
Detection of long transient dynamics in noisy time series

Reimer, J. R. · Arroyo-Esquivel, J. · Jiang, J. ·

Scharf, H. R. · Wolkovich, E. M. · Zhu K. ·

Boettiger, C.

Received: date / Accepted: date

Reimer, J. R.
Department of Mathematics, University of Utah
155 South 1400 East, Salt Lake City, UT, 84112, USA
Tel.: +801-585-1641
E-mail: reimer@math.utah.edu

Arroyo-Esquivel, J.
Department of Mathematics, University of California, Davis, USA
ORCID: 0000-0003-4399-954X

Jiang, J.
School of Electrical, Computer, and Energy Engineering, Arizona State University, USA
ORCID: 0000-0003-2930-7770

Scharf, H. R.
Department of Mathematics and Statistics, San Diego State University, USA
ORCID: 0000-0002-9266-4191

Wolkovich, E. M.
Forest and Conservation Sciences, Faculty of Forestry, University of British Columbia, Canada
ORCID: 0000-0001-7653-893X

Zhu K.
Department of Environmental Studies, University of California, Santa Cruz, USA
ORCID: 0000-0003-1587-3317

Boettiger, C.
Department of Environmental Science, Policy, and Management, University of California, Berkeley, USA
ORCID: 0000-0002-1642-628X

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 if noise qualitatively changes the expected dynamics of the system and the insights that four different methods could provide about the underlying dynamics. We found that the mean behavior of the stochastic model differs substantially from that of the deterministic model. This ghost attractor occurs for parameterizations very near to a bifurcation point in the deterministic model, and we also found that our results were consistent as we varied parameterizations to include a weaker ghost attractor or two alternative stable states—suggesting it is the presence of stochasticity, and not the ghost attractor, that causes challenges for estimating the underlying model. Despite this, 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.

key words: inference, model fitting, stochasticity, transient dynamics, non-parametric models

1 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 mechanisms. 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 (Figure 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).

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. This approach requires an understanding of a system’s underlying drivers. Provided these drivers 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. 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 mechanisms? (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 our chosen model of a ghost attractor was highly sensitive to multiplicative noise, with the mean behavior of the stochastic model differing substantially from that of the underlying deterministic model. In spite of this, we found that statistical inference on a single realization could still provide insight into model parameters and that inference improves for an increasing number of realizations of the process.

1.1 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).

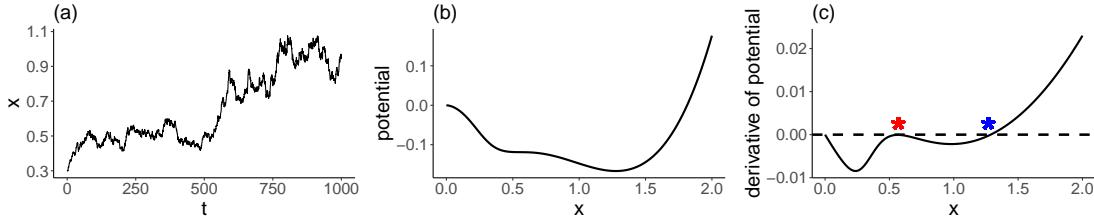


Fig. 1: (a) An example of a system that appears to be in a steady state for 500 time steps, but then suddenly jumps up to a different, apparently stable value. (b) The potential function for a model showing a ghost attractor for values of x , the state variable, around 0.5. (c) The derivative of the 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 Eq. (2) and (3), with parameters $r = 0.05$, $K = 2$, $a = 0.023$, $h = 0.38$, and $q = 5$, and $\sigma = 0.02$. The functions in (b) and (c) come from the deterministic herbivore-grazer model, Eq. (1), with the same parameters.

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

$$72 \quad \frac{dx}{dt} = rx \left(1 - \frac{x}{K}\right) - \frac{ax^q}{x^q + h^q}, \quad (1)$$

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

79 Insight and intuition can be gained by considering the potential function (Beisner et al., 2003;
 80 Nolting and Abbott, 2016) for this model (Figure 1b). For intuition about the potential function,
 81 imagine a ball placed on the curve. If the ball is placed on the curve at $x = 0$, it will roll downhill,
 82 lingering where the curve flattens (around $x = 0.5$, the ghost attractor) before coming to rest around
 83 $x = 1.3$ (the steady state). Steady states and ghost attractors occur where the derivative of this
 84 function is near 0 (Figure 1c). Note that the derivative of the potential function is just the right hand
 85 side of Eq. (1).

86 In this paper, we set the parameters of this model so that it has a ghost attractor and then
 87 added stochasticity around this deterministic core. We explored how this addition of stochasticity
 88 shifts the expected trajectory of the stochastic model away from the deterministic trajectory, and
 89 the implications of this shift for model inference. We considered the ability of phenomenological
 90 models to precisely describe the change in states, as well as the ability of model fitting to the known

⁹¹ deterministic model to return the parameters used for data generation. We also explored the insights
⁹² that may be gained by model fitting to a non-parametric model, which may increase increase our
⁹³ ability to draw inference from time series suspected to result from long transient dynamics.

⁹⁴ **2 Methods**

⁹⁵ **2.1 Time series generation**

⁹⁶ A population, $x(t)$, is assumed to evolve according to

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

$$dx(t) = d\mu(t) + dW(t). \quad (3)$$

⁹⁷ The right side of equation (2) is identical to (1) and captures the variation in $x(t)$ attributable to the
⁹⁸ deterministic ‘core’ process. The unobserved random variable $d\mu(t)$ can be interpreted as the incre-
⁹⁹ mental change in the population at time t attributable to the resource-consumer dynamics captured
¹⁰⁰ in (1), and thus $\tilde{x}(t + dt) = x(t) + d\mu(t)$ represents the predicted population due to the deterministic
¹⁰¹ core. To account for additional sources of variability, the true change in population, $dx(t)$, is defined
¹⁰² according to (3), in which a multiplicative stochastic process $dW(t)$ is specified to model random
¹⁰³ effects that impact the observed population. The effects of $dW(t)$ are assumed to be small relative to
¹⁰⁴ the core process and might be due, for example, to ephemeral changes in environmental conditions
¹⁰⁵ that impact the population size. Note that this noise affects the true population size, as opposed
¹⁰⁶ to noise in the data which may be introduced during sampling. We consider measurement error in
¹⁰⁷ Section 2.3.

¹⁰⁸ The probability density of $dW(t)$ is defined conditionally as

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

¹⁰⁹ which ensures that $x(t)$ is always non-negative. The cases in (4) represent a spike and slab-
¹¹⁰ type density function (Mitchell and Beauchamp, 1988) with support over $[-\tilde{x}(t), \infty)$. The density is
¹¹¹ equal to a Gaussian distribution with mean-scaled variance for values in $(-\tilde{x}(t), \infty)$, and has a point
¹¹² mass at $-\tilde{x}(t)$ equal to $\Phi(-(\sigma^2 dt)^{-1})$. Thus, when $x(t)$ is far from zero, $dW(t)$ resembles Gaussian

113 fluctuations, and when the population is close to extinction, the point mass at $-\tilde{x}(t)$ implies a non-
114 negligible probability of population collapse, after which $x(t) = 0$ for all t .

