

phenomenological and dynamical models for understanding long transients.

Arroyo-Esquivel, Boettiger, Jiang, Reimer, Scharf, Wolkovich, Zhu

October 21, 2019

## 1 Introduction

### 1.1 Brief introduction to long transients

- Dynamical (mechanistic) and statistical (phenomenological) models often focus on stationary dynamics
- However, increasing evidence in dynamical systems and from empirical data suggest transient dynamics may be important
- Here's some motivating examples (why this might be useful):

#### 1.1.1 Empirical examples from the literature

- Regime shifts
- Animal movement

#### 1.1.2 Theoretical (dynamical) examples from the literature

- examples of mechanisms: ghost attractors, saddle node crawl-bys, periodic behaviour with long periods, etc.

But there doesn't seem to be a lot of overlap between the statistical and dynamical models. Something about how Carl (and others?) have shown how to combine statistical and dynamical models in stationary systems ([We had this in our initial overview but I don't recall what we were referring to. Thoughts?](#)). Here we look at how statistical and dynamical models can contribute to our understanding of a (dynamically driven) long transient and discuss challenges for inference in this context.

Main idea: You've got a data set that seems to have some transient behaviour, what can (and can't) these different types of tools tell us about it?

### 1.1.3 Phenomenological modeling approaches

Briefly describe each approach, provide references of examples where people have used these methods in this context.

- changepoint analysis; quantify when we seem to have switched regimes (Jody)
- Hidden Markov Model (Jorge)
- ARIMA model (Jorge)
- look for changes in variance; looking for an increasing in variance or a “critical slowing down” (Junjie)
- look for changes in strength of auto-correlation; should increase before tipping points (Junjie)

### 1.1.4 Mechanistic modeling approach

Briefly describe approach, provide references of examples where people have tried to do this ([are there any?](#)). The main steps are:

1. construct a model encapsulating hypothesized mechanisms; here the correct model is known
2. estimate model parameters from time series data

## 1.2 Introduce the May model

- FYI: this model is, as far as I can tell, *not* from May 1976 (that paper is about chaos arising from discrete time models); it seems to be introduced in “Thresholds and breakpoints in ecosystems with a multiplicity of stable states”, by May in 1977.
- model motivation:  $x$  is a resource that has logistic growth, and that is being grazed by a consumer according to a generalized Holling type response ([can someone remind me what this functional form is actually called?](#))
- The full model is:

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

which May nondimensionalized to

$$dX/d\tau = X(1 - X) - \frac{\gamma X^q}{\alpha^q + X^q}$$

using

$$X = x/K, \quad \tau = rt, \quad \gamma = a/(rK), \quad \alpha = h/K$$

- note that we are all currently working with the full model, not the nondimensionalized one
- overview of what this model has been used for (original purpose and other applications)

- model used to show alternate stable states; it's well known and easily biologically interpretable
- show that we can get a ghost attractor for certain parameter values

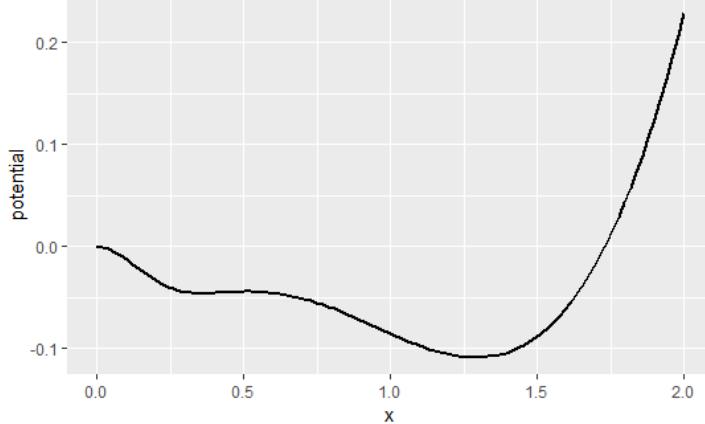


Figure 1: potential function for this model showing ghost attractor

### 1.3 Brief paper overview

- we consider a “simplest case”:
  - \* generate our own data using the May model
  - \* generate data where we observe a “full” trajectory (this should give us best chance of being able to say anything about the underlying model)
  - \* analyze the time series using standard phenomenological approaches, as well as try to get back the parameters of the May model when we already know the true underlying model structure

## 2 Methods

### 2.1 Data generation

- Describe data generation from May model
  - \* assume normally distributed noise term
  - \* variance scaled by the mean
- Generated single time series and also ensemble of 1000-ish independent time series
- Currently, parameters are fixed; Lizzie had mentioned possibly drawing the parameters from a distribution when generating replicates (e.g., a  $N(\mu_a, \sigma_a)$ ); do we want to do this, or stick with fixed values?

