quality and geographic coverage are influenced by research and funding biases (GBIF Secretariat and IAIA, 2020). More efforts may be needed to standardize and release pathogen occurrence data. Overall, the public health and economic significance of infectious diseases, availability of long-term environmental data, and extensive occurrence records are likely to promote more ENM studies in years to come.

The use of ENM methods, such as MaxEnt, have been commonly used to study the ecology of vector-borne and directly transmitted zoonotic diseases. For instance, MaxEnt has been effectively applied to map the potential distribution of schistosomiasis hosts and to guide eradication efforts for Tsetse flies and Trypanosomiasis (Dicko et al. 2014, Singleton et al. 2024). Nevertheless, ENM based solely on correlative relationships with climatic variables,

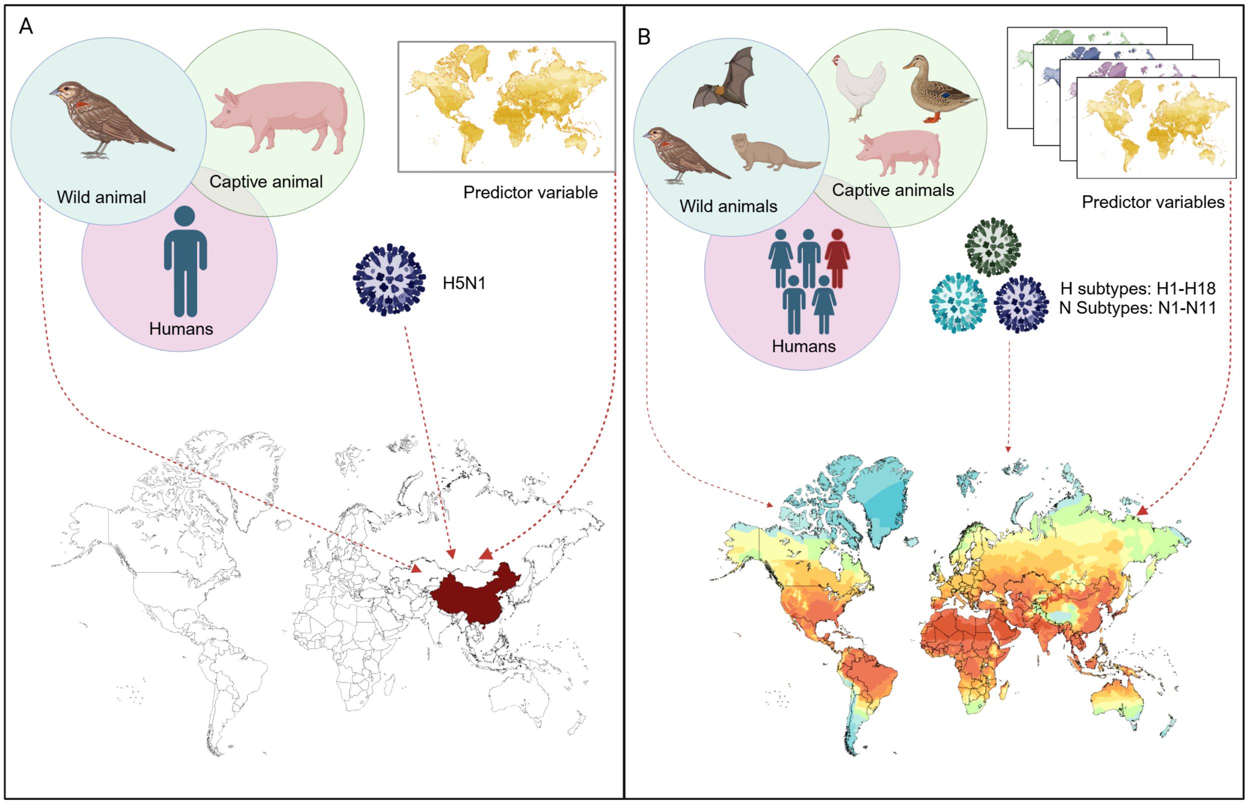


Figure 8: Schematic representation of simple and complex ecological niche model development. A. Building and geographical projection of ecological niche model using a subtype of avian influenza virus, a type of predictor, and a bird species may result in an incomplete understanding of the ecology of zoonotic avian influenza. **B.** Building and geographical projection of ecological niche using multiple subtypes of avian influenza virus, diverse predictors and multiple bird species for a more comprehensive understanding of the ecology of zoonotic avian influenza.

like those from WorldClim, may fall short in capturing the full complexity of disease ecology (Figure 8) (Cuervo et al. 2023). For example, a parsimonious study of avian influenza could include only one viral lineage (e.g., H5N1), one primary host (e.g., one bird), and one secondary host (e.g., pig) in a low-dimensional environmental space (e.g., one environmental variable). This simple model to forecast transmission from birds to other species including humans, could identify just a portion of the actual transmission risk (Figure 8A). To develop more comprehensive and biologically meaningful models, it is essential to understand the fundamental ecology of the disease systems and incorporate data on original hosts, secondary hosts, pathogen diversity, and relevant human features in ENM applications (Figure 8B). MaxEnt limitations derive from its vulnerability to extrapolate, which may produce overly simplistic or misleading results (Escobar et al. 2018; Qiao et al. 2019). For example, MaxEnt models can unrealistically predict the survival of malaria vectors beyond 100°C (boiling water) temperature (Owens et al 2013). Therefore, future ENM to understand the ecology of infectious disease, including pathogen, vector, and host, should consider the physiology of the organisms to generate sound and realistic forecast.

Caveats

The ODMAP primarily serves as a framework to enhance the reproducibility and transparency of spatial distribution models by standardizing the reporting of modeling procedures and data (Zurell et al. 2020). Its core purpose is to allow other researchers to evaluate, reproduce, and build upon published work rigorously, rather than inherently guaranteeing the ecological validity or comprehensive capture of complex disease ecological processes. Since pioneering applications to infectious diseases (i.e., Peterson et al. 2002), use of ENM in infectious disease research has emerged as a new theoretical and analytical framework (Srivastava et al. 2019; Escobar 2020; Vasconcelos et al. 2024). Nevertheless, key ENM components (e.g., study area, occurrence points, environmental variables, analytical framework) could mislead model results (Araújo et al. 2019; Barker and MacIsaac 2022). Our analysis focused solely on description of the analytical framework in the literature based on the ODMAP protocol. We did not account for the theoretical framework employed for the biological interpretation of the models.

ODMAP protocol suggests describing ENM parameterization in the main article, so that we focused on the details in the main articles. It is plausible that ENM tuning was rigorous for some studies but described only in the supplementary material, not in the main article. To address this caveat, we thoroughly examined the supplementary materials available in the selected articles and found no methodological descriptions as supplementary materials. Overall, we found a considerable lack of details to foster reproducibility and also minimum efforts in model parameterization to find the best model that reconstructs nature and allow the prediction of disease events (Levin 1992). This review covered articles published between 2000 and 2022 as a sample of current trends. Although, previous seminal work in the field and recent applications were not included, we argue that they likely suffer similar flaws in the analytical framework. Future ENM reviews should consider broader publication timeframes and future ENM applications should adopt ENM protocols to enhance reproducibility and transparency (Feng et al. 2019a; Zurell et al. 2020).