115 We simulated time series from (2) and (3) for 1000 time steps, with parameters $r = 0.05$, $K = 2$,
116 $a = 0.023$, $h = 0.38$, and $q = 5$, and $\sigma = 0.02$, except where otherwise indicated. Initial conditions
117 were in the vicinity of the ghost attractor, with $x_0 = 0.3$. To explore how the mean trajectory
118 changed with the variance, we calculated the mean of 5000 realizations for each of 4 levels of variance:
119 $\sigma = 0.005, 0.01, 0.015, 0.02$. When assessing possible inference, we considered both a single time series
120 as well as ensembles of realizations.

121 2.2 Hidden Markov Model and Changepoint analyses

122 We applied two standard time series approaches to our generated time series: Hidden Markov Mod-
123 els and Changepoint analysis. A Hidden Markov Model assumes observed dynamics result from an
124 underlying Markov process whose transition matrix is determined through model fitting. This model
125 classifies each point in the time series into one of two states (note that the number of states is deter-
126 mined *a priori* by the user). We interpreted the end of the long transient as the first time at which
127 the Hidden Markov Model assigned a change in state. Changepoint analysis determines the point in
128 the time series which most parsimoniously separates the data into two statistically different sets. We
129 applied both of these approaches to a single time series as well as an ensemble of 100 simulations. All
130 analyses were conducted in R (R Core Team, 2019) using packages **ecp** (James and Matteson, 2014)
131 and **depmixS4** (Visser and Speekenbrink, 2010).

132 2.3 Bayesian model fitting

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

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

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

¹⁴¹ We considered values of σ_{me}^2 ranging from 0.005 to 0.08.

¹⁴² *2.3.1 Fitting to the generating mechanistic model*

¹⁴³ We performed model fitting and parameter estimation in a Bayesian framework. Priors specified for
¹⁴⁴ all parameters in (2)–(5) are given in Table 1. The model was fit to the simulated data using a
¹⁴⁵ Markov chain Monte Carlo (MCMC) procedure with the R package **nimble** (de Valpine et al., 2017,
¹⁴⁶ see Supplement).

¹⁴⁷ One notable challenge to fitting the generating model via MCMC was the existence of strongly
¹⁴⁸ correlated parameters, which can impede the efficiency of traditional univariate samplers—the default
¹⁴⁹ sampler for most Bayesian model fitting software. To better explore the joint posterior distribution,
¹⁵⁰ we utilized adaptable block Metropolis-Hastings samplers and ran the MCMC algorithm for 100,000
¹⁵¹ iterations, discarding the first half of the iterations as burn-in and thinning to every 10th iteration
¹⁵² for a total of 5,000 posterior draws. We assessed convergence visually using traceplots and checking
¹⁵³ effective sample sizes. We refer readers to Gelman et al. (2020) for a general guideline to implementing
¹⁵⁴ Bayesian hierarchical models.

generating model priors		non-parametric model priors	
r	Gamma(2, 10)		
K	Gamma(1, 0.1)		
a	Gamma(2, 10)	$\beta_i, i = 1, \dots, m = 5$	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)

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

¹⁵⁵ *2.3.2 Kullback-Leibler divergence*

¹⁵⁶ To quantify Bayesian learning, we used the Kullback–Leibler (KL) divergence, which measures the
¹⁵⁷ information gain from prior to posterior distributions. For a generic parameter θ (any parameter in
¹⁵⁸ Table 1), the Bayesian model fitting can be summarized as

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

¹⁶⁰ where $\mathbf{x} = x(t)$ are the time-series data. The KL divergence is a measure of the discrepancy
¹⁶¹ between two probability densities based on information entropy (Kullback and Leibler, 1951). The
¹⁶² KL divergence between the prior ($[\theta]$) and the posterior ($[\theta|\mathbf{x}]$) distributions quantifies how much

¹⁶³ knowledge about θ has changed in light of information contained in the data (Itti and Baldi, 2006).

¹⁶⁴ It is defined mathematically as

$$\text{165} \quad D_{\text{KL}}([\theta|\mathbf{x}] \parallel [\theta]) = E_{\theta|\mathbf{x}} \log \left(\frac{[\theta|\mathbf{x}]}{[\theta]} \right) = \sum_{\theta} [\theta|\mathbf{x}] \log \left(\frac{[\theta|\mathbf{x}]}{[\theta]} \right). \quad (7)$$

¹⁶⁶ A KL divergence of 0 indicates that the prior and posterior distributions are identical. A large KL
¹⁶⁷ divergence indicates substantial information gain from prior to posterior. We computed KL divergence
¹⁶⁸ using a nearest neighbor search algorithm to calculate the distance from simulated samples of the
¹⁶⁹ prior distribution and the MCMC samples of the posterior distribution using the R package FNN
¹⁷⁰ (Beygelzimer et al., 2019).

¹⁷¹ *2.3.3 Importance of ghost attractor*

¹⁷² The ghost attractor in this model is found for a parameter set very near to a bifurcation point, where
¹⁷³ the deterministic model changes from having one non-zero stable state to having two alternative stable
¹⁷⁴ states. To explore how our model fitting results depended specifically on the ghost attractor, we also
¹⁷⁵ fit models for two additional scenarios on either side of the bifurcation by varying parameter a —one
¹⁷⁶ with a weaker ghost attractor ($a = 0.0225$) and one with two non-zero stable states ($a = 0.0235$).

¹⁷⁷ *2.3.4 Non-parametric model fitting*

¹⁷⁸ As mentioned in Section 2.3.1, a challenge to implementation for Bayesian model fitting is strong
¹⁷⁹ conditional dependence among the parameters in the model, especially those appearing in (2). Strong
¹⁸⁰ conditional dependence can indicate the potential for difficult to explore likelihood surfaces and
¹⁸¹ highly inefficient MCMC algorithms. We investigated an alternative, non-parametric specification of
¹⁸² 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
¹⁸³ m -dimensional functional space, such that

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

¹⁸⁴ By relaxing the assumed parametric form, it is possible to specify a model in which parameters
¹⁸⁵ do not suffer from the same high degree of dependence, leading to more efficient implementation. The
¹⁸⁶ trade-off for the improvements in model fitting are that direct interpretation of model parameters in
¹⁸⁷ (2) via their posterior distribution is no longer possible. However, important features of $\frac{d\mu(t)}{dt}$ such as
¹⁸⁸ curvature, and regions of $x(t)$ where the function is zero or nearly zero are still available through the
¹⁸⁹ non-parametric approach.

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

193 *2.3.5 Comparing parametric and non-parametric approaches*

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

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

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

215 **3 Results**

216 **3.1 The effect of variance on model dynamics**

217 The addition of stochasticity to the model with the ghost attractor resulted in mean behavior that
 218 differed from that of the deterministic core model. Even relatively small variance resulted in a sub-
 219 stantial reduction in the expected time spent in the area of the ghost attractor (Figure 2a). In spite
 220 of the fact that the variance term is symmetric around the deterministic model core, the asymmetry
 221 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 (Figure 2b, c).