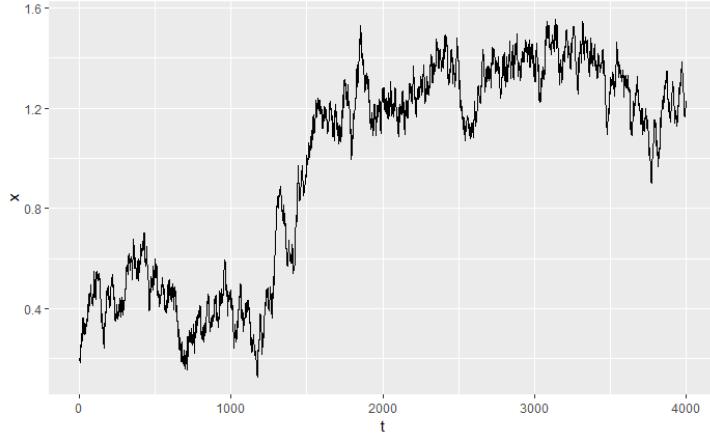


Figure 2: Simulated data looks like this

## 2.2 Phenomenological approaches taken

Describe technical details of these methods (packages used, any other methodological decisions made):

- changepoint analysis; provides way to say when we switched regimes (Jody)
- Hidden Markov Model (Jorge)
  - \* groups data points into different categories
  - \* number of categories specified *a priori*
- ARIMA model (Jorge)
  - \* We aren't sure if ARIMA models can be used on an ensemble; any updates on this Jorge?
- look for changes in variance (Junjie)
- look for changes in strength of auto-correlation (Junjie)

## 2.3 Model fitting approach taken

We should probably pick either nimble, ghost, or stan and move forward with just one of them. Preferences?

- Considered scenarios where (1) all parameters were free and (2) we fixed  $a$  and  $K$  their known values
- MCMC details:
  - \* R package(s) used
  - \* describe the non-informative priors used; uniform on [0,1] for everything?
- looked at posterior distributions of estimated parameters
- quantify how much learning about the parameters is possible using FNN method for KL divergence (Kai)

## 2.4 Comparing models

- Compare models; How do we want to do this? Originally we talked about how well they do perform forecasting phase shifts, but I think they'll all perform pretty poorly. Is there another way we want to compare them?
  1. Cross-validation (k-fold, or leave-one-out) on given data
  2. Compare forecast to model predictions
- We had also talked about comparing how models respond to perturbations; either perturbations of the state variable, or of parameters
- could compare how close the estimated “transition points” from the phenomenological models are to the expected escape time from the ghost attractor

## 3 Results

### 3.1 Results from phenomenological approaches

- changepoint analysis; way to say when we switched regimes (Jody; Figure 3)

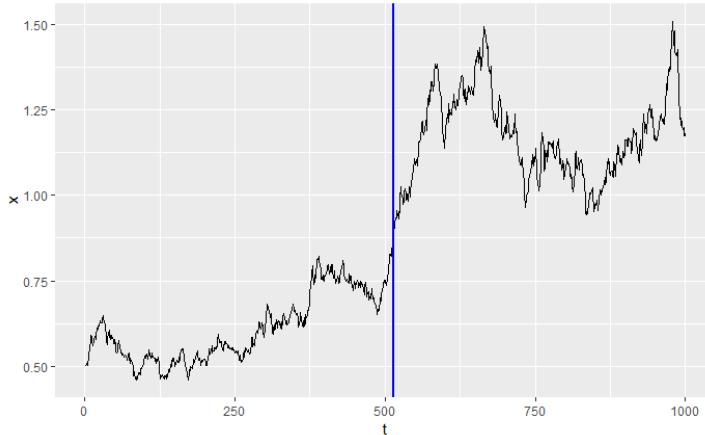


Figure 3: Results from a changepoint analysis looks something like this. It just provides the most likely place for a regime shift.

