Regime shifts in ecology and evolution

By

CARL BOETTIGER
B.S. (Princeton) 2007

DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Center for Population Biology

in the

OFFICE OF GRADUATE STUDIES

of the

UNIVERSITY OF CALIFORNIA

DAVIS

Approved:

Alan Hastings (Chair)

Peter Wainwright

Brian Moore

Committee in Charge

2012

DEDICATION

Ed Boettiger

Abstract

The most pressing issues of our time are all characterized by sudden regime shifts: the collapse of marine fisheries or stock-markets, the overthrow of governments, shifts in global climate. Regime shifts, or sudden transitions in dynamical behavior of a system, underly many important phenomena in ecological and evolutionary problems. How do they arise? How can we identify when a shift has occurred? Can we forecast these shifts? Here I address each of these central questions in the context of a particular system. First, I show how stochasticity in eco-evolutionary dynamics can give rise two different domains, or regimes, governing the behavior of evolutionary trajectories (Boettiger et al., 2010). In the next chapter, I turn to the question of identifying evolutionary shifts from data using phylogenetic trees and morphological trait data of extant species (Boettiger et al., 2012). In the last chapter, I adapt the approach of the previous section which allowed me to quantify the information available in a given data set that could detect a shift into an approach for detecting regime shifts in ecological time series data before the occur (Boettiger and Hastings, 2012).

Acknowledgments

They say UC Davis is a collaborative and interdisciplinary place, and perhaps nowhere is this more evident then in the making of a PhD student. Seven faculty members across three departments have been close mentors and collaborators, the students of many different graduate groups have been a constant support, and I have been particularly fortunate it many collaborators beyond the walls and fields of UC Davis as well. All this makes for a long acknowledgments section, and I thoroughly apologize to those who deserved more mention than they are given here.

First I would like to acknowledge my funding source. The four generous years of support from the Department of Energy Computational Science Graduate Fellowship program were more than financial – the required program of study and my practicum at Lawrence Berkeley National Lab helped transformed me from a pencil and paper theorist to a more fully fledged computational scientist. Jeana Gingery and the rest of the Krell Institute handling the fellowship were always a pleasure to work with, and I'm particularly thankful for the connections to computational science across other disciplines that were made possible through the annual conference and interactions with the other CSGF fellows and alumni, and the other members of my CSGF cohort in particular who have been both friends and inspiration to me.

I am very grateful for the opportunity to spend summers with Ulf Dieckmann at IIASA, Vienna, Austria, and Adam Arkin at Lawrence Berkeley National Lab, which introduced and gave me time to explore new methods and also exposed me to different ways of running research group.

I would like to thank the Population Biology graduate group, with its excellent faculty, supportive students and notorious Core sequence, and in particular the constant support, motivation, wondering questions and good fun of my cohort Chris Martin,

Matt McGee, Michelle Afkahami, Rebbecca Best, and Kirsten Sellheim. In particular I will never forgive Chris Martin for that day as a second year graduate students when he walked into my office asking questions I couldn't understand about adaptive landscapes and phylogenetic trees, thereby launching my parallel research program in phylogenetic comparative methods. I'd also like to thank the Theoretical Tea, or Teary group: the members and friends of Schreiber, Hastings, and Baskett labs for so many good conversations between cups of caffeine over the years. Marissa Baskett has been particularly generous with her time and support, feedback on manuscripts and proposals and welcoming me into her lab group meetings.

Sebastian Schreiber has been an unofficial adviser, mentor since I arrived at Davis. Sebastian has been a source of inspiration as an excellent teacher, a brilliant mathematician and impeccable ecologist, and I have appreciated his patience and generosity as much as his kind words and support over the past five years.

The Wainwright lab has treated me like one of there own, inviting me along to their lab meetings, conferences, weddings and an even Icthyosaur fossil excavation in Nevada. Peter Wainwright has been a teacher, mentor, collaborator and endless source of clever ideas and fascinating data sets. Samantha Price has been my kindred computational spirit submersed in a fish lab and a source of insight and inspiration. I would also like to thank Brian Moore, my thesis committee member, teacher, and organizer and co-instructor of the Bodega phylogenetics workshop. Brian's Bayesian world-view and empirical skepticism and fear of bad and miss-applied software have had a profound impact on me.

Graham Coop and Peter Ralph approached me with some objections to a talk I presented in CPB, sparking a discussion that evolved into a close collaboration that led to the publication in Chapter 2 and laid the groundwork for ideas that would remain central to many other chapters of my thesis. Graham is a brilliant scientist with a knack for finding nice ways to say "you're wrong." I am ever grateful Peter's patient, clear and reliable insight and explanations to countless thorny questions in statistics and probability.

Paul Armsworth from University of Tennessee, who cornered me at a nice dinner at ESA 2011 to bring me into the Pretty Darn Good Control Working Group, has had more influence on my future research at the interface of ecological management and economics than either of us expected or intended at the time. The rest of the working group has also been both an fun to work with and an inspiration towards new approaches. Likewise Jim Sanchirico has become another mentor to me as I have struggled through stochastic dynamic programming and other optimal control approaches in our recent ecological economics work.

Duncan Temple-Lang has been a teacher, mentor, co-author, co-developer and immensely influential to my thinking and practices in software development, scientific reproducibility, and access, manipulation, visualization and synthesis of large data sources. Duncan has a special talent for communicating more paradigm-shifting ideas in a single meeting than almost anyone I've met, and not solely for the speed of talking. For anyone who has had to endure me discussing one of these topics over a tea or beer, you can blame Duncan for at least part of that.

I cannot resist an unorthodox acknowledgment to the British Chemist Cameron Neylon, whose own example inspired me to keep an open lab notebook, and whose thinking about the scientific process has had a clear and profound impact on my approach to research and my engagement in the online scientific community. In particular, that engagement led me to Karthik Ram and Scott Chamberlain, ecologists at Berkeley and Rice, who have become both close collaborators and virtual lab mates.

My office mate and dear friend Alex Perkins,whose example I attempt to follow in all things within a year lag. Alex has had to endure just about every step of this PhD alongside me: the banging on the computer, the open science rants, kvetching over the latest papers and celebrating the little victories of some new understanding or the successful execution of some lines of code.

My adviser, Alan Hastings for giving me both the freedom to explore and the advice to succeed. Alan's perpetual confidence encouragement, and words of wisdom have only allowed me to make enough mistakes to learn from, while his razor-sharp

intuition and encyclopedic knowledge of the field have been my continual guide.

None of this would be possible without those dearest to me. My twin, Alistair Boettiger, has been my life-long collaborator, sounding board, moral support, and backup memory system. My parents, for their support and confidence, patience and understanding; for raising me with courage to be different and the will to learn from others, and my fiancée, Louise Berben, for being so supportive and encouraging as I finish this chapter in my education.

To all the folks at the Sunwise cooperative, and board of Solar Community Housing Association (SCHA): thanks for all the distractions that turned out more valuable than what I was doing anyway. How often do you get to help lead a volunteer-driven non-profit through a million dollar project to relocate and rebuild two historic homes to provide LEED certified green housing to low income residents, help foster and launch a new non-profit bicycle collective, or organize over 400 volunteers for a community build to rescue another historic local cooperative? I've learned non-profit law, and patience, managing a $200,000 annual budget and managing other community volunteers.

This thesis is dedicated to Ed Boettiger, my paternal grandfather, an insect physiologist by profession whose love of mathematics led him to start testing his new approach to teaching calculus on a ten-year old me. Grandpa Ed passed away shortly before I finished, but I will always remember his love and inspiration.

Table of Contents

Chapter 1
Introduction

.

Regime shifts are one of the most dramatic signatures of complex systems âĂŞ the ability to experience sudden transitions in their behavior and patterns is an example of the emergent properties that set such systems apart. These shifts, or tipping points, can be found in many different contexts and can occur for different reasons that may never be encapsulated by a single theory. Understanding and forecasting these events challenges the capacity of our models, our statistical methods and our available data. How much can we learn about these systems, what are the best predictions we can make with the knowledge we have, and how do we make sound management decisions in face of such uncertainty?

Ecological regime shifts may include the sudden collapse of fisheries or forests, eutrophication of lakes and rivers, or tipping points from carbon cycle feedback loops of global climate change. Evolutionary regime shifts may include major transitions such as the invasion of new habitat or the development of a novel trait. A regime of a complex system can be defined as a set of typical patterns or rules describing its behavior. While we may neither understand nor be able to predict everything about this regime, we can say something about what outcomes we expect and what we do not. A mathematical model may capture these patterns and behavior within the regime, matching what has already been observed and even predicting future events with reasonable accuracy. A shift in regime brakes these rules or patterns, driving the system into another patter of behavior, or alternative stable state. Although our ability to model and predict a complex system is greatest within a regime, many of the most significant dynamics may occur by large and sudden changes that are difficult to anticipate.

The chapters of my thesis represent rather than exhaust the suite of different contexts in which I have explored these phenomena. The selection of these chapters also captures the pro-

gression of my research trajectory in various lenses. The first trajectory is that of computational and data-driven research. A numerical simulation is present in Chapter 2, but pencil-and-paper calculations are in the focus. By Chapter 4, my analytical work is still essential, but I have learned to use it as a launching point for my computational approaches, rather than an ends by themselves. The second trajectory is in problem scope – from a technical question interesting to a niche community, to the larger and rapidly growing area of phylogenetics, to the more immediately relevant problem of suddenly collapsing ecosystems. In the process, I have become most attracted to problems not for pure mathematical elegance, but their potential relevance. The third trajectory has been towards increasing openness and reproducible research by making all the products of the research effort available. Chapters 3 and 4, and much else besides, are captured as they unfolded each day through the pages of my open laboratory notebook. The texts, code, and data accompanying each of these works are fully available as self-contained R packages.

If regime shifts are a unifying theme of my thesis, then they are an emergent one. I came to graduate school from a Bachelor's in physics and a loose background in population genetics in clonal interference and in adaptive dynamics. My statement of interest was something about stochasticity and the interface of ecology and evolution, and Chapter 2 is an excellent example of my thinking at the time. In Chapter 2 (Boettiger et al. 2010) I explore the stochastic deviations from a classic result, the canonical equation of adaptive dynamics. In what began as a nice analytical exercise, I was frustrated to find my approximations breaking down wildly when starting simulations a great distance from the evolutionary optimum. Like so many discoveries, what I assumed was a bug in my simulation code turned out to be the heart of the result: the emergence of domains where fluctuations were enhanced, rather than dissipated, by selective forces. The fluctuation-dissipation theorem had been central to my undergraduate physics training, and I was very happy to find an analytic expression for this boundary and a simple topological interpretation from the adaptive landscape. This gives rise to two regimes or domains of behavior on either side of the boundary. In the first domain evolutionary dynamics behave largely as expected, following the canonical equation of adaptive dynamics. In the second domain, evolutionary fluctuations are enhanced, leading to evolutionary trajectories that deviate widely from the expected canonical path. This can give rise to a bimodal distribution of outcomes, where the expected path is less common than either the much faster or much slower evolutionary trajectories.

Chapter 3 (Boettiger et al. 2012) is representative of my work in compartive phylogenetics that formed the basis of my dissertation proposal. Rather than derive phenomena from first principles, this chapter wrestles with the challenges of inferring evolutionary regime shifts from data. Analytic work will still play an important role, allowing me to derive the likelihood equation for a more general set of evolutionary transitions (Beaulieu et al 2012), but this chapter focuses instead on the side of statistical inference. In this context, Different evolutionary models represent different mechanisms or hypotheses for explaining the distribution of traits we observe in related species. The more complex models include regime shifts where the evolutionary parameters have changed in a particular group of lineages at some point in the evolutionary history. The challenge is to identify if, when, and how frequently such shifts have occured, using only present-day trait data and phylogenetic trees. Before adding my own models to the growing library of possibilities, I sought a robust method to identify just how well our data could distinguish between this complexity. The approaches of this chapter became central not only in establishing the validity of my later models, but also led me to see the need for more programmatic access to very large data sets, leading to the R packages and publications of rfishbase (Boettiger, Wainwright, and Temple Lang, accepted at Journal of Fish Biology) and treebase (Boettiger & Temple Lang, accepted at Methods in Ecology and Evolution) for the purpose.

This experience in teasing apart regime shift models in an evolutionary context echoed clearly in my mind as I encountered the growing literature on early warning signals of regime shifts in ecological collapses. Chapter 4 (Boettiger & Hastings 2012). Many aspects of the problem were quite different: while we have the benefit of using time-series data, rather than covariances on a phylogeny, now we would seek to identify regime shifts *before* they occur. Unlike the phylogenetics literature, none of the existing approaches were model based, leading me to start once more from pencil-and-paper to derive an appropriate model from the canonical form of the underlying bifurcation. Having learned the computational necessity of simplicity when inferring models from data, I went with the simplest approximation that could capture the essential features of the early warning. This not only allowed me to provide the first power analysis of early warning approaches, identifying the probabilities of false positive and false negatives in the detection, but also suggested to me a more powerful model-based indicator to provide early warning of a regime shift, also presented in Chapter 4.

Chapter 2
Fluctuation Domains in Adaptive Evolution

2.1 Introduction

Fitness landscapes have long been an important metaphor in evolution. Wright (1931) originally introduced the concept to explain his result that the mean rate of evolution of a quantitative trait is proportional to gradient of its fitness. The result and its accompanying metaphor continue to arise in evolutionary theory, having been derived independently in quantitative genetics (Lande, 1979), game-theoretic dynamics (Abrams, 1993; Hofbauer and Sigmund, 1998), and adaptive dynamics (Dieckmann and Law, 1996).ÂăThe metaphor creates a deterministic image of evolution as the slow and steady process of hill climbing. Other descriptions of evolution have focused on its more stochastic elements – the random chance events of mutations and the drift of births and deaths that underlie the process (Kimura, 1968, 1984; Ohta, 2002).ÂăIn this manuscript, we seek to characterize the deviations or fluctuations of evolutionary trajectories around the expected evolutionary path.

Our main result is that the size of deviations of evolutionary trajectories is determined largely by the *curvature* of the evolutionary landscape whose *gradient* is determining the mean rate of evolution. The landscape metaphor can be used to understand the interplay of stochastic and deterministic forces by identifying regimes where the selection will counterbalance or enhance such stochastic fluctuations. Because curvature can be positive (concave up) or negative (concave down), the evolutionary landscape can be divided into domains where fluctuations are enhanced or dissipated. To make this more precise, we focus on a Markov model of evolution used in the theory of adaptive dynamics. This Markov model gives rise to the familiar gradient equation – like that which first inspired the fitness landscapes metaphor – and also a more precise statement of the fluctuation dynamics. ÂăThe adaptive dynamics framework is general enough that we

can consider how these results apply in a variety of ecological scenarios. We illustrate the surprising consequence of bimodal distributions of expected phenotypes among parallel trajectories emerging from a fixed starting point on a single-peak adaptive landscape due to fluctuation enhancement. We also provide a landscape interpretation that suggests how the ideas of fluctuation dynamics apply to other ecological and evolutionary models.

2.2 Theoretical construction

Various definitions of fitness landscape have been used in the study of evolutionary dynamics. The conventional idea postulates a mapping between trait values and fitness (Levins, 1962, 1964; Rueffler et al., 2004). ÂăThe difficulty with this conception is that, in many cases, fitness of a given population type is mediated largely through competition with other populations, and is thus dependent on the current distribution of populations and their respective phenotypes in the environment. This dependence on densities or frequencies turns the static fitness landscape into a dynamic landscape, whose shape changes as the number of individuals, and their corresponding traits, changes.

It is possible to depict the fitness landscape that emerges even in the face of density- or frequency-dependent competition by returning to Wright's notion of a landscape. To do so, consider a resident population of individuals in an environment, each of which have an identical trait value, *i.e.*, a monomorphic population. Next, consider the per-capita fitness of a small number of mutant individuals with a similar, but different trait. ÂăThe per-capita fitness of mutants in the environment may be larger, smaller or identical to the per-capita fitness of residents. The gradient in fitness around the trait represented by the resident is termed the selective derivative (Geritz et al., 1997) in adaptive dynamics (analogous to the selective gradient in quantitative genetics). The fitness landscape is defined by integrating over these derivatives in trait space, to obtain a *local* picture of how fitness changes. ÂăIn the limit of small, slow mutation, it can be shown that the population will evolve as predicted by the shape of the landscape: by climbing towards a peak most rapidly when the slope is steep (Rueffler et al., 2004), see Fig. 2.1 for three illustrative examples. Âă The trait on the horizontal axis is that of the resident, not the mutant. ÂăThe lower panels show the slope of the mutant fitness as a function of the resident trait (up to a multiplicative factor)– the slope of the mutant fitness changes as the resident trait changes.