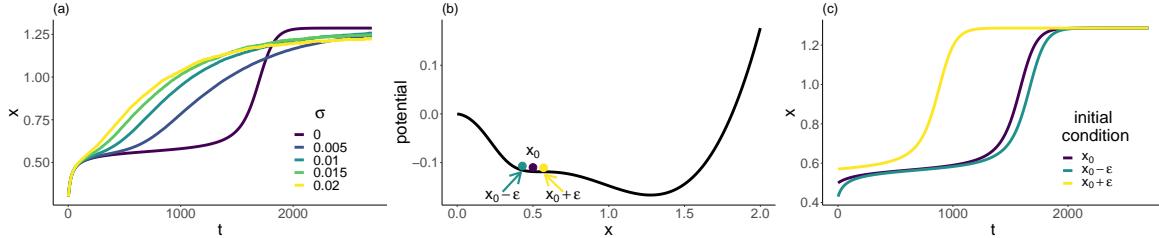


Fig. 2: (a) The effect of variance on the mean trajectory. We simulated 5000 realizations from Eq. (2)-(3) and took the mean for each of 4 different variance levels ($\sigma = 0.005, 0.01, 0.015, 0.02$). (b) Three initial conditions centered at the ghost attractor; x_0 , and $x_0 \pm \epsilon$ (here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

226 3.2 Results from Hidden Markov Model and changepoint approaches

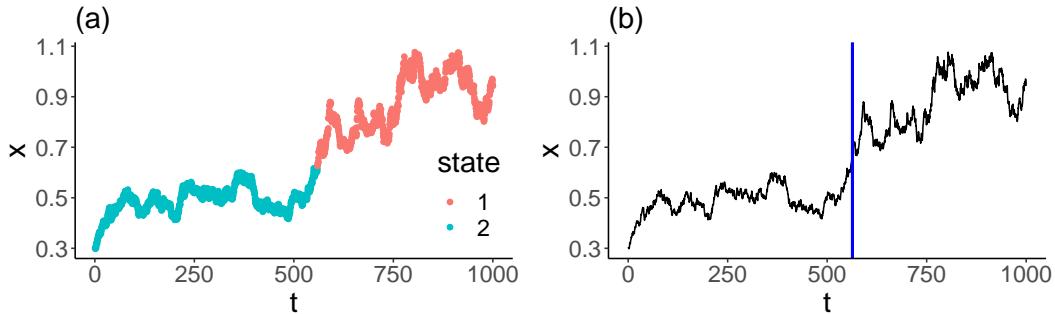


Fig. 3: Illustrative model results for a single time series. (a) Hidden Markov Model results, with two states represented by the two colors. (b) Changepoint analysis. The vertical blue line is the changepoint, signifying the most parsimonious break in the time series.

The Hidden Markov Model and changepoint analysis identified similar transition times from the ghost attractor to the stable state (Figure 3). Individual realizations varied in the timing of their transition away from ghost attractor (grey lines in Figure 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 (Figure 4), reflecting the behavior seen in Figure 2a.

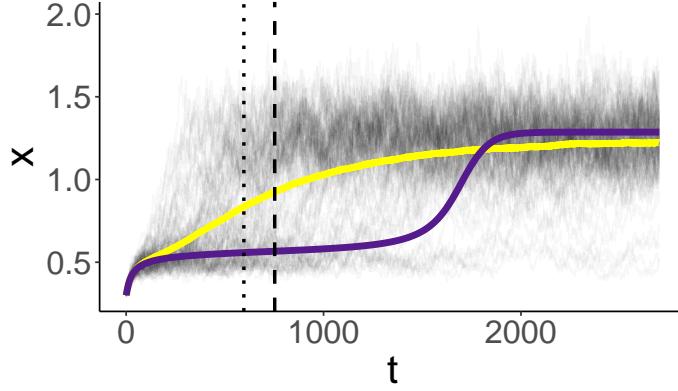


Fig. 4: Results from the Hidden Markov Model (dotted vertical line) and changepoint analysis (dashed vertical line) applied to an ensemble of 100 time series (each realization is plotted in grey in the background). The purple curve is the deterministic model core, and the yellow curve is the mean of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

²³⁴ 3.3 Parametric model fitting

²³⁵ Parameters from the generating mechanistic model appeared to be at least weakly identifiable from
²³⁶ realized population trajectories, and marginal posteriors showed evidence of concentrating around the
²³⁷ parameter values used to simulate the data as the number of realizations increased (Figure 5). The
²³⁸ derivative of the potential function was well-estimated in the vicinity of the ghost attractor ($x \approx 0.5$),
²³⁹ which suggests the simulated trajectories contain enough information about the presence of these
²⁴⁰ transient-inducing characteristics to allow us to identify them from data (Figure S1). Additional model
²⁴¹ fitting diagnostics confirm that while the parameters are only weakly identifiable, the derivative of the
²⁴² potential function ($\frac{d\mu(t)}{dt}$) is identifiable (Figures S2 and S3), providing insights into model dynamics
²⁴³ in spite of weak insights into parameter values.

²⁴⁴ 3.3.1 KL divergence

²⁴⁵ In most—but not all—cases, the KL divergence shows increasing trends from 1 to 10 realized trajec-
²⁴⁶ tories (top right corner of each sub-plot in Figure 5). This general pattern of increased KL divergence
²⁴⁷ for increasing realizations is visually consistent with the densities shown in Figure 5, where more
²⁴⁸ information is gained by using a larger number of realized population trajectories. Observe, however,
²⁴⁹ that for parameters r, q, a and h , more realizations did not always improve the KL divergence. It
²⁵⁰ should also be noted that the posteriors increasingly concentrate around the values used to simulate
²⁵¹ the data. Finally, changes in information about certain parameters as measured by KL divergence
²⁵² are non-linear in the number of realized trajectories. For example, much more is learned about h and
²⁵³ K in moving from 1 to 2 realizations than 2 to 5, while more is learned about q when we increase the
²⁵⁴ number of realizations from 5 to 10.

