Theoretical Ecology
https://doi.org/10.1007/s12080-021-00518-6

ORIGINAL PAPER

Noise can create or erase long transient dynamics

J. R. Reimer[1] · J. Arroyo-Esquivel[2] · J. Jiang[3] · H. R. Scharf[4] · E. M. Wolkovich[5] · K. Zhu[6] · C. Boettiger[7]

Received: 16 June 2020 / Accepted: 8 June 2021

Abstract
Recent theoretical work has highlighted several mechanisms giving rise to so-called long transient dynamics. These long transients tantalizingly appear to replicate dynamics seen in real systems—with one critical difference: ecological data is noisy, a reality theoretical work often ignores. In general, stochasticity is known to have important consequences: it can qualitatively alter model dynamics as well as impact our ability to infer underlying processes through statistical analysis. To explore the effect of stochasticity on qualitative model behavior and the implications for our ability to infer underlying mechanisms, we generated time series from a simple model of long transient behavior with multiplicative noise. We then examined whether noise qualitatively changes the expected dynamics of the system and the insights that four different statistical methods could provide about the underlying dynamics. We found that the expected duration of the long transient was significantly reduced in the stochastic model compared to the deterministic model. These transient dynamics arise for parameterizations very near to a bifurcation point in the deterministic model, and we also found that as we varied parameterizations to include two alternative stable states, stochasticity caused the population to jump from one basin of attraction to another, resulting in time series that suggest long transient dynamics. Despite challenges estimating the underlying model parameters, we illustrate that statistical inference on a single realization may still provide insight into the presence of a ghost attractor. Further, we highlight that inference improves, across parameterizations, for an increasing number of realizations of the process.

Keywords Inference · Model fitting · Stochasticity · Transient dynamics · Nonparametric models

Introduction

While theoretical ecology has often focused on asymptotic analyses that assume stationarity, there is growing recognition that many realistic ecological systems can experience long periods of transient dynamics, calling this common assumption of stationarity into question (Hastings and Higgins 1994; Hastings 2001; Hastings et al. 2018). Hastings et al. (2018) introduced several important mechanisms that can give rise to long transients, such as *ghost attractors* and *crawl-bys*. These mechanisms may be present in deterministic models that are based on an understanding of relevant ecological processes. In addition to these theoretical results, long transients have also been implicated in empirical studies, either explicitly (Gleeson and Tilman 1990; Van Geest et al. 2007) or implicitly, using the language of regime shifts (Ling et al. 2015).

Here, we seek to better understand the behavior of one such mechanism, the ghost attractor (Fig. 1b), in the presence of stochasticity. Stochasticity is ubiquitous in ecological systems (Bartlett 1960) and has important consequences both for our ability to infer an underlying model or process through statistical analysis and in qualitatively altering the dynamics the model produces relative to the deterministic skeleton. For example, stochasticity can create or disrupt persistence and co-existence, create oscillatory cycles or drive regime shifts (Boettiger 2018).

✉ J. R. Reimer
reimer@math.utah.edu

1 Department of Mathematics, University of Utah, 155 South 1400 East, Salt Lake City, UT 84112, USA

2 Department of Mathematics, University of California, Davis, USA

3 School of Electrical, Computer, and Energy Engineering, Arizona State University, Tempe, USA

4 Department of Mathematics and Statistics, San Diego State University, San Diego, California, USA

5 Forest and Conservation Sciences, Faculty of Forestry, University of British Columbia, Vancouver, Canada

6 Department of Environmental Studies, University of California, Santa Cruz, USA

7 Department of Environmental Science, Policy, and Management, University of California, Berkeley, USA

Published online: 09 July 2021

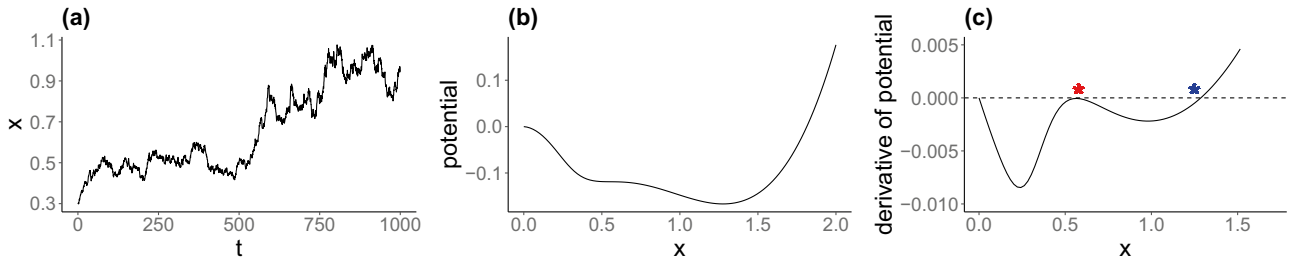

Fig. 1 (**a**) An example of a system that appears to be in a steady state for 500 time steps, but then suddenly jumps up to a different, apparently stable value. (**b**) The potential function for a model showing a ghost attractor for values of x, the state variable, around 0.5. (**c**) The derivative of the potential function shown in (**b**). Points that get near to the line $y = 0$ but do not cross correspond to the ghost attractor (red star), while those that cross the line correspond to the stable state (blue star). The time series in (**a**) was generated from Eqs. (2) and (3), with parameters $r = 0.05$, $K = 2$, $a = 0.023$, $h = 0.38$, $q = 5$, and $\sigma = 0.02$. The functions in (**b**) and (**c**) come from the deterministic herbivore-grazer model, Eq. (1), with the same parameters

These complex outcomes of noise may obscure how well we can use current statistical approaches to identify long transients in real ecological data. To date two major classes of statistical models—phenomenological and mechanistic—could be used to identify and understand long transients in empirical ecological systems. Phenomenological approaches are often used to describe the shift from one state to another. Several such methods have been used to identify transitions in dynamics in noisy data that seem to exhibit a regime shift or a long transient; e.g., changepoint analysis has been used to detect changes in climatic regimes (Beaulieu et al. 2012) and Hidden Markov Models may detect transitions in a disease state (Chen et al. 2016). These methods describe patterns in data without requiring any understanding of the underlying mechanisms. Alternatively, we may build a model that captures the hypothesized key mechanisms in the system of interest and then fit it to the time series to infer parameter values. Provided these key mechanisms are understood, this approach may yield greater insights into the system's behavior as well as greater predictive power. However, even if the relevant mechanisms are fully understood, it is unclear what the implications for inference may be in these systems, as there may be qualitative changes in the realized dynamics caused by the presence of stochasticity.

