LETTER

When mechanism matters: Bayesian forecasting using models of ecological diffusion

Trevor J. Hefley,[1]* Mevin B. Hooten,[2] Robin E. Russell,[3] Daniel P. Walsh[3] and James A. Powell[4]

Abstract

Ecological diffusion is a theory that can be used to understand and forecast spatio-temporal processes such as dispersal, invasion, and the spread of disease. Hierarchical Bayesian modelling provides a framework to make statistical inference and probabilistic forecasts, using mechanistic ecological models. To illustrate, we show how hierarchical Bayesian models of ecological diffusion can be implemented for large data sets that are distributed densely across space and time. The hierarchical Bayesian approach is used to understand and forecast the growth and geographic spread in the prevalence of chronic wasting disease in white-tailed deer (*Odocoileus virginianus*). We compare statistical inference and forecasts from our hierarchical Bayesian model to phenomenological regression-based methods that are commonly used to analyse spatial occurrence data. The mechanistic statistical model based on ecological diffusion led to important ecological insights, obviated a commonly ignored type of collinearity, and was the most accurate method for forecasting.

Keywords

Agent-based model, Bayesian analysis, boosted regression trees, dispersal, generalised additive model, invasion, partial differential equation, prediction, spatial confounding.

Ecology Letters (2017)

INTRODUCTION

Mathematical models that are specified as differential equations play an essential role in describing and understanding ecological processes. For example, Skellam (1951) used a reaction-diffusion partial differential equation (PDE) to understand the invasion of muskrats in central Europe; May (1973) made extensive use of differential equations, including the Fokker–Planck equation of diffusion, in his seminal work on stability and complexity of ecosystems; Levin (1992) noted the importance of diffusion models for understanding the problem of pattern and scale in ecology; Scheffer *et al.* (2001) employed a simple differential equation to explain catastrophic shifts in ecosystems. Hastings *et al.* (2005) noted that, for the problem of spatial spread and invasions, '[m]uch more data are becoming available, and new statistical techniques are being developed to match data with theory'. Many of the theories developed in ecology are phrased as differential equations, yet a fundamental gap exists between the mathematical models proposed by theoretical ecologists and the statistical models used by applied ecologists (Hilborn & Mangel 1997; Hobbs & Hooten 2015).

Ecological forecasting is the process of predicting the state of an ecological system with fully specified uncertainties (Clark *et al.* 2001). Although forecasting reduces uncertainty about future states, it does not eliminate uncertainty (Pielke & Conant 2003). Consequently, forecasts should be probabilistic

(Gneiting & Katzfuss 2014; Raftery 2016). Hierarchical modelling is a framework to specify uncertainties associated with ecological systems (Cressie *et al.* 2009) and Bayesian inference provides a coherent means for making probabilistic forecasts (Clark 2005).

Ecological theories expressed as PDEs are an essential component of statistical models capable of forecasting the temporal evolution of spatial processes (Holmes *et al.* 1994; Wikle 2003). By combining commonly used Bayesian estimation methods with mechanistic mathematical models, statistical implementations of PDEs facilitate defensible probabilistic forecasts of spatio-temporal processes (Wikle *et al.* 1998). Furthermore, there are connections between statistical models constructed using PDEs and other spatio-temporal modelling approaches such as agent (or individual-)-based models (Hooten & Wikle 2010) and basis function models (Hefley *et al.* 2017b).

Since the seminal work of Hotelling (1927), statisticians and mathematicians have made significant progress in developing the machinery necessary to fit differential equation models to ecological data (e.g. Wikle 2003; Cressie & Wikle 2011). As a result, the statistical and mathematical machinery required to fit PDEs to ecological data and obtain probabilistic forecasts is well developed and available for application. However, phenomenological regression-based spatio-temporal models are the most widely used method for many ecological applications. For example, species distribution models are commonly used to forecast the spread of invasive species (e.g. Uden *et al.* 2015),

[1]*Department of Statistics, Kansas State University, 205 Dickens Hall, 1116 Mid-Campus Drive North, Manhattan, KS 66506, USA*
[2]*U.S. Geological Survey, Colorado Cooperative Fish and Wildlife Research Unit, Department of Fish, Wildlife, and Conservation Biology, Department of Statistics, Colorado State University, 1484 Campus Delivery, Fort Collins, CO 80523*

[3]*U.S. Geological Survey, National Wildlife Health Center, 6006 Schroeder Road, Madison, WI 53711, USA*
[4]*Department of Mathematics and Statistics, Utah State University, 3900 Old Main Hill, Logan, Utah 84322*
Correspondence: E-mail: thefley@ksu.edu

but most modelling efforts rely on simple regression-based additive models (Elith & Leathwick 2009; Hefley & Hooten 2016). Although phenomenological regression models are convenient, modelling dynamic ecological processes using mechanistically motivated PDEs yields scientific insights that would not be possible otherwise (Wikle & Hooten 2010). Additionally, mechanistic spatio-temporal models offer an alternative to regression-based approaches that encounter difficulties associated with a type of collinearity that occurs when accounting for spatial and temporal autocorrelation (Kühn 2007; Bini *et al.* 2009; Hodges & Reich 2010; Hanks *et al.* 2015; Hefley *et al.* 2016; Hefley *et al.* 2017b). Although mechanistic spatio-temporal models may be the preferred method, Elith & Leathwick (2009) noted that 'these [mechanistic models] can require specialised mathematics and programming, and this currently hinders wider uptake despite apparent advantages', which is particularly true for large data sets.

We demonstrate how PDEs that describe ecological diffusion can be fit to large data sets within a hierarchical Bayesian framework to facilitate statistical inference and probabilistic forecasting. We begin by reviewing ecological diffusion, the hierarchical modelling framework, and the numerical methods required to increase computational efficiency for large data sets. Our work is motivated by the need to forecast and understand the mechanisms driving the geographic spread and growth in the prevalence of chronic wasting disease (CWD) in white-tailed deer (*Odocoileus virginianus*). We compare the accuracy of predictions and forecasts from our hierarchical Bayesian model to state-of-the-art regression-based statistical and machine learning methods using two out-of-sample validation data sets. In addition, we provide tutorials with the computational details, annotated computer code to assist readers implementing similar models, and the necessary code to reproduce all results and figures related to the analysis (Supporting Information).

MATERIALS AND METHODS

Ecological diffusion

Ecological diffusion describes the population-level distribution that emerges from individual random walks with movement probabilities based on local habitat information, resulting in variable residence times (Fig. 1; Chandrasekhar 1943; Patlak 1953; Turchin 1998; Okubo & Levin 2001). Although the initial motivation is a sequence of movements in discrete time and space, a continuous formulation of the process is desirable so that computationally efficient algorithms can be exploited and inference is not sensitive to an arbitrary choice of time step or grid cell size (Fig. 1). Ecological diffusion is described by the PDE

$$\frac{\partial}{\partial t} u(\mathbf{s}, t) = \left(\frac{\partial^2}{\partial s_1^2} + \frac{\partial^2}{\partial s_2^2} \right) [\mu(\mathbf{s}, t) u(\mathbf{s}, t)], \tag{1}$$

where $u(\mathbf{s}, t)$ is the density of the dispersing population (or disease), s_1 and s_2 are the spatial coordinates contained in the vector $\mathbf{s}$, and t is the time. The diffusion coefficient (or motility coefficient), $\mu(\mathbf{s}, t)$, is inversely related to residence time and could depend on covariates that vary over space and time (Garlick *et al.* 2011; Hooten *et al.* 2013). As a result of the variable residence times, the spatial distribution of $u(\mathbf{s}, t)$ is heterogeneous and captures local variability and abrupt changes in population densities (or disease prevalence) that occurs at the transition between habitat types.

Ecological diffusion is a simple mechanistic model that can be used to link the temporal dynamics of transient spatial processes. For example, Matthiopoulos *et al.* (2015) developed a conceptual framework to unite models of habitat selection and population dynamics but noted ten assumptions that would motivate future research, including developing models of colonisation that capture the transient dynamics preventing animals from instantaneously accessing all high-quality habitats. PDEs like eqn 1 connect processes interacting in space and time while providing a mechanism that naturally captures transient dynamics due to the explicit dependence on time (e.g. Wikle 2003; Williams *et al.* 2017).