The trait of the resident population evolves in the direction given by that slope. ÂăIntegrating the lower panel along the resident trait axis reveals the fitness landscape the resident trait climbs in an evolutionary process (top panels). Note that changes in the mutant fitness are captured in the changes in slope of the landscape; the fitness landscape itself remains fixed. We derive the expected evolutionary dynamics of the phenotypic trait of the resident and its variation among parallel trajectories in the following sections.

2.2.1 Model

Consider a population monomorphic for a particular phenotypic trait, x. ÂăThe abundance of the population, $N(x,t)$ of trait x, is governed by its ecological dynamics:

$$\frac{\mathrm{d}N(x)}{\mathrm{d}t} = f(x,N,E), \tag{2.1}$$

which may depend on the trait, population abundance, and the environmental conditions, E. ÂăThe evolutionary dynamics proceed in three steps (Champagnat et al., 2006; Dieckmann and Law, 1996). ÂăFirst, the population assumes its equilibrium-level abundance $N^*(x)$, determined by its phenotypic trait x. ÂăNext, mutants occur in the population at a rate given by the individual mutation rate μ times the rate of births at equilibrium, $b(x)$. ÂăThe mutant phenotype, y, is determined by a mutational kernel, $M(x,y)$. The success of the mutant strategy depends on the invasion fitness, $s(y,x)$, defined as the per-capita growth rate of a rare population. Mutants with negative invasion fitness die off, while mutants with positive invasion fitness will invade with a probability that depends on this fitness (Geritz et al., 1997).

The invasion fitness is calculated from the ecological dynamics, Eq. (2.1), and will in general depend on the trait x of the resident population as well as that of the mutant, y. ÂăWe assume that a successful invasion results in replacement of the resident (Geritz et al., 2002, 1997), and the population becomes monomorphically type y. ÂăThe details of this formulation can be found in Appendix A.1. ÂăThe important observation is that such a model can be represented by a Markov process on the space of possible traits, x. ÂăThe population can jump from any position x to any other position y at a transition rate $w(y|x)$ determined solely by the trait values x and y. ÂăThe probability that the process is at state x at time t then obeys the master equation for the Markov process,

$$\frac{\mathrm{d}}{\mathrm{d}t}P(x,t) = \int \mathrm{d}y \ \left[w(x|y)P(y,t) - w(y|x)P(x,t)\right]. \tag{2.2}$$

No general solution to the master equation (2.2) exists. However, if the jumps from state x to state y are sufficiently small (*i.e.*, if mutants are always close to the resident), we can obtain approximate solutions for $P(x,t)$ by applying a method known as the Linear Noise Approximation (van Kampen, 2001), Appendix A.2. ÂăDoing so yields a general solution for the probability $P(x,t)$ in terms of the transition rates, $w(y|x)$, Eq. (A.16). ÂăThe linear noise approximation *derives* a diffusion equation as an approximation to the original jump process, as is commonly postulated. ÂăWhile this diffusion can be written as a partial differential equation (PDE), following Kimura, or as a stochastic differential equation, the solution for the probability density can be proven to be Gaussian and hence it suffices to write down ordinary differential equations for the first two moments (Kurtz, 1971).

2.2.2 The Fluctuation Equation

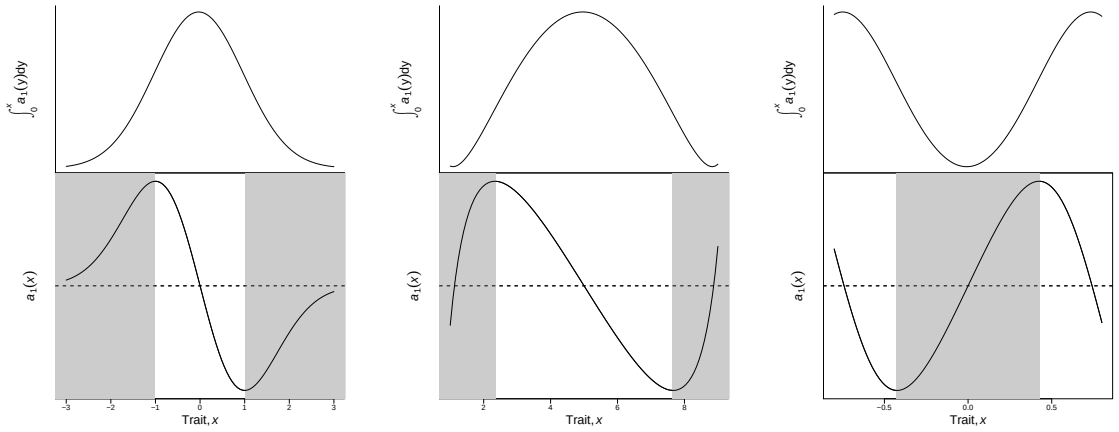
We assume the mutational kernel $M(y,x)$ is Gaussian in the difference between resident and mutant traits, $y-x$, with width σ_μ and that mutations occur at rate μ. A more thorough discussion of these quantities can be found in Appendix A.1, where they are developed in the process of deriving the transition probability $w(y|x)$ that specifies the underlying Markov process. Using Eq. (A.16) which results from the systematic expansion of the master equation (2.2), the mean trait $\hat{x}$ obeys

$$\frac{\mathrm{d}\hat{x}}{\mathrm{d}t} = \frac{1}{2}\mu\sigma_\mu^2 N^*(\hat{x})\partial_y s(y,\hat{x})|_{y=\hat{x}} \equiv a_1(\hat{x}), \tag{2.3}$$

with an expected variance σ^2 that obeys

$$\frac{\partial\sigma^2}{\partial t} = 2\sigma^2\partial_{\hat{x}}a_1(\hat{x}) + 2\sqrt{\tfrac{2}{\pi}}\sigma_\mu\left|a_1(\hat{x})\right|. \tag{2.4}$$

Eq. (2.3) recovers the familiar canonical equation of adaptive dynamics (Dieckmann and Law, 1996) for the mean trait. Eq. (2.4), which describes the variance, we term the *fluctuation equation* of adaptive dynamics. The linear noise approximation (see Appendix A.2) also predicts that the probability distribution itself is Gaussian, so that Eqs. (2.3) and (2.4) determine the entire distribution of possible trajectories in this approximation. To illustrate the local behavior of the model, we define the fitness landscape as $L(x) = \int_0^x a_1(y)\,dy$.



(a) Implicit Resource (b) Chemostat (c) Branching

Figure 2.1: **Fluctuation domains**. The lower panels show the rate of evolution given by Eq. (2.3) vs. the resident trait x for three ecological models: (a) implicit competition for a limiting resource (Section 2.4), (b) explicit resource competition (Appendix A.3), and (c) symmetric branching for dimorphic populations (with a resident population of trait x and another at $-x$, Appendix A.5). The horizontal dashed lines in the lower panel denote where $a_1(x) = 0$, *i.e.*, the selective derivative is zero, and correspond to evolutionarily singular points, *e.g.*, adaptive peaks and valleys. Shaded and unshaded regions in the lower panels correspond to trait values for which fluctuations are enhanced (shaded) and dissipated (unshaded). The adaptive landscape is defined as the integral of this expression (upper panels) – where populations climb the hills at a rate proportional to their steepness. Note that the adaptive landscape for the symmetric branching case has two peaks, and so evolutionary trajectories will lead to a stable dimorphism.

2.3 Fluctuation Domains

A closer look at Eq. (2.4) will help motivate our geometric interpretation of fluctuation domains. The approximation requires that the mutational step size σ_μ is very small, hence the second term will almost always be much smaller than the first (except when the fluctuations σ themselves are also very small). Note that the second term resembles the right-hand side of (2.3), differing only by a small scalar constant and the fact that it is always positive. Interestingly, this means that the second term vanishes at the singular point where $a_1(\hat{x}) = 0$ and $\sigma^2 = 0$: fluctuations cannot be introduced at an evolutionary equilibrium. That the first term is the gradient of the deterministic (mean) dynamics and the second term is smaller by a factor of the small parameter in the expansion are general features of the linear noise approximation.

The first term of (2.4) implies that the variance σ^2 will increase or decrease exponentially at a rate determined by $\partial_x a_1(x)$; *i.e.*, the *curvature* of the fitness landscape. This landscape and its gradient are depicted in Fig. 2.1 for several ecological scenarios. Plotting the gradient itself makes it easier to see when fluctuations will increase or decrease. On the gradient plot the singular point is found where the curve crosses the horizontal axis.

In the neighborhood of a singular point corresponding to an adaptive peak, the slope is negative, hence the first term in Eq. (2.4) (the coefficient of σ^2) is negative and fluctuations dissipate exponentially. This means that two populations starting with nearby trait values will converge to the same trajectory. In this region no path will stochastically drift far from the mean trajectory, hence the canonical equation will provide a good approximation of all observed paths. We term the part of the trait-space landscape where $\partial_x a_1(x) < 0$ the fluctuation-dissipation domain, analogous to the fluctuation-dissipation theorem found in other contexts (van Kampen, 2001).

Farther from the singular point corresponding to an adaptive peak, $\partial_x a_1(x)$ becomes positive (see Fig. 2.1(a) and Fig. 2.1(b)). While this part of trait-space still falls within the basin of attraction of the singular point, the variance σ^2 between evolutionary paths will grow exponentially. Initially identical populations starting with trait values in this region will experience divergent trajectories due to this enhancement. The variation can become quite large, with evolutionary trajectories that differ significantly from the mean. We call this region of trait-space where $\partial_x a_1(x) > 0$ the fluctuation enhancement domain. Eventually trajectories starting in this region

will be carried into the fluctuation-dissipation domain, where they will once again converge.

Evolutionary branching points (Dieckmann and Doebeli, 1999; Geritz et al., 1998) provide another example of a fluctuation enhancement domain. Until now, we have assumed that successful mutants replace the resident population, a result known as "invasion implies substitution" (Geritz et al., 2002). However, at an evolutionary branching point, the mutant invader no longer replaces the resident population, and the population becomes dimorphic. In dealing with monomorphic populations, we have been able to describe the evolutionary dynamics in terms of the change of a single trait, describing a single resident population. For dimorphic populations, two resident populations coexist, and both influence the shape of the landscape. Thus, an invader's fitness, $s(y; x_1, x_2)$, and its rate of evolution, $a_1(x_1, x_2)$, will depend on how it performs against both populations. The evolutionary dynamics after branching are two-dimensional and require a multivariate version of Eq. (2.4). Despite this, we can still gain qualitative insight using the intuition that connects fluctuation domains to curvature.

The existence of the stable dimorphism fundamentally distorts the landscape. While the population was monomorphic, being closer to the singular point always ensured a higher fitness. Once a resident population sits on either side of a branching point though, the singular point becomes a fitness *minimum*. This description of evolution towards points which become fitness minima after branching has been addressed extensively elsewhere (Geritz, 2004; Geritz et al., 1998, 1997). Here, what interests us is the effect of branching on fluctuations. If the landscape is smooth, a minimum must have positive curvature and therefore must be an enhancement domain. The farther a mutant gets from the branching point, the faster it can continue to move away, thus enhancing the initially small differences between mutational jumps. A rigorous description of this effect would require a multivariate version of Eq. (2.4) which is beyond the scope of this paper. Instead, we illustrate the enhancement effect by taking a slice of the two-dimensional landscape by assuming that the dimorphic populations have symmetric trait values about the singular strategy, $x_1 = -x_2 = x$. The fitness landscape for the case of symmetric branching is shown in Fig 2.1(c), and the derivation provided in Appendix A.4. Note that the region around the branching point is a fluctuation enhancement regime and the two new fitness maxima far from the branching point are in fluctuation dissipation regimes.

To provide a more concrete illustration of the fluctuation dynamics, we will focus on the de-

scription of the ecological competition model and compare the theoretical predictions of Eq. (2.3) and Eq. (2.4) to point-process simulations of the Markov process (Gillespie, 1977).

2.4 An Example of Fluctuation Domains in Resource Competition

2.4.1 An Ecological Model of Implicit Competition for a Limiting Resource

The logistic model of growth and competition in which populations compete for a limited resource is a standard model in ecological dynamics (Dieckmann and Doebeli, 1999). Here, we consider the population dynamics of $N(x,t)$ individuals each with trait x:

$$\frac{\mathrm{d}N(x,t)}{\mathrm{d}t} = rN(x,t)\left(1 - \frac{\sum_y N(y,t)C(x,y)}{K(x)}\right), \tag{2.5}$$

where r is the birth rate and $\sum_y rN(y)C(x,y)/K(x)$ is the density-dependent death rate. In the model, $K(x)$ is the equilibrium population density and $C(x,y)$ is a function which describes the relative change in death rate of individuals of type x due to competition by individuals of type y. Given $C(x,x) = 1$, the equilibrium density of a monomorphic population is $N^*(x) = K(x)$. Following Dieckmann and Doebeli (1999) we assume the following Gaussian forms,

$$K(x) = K_0 e^{-x^2/(2\sigma_k^2)}, \tag{2.6}$$

and

$$C(x,y) = e^{-(x-y)^2/(2\sigma_c^2)}, \tag{2.7}$$

where σ_k and σ_c are scale factors for the resource distribution and competition kernels respectively. We focus on the case $\sigma_c > \sigma_k$, for which the model has a convergence-stable evolutionarily stable strategy (ESS) at $x = 0$ (Geritz et al., 1998). Having specified the ecological dynamics, we must also specify the evolutionary dynamics, which we assume occur on a much slower time scale. We assume that with probability μ an individual birth results in a mutant offspring, and that the mutant trait is chosen from a Gaussian distribution centered at the trait value of the parents and with a variance σ_μ^2. From this we can assemble $w(y|x)$, and compute equations (2.3) and (2.4) (see Appendix A.2).

2.4.2 Numerical simulation of stochastic evolutionary trajectories

There are many ways to represent evolutionary ecology models in numerical simulations, including finite simulations, fast-slow dynamics, and point processes (Champagnat et al., 2006; Champagnat and Lambert, 2007; Dieckmann and Doebeli, 1999; Nowak and Sigmund, 2004). We choose to model the evolutionary trajectories of the ecological model in Eq. (2.5) by explicitly assuming a separation of ecological and evolutionary time scales. First, given a trait x, the ecological equilibrium density $N^*(x)$ is determined. Then, the time of the next mutant introduction is calculated based on a Poisson arrival rate of $\mu r N^*(x)$. In essence, we are assuming that ecological dynamics occur instantaneously when compared to evolutionary dynamics. The trait of the mutant, y, is equal to x plus a normal deviate with variance σ_μ^2. The mutant replaces the resident with probability given by the standard branching process result, $1 - d(y,x)/b(y,x)$ where d and b are the per-capita birth and death rate of the mutant, respectively, so long as $b > d$, and with probability 0 otherwise. The ecological steady state is recalculated and the process continues. Simulations are done using Gillespie's minimal process method (Gillespie, 1977). Estimates of the mean phenotypic trait $\hat{x}$ and the variance σ^2 are calculated from ensemble averages of at least 10^5 replicates.

2.4.3 Comparison of Theory and Simulation

We consider simulations with three different initial conditions: starting within the fluctuation-dissipation domain ($x = 1$), just into the fluctuation-enhancement domain ($x = 2$), and deep into the fluctuation enhancement regime ($x = 3$) in successive rows in Fig. 2.2. For each condition, the simulation is repeated 10^5 times and we compare the resulting distribution of trajectories to that predicted by theory. In the first column we plot the numerically calculated mean path taken to the ESS (shown in circles), and compare to the theoretical prediction of Eq. (2.3). In the second column, we plot the variance between paths, and compare to the predictions of Eq. (2.4). As each replicate begins under identical conditions, the initial variance is always zero. In the third column, we present the distribution of traits among the replicates at a single instant in time. We chose the moment of maximum variance so that deviations from the theoretically predicted Gaussian kernel will be most evident.