255 *3.3.2 Sensitivity of results to ghost dynamics*

256 When we compared scenarios with varied values of a corresponding to a very weak ghost attractor
 257 ($a = 0.0225$), a stronger ghost attractor ($a=0.023$), and two alternative stable states ($a = 0.0235$),
 258 we found negligible differences in our inference (Figure 6). In all three cases, the distribution of the

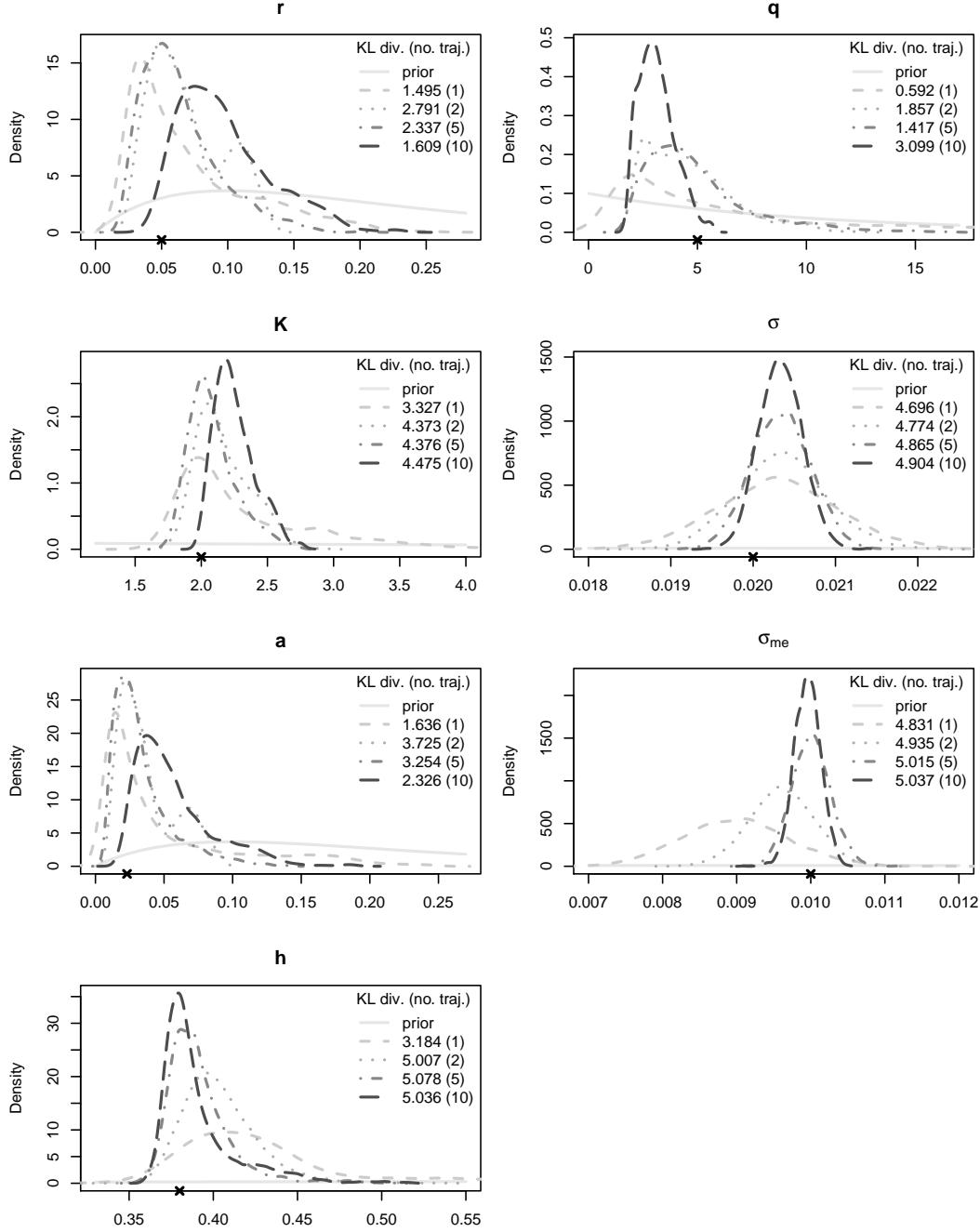


Fig. 5: Comparison of marginal posterior distributions for simulations with 1, 2, 5, and 10 realized trajectories. Each “ \times ” on the x-axis denotes the parameter value used to generate the data. KL divergences from the prior to the posterior are in the top right of each sub-plot. Parameters for the realizations were $r = 0.05$, $K = 2$, $a = 0.023$, $h = 0.38$, and $q = 5$, $\sigma = 0.02$, and $\sigma_{me} = 0.01$.

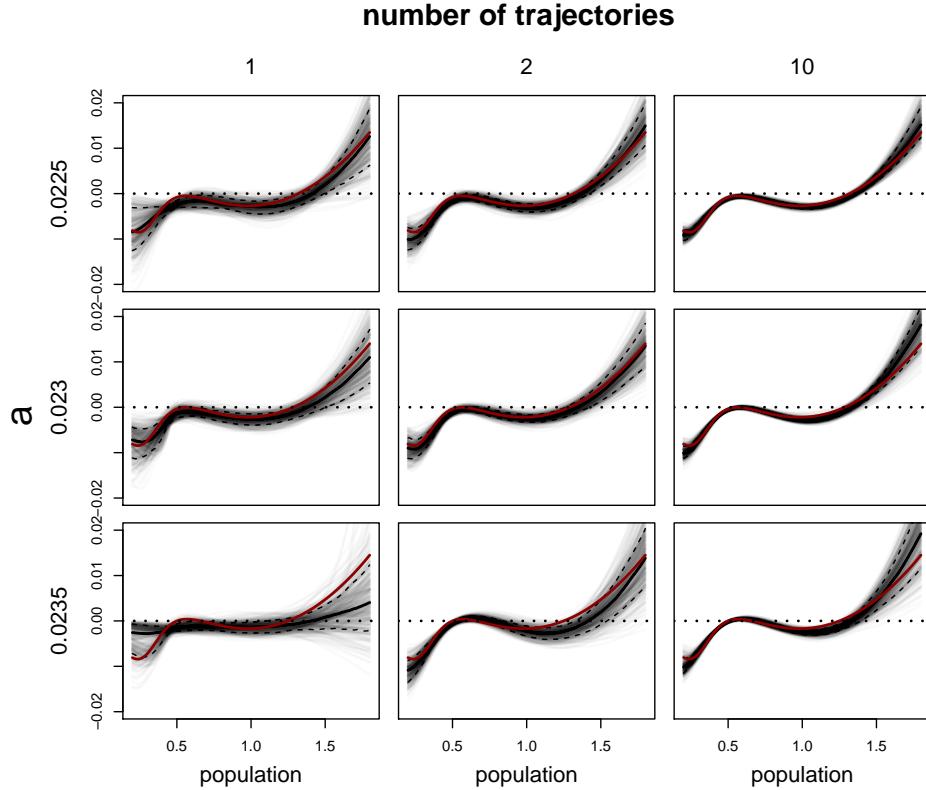


Fig. 6: Comparison of parametric model fitting results as a is varied through a scenario with a very weak ghost attractor (top row), a stronger ghost attractor (middle row), and two alternative stable states (bottom row). Curves were drawn from the posterior distribution of $\frac{d\mu(t)}{dt}$ in grey, 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 $r = 0.05$, $K = 2$, $a = 0.023$, $h = 0.38$, and $q = 5$, $\sigma = 0.02$, and $\sigma_{me} = 0.01$.

derivative of the potential functions concentrated around the true values as the number of realizations used in model fitting increased from 1 to 10.

3.4 Non-parametric model fitting

Results from the non-parametric 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 (Figure S4). Model diagnostics for the non-parametric model were an improvement from those of the parametric model (Figures S5, S6).