The ecological diffusion PDE can be modified to capture a wide range of spatio-temporal dynamics (Holmes *et al.* 1994). For example, eqn 1 can be combined with source and sink densities to describe important components of ecological processes resulting in reaction-diffusion equations of the form
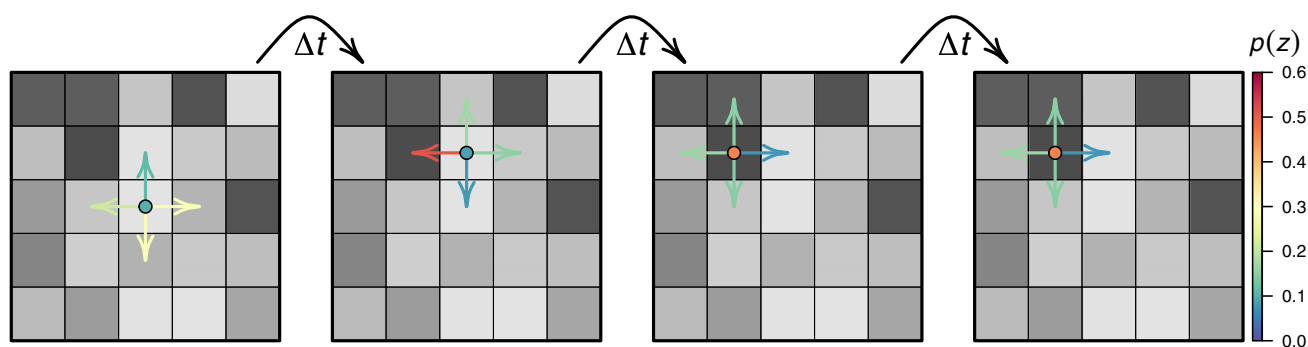

Figure 1 Illustration of an individual random walk based on local habitat information that results in the population-level distribution of individuals described by the ecological diffusion PDE (eqn 1). At each transition time (Δt), an individual ($\circ$) can move to a neighbouring habitat patch indicated by the arrows or stay in the current habitat patch. The probability $p(z)$ that an individual moves to a neighbouring patch or resides in the current patch, as indicated by the colour of arrows and dot, depends on the value of the habitat z. Darker habitat patches have a higher value of $p(z)$ and retain individuals longer, resulting in variable residence times. As Δt and the grid cell size become infinitesimally small $p(z)$ is proportional to the diffusion rate $\mu(\mathbf{s}, t)$ in eqn 1.

Table 1 Examples of ecological dynamics addressed by statistical implementation of diffusion models

Ecological dynamic	References
Animal disease	Hooten & Wikle (2010)
Animal resource selection	Moorcroft & Barnett (2008)
Fish migration	Arab (2007, ch. 2)
Insect dispersal	Powell & Bentz (2014)
Invasion/colonisation	Wikle (2003), Hooten & Wikle (2008), Broms et al. (2016), and Williams et al. (2017)
Plant disease	Zheng & Aukema (2010)

$$\frac{\partial}{\partial t} u(\mathbf{s}, t) = \left(\frac{\partial^2}{\partial s_1^2} + \frac{\partial^2}{\partial s_2^2} \right) [\mu(\mathbf{s}, t) u(\mathbf{s}, t)] + f(u(\mathbf{s}, t), \mathbf{s}, t), \qquad (2)$$

which contains a growth component ($f(u(\mathbf{s}, t), \mathbf{s}, t)$). For example, establishing the connection between habitat selection (i.e. the spatial distribution of organisms) and population dynamics is a prerequisite to understanding species-environment relationships (e.g. Matthiopoulos et al. 2015). When coupled with models of population (or disease) growth, the ecological diffusion PDE provides a mathematically appealing way to link non-separable spatial and temporal dynamics. For example, a logistic model could be used in eqn 2 (i.e. $f(u(\mathbf{s}, t), \mathbf{s}, t) \equiv \lambda(\mathbf{s}, t)(1 - u(\mathbf{s}, t)/\kappa(\mathbf{s}, t))$) where the growth rate, $\lambda(\mathbf{s}, t)$ and equilibrium population size $\kappa(\mathbf{s}, t)$, may vary over space and time (Hooten & Wikle 2008; Hefley & Hooten 2016). To illustrate short-term forecasting of a disease outbreak, we use a temporal forecast horizon of ≤ 3 years (Petchey et al. 2015) and an exponential growth model ($f(u(\mathbf{s}, t), \mathbf{s}, t) \equiv \lambda(\mathbf{s}, t) u(\mathbf{s}, t)$ in eqn 2) in what follows, but provide references to illustrate the breadth of statistical implementations of diffusion models (Table 1).

Hierarchical statistical modelling framework

Hierarchical models with PDE process components are flexible and can be tailored to match the specifics of the study. The generalised linear mixed model (GLMM) is a flexible and widely used hierarchical model in ecology (e.g. Bolker et al. 2009). We describe a modelling framework consistent with the GLMM terminology, but a wide range of hierarchical models could be employed (Table 1). Our model can be written as

$$y_i \sim [y_i | \pi_i, \psi] \qquad (3)$$

$$g(\pi_i) = h(\mathbf{x}_i' \boldsymbol{\beta}, u(\mathbf{s}_i, t_i)), \qquad (4)$$

where y_i is the ith observation ($i = 1, \ldots, n$) at the spatial location $\mathbf{s}_i$ within the study area $\mathcal{S}$ ($\mathbf{s}_i \in \mathcal{S} \subset \mathcal{R}^2$) and time t_i during the time period $\mathcal{T}$ ($t_i \in \mathcal{T}$), π_i is the conditional expected value of y_i, and ψ is the dispersion parameter from a probability distribution denoted by $[\cdot]$. For example, if binary data were observed, then $[\cdot]$ in eqn 3 would be a Bernoulli distribution with probability of success (i.e. $y_i = 1$) equal to π_i and the dispersion parameter would not be required. The inverse of the link function $g(\cdot)$ transforms the (possibly) nonlinear predictor, $h(\mathbf{x}_i' \boldsymbol{\beta}, u(\mathbf{s}_i, t_i))$, to match the support of π_i. The nonlinear predictor in eqn 4 has three components: $\mathbf{x}_i$, a vector of covariates associated with each observation y_i; a vector of regression coefficients $\boldsymbol{\beta}$; and the spatio-temporal effect $u(\mathbf{s}_i, t_i)$ that is a solution to a PDE. The function $h(\cdot)$ in eqn 4 combines the influence of individual-level effects, $\mathbf{x}_i' \boldsymbol{\beta}$, and the spatio-temporal effect, $u(\mathbf{s}_i, t_i)$. For example, the standard GLMM is additive: $g(\pi_i) = \mathbf{x}_i' \boldsymbol{\beta} + u(\mathbf{s}_i, t_i)$. Although useful in some contexts, linear forms of $h(\cdot)$ may not be the most appropriate. Determining how $\mathbf{x}_i' \boldsymbol{\beta}$ and $u(\mathbf{s}_i, t_i)$ are combined via $h(\cdot)$ will be problem-specific and requires knowledge of the process being modeled (e.g. Hooten et al. 2007; Cangelosi & Hooten 2009; Williams et al. 2017).

Numerical implementation

Statistical implementation of all but the simplest hierarchical models requires numerical integration (Hobbs & Hooten 2015). Consequently, implementation will require solving the PDE (i.e. finding $u(\mathbf{s}, t)$ in eqn 4) for each iteration of the algorithm used for numerical integration. Few PDEs with non-constant coefficients can be solved analytically, therefore numerical methods are often required.

The finite-difference method is a commonly used numerical approach to solve PDEs for use in hierarchical models (Farlow 1993, pp. 301–308; Wikle & Hooten 2010). Briefly, solving a PDE using finite-differencing involves partitioning the spatial domain $\mathcal{S}$ into a grid $\mathbf{S}$ ($\mathbf{S} \subseteq \mathcal{S}$) with q cells and the temporal domain $\mathcal{T}$ into r bins $\mathbf{T}$ of width Δt ($\mathbf{T} \subseteq \mathcal{T}$). The partitioning results in a linear equation $\mathbf{u}_t = \mathbf{H}_t \mathbf{u}_{t-\Delta t}$, where $\mathbf{u}_t$ approximates $u(\mathbf{s}, t)$ at the centre of the grid cells. The propagator matrix $\mathbf{H}_t$ has dimensions $q \times q$ and each element is determined from a discretised version of the PDE (Wikle & Hooten 2010). The accuracy of the numerical approximation of $u(\mathbf{s}, t)$ increases as the number of cells on the spatial grid increases and Δt becomes small. In some cases, formal uncertainty quantification associated with the error introduced by numerical approximation of the PDE will be needed (Chkrebtii et al. 2016). Alternatively, an assessment of the error introduced by discretisation on statistical inference can be made by checking the sensitivity of the results to changes in the resolution of the grid and Δt.