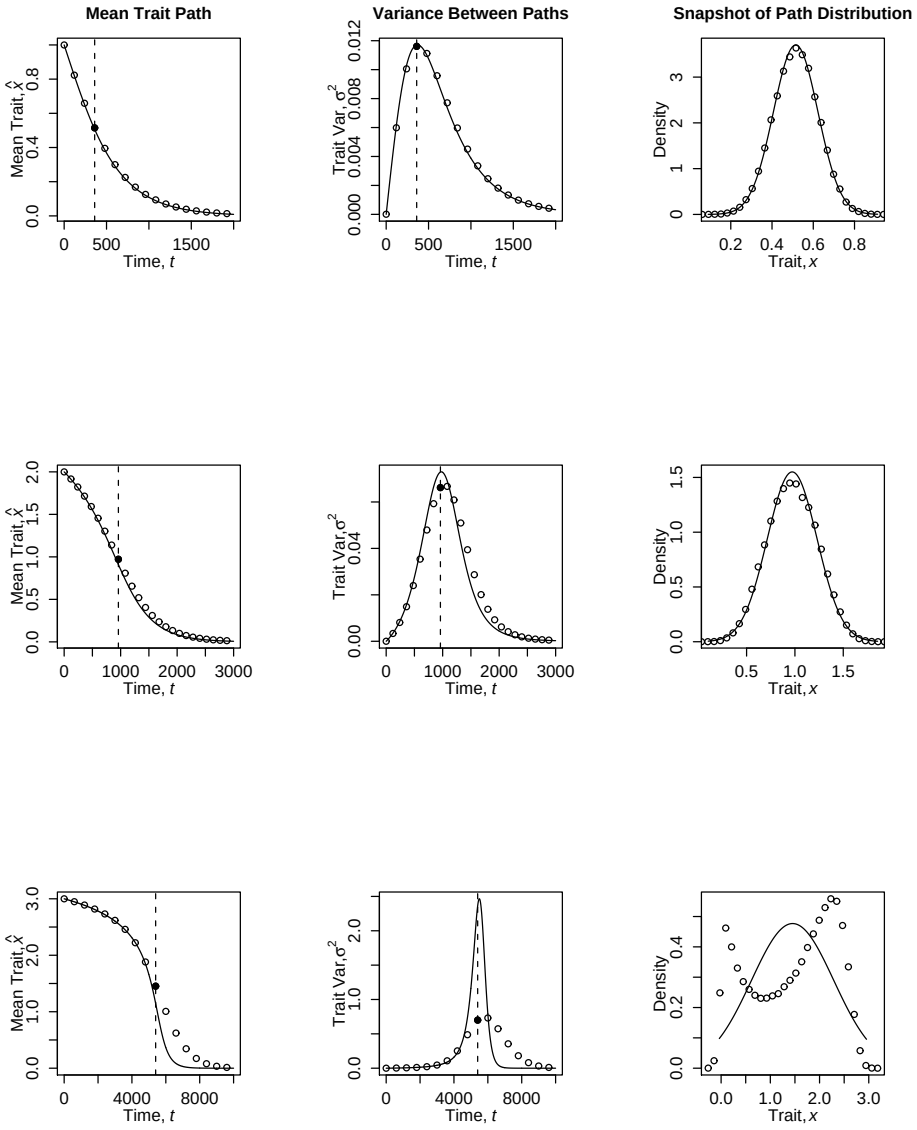


Figure 2.2: **Simulations over fluctuation domains**. The dynamics of the mean path, variance between paths, and snapshots of the trait distribution in time for three initial conditions: $x_0 = 1$, $x_0 = 2$, and $x_0 = 3$ (top to bottom successively). The three rows correspond to trajectories experiencing fluctuation dissipation (tow) and fluctuation enhancement (bottom) along with an intermediate case (middle). The mean (first column) and variance (second column) among trait values are plotted over time. Circles are simulation averages from 10^5 replicates and lines are theoretical predictions. The final column shows the theoretical Gaussian distribution (solid line) at the point in time indicated by the dashed line and closed circle in the first two columns, and the actual histogram (circles) of positions across the replicates at that time. The bottom right panel shows a bimodal distribution among replicates in the case of fluctuation enhancement. **Parameters:** $\sigma_k^2 = 1$, $\sigma_c^2 = 1.01$, $r = 10$, $\sigma_\mu = 0.0005$, $\mu = 1$.

Fluctuation-Dissipation Domain

Starting at the edge of the dissipation domain, fluctuations grow for only a short time before rapidly decaying, in the first row of Fig. 2.2. Since the simulation is sampled at regular time intervals, the vertical spacing of points indicate the rate of evolution. Theoretical predictions closely match simulation results for both the mean and variance, and the resulting distribution matches the theoretically predicted Gaussian. The variance represents an almost negligible deviation from the mean trajectory.

Fluctuation-Enhancement Domain

Our second ensemble at $x = 2$ begins clearly within the fluctuation-enhancement domain. In the second row of Fig. 2.2, the variance plot (center panel) begins by growing exponentially. Once the population crosses into the dissipation domain, at $x = 1$, at about $t = 1000$, the variance peaks and then dissipates exponentially. This transition appears as an inflection point in the mean path (left panel). Though the variation now represents significant deviations, the distribution still appears Gaussian (right panel).

Strong Fluctuations

In the third row of Fig. 2.2, starting deep within the fluctuation-enhancement domain at $x = 3$, theory and simulation agree initially. Remaining long enough in this domain, fluctuations are eventually on the same order of magnitude as the macroscopic dynamics themselves, and the linear noise approximation underlying the theory begins to break down. The theory and simulation variance diverge, while the mean path is substantially slower than predicted by the canonical equation. In the right panel, the reason for this becomes clear. At this point in the divergence, the distribution is far from Gaussian, rather it has become bimodal, with some replicates having reached the stable strategy at $x = 0$ and others having hardly left the initial state.

The mechanism behind the emergence of bimdal probability distributions for trajectories can be seen in the final row of Fig. 2.2. The theoretical trajectory predicted by Eq. (2.3) is sigmoidal, comprised of slow change at the beginning and end and rapid evolution in the middle. Ecologically, this arises at the beginning from the small population size available to generate mutations, and at the end from the weakening selection gradient. Populations with intermediate trait values hence experience rapid trait evolution, giving rise to this transiently bimodal distribution. The

theory provides insight into this case of strong fluctuations in two ways – first, it is driven by the existence of a fluctuation-enhancement domain, and second, it is anticipated by the theoretical prediction of fluctuations that are on the same order as the macroscopic dynamics. We observe that these macroscopic fluctuations can arise whenever the population remains long enough in the enhancement regime, independent of other model details. The fluctuation equation for the explicit resource competition (chemostat) model shown in Fig. 2.1(b) and described in Appendix A.3 never predicts fluctuations that reach macroscopic values, and simulations of this system (not shown) do not produce bimodal trait distributions.

2.5 Discussion

The metaphor of fitness landscapes has been a powerful tool for understanding evolutionary processes (Abrams, 1993; Dieckmann and Law, 1996; Lande, 1979; Wright, 1932). The concept of a landscape arises naturally from describing the rate of change of the mean trait as being proportional to the evolutionary gradient. We extend this metaphor to understand fluctuations away from this mean, where our analysis of a Markov process representation of evolution in the adaptive dynamics framework brings us back to the notion of a fitness landscape. In this case, it is the curvature of that landscape which informs the dynamics of these fluctuations away from the expected evolutionary path. Our result linking landscape curvature to fluctuation domains is consistent with similar results from quantitative genetics that relate the sign and magnitude of the quadratic selection coefficient in the Price equation to increases or decreases in trait variation (Chevin and Hospital, 2008; Lande and Arnold, 1983).

Though the mathematical analysis of the Markov process is somewhat technical, the result is generally intuitive. Graphs such as Fig. 2.1 provide a visualization of how fluctuations will behave on the landscape, while our fluctuation equation, Eq. (2.4), provides a more explicit description of those fluctuations. This result, like the gradient expression itself, relies on the underlying randomness being small relative to the scale of evolutionary change. We have seen how these approximations can break down in the case of very large fluctuations, resulting in a bimodal distribution of phenotypes. While Eq. (2.4) fails to describe this case accurately, it predicts the breakdown of the approximation by the explosion of fluctuations. Though we have focused on the adaptive dynamics model of evolution, we hope that the approach taken here can be extended

to other landscape representations. Further, we hope that our model will be broadly applicable to assessing the repeatability of evolution in experimental studies of microbes and viruses (Cooper et al., 2003; Wichman et al., 1999) and in ecological studies of rapid phenotypic change caused by biotic interactions (Duffy and Sivars-Becker, 2007) and/or anthropogenic factors (such as overfishing) (Olsen et al., 2005).

In this work, we have compared theoretical predictions to numerical simulations. Even when a closed form expression such as (2.4) exists for the deviations, some of the most dramatic examples in Fig. 2.2 emerge only when the diffusion approximation breaks down and stochastic forces become macroscopic. In the spirit of scientifically reproducible research (Gentleman and Lang, 2007; Schwab et al., 2000; Stodden, 2009), we freely provide all the source code required to replicate the simulations and figures shown in the text. Though the numerical simulations are written in C for computational efficiency, we provide a user interface and documentation by releasing all the code, figures, text, and examples as a software package for the widely used and freely available R statistical computing language.

2.6 Acknowledgements

We thank Sebastian Schreiber for many helpful discussions, and also thank Michael Turelli, Géza Meszéna and an anonymous reviewer for comments. This work is funded in part by the Defense Advanced Research Projects Agency under grant HR0011-05-1-0057. Joshua S. Weitz, Ph.D., holds a Career Award at the Scientific Interface from the Burroughs Wellcome Fund. Carl Boettiger is supported by a Computational Sciences Graduate Fellowship from the Department of Energy under grant number DE-FG02-97ER25308.

Chapter 3

Is your phylogeny informative? Measuring the power of comparative methods

3.1 Introduction

3.1.1 Are phylogenies informative?

Since their introduction into the comparative method over two and a half decades ago, phylogenetic methods have become increasingly common and increasingly complex. Despite this, concern persists about the ubiquitous use of these approaches (Losos, 2011; Price, 1997). From a statistical perspective these concerns can be divided into two categories: (a) Do we have appropriate models that reflect the biological reality of evolution and represent meaningful hypotheses, and (b) Do we have adequate data to fit these models and to choose between them? The models have been greatly improved since their introduction, and can now account for stabilizing selection (Hansen and Martins, 1996), multiple optima (Butler and King, 2004), and differing rates of evolution across taxa (O'Meara et al., 2006) or through time (Blomberg et al., 2003; Pagel, 1999); but little attention has been given to this second concern about data adequacy. In this paper, we highlight the importance of these concerns, and illustrate a method for addressing them.

It can be difficult to accurately interpret the results of comparative methods without quantification of uncertainty, model fit, or power. Most current comparative methods do not attempt to quantify this uncertainty; consequently it can be easy for inadequate power to lead to false biological conclusions. For instance, below we illustrate how estimates of phylogenetic signal (Gittleman and Kot, 1990) using the λ statistic (Pagel, 1999; Revell, 2010) can reach opposite conclusions (from no signal $\lambda = 0$ to approximately Brownian, $\lambda \approx 1$) when applied to different simulated realizations of the same process. We also show that model selection by information criteria can

prefer over-parameterized models by a wide margin. On the other hand, when a simpler model is chosen, it may be difficult to determine whether this merely reflects a lack of power. In both cases, the results can be correctly interpreted by estimating the uncertainty in parameter estimates and the statistical power (ability to distinguish between models) of the model selection procedure.

Here we provide one solution to these problems using a parametric bootstrapping approach which easily fits within the framework used by many comparative methods approaches. As comparative methods rely on explicit models, this is easily implemented by simulating under the specified models. For the problem of uncertainty in parameter estimation, the bootstrap is a well-established and straightforward method (Efron, 1987). A few areas of comparative methods have used a similar approach: for instance, phylogenetic ANOVA (Garland et al., 1993) calculates p values of the test statistic by simulation under Brownian motion. A similar approach was later introduced in the `Brownie` software (O'Meara et al., 2006) to generate the null distribution of likelihood ratios under Brownian motion, and applied in Revell and Harmon (2008), which showed the distribution can deviate substantially from χ^2, and a similar approach is applied in Revell and Collar (2009). Unfortunately, such approaches have never become a common in comparative analyses. Here, we describe a method due to Cox (1962) and used by others (Goldman, 1993; Huelsenbeck and Bull, 1996), that can be used in place of information criteria for model choice, allowing estimation of power and false positive rates, and can provide good estimates of confidence intervals on model parameter estimates. While simulations are often performed when a new method is first presented, this practice rarely becomes routine. By providing a simple R package ("`pmc`", phylogenetic Monte Carlo) for the method outlined, we hope Monte Carlo based model choice and estimates of power become common in comparative methods.

To set the stage, we will review common phylogenetic models and describe the Monte Carlo approach to model choice. We then present the results of our method applied to example data and discuss its consequences.

3.1.2 Common phylogenetic models

Comparative phylogenetics of continuous traits commonly uses a collection of simple stochastic models of evolution; we briefly review these here to fix ideas and notation. All models we consider take as given an ultrametric phylogenetic tree whose branch lengths represent evolutionary

divergence times; extant taxa are represented by the tips of the tree. We will assume that the tree is known without error. For convenience we will in all examples choose time units so that the tree height is one unit. For each extant taxon we have a trait value (say, the species mean) for some continuous trait such as body size, and represent the collection of trait values across extant taxa as the vector X. The joint distribution of these trait values is given by specifying the ancestral trait value X_0 at the root of the tree, by describing the stochastic process of trait evolution along branches of the tree, and assuming that evolution on separate branches proceeds independently.

Let Y_t be the value of our trait at time t along some branch. The simplest and most common model for the evolution of the trait Y_t is a scaled Brownian motion (Felsenstein, 1985), which can be represented by the stochastic differential equation:

$$dY_t = \sigma dB_t, \tag{3.1}$$

in which B_t is standard Brownian motion, and σ is the rate parameter. Under this model, the trait value evolves as a random walk starting from the ancestral state X_0, and upon reaching each node in the phylogeny, the process bifurcates into two independent Brownian walks. This Brownian motion (BM) model is completely defined given a phylogeny and two parameters: the initial state X_0 and the parameter σ, which is usually interpreted as the rate of increase in variance.

A closely related model introduced in a comparative phylogenetics context by Hansen (1997) is the Ornstein-Uhlenbeck (OU) model, for which trait evolution Y_t along each branch follows the Ornstein-Uhlenbeck process, which is described by the following stochastic differential equation

$$dY_t = -\alpha(Y_t - \theta)dt + \sigma dB_t. \tag{3.2}$$

Here Brownian motion is modified to have a central tendency towards a preferred trait value θ, usually interpreted as a optimum trait value under stabilizing selection. The strength of stabilizing selection increases linearly with distance from the optimum θ, controlled by the parameter α. When $\alpha = 0$, this model reduces to the BM model. Both evolutionary models are described in more detail elsewhere, *e.g.* Butler and King (2004).

Many variations of these basic models are also common – for instance, it may be desirable to allow the diversification rate parameter σ in the BM model to vary in some way over time (Blomberg et al., 2003; Harmon et al., 2010; Pagel, 1999) or across the phylogeny (O'Meara

et al., 2006). Similar extensions can be applied to the OU model – we will later consider the example of Butler and King (2004) which allows the optimum trait value θ to differ among different branches or clades. One can illustrate which branches of a phylogeny are permitted to have independently estimated values of the optimum trait by "painting" them different colors indicating where the model is allowed to change (Butler and King, 2004).

Another commonly used variation is Pagel's λ (Freckleton et al., 2002; Pagel, 1994), which was introduced as a test of phylogenetic signal – the degree to which correlations in traits reflect patterns of shared ancestry. The model underlying Pagel's λ is the simple Brownian motion along the phylogeny as above, except that the phylogeny is modified by shortening all internal edges by a multiplicative factor of λ, which reduces the resulting correlations between any pair of taxa by a factor λ, and adjusting terminal edges so the tree remains ultrametric. The parameter λ can then be estimated by maximum likelihood. Estimates near unity are taken to indicate high phylogenetic signal, while estimates near zero indicate that other processes such as natural selection have erased this "signal" of common descent.

3.2 Methods

3.2.1 Uncertainty in parameter estimates

To demonstrate the perils of inadequate data without estimates of uncertainty, we open with an example of a phylogenetic test using Pagel's λ statistic that also serves to illustrate the estimation of uncertainty in parameter estimates (e.g. confidence intervals). We illustrate that on a small tree, estimates of λ can differ greatly from the parameter used in the simulations. In practice, the danger is that an estimate of λ near zero may arise by chance because the tree is too small, not because the phylogeny is unimportant to the evolution of the trait. Larger phylogenies, on the other hand, generally allow greater accuracy.

In Figure 3.1(a) we show the empirical distribution of the maximum likelihood estimate of λ for 1000 data sets simulated under a model with moderate phylogenetic signal, $\lambda = 0.6$, and $\sigma = .03$. The estimates were performed on the *Geospiza* data using functions available in pmc in conjunction with the R package `geiger` (Harmon et al., 2008). The phylogeny, data, and script for the analysis are included in pmc. We see that for datasets coming from this small phylogeny,

the maximum likelihood statistic $\hat{\lambda}$ is a poor estimator for the true value of λ. The most common estimate is $\hat{\lambda} = 0$, which is usually interpreted to mean that the phylogeny contains little information. The next most common estimate is $\hat{\lambda} = 1$. Note that this is the upper bound set on λ by the fitting algorithm. It is clear that we must thus be cautious what we conclude based on values of λ estimated on this phylogeny.

Repeating this exercise on successively larger data sets makes it clear that this is a problem of insufficient data. With a simulated tree of 281 tips, the estimated values are closely centered around the true value, as shown in Figure 3.1(b).