Conclusions

This study highlights flaws in ENM applications in disease ecology and disease biogeography. Specifically, many published articles provide insufficient explanations of key model components, such as spatial and temporal extent, biodiversity data sources, and mitigation of collinearity, which reduce study reproducibility and transparency. While ENM applications have advanced our understanding of disease ecology and disease biogeography, our findings underscore the importance of adopting standard protocols to ensure consistency and reproducibility. Our study aimed to lay the groundwork for subsequent discussions and research on effectively incorporating important parameters in ENM applications to disease ecology and disease biogeography.

Acknowledgements

This study was supported by the National Science Foundation Human-Environment and Geographical Sciences Program (2116748) and CAREER (2235295) awards, and by the National Institute of Allergy and Infectious Diseases of the National Institutes of Health (K01AI168452). This project was also supported by seed grants from the Virginia Tech Institute for Critical Technology and Applied Science, Pandemic Prediction and Prevention Destination Area, and the Center for Emerging, Zoonotic, and Arthropod-borne Pathogens. The content is solely the authors' responsibility and does not necessarily represent the official views of the National Institutes of Health. The authors would like to thank Dana Hawley, William Mark Ford, Sarah Karpanty, Paanwaris Paansri, and Paige Van de Vuurst for their support on the development and review of this manuscript.

Data Availability

The data table on which the analyses in this study are based is available via KU Scholarworks.[1]

Competing Interests

The authors have declared that no competing interests exist.

Author Contributions

S.I. provided substantial input on the conceptualization, performed the formal analyses and wrote the first version of the manuscript and headed the review editing. M.C.G., D.S.T., L.E.E. provided methodological revision and gave considerable suggestions on writing – review and editing. L.E.E. developed the idea. All authors approved the last version of this article.

References

Adler, Robert F., George J. Huffman, David T. Bolvin, Scott Curtis, and Eric J. Nelkin. 2000. "Tropical Rainfall Distributions Determined Using TRMM Combined with Other Satellite and Rain Gauge Information." *Journal of Applied Meteorology* 39(12): 2007-2023. https://doi.org/10.1175/1520-0450(2001)040%3C2007:TRDDUT%3E2.0.CO;2

Anderson, Robert P., and Ali Raza. 2010. "The Effect of the Extent of the Study Region on GIS Models of Species Geographic Distributions and Estimates of Niche Evolution: Preliminary Tests with Montane Rodents (Genus *Nephelomys*) in Venezuela." *Journal of Biogeography* 37(7): 1378–1393. https://doi.org/10.1111/j.1365-2699.2010.02290.x

Araújo, Miguel B., Robert P. Anderson, A. Márcia Barbosa, Colin M. Beale, Carsten F. Dormann, Regan Early, Raquel A. Garcia, et al. 2019. "Standards for Distribution Models in Biodiversity Assessments." *Science Advances* 5(1): eaat4858. https://doi.org/10.1126/sciadv.aat4858

Astorga, Francisca, Luis E. Escobar, Daniela Poo-Muñoz, Joaquin Escobar-Dodero, Sylvia Rojas-Hucks, Mario Alvarado-Rybak, Melanie Duclos et al. 2018. "Distri-

---

[1] https://hdl.handle.net/1808/36112.

butional Ecology of Andes Hantavirus: A Macroecological Approach." *International Journal of Health Geographics* 17: 1-12. https://doi.org/10.1186/s12942-018-0142-z

Barbet-Massin, Morgane, Frédéric Jiguet, Cécile Hélène Albert, and Wilfried Thuiller. 2012. "Selecting Pseudo-Absences for Species Distribution Models: How, Where and How Many?" *Methods in Ecology & Evolution* 3(2): 327–338. https://doi.org/10.1111/j.2041-210X.2011.00172.x

Barker, Justin R., and Hugh J. MacIsaac. 2022. "Species Distribution Models Applied to Mosquitoes: Use, Quality Assessment, and Recommendations for Best Practice." *Ecological Modelling* 472(10):110073. https://doi.org/10.1016/j.ecolmodel.2022.110073

Beery, Sara, Elijah Cole, Joseph Parker, Pietro Perona, and Kevin Winner. 2021. "Species Distribution Modeling for Machine Learning Practitioners: A review." In ACM SIGCAS Conference on Computing and Sustainable Societies (COMPASS) (COMPASS '21), June 28-July 2, 2021, Virtual Event, Australia. ACM, New York, NY, USA, 20 pages. https://doi.org/10.1145/3460112.3471966

Blonder, Benjamin, Cecina Babich Morrow, Brian Maitner, David J. Harris, Christine Lamanna, Cyrille Violle, Brian J. Enquist, and Andrew J. Kerkhoff. 2018. "New Approaches for Delineating n-dimensional Hypervolumes." *Methods in Ecology and Evolution* 9(2): 305-319. https://doi.org/10.1111/2041-210X.12865

Bobrowski, Maria, Johannes Weidinger, and Udo Schickhoff. 2021. "Is New Always Better? Frontiers in Global Climate Datasets for Modeling Treeline Species in the Himalayas." *Atmosphere* 12(5): 543. https://doi.org/10.3390/atmos12050543

Booth, Trevor H. 2022. "Checking Bioclimatic Variables that Combine Temperature and Precipitation Data Before their Use in Species Distribution Models." *Austral Ecology* 47(7): 1506-1514. https://doi-org.ezproxy.lib.vt.edu/10.1111/aec.13234

Boria, Robert A., Link E. Olson, Steven M. Goodman, and Robert P. Anderson. 2014. "Spatial Filtering to Reduce Sampling Bias Can Improve the Performance of Ecological Niche Models." *Ecological Modelling* 275(3): 73–77. https://doi.org/10.1016/j.ecolmodel.2013.12.012

Braga, Guilherme Basseto, José Henrique Hildebrand Grisi-Filho, Bruno Meireles Leite, Elaine Fátima de Sena, and Ricardo Augusto Dias. 2014. "Predictive Qualitative Risk Model of Bovine Rabies Occurrence in Brazil." *Preventive Veterinary Medicine* 113(4): 536-546. https://doi.org/10.1016/j.prevetmed.2013.12.011

Brito-Hoyos, Diana Marcela, Edilberto Brito Sierra, and Rafael Villalobos Alvarez. 2013. "Geographic Distribution of Wild Rabies Risk and Evaluation of the Factors Associated with its Incidence in Colombia, 1982-2010." *Pan American Journal of Public Health* 33(1): 8–14. https://doi.org/10.1590/s1020-49892013000100002

Bucklin, David N., Mathieu Basille, Allison M. Benscoter, Laura A. Brandt, Frank J. Mazzotti, Stephanie S. Romañach, Carolina Speroterra, James I. Watling, and Wilfried Thuiller. 2015. "Comparing Species Distribution Models Constructed with Different Subsets of Environmental Predictors." *Diversity & Distributions* 21(1): 23–35. https://doi.org/10.1111/ddi.12247

Campos, João C., Nuno Garcia, João Alírio, Salvador Arenas-Castro, Ana C. Teodoro, and Neftalí Sillero. 2023. "Ecological Niche Models Using MaxEnt in Google Earth Engine: Evaluation, Guidelines and Recommendations." *Ecological Informatics* 76(9):102147. https://doi.org/10.1016/j.ecoinf.2023.102147