Although discretisation of the PDE results in a series of discrete time matrix equations (i.e. $\mathbf{u}_t = \mathbf{H}_t \mathbf{u}_{t-\Delta t}$), it is important to maintain the connection to the PDE that is defined in continuous space and time to facilitate efficient implementation via homogenisation (see below) and because the accuracy of the approximation can be formally assessed (or modeled) so that inference will not be sensitive to the choice of the time step or size of the grid cells. Furthermore, if modelling fine-scale spatial variability is important to the goals of the study, it is necessary to have the same (or finer) resolution grid as the desired scale of inference. Although several numerical methods exist to solve PDEs, they become computationally prohibitive at the discretisation required for fine-scale inference over expansive spatial domain because the matrix multiplication required to evaluate $\mathbf{u}_t = \mathbf{H}_t \mathbf{u}_{t-\Delta t}$ results in a computational cost that scales by a factor $q^2 r$ (Garlick et al. 2011).

To address this computational challenge, we apply the homogenisation technique (Powell & Zimmermann 2004; Garlick et al. 2011; Hooten et al. 2013; Appendix S1). Homogenisation of PDEs leads to more efficient numerical solutions

when the goal is to capture broad scale dynamics with rapid fine-scale spatial variability and has, therefore, recently been employed for statistical implementation of ecological models (Hooten *et al.* 2013; Powell & Bentz 2014). Homogenisation is an analytical approach that takes advantage of multiple scales to solve a governing PDE on small scales and derive a related PDE at the broad scale that accurately represents the integrated consequences of the small-scale solution behaviour. Consequently, the accuracy of homogenisation depends on the ratio between the scale at which habitat type varies and the mean dispersal distance of individuals over a characteristic time period. This is a degree of accuracy that is intrinsic to the process and is independent of the grid sizes, provided fine- and broad-scale grids accurately capture the associated variability. How broad-scale grids are chosen depends on dispersal rates and should be approximately an order of magnitude larger than the fine-scale resolution (see Garlick *et al.* 2011 for details and Appendix S1).

Homogenisation is a useful approach for ecological diffusion models because the implementation is conceptually simple and involves solving a regularised diffusion PDE with analytically averaged diffusion coefficients over a much coarser grid (Garlick *et al.* 2011; Hooten *et al.* 2013). Another benefit of the ecological diffusion model is that homogenisation can be accomplished using 'weak' (e.g. piecewise continuous) solutions, so the approach is readily applicable to real-world situations with habitat discontinuities (see Garlick *et al.* 2011). Details on implementing the homogenised ecological diffusion PDE and step-by-step R code are provided in Appendix S1. In the next section, we illustrate a hierarchical Bayesian implementation of the ecological diffusion PDE to understand and forecast the geographic spread and growth in the prevalence of a disease.

Mechanistic forecasting of CWD

Chronic wasting disease is a fatal transmissible spongiform encephalopathy that occurs in cervids (Williams & Young 1980). In the state of Wisconsin, CWD was first detected in white-tailed deer in 2001 as a result of the Wisconsin Department of Natural Resources surveillance efforts. Continued surveillance of CWD has resulted in a large spatio-temporal data set with dense spatial coverage (Fig. 2). Using this data set, we obtained records from 103,256 tested deer (2562
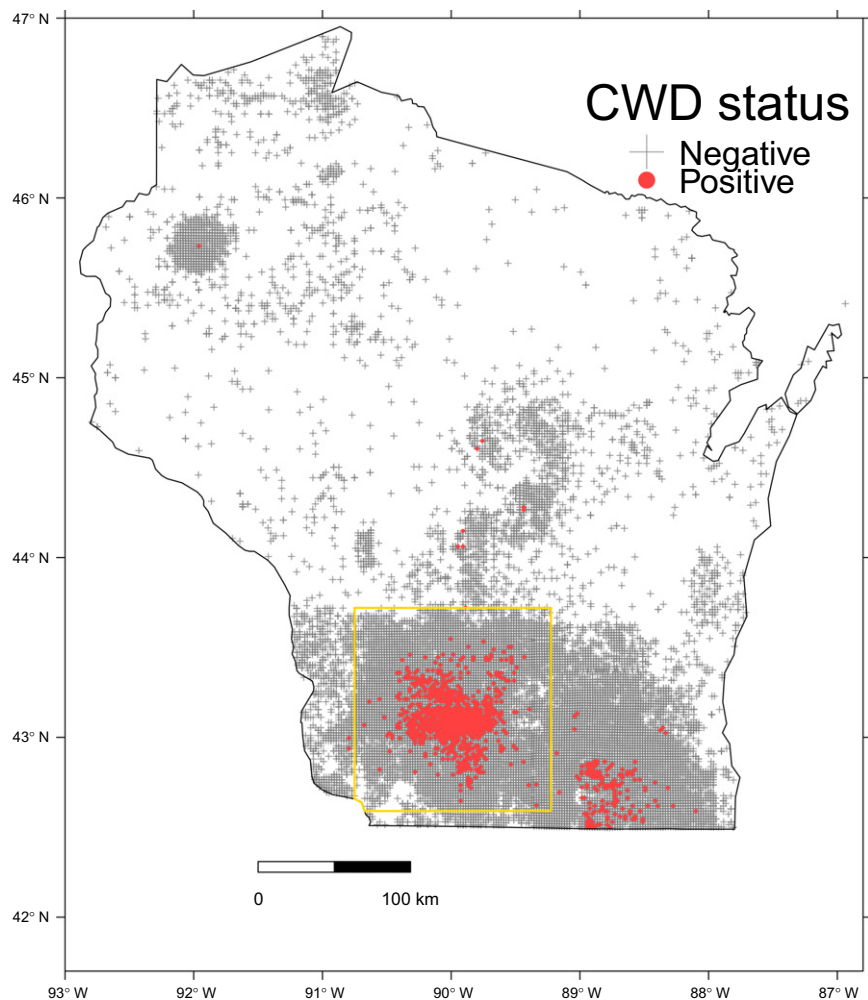

Figure 2 Location and infection status of white-tailed deer tested for chronic wasting disease in Wisconsin, USA from 2002–2014. Our models and out-of-sample validations were applied to 103,526 tested deer (2562 positive cases) with locations contained within the golden box. Only tested deer with location accuracy at the section-level (or better) are displayed. See Appendix S3 for an animation of this figure.
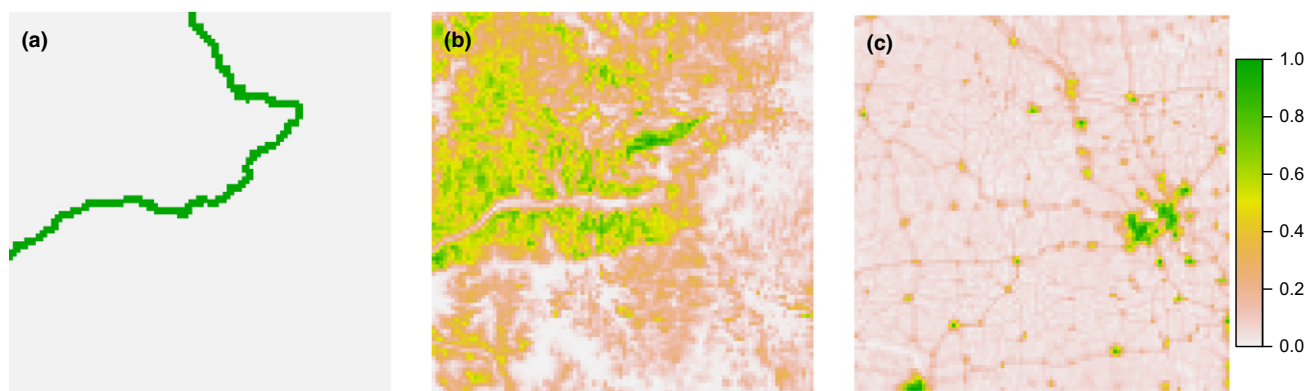
Figure 3 Spatial covariates used to model the growth and diffusion in the prevalence of chronic wasting disease in white-tailed deer. The three covariates include a categorical covariate of the Wisconsin River corridor (a), the proportion of hardwood forest (b), and human development (c; roads, houses, etc) within each $2.6\,\mathrm{km}^2$ grid cell.

positive deer) that were sampled within a 15,539 km^2 region in the southwestern portion of Wisconsin and had location information collected at the public land survey system section-level or better (Fig. 2; note a section of land is a square with an area of 2.6 km^2). We used the 2011 National Land Cover Dataset to calculate two landscape risk factors: proportion of hardwood forest and human development within a grid cell that was the same area as a section of land (2.6 km^2, hereafter forest and development covariates; Fig. 3; Homer *et al.* 2015). Based on a temporal animation of the spatial distribution of tested deer, the Wisconsin River appeared to potentially affect the movement and intensity of CWD-positive deer (Animation C1, Appendix S3); therefore, we included a categorical covariate that indicated if a 2.6 km^2 grid cell contained the Wisconsin River corridor (hereafter river covariate; Fig. 3; Smith *et al.* 2002; Wheeler & Waller 2008). In addition to spatial covariates, we also used the sex and age of the sampled deer as individual-level covariates.