3.4.1 Comparing parametric and non-parametric results

The ESS is much higher for the non-parametric model than the generating model (Figure 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 non-parametric model is nearly 5000, while that

271 of the generating model is near 0. A look at the IMSE shows no clear indication which model fit the
 272 data better (Figure 7b).

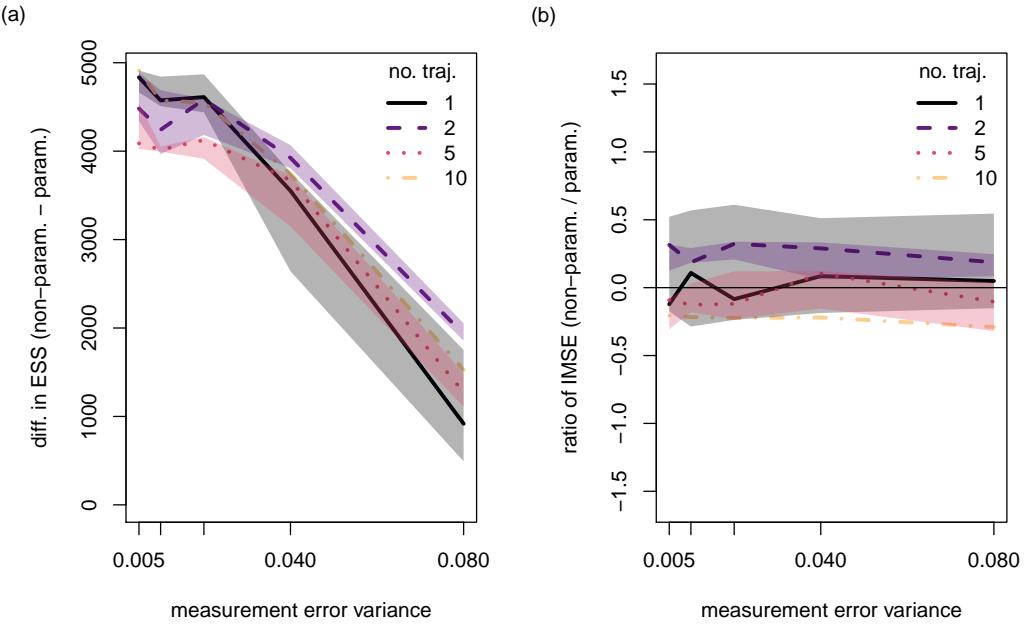


Fig. 7: (a) The difference in effective sample size (ESS) between the non-parametric model and the parametric (generating) model decreases for larger measurement error variance, σ_{me}^2 . (b) No clear trend emerges when comparing the integrated mean squared error (IMSE) for the two model types, suggesting 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 $r = 0.05$, $K = 2$, $a = 0.023$, $h = 0.38$, and $q = 5$, and $\sigma = 0.02$.

273 4 Discussion

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

282 Identifying long transients in ecological systems requires connecting theoretical advances, which
 283 have generally been derived from deterministic models, to the stochasticity of the real world. Stochas-
 284 ticity enters at myriad levels, from the relative simplicity of measurement error, to environmental and
 285 demographic stochasticity. Whichever its sources, stochasticity—depending on its type, magnitude

286 and also how accurately it is accounted for in modeling approaches—can obscure the deterministic
287 skeleton that we wish to uncover. Thus, advancing ecological theory on long-transients requires bridg-
288 ing stochastic versions of common models producing long transients to statistical inference methods
289 that can robustly match data to pattern and process.

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

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

303 Additional modelling methods have been proposed elsewhere to study similar systems. For exam-
304 ple, in the literature on so-called tipping points, changes in the standard deviation and autocorrelation
305 coefficient of a time series may provide early warning signs of tipping points from one regime to an-
306 other (Scheffer et al., 2009; Dakos et al., 2008; Wissel, 1984). Consideration of a time series’ spectral
307 density has also been proposed (Biggs et al., 2009). This provides a method to identify earlier warning
308 signs of tipping points than those identified by changes in the standard deviation and autocorrelation.
309 However the system studied in this paper is different from those typically discussed in the literature
310 on tipping points, which are usually attributed to a slowly shifting parameter resulting in a sudden
311 bifurcation of stable states. Since all parameters are fixed in time in our model, there is no reason
312 to expect that the tools developed to analyze systems experiencing tipping points should provide
313 insights here. Indeed, while performing exploratory analyses of changes in standard deviation, auto-
314 correlation, and spectral density of our generated time series, we found that all three methods failed
315 to consistently identify the shift from the ghost attractor to the steady state across the realizations.

316 Our Bayesian inference approach to estimating the underlying mechanistic model suggested that
317 parameters were weakly identifiable. Posterior estimates from MCMC increased in accuracy and
318 decreased in uncertainty as the number of model realizations increased. KL divergences further confirm
319 Bayesian learning from prior to posterior distributions driven by data. This estimation performance is
320 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 stable states. This suggests 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., Figure 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 Figures 5 and 6.

The findings of the non-parametric model fitting were very similar to that of the parametric model fitting, as seen through their similar IMSE values (Figure 7b). While the ESS of the non-parametric model was considerably higher than that of the parametric model, especially for low measurement error (Figure 7a), scientific knowledge about the parameter values cannot be gained through consideration of the non-parametric 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 non-parametric models together may yield the greatest insights; the more efficient non-parametric 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.

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. An important next step may be to explore whether it is possible to differentiate between time series generated from a model with a ghost attractor or from a model with a time varying parameter that crosses a bifurcation threshold (e.g., parameter a in our model).

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.

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.

354 **Declarations**

355 **Funding**

356 KZ acknowledges support from the US National Science Foundation, DEB-1926438 and the University
357 of California, Santa Cruz, Committee on Research, Faculty Research Grant.

358 **Conflict of interest**

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

360 **Ethics approval**

361 Ethics approval was not required for this study.

362 **Consent to participate and publication**

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

364 **Availability of data and material (data transparency)**

365 No unpublished data were used in this study.

366 **Code availability (software application or custom code)**

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

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. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 370(1962):1228–1249
- Beisner BE, Haydon DT, Cuddington K (2003) Alternative stable states in ecology. *Frontiers in Ecology and the Environment* 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. URL <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. *Proceedings of the National academy of Sciences* 106(3):826–831
- Boettiger C (2018) From noise to knowledge: how randomness generates novel phenomena and reveals information. *Ecology Letters* DOI 10.1111/ele.13085, URL <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. *Proceedings of the National Academy of Sciences* 105(38):14308–14312
- Dulvy NK, Freckleton RP, Polunin NV (2004) Coral reef cascades and the indirect effects of predator removal by exploitation. *Ecology letters* 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. *Ecology Letters* 4(3):215–220, DOI 10.1046/j.1461-0248.2001.00220.x, URL <http://doi.wiley.com/10.1046/j.1461-0248.2001.00220.x>