The amount of data required to be informative will depend not only on the size and topology of the tree but also on the question being asked. For instance, it may be impossible to distinguish moderately different values of λ, which is very difficult to estimate accurately. However, it may be feasible to estimate other parameters on smaller phylogenies than this 281 taxa example. For instance, using the same 13 taxa *Geospiza* phylogeny, we can estimate the diversification rate parameter σ much more precisely, as shown in Figure 3.1(c).

A natural way to report the uncertainty associated with a parameter estimate is construct a confidence interval, which is rarely performed in the literature but can easily be done by parametric bootstrapping. Given the parameter estimate, a confidence interval can be estimated by simulating a large number of datasets using the known phylogeny and the estimated parameter, and re-estimating the parameter on each simulated dataset (*e.g.* see Diciccio and Efron, 1996). The distribution of the re-estimated parameters is used to construct the confidence interval; e.g. the 2.5 to the 97.5 percentile gives a 95% confidence interval. For the example shown in Figure 3.1(b), our estimate of λ on the Yule tree with 281 tips, the 95% confidence interval would be $(0.45, 0.69)$. For the parameter σ, Figure 3.1(c) shows that the confidence interval is $(0.007, 0.059)$. Given the noisy nature of parameters estimated from phylogenies we recommend that confidence interval should routinely be reported, and to facilitate this, have implemented this as `pmc::confidenceIntervals.pow`. Confidence intervals could also be estimated from the curvature of the likelihood surface, but these can be unreliable and problematic to compute.

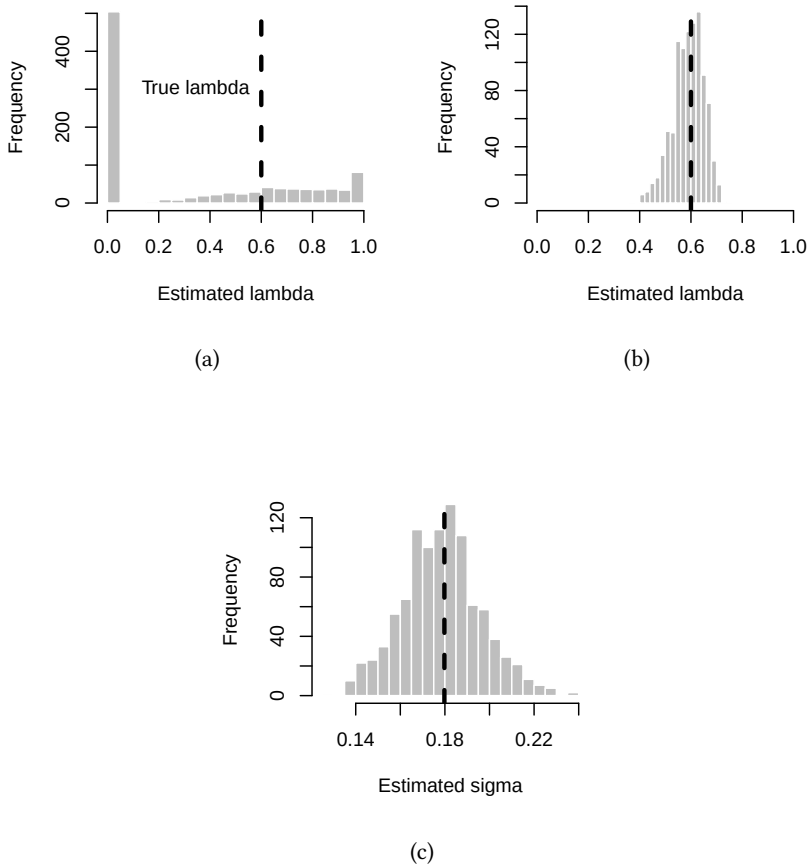


Figure 3.1: (a) Empirical distribution of maximum likelihood estimates of λ for 1000 sets of trait values simulated on the *Geospiza* phylogeny with 13 taxa transformed with $\lambda = 0.6$, using $\sigma = 0.18$. Most such datasets yielded a maximum likelihood estimate of 0; the mean estimate is $\bar{\lambda} = 0.35$. (b) As above, but simulating trait values on a much larger phylogenetic tree (a single, simulated Yule tree with 281 tips), again transformed with $\lambda = 0.6$. The estimated values now cluster around the true value, and have mean $\bar{\lambda} = 0.59$. (c) The data can be more informative about some parameters than others: shown is the empirical distribution of maximum likelihood estimates of the diversification rate σ for the same simulations as in (a). The mean of the distribution is $\bar{\sigma} = 0.18$, matching the value used in the simulations.

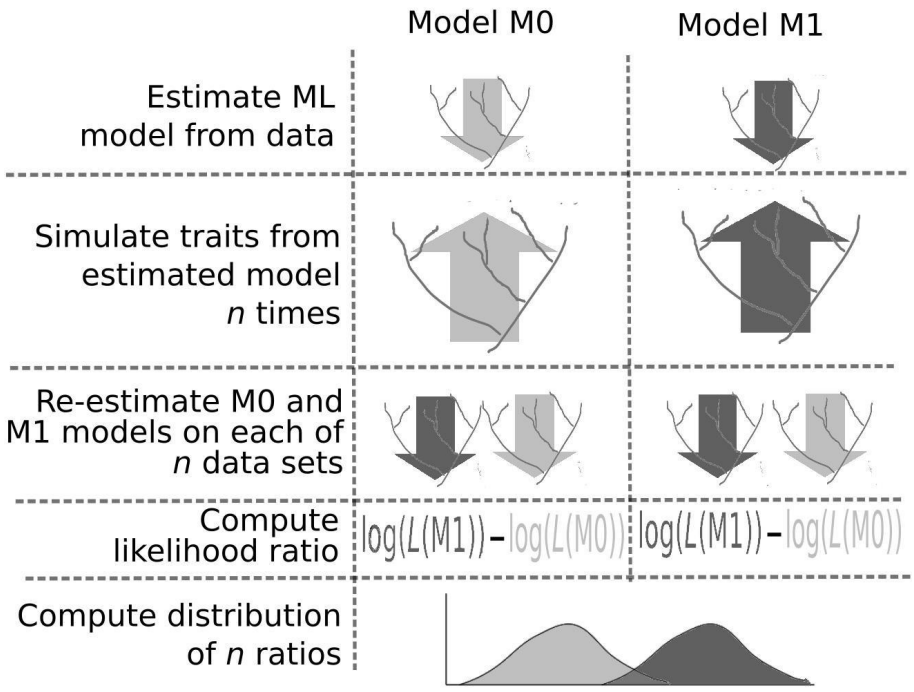


Figure 3.2: Conceptual diagram of the Monte Carlo method for model choice. First, parameters for both models are estimated from the original data. Then, n simulated datasets are created from each model at these parameters, and on each dataset, the parameters for both models are re-estimated and the likelihood ratio statistic is computed. The collection of likelihood ratio statistics generates the corresponding distribution. This involves a process of $4n$ fits by maximum likelihood, instead of only 2 fits required for information criteria.

3.2.2 The Monte Carlo approach

Knowing when the data are sufficiently informative is also crucial when comparing different models. To do this, we introduce a Monte Carlo-based method, described below. Suppose we have a dataset X for which we wish to determine which of two models, model 0 or model 1, is the better description. Each model is specified by a vector of parameters, Θ_0 and Θ_1 respectively, which can assume values in the spaces Ω_0 and Ω_1 respectively. We tend to imagine that model 1 is the more complex model, though in general they need not be nested. Let $\mathcal{L}_0$ be the likelihood function for model 0, let $\hat{\Theta}_0 = \text{argmax}_{\Theta_0 \in \Omega_0}(\mathcal{L}_0(\Theta_0|X))$ be the maximum likelihood estimator for Θ_0 given X, and let $L_0 = \mathcal{L}_0(\hat{\Theta}_0|X)$; and define $\mathcal{L}_1$, $\hat{\Theta}_1$, L_1 similarly for model 1.

The statistic we will use is δ, defined to be twice the difference in log likelihood of observing

the data under the two MLE models,

$$\delta = -2\left(\log L_0 - \log L_1\right) \tag{3.3}$$

For simplicity we will refer to this as the likelihood ratio. Larger values of δ indicate more support for model 1 relative to model 0. It is natural to use the difference in log-likelihoods as a statistic to choose between the models (Neyman and Pearson, 1933), as do information criteria such as AIC. To do this we need to know, for instance, how large should δ be before we decide that model 1 is much closer to the truth than is model 0. Many common methods proceed to approximate the distribution of δ asymptotically. For instance, if the models are nested in a manner that does not force a parameter to its boundary value, this statistic has asymptotically the χ^2 distribution with degrees of freedom equal to the difference in the number of parameters. These asymptotic approximations for phylogenetic comparative analyses are often inadequate for phylogenetic comparisons. Instead, we can estimate the distribution of δ under either model directly from Monte Carlo simulation. This method seems to have been first suggested in the statistical literature by Cox (1961, 1962) and applied to mixture models by McLachlan (1987). It has been previously applied to the case of estimating phylogenies from sequence data by Huelsenbeck and Bull (1996); see also Goldman (1993).

To estimate the distribution of δ under model 0 and the estimated parameters ($\hat{\Theta}_0$), we proceed as follows. First simulate n datasets $X^1, \ldots, X^n$ independently from model 0 with parameters $\hat{\Theta}_0$. For each $1 \leq k \leq n$, let $\hat{\Theta}_0^k$ be the maximum likelihood estimator of the parameters Θ_0 of model 0 for dataset X^k, and likewise let $\hat{\Theta}_1^k$ be the MLE under model 1. Then we compute the likelihood ratio statistic for the k^{th} data set, $\delta_k = -2(\log \mathcal{L}_0(X^k|\hat{\Theta}_0^k) - \log \mathcal{L}_1(X^k|\hat{\Theta}_1^k))$, and examine the empirical distribution of $\delta_1, \ldots, \delta_n$. We can also estimate the distribution of δ under model 1 in the same way.

There are two things to note about this procedure. First, the Monte Carlo datasets are simulated at the maximum likelihood parameters $\hat{\Theta}_0$ and $\hat{\Theta}_1$, which are in turn estimated from the same dataset X. So if, for instance, the models are nested and the simpler is correct, then one would expect model 0 at $\hat{\Theta}_0$ to be quite similar to model 1 at $\hat{\Theta}_1$. Secondly, it is necessary when computing the Monte Carlo values δ_k to *re-estimate* the maximum likelihood parameters, rather than using the original parameters $\hat{\Theta}_0$ and $\hat{\Theta}_1$ – simply computing $\delta_k = -2(\log \mathcal{L}_0(X^k|\hat{\Theta}_0) -$

$\log \mathcal{L}_0(X^k|\hat{\Theta}_1))$ would lead to a much less powerful test (Hall and Wilson, 1991). The reason for this is somewhat subtle (see McLachlan, 1987), and is related to the first point. For further suggestions on obtaining a reliable estimate of the distributions, see Efron (1987) and Diciccio and Efron (1996).

3.2.3 Model selection

If we suppose model 0 is "simpler" than model 1, it is natural to regard model 0 as the "null" and test the hypothesis that the data came from model 0. To do this, we would compare where the observed difference in log likelihoods δ for the original data falls relative to the distribution under model 0. The proportion of the simulated values larger than δ provides an approximation to the p–value for the test, the probability that a difference at least as large would be seen under model 0. (Because the datasets X^k are all simulated at the estimated parameters $\hat{\Theta}_0$ this strictly applies only for the hypothesis test between the maximum-likelihood estimated models, and is not the p-value when comparing the composite hypothesis represented by the original model with unspecified parameters (see McLachlan, 1987).) If we choose, say, δ_* so that 95% of the simulated values $\delta_1, \ldots, \delta_n$ fall below δ_*, and choose to reject model 0 if $\delta > \delta_*$, then we have a test of the null hypothesis that model 0 is true, with a false positive probability of approximately .05 under model 0. If we then want to know about the statistical power of this test – the probability that we correctly reject model 0 when the data came from model 1 – we would turn to the distribution of δ under model 1. If we have chosen δ_* as above, then the amount of this distribution to the left of δ_* approximates the probability of rejecting model 0 when the data are produced by model 1 – the power of the test.

The procedure we have described is motivated by classical hypothesis testing, but is only one way to use the information provided by the empirical distributions of δ.

3.3 An example using *Anolis* data

3.3.1 The anoles data

To illustrate the concerns about phylogenetic information in comparative methods, we shall revisit a classic data set of mean body size for 23 species of *Anolis* lizards from the Lesser Antilles, which has been used to introduce other comparative phylogenetic approaches (e.g. Butler

and King, 2004, familiar to many who have used the `ouch` package). The phylogeny reconstruction used here (Losos, 1990) is based upon morphological (Lazell, 1972) and protein-electrophretic (Gorman and Kim, 1976) techniques rather than the more recent phylogenies based on mitochondrial sequences (Schneider et al., 2001; Stenson et al., 2004), which have substantial differences. As our purpose is simply to illustrate the approach, we continue to use older tree familiar to the readers of earlier work (Butler and King, 2004; Losos, 1990).

Identification of branches or clades of a phylogenetic tree that show significantly different evolutionary patterns can illuminate key elements about the origin and maintenance of biodiversity. Butler and King (2004) demonstrated how the existence of different adaptive optima in character traits on different parts of a phylogenetic tree could be detected. They assumed that evolution of the trait along each branch followed the Ornstein-Uhlenbeck model, but that different branches could have different optima (the parameter θ). The branches that must share a common value of θ are represented by a "painting" of the tree; three possibilities for the *Anolis* tree that we later investigate are shown in Figure 3.3. Any branch of a given color must have the same optimum trait value, each of which is estimated by the fitting algorithm. The remaining parameters α and σ are shared across the entire tree.

To confirm that the proposed pattern of heterogeneity (the painting) is justified by the data, it is necessary to compare between possible paintings and possible assignments of model parameters to each part of the painting. We seek to identify (a) which model best describes the data and (b) whether we have sufficient data to resolve that difference?

3.3.2 Models for the *Anolis* phylogeny

To illustrate the approach we consider a total of five models for the *Anolis* data set. The first two models apply the same model of evolution to the entire tree (i.e. a one-color painting) – either Brownian motion (BM) (Edwards and Cavalli-Sforza, 1964; Felsenstein, 1985), with two parameters; or the Ornstein-Uhlenbeck process (OU.1) (Felsenstein, 1985; Hansen, 1997), with three.

The remaining three models extend these simple cases by introducing heterogeneity in the model, allowing the trait optimum to vary across the tree as indicated in Fig 3.3. The OU.3 model of Figure 3.3(a) has three optima, and corresponds to the *character displacement hypothesis* (Losos,

Figure 3.3: "Paintings" of the *Anolis* phylogeny specifying which branches are assumed to have a common value of the trait optimum θ for three different models: (a) **OU.3**, with three possible optima from Butler and King (2004); (b) **OU.4**, with four possible optima; and (c) **OU.15**, with a unique optimum for each branch in the upper two clades. The remaining models, BM and OU.1, fit the same parameters across the entire phylogeny and so are not shown. Estimated model parameters for each are shown in Table 3.3.2.

Model	$\log(L)$	MLE Parameters	95% CI
BM	17.33	$X_0 = 2.9$, $\sigma^2 = 0.043$	(0.14, 0.26) (2.74, 3.16)
OU.1	15.69	$\theta = 3.0$, $\sigma^2 = 0.048$, $\alpha = 0.19$	(2.36, 3.56) (0.028, .13) (.24, 4.41)
OU.3	24.82	$\theta = \{3.36, 3.04, 2.56\}$, $\sigma^2 = 0.05$, $\alpha = 2.61$	$\{(3.20, 3.47), (2.94, 3.11), (2.41, 2.76)\}$ (0.025, 0.19) (1.77, 17.98)
OU.4	26.69	$\theta = \{2.97, 3.31, 3.12, 2.63\}$, $\sigma^2 = 0.06$, $\alpha = 4.68$	$\{(2.87, 3.05), (3.22, 3.38), (3.02, 3.21), (2.53, 2.74)\}$ (0.031, 3.39) (3.34, 384.16)
OU.15	44.17	$\theta = \{2.91, 2.99, 2.98, 3.04,$ $3.11, 3.35, 2.97, 3.08,$ $3.19, 3.15, 3.17, 2.81,$ $3.30, 3.05, 2.62\}$, $\sigma^2 = 0.06$, $\alpha = 24.3$	{ (2.84,2.98), (2.91,3.22), (2.81,3.46), (2.85,3.57) {(3.04,3.20), (3.28,3.53), (2.80,3.42), (2.30 , 352), (2.34, 352), (2.94, 3.84) (2.94,3.85), (2.66, 1.5e6), (3.27,3.38), (2.98, 3.12), (2.55, 2.67) } (0.0036, 0.44) (7.29, 322.92)

Table 3.1: Parameter values estimated for the *Anolis* dataset by maximum likelihood for models with varying number and location of optima (also see Figure 3.3), used in the comparisons in Figure 3.4. These parameter values were used to produce the simulated datasets in Figures 3.4 and 3.5. The values of θ are in order of first appearance, left-to-right, in Figure 3.3. The corresponding 95% confidence intervals calculated from the 2000 replicates are also shown. Note that the optima θ of OU.15 just represent a finer partition of the optima in OU.3.