- Hidden Markov Model (Jorge)
  - \* initial results: when looking at one time series, it tended to put the breaking point between groups in a suitable spot half the time, and somewhere weird the rest of the time
  - \* when it was run on 1000 replicates, however, it did a better job predicting a suitable breaking point
- ARIMA model (Jorge)
  - \* ARIMA in a predictive capacity doesn't do well; it keeps the shifted behaviour within the 80% CI, even though it doesn't capture the “regime shift” at all

- \* results are conservative, in that they don't predict wild swings per se, but the rapid shift away from the ghost still falls within the ARIMA 80% CI so the model is "open to the possibility" of transient behaviour
- look for changes in variance (Junjie)
- look for changes in strength of auto-correlation (Junjie)

### 3.2 Results from model fitting

- identifiability of all model parameters based on a single observed trajectory does not look promising
- get biased posteriors; tends to imply a weaker ghost (Figure 4)

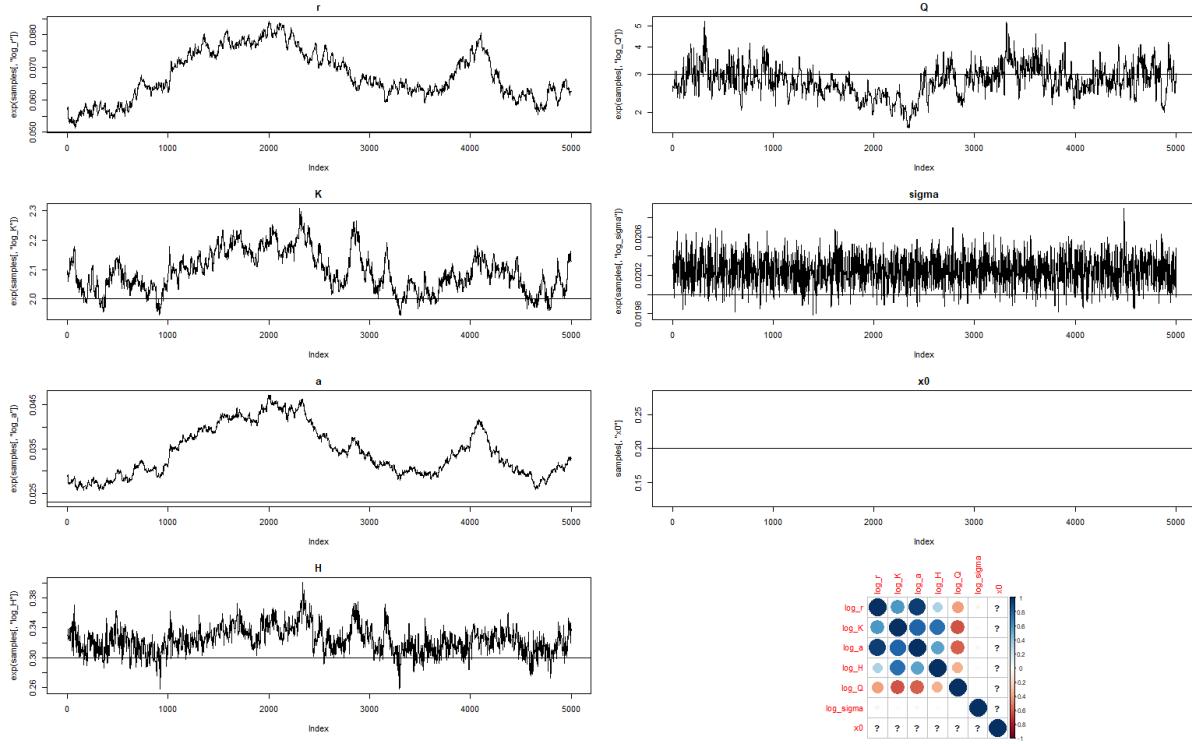


Figure 4: mcmc chains (when all parameters are allowed to vary); note that the parameter values used to generate the data are  $r = 0.05$ ,  $K = 2$ ,  $a = 0.023$ ,  $H = 0.3$ ,  $q = 3$ , and  $\sigma = 0.02$ .

