Received: 30 September 2024 | Accepted: 2 December 2024

DOI: 10.1002/ecm.70003

ECOLOGICAL MONOGRAPHS
ECOLOGICAL SOCIETY OF AMERICA

METHOD

A mechanistic statistical approach to infer invasion characteristics of human-dispersed species with complex life cycle

Nikunj Goel[1] | **Andrew M. Liebhold**[2,3] | **Cleo Bertelsmeier**[4] |
Mevin B. Hooten[5] | **Kirill S. Korolev**[6] | **Timothy H. Keitt**[1]

[1]Department of Integrative Biology, The University of Texas at Austin, Austin, Texas, USA

[2]USDA Forest Service Northern Research Station, Morgantown, West Virginia, USA

[3]Czech University of Life Sciences Prague, Forestry and Wood Sciences, Prague 6, Czech Republic

[4]Department of Ecology and Evolution, Biophore, UNIL-Sorge, University of Lausanne, Lausanne, Vaud, Switzerland

[5]Department of Statistics and Data Sciences, The University of Texas at Austin, Austin, Texas, USA

[6]Department of Physics, Graduate Program in Bioinformatics, and Biological Design Center, Boston University, Boston, Massachusetts, USA

Correspondence
Nikunj Goel
Email: nikunj410@gmail.com

Funding information
National Institutes of Health, Grant/Award Number: R01-GM138530; Operational Programme Research, Development and Education, Grant/Award Number: CZ.02.1.01/0.0/0.0/16_019/0000803; The State Secretariat for Education, Research and Innovation, Grant/Award Number: MB22.00086; National Science Foundation, Grant/Award Numbers: IOS-2222525, BCS-2009669; Stengl-Wyer Endowment, The University of Texas at Austin; Office of the Vice President for Research, Scholarship, and Creative Endeavors, The University of Texas at Austin

Handling Editor: Charles Yackulic

Abstract

The rising introduction of invasive species through trade networks threatens biodiversity and ecosystem services. Yet, we have a limited understanding of how transportation networks determine spatiotemporal patterns of range expansion. This knowledge gap may stem from two reasons. First, current analytical models fail to integrate the invader's life-history dynamics with heterogeneity in human-mediated dispersal patterns. Second, classical statistical methods often fail to provide reliable estimates of model parameters, such as the time and place of species introduction and life-history characteristics, due to spatial biases in the presence-only records and lack of informative demographic data. To address these gaps, we first formulate an age-structured metapopulation model that uses a probability matrix to emulate human-mediated dispersal patterns. The model reveals that an invader spreads radially along the shortest network path, such that the inter-patch network distances decrease with increasing traffic volume and reproductive value of hitchhikers. Next, we propose a hierarchical Bayesian statistical method to estimate model parameters using presence-only data and prior demographic knowledge. To show the utility of the statistical approach, we analyze zebra mussel (*Dreissena polymorpha*) expansion in North America through the inland commercial shipping network. Our analysis suggests that zebra mussels might have been introduced before 1981, indicating a lag of 5 years between the time of introduction and first detection in late 1986. Furthermore, using our statistical model, we estimated a one in three chance that they were introduced near Kingsville (Ontario, Canada), where they were first reported. We also find that survival, fecundity, and dispersal during early life (1–2 years) play a critical role in determining the expansion success of these mollusks. These results underscore the importance of fusing prior scientific knowledge with observation and demographic processes in a Bayesian framework for conceptual and practical understanding of how invasive species spread by human agency.

KEYWORDS
invasion, transportation network, metapopulation, age-structure, bayesian statistics, elasticity analysis

INTRODUCTION

Biological invasions are a major component of the global environmental crisis. Burgeoning trade and commerce have led to a surge in the establishment and spread of invasive species (Hulme, 2009; Meyerson & Mooney, 2007), altering the functioning and composition of global ecosystems. Invasions have resulted in trillions of dollars in damage to human health, infrastructure, and production services (Diagne et al., 2021; Turbelin et al., 2023) and are one of the leading drivers of species extinctions (Bellard et al., 2016). Even more alarming is that the pool of invasive species is expected to increase over the next century as new trade routes emerge (Seebens et al., 2018). Therefore, a major theoretical and practical challenge in invasion science is to understand how human transportation patterns interact with life-history dynamics to determine the spread of an invasive species (Lewis et al., 2016).

Early efforts to model invasions were based on integrodifference equations (IDE) (Kot et al., 1996; Lutscher, 2019) that used fat-tailed redistribution kernels to emulate long-distance dispersal by human transportation (Hallatschek & Fisher, 2014). The IDE models are conceptually attractive because they can be combined with structured demographic models (Neubert & Caswell, 2000) and permit analytical solutions (Bateman et al., 2015). This allows researchers to understand invasion patterns intuitively and provides valuable insights into managing invaders with complex life histories. For example, a Gaussian IDE model predicts a radially (linearly) expanding 2D (1D) invasion front (Kot et al., 1996), such that the species' arrival time is a linear function of distance from the origin (Skellam, 1951). Moreover, managers can conduct sensitivity analyses to identify which life-history parameters should be targeted to slow or stop invasion (Neubert & Caswell, 2000).

Despite the conceptual simplicity of IDE models, they have two major limitations. IDE models assume that species dispersal is directionally random (i.e., isotropic) (Thompson et al., 2021), and the shape of the dispersal kernel does not change from location to location (i.e., spatially invariant) (Hallatschek & Fisher, 2014). However, human-mediated dispersal patterns are highly heterogeneous owing to the complex topology of transportation networks (Banks et al., 2015). Transportation networks are characterized by a local cluster of densely connected neighbors and a few long-distance connections between clusters. This feature creates a small-world structure (Hu & Zhu, 2009), leading to anisotropy and spatial invariance in dispersal patterns (Wolf et al., 2019). Consequently, the patterns of invasions may appear spatially less coherent, and, as such, distance from the origin might not be a suitable predictor of species arrival time (Horvitz et al., 2017).

These limitations of IDE models are well recognized (Hallatschek & Fisher, 2014; Thompson et al., 2021), and several mathematical approaches have been developed to address them (Hui & Richardson, 2017). Among these approaches, metapopulation models are gaining popularity, in part due to their success in describing the spread of infectious diseases through human transportation networks (Brockmann & Helbing, 2013; Colizza et al., 2006; Gautreau et al., 2008; Iannelli et al., 2017; Rvachev & Longini, 1985). Broadly, metapopulation models consist of discrete patches (such as a lake or a county) within which the population changes according to a demographic model and edges that connect these patches to capture heterogeneous dispersal patterns of invaders. Because of this flexibility, metapopulation models have been used to study the expansion of invasive species through a wide range of human transportation mechanisms (Carrasco et al., 2010; Chapman et al., 2016; Ferrari et al., 2014; Floerl et al., 2009; Seebens et al., 2019).

However, these implementations of metapopulation models are conceptually impenetrable because researchers often err on the side of realism by creating parameter-rich computational models that incorporate minute species and transportation-specific details. As a result, it becomes challenging to decipher what aspects of species biology and transportation networks determine the broad-scale patterns of invasions and what general underlying ecological principles characterize these patterns (Brockmann & Helbing, 2013; May, 2004).

Building simple conceptual models of human-mediated invasions is only one part of the challenge. To make these models useful, we also need statistical methods to estimate model parameters (and corresponding uncertainty) to answer questions such as where and when the species was introduced and what makes a given species a successful invader. Answering such questions is critical for formulating management actions to contain and eradicate invaders. However, estimating model parameters is challenging because species distribution data are often fraught with errors and biases (Isaac et al., 2014; Johnston et al., 2023). For instance, most commonly available species occurrence records are gathered by opportunistic sampling rather than structured surveys that report both the presence and absence of species. Such types of opportunistically collected data, referred to as presence-only data, typically suffer from irregular spatial sampling, making it hard to narrow down the location of species introduction. Even when sampling is structured, a species may remain undetected during the establishment phase, making it challenging to infer the time of species introduction (Crooks, 2005).

Another potential issue in estimating parameters may arise if multiple combinations of two or more parameters produce the same ecological outcome. For example, two species with different life-history strategies can generate the same spread patterns (Bateman et al., 2015). This issue of non-identifiability of parameters makes it statistically challenging to infer the life-history tactics of an invader, which, in turn, can limit our ability to determine which life-history parameters play an outsized role in the species spread and, therefore, should be the target of management efforts (Crouse et al., 1987).

Recent theoretical and computational advances in applied Bayesian statistics provide avenues to overcome these hurdles (Gelman et al., 1995; Hobbs & Hooten, 2015; Hooten & Hefley, 2019; McElreath, 2020). Bayesian models learn parameters using a data model and a parameter model. The data model (or the likelihood) is a probabilistic generative model of observed data. For invasions, this generative model is a multi-level process in which the occurrence data are linked to latent parameters, such as invaders' abundance, via an observation model that describes how the data were collected. These abundances are then, in turn, linked to demographic parameters via a mechanistic model of invasion (Hooten et al., 2007; Wikle, 2003). This hierarchy offers two advantages. First, the information from data to demographic parameters flows through an observation model where one can explicitly express the generative process of how spatial biases and detection errors arise in the presence-only data. Second, because the information obtained from the data is imprecise due to the probabilistic nature of the generative process, the hierarchy allows biogeographers to propagate the uncertainty in the data model to inform parameter estimates. This may allow one to obtain a reliable estimate of parameters, such as where and when the species was introduced in the alien territory, even when sampling is spatially unstructured.

In contrast to the data model, the parameter model (i.e., prior) characterizes our uncertainty about the parameters before the data are collected. Intuitively, a parameter model is an expression of our prior scientific knowledge. For example, population biology dictates that a successful invader will have a positive growth rate. Therefore, we can constrain the growth rate to take only positive values a priori. Similarly, a biologist may have additional information about the invader or a closely related species, such as its life-history parameters, based on previous studies on the invader in its native range or a closely related species. By including this information via a parameter model, Bayesian models can disentangle otherwise non-identifiable parameters (Neath & Samaniego, 1997). Together, the data and parameter models provide a rigorous framework to estimate model parameters and quantify the uncertainty in those parameters, which can be used to inform probabilistic predictions and decision-making.

In this study, we aim to fill the aforementioned theoretical and statistical gaps to understand the spread of invasive species by human movement. We first formulate an analytically tractable age-structured metapopulation model to gain a conceptual understanding of how complexity in human transportation networks determines the spatiotemporal patterns of species spread. We ask how the properties of the transportation network and species' life-history tactics determine the invasion path and arrival time. Next, we propose a hierarchical Bayesian model to estimate the parameters of the metapopulation model using presence-only data and prior scientific knowledge. We tailor our mathematical and statistical models to analyze the expansion of zebra mussels (*Dreissena polymorpha*) in the inland waterways of North America via the commercial shipping network. We identify the time and place of zebra mussel introduction, the invasion route, and the life-history parameters that are most critical for their spread. Despite our focus on zebra mussels, the conceptual insights offered by the metapopulation model and the statistical guidance to analyze empirical data are general and broadly applicable.