To explore the effect of stochasticity on realized dynamics and the implications for our ability to infer underlying mechanisms, we generated time series from a simple ghost attractor model with multiplicative noise. The ghost attractor occurs for parameter values very near to a bifurcation point; shifting parameters across that bifurcation point results in a deterministic model with two alternative stable states. For comparison, we also generated time series from this model with alternative stable states. We then addressed three questions: (1) Does the addition of noise qualitatively change the expected dynamics of the system? (2) In this idealized scenario with abundant (simulated) data, which statistical approaches can provide information on the underlying deterministic model? (3) How can this inform the way we approach time series with suspected long transient behavior in real, and often limited, ecological datasets?

We found that the ghost attractor model was highly sensitive to multiplicative noise, with the mean behavior of the stochastic model displaying a significantly reduced transient period. The addition of noise also caused qualitative changes to the model parameterized to have alternative stable states, as the noise was able to push the population from one basin of attraction to the other, resulting in realized behavior that displays a long transient period. In spite of this, we found that statistical inference on a single realization could still provide insight into model parameters and that our ability to distinguish between underlying deterministic models improves for an increasing number of realizations of the process.

A model with a ghost attractor

To explore the effect of stochasticity on the realized dynamics of a model with long transients, as well as our ability to gain insight into the system, we added noise to a model first introduced in May (1977). This model has been used to understand the potential consequences of herbivore overpopulation in several ecosystems including North American forests (Côté et al. 2004), Caribbean coral reefs (Dulvy et al. 2004), and semi-arid regions (Rietkerk and van de Koppel 1997).

Consider a plant population of size x which grows logistically with intrinsic growth rate r and carrying capacity K. Suppose this plant species is consumed by a grazer which is assumed to remain at a constant density with consumption of the plant resource following a Holling Type III functional response. Then the population density of the plant follows the equation

$$\frac{dx}{dt} = rx\left(1 - \frac{x}{K}\right) - \frac{ax^q}{x^q + h^q}, \tag{1}$$

where h is the half-saturation constant (i.e., the plant density at which the realized grazing rate is $a/2$), a is the maximum grazing rate, and q determines the sigmoidal shape of the type III functional response term. For certain parameter regimes, this model has two alternative stable states; a vegetation-dominated state and a herbivore-dominated state (Ludwig et al. 1978). However, if this parameter regime is changed slightly, one of these stable states (the herbivore-dominated state) becomes a ghost attractor (Fig. 1).

Insight and intuition can be gained by considering the potential function (Beisner et al. 2003; Nolting and Abbott 2016) for this model (Fig. 1b). For intuition about the potential function, imagine a ball placed on the curve. If the ball is placed on the curve at $x = 0$, it will roll downhill, lingering where the curve flattens (around $x = 0.5$, the ghost attractor) before coming to rest around $x = 1.3$ (the steady state). Steady states and ghost attractors occur where the derivative of this function is near 0 (Fig. 1c). Note that the derivative of the potential function is just the right hand side of Eq. (1).

In this paper, we set the parameters of this model so that it has a ghost attractor and then added stochasticity around this deterministic core. We explored how this addition of stochasticity shifts the expected trajectory of the stochastic model away from the deterministic trajectory, and the implications of this shift for model inference. We considered the ability of phenomenological models to precisely describe the change in states, as well as the ability of model fitting to the known deterministic model to return the parameters used for data generation. We also explored the insights that may be gained by model fitting to a nonparametric model, which may increase our ability to draw inference from time series suspected to result from long transient dynamics.

Methods

Time series generation

A population, $x(t)$, is assumed to evolve according to

$$\frac{\mathrm{d}\mu(t)}{\mathrm{d}t} = rx(t)\left(1 - \frac{x(t)}{K}\right) - \frac{ax(t)^q}{x(t)^q + h^q} \tag{2}$$

$$\mathrm{d}x(t) = \mathrm{d}\mu(t) + \mathrm{d}W(t). \tag{3}$$

The right side of Eq. (2) is identical to Eq. (1) and captures the variation in $x(t)$ attributable to the deterministic 'core' process. The unobserved random variable $\mathrm{d}\mu(t)$ can be interpreted as the incremental change in the population at time t attributable to the resource-consumer dynamics captured in Eq. (1), and thus $\tilde{x}(t + \mathrm{d}t) = x(t) + \mathrm{d}\mu(t)$ represents the predicted population due to the deterministic core. To account for additional sources of variability, the true change in population, $\mathrm{d}x(t)$, is defined according to Eq. (3), in which a multiplicative stochastic process $\mathrm{d}W(t)$ is specified to model random effects that impact the observed population. The effects of $\mathrm{d}W(t)$ are assumed to be small relative to the core process and might be due, for example, to ephemeral changes in environmental conditions that impact the population size. Note that this noise affects the true population size, as opposed to noise in the data which may be introduced during sampling. We consider measurement error in "Bayesian model fitting".

The probability density of $\mathrm{d}W(t)$ is defined conditionally as

$$p(dW(t) = w|\tilde{x}(t)) = \begin{cases} \mathrm{N}\left(w;0,\sigma^2\tilde{x}^2(t)\mathrm{d}t\right), & \tilde{x}(t) + w > 0 \\ \Phi\left(-(\sigma^2\mathrm{d}t)^{-1}\right), & \tilde{x}(t) + w = 0 \\ 0, & \tilde{x}(t) + w < 0, \end{cases} \tag{4}$$

which ensures that $x(t)$ is always nonnegative. The cases in Eq. (4) represent a spike and slab-type density function (Mitchell and Beauchamp 1988) with support over $[-\tilde{x}(t), \infty)$. The density is equal to a Gaussian distribution with mean-scaled variance for values in $(-\tilde{x}(t), \infty)$, and has a point mass at $-\tilde{x}(t)$ equal to $\Phi\left(-(\sigma^2\mathrm{d}t)^{-1}\right)$. Thus, when $x(t)$ is far from zero, $dW(t)$ resembles Gaussian fluctuations, and when the population is close to extinction, the point mass at $-\tilde{x}(t)$ implies a non-negligible probability of population collapse, after which $x(t) = 0$ for all t.

We simulated time series from Eqs. (2) and (3) for 1000 time steps, with parameters $r = 0.05$, $K = 2$, $a = 0.023$, $h = 0.38$, and $q = 5$, and $\sigma = 0.02$, except where otherwise indicated. Initial conditions were in the vicinity of the ghost attractor, with $x_0 = 0.3$. To explore how the mean trajectory changed with the variance, we calculated the mean of 5000 realizations for each of 4 levels of variance: $\sigma = 0.005, 0.01, 0.015, 0.02$. When assessing possible inference, we considered both a single time series as well as ensembles of realizations.