- when we allow  $a$ ,  $r$ , and  $K$  to vary, they are highly correlated (bottom right of Figure 4)
- In response to the biased estimators, Lizzie asked Andrew Gelman and Mike Betancourt: "What diagnostic tools do you recommend for assessing the information contained about a particular parameter, or set of parameters, in the data? What tools do you recommend for designing efficient data collection procedures based on a desire to learn about specific sets of parameters (e.g., how can we determine whether it's better to make observations of trajectories for longer periods of time vs. at higher

resolution time scales if we want to learn about  $a$ )? Under what conditions will the posterior mean/mode be a biased estimator of the “true” parameter?”

- \* Lizzie: both Andrew and Mike honed in on the “jumps” in the data ([I don’t quite know what these jumps are referring to. Did they think we needed higher temporal resolution?](#))
- \* Mike suggested following a standard Bayesian workflow to identify what is and is not possible with the model ([Someone stats-y could do this.](#))
- \* Mike said these types of issues come up a lot in systems where small noise or differences in initial conditions produce wildly different trajectories; one way to demonstrate this is to do prior predictive checks on super tight priors (delta-function priors). See , step 7. ([Someone stats-y could do this.](#))
- \* Mike went through the steps we’d need to take: (1) prior predictive checks (2) fit the data from the prior predictive checks (3) check the z-score contraction plot. If the posteriors all collapse to extreme values, your model is non-identifiable. ([Again, someone stats-y could do this.](#))
- \* [Do we want to do the above things ourselves, or Lizzie also asked if Mike might be willing to join project as co-author and he said sure, if he has time and if not he can recommend good people. Thoughts?](#)
- \* Andrew: Said we should try to understand the jumps; that’s why we replicates are good; otherwise we just have  $n = 1$  on the jumps; Suggested we get a distribution of timing of jumps and try to understand it. [Would this be similar to a distribution of escape times from the ghost?](#)
- Kai was looking into measuring the amount of learning about parameters that happens (i.e., quantifying the difference between the prior and posterior distribution); Using the the FNN method for KL divergence seemed to be a good approach [Kai, what do you need to go ahead with that?](#)

### 3.3 Result from comparing models

## 4 The effect of variance on model dynamics

[I’m not sure where this should go, but it seems worth thinking about a bit more.](#)

- increasing variance reduces the strength of the ghost attractor (Figure 5)
- seems to be that the asymmetry of the deterministic curve leads to the positive relationship between higher variance and quicker escapes (Figure 6)

## 5 Discussion

Stuff we’ve mentioned:

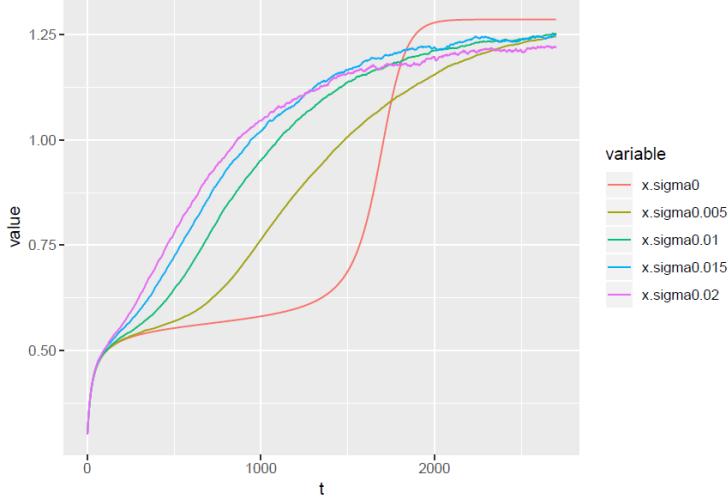


Figure 5: The effect of variance on the mean trajectory. I plotted the deterministic model core, and then simulated 1000 replicates for 4 different variance levels ( $\sigma = 0.005, 0.01, 0.015, 0.02$ ) and took the mean of each. You can see that even small variance really affects the average trajectory and reduces time spent in the ghost.