Zebra mussels

Zebra mussels are freshwater bivalves native to Eastern Europe and were likely introduced in the Great Lakes by the release of contaminated ballast water (Nalepa & Schloesser, 2013). Post introduction, zebra mussels have dramatically altered North America's water ecology and local economy. These mollusks are filter feeders that primarily consume planktonic algae and microzooplankton, thereby changing the energy flow and structure of food webs in water bodies where they reside. They are notorious biofouling organisms (Ludyanskiy et al., 1993)—their active byssal apparatus allows them to attach to hard substrates at very high densities that can impede the function of human infrastructure, resulting in an annual loss of half a billion dollars (Warziniack et al., 2021). Due to these negative impacts on biodiversity and the economy, zebra mussels have earned a spot in the International

Union for Conservation of Nature's list of 100 of the world's worst invaders (Lowe et al., 2004). Therefore, understanding the expansion characteristics of zebra mussels can help us slow or stop their spread and may provide insights into managing other emerging invasions.

However, due to detection errors and unstructured sampling of zebra mussels, tracing the invasion pathway and identifying the time and place of their introduction are challenging. They were first detected during a large-scale survey in 1988 in Lake St. Clair (Hebert et al., 1989). However, the rediscovery of a 1986 letter exchange between Pembina Explorations Ltd. and the Ontario Ministry of Natural Resources suggests that a colony of mussels was present in Lake Erie near Kingsville, Ontario, Canada (Carlton, 2008). In subsequent years, zebra mussel detections spiked at distant locations across the Great Lakes (GBIF.org, 2022; USGS, 2022) (Appendix S1: Figure S1). By 2010, they were detected and established in most of the inland commercial waterways from the Great Lakes to the Gulf of Mexico, implicating hitchhiking by hull fouling of barges and tugs as putatively the primary mechanism of continental-scale dispersal (Johnson & Padilla, 1996). These established populations in commercial waterways are, in turn, fueling over-land range expansion to isolated water bodies through recreational boating (Johnson & Carlton, 1996).

FORMULATION AND RESULTS OF THE METAPOPULATION MODEL

To model the continental-scale range expansion of zebra mussels in North America, we consider an environmentally homogeneous metapopulation with k patches (or locations) connected by the commercial shipping network. For zebra mussels, these locations can correspond to a port or a collection of ports in some fixed area. At each location, the zebra mussel population is structured by age, such that the population dynamics are governed by the following matrix difference equation (Hunter & Caswell, 2005):

$$\boldsymbol{N}_{t+1,i} = \boldsymbol{A}\left[\boldsymbol{N}_{t,i} + \gamma \boldsymbol{B} \sum\nolimits_{j \neq i}^{k} q_{ji} \boldsymbol{N}_{t,j} - q_{ij} \boldsymbol{N}_{t,i}\right]. \tag{1}$$

$\boldsymbol{N}_{t,i}$ is a vector of the population in different age classes at location i and time t, γ is the total number of voyages that ships take in one-time step, $\boldsymbol{B}$ is a diagonal matrix with elements corresponding to age-dependent hitchhiking rate, and $\boldsymbol{A}$ is the Leslie matrix that captures age-dependent life-history schedules of zebra mussels such as reproduction, survival, and mortality (Leslie, 1945). When the invader population is structured by ontogenetic stages, we replace the Leslie matrix with the Lefkovitch matrix to capture stage-dependent life-history events (Lefkovitch, 1965). Dispersal of propagules is weighted by q_{ij}, where q_{ij} is the fraction of voyages from patch i to patch j, such that $\sum_i \sum_{j \neq i} q_{ji} = 1$.

Using port call data from MarineTraffic, we estimate weights, q_{ij}, at a 50-km resolution and simulate the hypothetical spread of zebra mussels originating near Kingsville (Ontario, Canada) (brown point in Figure 1C), where mussels were first detected in late 1986 (Carlton, 2008). We find that the arrival time of the species is correlated with river distance from Kingsville ($R^2 = 0.89$; Figure 1A), with an expansion from the Great Lakes to the Gulf of Mexico (Figure 1C). Although these patterns are broadly consistent with the predictions of 1D Gaussian IDE model (Kot et al., 1996), we can gain a better understanding of invasion patterns (e.g., route of invasion and arrival time) by exploiting the hidden underlying structure of the transportation network (Brockmann & Helbing, 2013; Iannelli et al., 2017) and elucidating how the network structure interacts with species life-history characteristics.

To do that, we assume that for human-dispersed species, the dispersal inputs to a population contribute much less than local reproduction. This allows us to approximate $\boldsymbol{N}_{t,i}$ as a scalar multiple of the stable-age distribution ($\boldsymbol{w}$; proportion of individuals in each age class), $\boldsymbol{N}_{t,i} \approx x_{t,i} \boldsymbol{w}$, where

$$\boldsymbol{A}\boldsymbol{w} = \lambda \boldsymbol{w}, \tag{2}$$

and λ is the species growth rate (Caswell, 2001). Intuitively, this approximation allows us to project Equation (1) onto $\boldsymbol{w}$, which simplifies the matrix difference equation to a scalar difference equation:

$$x_{t+1,i} = \lambda\left[x_{t,i} + D_e \sum\nolimits_{j \neq i}^{k} q_{ji} x_{t,j} - q_{ij} x_{t,i}\right]. \tag{3}$$

In Equation (3),

$$D_e = \gamma \frac{\boldsymbol{v}^T \boldsymbol{B} \boldsymbol{w}}{\boldsymbol{v}^T \boldsymbol{w}} \tag{4}$$

is effective dispersal, and $\boldsymbol{v}^T$ is the reproductive value (total number of offspring born to an individual from their current age onward), which is given by

$$\boldsymbol{v}^T \boldsymbol{A} = \lambda \boldsymbol{v}^T. \tag{5}$$

We use the superscript T to denote the transpose of a vector or a matrix (see Appendix S1).

The scalar approximation provides an intuitive way to interpret the metapopulation model by partitioning the population dynamics into two hierarchical levels. The upper

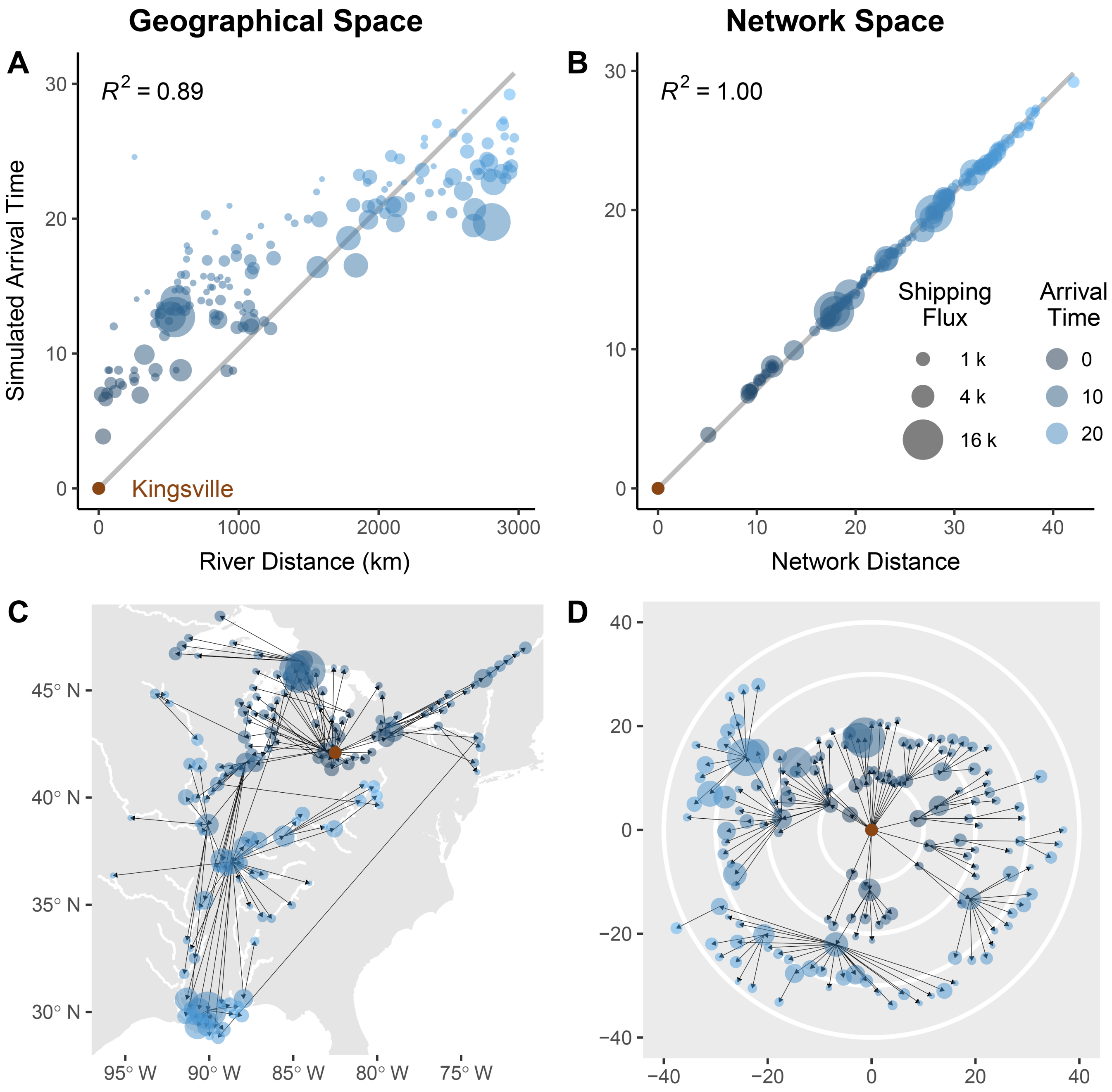

FIGURE 1 A deterministic simulation of the spread of zebra mussels originating from Kingsville, Ontario (brown point), on the geographical and network space. On the network space, (B) the shortest network distance from the origin (Equation 9) shows an almost perfect linear association with the species arrival time (Equation 10). (D) Visualizing species spread on the network space yields radially expanding wavefront—reminiscent of integrodifference equation models (Kot et al., 1996). In contrast, on the geographical space, (A) the river distance from Kingsville is a relatively weaker predictor of the time of zebra mussel arrival, and (C) the pattern of expansion is spatially less coherent. In plots (A) and (B), the best-fit line is constrained to pass from zero because, at time zero, the invader is at the location of the introduction.

level of the hierarchy is the spatial dynamics model (Equation 3), describing how a population with no age structure spreads along the transportation network. The lower level of the hierarchy is the life-history model (Equations 2, 4, and 5), describing the age-dependent life-history strategies of an individual, such as the probability of survival, probability of dispersal, and fecundity. The conceptual advantage of using the scalar approximation is the realization that the age-structured metapopulation model in Equation (1) effectively yields the same spread patterns as the spatial dynamics model with no age structure in which the growth (λ) and dispersal (D_e) rates correspond to the dominant eigenvalue of the Leslie matrix (Equations 2 and 5) and mean age-specific propagule pressure weighted by the reproductive value (Equation 4), respectively. Thus, the life-history model, expressed in terms of age-dependent life-history parameters, produces scalar outputs λ and D_e, which are inputs to the spatial dynamics model. This link between population level (D_e and λ) and life-history parameters ($\boldsymbol{A}$ and $\boldsymbol{B}$) highlights the importance of incorporating age or stage structure in invasion models (Bosch et al., 1990; Bosch et al., 1992; Neubert & Caswell, 2000). For example, everything else being equal, species for which individuals disperse at their peak reproductive value (e.g., age of maturation) spread faster due to higher effective dispersal (D_e).