1990), which predicts three different optimum body sizes – an intermediate optimum on islands having only one species, and a larger and a smaller optimum for islands with two species of lizards. The island size determines to which optimum the tips or extant species are assigned, while the ancestral states are constructed by parsimony as per Butler and King (2004). To these three models (BM, OU.1, and OU.3) analyzed by Butler and King (2004) we add two more to illustrate possible outcomes. OU.4, Figure 3.3(b) hypothesizes four optima corresponding to four separate clades. The fifth model OU.15 is intentionally arbitrary and overly complex, assigning a unique optimum to each species in the top two clades for a total of 15 optima. We apply these methods to determine which model best fits the data and whether the data are sufficiently informative to distinguish between them.

3.4 Results

We illustrate several points with four different comparisons, depicted in Figure 3.4. In each case, the distribution of δ under each of the two models is shown as the dark-shaded and the light-

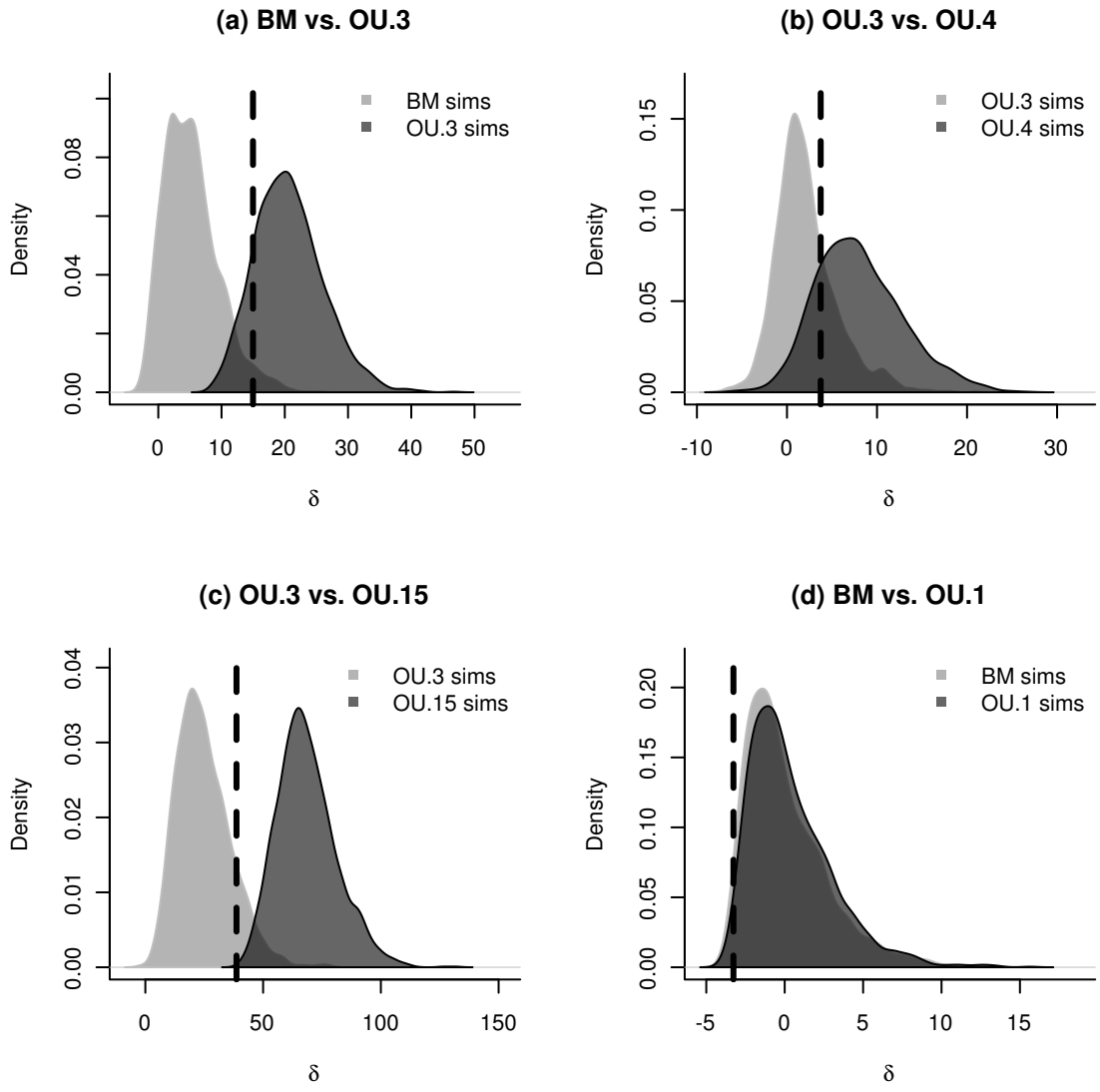


Figure 3.4: Distributions of the likelihood ratio statistic of equation (3.3) for four different model comparisons. In each case the lighter distribution shows the distribution of δ values obtained by bootstrapping under the simpler of the two models, while the darker distribution shows the distribution under the more complicated of the two models. 2000 replicates are used for each distribution. The dashed vertical line indicates the observed value of δ when the models are fit to the *Anolis* dataset. **(a) BM versus OU.3:** the observed likelihood ratio is much more likely under OU.3. **(b) OU.3 versus OU.4:** here the distributions overlap more, indicating that the data are less informative about this more subtle comparison. **(c) OU.3 versus OU.15:** these distributions have little overlap and the observed ratio falls clearly in the range of the simpler model. We can conclude that this support for OU.3 is not merely due to lack of power. **(d) BM versus OU.1:** the data contain almost no information to distinguish between these two models at the estimated (small) level of selection α.

shaded curves, and the observed value of δ is marked by the dashed vertical line. We also construct confidence intervals for the parameters in the same way as we did for the λ estimates, shown in Table 3.3.2. The maximum likelihood parameter values for each model, estimated from the anoles data, are given in Table 3.3.2, and are computed from the original body size data described in Section 3.3 using the `ouch` package of Butler and King (2004) together with tools from our `pmc` package. Scripts to perform all analyses shown here are included in the `pmc` package. We will be able to determine not only which model is preferred, but also the certainty of the model choice.

3.4.1 Quantification of model choice

For a first example, comparing BM to OU.3 (Figure 3.4(a)), we see that only 2.5% of simulations under BM have a likelihood ratio δ more extreme than the observed ratio of 15 units seen in the real data (*i.e.* $p = 0.025$). The degree of overlap in the distributions reflects the extent to which the phylogeny is useful to discriminate between the two hypotheses at these parameter values; in this case the test that rejects the BM model with 5% false positive rate has a power of 93.6%. Thus we have a direct estimate of both which model is a better fit and of our power to choose between the models. Note that in our framework we are free to choose the tradeoff between the false positive and false negative rates. For instance, a 5% cutoff may be too stringent if it is unnatural to treat either model as a null.

3.4.2 Information criteria often fail to choose the correct model

For a second example, we compare OU.3 to the over-parameterized model OU.15 (Figure 3.4(c)). Table 3.3.2 shows that the maximum likelihood optimum trait values θ and rate of divergence σ are similar for the two models, but that the strength of selection α is much larger for OU.15. From the table of estimated values and confidence intervals, it is clear that OU.15 has simply divided up each of these broader peaks into finer optima clustered around the original estimates. The higher value of α in the OU.15 model indicates narrow peaks of strong selection that result in the much higher likelihood. Despite this, our method will not select OU.15, since the observed likelihood ratio δ falls below value of δ seen in 18.8% of simulations under OU.3. Furthermore, this is a powerful test: 98.8% of simulations under OU.15 produce a δ that falls beyond the 95% quantile of the OU.3 distribution.

We can compare this method to information criteria (e.g. AIC, BIC), which are the standard tools for model comparison in comparative methods of continuous traits (Butler and King, 2004). Because we have generated simulated datasets under both hypothesized models, it is straightforward to estimate how often these datasets are misclassified by various information criteria. The same distributions from Figure 3.4 are shown with the cutoff given by AIC for choosing the more complex model in Figure 3.5. We see that AIC would assign nearly half (47.7%) of the simulations done under OU.3 incorrectly to the OU.15 model, and that the observed data would also be assigned to OU.15. If we evaluate the performance of AIC when comparing two reasonable models, OU.3 and OU.4, information criteria still prefer the more complicated model (AIC(OU.3) =-39.6; AIC(OU.4) =-41.3, and BIC(OU.3) =-33.9; BIC(OU.4) = -34.6), but here we know this may be illusory, since Figure 3.5 shows that AIC falsely assigns 44% of simulations produced under OU.3 as coming from OU.4. Sample-size correction of AIC (AICc, not shown) can be similarly misleading. See the online appendix for example code to reproduce this figure under each of the different information criteria.

3.4.3 Applied to non-nested models

The next example compares OU.3 to OU.4, where as mentioned above, the degree of overlap between the distributions of δ under the two models seen in Figure 3.4(b) shows that we have relatively little power to distinguish between the two. Note that since the painting defining the OU.4 model is not a refinement of the painting defining the OU.3 model, the two models are not nested. The Monte Carlo approach applies equally well to non-nested models, unlike the asymptotic derivations commonly used to justify information criteria. We furthermore do not have to determine the difference in number of parameters, as is required by AIC, which in some situations is not obvious.

3.4.4 When the data are insufficient to distinguish between models

The fourth comparison is between the simplest models, BM and OU.1. Figure 3.4(d) shows that there is essentially no information to adequately distinguish between them. This should not be taken as evidence that BM is a better fit, but rather that given the small selection parameter estimated from the anoles data, we have low power to distinguish OU.1 from BM on this phylogeny.

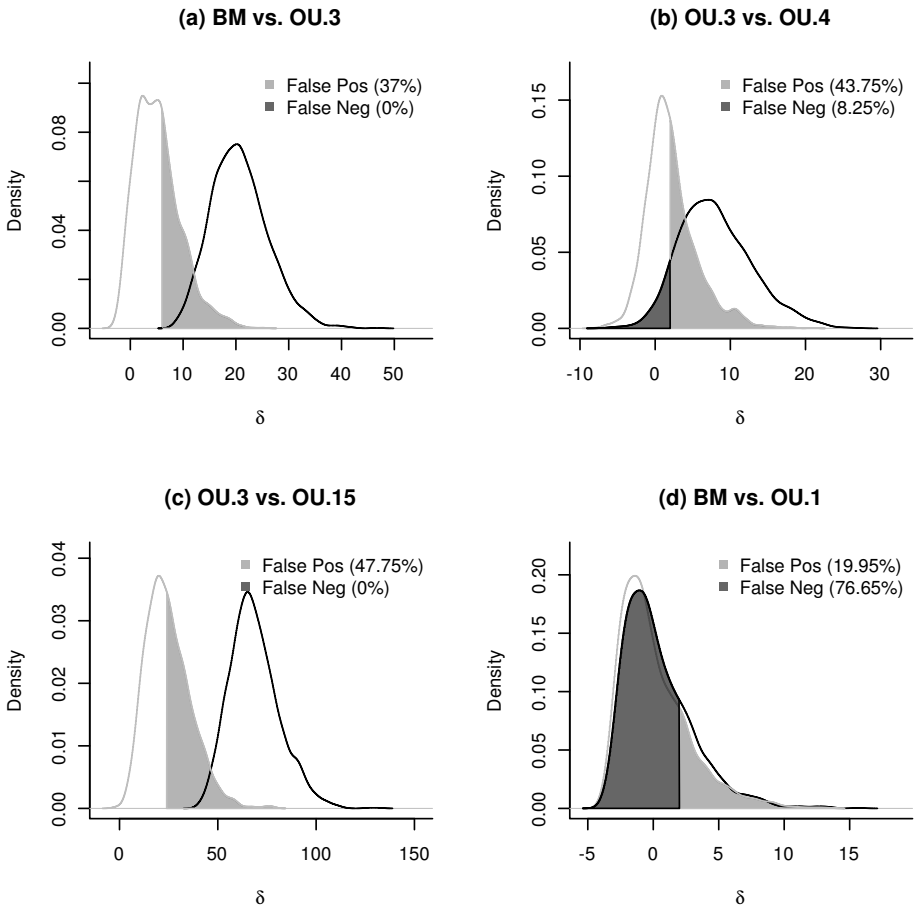
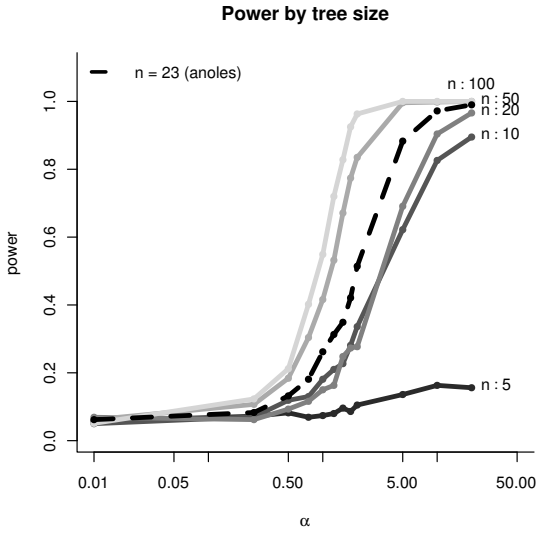


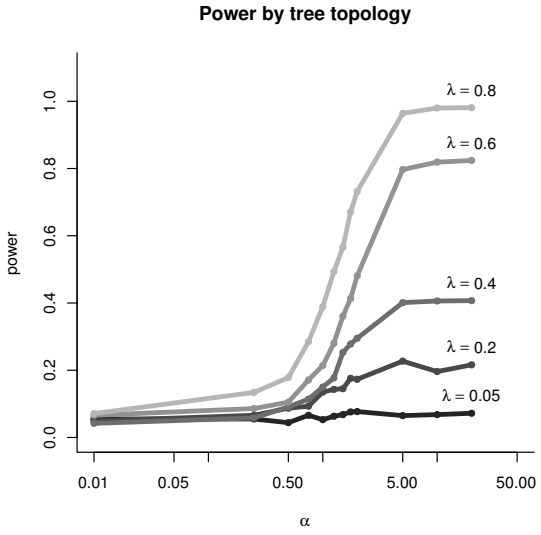
Figure 3.5: Error rates for model choice by AIC based on simulation. Shown are the same distributions of the likelihood ratio statistic δ as in Figure 3.4. Also shown is the probability that AIC selects the more complicated model when the simpler is true ("False Positives", light shading); and the probability that AIC selects the more simpler model when the more complicated is true ("False Negatives" error, dark shading).

	AIC errors (%)		BIC errors (%)		AICc errors (%)	
Comparison	Type I	Type II	Type I	Type II	Type I	Type II
BM vs. OU.3	37.00	0.00	15.90	0.45	13.05	1.05
OU.3 vs. OU.4	43.75	8.25	29.35	14.5	2.30	73.55
OU.3 vs. OU.15	47.75	0.00	13.65	0.00	0.00	100
BM vs. OU.1	19.95	76.65	11.95	86.05	8.90	89.7

Table 3.2: A comparison of error rates across various information criteria. In the comparisons that have high overlap between the distributions (BM vs. OU.1, OU.3 vs. OU.4, Figure 3.4), at least one of the rates will be high for any method. In cases with adequate power (OU.3 vs. OU.15, BM vs. OU.3), information criteria can still have high error rates. The methods we describe allow the researcher not only to estimate these rates, but to specify a tradeoff between the error types.



(a)



(b)

Figure 3.6: Power to identify stabilizing selection α at a given strength on different phylogenies. Shown is the empirical probability that data generated with a given α on a given tree will favor a OU.1 model over BM, based on a cutoff of the likelihood ratio statistic δ chosen to have a false positive probability of 5%, based on 1000 simulations with $\sigma = 1$. (a) Increasing the number of taxa in the tree (simulated under a birth-death model) increases the power to detect a given strength of selection. (b) Fixing the number of taxa to 50, we distort the shape of the simulated tree to one in which most of the branching events occur farther and farther in the past using Pagel's λ transformation. On trees that are highly distorted (smaller λ) we have substantially less power to detect any given strength of selection.

The strength of selection in the OU model is represented by α in equation (3.2), and is measured in units of inverse time since the common ancestor (when the tree height has been normalized to unity). Hence the maximum likelihood estimate for this model with a value of $\alpha = 0.2$ means that correlations between traits that diverged at that common ancestor will have decayed to only $e^{-0.2} = .81$ of what is expected under BM. The chance we could detect this level of selection at 95% false positive rate (*i.e.* the power) was only 7%.