- big idea: phenomenological models are bad for forecasting and understanding transient dynamics; mechanistic models may be lauded as a better option to try to get at these things; but even if you have the right mechanisms, there may be tricky estimability issues
- Each different single time series effectively looks like a slightly different deterministic version of the model, making inference tricky.
- changing the seed for the single trajectory made a big difference in our ability to say anything meaningful about the underlying processes; a possible take home message from Henry: observing a single trajectory is probably not enough to help us learn much about these sophisticated parametric models, but maybe an ensemble of observations is. That general principle is pretty familiar to spatial/temporal statisticians: it's hard to estimate complicated covariance functions for a single observation of a process. They way people overcome the limitation there is typically to just fix or place very informative priors on the hardest to estimate parameters (which also sometimes have the weakest interpretability anyway). Maybe our recommendation for how to address the issue in practice could be similar, with some advice on how to solicit priors or picked fixed values.
- The reduction in the strength of the ghost attractor with increasing variance is likely what makes it difficult to get back parameters
- Challenges in predicting transient behaviour from real data.
- Coming up with new ways to describe phenomenon that perhaps could be better described with another method. For example, for animal movement data, where hidden Markov models are mostly used maybe you should be using dynamical models.

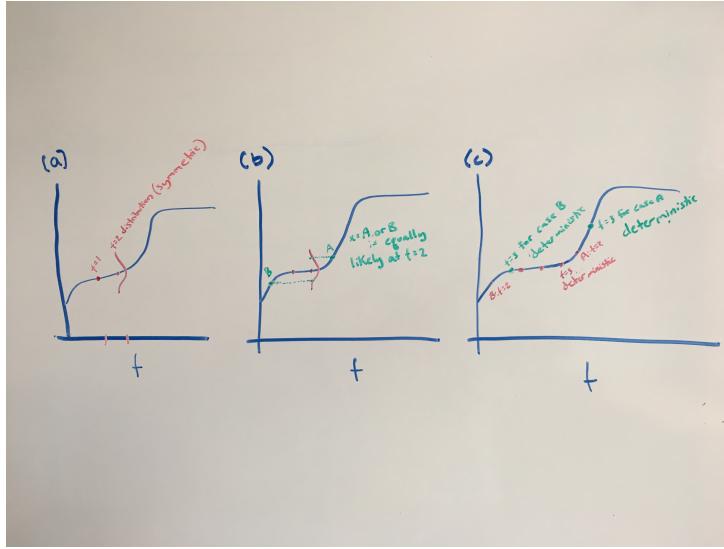


Figure 6: State of the art schematic helping me (Jody), at least, understand how symmetric noise reduces time spent in the ghost rather than just adding wiggles around the deterministic core. Henry suggested adding another pair of those red Gaussian curves with even larger (or smaller, depending on what's easier to draw) variance, doing the same A and B experiment and seeing where we wind up. If higher variance leads to quicker escape times, then the average of the two vertical values of A and B should be greater for the higher variance scenario. Higher variance will make it much more likely to reach higher populations, but only slightly more likely to reach lower ones.

- Focusing on phase shifts ... understanding if the behaviour is transient versus phase shift versus observing asymptotic behaviour ...
- If you don't think about transients (by going in with a preformed idea of a system of asymptotic or a mixture model) you can miss important transient behaviour which can be critical prediction.
- Are some dynamical models totally intractable statistically?
- From all this, get to what you need to know about system (helping improve experiments) ... If you can observe the system at transient dynamics you can perhaps learn more.
- Transients can arise from inside the system (autonomous) ... so if you cannot find an exogenous factor, maybe it's autonomous to your system
- Some other dynamical models with transients (that we dreamed of working on):
  1. Linear (matrix model)
  2. Chaotic
  3. Periodic (fast-slow)

## Issues

- Do we have any hope of detecting a ghost attractor in spite of the biased estimators?
- Can we look at the distribution of expected escape times for the posteriors?
- As Henry mentioned, it would be nice to have a bit of a positive message to our paper, such as “in these circumstances, we can say something about transient dynamics”, or something else constructive
- from Henry: while the individual parameters may not be strongly identifiable from these types of data, it might be that some other useful quantities are. For instance, how well are we able to resolve the potential function, which encapsulates important information about equilibria and any possible ghosts (Figure fig:posterior<sub>p</sub>otentials).
- From when Lizzie talked to Mike: Sometimes with these models it’s NOT about needing more time points on a grid, or more points across a grid, it’s about needing more points at certain places in the grid. In our case that’s probably where the jumps happen (low x to high x).
- We can also think about playing with the temporal resolution (e.g., time step of 1/2 instead of 1) or  $\sigma$