Cuervo, Pablo Fernando, Patricio Artigas, Jacob Lorenzo-Morales, María Dolores Bargues, and Santiago Mas-Coma. 2023. "Ecological Niche Modelling Approaches: Challenges and Applications in Vector-Borne Diseases." *Tropical Medicine and Infectious Disease* 8(4): 187. https://doi.org/10.3390/tropicalmed8040187

Datta, Arunava, Oliver Schweiger, and Ingolf Kuhn. 2020. "Origin of Climatic Data Can Determine the Transferability of Species Distribution Models." *NeoBiota* 59: 61-76. https://doi.org/10.3897/neobiota.59.36299

Dicko, Ahmadou H., Renaud Lancelot, Momar T. Seck, Laure Guerrini, Baba Sall, Mbargou Lo, Marc JB Vreysen et al. 2014. "Using Species Distribution Models to Optimize Vector Control in the Framework of the Tsetse Eradication Campaign in Senegal." *Proceedings of the National Academy of Sciences USA* 111(28): 10149-10154. https://doi.org/10.1073/pnas.1407773111

Dobson, Andrew P., Stuart L. Pimm, Lee Hannah, Les Kaufman, Jorge A. Ahumada, Amy W. Ando, Aaron Bernstein, et al. 2020. "Ecology and Economics for Pandemic Prevention." *Science* 369(6502): 379–381. https://doi.org/10.1126/science.abc3189

Drake, John M. 2015. "Range Bagging: A New Method for Ecological Niche Modelling from Presence-only Data." *Journal of the Royal Society Interface* 12(107): 20150086. http://dx.doi.org/10.1098/rsif.2015.0086

Elith, Jane, Catherine H. Graham*, Robert P. Anderson, Miroslav Dudík, Simon Ferrier, Antoine Guisan, Robert J. Hijmans, et al. 2006. "Novel Methods Improve

Prediction of Species' Distributions from Occurrence Data." *Ecography* 29(2): 129–151. https://doi.org/10.1111/j.2006.0906-7590.04596.x

Elith, Jane, John R. Leathwick, and Trevor Hastie. 2008. "A Working Guide to Boosted Regression Trees." *Journal of Animal Ecology* 77(4): 802-813. https://doi.org/10.1111/j.1365-2656.2008.01390.x

Escobar, Luis E. 2020. "Ecological Niche Modeling: An Introduction for Veterinarians and Epidemiologists." *Frontiers in Veterinary Science* 7: 519059. https://www.frontiersin.org/articles/10.3389/fvets.2020.519059

Escobar, Luis E., and Serge Morand. 2021. "Disease Ecology and Biogeography." *Frontiers in Veterinary Science* 8: 765825. https://doi.org/10.3389/fvets.2021.765825

Escobar, Luis E., Andrés Lira-Noriega, Gonzalo Medina-Vogel, and A. Townsend Peterson. 2014. "Potential for Spread of the White-Nose Fungus (*Pseudogymnoascus Destructans*) in the Americas: Use of Maxent and NicheA to Assure Strict Model Transference." *Geospatial Health* 9(1): 221. https://doi.org/10.4081/gh.2014.19

Escobar, Luis E., and Meggan E. Craft. 2016. "Advances and Limitations of Disease Biogeography Using Ecological Niche Modeling." *Frontiers in Microbiology* 07(8):1174. https://doi.org/10.3389/fmicb.2016.01174

Escobar, Luis E., and A. Townsend Peterson. 2013. "Spatial Epidemiology of Bat-Borne Rabies in Colombia." *Pan American Journal of Public Health* 34(2): 135–136.

Escobar, Luis E., Huijie Qiao, Javier Cabello, and A. Townsend Peterson. 2018. "Ecological Niche Modeling Re-Examined: A Case Study with the Darwin's Fox." *Ecology and Evolution* 8(10): 4757–4770. https://doi.org/10.1002/ece3.4014

Escobar, Luis E., Huijie Qiao, Christine Lee, and Nicholas BD Phelps. 2017. "Novel Methods in Disease Biogeography: A Case Study with Heterosporosis." *Frontiers in Veterinary Science* 4: 105. https://doi.org/10.3389/fvets.2017.00105

Feng, Xiao, Daniel S. Park, Cassondra Walker, A. Townsend Peterson, Cory Merow, and Monica Papeş. 2019a. "A Checklist for Maximizing Reproducibility of Ecological Niche Models." *Nature Ecology & Evolution* 3(10): 1382–1395. https://doi.org/10.1038/s41559-019-0972-5

Feng, Xiao, Daniel S. Park, Ye Liang, Ranjit Pandey, and Monica Papeş. 2019b. "Collinearity in Ecological Niche Modeling: Confusions and Challenges." *Ecology and Evolution* 9(18): 10365–10376. https://doi.org/10.1002/ece3.5555

Fick, Stephen E., and Robert J. Hijmans. 2017. "WorldClim 2: New 1-Km Spatial Resolution Climate Surfaces for Global Land Areas." *International Journal of Climatology* 37(12): 4302–4315. https://doi.org/10.1002/joc.5086

Fitzpatrick, Matthew C., Susanne Lachmuth, and Natalie T. Haydt. 2021. "The ODMAP Protocol: A New Tool for Standardized Reporting that Could Revolutionize Species Distribution Modeling*." Ecography* 44(7): 1067-1070. https://doi.org/10.1111/ecog.05700

Frans, Veronica F., and Jianguo Liu. 2024. "Gaps and Opportunities in Modelling Human Influence on Species Distributions in the Anthropocene." *Nature Ecology & Evolution* 8(7): 1365-1377. https://doi-org.ezproxy.lib.vt.edu/10.1038/s41559-024-02435-3

GBIF Secretariat and International Association for Impact Assessment. 2020. "Best Practices for Publishing Biodiversity Data from Environmental Impact Assessments." https://doi.org/10.35035/DOC-5XDM-8762

Grimm, Volker, Uta Berger, Donald L. DeAngelis, J. Gary Polhill, Jarl Giske, and Steven F. Railsback. 2010. "The ODD Protocol: A Review and First Update." *Ecological Modelling* 221(23): 2760-2768. https://doi.org/10.1016/j.ecolmodel.2010.08.019

Hallgren, W., F. Santana, S. Low-Choy, Y. Zhao, and B. Mackey. 2019. "Species Distribution Models can be Highly Sensitive to Algorithm Configuration." *Ecological Modelling* 408(9): 108719. https://doi.org/10.1016/j.ecolmodel.2019.108719

Hijmans, Robert J., Susan E. Cameron, Juan L. Parra, Peter G. Jones, and Andy Jarvis. 2005. "Very High Resolution Interpolated Climate Surfaces for Global Land Areas." *International Journal of Climatology* 25(15): 1965–1978. https://doi.org/10.1002/joc.1276

Jones, Lucas A., Craig R. Ferguson, John S. Kimball, Ke Zhang, Steven Tsz K. Chan, Kyle C. McDonald, Eni G. Njoku, and Eric F. Wood. 2010. "Satellite Microwave Remote Sensing of Daily Land Surface Air Temperature Minima and Maxima From AMSR-E." *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 3(1): 111–23. https://doi.org/10.1109/JSTARS.2010.2041530