Hidden markov model and changepoint analyses

We applied two standard time series approaches to our generated time series: Hidden Markov Models and Changepoint analysis. A Hidden Markov Model assumes observed dynamics result from an underlying Markov process whose transition matrix is determined through model fitting. This model classifies each point in the time series into one of two states (note that the number of states is determined *a priori* by the user). We interpreted the end of the long

transient as the first time at which the Hidden Markov Model assigned a change in state. Changepoint analysis determines the point in the time series which most parsimoniously separates the data into two statistically different sets. We applied both of these approaches to a single time series as well as an ensemble of 100 simulations. All analyses were conducted in R (R Core Team 2019) using packages ecp (James and Matteson 2014) and depmixS4 (Visser and Speekenbrink 2010).

Bayesian model fitting

We explored a Bayesian approach to inference by specifying moderately to weakly informative priors for all parameters in the model and obtaining realizations from the joint posterior distribution given the simulated data.

To explore the impact of measurement error in parameter estimation, we introduced an additional layer to the hierarchical model structure in Eqs. (2) and (3). We modeled $y(t)$ as observations of the true population $x(t)$ contaminated with the same spike and slab-type density function in Eq. (4) to ensure observed populations are nonnegative such that,

$$p(y(t)=y|x(t)) = \begin{cases} \mathrm{N}\left(y; x(t), \sigma^2_{\mathrm{me}}\right), & x(t)+y>0 \\ \Phi\left(\frac{-x(t)}{\sigma_{\mathrm{me}}}\right), & x(t)+y=0 \\ 0, & x(t)+y<0, \end{cases} . \quad (5)$$

We considered values of σ^2_{me} ranging from 0.005 to 0.08.

Fitting to the generating mechanistic model

We performed model fitting and parameter estimation in a Bayesian framework. Priors specified for all parameters in Eqs. (2)–(5) are given in Table 1. The model was fit to the simulated data using a Markov chain Monte Carlo (MCMC) procedure with the R package nimble (de Valpine et al. 2017).

One notable challenge to fitting the generating model via MCMC was the existence of strongly correlated parameters, which can impede the efficiency of traditional univariate samplers—the default sampler for most Bayesian model fitting software. To better explore the joint posterior distribution, we utilized adaptable block Metropolis-Hastings samplers and ran the MCMC algorithm for 100,000 iterations, discarding the first half of the iterations as burn-in and thinning to every 10th iteration for a total of 5,000 posterior draws. We assessed convergence visually using traceplots and checking effective sample sizes. We refer readers to Gelman et al. (2020) for a general guideline to implementing Bayesian hierarchical models.

Table 1 Priors specified for all parameters in the generating mechanistic model, Eqs. (2)–(5), and nonparametric model Eq. (8), during model fitting

generating model priors		nonparametric model priors	
r	Gamma(2, 10)		
K	Gamma(1, 0.1)		
a	Gamma(2, 10)	$\beta_i, i = 1, \ldots, m = 5$	N(0, 10)
h	Gamma(2, 1)		
Q	Gamma(1, 0.1)		
σ	Gamma(1, 10)	σ	Gamma(1, 10)
σ_{me}	Gamma(1, 10)	σ_{me}	Gamma(1, 10)

Kullback–Leibler divergence

To quantify Bayesian learning, we used the Kullback–Leibler (KL) divergence, which measures the information gain from prior to posterior distributions. For a generic parameter θ (any parameter in Table 1), the Bayesian model fitting can be summarized as

$$\underbrace{[\theta|\mathbf{x}]}_{\text{posterior}} \propto \underbrace{[\mathbf{x}|\theta]}_{\text{likelihood}} \underbrace{[\theta]}_{\text{prior}}, \quad (6)$$

where $\mathbf{x} = x(t)$ are the time-series data. The KL divergence is a measure of the discrepancy between two probability densities based on information entropy (Kullback and Leibler 1951). The KL divergence between the prior ($[\theta]$) and the posterior ($[\theta|\mathbf{x}]$) distributions quantifies how much knowledge about θ has changed in light of information contained in the data (Itti and Baldi 2006). It is defined mathematically as

$$D_{\mathrm{KL}}([\theta|\mathbf{x}]\|[\theta]) = |\mathrm{E}_{\theta|\mathbf{x}} \log\left(\frac{[\theta|\mathbf{x}]}{[\theta]}\right) = \sum_{\theta}[\theta|\mathbf{x}] \log\left(\frac{[\theta|\mathbf{x}]}{[\theta]}\right). \quad (7)$$

A KL divergence of 0 indicates that the prior and posterior distributions are identical. A large KL divergence indicates substantial information gain from prior to posterior. We computed KL divergence using a nearest neighbor search algorithm to calculate the distance from simulated samples of the prior distribution and the MCMC samples of the posterior distribution using the R package FNN (Beygelzimer et al. 2019).

Importance of ghost attractor

The ghost attractor in this model is found for a parameter set very near to a bifurcation point, where the deterministic model changes from having one nonzero stable state to having two alternative stable states. We have parameterized the model with a value of $a = 0.023$, which is very near the bifurcation point, $a = 0.02306$. To explore how our model fitting results depended

specifically on the ghost attractor, we also fit models for two additional scenarios on either side of the bifurcation by varying parameter a—one with a weaker ghost attractor ($a = 0.0225$) and one with two nonzero stable states ($a = 0.0235$).

Nonparametric model fitting

As mentioned in "Fitting to the generating mechanistic model", a challenge to implementation for Bayesian model fitting is strong conditional dependence among the parameters in the model, especially those appearing in Eq. (2). Strong conditional dependence can indicate the potential for difficult to explore likelihood surfaces and highly inefficient MCMC algorithms. We investigated an alternative, nonparametric specification of the model in which the form of $\frac{d\mu(t)}{dt}$ is expressed as a generic polynomial function of $x(t)$ in a linear m-dimensional functional space, such that

$$\frac{d\mu(t)}{dt} = \sum_{i=1}^{m} \beta_i \phi_i(x(t)). \tag{8}$$

By relaxing the assumed parametric form, it is possible to specify a model in which parameters do not suffer from the same high degree of dependence, leading to more efficient implementation. The trade-off for the improvements in model fitting are that direct interpretation of model parameters in Eq. (2) via their posterior distribution is no longer possible. However, important features of $\frac{d\mu(t)}{dt}$ such as curvature, and regions of $x(t)$ where the function is zero or nearly zero are still available through the nonparametric approach.

Several possible bases, $\phi_i(x(t))$, could be used to approximate the non-linear space of functions defined by Eq. (2). We chose an approximately orthogonal basis that spans the space of fourth degree polynomials.