For initial model fitting and statistical inference, we randomly selected 67% of the data from the first 10 years (2002–2011) of surveillance (n = 61,910; hereafter in-sample data). We evaluated the ability of our model to make accurate forecasts using data collected during the last 3 years of surveillance (2012–2014; n = 10,661; hereafter out-of-sample forecast data). We also evaluated the predictive ability of our model, using the 33% of the data from the first 10 years that were not used for fitting our model (n = 30,955; hereafter out-of-sample prediction data).

When evaluating the predictive ability or forecast proficiency of a model (Petchey *et al.* 2015), it is important to use scoring functions that are local and proper to guarantee an honest comparison (Gneiting & Katzfuss 2014). To facilitate comparison of our ecological diffusion model with other non-Bayesian approaches, we calculated the local and proper Bernoulli deviance using the posterior mean of the probability of infection (i.e. $-2\sum_{j=1}^{J}(y_j\log(p_j) + (1 - y_j)\log(1 - p_j))$, where y_j is the jth out-of-sample observation and p_j is the posterior mean probability of infection; Gneiting & Raftery 2007; Gneiting 2011; Hooten & Hobbs 2015). Akin to in-sample predictive scores (e.g. Akaike information criterion), lower values of the deviance indicate a model with superior

predictive ability (or forecast proficiency; unlike in-sample predictive scores, the deviance measures the predictive ability of a model against out-of-sample data rather than estimating the predictive ability, using a penalty term and in-sample data that were also used to fit the model (Hooten & Hobbs 2015).

As an initial hypothesis, we expect that the geographic spread in the prevalence of CWD among white-tailed deer was driven by the movement of individuals away from a central location. For example, white-tailed deer demonstrate habitat preferences that, from an Eulerian perspective, result in a heterogeneous geographic distribution in the abundance. When viewed from a Lagrangian perspective, the heterogeneous geographic distribution of white-tailed deer is a result of individuals moving quickly through habitat of poor quality and congregating in more favourable habitat. During the outbreak of a disease such as CWD, the transmission and spread may be driven by the movement of individuals; thus, ecological diffusion is a realistic model for a population-level spatio-temporal process representing the latent risk of CWD (Garlick *et al.* 2014).

To understand and forecast the growth and geographic spread in the prevalence of CWD, we used the hierarchical model:

$$y_i \sim \mathrm{Bernoulli}(p_i) \tag{5}$$

$$g(p_i) = u(\mathbf{s}_i, t_i)e^{\mathbf{x}_i'\beta} \tag{6}$$

$$\frac{\partial}{\partial t}u(\mathbf{s}, t) = \left(\frac{\partial^2}{\partial s_1^2} + \frac{\partial^2}{\partial s_2^2}\right)[\mu(\mathbf{s})u(\mathbf{s}, t)] + \lambda(\mathbf{s})u(\mathbf{s}, t) \tag{7}$$

$$\log(\mu(\mathbf{s})) = \alpha_0 + \mathbf{z}(\mathbf{s})'\boldsymbol{\alpha} \tag{8}$$

$$\lambda(\mathbf{s}) = \gamma_0 + \mathbf{w}(\mathbf{s})'\boldsymbol{\gamma}, \tag{9}$$

where y_i is equal to 1 if the ith deer is CWD-positive and 0 otherwise. The probability that a deer is CWD-positive (p_i) depends on the spatio-temporal effect $u(\mathbf{s}_i, t_i)$ defined by the PDE in eqn 7. For the ecological diffusion PDE, the spatio-temporal effect $u(\mathbf{s}_i, t_i)$ is greater than zero for all $\mathbf{s}$, t; the vector $\mathbf{x}_i$, includes covariates for sex and age of the tested deer and the quantity $e^{\mathbf{x}_i'\beta}$ scales $u(\mathbf{s}_i, t_i)$ depending on characteristics of the individual deer. Scaling $u(\mathbf{s}_i, t_i)$ by a quantity that depends on covariates specific to the infected and non-infected individuals mimics the dynamics of CWD because individual-

level risk factors should not influence the probability of infection unless the disease process has spread to the location where the individual is located (i.e. $u(\mathbf{s}_i, t_i) > 0$). The product, $u(\mathbf{s}_i, t_i)e^{\mathbf{x}_i'\boldsymbol{\beta}}$, is greater than zero for all $\mathbf{s}$, t, and therefore the inverse of the link function $g(\cdot)$ maps the positive real line $[0, \infty)$ to a probability between 0 and 1. We used a truncated cumulative normal distribution as a link function (i.e. the inverse of $g(\cdot)$ is $g^{-1}(x) = \sqrt{\frac{2}{\pi}}\int_0^x e^{-\frac{1}{2}x^2} dx$). The diffusion rate ($\mu(\mathbf{s})$) in eqn 7 is inversely related to residence times in various habitats and depends on an intercept term (α_0), coefficients ($\boldsymbol{\alpha} \equiv (\alpha_1, \ldots, \alpha_{p_\alpha})'$), and $\mathbf{z}(\mathbf{s})$, which is a vector that contains the spatial covariates forest, development, and river. Expected residence times are always positive, thus $\mu(\mathbf{s}) \geq 0$, for all $\mathbf{s}$ which motivates the log link function in eqn 8. The growth rate $\lambda(\mathbf{s})$ depends on an intercept term (γ_0), coefficients ($\boldsymbol{\gamma} \equiv (\gamma_1, \ldots, \gamma_{p_\gamma})'$), and $\mathbf{w}(\mathbf{s})$, a vector that also contains the spatial covariates forest, development, and river. To allow the growth rate in eqn 9 to be positive or negative, we used the identity link so that the support of $\lambda(\mathbf{s})$ was $(-\infty, \infty)$.

For regression coefficients $\boldsymbol{\beta}$, $\boldsymbol{\alpha}$, and $\boldsymbol{\gamma}$, we used the following priors: $\boldsymbol{\beta} \sim N(0, 10\mathbf{I})$, $\boldsymbol{\alpha} \sim N(0, 10\mathbf{I})$, and $\boldsymbol{\gamma} \sim N(0, 10\mathbf{I})$, where $\mathbf{I}$ is the identity matrix. For the intercept terms, we used the prior $\alpha_0 \sim N(0, 10)$ and $\gamma_0 \sim N(0, 10)$. We used the normal distribution with mean 0 and variance 10 as priors to result in a minimal amount of regularisation (shrinkage) of the intercept and regression coefficients. Alternatively, to perform covariate selection and possibly obtain a model with a higher accuracy for prediction and forecasting, one could choose the optimal amount of regularisation (i.e. variance of the normal prior distribution) using methods discussed by Hooten & Hobbs (2015).

Solving PDEs require that boundary conditions and the initial state be specified or estimated. Within the cells that occur at the boundary, we assume $u(\mathbf{s}, t) = 0$ for all t. For the initial state, we use

$$u(\mathbf{s}, 0) = \frac{\theta e^{\frac{-|\mathbf{s}-\mathbf{d}|^2}{\phi^2}}}{\int_{\mathcal{S}} e^{\frac{-|\mathbf{s}-\mathbf{d}|^2}{\phi^2}} d\mathbf{s}}, \tag{10}$$

which is a scaled bivariate Gaussian kernel with compact (truncated) support centred at a point with coordinate $\mathbf{d}$ where $|\mathbf{s} - \mathbf{d}|$ is the Euclidean distance in meters and the dispersion and scale are controlled by ϕ and θ respectively. We assigned ϕ and θ the priors $\phi \sim TN(0, 10^6)$ and $\theta \sim TN(0, 10^6)$ (TN refers to a normal distribution truncated below zero). The coordinate $\mathbf{d}$ in eqn 10 was assumed to be the centroid of the locations where positive deer were sampled in 2002, however, this parameter could be treated as unknown and then estimated along with other parameters in eqns 5–10. Alternatively, the initial state could be specified, using a point process distribution and the location(s) where CWD was introduced within the study area could be estimated (Hefley *et al.* 2017c).

The location accuracy of the sampled deer (section level) results in a natural spatial grid with $2.6\,\text{km}^2$ cells within the study area (Fig. 2). Using traditional finite-difference methods with an enlarged grid of $25,900\,\text{km}^2$ to solve eqn 7 results in

10,000 cells and would be computationally challenging to implement. Instead, we use homogenisation and solve the piecewise constant diffusion PDE (eqn S7) over a spatial grid with $65\,\text{km}^2$ cells (400 total cells). The homogenisation technique allows the solution from eqn S7 to be analytically downscaled to the $2.6\,\text{km}^2$ cells using eqn S10. We use a time step (Δt) of 3 months in the finite-difference implementation. To assess the sensitivity of statistical inference to errors introduced by discretisation and homogenisation, we evaluated several different resolutions of the grid and Δt (results not shown).