Joppa, Lucas N., Greg McInerny, Richard Harper, Lara Salido, Kenji Takeda, Kenton O'Hara, David Gavaghan, and Stephen Emmott. 2013. "Troubling Trends in Scientific Software Use." *Science* 340(6134): 814-815. https://doi.org/10.1126/science.1231535

Kass, Jamie M., Adam B. Smith, Dan L. Warren, Sergio Vignali, Sylvain Schmitt, Matthew E. Aiello-Lammens, Eduardo Arlé et al. 2025. "Achieving Higher Standards in Species Distribution Modeling by Leveraging the Diversity of Available Software." *Ecography* 2(2025): e07346. https://doi.org/10.1111/ecog.07346

Keesing, Felicia, and Richard S. Ostfeld. 2024. "Emerging Patterns in Rodent-Borne Zoonotic Diseases." *Science* 385(6715): 1305–1310. https://doi.org/10.1126/science.adq7993

Levin, Simon A. 1992. "The Problem of Pattern and Scale in Ecology: The Robert H. Macarthur Award Lecture." *Ecology* 73(6): 1943-1967. https://doi.org/10.2307/1941447

Lippi, Catherine A., Stephanie J. Mundis, Rachel Sippy, J. Matthew Flenniken, Anusha Chaudhary, Gavriella Hecht, Colin J. Carlson, and Sadie J. Ryan. 2023a. "Trends in Mosquito Species Distribution Modeling: Insights for Vector Surveillance and Disease Control." *Parasites & Vectors* 16(1): 302. https://doi.org/10.1186/s13071-023-05912-z

Lippi, Catherine A, Samuel S C Rund, and Sadie J Ryan. 2023b. "Characterizing the Vector Data Ecosystem." *Journal of Medical Entomology* 60(2): 247–254. https://doi.org/10.1093/jme/tjad009

Maher, Sean P., Christine Ellis, Kenneth L. Gage, Russell E. Enscore, and A. Townsend Peterson. 2010. "Range-wide Determinants of Plague Distribution in North America." *The American Journal of Tropical Medicine and Hygiene* 83(4): 736. https://doi.org/10.4269/ajtmh.2010.10-0042

Maria, Bobrowski, and Schickhoff Udo. 2017. "Why Input Matters: Selection of Climate Data Sets for Modelling the Potential Distribution of a Treeline Species in the Himalayan Region." *Ecological Modelling* 359(9): 92-102. https://doi.org/10.1016/j.ecolmodel.2017.05.021

McGill, Brian J. 2010. "Matters of Scale." *Science* 328(5978): 575–576. https://doi.org/10.1126/science.1188528

Merkenschlager, Christian, Freddy Bangelesa, Heiko Paeth, and Elke Hertig. 2023. "Blessing and Curse of Bioclimatic Variables: A Comparison of Different Calculation Schemes and Datasets for Species Distribution Modeling within the Extended Mediterranean Area." *Ecology and Evolution* 13(10): e10553. https://doi.org/10.1002/ece3.10553

Merow, Cory, Matthew J. Smith, and John A. Silander Jr. 2013. "A Practical Guide to Maxent for Modeling Species' Distributions: what it Does, and why Inputs and Settings Matter." *Ecography* 36(10): 1058-1069. https://doi.org/10.1111/j.1600-0587.2013.07872.x

Merow, Cory, Brian S. Maitner, Hannah L. Owens, Jamie M. Kass, Brian J. Enquist, Walter Jetz, and Rob Guralnick. 2019. "Species' Range Model Metadata Standards: RMMS." *Global Ecology and Biogeography* 28(12): 1912-1924. https://doi-org.ezproxy.lib.vt.edu/10.1111/geb.12993

Moher, David, Alessandro Liberati, Jennifer Tetzlaff, Douglas G. Altman, and The PRISMA Group. 2009. "Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement." *PLOS Medicine* 6(7): e1000097. https://doi.org/10.1371/journal.pmed.1000097

Morales-Barbero, Jennifer, and Julia Vega-Álvarez. 2019. "Input Matters Matter: Bioclimatic Consistency to Map More Reliable Species Distribution Models." *Methods in Ecology and Evolution* 10(2): 212-224. https://doi.org/10.1111/2041-210X.13124

Oliver, Tom H., and Mike D. Morecroft. 2014. "Interactions between Climate Change and Land Use Change on Biodiversity: Attribution Problems, Risks, and Opportunities." *WIREs Climate Change* 5(3): 317–335. https://doi.org/10.1002/wcc.271

Olivera, Leonela, Eugenia Minghetti, and Sara I. Montemayor. 2021. "Ecological Niche Modeling (ENM) of *Leptoglossus clypealis* a New Potential Global Invader: Following in the Footsteps of *Leptoglossus occidentalis*?" *Bulletin of Entomological Research* 111(3): 289-300. https://doi.org/10.1017/S0007485320000656

Owens, Hannah L., Lindsay P. Campbell, L. Lynnette Dornak, Erin E. Saupe, Narayani Barve, Jorge Soberón, Kate Ingenloff et al. 2013. "Constraints on Interpretation of Ecological Niche Models by Limited Environmental Ranges on Calibration Areas." *Ecological Modelling* 263(8): 10-18. https://doi.org/10.1016/j.ecolmodel.2013.04.011

Park, Daniel S., and Charles C. Davis. 2017. "Implications and Alternatives of Assigning Climate Data to Geographical Centroids." *Journal of Biogeography* 44(10): 2188–2198. https://www.jstor.org/stable/26626941

Pearson, Richard G., and Terence P. Dawson. 2003. "Predicting the Impacts of Climate Change on the Distribution of Species: Are Bioclimate Envelope Models Useful?" *Global Ecology and Biogeography* 12(5): 361–371. https://doi.org/10.1046/j.1466-822X.2003.00042.x

Peterson, A. Townsend. 2008. "Biogeography of Diseases: A Framework for Analysis." *Naturwissenschaften* 95(6): 483–491. https://doi.org/10.1007/s00114-008-0352-5

Peterson, A. Townsend, Robert P. Anderson, Marlon E. Cobos, Martín Cuahutle, Angela P. Cuervo-Robayo, Luis E. Escobar, Marc Fernández et al. 2019. "Curso Modelado de Nicho Ecológico, Versión 1.0." *Biodiversity Informatics* 14: 1-7. https://doi.org/10.17161/bi.v14i0.8189

Peterson, A. Townsend, and Kate Ingenloff. 2016. "Biodiversity Informatics Training Curriculum, version 1.2." *Biodiversity Informatics* 11:1. https://doi.org/10.17161/bi.v11i1.5008

Peterson, A. T., and Y. Nakazawa. 2008. "Environmental data sets matter in ecological niche modelling: an example with *Solenopsis invicta* and *Solenopsis richteri*." *Global Ecology and Biogeography* 17(1): 135-144. https://doi-org.ezproxy.lib.vt.edu/10.1111/j.1466-8238.2007.00347.x

Peterson, A. Townsend. 2014. *Mapping Disease Transmission Risk*. Baltimore: Johns Hopkins University Press. https://doi.org/10.1353/book.36167

Peterson, A. T., Victor Sánchez-Cordero, C. Ben Beard, and Janine M. Ramsey. 2002. "Ecologic Niche Modeling and Potential Reservoirs for Chagas Disease, Mexico." *Emerging Infectious Diseases* 8(7): 662–667. https://doi.org/10.3201/eid0807.010454