What is the weakest level of stabilizing selection on a trait we could reliably detect using this *Anolis* phylogeny? To answer this, we repeat the analysis on data simulated using OU.1 models with progressively larger α and estimate the power for each. The results are shown as the dashed curve in Figure 3.6(a). Power increases with increasing strength of selection α, which we can visualize by imagining the darker distribution of Figure 3.4(d) moving farther to the right. In the next section, we use this approach of power simulation to understand what aspects of phylogeny (*i.e.* shape and size) influence its power to detect a given strength of selection.

3.5 Understanding the role of phylogeny shape and size on estimates of selection

The shape and size of the phylogeny is key to understanding how much information about evolutionary processes it is possible to extract from characters of taxa at the leaves of the tree. As an application of the method of obtaining a power curve for the strength of selection described in section 3.4.4 we can compare the power curves for trees of different shapes. As before, we are comparing the single–optimum Ornstein-Uhlenbeck (OU.1) model to the Brownian motion (BM) model without selection, and computing the power to correctly choose the OU.1 model at different values of α, if we choose models based on the 95% quantile of δ under the BM model. Figure 3.6(a) compares trees simulated from a pure-birth process with increasing number of taxa, scaled to unit height.

Number of taxa is not all that matters; Figure 3.6(b) considers a single (simulated pure-birth) tree of 50 taxa rescaled so that successively more of the time occurs in the tips and so that the speciation events occur more distantly in the past. The farther in the past diversification has occurred, the less informative the tree. This is the rescaling performed by the λ transformation

described in section 3.1.2. Covariances introduced by different amounts of shared evolution are crucial for distinguishing slower character diversification rates σ from stronger selection α. We see that as the branching events occur earlier (smaller λ transformations), these correlations are harder to detect, so the phylogeny becomes less informative.

We note that many simulation studies (*e.g.* Freckleton et al., 2002) are conducted using trees generated by a pure-birth (Yule) process, which generates phylogenies with more very shallow nodes than are generally seen in practice. Perhaps counter-intuitively, the presence of these highly-correlated points makes the phylogenies particularly informative relative to branching patterns resulting from any density-dependent or niche-filling models. Early bursts of speciation such as adaptive radiations will tend to generate phylogenies that are less informative of parameters such as the strength of selection, α. These examples show that the ability to distinguish between models can depend strongly on the value of the parameters, the number of taxa, and the shape of the tree. Rather than attempt to draw rules of thumb from such exercises, we suggest that it is best to perform a power analysis that is specific to the phylogeny and estimated model parameters being compared.

3.6 Discussion

We have introduced a general, simulation-based method to choose between models of character evolution and quantify the power of such choices on a particular phylogeny. While the methodological underpinnings of this approach are not new, the field of comparative methods continues to rely almost universally on information criteria. We have illustrated that the performance of these methods can be remarkably poor, particularly with insufficiently large or structured phylogenies. The results can provide a clear indication of when a phylogenetic tree is either too small or too unstructured to resolve differences in the proposed models.

Though our analysis selects the same model (OU.3) for the anoles dataset as does Butler and King (2004), we have shown that existing approaches such as AIC Butler and King (as used in 2004) would have preferred either of our more complex models (OU.4 and OU.15). Our models are chosen to illustrate various possible outcomes: not only can we choose either the simpler or the more complex model, but through power simulations we can determine if choice of simpler model is due to poor fit of the data by the complex model, or simply due to insufficient data.

Since their introduction in a modeling framework in Felsenstein (1985), phylogenetic comparative methods have continued to increase in complexity. We provide a simple method to reliably indicate if the informativeness of the datasets is keeping pace with this increase in complex models. Through these methods, we can know when the comparison we are making is too fine for the resolution of the data, as in the BM vs. OU.1 comparison Figure 3.4(d), and when increased model complexity is clearly unsupported, as in OU.3 vs. OU.15 comparison, Figure 3.4(c). Model choice plays a similar role in many other models in comparative phylogenetics, such as deciding between the various tree transforms such as λ, δ, γ, or ACDC, which can benefit from the same attention to whether the data are adequately informative.

As shown in section 3.5, the power to distinguish between two models can depend strongly on the parameter values, which can be a subtle point and pose difficulties for interpretation. For instance, if a power analysis is done by simulating under a certain set of parameter values, but the test is applied to datasets consistent with very different parameter values (a situation found in Harmon et al., 2010), then it remains a possibility that failure to find evidence for more complex models results from a lack of power.

Our results cast doubt on the use of AIC for phylogenetic model selection; however, mathematically our methods are very similar to information criteria. When applied to a pair of models, the various information criteria (AIC, BIC, AICc, etc.) give a cutoff for the likelihood ratio statistic δ that determines which model to choose. Our method can provide such a cutoff as well, but also allows choice of such a cutoff based on the power–false positive tradeoff. One use for our methods would be to simply quantify the resolving power of an AIC-based model choice. A drawback of our method over AIC is that it does not compare simultaneously many models, instead relying on a collection of pairwise comparisons. This is a disadvantage particularly when AIC is applied to find the best model out of many, and the goal is to find a parsimonious predictive model of more complex reality. However, it seems to us that comparative methods are usually concerned with rigorously distinguishing between alternative models, and so the goal of model choice is to describe underlying process rather than to provide plausible predictions. See Burnham and Anderson (2002) for discussion of a philosophy of model selection using AIC in a predictive framework.

The procedure we describe is grounded in a familiar maximum-likelihood framework of

model comparison, and the dependence on certain estimated parameter values for each model poses one of the difficulties for interpretation. A Bayesian approach might compare models using Bayes factors, thus integrating over all parameter values for each model, and could be implemented using a reversible jump Markov chain Monte Carlo scheme (Green, 1995). Note, however, that the restriction to fixed parameter values is not necessarily a limitation, as it allows us to perform such analyses as identifying the weakest level of selection detectable on a given phylogeny, as in the power curves of Figure 3.6.

Comparative data, while an integral and powerful tool in evolutionary biology, sometimes holds only limited information about the evolutionary process. We suggest that the application of these approaches to specific dataset should routinely be guided by the use of simulation to assess model choice and power.

3.6.1 A parallelized package for the computational methods

To compare models using information criteria it is only necessary to fit each model to the observed data once, while the Monte Carlo approach we describe requires $2n$ model simulations and $4n$ model fits, where n is the number of replicates used. Fortunately, fitting is both fast and easy to parallelize on modern architectures. Our R package `pmc` integrates parallel computation (from the `snowfall` package) with commonly used phylogenetic model fitting tools provided in the `geiger`, `ape` and `ouch` packages. The analyses presented in this paper are included as examples, most of which can be run in minutes when spread over many processors.

3.6.2 Guidelines for analysis

We have discussed how to compare models pairwise, and applied the methods to a series of models for the *Anolis* dataset. However, we have not discussed what one is to do when faced with a multitude of models. Here, as in the situation of choosing which variables to use in a multiple linear regression, there is no single best answer. If there are few enough models, by analogy to stepwise addition for linear regression, one could arrange the models in rough order of complexity, begin with the simplest, and compare each to the next more complex, stopping when there is insufficient support to choose a more complex model. Alternatively, one could do all pairwise comparisons, although the results may be difficult to interpret if there no single model

is clearly best. If there are many models, one option would be to rank all models according to AIC score, and evaluate uncertainty by comparing each model to the top-ranking few models by our methods. There are many methods and philosophies of model choice; it is our opinion that a good method of evaluating uncertainties behind model choice can only aid in this process.

3.7 Acknowledgements

CB thanks P. Wainwright and the rest of the Wainwright lab for valuable input and inspiration and the generous support of the Computational Sciences Graduate Fellowship from the Department of Energy under grant number DE-FG02-97ER25308. PR was supported by funds from G. Coop and S. Schreiber, and by a NIH fellowship under grant number F32GM096686. GC is supported in part by a Sloan Fellowship in Computational and Evolutionary Molecular Biology and by University of California Davis start-up funds. We also thank Luke Harmon and an anonymous reviewer for their insightful comments on an earlier version of the manuscript.

Chapter 4

Quantifying Limits to Detection of Early Warning for Critical Transitions

ectionIntroduction There is an increasing recognition of the importance of regime shifts or critical transitions at a variety of scales in ecological systems (Carpenter, 2011; Drake and Griffen, 2010; Holling, 1973; Scheffer et al., 2009, 2001; Wissel, 1984)âĄă. Many important ecosystems may currently be threatened with collapse, including corals (Bellwood et al., 2004), fisheries (Berkes et al., 2006)âĄă, lakes (Carpenter, 2011), and semi-arid ecosystems (Kéfi et al., 2007)âĄă. Given the potential impact of these shifts on the sustainable delivery of ecosystem services (Folke et al., 2004) and the need for management to either avoid an undesirable shift or else to adapt to novel conditions, it is important to develop the ability to predict impending regime shifts based on early warning signs.

A number of particular systems have demonstrated the kinds of relationships that would produce regime shifts, including dynamics of coral reefs (Mumby et al., 2007), and simple models of metapopulations with differing local population sizes (Hastings, 1991). In cases like these one sensible approach to understanding whether a regime shift would be likely would be to fit the model using either a time series or else independent estimates of parameters. More generally, with a good model of the system, detail-oriented approaches could be useful (Lade and Gross, 2012). In this treatment we focus on the situation where these more detailed models are not available.

Indeed, for many ecological systems specific models are not available and general approaches are needed (Lade and Gross, 2012; Scheffer et al., 2009) that do not depend on estimating the parameters of a known model of a specific system. This has led to a variety of approaches based on summary statistics (*e.g.* Biggs et al., 2009; Carpenter, 2011; Carpenter and Brock, 2006; Dakos et al., 2008; Guttal and Jayaprakash, 2008b; Held, 2004; Seekell et al., 2011) that look for generic

signs of impending regime shifts. Here we extend earlier work by providing estimates of the ability of different potential indicators to accurately signal impending regime shifts, and develop new approaches that both are more efficient and also lay bare some of the important assumptions underlying attempts to find general warning signs of regime shifts. We distinguish this question from the extensive literature involving change-point analysis for the post-hoc identification of if and when a regime shift has occurred (Easterling and Peterson, 1995; Lenton et al., 2009; Rodionov, 2004), which is of little use if the goal is the advanced detection of the shift.

We begin by discussing the limitations of current approaches that rely on summary statistics and provide a description of assumptions through the introduction of a model based approach to detect early warning signals. We then illustrate how stochastic differential equation (SDE) models can be used to reflect the uncertainty inherent in the detection of early warning signals. We caution against paradigms that are not useful for capturing uncertainty in a model-selection based approach, such as information criteria. Finally we use receiver-operating characteristics (Green and Swets, 1989; Keller et al., 2009) as a way to illustrate the sensitivity different data sets and different indicators have in detecting early warning signals and use this to explore a number of examples. This approach provides a visualization of the types of errors that arise and how one can trade off between them, and is important for framing the problem as one focused on prediction.

4.1 The summary statistics approach

Foundational work on early warning signals has operated under the often-implicit assumption that the system dynamics contain a saddle-node bifurcation by looking for patterns that are associated with this kind of transition. A saddle-node bifurcation occurs when a parameter changes and a stable equilibrium (node) and an unstable equilibrium (saddle) coalesce and dissapear. The system then moves to a more distant equilbrium. Guckenheimer and Holmes (1983) or any other textbook on dynamical systems will provide precise definitions and further explanation.

Typical patterns used as warning signals include an increasing trend in a summary statistic such as variance (Carpenter and Brock, 2006), autocorrelation (Dakos et al., 2008; Held, 2004), skew (Guttal and Jayaprakash, 2008b), spectral ratio (Biggs et al., 2009). While attractive for their simplicity, such approaches must confront numerous challenges. In this paper we argue for a model-based approach to warning signals, and describe how this can be done in a way that best

addresses these difficulties. We begin by enumerating several of the difficulties encountered in approaches lacking an explicit model.

Hidden assumptions

The underlying assumption that the system contains a saddle-node bifurcation can be easily overlooked in common summary-statistics based approaches. For instance, variance may increase for reasons that do not signal an approaching transition (Schreiber, 2003; Schreiber and Rudolf, 2008). Alternatively, variance may not increase as a bifurcation is approached (Dakos et al., 2011b; Livina et al., 2012). Some classes of sudden transitions may exhibit no warning signals Hastings and Wysham (2010). Like saddle-node bifurcations, transcritical bifurcations involve an eigenvalue passing through zero, and exhibit the patterns of critical slowing down and increased variance (Drake and Griffen, 2010). However, transcritical bifurcations involve a change in stability of a fixed point, rather than the sudden dissapearance of a fixed point that has made critical transitions so worrisome.While no approach will be applicable to all classes of sudden transitions, it is certainly still useful to have an approach that detects transitions driven by saddle-node bifurcations, which have been found in many contexts (*e.g.*, see Scheffer et al., 2001).

Even when we can exclude or ignore other dynamics and restrict ourselves to systems that can produce a saddle-node bifurcation, approaches based on critical slowing down or rising variance (*e.g.* Carpenter, 2011; Held, 2004; Scheffer et al., 2009) must further assume that a changing parameter has brought the system closer to the bifurcation. This assumption excludes at least three alternative explanations for the transition in system behavior. The first possibility is that a large perturbation of the system state has moved the system into the alternative basin of attraction (Scheffer et al., 2001). This is an exogenous forcing that does not arise from the system dynamics, so it is not the kind of event we can expect to forecast. (An example might be a sudden dramatic increase in fishing effort that pushes a harvested population past a threshold.) The second scenario is a purely noise-induced transition, a chance fluctuation that happens to carry the system across the boundary (Ditlevsen and Johnsen, 2010). Livina et al. (2012) indicate that such noise induced transitions cannot be predicted through early warning signals – at least they are not expected to exhibit the same early warning patterns of increased variance and increased autocorrelation anticipated in the case of a saddle-node bifurcation. The third scenario is that the

system does pass through a saddle-node bifurcation, but rather than gradually and monotonically approaching the critical point, the bifurcation parameter moves in a rapid or highly non-linear way, making the detection of any gradual trend impossible.

Arbitrary windows

In addition to the assumption of a saddle-node bifurcation, the calculation of statistics that would be used to detect an impending transition is subject to several arbitrary choices. A basic difficulty arises from the need to assume a time-series is *ergodic*: that averaging over time is equivalent to averaging over replicate realizations, while trying to test if it is not. Theoretically, the increasing trend in variance, autocorrelation, or other statistics is something that would be measured across an ensemble – across replicates. As true replicates are seldom available in systems for which developing warning signals would be most desirable, typical methods average across a single replicate using a moving window in time. The selection of the size of this window and whether and by how much to overlap consecutive windows varies across the literature. Lenton et al. (2012) demonstrates that these differences can influence the results, and that the different choices each carry advantages and disadvantages.

In addition to introducing the challenge of selecting a window size, this ergodic assumption raises further difficulties. While appropriate for a system that is stationary, or changing slowly enough in the window that it may appear stationary, the assumption is at odds with the original hypothesis that the system is approaching a saddle-node bifurcation.

Further, certain statistics such as the critical slowing down measured by autocorrelation require data that is evenly sampled in time. Interpolating from existing data to create evenly spaced points is particularly problematic, as this introduces an artificial autocorrelation into the data.

No quantitative measures

Summary statistics typically invoke qualitative patterns such as an increase in statistic x, rather than a quantitative measure of the early warning pattern. This makes it difficult to compare between signals or to attribute a statistical significance to the detection. Some authors have suggested Kendall's correlation coefficient, τ, could be used to quantify an increase (Dakos et al., 2011a, 2008) in autocorrelation or variance. Other measures of increase, such as Pearson's corre-

lation coefficient have also been proposed (Drake and Griffen, 2010), while most of the literature simply forgoes quantifying the increase or estimating significance. While adequate in experimental systems that can compare patterns between controls and replicates (*e.g.* Carpenter, 2011; Drake and Griffen, 2010), any real-world application of these approaches must be useful on a single time-series of observations. In these cases a quantitative definition of a statistically significant detection is essential. Without this, we have no assurance that a purported detection is not, in fact, a false positive. By focusing primarily on examples known to be approaching a transition when testing warning signals, the probability of false positives has largely been overlooked.

Problematic null models

Specifying an appropriate null model is also difficult. Non-parametric null hypotheses seem to require the fewest assumptions but in fact can be the most problematic. For instance, the standard non-parametric hypothesis test with Kendall's tau rank correlation coefficient assumes only that the two variables are independent, but this is an assumption that is violated by the very experimental design: temporal correlations will exist in any finely-enough sampled time series, and moving windows introduce temporal correlations in the statistics. Under such a test any adequately large data set will find a significant result, regardless of whether a warning signal exists. A similar problem arises when the points in the time series are reordered to create a null hypothesis – this destroys the natural autocorrelation in the time series. More promising parametric null models have been proposed, such as autoregressive models in Dakos et al. (2008), bringing us closer to a model-based approach with explicit assumptions. Others have looked for alternative summary statistics where reasonable null models are more readily available, such as Seekell et al. (2011)'s proposal to test for conditional heteroscedasticity.