To simplify the model further, we define $\boldsymbol{x}_t^T = [x_{t,1}\ x_{t,2}\ \dots\ x_{t,k}]$, which allows us to express the joint population change in all locations as

$$\boldsymbol{x}_{t+1}^T = \lambda \boldsymbol{x}_t^T \boldsymbol{P}, \tag{6}$$

where $\boldsymbol{P}$ is a probability matrix. The non-diagonal elements of the matrix, $P_{ij} = D_e q_{ij}$, correspond to the proportion of mussels dispersing from patch i to j in one time step. Note that the structure of the metapopulation model in Equation (6) bears a remarkable resemblance with IDE models (Kot et al., 1996). In both models, the population at the next time step is a product of population growth rate and dispersal. In IDE models, dispersal is captured using a redistribution kernel, which integrates to unity. Similarly, in the metapopulation model, we capture dispersal using a probability matrix $\boldsymbol{P}$ (hereafter referred to as the redistribution matrix) in which each row sums to unity. But, unlike redistribution kernels, the redistribution matrix can capture spatial invariance and anisotropy in dispersal patterns.

Next, assuming $\boldsymbol{N}_{t_0,n}$ mussels are introduced at location n at time t_0, the population after t time steps is given by

$$\boldsymbol{x}_t^T = x_{t_0} \lambda^t \boldsymbol{e}_n^T \boldsymbol{P}^t, \tag{7}$$

where $x_{t_0} = \boldsymbol{v}^T \boldsymbol{N}_{t_0,n} / \boldsymbol{v}^T \boldsymbol{w}$ is the effective initial population and $\boldsymbol{e}_n^T$ is a row vector with all zero elements except at index n, where it takes value one. The population change is a product of exponential growth, $x_{t_0}\lambda^t$, and a t-step transition probability, $\boldsymbol{e}_n^T \boldsymbol{P}^t$, that encapsulates the dispersal trajectory of zebra mussels by the inland shipping network (Kulkarni, 2016). Intuitively, $\boldsymbol{e}_n^T \boldsymbol{P}^t$ is the probability of finding a zebra mussel at a given location after t time steps via any path originating from n.

A natural question is whether multiple paths contribute to zebra mussel arrival because they might be equally likely or if a single path has the dominant contribution to the expansion process? And, if the latter were the case, what would determine this path? In principle, the invader can arrive at a location by infinite paths. However, not all paths are equally probable. Human transport networks are highly heterogeneous, so there is usually a single fastest path to the destination (Brockmann & Helbing, 2013; Gautreau et al., 2008).

To understand what determines this path, we define network distance from location i to j as

$$d_{ij} = \mathbb{W}\left(\frac{\log\lambda}{P_{ij}}\right) \approx \log\left(\frac{\log\lambda}{D_e}\right) - \log q_{ij} - \gamma_e, \tag{8}$$

where $\mathbb{W}$ is the Lambert W function and γ_e is the Euler–Mascheroni constant (Gautreau et al., 2008). This definition of distance emerges naturally from the theory of discrete stochastic processes (Kulkarni, 2016) and provides an intuitive way to visualize and understand human-mediated invasions. For one, this notion of distance suggests that, on the network space, two locations are close if they exchange a high traffic volume and the hitchhikers have high reproductive value (i.e., high D_e).

Now consider a set of all paths from n (origin) to m (destination)—Γ_{nm}. We postulate that the zebra mussel will arrive at m via the shortest network path $\left(\Gamma_{nm}^{SP_1}\right)$, which maximizes the product of dispersal probabilities along the path (Bagnara et al., 2022; Brockmann & Helbing, 2013; Gautreau et al., 2008; Iannelli et al., 2017). As such, the total network distance traveled is given by

$$D_{nm}^{SP_1} = \sum_{(i,j)\,\in\,\Gamma_{nm}^{SP_1}} d_{ij}, \tag{9}$$

and the time of the species' arrival is

$$\tau_{nm}^{SP_1} = \frac{D_{nm}^{SP_1}}{\log\lambda}. \tag{10}$$

Although mussels can arrive by alternate paths, their contribution diminishes exponentially with increasing network distance such that the shortest path is the major contributor to the most probable distance.

$$D_{nm}^{MP} = -\log\sum\nolimits_i e^{-D_{nm}^{SP_i}}, \tag{11}$$

where $D_{nm}^{SP_i}$ is the network distance along the ith shortest path from n to m (Iannelli et al., 2017).

Indeed, when we simulate the arrival time of zebra mussels from all possible paths, we find a strong linear relationship between the shortest network distance from the origin and arrival time ($R^2 = 1.00$; Figure 1B), with a mean error of about 2% (Appendix S1: Figure S2A). This supports our assumption that even though a species may travel along multiple paths due to the heterogeneity in the human transportation networks, the shortest network path is the primary route of invasion. In fact, the linear relation also holds even when we consider an individual-based model with stochastic age-dependent dispersal, reproduction, and survival to simulate arrival time (Appendix S1: Figure S2B) (also see Jamieson-Lane & Blasius, 2020 for theoretical results on the stochastic version of the model). These findings explain why the physical distances may be less appropriate to infer species arrival time (Padilla et al., 1996) (Figure 1A). And, interestingly, visualizing species spread on the network space yields a radially expanding wavefront (Figure 1D)—reminiscent of IDE models—even though on the geographical space, the spread patterns may appear less coherent (Brockmann & Helbing, 2013; Horvitz et al., 2017).

FORMULATION OF THE HIERARCHICAL BAYESIAN MODEL

To estimate the parameters of the metapopulation model, we specify a hierarchical Bayesian model with three levels—(1) an observer model that describes how variation in "true" abundance translates to invader presence data; (2) a spatial dynamics model that describes spatio-temporal variation in "true" abundances of the invader; (3) a life-history model to describe survivorship, fecundity, and dispersal schedules of the species. These sub-models form a hierarchy in that the output of the life-history model is the input for the spatial dynamics model, whose output is the input of the observer model (Pagel & Schurr, 2012).

In a Bayesian framework, this hierarchy is implemented by partitioning the likelihood into the product of probabilities corresponding to each sub-model by conditioning (Berliner, 1996). Then, using the Bayes rule, the posterior distribution of parameters is expressed as the product of the likelihood (informed by data) and the prior (informed by beliefs based on expert opinion or scientific experiments conducted by others). This posterior is a high-dimensional joint probability distribution of parameters, which is used to infer parameter expectations and uncertainties (Gelman et al., 1995). We follow the recommended Bayesian workflow guidelines to specify the model, evaluate the output, and make inferences (Gelman et al., 2020). We outline the main components of the three sub-models in Figure 2 and summarize how these sub-models are linked hierarchically in the section below. For the full statistical model, including implementation details, statements, relationships between parameters, and priors, see Appendix S1: Table S1.

Observer model

To characterize the presence-only data for zebra mussels (GBIF.org, 2022; USGS, 2022), we define an indicator data variable $y_{t,i}$ that takes value one if mussels are reported within a 25-km radius of location i at time t and zero otherwise. When $y_{t,i}=1$, the location is sampled, and mussels are detected. However, when $y_{t,i}=0$, it can imply one of two things—either a location was not sampled or, when sampled, the mussels were not detected. Implicitly, this data model neglects false positives, which may be reasonable as these datasets are subject to quality controls before cataloging. We model this data-generative process as a mixture of two components (see Miller et al., 2019). Location i is sampled with probability p_s^i and, given that the location is sampled, mussels are detected with probability $p_d^{t,i}$ (see observer model in Figure 2). Cumulatively, this yields

$$y_{t,i} \sim \text{Bernoulli}\left(p_s^i p_d^{t,i}\right). \tag{12}$$

We assume sampling probability is constant over time but can vary with location, such that nearby locations have similar sampling probabilities. To account for this spatial information, we model the joint distribution of sampling probabilities across locations as a multivariate normal distribution.

$$[\beta_1 \, \beta_2 \ldots \beta_k]^T \sim \text{MVNormal}\left(\mu_\beta, \boldsymbol{K}\right), \tag{13}$$

where $\text{logit}^{-1}(\beta_i)=p_s^i$ is the sampling probability at location i, $\text{logit}^{-1}\left(\mu_\beta\right)$ is the mean sampling probability on the logit scale, and K_{ij} is the covariance between location i and j. We use an RBF kernel

$$K_{ij}=\eta^2 e^{-D_{ij}^2/2\rho^2}+\delta_{ij}\sigma^2, \tag{14}$$

to model covariance, D_{ij} is the geographical distance between locations i and j, and δ_{ij} is the Kronecker delta function. The hyperparameter ρ is the length scale of correlation, η is the maximum covariance between two locations, and σ provides extra covariance when $i=j$. Modeling spatial correlation in sampling probabilities using a multivariate normal distribution allows partial pooling, thus providing a more robust estimate of sampling probabilities for locations where data are insufficient. Assuming π_p is the per-mussel detection