Peterson, A.T., J. Soberón, R.G. Pearson, R.P. Anderson, E. Martínez-Meyer, M Nakamura, and M.B. Araújo. 2011. "Ecological Niches and Geographic Distributions." Princeton University Press, New Jersey, USA. https://press.princeton.edu/books/paperback/9780691136882/ecological-niches-and-geographic-distributions-mpb-49

Phillips, Steven J., Miroslav Dudík, Jane Elith, Catherine H. Graham, Anthony Lehmann, John Leathwick, and Simon Ferrier. 2009. "Sample Selection Bias and Presence-Only Distribution Models: Implications for Background and Pseudo-Absence Data." *Ecological Applications* 19(1): 181–197. https://doi.org/10.1890/07-2153.1

Phillips, Steven J., Robert P. Anderson, and Robert E. Schapire. 2006. "Maximum Entropy Modeling of Species Geographic Distributions." *Ecological Modelling* 190(3-4): 231-259. https://doi.org/10.1016/j.ecolmodel.2005.03.026

Pliscoff, Patricio, Federico Luebert, Hartmut H. Hilger, and Antoine Guisan. 2014. "Effects of Alternative Sets of Climatic Predictors on Species Distribution Models and Associated Estimates of Extinction Risk: A Test with Plants in an Arid Environment." *Ecological Modelling* 288(9):166–177. https://doi.org/10.1016/j.ecolmodel.2014.06.003

Poggio, Laura, Enrico Simonetti, and Alessandro Gimona. 2018. "Enhancing the WorldClim Data Set for National and Regional Applications." *Science of the Total Environment* 625: 1628-1643. https://doi.org/10.1016/j.scitotenv.2017.12.258

Qiao, Huijie, Xiao Feng, Luis E. Escobar, A. Townsend Peterson, Jorge Soberón, Gengping Zhu, and Monica Papeş. 2019. "An Evaluation of Transferability of Ecological Niche Models." *Ecography* 42(3): 521–534. https://doi.org/10.1111/ecog.03986

Qiao, Huijie, Jorge Soberón, and A. Townsend Peterson. 2015. "No Silver Bullets in Correlative Ecological Niche Modelling: Insights from Testing among Many Potential Algorithms for Niche Estimation." *Methods in Ecology and Evolution* 6(10): 1126–1136. https://doi.org/10.1111/2041-210X.12397

R Core Team. 2024. "R: A Language and Environment for Statistical Computing." R Foundation for Statistical Computing. 2024. https://www.R-project.org/

Regos, Adrián, Laura Gagne, Domingo Alcaraz-Segura, João P. Honrado, and Jesús Domínguez. 2019. "Effects of Species Traits and Environmental Predictors on Performance and Transferability of Ecological Niche Models." *Scientific Reports* 9(1): 4221. https://doi.org/10.1038/s41598-019-40766-5

Riccaboni, Massimo, and Luca Verginer. 2022. "The Impact of the COVID-19 Pandemic on Scientific Research in the Life Sciences." *PLoS One* 17(2): e0263001. https://doi.org/10.1371/journal.pone.0263001

Romero-Alvarez, Daniel, Luis E. Escobar, Albert J. Auguste, Sara Y. Del Valle, and Carrie A. Manore. 2023. "Transmission Risk of Oropouche Fever Across the Americas." *Infectious Diseases of Poverty* 12(1): 47. https://doi.org/10.1186/s40249-023-01091-2

Rosenberg, Ronald, Michael A. Johansson, Ann M. Powers, and Barry R. Miller. 2013. "Search Strategy Has Influenced the Discovery Rate of Human Viruses." *Proceedings of the National Academy of Sciences USA* 110(34): 13961–13964. https://doi.org/10.1073/pnas.1307243110

Sandel, Brody, L. Arge, Bo Dalsgaard, R. G. Davies, K. J. Gaston, W. J. Sutherland, and J-C. Svenning. 2011. "The influence of Late Quaternary Climate-Change Velocity on Species Endemism." *Science* 334(6056): 660-664. https://doi.org/10.1126/science.1210173

Saupe, E. E., V. Barve, C. E. Myers, J. Soberón, N. Barve, C. M. Hensz, A. T. Peterson, H. L. Owens, and A. Lira-Noriega. 2012. "Variation in Niche and Distribution Model Performance: The Need for a Priori Assessment of Key Causal Factors." *Ecological Modelling* 237–238(7):11–22. https://doi.org/10.1016/j.ecolmodel.2012.04.001

Schmolke, Amelie, Pernille Thorbek, Donald L. DeAngelis, and Volker Grimm. 2010. "Ecological Models Supporting Environmental Decision Making: A Strategy for the Future." *Trends in Ecology & Evolution* 25(8): 479–486. https://doi.org/10.1016/j.tree.2010.05.001

Sillero, Neftalí, Salvador Arenas-Castro, Urtzi Enriquez-Urzelai, Cândida Gomes Vale, Diana Sousa-Guedes, Fernando Martínez-Freiría, Raimundo Real, and A. Márcia Barbosa. 2021. "Want to Model a Species Niche? A Step-By-Step Guideline on Correlative Ecological Niche Modelling." *Ecological Modelling* 456(9): 109671. https://doi.org/10.1016/j.ecolmodel.2021.109671

Singleton, Alyson L., Caroline K. Glidden, Andrew J. Chamberlin, Roseli Tuan, Raquel GS Palasio, Adriano Pinter, Roberta L. Caldeira et al. 2024. "Species Distribution Modeling for Disease Ecology: A Multi-Scale Case Study for Schistosomiasis Host Snails in Brazil." *PLoS Global Public Health* 4(8): e0002224. https://doi.org/10.1371/journal.pgph.0002224

Soberón, Jorge, and Miguel Nakamura. 2009. "Niches and Distributional Areas: Concepts, Methods, and Assumptions." *Proceedings of the National Academy of Sciences USA* 106(supplement_2): 19644–19650. https://doi.org/10.1073/pnas.0901637106

Soley-Guardia, Mariano, Diego F. Alvarado-Serrano, and Robert P. Anderson. 2024. "Top Ten Hazards to Avoid When Modeling Species Distributions: A Didactic Guide of Assumptions, Problems, and Recommendations." *Ecography* 2024(4): e06852. https://doi.org/10.1111/ecog.06852

Springer, Yuri P., David Hoekman, Pieter T. J. Johnson, Paul A. Duffy, Rebecca A. Hufft, David T. Barnett, Brian F. Allan, et al. 2016. "Tick-, Mosquito-, and Rodent-Borne Parasite Sampling Designs for the National Ecological Observatory Network." *Ecosphere* 7(5): e01271. https://doi.org/10.1002/ecs2.1271

Srivastava, Vivek, Valentine Lafond, and Verena C. Griess. 2019. "Species Distribution Models (SDM): Applications, Benefits and Challenges in Invasive Species Management." *CABI Reviews* 2019: 1-13. https://doi.org/10.1079/PAVSNNR201914020

Sunday, Jennifer M., Amanda E. Bates, and Nicholas K. Dulvy. 2012. "Thermal Tolerance and the Global Redistribution of Animals." *Nature Climate Change* 2(9): 686–690. https://doi.org/10.1038/nclimate1539