We implemented the hierarchical model from eqns 5–10 using a Markov chain Monte Carlo (MCMC) algorithm coded in R (R Core Team 2017). We obtained 250,000 samples from a single chain using the MCMC algorithm. Fitting the model at the $2.6\,\text{km}^2$ scale using homogenisation required approximately 10 min to acquire 1000 MCMC samples on a laptop computer with a 2.8 GHz quad-core processor, 16 GB of RAM, and optimised basic linear algebra subprograms.

To assess the fit of our model, we used a posterior predictive check that involved comparing the posterior prediction and forecast distributions of the probabilities of infection to the empirical distributions obtained from the out-of-sample data (Gelman *et al.* 1996). We calculated the empirical distributions using only out-of-sample data and a beta-binomial model with a uniform prior on the probability of infection (see Appendix S2 for details).

Comparison with generalised additive models

We compared our mechanistic Bayesian hierarchical model to a generalised additive model (GAM). GAMs use a phenomenological regression-based framework that is well-developed within the statistics literature (e.g. Hastie & Tibshirani 1986, 1990; Wood 2006) and widely used in ecology (e.g. Guisan *et al.* 2002; Wood & Augustin 2002). We compared a GAM to our hierarchical model because (1) the linear structure of the GAM is easy to interpret for statistical inference, (2) basis functions can be used to explicitly model the spatial and temporal process (Hefley *et al.* 2017b), and (3) new dimension reduction and estimation techniques facilitate efficient implementation for large data sets (Wood *et al.* 2015, 2017). For comparison with our mechanistic Bayesian hierarchical model, we implemented a GAM that can be formally written as follows:

$$\begin{aligned} y_i &\sim \text{Bernoulli}(p_i) \\ g(p_i) &= \mathbf{x}_i\boldsymbol{\beta} + \eta_t + \eta_s, \end{aligned} \tag{11}$$

where, as before, y_i is equal to 1 if the ith deer is CWD-positive and 0 otherwise. The probability that a deer is CWD-positive (p_i) is transformed, using a link function $g(\cdot)$ and depends on both the individual level (sex and age) and spatial covariates (river, forest and development) which are included in the vector $\mathbf{x}_i$. Unlike our mechanistic hierarchical model, the GAM framework does not explicitly distinguish between individual-level and spatial covariates. The effect of time (η_t) and spatial location (η_s) are modeled using reduced dimension thin plate regression splines. Briefly, thin plate regression splines estimate a smooth function that emulates the spatial and temporal dynamics by specifying flexible basis functions

(Wood 2006 pp. 150–156; Hefley *et al.* 2017b). Details associated with the exact model specifications and computations are reported in Appendix S2.

Because GAMs provide a flexible semiparametric approach, they can emulate the spatio-temporal dynamics generated by the growth and diffusion of CWD. As a result, we expect the GAM to have superior predictive ability compared to our Bayesian hierarchical model when tested against the out-of-sample prediction data. GAMs, however, are phenomenological and lack a foundation for providing principled forecasts. For example, our choice of thin plate regression splines results in an arbitrary linear trend when forecasting to out-of-sample data from 2012–2014. As with any phenomenological modelling approach that uses basis functions (e.g. polynomial regression), forecasts (or extrapolations) beyond the range of the data are not reliable.

Unlike the spatio-temporal effect that was specified using the ecological diffusion PDE, the spatial and temporal effects in eqn 11 are separable (modeled individually) and do not depend on covariates (e.g. river, forest, and development), which can have important implications for statistical inference. When implementing the GAM, covariates that are spatially (or temporally) indexed and structured are sometimes collinear with the smooth spatial (or temporal) effect (Wood 2006, p. 176). When using basis functions to model a spatial (or temporal) process, it is important to check for collinearity between the basis vectors and the spatial (or temporal) covariates (Hefley *et al.* 2017ab). Collinearity between the basis vectors and the spatial (or temporal) covariates can produce surprising and counterintuitive results (e.g. Hodges & Reich 2010; Fieberg & Ditmer 2012; Hanks *et al.* 2015; Hefley *et al.* 2016). For example, Kühn (2007) and Bini *et al.* (2009) noted that the coefficient estimates for some spatial covariates obtained from regression models that account of autocorrelation can 'shift' or change sign when compared to non-spatial models. To illustrate this, we removed the spatial and temporal effects from the GAM and fit a model that contained only the individual-level and spatial covariates (i.e. $g(p_i) = \mathbf{x}_i\boldsymbol{\beta}$ in eqn 11). For comparison, we also report the predictive and forecasting ability of the GAM when the spatial and temporal effects were removed.

Comparison with boosted regression trees

We compared our Bayesian hierarchical model to boosted regression trees (BRTs). BRTs are a well-studied machine learning method in the statistics literature (e.g. Friedman *et al.* 2000; Bühlmann & Hothorn 2007) and widely used in ecology (e.g. Elith *et al.* 2008). We compared BRTs to our Bayesian hierarchical model because (1) BRTs are considered one of the best off-the-shelf methods for binary data (Hastie *et al.* 2009, p. 340), (2) unlike GAMs, BRTs can handle sharp discontinuities and may capture abrupt transitions that are characteristic of ecological diffusion (Elith *et al.* 2008), and (3) BRTs are computationally feasible for large data sets (Hastie *et al.* 2009).

Briefly, BRTs are an additive modelling approach where, in our example, the probability that a deer is CWD-positive (p_i) is transformed using a link function $g(\cdot)$ and specified as

$$g(p_i) = \sum_{m=1}^{M} \beta_m T(\mathbf{x}_i, \mathbf{s}_i, t_i, \boldsymbol{\zeta}_m), \tag{12}$$

where β_m is the *m*th basis coefficient for a simple regression tree, $T(\mathbf{x}_i, \mathbf{s}_i, t_i, \boldsymbol{\zeta}_m)$, that depends on the parameters $\boldsymbol{\zeta}_m$. Each regression tree partitions the spatial and temporal domains as well as the domains of the individual-level and spatial covariates. Estimation of the basis coefficients proceeds, using a forward stagewise algorithm. Elith *et al.* (2008) provides a lucid guide to BRTs for ecological data and a technical introduction can be found in Hastie *et al.* (2009, ch. 10). Details associated with tuning and computation are reported in Appendix S2.

Similar to GAMs, BRTs are phenomenological and lack a mechanism for providing principled forecasts. As evident from eqn 12, predictions and forecasts obtained from the BRTs depend on a sum of *M* simple regression trees weighted by the basis coefficients. As a result, each simple regression tree contributes an arbitrary static prediction (or forecast) for all covariates, spatial locations, or times that do not occur within the domain of the data used for estimation. Unlike GAMs, BRTs do not have interpretable parameters, rather, inference is obtained from plots of partial dependence which are constructed using predictions from eqn 12 for one or two features of interest (i.e. covariates, space, time) obtained while holding the remaining features constant.

RESULTS

Mechanistic forecasting of CWD

The prevalence of CWD was highest for male deer and increased with age (Fig. B2, Appendix S2). For all deer, the prevalence of CWD was highest in the centre of the study region and exhibited both diffusion and growth over time (Fig. 4; Animation C2, Appendix S3). The diffusion and growth in the prevalence of CWD within the study area were most heavily influenced by the forest and river covariates (Fig. 5). Most notably, our results demonstrate that CWD would diffuse at a rate 1.93 (1.19, 3.02; 95% equal-tail credible interval) times faster within the Wisconsin River corridor when compared to an area outside of the corridor. Our results also indicate that the annual growth rate was 0.07 (−0.02, 0.14) in areas where the forest cover was 19% (average value in the study area), but was 0.87 (0.59, 1.21) at the maximum amount of forest within the study area (98%; Fig. 5). The Bernoulli deviance for the out-of-sample forecast was 5753, while the deviance for the out-of-sample prediction was 4564. Based on the posterior predictive check, our model-based prediction and forecast distributions of the mean probability of infection appear to match the out-of-sample empirical estimates well (Fig. 6). The match between empirical and model-based statistics, suggests that eqns 5–10 captured important spatio-temporal dynamics and are a good fit to the data.