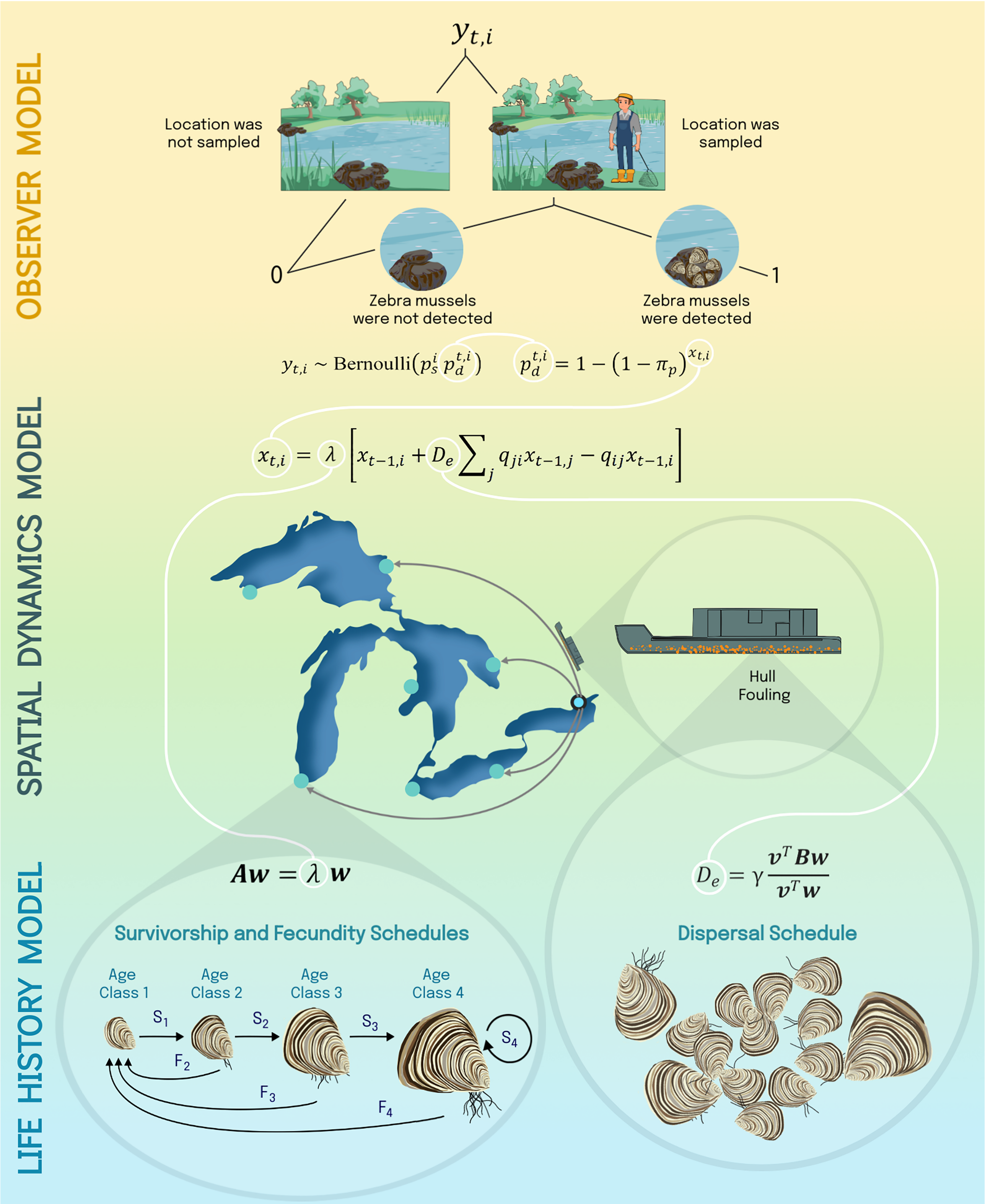

FIGURE 2 A conceptual diagram describing the hierarchical structure of the Bayesian model. The observer model (top) links the presence data, $y_{t,i}$, with zebra mussel abundance, $x_{t,i}$ (Equations 12 and 15). The spatial dynamics model (middle) describes population change due to local growth and dispersal by the shipping network (Equation 16). Finally, the life-history model (bottom) describes how zebra mussel survival, fecundity, and dispersal schedules ($\boldsymbol{A}$ and $\boldsymbol{B}$) determine the population level parameters—growth rate, λ, and effective dispersal, D_e (Equations 2 and 4). These sub-models form a hierarchy in the sense that the output of the lower level is the input of the level above (Pagel & Schurr, 2012). The illustration was made by Upasana Sarraju.

probability, the detection probability of the species at location i and time t is

$$p_d^{t,i} = 1 - \left(1 - \pi_p\right)^{x_{t,i}}, \tag{15}$$

where $x_{t,i}$ is the local abundance of mussels. The joint model of sampling and detection allows us to estimate heterogeneity in sampling as well as the zebra mussel abundance as it expands.

Spatial dynamics model

To model changes in the abundance of mussels (via all routes), we use the spatial dynamics model in Equation (6) (see spatial dynamics model in Figure 2), with a slight modification—at high density, the population saturates to the carrying capacity. We achieve this by reformulating the spatial dynamics model as

$$\boldsymbol{x}_{t+1}^T = \boldsymbol{x}_t^T \boldsymbol{G}_t \boldsymbol{P}, \tag{16}$$

where $\boldsymbol{G}_t = \mathrm{diag}\left(\left[\lambda^{1-x_{t,1}/K} \ \lambda^{1-x_{t,2}/K} \ \ldots \ \lambda^{1-x_{t,k}/K}\right]^T\right)$ is the Ricker growth matrix (Ricker, 1954) with carrying capacity K. Although density dependence does not affect arrival statistics because of exponential growth at the expanding front, it helps avoid numerical pathologies during inference. Recall, $P_{i \neq j} = D_e q_{ij}$, where $D_e = \gamma \boldsymbol{v}^T \boldsymbol{B} \boldsymbol{w} / \boldsymbol{v}^T \boldsymbol{w}$. We assume x_{t_0} mussels are introduced at n such that $\boldsymbol{x}_{\boldsymbol{t_0}}^T = x_{t_0} \boldsymbol{e}_n^T$.

Next, by setting the effective initial population to one (i.e., $x_{t_0} = 1$), we can estimate the earliest time of introduction. Unfortunately, the nature of the observation and spatial dynamics models do not allow us to infer t_0. At carrying capacity, the detection probability is approximately $1 - e^{-\pi_p K}$. Therefore, even though the product of π_p and K is identifiable, these parameters are individually non-identifiable, which implies that we can only estimate local mussel abundances relative to carrying capacity. In other words, we cannot estimate x_{t_0} and, consequently, cannot identify t_0. We can remedy this situation through a parameter model by constraining carrying capacity. Studies estimate that the carrying capacity of quagga mussels (a species similar to zebra mussel in its biological characteristics) varies between 10 and 100 thousand individuals/m^2 (Cross et al., 2011). This estimate of carrying capacity is consistent with the observed fouling density of zebra mussels (Griffiths et al., 1991). Using quagga mussel carrying capacity as a prior for zebra mussel carrying capacity, we can calculate the time, τ_0, when the population at origin is $10^{-9}K$ or 1–10 individuals/km^2. Assuming $\tau_0 > t_0$, we use τ_0 as a proxy for the time of zebra mussel introduction.

Life-history model

We model local patch dynamics using a Leslie matrix $\boldsymbol{A}$ with four age classes of width 1 year,

$$\boldsymbol{A} = \begin{pmatrix} F_1 & F_2 & F_3 & F_4 \\ S_1 & 0 & 0 & 0 \\ 0 & S_2 & 0 & 0 \\ 0 & 0 & S_3 & S_4 \end{pmatrix}, \tag{17}$$

and diagonal hitchhiking matrix $\boldsymbol{B} = \mathrm{diag}\left(\left[B_{11} \ B_{22} \ B_{33} \ B_{44}\right]^T\right)$ (see the life-history model in Figure 2). Because multiple combinations of life-history parameters yield the same λ and D_e, the presence records cannot constrain the elements of the Leslie ($\boldsymbol{A}$) and hitchhiking ($\boldsymbol{B}$) matrices. We circumvent this limitation by assigning informative priors on age-specific survivorship probability (S_i) and fecundity (F_i) based on estimates from literature (Casagrandi et al., 2007) (Appendix S1: Table S1). We also constrain the estimates of age-specific hitchhiking rate ($\boldsymbol{B}$) using published data on the shell length of randomly sampled zebra mussels on a commercial vessel (Keevin et al., 1992). Based on the allometric relationship between zebra mussels' age and shell length, the data suggest that attached mussels were older than 1 year. Therefore, we assume that the hitchhiking rate is very low for age class 1 ($B_{11} \approx 0$) and remains constant after ($B_{22} = B_{33} = B_{44} = B$). We also consider an alternate hypothesis of a constant hitchhiking rate for all age classes (Appendix S1: Figure S3).

RESULTS OF THE HIERARCHICAL BAYESIAN MODEL

We implement the above statistical model in the Stan language (Carpenter et al., 2017). Internally, Stan samples from a probability distribution using the no-U-turn sampler (Hoffman & Gelman, 2014), which is an efficient implementation of the Hamiltonian Monte Carlo MCMC algorithm (Neal, 2011).

Model testing

The choice of prior can play a notable role in Bayesian models (Gelman et al., 2020). On the one hand, strong priors can resolve non-identifiable parameters such as the elements of the hitchhiking and Leslie matrix. On the other hand, vague priors on transformed parameters can lead to unrealistic expectations on the outcome scale. For example, vague hyperpriors of the covariance function (Equation 13) can lead to a highly informative U-shape prior on the probability scale, thereby biasing estimates of sampling probabilities. For these reasons, we evaluate and refine our priors by simulating invasion outcomes using parameters drawn from the prior distribution. Our choice of prior yields biologically realistic invasion outcomes and parameter values.

Next, to test if the statistical model works as intended, we generated synthetic data according to the generative model described above (Equations 12–17) and then fit the statistical model to the synthetic data. Diagnostics reveal that the MCMC chains converge and yield reliable estimates of means and interquartile ranges. For most parameters, the actual value lies within the 95% credible interval. Posterior predictive checks confirm that the predicted time of first detection (Figure 3A) and the number of zebra mussel detections at a location (Figure 3C) are consistent with synthetic data. Importantly, we can reasonably predict the time (using τ_0 as a proxy; Figure 4A) and the location of the introduction (Figure 4C; actual values are marked with an asterisk).

Data analysis

We fit the statistical model to zebra mussel occurrence data recorded from January 1980 to December 2019 (GBIF.org, 2022; USGS, 2022). Diagnostics reveal that the MCMC chains converge and yield reliable estimates of parameter means and interquartile ranges. To evaluate the model fit, we performed posterior predictive checks. We found that the predicted time of first detection and the number of zebra mussel detections at a location are consistent with real data (Figure 3B,D).

The model fit provides two insights about the initial phases of zebra mussel expansion. First, the bulk of the posterior distribution of τ_0—the time when the effective population size ($x_{\tau_0,n}$ or $\boldsymbol{v}^T \boldsymbol{N}_{\tau_0,n} / \boldsymbol{v}^T \boldsymbol{w}$) of the zebra mussels at the location of introduction is between 2.5 and 25 k (or 1–10 individuals/km^2)—lies before 1981 (Figure 4B). Assuming that the initial size of propagules introduced at the location of origin was less than the number of zebra mussels at time τ_0, the statistical model suggests that these mollusks might have established before 1981 and went undetected for at least 5 years until late 1986 (Carlton, 2008). Based on the model assumption of deterministic exponential growth, the observed lag between the time of introduction (inferred using τ_0 as a proxy) and first detection can be attributed to detection failure at low abundances (Crooks, 2005).