- Hastings A, Higgins K (1994) Persistence of transients in spatially structured ecological models. *Science* 263(5150):1133–6, DOI 10.1126/science.263.5150.1133, URL <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), DOI 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. *Journal of Statistical Software* 62(7):1–25, URL <http://www.jstatsoft.org/v62/i07/>
- Kullback S, Leibler RA (1951) On information and sufficiency. *The annals of mathematical statistics* 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. *Philosophical Transactions of the Royal Society B: Biological Sciences* 370(1659):20130269
- Ludwig D, Jones DD, Holling CS, et al. (1978) Qualitative analysis of insect outbreak systems: the spruce budworm and forest. *Journal of animal ecology* 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. *Journal of the American Statistical Association* 83(404):1023–1032
- Nolting BC, Abbott KC (2016) Balls, cups, and quasi-potentials: quantifying stability in stochastic systems. *Ecology*
- R Core Team (2019) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria, URL <https://www.R-project.org/>
- Rietkerk M, van de Koppel J (1997) Alternate stable states and threshold effects in semi-arid grazing systems. *Oikos* pp 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. *Journal of Computational and Graphical Statistics* 26:403–417, [arXiv:1505.05093v1](https://arxiv.org/abs/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. *Journal of Statistical Software* 36(7):1–21, URL <http://www.jstatsoft.org/v36/i07/>
- Wissel C (1984) A universal law of the characteristic return time near thresholds. *Oecologia* 65(1):101–107

A 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 (plummer2006). 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 (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 non-parametric model given by (8) evaluated at $\hat{\beta}$. We defined the IMSE for either the parametric (P) or non-parametric (NP) approach as

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

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.

B Supplementary Figures

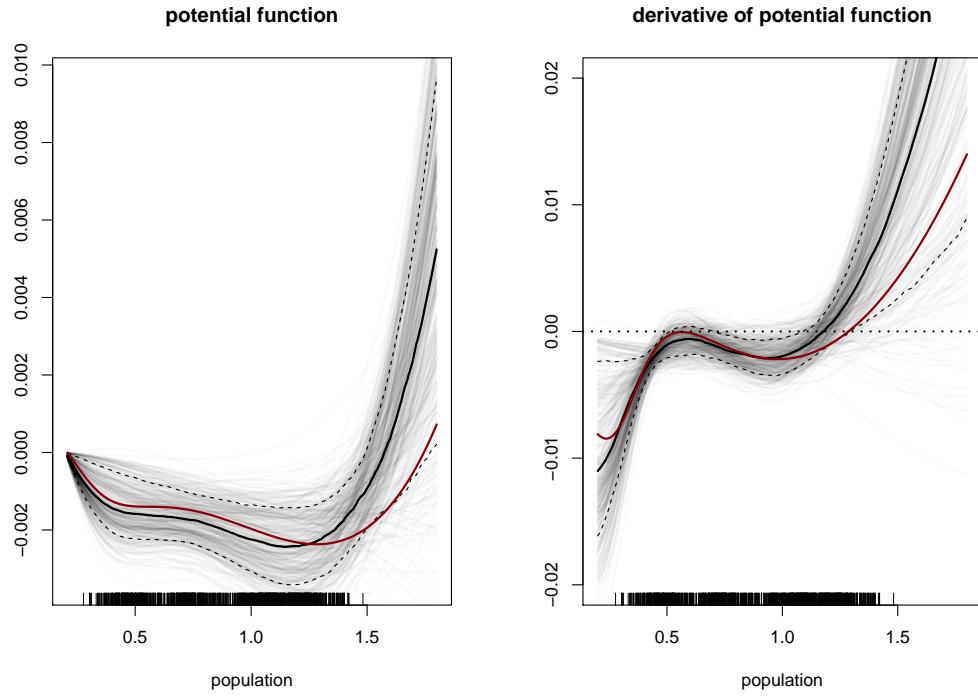


Fig. S1: Fitted potential curves (left) and the derivative of the potential curves (right) from the parametric model fitting to a single realized population trajectory. Curves drawn from the posterior distribution are in grey, pointwise median value in black, pointwise equal-tailed 95% credible intervals as dashed, and true curves are red. Points near 0 represent ‘stable’ population sizes (i.e., the population equilibrium and ghost attractor). Model was generated using parameters $r = 0.05$, $K = 2$, $a = 0.023$, $h = 0.38$, and $q = 5$, $\sigma = 0.02$, and $\sigma_{me} = 0.04$. Model fitting diagnostics in Figures S2 and S3

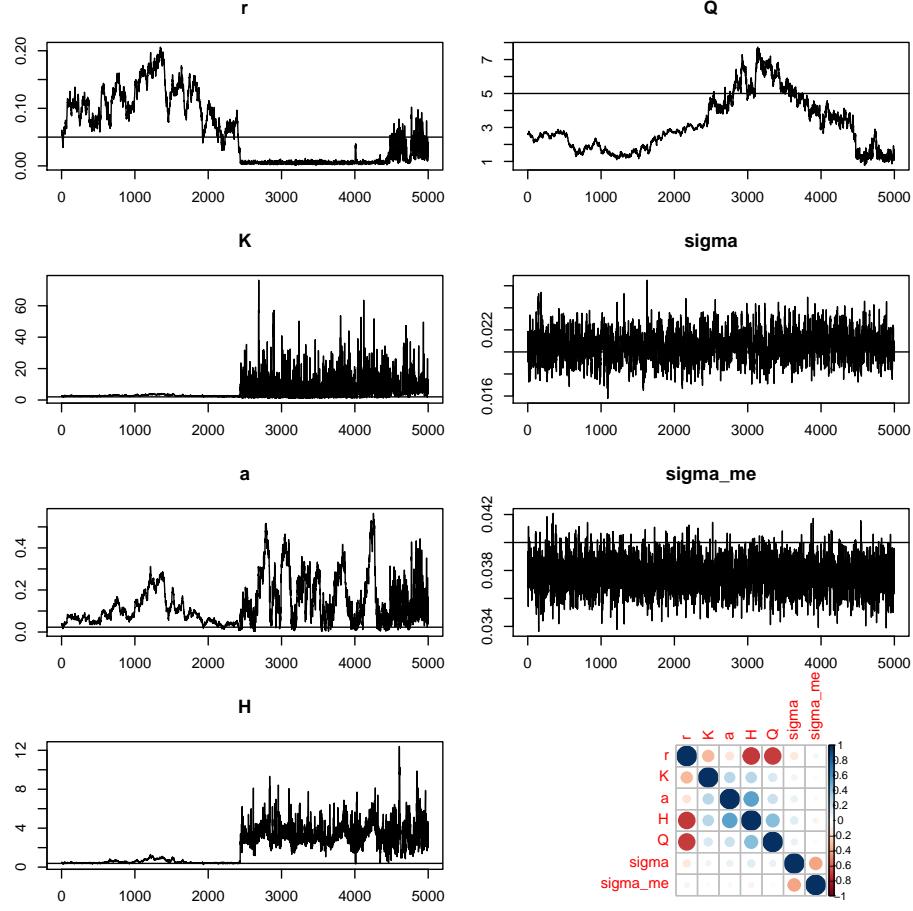


Fig. S2: The trace plots of MCMC chains for simulation with 1 realization, corresponding to Figure S1. Bottom right shows sample correlations among parameters. Parameter values used to generate the data are shown as horizontal black lines ($r = 0.05$, $K = 2$, $a = 0.023$, $h = 0.38$, $q = 5$, $\sigma = 0.02$, and $\sigma_{me} = 0.04$). While these parameters appear only weakly identifiable from the data, corresponding trace plots for $\frac{d\mu(t)}{dt}$ suggest, however, that the potential function is identifiable (Figure S3).