Comparing parametric and nonparametric approaches

To compare the performance of the nonparametric and generating models, we investigated two metrics intended to reveal the efficiency and accuracy of each approach. To assess efficiency, we calculated the median effective sample size (ESS) for the derived quantity $\frac{d\mu(t)}{dt}$ over a grid of values for $x(t) = 0.2$ and 1.8, and then took the median across the grid. ESS is an estimate of the number of independent samples drawn from a target distribution. It is used to measure the efficiency of samplers whose realizations are not independent, such as MCMC algorithms that sample from posterior distributions in Bayesian analyses (Gelman et al. 2013). When comparing two algorithms with the same target distribution, in general, the more computationally efficient algorithm will provide a larger number of independent samples per total number of iterations (or per unit time, depending on the most relevant definition of efficiency).

To assess accuracy, we computed the mean squared error between the function $\frac{d\mu(t)}{dt}$ of $x(t)$ used to generate the data, and the associated marginal posterior distribution. We then integrated these mean errors over the range $x(t) = 0.2$ and 1.8 numerically to yield the integrated mean squared error (IMSE). Intuitively, the model with smaller IMSE yields a posterior distribution of $\frac{d\mu(t)}{dt}$ that more closely resembles the true data-generating function.

We fit both models to subsets of the ten simulated population trajectories such that all 10 single trajectories were fit, 10 randomly selected 2- and 5-trajectory subsets were fit, and the unique complete subset was fit. In the results section, ESS and IMSE are aggregated by number of trajectories and measurement error variance. The nonparametric approach was expected to exhibit improved ESS, perhaps at the expense of IMSE, because it used a linear approximation to a non-linear function. For further details on both ESS and IMSE see Appendix.

Results

The effect of variance on model dynamics

The addition of stochasticity to the model with the ghost attractor resulted in mean behavior that differed from that of the deterministic core model. Even relatively small variance resulted in a substantial reduction in the expected time spent in the area of the ghost attractor (Fig. 2a). In spite of the fact

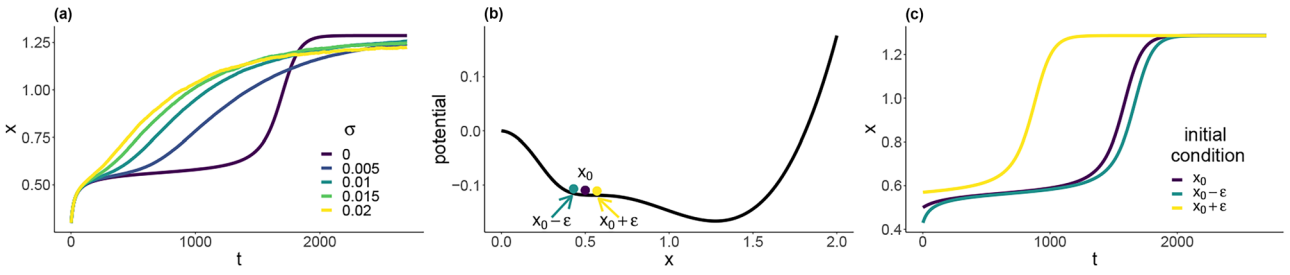

Fig. 2 (**a**) The effect of variance on the mean trajectory. We simulated 5000 realizations from Eqs. (2)-(3) and took the mean for each of 4 different variance levels ($\sigma = 0.005, 0.01, 0.015, 0.02$). (**b**) Three initial conditions centered at the ghost attractor; x_0, and $x_0 \pm \epsilon$ (here $x_0 = 0.5, \epsilon = 0.07$). (**c**) Deterministic trajectories resulting from the initial conditions in (**b**)

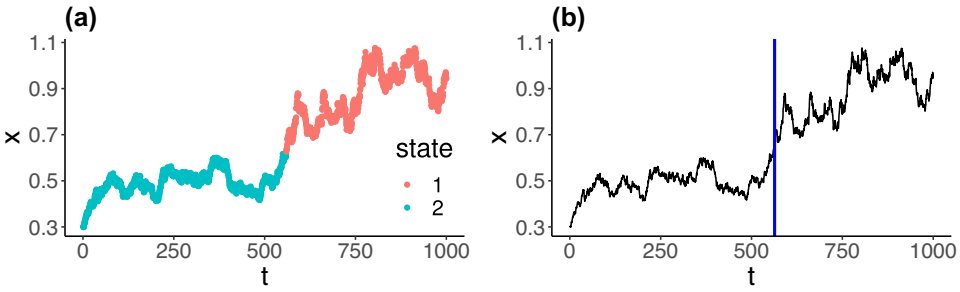

Fig. 3 Illustrative model results for a single time series. (**a**) Hidden Markov Model results, with two states represented by the two colors. (**b**) Changepoint analysis. The vertical blue line is the changepoint, signifying the most parsimonious break in the time series

that the variance term is symmetric around the deterministic model core, the asymmetry in the potential function appears to cause stochastic trajectories to leave the ghost attractor sooner, on average, than the deterministic model. A small positive perturbation to x in the vicinity of the ghost attractor results in a subsequent deterministic trajectory which leaves the ghost much more rapidly, while a small negative perturbation to x results in a deterministic trajectory very similar to that without the perturbation (Fig. 2b, c).

Results from Hidden Markov Model and changepoint approaches

The Hidden Markov Model and changepoint analysis identified similar transition times from the ghost attractor to the stable state (Fig. 3). Individual realizations varied in the timing of their transition away from ghost attractor (grey lines in Fig. 4), resulting in considerable variability in the transition timing for a set of 100 model realizations; the standard deviation of the timing of the shift for the Hidden Markov Models and changepoint analysis were 459 and 487 time steps, respectively. When averaged over the ensemble of time series, the average shift occurred earlier than in the deterministic core model (Fig. 4), reflecting the behavior seen in Fig. 2a.