Second, we find that zebra mussels were likely introduced in Lake Erie (Figure 4D) and not in Lake St. Clair (Hebert et al., 1989). However, within Lake Erie, the statistical model and occurrence data are consistent with multiple regions as possible locations for origin (Figure 4C). In fact, there is only a one in three chance that mussels were introduced near Kingsville, where they were first detected (Carlton, 2008) (Figure 4D). This feature is also present in synthetic data simulations: The statistical model attributed a nonzero probability of introduction at a location that was not the true location of introduction (true location is marked by an asterisk in Figure 4C). This uncertainty in the location of origin stems from uneven spatial sampling, as zebra mussels could have been introduced at a less frequently sampled location and later dispersed and detected at a more frequently sampled location.

Next, we numerically performed elasticity analysis to ask how small perturbations in life-history parameters ($\boldsymbol{A}$ and $\boldsymbol{B}$) change the arrival time. We find that the arrival time (Equation 10) is most sensitive to the transition probability from age class 1–2 (S_1), the hitchhiking rate of individuals in age class 2 (B_{22}), and the fecundity of individuals in age class 2 (F_2). Specifically, a 1% change in these parameters (S_1, B_{22}, and F_2) will slow the expansion rate by approximately 0.1% (Figure 5A). Managers can use this information to formulate targeted management policies that maximize positive ecological outcomes (e.g., slow the spread rate) for a specified amount of monetary resources. Elasticity analysis also reveals another important and often neglected aspect of species spread—not all age groups contribute equally to species dispersal. The stable age distribution for zebra mussels is skewed toward early life stages, while reproductive value is skewed toward later life stages (Appendix S1: Figure S4). Cumulatively, these two trends, with the assumption that the hitchhiking rate of mussels is very low for age class 1 (Keevin et al., 1992), suggest that age class 2 is the main contributor to effective dispersal (Figure 5B).

DISCUSSION

We show that the metapopulation dynamics of an invader (Equation 1) can be simplified into a redistribution matrix with an underlying exponential growth model (Equation 6), such that the effective dispersal (D_e) and growth rates (λ) are a function of life-history parameters (Equations 2 and 4). This hierarchical formulation of the metapopulation model offers several advantages for conceptually understanding drivers of human-mediated invasions.

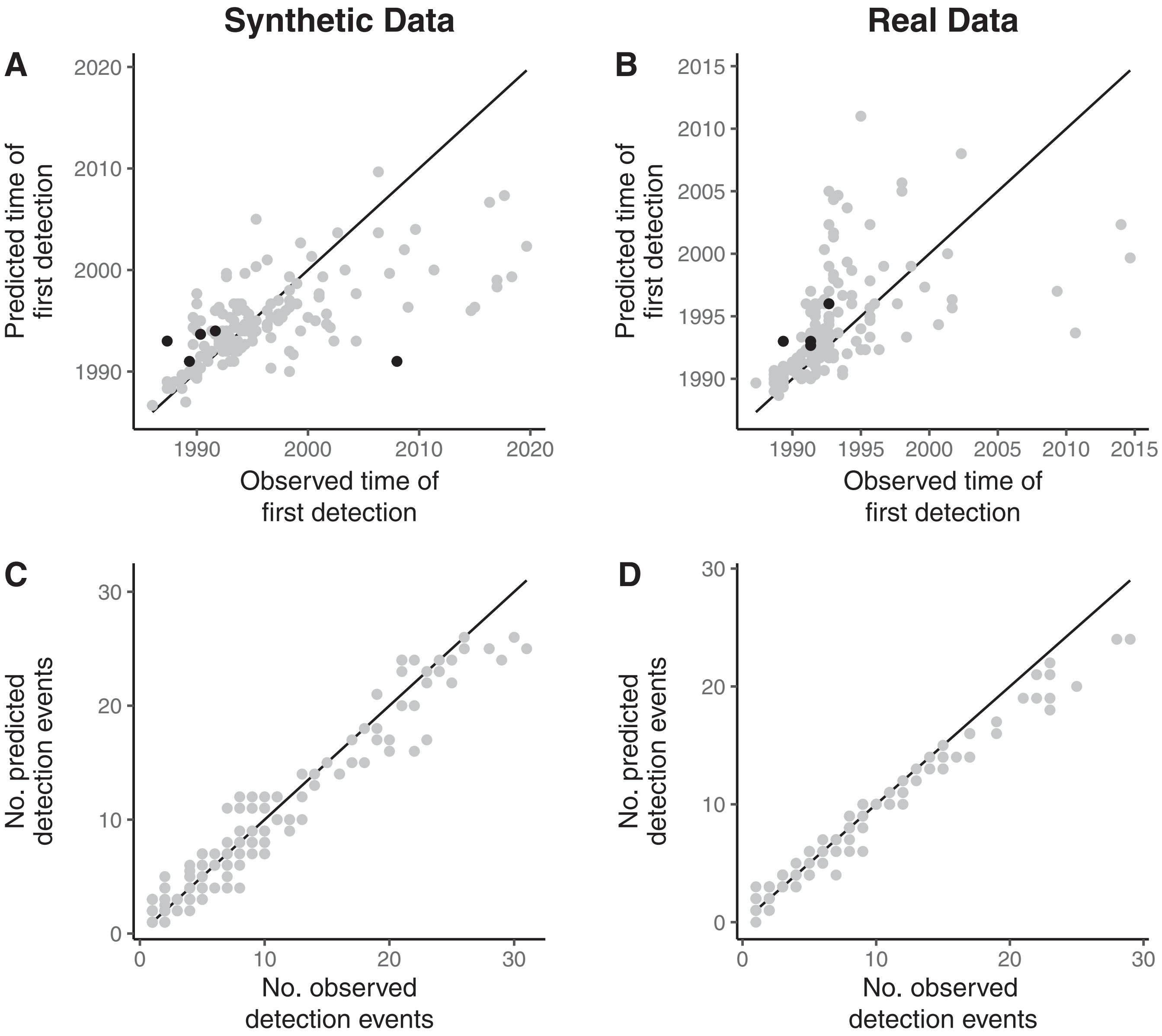

FIGURE 3 Posterior predictive checks for synthetic data (left column) and real data (right column) suggest that the predicted time of first detection (A and B) and the number of detections of zebra mussels (C and D) at a particular location are consistent with the observed time of first detection and the number of confirmed detections, respectively. In both plots, the points represent the median value of the estimate. For gray (black) points, the observed data lie inside (outside) of the 95% credible interval. The black line represents the perfect match between the data and the statistical estimate.

First, by expressing the dispersal process as a redistribution matrix, we can capture the spatial heterogeneity in the dispersal process and exploit the properties of probability matrices to describe the expansion patterns using the concept of network distances (Equation 8). Specifically, the invasion path corresponds, with high probability, to the shortest network path from the origin such that the arrival time is a linear function of the network distance along this path (Equation 9; Figure 1B). Furthermore, visualizing the invasion path on the network space produces a radial wavefront (Figure 1D), which is geometrically analogous to the radial spread patterns produced by the Gaussian IDE models in the geographical space (Brockmann & Helbing, 2013; Kot et al., 1996).

Second, splitting the metapopulation model into spatial and life-history dynamics provides new insights into what makes an invader successful. It is widely accepted that increasing propagule pressure decreases demographic extinctions (Dennis, 1989) and increases the chances of successful establishment. Therefore, everything else being equal, introductions with larger propagule pressure have a higher probability of establishment

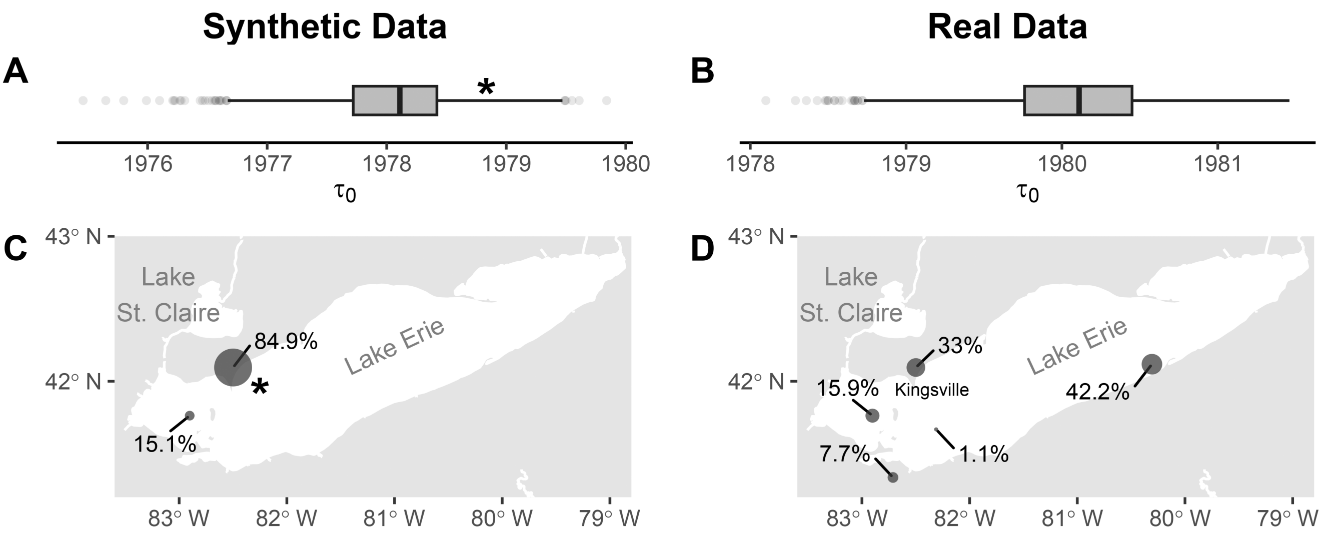

FIGURE 4 The plots show the posterior distribution of time (A and B; using τ_0 as proxy) and place of zebra mussel introduction (C and D) for synthetic (left column) and real data (right column). Fitting the Bayesian model to synthetic data generated from known parameter values recovers the τ_0 (A) and location of the introduction (C) (true parameter values are marked with an asterisk). Analyzing the real data suggests that zebra mussels might have been introduced before 1981 in Lake Erie (B)—at least 5 years before the first confirmed detection in late 1986 (Carlton, 2008). However, within Lake Erie, the statistical model identifies multiple locations where the mussels could have been introduced (D). Interestingly, there is only a one in three chance that mussels were introduced in Kingsville, where they were first detected in 1986.