Comparison with generalised additive models

The GAM emulates the spatio-temporal trends within the data generated by the dynamics of CWD (Fig. B3; Appendix S2).
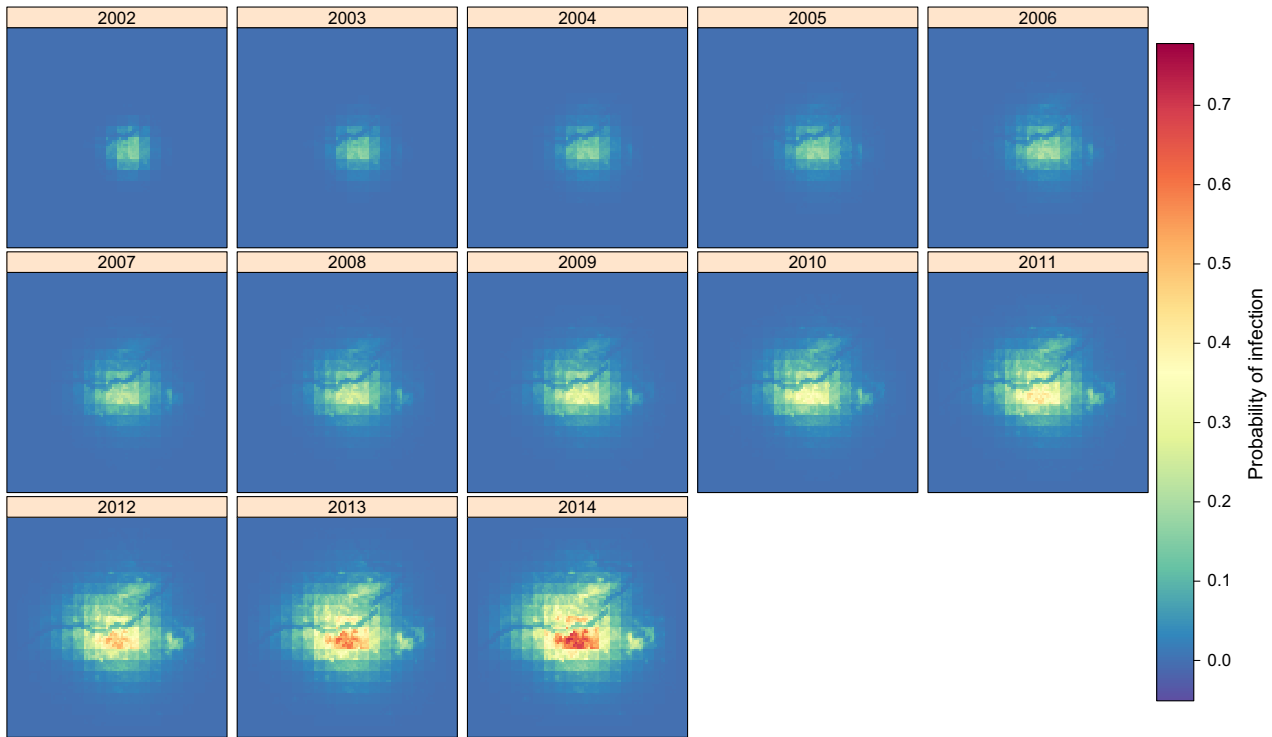
Figure 4 Spatial time series showing the predicted (2002–2011) and forecasted (2012–2014) probability of chronic wasting disease infection in male white-tailed deer that are 4 years of age or older (see Fig. 2 for study area). The predicted and forecasted probability of infection was obtained from the ecological diffusion and growth model (eqns 5–10), which depended on the covariates shown in Fig. 3 and was fit to a portion of the data shown in Fig. 2. Notable features include the lower observed probability of infection within the Wisconsin River corridor (Fig. 3a) due to a increased rate of diffusion (Fig. 5a). Note that the observed data used to fit the model were collected during the period 2002–2011.
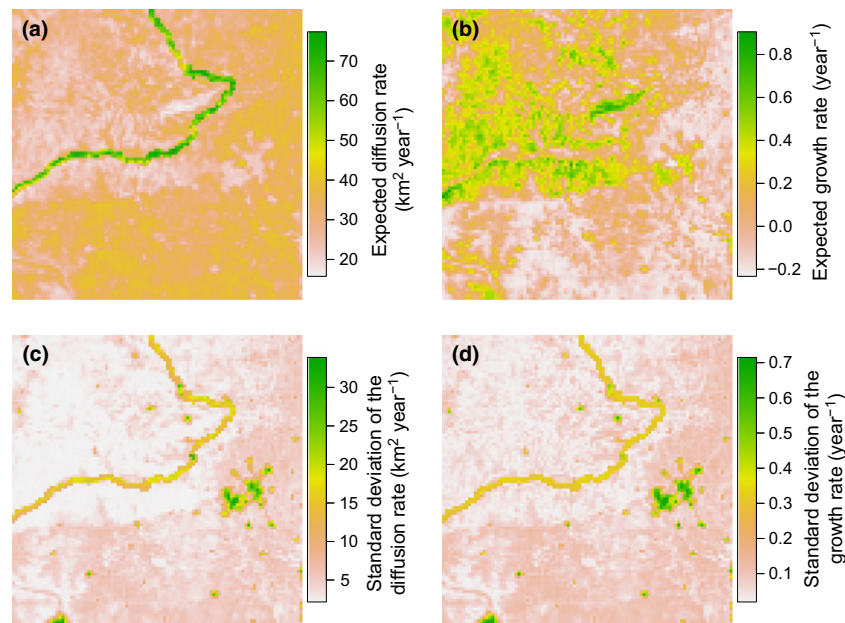
Figure 5 Posterior expectation and standard deviation of the diffusion rate $\mu(\mathbf{s})$ (panels (a) and (c)) and growth rate $\lambda(\mathbf{s})$ (panels (b) and (d)) from the ecological diffusion and growth model presented in eqns 5–10 and fit to the data in Fig. 2. Both the diffusion rate and growth rate vary due to the spatial covariates shown in Fig. 3.

The Bernoulli deviance for the out-of-sample forecast was 6103, while the deviance for the out-of-sample prediction was 4431; thus, the GAM has a lower accuracy for forecasting when compared to our Bayesian hierarchical model, but a higher accuracy for prediction. The regression coefficient estimates for the spatial covariates forest, development, and river were −0.32 (−0.90, 0.24; 95% confidence interval), 0.42 (−1.89, 2.72), and −0.70 (−1.14, −0.25), respectively.
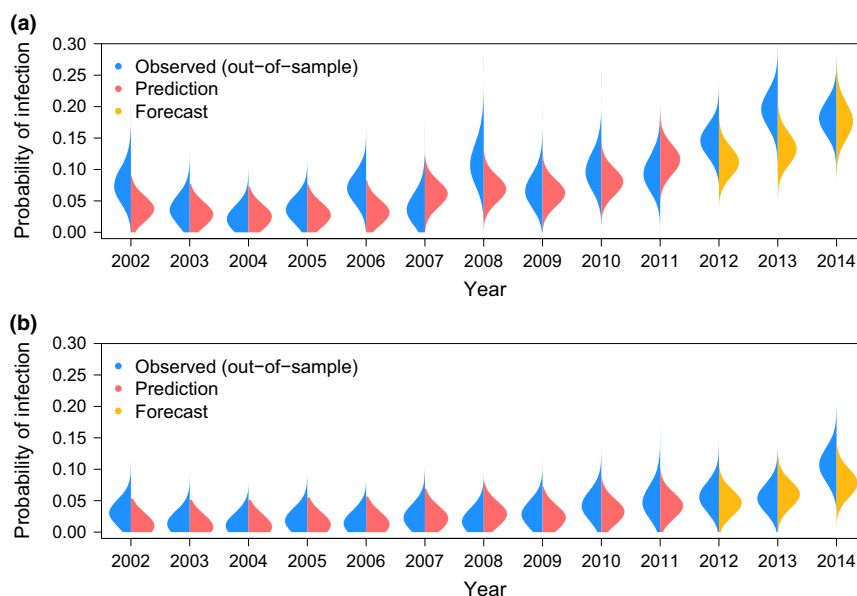
Figure 6 Empirical distributions obtained from out-of-sample data (blue) and posterior prediction (red) and forecast (yellow) distributions of the mean probability of infection for male (panel a) and female (panel b) white-tailed deer that are 3 years of age within the study area. This plot demonstrates agreement between the empirical estimates and predictions and forecasts from our model and can be used to assess goodness of fit. Note that the observed data used to fit the model were collected during the period 2002–2011.

When we removed the spatial and temporal effects from the GAM and fit a model that contained only the individual-level and spatial covariates, we found that the coefficient estimates for the spatial covariates forest, development, and river were 2.86 (2.50, 3.22), 1.72 (0.42, 3.03), and −0.30 (−0.70, 0.10), respectively; thus, the coefficient estimate associated with the forest covariate shifted from a small negative number when using the GAM to a larger positive and statistically significant estimate due to correlation with the basis vectors (see Appendix S2 for details). When we removed the spatial and temporal effects from the GAM, the Bernoulli deviance calculated using the out-of-sample prediction and forecast data was 5176 and 7447 respectively; thus, indicating that the spatial and temporal effect increases both the predictive and forecasting ability of the model.