Synes, Nicholas W., and Patrick E. Osborne. 2011. "Choice of Predictor Variables as a Source of Uncertainty in Continental-Scale Species Distribution Modelling under Climate Change." *Global Ecology and Biogeography* 20(6): 904–914. https://doi.org/10.1111/j.1466-8238.2010.00635.x

United States Geological Survey. 2011. "USGS OFR 2011-1127: Construction of a 3-Arcsecond Digital Elevation Model for the Gulf of Maine, Conversion Factors." Conversion Factors. 2011. https://pubs.usgs.gov/of/2011/1127/convert.html

Valavi, Roozbeh, Gurutzeta Guillera-Arroita, José J. Lahoz-Monfort, and Jane Elith. 2022. "Predictive Performance of Presence-Only Species Distribution Models: A Benchmark Study with Reproducible Code." *Ecological Monographs* 92(1): e01486. https://doi.org/10.1002/ecm.1486

Van de Vuurst, Paige, and Luis E. Escobar. 2023. "Climate Change and Infectious Disease: A Review of Evidence and Research Trends." *Infectious Diseases of Poverty* 12(1): 51. https://doi-org.ezproxy.lib.vt.edu/10.1186/s40249-023-01102-2

Vasconcelos, Rodrigo N., Taimy Cantillo-Pérez, Washington JS Franca Rocha, William Moura Aguiar, Deorgia Tayane Mendes, Taíse Bomfim de Jesus, Carolina Oliveira de Santana, Mariana MM de Santana, and Reyjane Patrícia Oliveira. 2024. "Advances and Challenges in Species Ecological Niche Modeling: A Mixed Review." *Earth* 5(4): 963-989. https://doi.org/10.3390/earth5040050

Waltari, Eric, Ronny Schroeder, Kyle McDonald, Robert P. Anderson, and Ana Carnaval. 2014. "Bioclimatic Variables Derived from Remote Sensing: Assessment and Application for Species Distribution Modelling." *Methods in Ecology and Evolution* 5(10): 1033–1042. https://doi.org/10.1111/2041-210X.12264

Wickham, Hadley. 2016. *ggplot2: Elegant Graphics for Data Analysis*. Second edition. Use R! Cham: Springer International Publishing.

Wieczorek, John, David Bloom, Robert Guralnick, Stan Blum, Markus Döring, Renato Giovanni, Tim Robertson, and David Vieglais. 2012. "Darwin Core: An Evolving Community-Developed Biodiversity Data Standard." *PLoS One* 7(1): e29715. https://doi.org/10.1371/journal.pone.0029715

Wiens, John A., Diana Stralberg, Dennis Jongsomjit, Christine A. Howell, and Mark A. Snyder. 2009. "Niches, Models, and Climate Change: Assessing the Assumptions and Uncertainties." *Proceedings of the National Academy of Sciences USA* 106(2): 19729-19736. https://doi.org/10.1073/pnas.0901639106

Xu, Tingbao, and Michael Hutchinson. 2011. "ANUCLIM Vversion 6.1 User Gguide." Fenner School of Environment and Society, Australian National University.

Xu, Quanli, Xiao Wang, Junhua Yi, and Yu Wang. 2024. "Bias Correction in Species Distribution Models

Based on Geographic and Environmental Characteristics." *Ecological Informatics* 81(7): 102604. https://doi.org/10.1016/j.ecoinf.2024.102604

Zarzo-Arias, Alejandra, Britta Uhl, Daniel S. Maynard, and Manuel B. Morales. 2023. "The Ecological Niche at Different Spatial Scales." *Frontiers in Ecology and Evolution* 11: 1296340. https://doi.org/10.3389/fevo.2023.1296340

Zeng, Yiwen, Bi Wei Low, and Darren C. J. Yeo. 2016. "Novel Methods to Select Environmental Variables in MaxEnt: A Case Study Using Invasive Crayfish." *Ecological Modelling* 34(12):5–13. https://doi.org/10.1016/j.ecolmodel.2016.09.019

Zurell, Damaris, Janet Franklin, Christian König, Phil J. Bouchet, Carsten F. Dormann, Jane Elith, Guillermo Fandos, et al. 2020. "A Standard Protocol for Reporting Species Distribution Models." *Ecography* 43(9): 1261–1277. https://doi.org/10.1111/ecog.04960

Supplemental Material

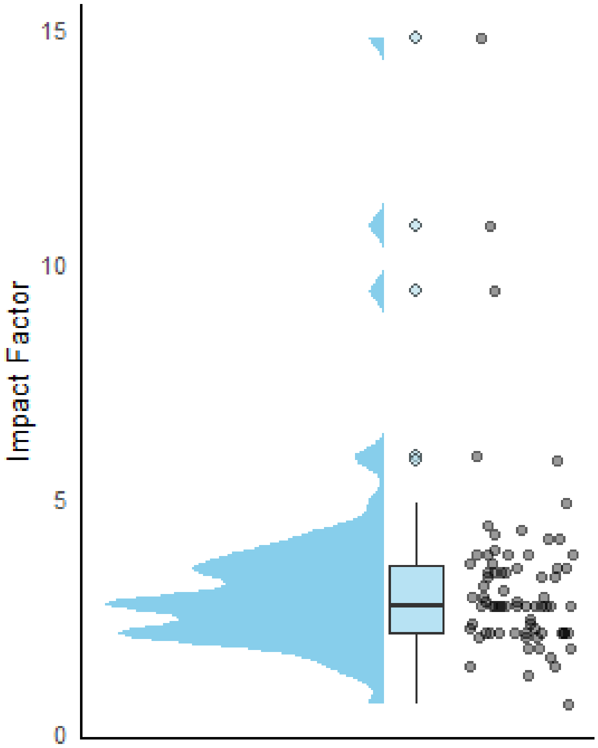


Figure S1: Distribution of Impact Factors of the journals where selected articles for this study were published.