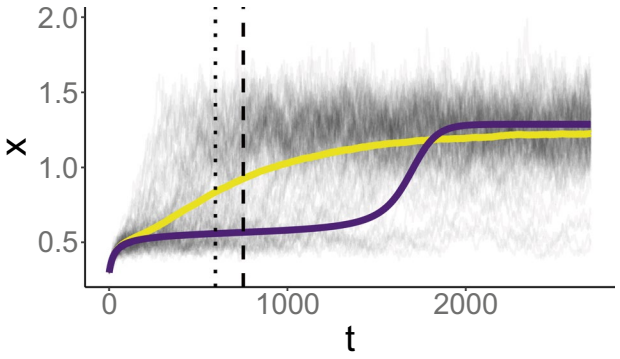

Fig. 4 Results from the Hidden Markov Model (dotted vertical line) and changepoint analysis (dashed vertical line) applied to an ensemble of 100 time series (each realization is plotted in grey in the background). The purple curve is the deterministic model core, and the yellow curve is the mean of all realizations, as in Fig. 2a. Parameters for the realizations were $r = 0.05$, $K = 2$, $a = 0.023$, $h = 0.38$, $q = 5$, and $\sigma = 0.02$

Parametric model fitting

Parameters from the generating mechanistic model appeared to be at least weakly identifiable from realized population trajectories, and marginal posteriors showed evidence of concentrating around the parameter values used to simulate the data as the number of realizations increased (Fig. 5). The derivative of the potential function was well-estimated in the vicinity of the ghost attractor ($x \approx 0.5$), which suggests the simulated trajectories contain enough information about the presence of these transient-inducing characteristics to allow us to identify them from data (Fig. S1). Additional model fitting diagnostics confirm that while the parameters are only weakly identifiable, the derivative of the potential function ($\frac{d\mu(t)}{dt}$) is more strongly identifiable for some regions of $x(t)$ (Figs. S2 and S3), providing insights into model dynamics in spite of weak insights into parameter values.

KL divergence

In most—but not all—cases, the KL divergence shows increasing trends from 1 to 10 realized trajectories (top right corner of each sub-plot in Fig. 5). This general pattern of increased KL divergence for increasing realizations is visually consistent with the densities shown in Fig. 5, where more information is gained by using a larger number of realized population trajectories. Observe, however, that for parameters r, q, a and h, more realizations did not always improve the KL divergence. It should also be noted that the posteriors increasingly concentrate around the values used to simulate the data. Finally, changes in information about certain parameters as measured by KL divergence are non-linear in the number of realized trajectories. For example, much more is learned about h and K in moving from 1 to 2 realizations than 2 to 5, while more is learned about q when we increase the number of realizations from 5 to 10.

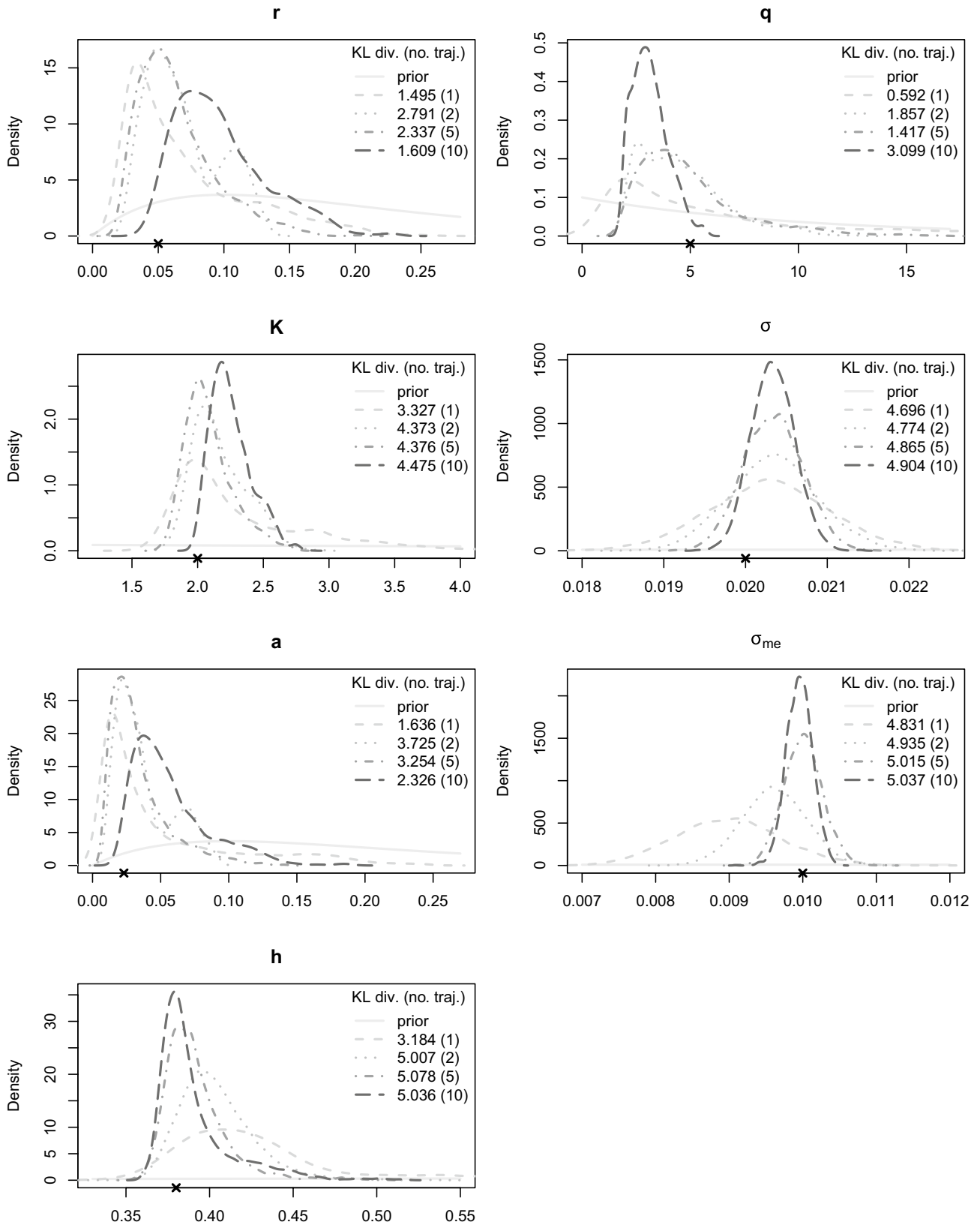

Fig. 5 Comparison of marginal posterior distributions for simulations with 1, 2, 5, and 10 realized trajectories. Each "×" on the x-axis denotes the parameter value used to generate the data. KL divergences from the prior to the posterior are in the top right of each subplot. Parameters for the realizations were $r = 0.05$, $K = 2$, $a = 0.023$, $h = 0.38$, $q = 5$, $\sigma = 0.02$, and $\sigma_{me} = 0.01$. Note that $a = 0.02306$ is the bifurcation value of a; for larger a values, the deterministic model has two alternative stable states

Sensitivity of results to ghost dynamics

When we compared scenarios with varied values of a corresponding to a very weak ghost attractor ($a = 0.0225$), a stronger ghost attractor ($a = 0.023$), and two alternative stable states ($a = 0.0235$), we found negligible differences in our inference (Fig. 6). In all three cases, the distribution of the derivative of the potential functions concentrated around the true values as the number of realizations used in model fitting increased from 1 to 10.