Comparison with boosted regression trees

The BRTs emulate the spatio-temporal trends within the data, but the predicted probability of infection is more variable between adjacent cells when compared to the GAM and forecasts are temporally static for any year after 2011 (Fig. B4; Appendix S2). The Bernoulli deviance for our out-of-sample forecast was 6307, while the deviance for out of sample prediction was 4408; thus, BRTs have the lowest accuracy for forecasting, but the highest accuracy for prediction.

DISCUSSION

Modelling dynamic ecological processes based on mechanistically motivated PDEs yields more accurate forecasts and scientific insights that would not be possible when using phenomenological regression-based models, such as GAMs or BRTs. For example, when using the ecological diffusion PDE,

the temporal change in prevalence of CWD can be separated into growth and diffusion components. The ability to obtain inference for the growth and diffusion components has important implications for the management of CWD. For example, a positive growth rate suggests that an area is a 'source' in the sense that an area with a positive growth rate will contribute to an overall increase in the prevalence of CWD within the study area over time. Similarly, a negative growth rate suggests that an area is a 'sink' in that the area will contribute to an overall decrease in the prevalence of CWD within the study area. As another example, the diffusion rate shown in Fig. 5 has physically interpretable units ($km^2 year^{-1}$) that are inversely related to the residence time ($km^{-2} year$) of the disease process (Garlick *et al.* 2011).

Partial differential equations play an important role in applied and theoretical ecology (Holmes *et al.* 1994). Diffusion models have been used to understand spatio-temporal patterns in ecology and can be used to describe a range of phenomena such as population growth, invasion, disease transmission, and dispersal (Table 1). Diffusion models, however, are just one of the many manifestations of hypotheses proposed by ecologists as PDEs. Testing such hypotheses requires reconciling theory with observations. Since the seminal work of Wikle (2003), ecologists have possessed the tools to fit mechanistically motivated PDEs to ecological data within a rigorous hierarchical modelling framework, yet the use of such models for testing ecological theory and providing ecological forecasts has been gradual. In our experience, implementing statistical PDE models is challenging because doing so requires (1) development of ecological theory, (2) large spatio-temporal data sets with dense coverage in both the spatial and temporal domains, and (3) specialised computational techniques. Although new software is making statistical inference from PDEs more accessible by lowering the need

for specialised mathematics and programming (e.g. Pienaar & Varughese 2016), fitting statistically rigorous PDE models to ecological data will most often require interdisciplinary collaboration between theoretical ecologists, applied ecologists, statisticians and mathematicians. The reward of such collaborations is that defensible forecasts can be made and theories can be tested when we confront models with data.

ACKNOWLEDGEMENTS

AUTHORSHIP

TJH, MBH, RER, DPW, and JAP conceived the study. JAP applied the mathematical methods. TJH and MBH applied the statistical methods. TJH conducted the statistical analysis and wrote the manuscript. All authors contributed substantially to revisions.

DATA ACCESSIBILITY STATEMENT

The data set that contains the location and disease status of the white-tailed deer used in our analysis is owned by the Wisconsin Department of Natural Resources. Please contact Tamara Ryan (tamara.ryan@wisconsin.gov) for access.

REFERENCES

Arab, A. (2007). Hierarchical spatio-temporal models for environmental processes. PhD Thesis, University of Missouri-Columbia.

Bini, M., Diniz-Filho, J.A.F., Rangel, T.F., Akre, T.S., Albaladejo, R.G., Albuquerque, F.S. *et al.* (2009). Coefficient shifts in geographical ecology: an empirical evaluation of spatial and non-spatial regression. *Ecography*, 32(2), 193–204.

Bolker, B.M., Brooks, M.E., Clark, C.J., Geange, S.W., Poulsen, J.R., Stevens, M.H.H. *et al.* (2009). Generalized linear mixed models: a practical guide for ecology and evolution. *Trends Ecol. Evol.*, 24(3), 127–135.

Broms, K.M., Hooten, M.B., Johnson, D.S., Altwegg, R. & Conquest, L.L. (2016). Dynamic occupancy models for explicit colonization processes. *Ecology*, 97(1), 194–204.

Bühlmann, P. & Hothorn, T. (2007). Boosting algorithms: regularization, prediction & model fitting. *Stat. Sci.*, 22(4), 477–505.

Cangelosi, A.R. & Hooten, M.B. (2009). Models for bounded systems with continuous dynamics. *Biometrics*, 65(3), 850–856.

Chandrasekhar, S. (1943). Stochastic problems in physics and astronomy. *Rev. Mod. Phys.*, 15(1), 1.

Chkrebtii, O.A., Campbell, D.A., Calderhead, B. & Girolami, M.A. (2016). Bayesian solution uncertainty quantification for differential equations. *Bayesian Anal.*, 11(4), 1239–1267.

Clark, J.S. (2005). Why environmental scientists are becoming Bayesians. *Ecol. Lett.*, 8(1), 2–14.

Clark, J.S., Carpenter, S.R., Barber, M., Collins, S., Dobson, A., Foley, J.A. *et al.* (2001). Ecological forecasts: an emerging imperative. *Science*, 293(5530), 657–660.

Cressie, N. & Wikle, C.K. (2011). *Statistics for Spatio-Temporal Data*. John Wiley & Sons, Hoboken.

Cressie, N., Calder, C.A., Clark, J.S., Ver Hoef, J.M. & Wikle, C.K. (2009). Accounting for uncertainty in ecological analysis: the strengths and limitations of hierarchical statistical modeling. *Ecol. Appl.*, 19(3), 553–570.

Elith, J. & Leathwick, J.R. (2009). Species distribution models: ecological explanation and prediction across space and time. *Annu. Rev. Ecol. Evol. Syst.*, 40(1), 677.

Elith, J., Leathwick, J.R. & Hastie, T. (2008). A working guide to boosted regression trees. *J. Anim. Ecol.*, 77(4), 802–813.

Farlow, S. (1993). *Partial Differential Equations for Scientists and Engineers*. Dover Publicaitons, New York.

Fieberg, J. & Ditmer, M. (2012). Understanding the causes and consequences of animal movement: a cautionary note on fitting and interpreting regression models with time-dependent covariates. *Methods Ecol. Evol.*, 3(6), 983–991.

Friedman, J., Hastie, T. & Tibshirani, R. (2000). Additive logistic regression: a statistical view of boosting. *Ann. Stat.*, 28(2), 337–407.

Garlick, M.J., Powell, J.A., Hooten, M.B. & McFarlane, L.R. (2011). Homogenization of large-scale movement models in ecology. *Bull. Math. Biol.*, 73(9), 2088–2108.

Garlick, M.J., Powell, J.A., Hooten, M.B. & MacFarlane, L.R. (2014). Homogenization, sex, and differential motility predict spread of chronic wasting disease in mule deer in southern Utah. *J. Math. Biol.*, 69(2), 369–399.

Gelman, A., Meng, X.-L. & Stern, H. (1996). Posterior predictive assessment of model fitness via realized discrepancies. *Stat. Sin.*, 6(4), 733–760.

Gneiting, T. (2011). Making and evaluating point forecasts. *J. Am. Stat. Assoc.*, 106(494), 746–762.

Gneiting, T. & Katzfuss, M. (2014). Probabilistic forecasting. *Ann. Rev. Stat. Appl.*, 1, 125–151.

Gneiting, T. & Raftery, A.E. (2007). Strictly proper scoring rules, prediction, and estimation. *J. Am. Stat. Assoc.*, 102(477), 359–378.

Guisan, A., Edwards, T.C., & Hastie, T. (2002). Generalized linear and generalized additive models in studies of species distributions: setting the scene. *Ecol. Model.*, 157(2), 89–100.

Hanks, E.M., Schliep, E.M., Hooten, M.B. & Hoeting, J.A. (2015). Restricted spatial regression in practice: geostatistical models, confounding, and robustness under model misspecification. *Environmetrics*, 26(4), 243–254.

Hastie, T. & Tibshirani, R. (1986). Generalized additive models. *Stat. Sci.*, 3(3), 297–310.

Hastie, T.J. & Tibshirani, R.J. (1990). *Generalized Additive Models*. CRC Press, Boca Raton, Florida.

Hastie, T., Tibshirani, R. & Friedman, J. (2009). *The Elements of Statistical Learning*. Springer, Berlin.

Hastings, A., Cuddington, K., Davies, K., Dugaw, C.J., Elmendorf, S., Freestone, A. *et al.* (2005). The spatial spread of invasions: new developments in theory and evidence. *Ecol. Lett.*, 8(1), 91–101.

Hefley, T.J. & Hooten, M.B. (2016). Hierarchical species distribution models. *Current Landsc. Ecol. Reports*, 1(2), 87–97.