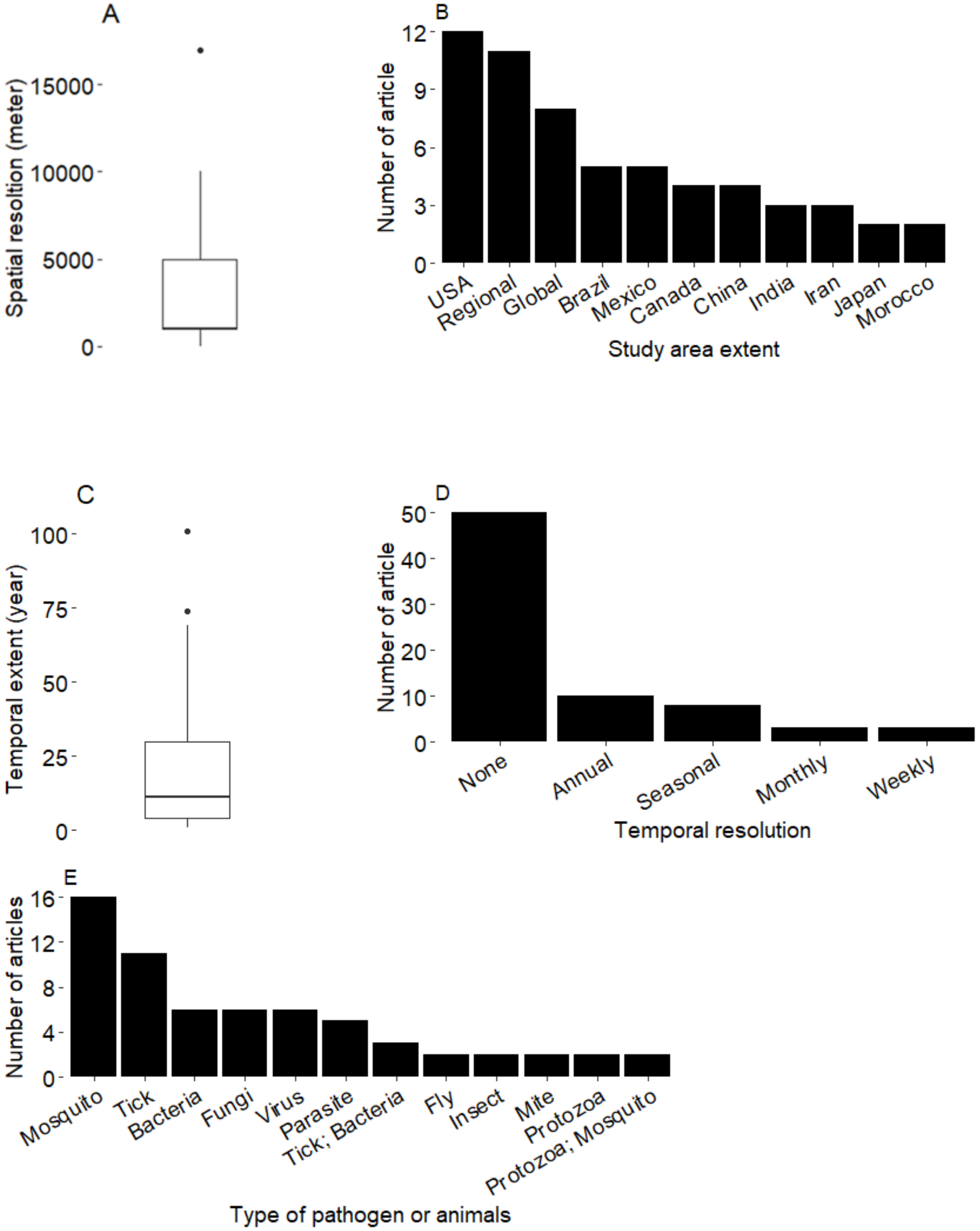


Figure S2: Some of the selected components from the Overview section for a comprehensive understanding. A. Boxplot shows the ranges of spatial resolutions (meters) used in the 78 articles; B. Bar plot shows the spatial extent of the study area mentioned in 78 articles; C. Boxplot shows the temporal extent considered in different articles; D. Bar plot shows the temporal resolutions used in different articles; and E. Bar plot shows the study subjects considered in different articles to study the ecological niche. We plotted the bars having value more than one only.

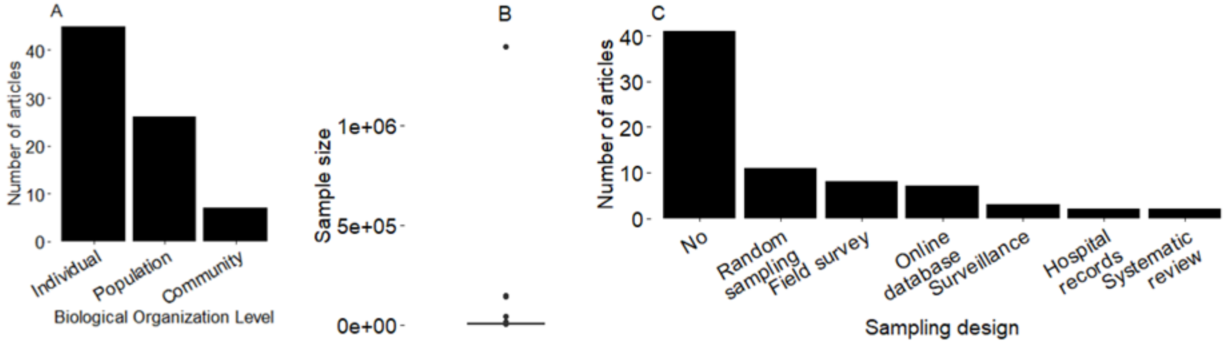
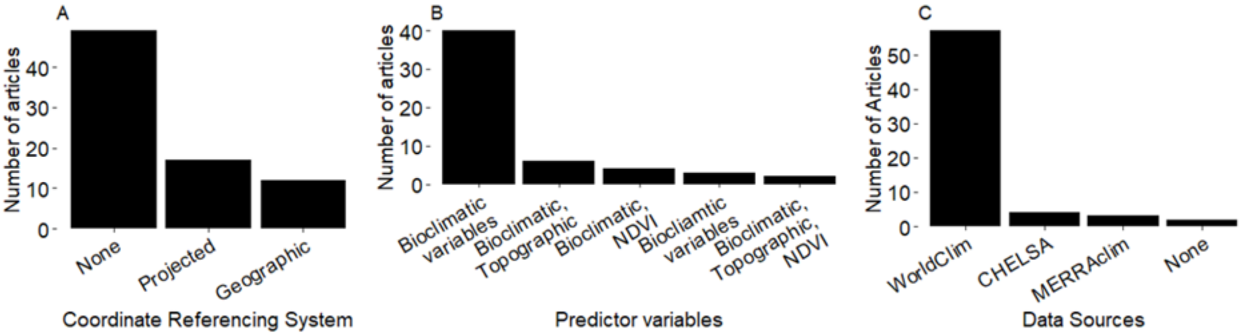


Figure S3.1: Some of the selected components related to biodiversity data from the Data section for a comprehensive understanding. A. Bar plot shows the articles using ecological niche models at different biological organization levels; B. Box plot shows the ranges of sample sizes used in different articles; and C. Bar plot shows the different sampling techniques used in different articles. We plotted the bars having value more than one only.



igure S3.2: Some of the selected components related to predictor data from the Data section for a comprehensive understanding. A. Bar plot shows the articles using different coordinate reference system; B. Bar plot shows the different predictor variables used in different articles for ecological niche model development; and C. Bar plot shows the different sources for the predictor data used in different articles. We plotted the bars having value more than one only.