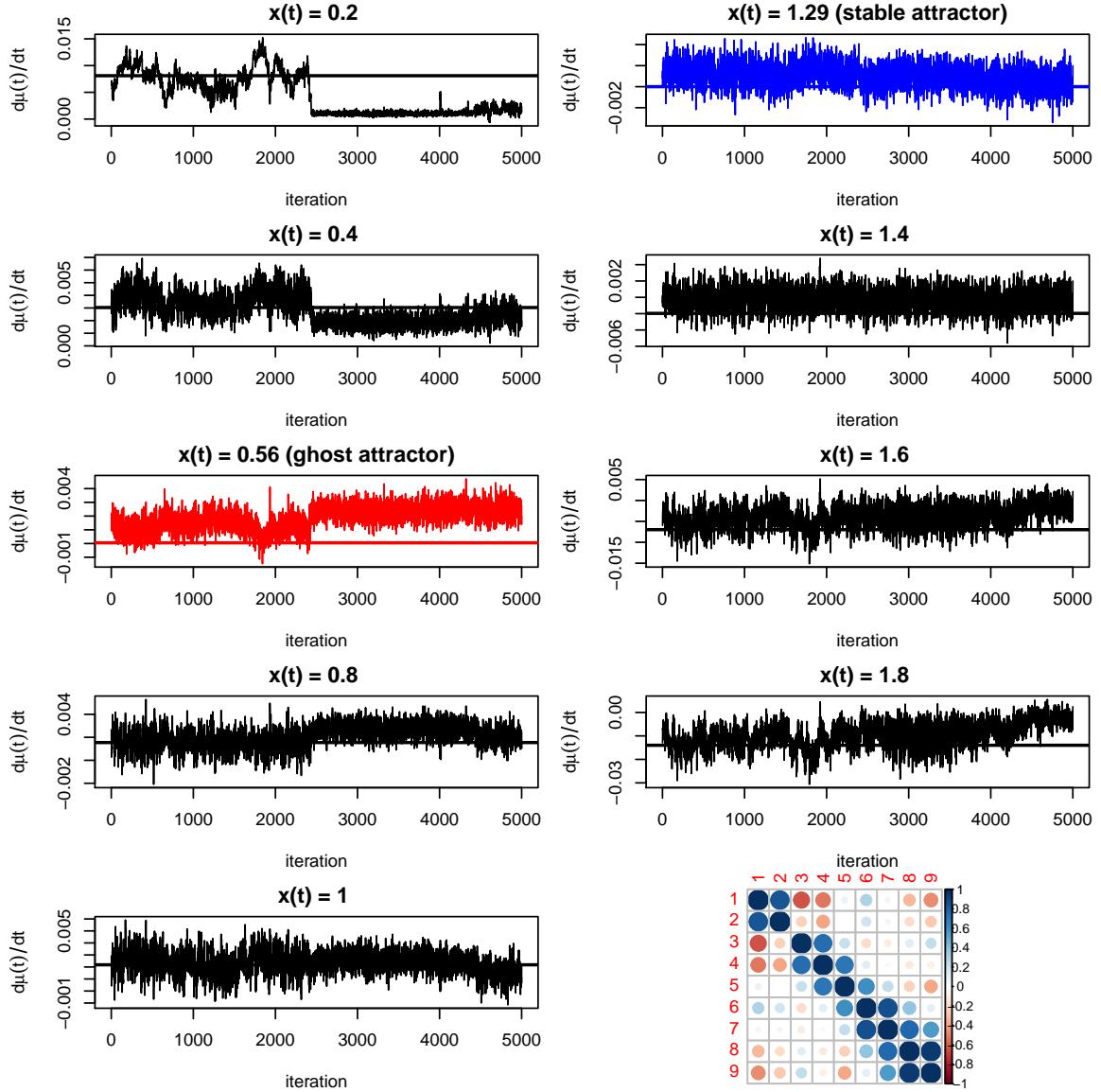


Fig. S3: Trace plots for points along the function $\frac{d\mu(t)}{dt}$. Each plot represents one value of $x(t)$, including the locations of the ghost attractor and the stable attractor.

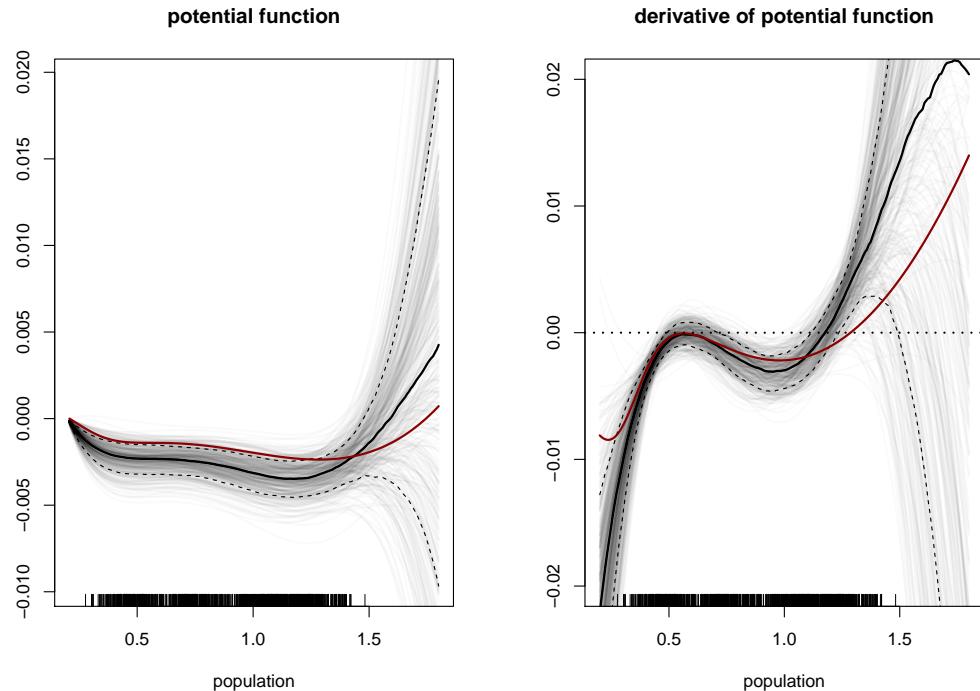


Fig. S4: Fitted potential curves (left) and the derivative of the potential curves (right) from the non-parametric model fitting to the same single realized population trajectory as Figure S1. All other details are as in Figure S1. Model fitting diagnostics in Figures S5 and S6.