Nonparametric model fitting

Results from the nonparametric model appeared very similar to those of the parametric model, with the derivative of the potential function similarly well-estimated in the vicinity of the ghost attractor (Fig. S4). Model diagnostics for the nonparametric model were an improvement from those of the parametric model (Figs. S5 and S6).

Comparing parametric and nonparametric results

The ESS is much higher for the nonparametric model than the generating model (Fig. 7a). This is especially true for lower values of measurement error; as measurement error increases, the difference between the ESS of the two models decreases. For context, there were 5,000 posterior draws, total, so for low measurement error, the sample size of the nonparametric model is nearly 5000, while that of the generating model is near 0. A look at the IMSE shows no clear indication which model fit the data better (Fig. 7b). Additionally, consideration of the trace plots between the parametric model (Figs. S1 and S3) and the nonparametric model (Figs. S5 and S6) suggests identifiability issues with the parametric model which are not present for the nonparametric model.

Discussion

As ecological theory on long transients has developed, so too has the desire to identify them in real systems. Identifying long transients in natural systems would anchor the growing theoretical literature (Hastings 2001; Hastings et al. 2018), and could help narrow in on common attributes of such systems that may represent areas for more research or highlight gaps in current theory. Further, it would have important implications for those systems—as the presence of long transients means managers would have to consider complex possible trajectories. Depending on the goal, managers may want to maintain a system in a long transient state, or they may desire to push a system out of its transient state and into the stable state.

Identifying long transients in ecological systems requires connecting theoretical advances, which have generally been derived

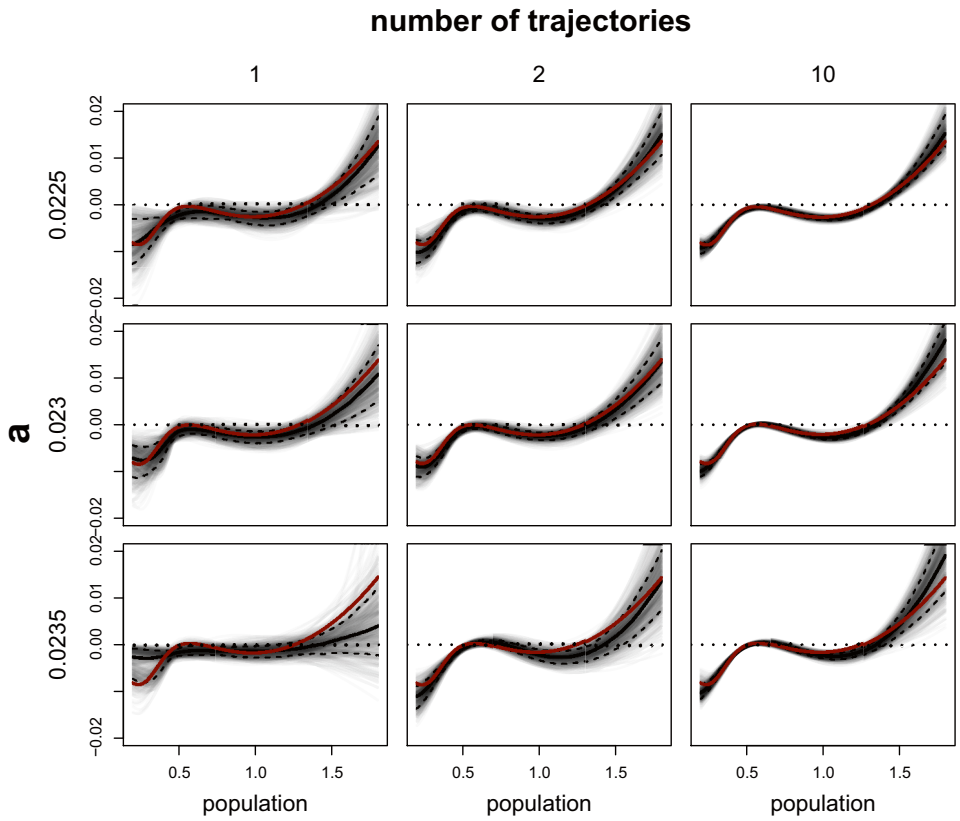

Fig. 6 Comparison of parametric model fitting results as a is varied through a scenario with a very weak ghost attractor (top row), a stronger ghost attractor (middle row), and two alternative stable states (bottom row). Curves were drawn from the posterior distribution of $\frac{d\mu(t)}{dt}$ in grey, pointwise median value in black, pointwise equal-tailed 95% credible intervals as dashed, and true potential curve in red. Inference based on 1, 2, or 10 realized population trajectories, from left to right. Parameters for the realizations were $r = 0.05$, $K = 2$, $h = 0.38$, $q = 5$, $\sigma = 0.02$, and $\sigma_{me} = 0.01$

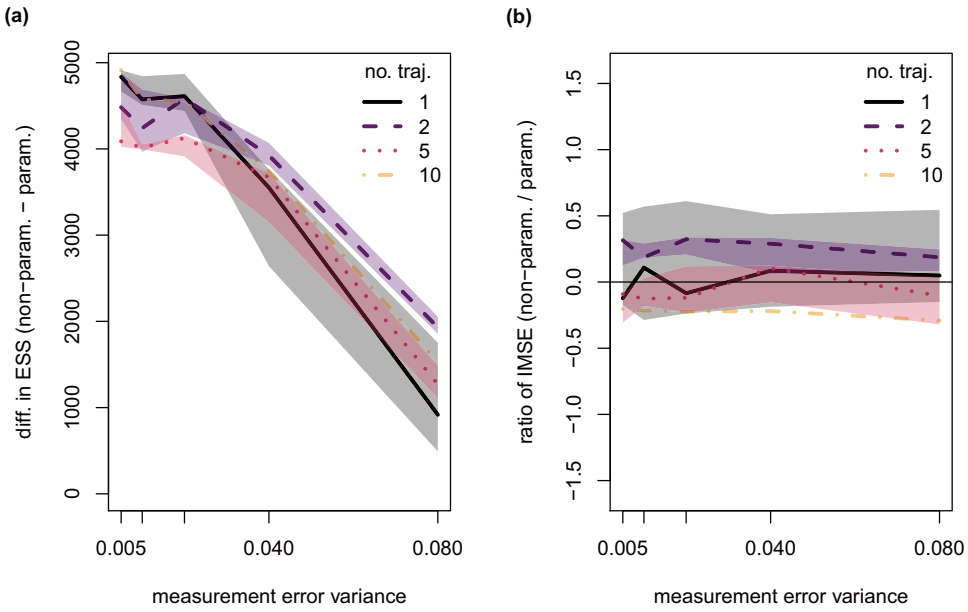

Fig. 7 (**a**) The difference in effective sample size (ESS) between the nonparametric model and the parametric (generating) model decreases for larger measurement error variance, σ^2_{me}. (**b**) No clear trend emerges when comparing the integrated mean squared error (IMSE) for the two model types, suggesting the two models fit similarly well. Legend entries refer to the number of realizations used for model fitting. Shaded polygons represent pointwise interquartile ranges across all subsets of a given size. Parameters used in time series generation were $r = 0.05$, $K = 2$, $a = 0.023$, $h = 0.38$, $q = 5$, and $\sigma = 0.02$

from deterministic models, to the stochasticity of the real world. Stochasticity enters at myriad levels, from the relative simplicity of measurement error, to environmental and demographic stochasticity. Whichever its sources, stochasticity—depending on its type, magnitude and also how accurately it is accounted for in modeling approaches—can obscure the deterministic skeleton that we wish to uncover. Thus, advancing ecological theory on long-transients requires bridging stochastic versions of common models producing long transients to statistical inference methods that can robustly match data to pattern and process.