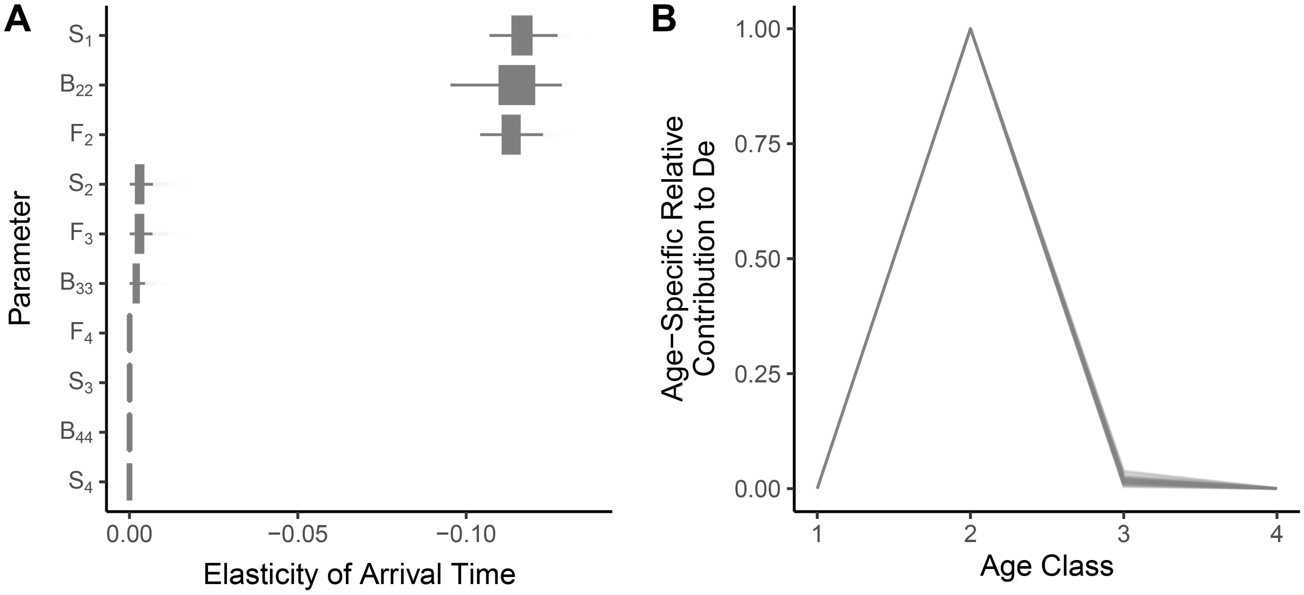

FIGURE 5 (A) Elasticity or proportional sensitivity analysis shows that the arrival time (Equation 10) of zebra mussels is most sensitive to transition probability from age class 1–2 (S_1), hitchhiking rate of individuals in age class 2 (B_{22}), and fecundity of individuals in age class 2 (F_2). (B) Not all age classes contribute equally to zebra mussel dispersal (Equation 4). Age class 2 has the highest relative contribution to effective dispersal, D_e.

(Lockwood et al., 2005; Simberloff, 2009). However, this assertion ignores that not all propagules are equal. Propagules in different age classes (or ontogenetic stages) have different reproductive values—the number of offspring an individual produces from their present age onward (Caswell, 2001). Therefore, even when two introductions have the same propagule pressure, the age composition of propagules can vary, leading to variation in the probability of establishment (Engen et al., 2005). Corollarily, an invader that disperses at peak reproductive value (e.g., age of maturity) is more likely to overcome demographic stochasticity and, consequently, expand faster (Equation 4). This logic also provides quantitative predictions about how evolution by spatial sorting might proceed at invasion fronts (Shine et al., 2011)—evolution will select for individuals that hitchhike with transportation vessels at the age of peak reproductive value or individuals with higher reproductive value when they hitchhike or both.

Finally, a deeper understanding of the synergy between life-history and spatial dynamics allows one to ask novel ecological questions such as "By how much is it necessary to decrease fecundity at age two to stop an invasion?" or "Is it better to inspect vessels for dispersers at early or later life stages?" (Neubert & Caswell, 2000). This knowledge allows managers to identify and target the most responsive life-history parameters to maximize resources (Buhle et al., 2005).

Next, we propose a hierarchical Bayesian model to estimate the parameters of the metapopulation model using commonly available presence-only data. We showed that when dealing with such data, one needs to account explicitly for spatial biases in the data collection procedure (Equations 12–14). Failure to do so can lead to spurious parameter estimates and inferences. For instance, because of uneven spatial sampling, some locations are sampled more often. As a result, it is possible that an invader could have been introduced at a less frequently sampled location and later dispersed and, subsequently, detected for the first time at a location where the sampling was relatively frequent. In such a scenario, using the location of first detection as the location of origin may be inaccurate and produce misleading projections of an invader's path and arrival time. Our analysis of zebra mussel expansion in North America confirms that this scenario is very likely. As expected, based on early detections of zebra mussels (GBIF.org, 2022; USGS, 2022) (Appendix S1: Figure S1), the Bayesian model suggests that zebra mussels were likely introduced in Lake Erie (Figure 4D). However, within Lake Erie, the origin of zebra mussels is uncertain—the statistical model suggests that there is only a one in three chance that these mollusks were initially introduced in Kingsville, where they were first detected (Carlton, 2008). Remarkably, the statistical model facilitated these inferences with sparse detection data—the survey data only had ~6% confirmed detections—demonstrating the value of a mechanistic statistical approach to learning from limited and unstructured occurrence datasets.

Our statistical approach also highlights that some parameters cannot be inferred from the occurrence data when multiple combinations of parameters yield the same invasion outcome. We show that these non-identifiable parameters can be resolved by imputing knowledge from previous studies and the dynamics of sister species in the form of priors. We use this technique to constrain zebra mussel's life-history parameters and carrying capacity, which, in turn, allowed us to make two important inferences. First, the expansion rate of zebra mussels is most sensitive to values of fecundity and dispersal in age class two and survivorship from age one to two (Figure 5A). Second, zebra mussels might have been introduced in North America before 1981 (inferred using τ_0 as a proxy for the time of introduction; see Figure 4B), indicating a lag of 5 years between the introduction and first confirmed detection in 1986. Although such lags are a common feature of invasions (Crooks, 2005), they are notoriously hard to estimate from data. Thus, using informative priors and explicitly modeling the observation and population processes may allow us to estimate invasion parameters that are otherwise challenging to infer (Hobbs & Hooten, 2015).

There are several possible avenues to build on our work. In the metapopulation model, we ignore niche constraints imposed by spatial variation in environmental conditions (Peterson et al., 2011). The environment can also vary in time due to diurnal and seasonal fluctuations (Lande & Orzack, 1988). Variation can also stem from demographic stochasticity associated with reproduction and dispersal (Engen et al., 2005). Mechanistically incorporating these sources of variation in population dynamics may provide us with a more realistic understanding of invasion patterns.

On the statistical front, too, many opportunities exist to improve inferences. We showed that one can obtain reliable parameter estimates by explicitly modeling the observation process. However, these improvements come at the expense of precision. To improve precision, we need standardized biodiversity sampling, which currently exists for only a few taxa due to high cost (Pardieck et al., 2020; Pollard et al., 1994). Standardized species observations can be automated with developments in e-DNA, audio and video sensor technology, and advances in computation (Besson et al., 2022; Keitt & Abelson, 2021). Such automated schemes are scalable, cost-effective, and can collect standardized species data at broad spatial scales with high temporal resolution. In congruence, to use these data sources in our Bayesian models, we also need to develop observation models that can link these novel datasets to latent parameters of the invasion model.

There is also a need to improve data collection on human transportation pathways for species whose dispersal is tightly associated with specific transportation mechanisms. Examples of such pathways include the spread of insects through the transport of firewood (Solano et al., 2021) and plant material (Fenn-Moltu et al., 2023), the introduction of exotic animals through international pet trade (Gippet & Bertelsmeier, 2021), and the transport of mosquitoes with used tires (Yee, 2008). However, when transportation data are absent, we can model the weights of the dispersal network using a mechanistic model of human movement (Simini et al., 2012; Stouffer, 1940; Zipf, 1946).

Finally, the most useful extension of our statistical framework is hierarchically linking the metapopulation model of invasion with a management model (or policy) (Yemshanov et al., 2009), such as establishing checkpoints to intercept hitchhikers (Zook & Phillips, 2012) or chemically eradicating an invader (Fernald & Watson, 2013). Because Bayesian models provide a principled way to quantify uncertainty in parameters (e.g., vital rates and abundances), adding a management model within the Bayesian hierarchy allows managers to propagate this uncertainty while evaluating the cost of implementing different management strategies and then use decision theory to choose the most cost-effective strategy (Dorazio & Johnson, 2003; Williams & Hooten, 2016).

AUTHOR CONTRIBUTIONS

Nikunj Goel conceived and designed the study. Andrew M. Liebhold and Nikunj Goel procured the port call data from Marine Traffic. Nikunj Goel constructed and analyzed the metapopulation and statistical models with inputs from Timothy H. Keitt, Kirill S. Korolev, and Mevin B. Hooten. Nikunj Goel wrote the paper with feedback from Andrew M. Liebhold, Timothy H. Keitt, Kirill S. Korolev, Cleo Bertelsmeier, and Mevin B. Hooten.

ACKNOWLEDGMENTS

We thank Upasana Sarraju for creating the illustration (Figure 2). We also thank Joseph Carver and Randal Holt for collecting metadata on the shipping network from MarineTraffic. Nikunj Goel acknowledges support from the Stengl-Wyer and continuing fellowships from The University of Texas at Austin. Timothy Keitt acknowledges support from the Stengl-Wyer Endowment and the Office of the Vice President for Research, Scholarship, and Creative Endeavors at The University of Texas at Austin and from the National Science Foundation (BCS-2009669). Mevin B. Hooten acknowledges support from National Science Foundation (IOS-2222525) for methodological aspects of this research. Kirill S. Korolev was supported through an National Institute of Health grant R01-GM138530. Cleo Bertelsmeier was supported by The State Secretariat for Education, Research and Innovation-funded project SPREAD (MB22.00086). Andrew M. Liebhold was supported by grant EVA4.0, no. CZ.02.1.01/0.0/0.0/16_019/0000803 financed by Operational Programme Research, Development and Education. MarineTraffic provided the port calls.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

Data and code (Goel, 2025) are available in Zenodo at https://doi.org/10.5281/zenodo.14708765.