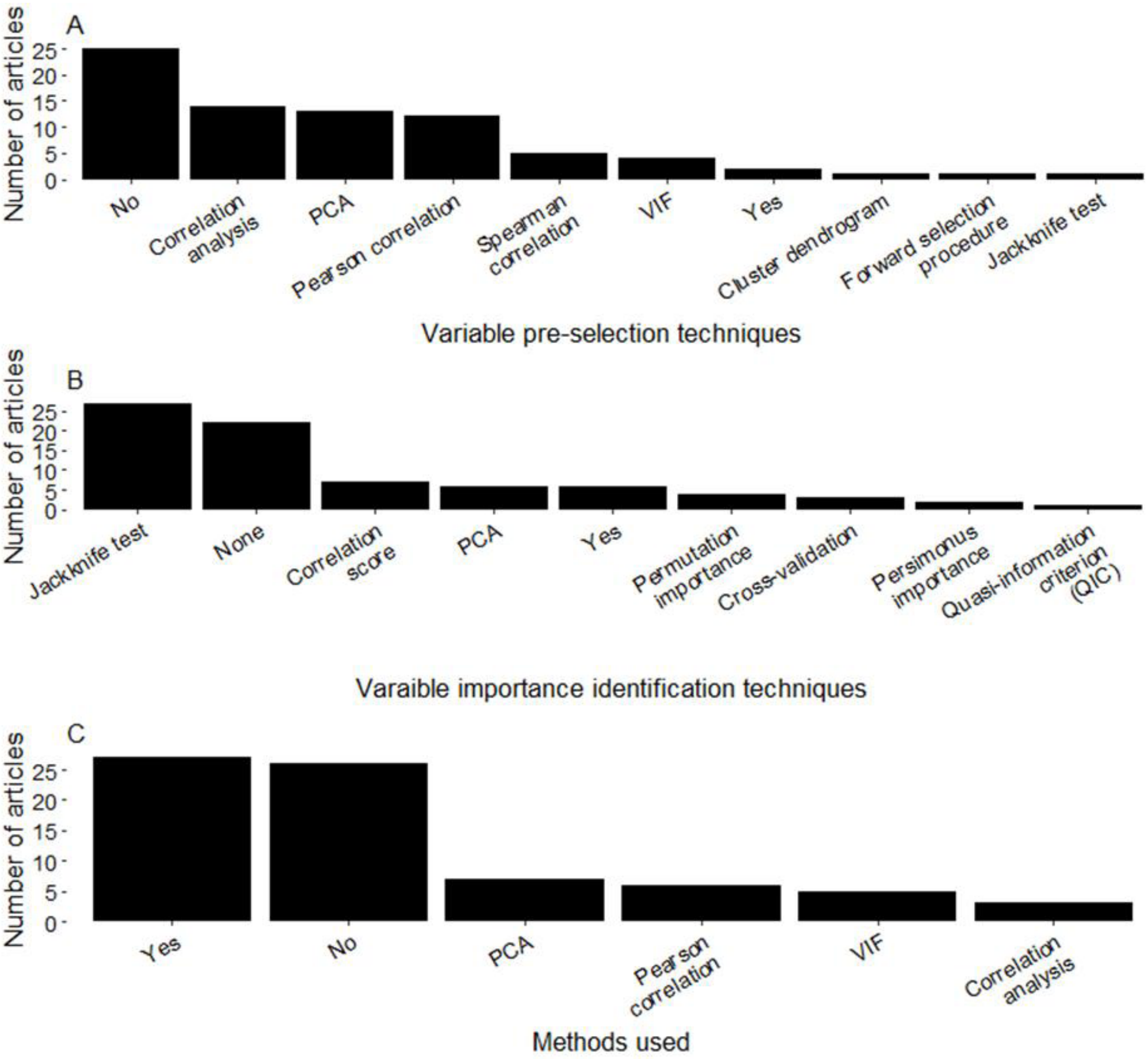


Figure S4: Some of the selected components from the Modeling section for a comprehensive understanding. A. Bar plot shows the articles using different techniques used to preselect the predictor variables; B. Bar plot shows the different techniques used to identify the variable importance in ecological niche modeling in different articles; and C. Bar plot shows the different techniques used to check for multicollinearity among the predictor variable during ecological niche modeling in different articles. We plotted the bars having value more than one only.

Table S1: Different ecological niche modeling techniques used to study the ecological niches of diseases

Serial no	Modeling techniques	Number of articles
1	Maximum Entropy Modeling (MaxEnt)	60
2	Artificial Neural Networks (ANN), Surface Range Envelope (SRE), Flexible Discriminant Analysis (FDA), General Linear Models (GLM), General Additive Models (GAM), General Boosted Models (GBM), Classification Tree Analysis (CTA), Multiple Adaptive Regression Splines (MARS), Random Forests (RF), and MaxEnt	1
3	ANN, SRE, FDA, GLM, GAM, GBM, CTA, MARS, RF, and MaxEnt	1
4	Bayesian additive regression trees (BARTs)	1
5	Bioclimatic envelope, MaxEnt, Logistic regression, and Domain model	1
6	Bioclimatic envelope (BIOCLIM), MaxEn, GLM, MARS, CART, mixture discriminant analysis (MDA), RF, BRT, and GAM	1
7	BRT, GAM, MARS, and Maxent.	1
8	BIOCLIM, and BIOCLIM True/False	1
9	SRE, MaxEnt, GLM, GAM, MARS, CTA, RF, and GBM	1
10	GBM, GLM, RF, and MARS	1
11	GLM, MARS, RF, and MaxEnt	1
12	GLM, GBM, RF, and MaxEnt	1
13	Logistic regression model	1
14	MaxEnt, SVM, Environmental Distance (ED) and Climate Space Model (CSM)	1
15	MAXENT, GLM, CTA, MARS, SVM, GAM, GBM, and ANN	1
16	MaxEnt, RF, and BRT	1
17	GLM, GAM, MaxEnt, R), and BRT	1
18	Simple logistic regression model; RF	1
19	None	1

Serial No	Sources of the predictor variables	Number of articles
	WorldClim	36
	CHELSA	2
	CHELSA; ISRO	1
	CHELSA; ISRIC; MODIS; NASA; SoilGrid; WorldGrids	1
	MERRAclim	1
	MERRAclim; SoilGrids	1
	MERRAclim; NASA; SoilGrids	1
	AAFC; CWI; NRCan	1
	Agri-Geomatics Service of Agriculture and Agri-Food Canada; Agriculture and Agri-Food Canada; Canada Centre for Mapping and Earth Observation	1
	Argentinian Space Agency	1
	AWS Open Data Terrain Tiles; IUCN; PRISM; SRIC SoilGrids	1
	Brazilian Geomorphometric Database; WorldClim	1
	CalEnviroscreen; GAP; USGS	1
	CGIAR-CSI; CGLS; DHS; Ethiopia’s survey data in 2016; ISRIC; SDSM; WorldClim	1
	CGIAR-CSI; Gridded Population of the World, MODIS; WorldClim; WorldGrids	1
	CGIAR-CSI; FAO; HydroSHEDS; WorldClim	1
	CGIAR-CSI; WorldClim	1
	China Resource and Environment Science and Data Center; WorldClim	1
	ClimateNA	1
	Copernicus Global Land Service archive; WorldClim	1
	DEM; DIVA-GIS; ISRIC; MODIS; NASA; NOAA; WorldClim	1
	DEM; DIVA-GIS; IUCN; WorldClim	1
	Digital Soil Mapof the World; SGS; SGS FEWS NET; WorldClim	1
	DTM; WorldClim	1
	Geographical Survey Institute	1
	Global land cover database; WorldClim	1
	Government databases; WorldClim	1
	ISRIC; SoilGrids; WorldClim	1
	MODIS; NLCD; PRISM; TIGER	1
	NASA; SPOT4; WorldClim	1
	National Land Numerical Information	1
	National Mapping Organization; WorldClim	1
	Pakistan Bureau of Statistics; WorldClim	1
	PaleoClim	1
	SEDAC; SRT; WorldClim	1
	Shuttle Radar Topography Mission; WorldClim	1
	Soilgrids; Worldclim	1
	SRTM	1
	SRTM; Worldlcim	1
	USGS; WorldClim	1
	None	2