In a purely deterministic framework, the dynamics resulting from a model with a long transient can differ substantially from a model with alternative stable states. However, the introduction of noise into this framework blurs this distinction between deterministic behaviors. In the ghost attractor model, noise shortens the expected time spent in the long transient state. In the model with alternative stable states, noise can result in dynamics with features that look like long transients, as the population is bumped from one basin of attraction to another. This work is a first step towards understanding whether we can infer the underlying deterministic behavior from a noisy time series by estimating the relevant parameter values.

To explore the performance of statistical inference methods, we purposefully simplified inference here by using a known underlying model with given parameters to generate time series of the system's dynamics. This approach allowed us to explore the performance of various modeling approaches and to interrogate deterministic versus stochastic versions of our model to understand the limitations of these approaches.

Both the Hidden Markov Model and changepoint analysis were able to identify the key pattern of the shifting states of the ghost attractor model. As the timing of the shift varied considerably between randomly generated time series, the timing of the shift implicated by both models also varied considerably between individual time series. When we considered an ensemble of time series, however, the mean modelled shift converged to a value consistent with the mean of the simulated trajectories (Fig. 4). While these methods can categorize data into multiple states (with the number of states being set *a priori* for both models), they provide little insight into the underlying dynamics of the system.

Our Bayesian inference approach to estimating the underlying mechanistic model suggested that parameters were weakly identifiable. Posterior estimates from MCMC increased in accuracy and decreased in uncertainty as the number of model realizations increased. KL divergences further confirm Bayesian learning from prior to posterior distributions driven by data. This estimation performance is somewhat surprising given that stochasticity shifts the expected behavior away from the deterministic core. These findings (using a parameterization with a strong ghost attractor) were consistent across other formulations of the model—from a weak ghost attractor to a system with two alternative stables states. This suggests that that the ghost attractor does not present a unique estimation challenge compared to other models, but instead highlights the challenge of connecting models from the theoretical ecological literature to real-world noisy conditions.

As is often the case for models with high dependency across parameters, certain parameter estimates improved much more given additional realization than others, e.g., K, σ, and σ_{me}. However, even if precise estimates for some parameters are not possible, the presence and location of a ghost attractor may still be identified (e.g., Fig. 6) by

considering regions where $\frac{d\mu(t)}{dt}$ is very near to 0. The shape of $\frac{d\mu(t)}{dt}$ resolves better for fewer realizations than the values of each specific parameter, as can be seen by comparing the results for 10 realizations in Figs. 5 and 6.

The findings of the nonparametric model fitting were very similar to that of the parametric model fitting, as seen through their similar IMSE values (Fig. 7b). While the ESS of the nonparametric model was considerably higher than that of the parametric model, especially for low measurement error (Fig. 7a), scientific knowledge about the parameter values cannot be gained through consideration of the nonparametric model alone. Fitting to the parametric model, which incorporates the main mechanisms thought to determine the population size, provides important information into biologically relevant quantities and dynamics. Thus consideration of the parametric and nonparametric models together may yield the greatest insights; the more efficient nonparametric model can provide confirmation of the shape of the potential function found by the parametric function, and thus for the parameter values that accompany the parametric inference.

Additional modeling methods have been proposed elsewhere to study seemingly similar systems. For example, in the literature on so-called tipping points, changes in the standard deviation and autocorrelation coefficient of a time series may provide early warning signs of tipping points from one regime to another (Scheffer et al. 2009; Dakos et al. 2008; Wissel 1984). Consideration of a time series' spectral density has also been proposed (Biggs et al. 2009). This provides a method to identify earlier warning signs of tipping points than those identified by changes in the standard deviation and autocorrelation. However, the system studied in this paper is different from those typically discussed in the literature on tipping points, which are usually attributed to a slowly shifting parameter resulting in a sudden bifurcation of stable states. Allowing a model parameter to vary explicitly with time makes the system non-autonomous. Since the model considered here is autonomous, with all parameters fixed, there is no reason to expect that the tools developed to analyze systems experiencing tipping points should provide insights here. Indeed, while performing exploratory analyses of changes in standard deviation, autocorrelation, and spectral density of our generated time series, we found that all three methods failed to consistently identify the shift from the ghost attractor to the steady state across the realizations.

These findings suggest several areas for future theoretical work. By fitting generated data to a known underlying model, we have avoided the difficult step in inference of comparing different underlying models. Future work should focus on the open question of question of whether it is possible to differentiate between non-stationary behavior that arises from a long transient and non-stationary behavior that arises from a non-autonomous model with a time-dependent parameter that crosses a bifurcation threshold.

Our finding that the mean behavior over many realizations of the ghost attractor model with added process noise differs from the deterministic mean also raises the question of whether this is generalizable to other systems with transient dynamics. Further work may provide insight into explaining and predicting the expected deviance between the ensemble mean and the deterministic skeleton.

Appendix

Effective sample size and integrated mean squared error

We computed median ESS using the effectiveSize() function from the coda package for the R statistical programming environment (Plummer 2006). We first computed the ESS of $\frac{d\mu(t)}{dt}$ over a grid of 100 equally-spaced values between $x(t) = 0.2$ and 1.8. We then took the median value across the grid.

Let $\frac{d\hat{\mu}_P(t)}{dt}$ represent the functional $\frac{d\mu(t)}{dt}$ in Eq. (2) evaluated at the values $\hat{\theta} = (\hat{r}, \hat{K}, \hat{a}, \hat{h}, \text{and } \hat{Q})$. Analogously, let $\frac{d\hat{\mu}_{NP}(t)}{dt}$ represent the functional in the nonparametric model given by Eq. (8) evaluated at $\hat{\beta}$. We defined the IMSE for either the parametric (P) or nonparametric (NP) approach as

$$\text{IMSE}_M = \int_{0.2}^{1.8} E_{\hat{\theta}}\left(\frac{d\hat{\mu}_M(t)}{dt} - \frac{d\mu(t)}{dt}\right)^2 dx(t), \quad M \in \{P, NP\}. \tag{9}$$

The expectation was approximated using a Monte Carlo approximation via samples from the joint posterior distribution of $\hat{\theta}$, and the integral was approximated numerically using a simple Riemannian quadrature.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s12080-021-00518-6.