Hefley, T.J., Hooten, M.B., Drake, J.M., Russell, R.E. & Walsh, D.P. (2016). When can the cause of a population decline be determined? *Ecol. Lett.*, 19(11), 1353–1362.

Hefley, T.J., Hooten, M.B., Hanks, E.M., Russell, R.E., & Walsh, D.P. (2017a). The Bayesian group lasso for confounded spatial data. *J. Agric. Biol. Environ. Stat.*, 22(1), 42–59.

Hefley, T.J., Broms, K.M., Brost, B.M., Buderman, F.E., Kay, S., Scharf, H.R., Tipton, J.R. & P.J., Williams, M.B. Hooten, (2017b). The basis function approach for modeling autocorrelation in ecological data. *Ecology*, 98(3), 632–646.

Hefley, T.J., Hooten, M.B., Hanks, E.M., Russell, R.E. & Walsh, D.P. (2017c). Dynamic spatio-temporal models for spatial data. *Spat. Stat.* doi: 10.1016/j.spasta.2017.02.005, in press.

Hilborn, R. & Mangel, M. (1997). *The Ecological Detective: Confronting Models with Data*. Princeton University Press, Princeton, New Jersey, USA.

Hobbs, N.T. & Hooten, M.B. (2015). *Bayesian Models: A Statistical Primer for Ecologists*. Princeton University Press, Princeton, New Jersey, USA.

Hodges, J.S. & Reich, B.J. (2010). Adding spatially-correlated errors can mess up the fixed effect you love. *Am. Stat.*, 64(4), 325–334.

Holmes, E.E., Lewis, M.A., Banks, J. & Veit, R. (1994). Partial differential equations in ecology: spatial interactions and population dynamics. *Ecology*, 75(1), 17–29.

Homer, C.G., Dewitz, J.A., Yang, L., Jin, S., Danielson, P., Xian, G. *et al.* (2015). Completion of the 2011 National Land Cover Database for the conterminous United States-Representing a decade of land cover change information. *Photogramm. Eng. Remote Sensing*, 81(5), 345–354.

Hooten, M. & Hobbs, N. (2015). A guide to Bayesian model selection for ecologists. *Ecol. Monogr.*, 85(1), 3–28.

Hooten, M.B. & Wikle, C.K. (2008). A hierarchical Bayesian non-linear spatio-temporal model for the spread of invasive species with application to the Eurasian Collared-Dove. *Environ. Ecol. Stat.*, 15(1), 59–70.

Hooten, M.B. & Wikle, C.K. (2010). Statistical agent-based models for discrete spatio-temporal systems. *J. Am. Stat. Assoc.*, 105(489), 236–248.

Hooten, M.B., Wikle, C.K., Dorazio, R.M. & Royle, J.A. (2007). Hierarchical spatiotemporal matrix models for characterizing invasions. *Biometrics*, 63(2), 558–567.

Hooten, M.B., Garlick, M.J. & Powell, J.A. (2013). Computationally efficient statistical differential equation modeling using homogenization. *J. Agric. Biol. Environ. Stat.*, 18(3), 405–428.

Hotelling, H. (1927). Differential equations subject to error, and population estimates. *J. Am. Stat. Assoc.*, 22(159), 283–314.

Kühn, I. (2007). Incorporating spatial autocorrelation may invert observed patterns. *Divers. Distrib.*, 13(1), 66–69.

Levin, S.A. (1992). The problem of pattern and scale in ecology. *Ecology*, 73(6), 1943–1967.

Matthiopoulos, J., Fieberg, J., Aarts, G., Beyer, H.L., Morales, J.M. & Haydon, D.T. (2015). Establishing the link between habitat selection and animal population dynamics. *Ecol. Monogr.*, 85(3), 413–436.

May, R.M. (1973). *Stability and Complexity in Model Ecosystems*. Princeton University Press, Princeton, New Jersey, USA.

Moorcroft, P.R. & Barnett, A. (2008). Mechanistic home range models and resource selection analysis: a reconciliation and unification. *Ecology*, 89(4), 1112–1119.

Okubo, A. & Levin, S.A. (2001). *Diffusion and Ecological Problems: Modern Perspectives*. Springer-Verlag, New york.

Patlak, C.S. (1953). Random walk with persistence and external bias. *Bull. Math. Biophys.*, 15(3), 311–338.

Petchey, O.L., Pontarp, M., Massie, T.M., Kéfi, S., Ozgul, A., Weilenmann, M. *et al.* (2015). The ecological forecast horizon, and examples of its uses and determinants. *Ecol. Lett.*, 18(7), 597–611.

Pielke, R.A. & Conant, R.T. (2003). Best practices in prediction for decision-making: lessons from the atmospheric and earth sciences. *Ecology*, 84(6), 1351–1358.

Pienaar, E.A. & Varughese, M.M. (2016). Diffusionrgqd: an R package for performing inference and analysis on time-inhomogeneous quadratic diffusion processes. Available at: https://cran.r-project.org/web/packages/DiffusionRgqd/vignettes/DiffusionRgqd_Paper.pdf. Last accessed 28 December 2016.

Powell, J.A. & Bentz, B.J. (2014). Phenology and density-dependent dispersal predict patterns of mountain pine beetle (*Dendroctonus ponderosae*) impact. *Ecol. Model.*, 273, 173–185.

Powell, J.A. & Zimmermann, N.E. (2004). Multiscale analysis of active seed dispersal contributes to resolving Reid's paradox. *Ecology*, 85(2), 490–506.

R Core Team (2017). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Raftery, A.E. (2016). Use and communication of probabilistic forecasts. *Stat. Anal. Data Mining*, 9(6), 397–410.

Scheffer, M., Carpenter, S., Foley, J.A., Folke, C. & Walker, B. (2001). Catastrophic shifts in ecosystems. *Nature*, 413, 591–596.

Skellam, J.G. (1951). Random dispersal in theoretical populations. *Biometrika*, 38(1:2), 196–218.

Smith, D.L., Lucey, B., Waller, L.A., Childs, J.E. & Real, L.A. (2002). Predicting the spatial dynamics of rabies epidemics on heterogeneous landscapes. *Proc. Natl Acad. Sci.*, 99(6), 3668–3672.

Turchin, P. (1998). *Quantitative Analysis of Movement: Measuring and Modeling Population Redistribution in Animals and Plants*. Sinauer Associates, Sunderland, MA.

Uden, D.R., Allen, C.R., Angeler, D.G., Corral, L. & Fricke, K.A. (2015). Adaptive invasive species distribution models: a framework for modeling incipient invasions. *Biol. Invasions*, 17(10), 2831–2850.

Wheeler, D.C. & Waller, L.A. (2008). Mountains, valleys, and rivers: the transmission of raccoon rabies over a heterogeneous landscape. *J. Agric. Biol. Environ. Stat.*, 13(4), 388–406.

Wikle, C.K. (2003). Hierarchical Bayesian models for predicting the spread of ecological processes. *Ecology*, 84(6), 1382–1394.

Wikle, C.K. & Hooten, M.B. (2010). A general science-based framework for dynamical spatio-temporal models. *Test*, 19(3), 417–451.

Wikle, C.K., Berliner, L.M. & Cressie, N. (1998). Hierarchical Bayesian space-time models. *Environ. Ecol. Stat.*, 5(2), 117–154.

Williams, E. & Young, S. (1980). Chronic wasting disease of captive mule deer: a spongiform encephalopathy. *J. Wildl. Dis.*, 16(1), 89–98.

Williams, P.J., Hooten, M.B., Womble, J.N., Esslinger, G.G., Bower, M.R. & Hefley, T.J. (2017). An integrated data model to estimate spatio-temporal occupancy, abundance, and colonization dynamics. *Ecology*, 98, 328–336.

Wood, S. (2006). *Generalized Additive Models: An Introduction with R*. Chapman & Hall/CRC Texts in Statistical Science, Taylor & Francis, Boca Raton, Florida.

Wood, S.N. & Augustin, N.H. (2002). GAMs with integrated model selection using penalized regression splines and applications to environmental modelling. *Ecol. Model.*, 157(2), 157–177.

Wood, S.N., Goude, Y. & Shaw, S. (2015). Generalized additive models for large data sets. *J. Roy. Stat. Soc.: Ser. C (Appl. Stat.)*, 64(1), 139–155.

Wood, S.N., Li, Z., Shaddick, G. & Augustin, N.H. (2017). Generalized additive models for gigadata: modelling the UK black smoke network daily data. *J. Am. Stat. Assoc.*, doi: 10.1080/01621459.2016.1195744, in press.

Zheng, Y. & Aukema, B.H. (2010). Hierarchical dynamic modeling of outbreaks of mountain pine beetle using partial differential equations. *Environmetrics*, 21(7–8), 801–816.

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.