Summary-statistic approaches have less statistical power.

Methods for the detection of early warning signals are continually challenged by inadequate data (Bestelmeyer et al., 2011; Carpenter, 2011; Dakos et al., 2008; Guttal and Jayaprakash, 2008b; Held, 2004; Inman, 2011; Scheffer, 2010; Scheffer et al., 2009). Despite the widespread recognition of the this need for large data sets, there has been very few studies quantitative studies of power to determine at how much data is required (Contamin and Ellison, 2009), how often a particular method would produce a false alarm or fail to detect a signal, and which tests will be the most

powerful or sensitive. The Neyman-Pearson Lemma demonstrates that the most powerful test between hypotheses compares the likelihood that the data was produced under each (Neyman and Pearson, 1933). Such likelihood calculations require a model-based approach.

4.2 A model based approach

Model-based approaches are beginning to play a larger role in early warning signal detection, though we have not as yet seen the direct fitting and simulation of models to compare hypotheses. While choosing appropriate models without system-specific knowledge is challenging, much can be accomplished by framing the implicit assumptions into equations. Lade and Gross (2012) introduce the idea of generalized models for early warning signals, and Kuehn (2011) presents normal forms for bifurcation processes that can give rise to critical transitions. Carpenter and Brock (2011) and Dakos et al. (2011b) start by assuming the dynamics obey a generic stochastic differential equation (SDE), but use this only to derive or define the summary statistics of interest.

In this section we outline how the detection of early warning signals may be thought of as a problem of model choice. We next show generic models can be constructed under the assumptions discussed above and estimated from the data in a maximum likelihood framework. We highlight the disadvantages of comparing these estimates by information criteria, and instead introduce a simulation or bootstrapping approach rooted in Cox (1961) and McLachlan (1987) that characterizes the rate of missed detections and false alarms expected in the estimate.

Early warning signals as model choice

It may be useful to think of the detection of early warning signals as a problem of model choice rather than one of pattern recognition. The model choice approach attempts to frame each of the possible scenarios as structurally different equations, each with unknown parameters that must be estimated from the data. In any model choice problem, it is important to identify the goal of the exercise – such as the ability to generalize, to imitate reality, or to predict (Levins, 1966). In this case generality is more important than realism or predictive capability: we will write down a general model that is capable of approximating a wide class of models in which regime shifts are characterized by a saddle-node bifurcation, and a second generic model that is capable of representing the behavior of such systems when they are not approaching a bifurcation. These

may be thought of as the hypothesis and null hypothesis, though they are in fact compound hypotheses, as we must first estimate the model parameters from the data. In this approach it is not assumed that "reality" is included in the models being tested, but that one of the models is a better approximation of the true dynamics than the other. System whose dynamics violate the assumptions common to both models, such as in the examples of Hastings and Wysham (2010) where systems exhibit sudden transitions without warning, fall outside the set of cases where this approach would be valid; though the inability of either model to match the system dynamics could be an indication of such a violation.

Models

In the neighborhood of a bifurcation a system can be transformed into its *normal form* by a change of variables to facilitate analysis (Guckenheimer and Holmes, 1983). The normal form (Guckenheimer and Holmes, 1983; Kuehn, 2011) for the saddle-node bifurcation is

$$\frac{\mathrm{d}x}{\mathrm{d}t} = r_t - x^2. \tag{4.1}$$

where x is the state variable and r_t our bifurcation parameter. We have added a subscript t to the bifurcation parameter as a reminder that it is the value which may be slowly varying in time and consequently moving the system closer to a critical transition or regime shift (Scheffer et al., 2009). Transforming this canonical form to allow for an arbitrary mean in the state variable θ, the system near the bifurcation looks like $dx/dt = r_t - (\theta - x)^2$, with fixed point $\hat{x} = \sqrt{r_t} + \theta =: \phi(r_t)$. We expand around the fixed point and express as a stochastic differential equation (*e.g.* Gardiner, 2009):

$$\mathrm{d}X = \sqrt{r_t}(\phi(r_t) - X_t)\mathrm{d}t + \sigma\sqrt{\phi(r_t)}\mathrm{d}B_t \tag{4.2}$$

where B_t is the standard Brownian motion. This expression captures the behavior of the system near the stable point as it approaches the bifurcation. Allowing the stochastic term to scale with the square root of ϕ follows from the assumption that of an internal-noise process, such as demographic stochasticity, that arises in deriving the SDE from a Markov process, see Kampen (2007) or Black and McKane (2012). The square root could be removed for an external noise process, such as environmental noise. In practice it will be difficult to descriminate between the

square root and linear scaling in these applications, since the average value of the state changes little before the bifurcation.

As we discuss above, in this paradigm we must include an assumption on how the bifurcation parameter, r_t, is changing. We assume a gradual, monotonic change which we approximate to first order:

$$r_t = r_0 - mt. \tag{4.3}$$

Detecting accelerating or otherwise nonlinear approaches to the bifurcation will generally require more power. When the underlying system is not changing, r_t is constant ($m = 0$) and Equation (4.2) will reduce to a simple Ornstein-Uhlenbeck process,

$$\mathrm{d}X_t = r(\theta - X_t)\mathrm{d}t + \sigma \mathrm{d}B_t \tag{4.4}$$

This is the continuous time analog of the first-order autoregressive model considered as a null model elsewhere (*e.g.* Dakos et al., 2008; Guttal and Jayaprakash, 2008a).

Likelihood calculations

The probability $P(X|M)$ of the data X given the model M is the product of the probability of observing each point in the time series given the previous point and the length of the interval,

$$\log P(X|M) = \sum_i \log P(x_i|x_{i-1}, t_i) \tag{4.5}$$

For (4.2) or (4.4) it is sufficient (Gardiner, 2009) to solve the moment equations for mean and variance respectively:

$$\frac{\mathrm{d}}{\mathrm{d}t}E(x|M) = f(x) \tag{4.6}$$

$$\frac{\mathrm{d}}{\mathrm{d}t}V(x|M) = -\partial_x f(x)V(x|M) + g(x)^2 \tag{4.7}$$

For the OU process, we can solve this in closed form over an interval of time t_i between subsequent observations:

$$E(x_i|M = \text{OU}) = X_{i-1}e^{-rt_i} + \theta\left(1 - e^{-rt_i}\right) \tag{4.8}$$

$$V(x_i|M = \text{OU}) = \frac{\sigma^2}{2r}\left(1 - e^{-2rt_i}\right) \tag{4.9}$$

For the time dependent model, we have analytic forms only for the dynamical equations of these moments from equation (4.7), which we must integrate numerically over each time interval. The moments of Equation (4.2) are given by

$$\frac{\mathrm{d}}{\mathrm{d}t}E(x_i|M = \text{LSN}) = 2\sqrt{r(t)}(\sqrt{r(t)} + \theta - x_i) \tag{4.10}$$

$$\frac{\mathrm{d}}{\mathrm{d}t}V(x_i|M = \text{LSN}) = -2\sqrt{r(t)}V(x_i) + \sigma^2(\sqrt{r(t)} + \theta) \tag{4.11}$$

These are numerically integrated using `lsoda` routine available in `R` for the likelihood calculation.

Comparing Models

Likelihood methods form the basis of much of modern statistics in both Frequentist and Bayesian paradigms. The ability to evaluate likelihoods directly by computation has made it possible to treat cases that do not conform to traditional assumptions more directly. The basis of likelihood comparisons has its roots in the Neyman-Pearson Lemma, which essentially asserts that comparing likelihoods is the most powerful test of a choice between two hypotheses (Neyman and Pearson, 1933), and motivates tests from the simple likelihood ratio test up through modern model adequacy methods.

The hypotheses considered here are more challenging then the original lemma provides for, as they are composite in nature: they specify two model forms (stable and changing stability) but with model parameters that must be first estimated from the data. Comparing models whose parameters have been estimated by maximum likelihood is first treated by Cox (1961, 1962), and has since been developed in this simulation estimation of the null distribution (McLachlan, 1987), by parametric bootstrap estimate (Efron, 1987). Cox's δ statistic (often called the deviance between models) is simply the difference between the log likelihoods of these maximum likelihood estimates, defined as follows.

Let L_0 be the likelihood function for model 0, let $\theta_0 = \arg\max \theta_0 \in \Omega_0$, $(L_0(\theta_0|X))$ be the maximum likelihood estimator for θ_0 given X, and let $L_0 = L_0(\theta_0|X)$; and define L_1, θ_1, L_1 similarly for model 1. The statistic we will use is δ, defined to be twice the difference in log likelihood of observing the data under the two MLE models,

$$\delta = -2(\log L_0 - \log L_1). \tag{4.12}$$

This approach has been applied to the problem of model adequacy (Goldman, 1993) and model choice (Huelsenbeck and Bull, 1996) in other contexts. We have extended the approach by generating the test distribution as well as a null distribution of the statistic δ.

4.2.1 Simulation-based comparisons

We perform the identical analysis procedure described above on each of these three data sets. First, we estimate parameters for the null and test model to each data set by maximum likelihood. Comparing the likelihood of these fits directly gives us only a minimal indication of which model fits better. To identify if these differences are significant, and by what probability they could arise as a false alarm or a missed event, we simulate 500 replicate time series from each estimated model.

The model parameters of both models are re-estimated on both families of replicates (the null and test, *i.e.* $2 \times 2 \times 500$ fits). The differences in the likelihood values between the model estimates produced from the first set of simulations determines the null distribution for the deviance statistic δ. As the constant OU process model is nested within the time-heterogeneous model, these values are always positive, but tend to be not as large as those produced when the models are fit to the second family of data.

The extent to which these distributions overlap indicates our inability to distinguish between these scenarios. The tendency of the observed deviance to fall more clearly in the domain of one distribution or the other indicates the probability our observed data corresponds best with that model – either approaching a critical transition or remaining stable. While it trivial to assign a statistical significance to this observation based on how far into the tail of the null distribution it falls, for the reasons we discussed we prefer the more symmetric comparison of the probability that this value was observed in either distribution. We visualize the trade-off between false alarms

and failed detection using the ROC curves introduced above.

Information criteria will not serve.

One will commonly observe models representing alternative processes being compared through the use of various information criteria such as the Akaike information criterion. While tempting to apply in this situation, such approaches are not suited to this problem for several reasons. The first is that information criteria are not concerned with the model choice objective we have in mind, as they are typically applied to find an adequate model description without too many parameters that the system may be over-fit. More pointedly, information criteria have no inherent notion of uncertainty. Information criteria tests alone will not tell us our chances of a false alarm, of missing a real signal, or how much data we need to be confident in our ability to detect transitions.

Beyond hypothesis testing

It is possible to frame the question of sensitivity, reliability, and adequate data in the language of hypothesis testing. This introduces the need for selecting a statistical significance criterion. In the hypothesis testing framework, a false positive is a Type I error, which is defined relative to this arbitrary statistical significance criterion, most commonly 0.05. By changing the criterion, one can increase or decrease the probability of the Type I error at the cost of decreasing or increasing false negative or Type II error, which must also be defined relative to this criterion.

The language of hypothesis testing is built around a bias that false positives are worse than false negatives, and consequently an emphasis on p-values rather than power. In the context of early warning signals this is perilous – it suggests that we would rather fail to predict a catastrophe than to sound a false alarm. To avoid this linguistic bias and the introduction of an nuisance parameter on which to define statistical significance, we propose the use of receiver operating characteristic (ROC) curves.

ROC Curves

We illustrate the trade-off between false alarms and failed detection using receiver-operating characteristic curves first developed in signal-processing literature (Green and Swets, 1989; Keller

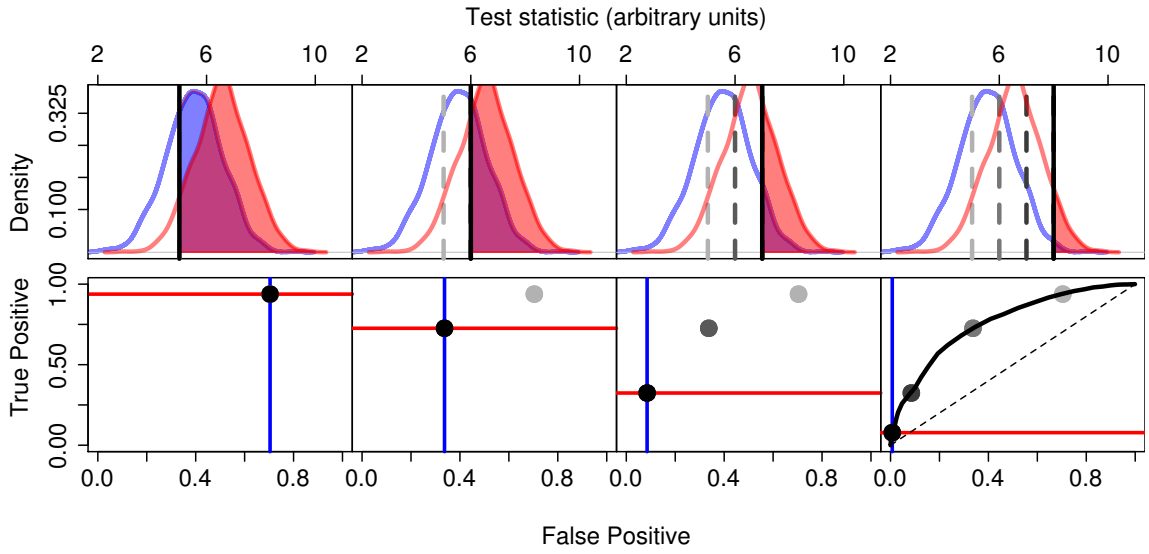


Figure 4.1: Top row: The distributions of a hypothetical warning indicator are shown under the case of a stable system (blue) and a system approaching a critical transition (red). Bottom row: Points along the ROC curve are calculated for each possible threshold indicated in the top row. The false positive rate is the integral of the distribution of the test statistic under the stable system right of the threshold (blue shaded area, corresponding to blue vertical line). The true positive rate is the integral of the system approaching a transition left of the threshold (red shaded area, corresponds to the red line). Successive columns show the threshold increasing, tracing out the ROC curve.

et al., 2009)âĄă. The curves represent the corresponding false alarm rate at any detection sensitivity (true positive rate), Fig 4.1. The closer these distributions are to one-another, the more severe the trade-off. If the distributions overlap exactly, the ROC curve has a constant slope of unity. The ROC curve demonstrates this trade-off between accuracy and sensitivity. Different early-warning indicators will vary in their sensitivity to detect differences between stable systems and those approaching a critical transition, making the ROC curves a natural way to compare their performance. Since the shape of the curve will also depend on the duration and frequency of the time-series observations, we can use these curves to illustrate by how much a given increase in sampling effort can decrease the rate of false alarms or failed detections.

4.3 Example Results

We illustrate this approach on simulated data as well as several natural time-series that have been previously analyzed for early warning signals. All data and code for simulations and analysis are

found in the accompanying R package, `earlywarning`.

Data

The simulation implements an individual, continuous-time stochastic birth-death process with rates given by the master equation (Gardiner, 2009),

$$\frac{dP(n,t)}{dt} = b_{n-1}P(n-1,t) + d_{n+1}P(n+1,t) - (b_n + d_n)P(n,t) \tag{4.13}$$

$$b_n = \frac{eKn^2}{n^2 + h^2} \tag{4.14}$$

$$d_n = en + a_t \tag{4.15}$$

where $P(n,t)$ is the probability of having n individuals at time t, b_n is the probability of a birth event occurring in a population of n individuals, d_n the probability of a death. e, K, h and a_t are parameters. This corresponds to the well-studied ecosystem model of over-exploitation (May, 1977; Noy-Meir, 1975), with stochasticity introduced directly through the demographic process. We select this model since it is has discrete numbers of individuals, nonlinear processes, and the noise is driven by Poisson process of births and deaths instead of a Gaussian, and thus provides an illustration that our approach is robust to the violations of those assumptions in model (4.2).

This model is forced through a bifurcation by gradually increasing the a parameter, which increases can be thought of as an increasing toxicity of the environment (from $a_0 = 100$ increasing at constant rate of 0.09 units/unit time). Other parameters are: $Xo = 730$, $e = 0.5$, $K = 1000$, $h = 200$. We run this model over a time interval from 0 to 500 and sample at 40 evenly spaced time points, which were used for subsequent analysis. This sampling frequency was chosen to be representative of reasonable sampling in biological time-series, and provides enough points to detect a signal while not too many that errors can be avoided entirely. For the convenience of the inquisitive reader, we have also provided a simple function in the associated R package where the user can vary the sampling scheme and parameter values and rerun this analysis. This time series is shown in the top panel of Figure 4.2.