Acknowledgements The authors would like to thank the organizers of the Transients in Biological Systems workshop in May 2019 in NIMBioS at the University of Tennessee, Knoxville for the space to develop this article.

Funding KZ acknowledges support from the US National Science Foundation, DEB1926438 and the University of California, Santa Cruz, Committee on Research, Faculty Research Grant.

Data Availability No unpublished data were used in this study.

Code availability Code and simulated data used to generate all figures and analysis is freely available at https://doi.org/10.5281/zenodo.3897393.

Declarations

Ethics approval Ethics approval was not required for this study.

Consent to participate and publication Consent to participate and consent for publication were not required for this study.

Conflict of interest The authors declare that they have no conflict of interest.

References

Bartlett MS (1960) Stochastic population models in ecology and epidemiology. Methuen and Wiley, London

Beaulieu C, Chen J, Sarmiento JL (2012) Change-point analysis as a tool to detect abrupt climate variations. Philos Trans R Soc A Math Phys Eng Sci 370(1962):1228–1249

Beisner BE, Haydon DT, Cuddington K (2003) Alternative stable states in ecology. Front Ecol Environ 1(7):376–382

Beygelzimer A, Kakadet S, Langford J, Arya S, Mount D, Li S (2019) FNN: Fast Nearest Neighbor Search Algorithms and Applications. https://CRAN.R-project.org/package=FNN, r package version 1.1.3

Biggs R, Carpenter SR, Brock WA (2009) Turning back from the brink: detecting an impending regime shift in time to avert it. Proc Natl Acad Sci 106(3):826–831

Boettiger C (2018) From noise to knowledge: how randomness generates novel phenomena and reveals information. Ecol Lett. http://doi.wiley.com/10.1111/ele.13085

Chen P, Liu R, Li Y, Chen L (2016) Detecting critical state before phase transition of complex biological systems by hidden markov model. Bioinformatics 32(14):2143–2150

Côté SD, Rooney TP, Tremblay JP, Dussault C, Waller DM (2004) Ecological impacts of deer overabundance. Annu Rev Ecol Evol Syst 35:113–147

Dakos V, Scheffer M, van Nes EH, Brovkin V, Petoukhov V, Held H (2008) Slowing down as an early warning signal for abrupt climate change. Proc Natl Acad Sci 105(38):14308–14312

Dulvy NK, Freckleton RP, Polunin NV (2004) Coral reef cascades and the indirect effects of predator removal by exploitation. Ecol Lett 7(5):410–416

Gelman A, Carlin JB, Stern HS, Dunson DB, Vehtari A, Rubin DB (2013) Bayesian data analysis. CRC Press

Gelman A, Vehtari A, Simpson D, Margossian CC, Carpenter B, Yao Y, Kennedy L, Gabry J, Bürkner PC, Modrák M (2020) Bayesian workflow. arXiv preprint arXiv:201101808

Gleeson SK, Tilman D (1990) Allocation and the transient dynamics of succession on poor soils. Ecology 71(3):1144–1155

Hastings A (2001) Transient dynamics and persistence of ecological systems. Ecol Lett 4(3):215–220. https://doi.org/10.1046/j.1461-0248.2001.00220.x

Hastings A, Higgins K (1994) Persistence of transients in spatially structured ecological models. Science 263(5150):1133–1136. https://doi.org/10.1126/science.263.5150.1133, http://www.ncbi.nlm.nih.gov/pubmed/17831627

Hastings A, Abbott KC, Cuddington K, Francis T, Gellner G, Lai YC, Morozov A, Petrovskii S, Scranton K, Zeeman ML (2018) Transient phenomena in ecology. Science 361(6406). https://doi.org/10.1126/science.aat6412

Itti L, Baldi PF (2006) Bayesian surprise attracts human attention. In: Advances in neural information processing systems, pp 547–554

James NA, Matteson DS (2014) ecp: An R package for nonparametric multiple change point analysis of multivariate data. J Stat Softw 62(7):1–25. http://www.jstatsoft.org/v62/i07/

Kullback S, Leibler RA (1951) On information and sufficiency. Ann Math Stat 22(1):79–86

Ling S, Scheibling R, Rassweiler A, Johnson C, Shears N, Connell S, Salomon A, Norderhaug K, Pérez-Matus A, Hernández J et al (2015) Global regime shift dynamics of catastrophic sea urchin overgrazing. Philos Trans R Soc B 370(1659):20130269

Ludwig D, Jones DD, Holling CS et al (1978) Qualitative analysis of insect outbreak systems: the spruce budworm and forest. J Anim Ecol 47(1):315–332

May RM (1977) Thresholds and breakpoints in ecosystems with a multiplicity of stable states. Nature 269(5628):471

Mitchell TJ, Beauchamp JJ (1988) Bayesian variable selection in linear regression. J Am Stat Assoc 83(404):1023–1032

Nolting BC, Abbott KC (2016) Balls, cups, and quasi-potentials: quantifying stability in stochastic systems. Ecology

Plummer M, Best N, Cowles K, Vines K (2006) CODA: convergence diagnosis and output analysis for MCMC. *R news*, *6*(1), 7-11

R Core Team (2019) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

Rietkerk M, van de Koppel J (1997) Alternate stable states and threshold effects in semi-arid grazing systems. Oikos 69–76

Scheffer M, Bascompte J, Brock WA, Brovkin V, Carpenter SR, Dakos V, Held H, Van Nes EH, Rietkerk M, Sugihara G (2009) Early-warning signals for critical transitions. Nature 461(7260):53

de Valpine P, Turek D, Paciorek CJ, Lang DT, Bodik R (2017) Programming with models: Writing statistical algorithms for general model structures with NIMBLE. J Comput Graph Stat 26:403–417 arXiv:1505.05093v1

Van Geest G, Coops H, Scheffer M, Van Nes E (2007) Long transients near the ghost of a stable state in eutrophic shallow lakes with fluctuating water levels. Ecosystems 10(1):37–47

Visser I, Speekenbrink M (2010) depmixS4: An R package for hidden markov models. J Stat Softw 36(7):1–21. http://www.jstatsoft.org/v36/i07/

Wissel C (1984) A universal law of the characteristic return time near thresholds. Oecologia 65(1):101–107