ORCID

Nikunj Goel https://orcid.org/0000-0002-5720-7761
Andrew M. Liebhold https://orcid.org/0000-0001-7427-6534
Cleo Bertelsmeier https://orcid.org/0000-0003-3624-1300
Mevin B. Hooten https://orcid.org/0000-0002-1614-723X
Kirill S. Korolev https://orcid.org/0000-0003-1988-0645
Timothy H. Keitt https://orcid.org/0000-0002-4587-1083

REFERENCES

Bagnara, M., L. Nowak, H. J. Boehmer, F. Schöll, F. M. Schurr, and H. Seebens. 2022. "Simulating the Spread and Establishment of Alien Species along Aquatic and Terrestrial Transport Networks: A Multi-Pathway and High-Resolution Approach." *Journal of Applied Ecology* 59: 1769–80.

Banks, N. C., D. R. Paini, K. L. Bayliss, and M. Hodda. 2015. "The Role of Global Trade and Transport Network Topology in the Human-Mediated Dispersal of Alien Species." *Ecology Letters* 18: 188–199.

Bateman, A. W., M. G. Neubert, M. Krkošek, and M. A. Lewis. 2015. "Generational Spreading Speed and the Dynamics of Population Range Expansion." *The American Naturalist* 186: 362–375.

Bellard, C., P. Cassey, and T. M. Blackburn. 2016. "Alien Species as a Driver of Recent Extinctions." *Biology Letters* 12: 20150623.

Berliner, L. M. 1996. *Hierarchical Bayesian Time Series Models* 15–22. Dordrecht: Springer Netherlands.

Besson, M., J. Alison, K. Bjerge, T. E. Gorochowski, T. T. Høye, T. Jucker, H. M. Mann, and C. F. Clements. 2022. "Towards the Fully Automated Monitoring of Ecological Communities." *Ecology Letters* 25: 2753–75.

van den Bosch, F., R. Hengeveld, and J. A. J. Metz. 1992. "Analysing the Velocity of Animal Range Expansion." *Journal of Biogeography* 19: 135–150.

Van den Bosch, F., J. A. J. Metz, and O. Diekmann. 1990. "The Velocity of Spatial Population Expansion." *Journal of Mathematical Biology* 28: 529–565.

Brockmann, D., and D. Helbing. 2013. "The Hidden Geometry of Complex, Network-Driven Contagion Phenomena." *Science* 342: 1337–42.

Buhle, E. R., M. Margolis, and J. L. Ruesink. 2005. "Bang for Buck: Cost-Effective Control of Invasive Species with Different Life Histories." *Ecological Economics* 52: 355–366.

Carlton, J. T. 2008. "The Zebra Mussel *Dreissena Polymorpha* Found in North America in 1986 and 1987." *Journal of Great Lakes Research* 34: 770–73.

Carpenter, B., A. Gelman, M. D. Hoffman, D. Lee, B. Goodrich, M. Betancourt, M. A. Brubaker, J. Guo, P. Li, and A. Riddell.

2017. “Stan: A Probabilistic Programming Language.” *Journal of Statistical Software* 76: 1–32.

Carrasco, L. R., J. Mumford, A. MacLeod, T. Harwood, G. Grabenweger, A. Leach, J. Knight, and R. Baker. 2010. “Unveiling Human-Assisted Dispersal Mechanisms in Invasive Alien Insects: Integration of Spatial Stochastic Simulation and Phenology Models.” *Ecological Modelling* 221: 2068–75.

Casagrandi, R., L. Mari, and M. Gatto. 2007. “Modelling the Local Dynamics of the Zebra Mussel (*Dreissena Polymorpha*).” *Freshwater Biology* 52: 1223–38.

Caswell, H. 2001. *Matrix Population Models*. Sunderland, MA: Sinauer.

Chapman, D. S., L. Makra, R. Albertini, M. Bonini, A. Páldy, V. Rodinkova, B. Šikoparija, E. Weryszko-Chmielewska, and J. M. Bullock. 2016. “Modelling the Introduction and Spread of Non-native Species: International Trade and Climate Change Drive Ragweed Invasion.” *Global Change Biology* 22: 3067–79.

Colizza, V., A. Barrat, M. Barthélemy, and A. Vespignani. 2006. “The Role of the Airline Transportation Network in the Prediction and Predictability of Global Epidemics.” *Proceedings of the National Academy of Sciences of the United States of America* 103: 2015–20.

Crooks, J. A. 2005. “Lag Times and Exotic Species: The Ecology and Management of Biological Invasions in Slow-Motion.” *Ecoscience* 12: 316–329.

Cross, C. L., W. H. Wong, and T. Che. 2011. “Estimating Carrying Capacity of Quagga Mussels (*Dreissena Rostriformis Bugensis*) in a Natural System: A Case Study of the Boulder Basin of Lake Mead, Nevada-Arizona.” *Aquatic Invasions* 6: 141–47.

Crouse, D. T., L. B. Crowder, and H. Caswell. 1987. “A Stage-Based Population Model for Loggerhead Sea Turtles and Implications for Conservation.” *Ecology* 68: 1412–23.

Dennis, B. 1989. “Allee Effects: Population Growth, Critical Density, and the Chance of Extinction.” *Natural Resource Modeling* 3: 481–538.

Diagne, C., B. Leroy, A.-C. Vaissière, R. E. Gozlan, D. Roiz, I. Jarić, J.-M. Salles, C. J. Bradshaw, and F. Courchamp. 2021. “High and Rising Economic Costs of Biological Invasions Worldwide.” *Nature* 592: 571–76.

Dorazio, R. M., and F. A. Johnson. 2003. “Bayesian Inference and Decision Theory—A Framework for Decision Making in Natural Resource Management.” *Ecological Applications* 13: 556–563.

Engen, S., R. Lande, B.-E. Sæther, and H. Weimerskirch. 2005. “Extinction in Relation to Demographic and Environmental Stochasticity in Age-Structured Models.” *Mathematical Biosciences* 195: 210–227.

Fenn-Moltu, G., S. Ollier, B. Caton, A. M. Liebhold, H. Nahrung, D. S. Pureswaran, R. M. Turner, T. Yamanaka, and C. Bertelsmeier. 2023. “Alien Insect Dispersal Mediated by the Global Movement of Commodities.” *Ecological Applications* 33: e2721.

Fernald, R. T., and B. T. Watson. 2013. “Eradication of Zebra Mussels (*Dreissena Polymorpha*) from Millbrook Quarry, Virginia.” In *Quagga and Zebra Mussels: Biology, Impacts and Control*, edited by T. F. Nalepa and D. W. Schloesser, 195–214. Boca Raton, FL: CRC Press.

Ferrari, J. R., E. L. Preisser, and M. C. Fitzpatrick. 2014. “Modeling the Spread of Invasive Species Using Dynamic Network Models.” *Biological Invasions* 16: 949–960.

Floerl, O., G. Inglis, K. Dey, and A. Smith. 2009. “The Importance of Transport Hubs in Stepping-Stone Invasions.” *Journal of Applied Ecology* 46: 37–45.

Gautreau, A., A. Barrat, and M. Barthelemy. 2008. “Global Disease Spread: Statistics and Estimation of Arrival Times.” *Journal of Theoretical Biology* 251: 509–522.

GBIF.org. 2022. *Occurrence Download*. Denmark: The Global Biodiversity Information Facility.

Gelman, A., J. B. Carlin, H. S. Stern, and D. B. Rubin. 1995. *Bayesian Data Analysis*. Boca Raton, FL: CRC Press.

Gelman, A., A. Vehtari, D. Simpson, C. C. Margossian, B. Carpenter, Y. Yao, L. Kennedy, J. Gabry, P.-C. Bürkner, and M. Modrák. 2020. “Bayesian Workflow.” arXiv Preprint arXiv: 2011.01808.

Gippet, J. M., and C. Bertelsmeier. 2021. “Invasiveness Is Linked to Greater Commercial Success in the Global Pet Trade.” *Proceedings of the National Academy of Sciences of the United States of America* 118: e2016337118.

Goel, N. 2025. “A Mechanistic Statistical Approach to Infer Invasion Characteristics of Human-Dispersed Species with Complex Life Cycle. In Ecological Monographs (Version V1).” Zenodo. https://doi.org/10.5281/zenodo.14708765.

Griffiths, R. W., D. W. Schloesser, J. H. Leach, and W. P. Kovalak. 1991. “Distribution and Dispersal of the Zebra Mussel (*Dreissena Polymorpha*) in the Great Lakes Region.” *Canadian Journal of Fisheries and Aquatic Sciences* 48: 1381–88.

Hallatschek, O., and D. S. Fisher. 2014. “Acceleration of Evolutionary Spread by Long-Range Dispersal.” *Proceedings of the National Academy of Sciences of the United States of America* 111: E4911–E4919.

Hebert, P. D., B. Muncaster, and G. Mackie. 1989. “Ecological and Genetic Studies on *Dreissena Polymorpha* (Pallas): A New Mollusc in the Great Lakes.” *Canadian Journal of Fisheries and Aquatic Sciences* 46: 1587–91.

Hobbs, N. T., and M. B. Hooten. 2015. *Bayesian Models: A Statistical Primer for Ecologists*. Princeton, NJ: Princeton University Press.

Hoffman, M. D., and A. Gelman. 2014. “The No-U-Turn Sampler: Adaptively Setting Path Lengths in Hamiltonian Monte Carlo.” *Journal of Machine Learning Research* 15: 1593–1623.

Hooten, M. B., and T. J. Hefley. 2019. *Bringing Bayesian Models to Life*. Boca Raton, FL: CRC Press.

Hooten, M. B., C. K. Wikle, R. M. Dorazio, and J. A. Royle. 2007. “Hierarchical Spatiotemporal Matrix Models for Characterizing Invasions.” *Biometrics* 63: 558–567.

Horvitz, N., R. Wang, F.-H. Wan, and R. Nathan. 2017. “Pervasive Human-Mediated Large-Scale Invasion: Analysis of Spread Patterns and their Underlying Mechanisms in 17 of China’s Worst Invasive Plants.” *Journal of Ecology* 105: 85–94.

Hu, Y., and D. Zhu. 2009. “Empirical Analysis of the Worldwide Maritime Transportation Network.” *Physica A: Statistical Mechanics and its Applications* 388: 2061–71.

Hui, C., and D. M. Richardson. 2017. *Invasion Dynamics*. Oxford: Oxford University Press.

Hulme, P. E. 2009. “Trade, Transport and Trouble: Managing Invasive Species Pathways in an Era of Globalization.” *Journal of Applied Ecology* 46: 10–18.

Hunter, C. M., and H. Caswell. 2005. “The Use of the Vec-Permutation Matrix in Spatial Matrix Population Models.” *Ecological Modelling* 188: 15–21.

Iannelli, F., A. Koher, D. Brockmann, P. Hovel, and I. M. Sokolov. 2017. “Effective Distances for Epidemics Spreading on Complex Networks.” *Physical Review E* 95: 012313.

Isaac, N. J., A. J. van Strien, T. A. August, M. P. de Zeeuw, and D. B. Roy. 2014. “Statistics for Citizen Science: Extracting Signals of Change from Noisy Ecological Data.” *Methods in Ecology and Evolution* 5: 1052–60.