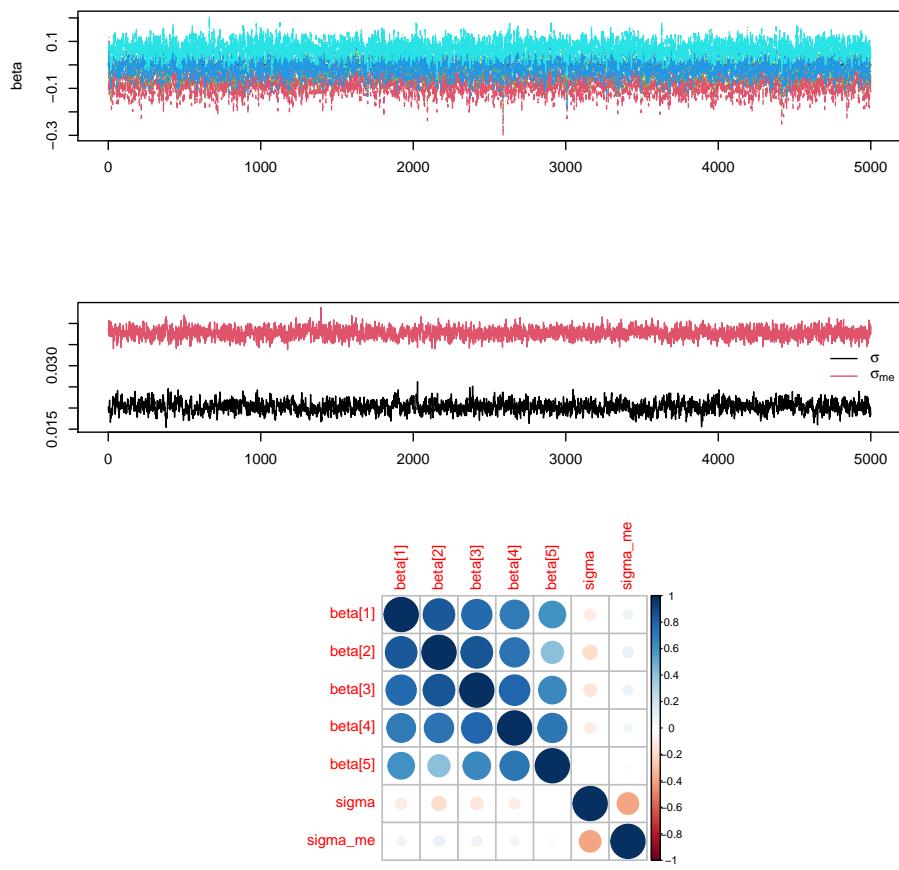


Fig. S5: Left: trace plots of MCMC chains for simulation with 1 realization, corresponding to Figure S4. Note that y-axis is on log scale. Right: sample correlations among parameters.

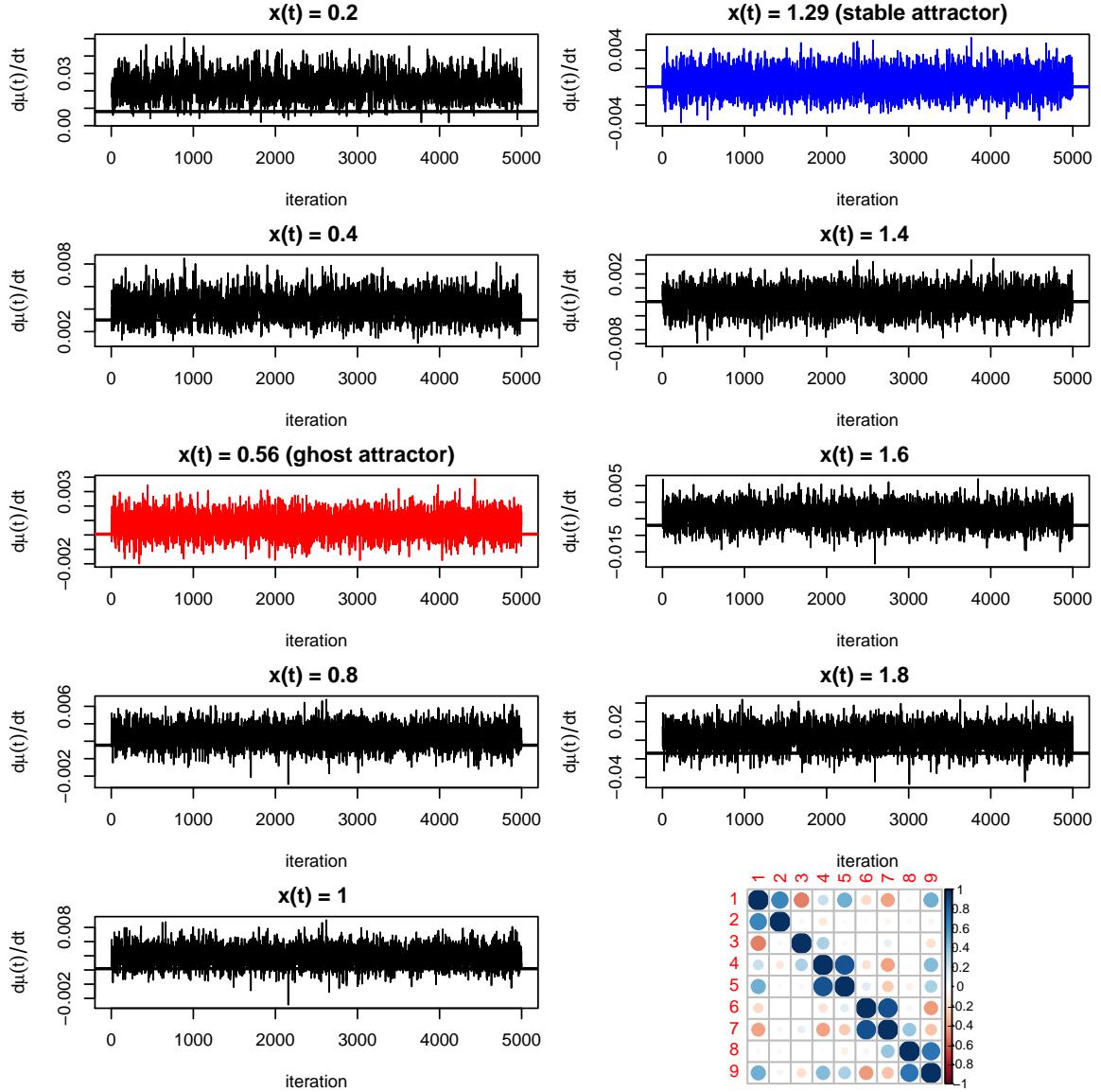


Fig. S6: Trace plots for points along the function $\frac{d\mu(t)}{dt}$. Each plot represents one value of $x(t)$, including the locations of the ghost attractor and the stable attractor.