Jamieson-Lane, A., and B. Blasius. 2020. “Calculation of Epidemic Arrival Time Distributions Using Branching Processes.” *Physical Review E* 102: 042301.

Johnson, L. E., and J. T. Carlton. 1996. “Post-Establishment Spread in Large-Scale Invasions: Dispersal Mechanisms of the Zebra Mussel *Dreissena Polymorpha*.” *Ecology* 77: 1686–90.

Johnson, L. E., and D. K. Padilla. 1996. “Geographic Spread of Exotic Species: Ecological Lessons and Opportunities from the Invasion of the Zebra Mussel *Dreissena Polymorpha*.” *Biological Conservation* 78: 23–33.

Johnston, A., E. Matechou, and E. B. Dennis. 2023. “Outstanding Challenges and Future Directions for Biodiversity Monitoring Using Citizen Science Data.” *Methods in Ecology and Evolution* 14: 103–116.

Keevin, T. M., R. E. Yarbrough, and A. C. Miller. 1992. “Long-Distance Dispersal of Zebra Mussels (*Dreissena Polymorpha*) Attached to Hulls of Commercial Vessels.” *Journal of Freshwater Ecology* 7: 437.

Keitt, T. H., and E. S. Abelson. 2021. “Ecology in the Age of Automation.” *Science* 373: 858–59.

Kot, M., M. A. Lewis, and P. vandenDriessche. 1996. “Dispersal Data and the Spread of Invading Organisms.” *Ecology* 77: 2027–42.

Kulkarni, V. G. 2016. *Modeling and Analysis of Stochastic Systems*. London: Chapman and Hall/CRC.

Lande, R., and S. H. Orzack. 1988. “Extinction Dynamics of Age-Structured Populations in a Fluctuating Environment.” *Proceedings of the National Academy of Sciences of the United States of America* 85: 7418–21.

Lefkovitch, L. 1965. “The Study of Population Growth in Organisms Grouped by Stages.” *Biometrics* 21: 1–18.

Leslie, P. H. 1945. “On the Use of Matrices in Certain Population Mathematics.” *Biometrika* 33: 183–212.

Lewis, M. A., S. V. Petrovskii, and J. R. Potts. 2016. *The Mathematics behind Biological Invasions*. New York: Springer.

Lockwood, J. L., P. Cassey, and T. Blackburn. 2005. “The Role of Propagule Pressure in Explaining Species Invasions.” *Trends in Ecology & Evolution* 20: 223–28.

Lowe, S., M. Browne, S. Boudjelas, and M. De-Poorter. 2004. *One Hundred of the World’s Worst Invasive Alien Species: A Selection from the Global Invasive Species Database*. Geneva: Invasive Species Specialist Group, IUCN.

Ludyanskiy, M. L., D. McDonald, and D. MacNeill. 1993. “Impact of the Zebra Mussel, a Bivalve Invader: *Dreissena Polymorpha* Is Rapidly Colonizing Hard Surfaces throughout Waterways of the United States and Canada.” *BioScience* 43: 533–544.

Lutscher, F. 2019. *Integrodifference Equations in Spatial Ecology*. Cham: Springer.

May, R. M. 2004. “Uses and Abuses of Mathematics in Biology.” *Science* 303: 790–93.

McElreath, R. 2020. *Statistical Rethinking: A Bayesian Course with Examples in R and Stan*. Boca Raton, FL: CRC Press.

Meyerson, L. A., and H. A. Mooney. 2007. “Invasive Alien Species in an Era of Globalization.” *Frontiers in Ecology and the Environment* 5: 199–208.

Miller, D. A., K. Pacifici, J. S. Sanderlin, and B. J. Reich. 2019. “The Recent Past and Promising Future for Data Integration Methods to Estimate Species’ Distributions.” *Methods in Ecology and Evolution* 10: 22–37.

Nalepa, T. F., and D. W. Schloesser. 2013. *Quagga and Zebra Mussels: Biology, Impacts, and Control*. Boca Raton, FL: CRC Press.

Neal, R. M. 2011. “MCMC Using Hamiltonian Dynamics.” In *Handbook of Markov Chain Monte Carlo*, edited by S. Brooks, A. Gelman, G. Jones, and X.-L. Meng, 2. New York: CRC Press.

Neath, A. A., and F. J. Samaniego. 1997. “On the Efficacy of Bayesian Inference for Nonidentifiable Models.” *The American Statistician* 51: 225–232.

Neubert, M. G., and H. Caswell. 2000. “Demography and Dispersal: Calculation and Sensitivity Analysis of Invasion Speed for Structured Populations.” *Ecology* 81: 1613–28.

Padilla, D. K., M. A. Chotkowski, and L. A. Buchan. 1996. “Predicting the Spread of Zebra Mussels (*Dreissena Polymorpha*) to Inland Waters Using Boater Movement Patterns.” *Global Ecology and Biogeography Letters* 5: 353–59.

Pagel, J., and F. M. Schurr. 2012. “Forecasting Species Ranges by Statistical Estimation of Ecological Niches and Spatial Population Dynamics.” *Global Ecology and Biogeography* 21: 293–304.

Pardieck, K. L., D. Ziolkowski, M. Lutmerding, V. Aponte, and M.-A. R. Hudson. 2020. *North American Breeding Bird Survey Dataset 1966—2019*. Kearneysville, WV: U.S.G.S.

Peterson, A. T., J. Soberón, R. G. Pearson, R. P. Anderson, E. Martinez-Meyer, M. Nakamura, and M. B. Arau’jo. 2011. *Ecological Niches and Geographic Distributions*. Princeton, NJ: Princeton University Press.

Pollard, E., T. J. Yates, and S. P. Courtney. 1994. “Monitoring Butterflies for Ecology and Conservation: The British Butterfly Monitoring Scheme.” *Ecology* 75: 1851.

Ricker, W. E. 1954. “Stock and Recruitment.” *Journal of the Fisheries Research Board of Canada* 11: 559–623.

Rvachev, L. A., and I. M. Longini, Jr. 1985. “A Mathematical Model for the Global Spread of Influenza.” *Mathematical Biosciences* 75: 3–22.

Seebens, H., T. M. Blackburn, E. E. Dyer, P. Genovesi, P. E. Hulme, J. M. Jeschke, S. Pagad, et al. 2018. “Global Rise in Emerging Alien Species Results from Increased Accessibility of New Source Pools.” *Proceedings of the National Academy of Sciences of the United States of America* 115: E2264–E2273.

Seebens, H., E. Briski, S. Ghabooli, T. Shiganova, H. J. MacIsaac, and B. Blasius. 2019. “Non-Native Species Spread in a Complex Network: The Interaction of Global Transport and Local Population Dynamics Determines Invasion Success.” *Proceedings of the Royal Society B* 286: 20190036.

Shine, R., G. P. Brown, and B. L. Phillips. 2011. “An Evolutionary Process that Assembles Phenotypes through Space Rather than through Time.” *Proceedings of the National Academy of Sciences of the United States of America* 108: 5708–11.

Simberloff, D. 2009. “The Role of Propagule Pressure in Biological Invasions.” *Annual Review of Ecology, Evolution, and Systematics* 40: 81–102.

Simini, F., M. C. González, A. Maritan, and A.-L. Barabási. 2012. "A Universal Model for Mobility and Migration Patterns." *Nature* 484: 96–100.

Skellam, J. G. 1951. "Random Dispersal in Theoretical Populations." *Biometrika* 38: 196–218.

Solano, A., S. L. Rodriguez, L. Greenwood, K. J. Dodds, and D. R. Coyle. 2021. "Firewood Transport as a Vector of Forest Pest Dispersal in North America: A Scoping Review." *Journal of Economic Entomology* 114: 14–23.

Stouffer, S. A. 1940. "Intervening Opportunities: A Theory Relating Mobility and Distance." *American Sociological Review* 5: 845–867.

Thompson, B. K., J. D. Olden, and S. J. Converse. 2021. "Mechanistic Invasive Species Management Models and their Application in Conservation." *Conservation Science and Practice* 3: e533.

Turbelin, A. J., R. N. Cuthbert, F. Essl, P. J. Haubrock, A. Ricciardi, and F. Courchamp. 2023. "Biological Invasions Are as Costly as Natural Hazards." *Perspectives in Ecology and Conservation* 21: 143–150.

USGS. 2022. *Dreissena polymorpha (Pallas, 1771): U.S. Geological Survey, Nonindigenous Aquatic Species Database*. Gainesville, FL: United States Geological Survey.

Warziniack, T., R. G. Haight, D. Yemshanov, J. L. Apriesnig, T. P. Holmes, A. M. Countryman, J. D. Rothlisberger, and C. Haberland. 2021. "Economics of Invasive Species." In *Invasive Species in Forests and Rangelands of the United States: A Comprehensive Science Synthesis for the United States Forest Sector*, edited by T. M. Poland T. Patel-Weynand, D. M. Finch, C. F. Miniat, D. C. Hayes, and V. M. Lopez, 305–320. Cham: Springer International Publishing.

Wikle, C. K. 2003. "Hierarchical Bayesian Models for Predicting the Spread of Ecological Processes." *Ecology* 84: 1382–94.

Williams, P. J., and M. B. Hooten. 2016. "Combining Statistical Inference and Decisions in Ecology." *Ecological Applications* 26: 1930–42.

Wolf, F., C. Kirsch, and R. V. Donner. 2019. "Edge Directionality Properties in Complex Spherical Networks." *Physical Review E* 99: 012301.

Yee, D. A. 2008. "Tires as Habitats for Mosquitoes: A Review of Studies within the Eastern United States." *Journal of Medical Entomology* 45: 581–593.

Yemshanov, D., D. W. McKenney, J. H. Pedlar, F. H. Koch, and D. Cook. 2009. "Towards an Integrated Approach to Modelling the Risks and Impacts of Invasive Forest Species." *Environmental Reviews* 17: 163–178.

Zipf, G. K. 1946. "The P 1 P 2/D Hypothesis: On the Intercity Movement of Persons." *American Sociological Review* 11: 677–686.

Zook, B., and S. Phillips. 2012. *Uniform Minimum Protocols and Standards for Watercraft Interception Programs for Dreissenid Mussels in the Western United States (UMPS)* 77. Portland, OR: Pacific States Marine Fisheries Commission.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Goel, Nikunj, Andrew M. Liebhold, Cleo Bertelsmeier, Mevin B. Hooten, Kirill S. Korolev, and Timothy H. Keitt. 2025. "A Mechanistic Statistical Approach to Infer Invasion Characteristics of Human-Dispersed Species with Complex Life Cycle." *Ecological Monographs* 95(1): e70003. https://doi.org/10.1002/ecm.70003