The first empirical data set comes from the population dynamics of *Daphnia* living in the chemostat "H6" in the experiments of Drake & Griffen (Drake and Griffen, 2010). This individual replicate was chosen as an example that showed a pattern of increasing variance over the 16 data

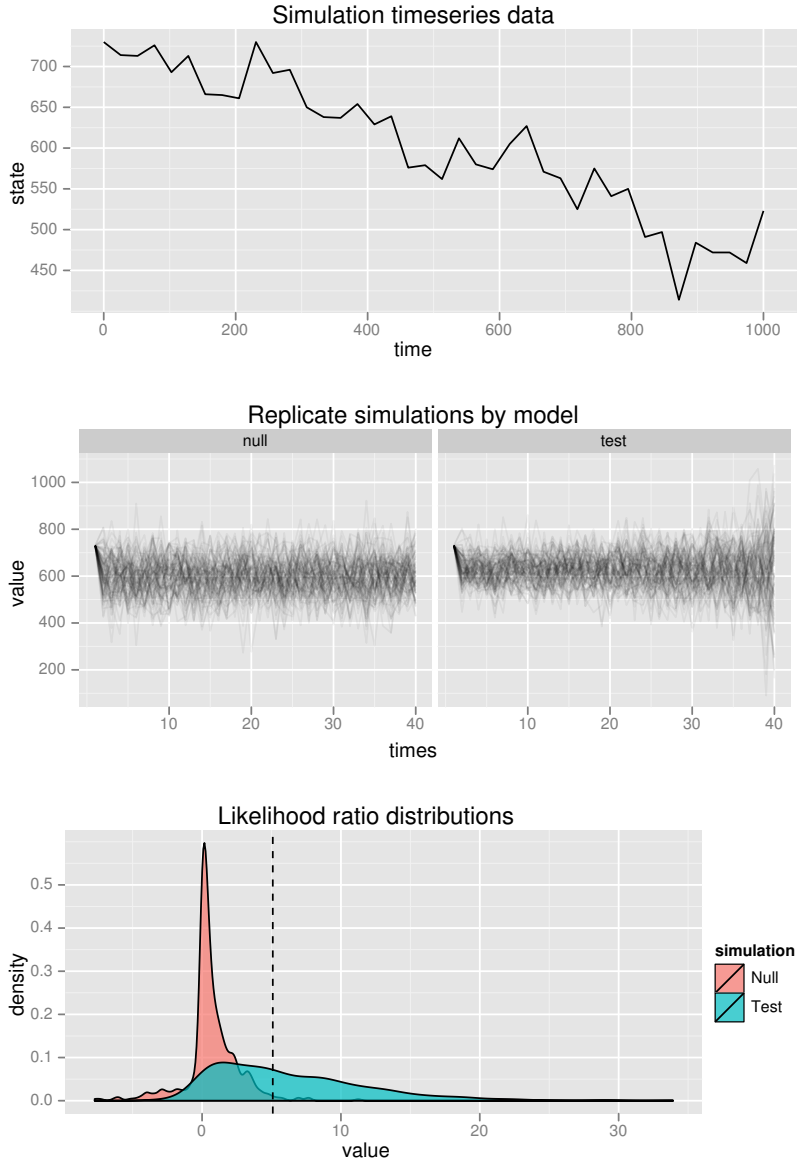


Figure 4.2: A model-based calculation of warning signals for the simulated data example. Top panel: The original time series data on which model parameters for Equations (4.4) and (4.2) are estimated. Middle panel: replicate simulations under the maximum likelihood estimated (MLE) parameters of the null model, Equation (4.4) and test model, Equation (4.2). Bottom panel: The distribution of deviances (differences in log likelihood, Equation (4.12)), when both null and test models are fit to each of the replicates from the null model, "null," in red, and these differences when estimating for each of the replicates from the test model, in blue. The overlap of distributions indicate replicates that will be difficult to tell apart. The observed differences in the original data are indicated by the vertical line.

points where the system was being manipulated towards a crash. This time series is shown in the top panel of Figure 4.3.

Our second empirical data set comes from the glaciation record seen in deuterium levels in Antarctic ice cores (Petit et al., 1999), as analyzed by Dakos et al. (2008). The data are preprocessed by linear interpolation and de-trending by Gaussian kernel smoothing to be as consistent as possible with the original analysis. We focus on the third glaciation event, consisting of 121 sample points. The match is not exact since Dakos et al. (2008) estimates the de-trending window size manually, but the estimated correlations in the first-order auto-regression coefficients are in close agreement with that analysis. De-trending is intended to make the data consistent with the assumptions of the warning signal detection (Dakos et al., 2008), which did not apply to the other data sets (Drake and Griffen, 2010). This time series is shown in the top panel of Figure 4.4.

Analysis

The deviances δ observed are 5.1, 6.0, 83.9 for the simulation, the chemostat data and the glaciation data, respectively. Based on AIC score each is large enough to reject the null hypothesis of a stable model with its one extra parameter, but this does not give the full picture of the anticipated error rates. The size of these differences reflects not only the magnitude of the difference in fit between the models but also the arbitrary units of the raw likelihoods, which are smaller for larger datasets. Consequently the glaciation score reflects as much the greater length of its time series as it does anything else.

Our simulation approach can provide a better sense of the relative trade-off in error rates associated with these estimates. As described above (Section 4.2.1), we simulate 500 replicates under each model, shown in the middle panels of Figures 4.2, 4.3 and 4.4, and determine the distributions in likelihood ratio under each, shown in the lower panels. The observed deviance from the original data is also indicated (vertical line).

The ROC curves for each of these data sets are plotted in Figure 4.5. While differences in the rate at which the system approaches a transition will also improve the ratio of true positives to false positives, here we see the best-sampled data set, Glaciation, with 121 points, also has the clearest signal with no observed errors in the 500 replicates of each type. Comparing the chemostat and simulation curves illustrate how the trade-off between false positives and true

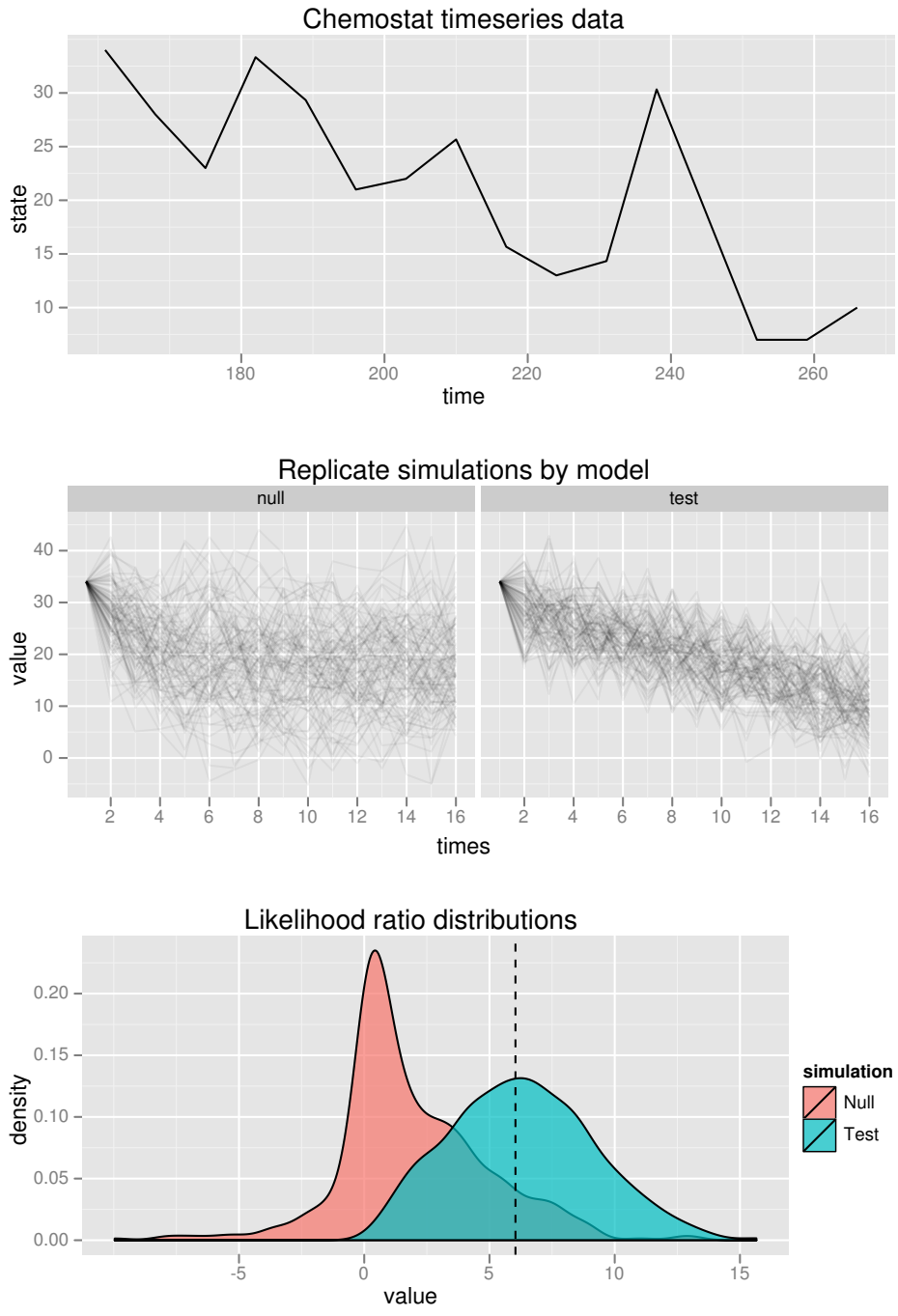


Figure 4.3: A model-based calculation of warning signals for the Daphnia data analyzed in Drake and Griffen (2010) (Chemostat H6). Panels as in Figure 4.2.

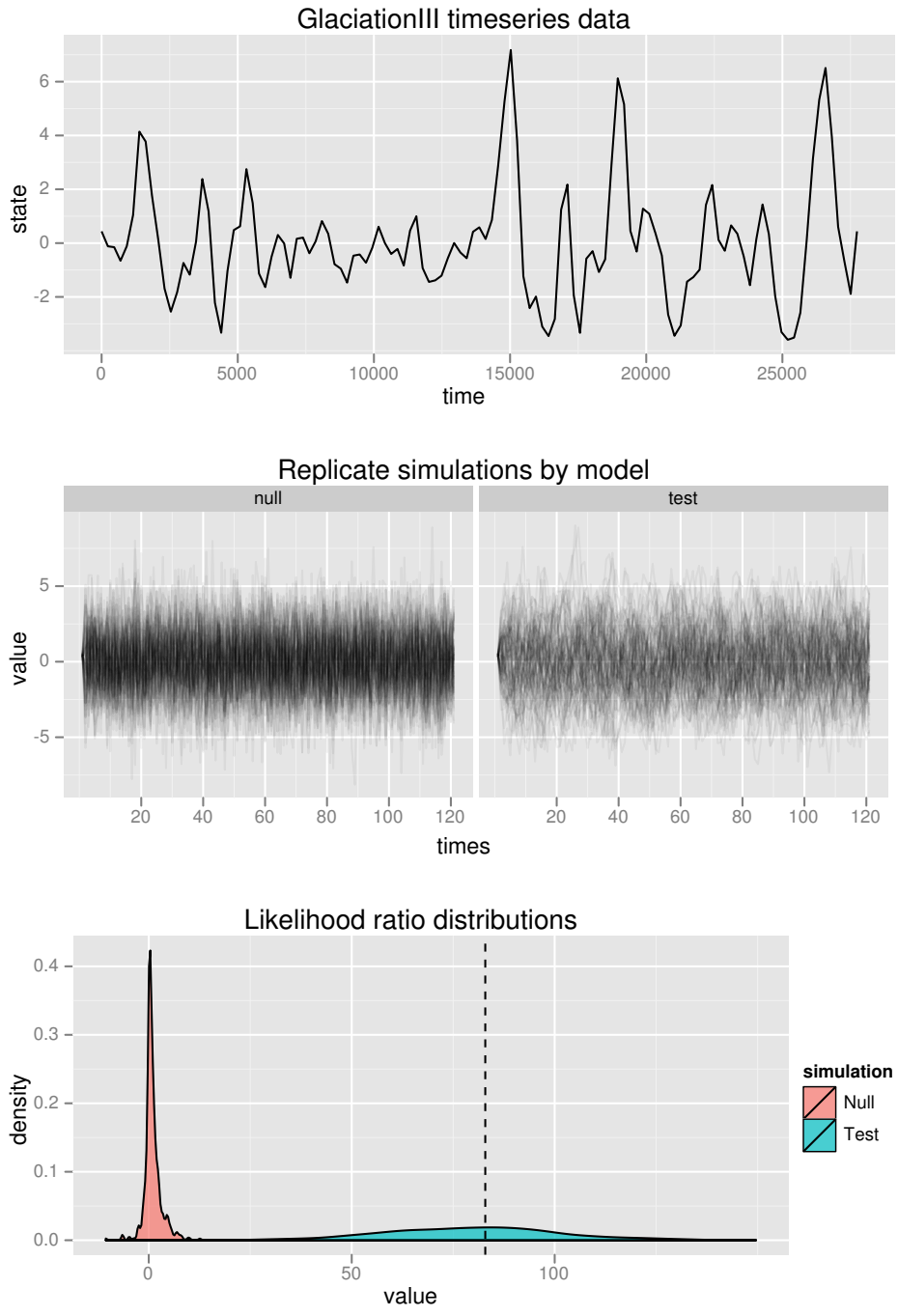


Figure 4.4: A model-based calculation of warning signals for the Glaciation data analyzed in Dakos et al. (2008) (Glaciation III). Panels as in Figure 4.2.

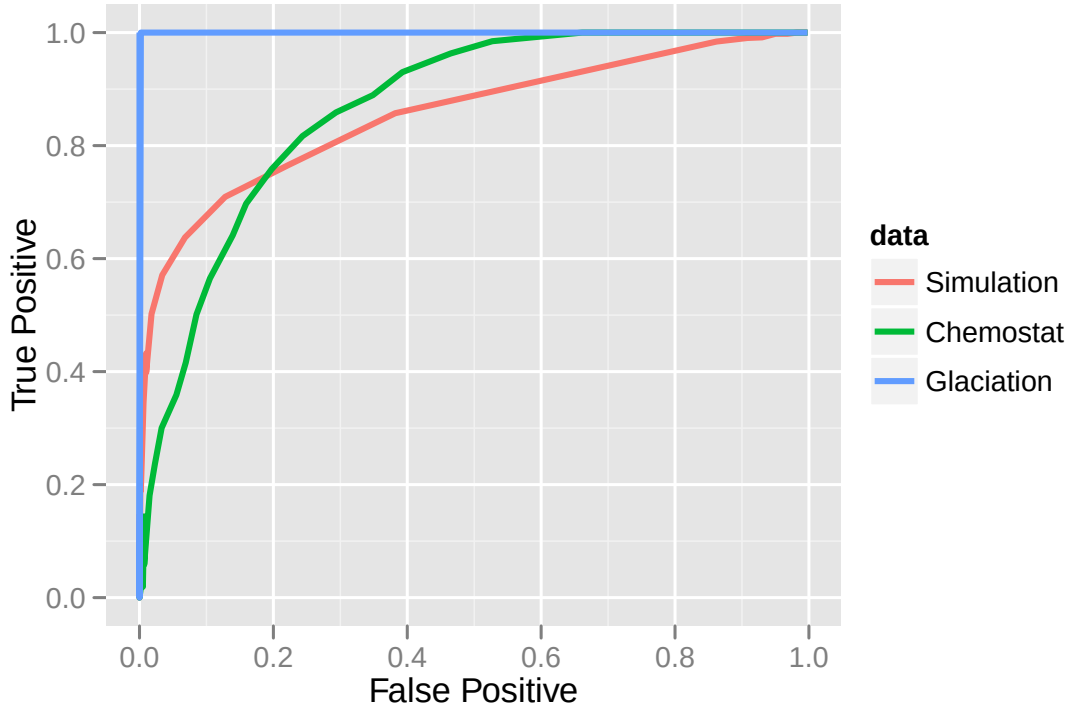


Figure 4.5: ROC curves for the Simulation, Chemostat, and Glaciation data, computed from the distributions shown in Figures 4.2, 4.3 and 4.4, bottom panel.

positives can vary between data. The chemostat signal, which estimates a relatively rapid rate of change but has less data, captures a higher rate of true positives for a given rate of false positives than the simulation data set with a weaker rate of change but more data, for false positive rates above 20%. However, the simulated set with more data performs better if lower false-positive rates are desired.

4.4 Comparing the performance of summary statistics and model-based approaches

Due to the variety of ways in which early warning signals based on summary statistics are implemented and evaluated it is difficult to give a straight-forward comparison between them and the performance of this model-based approach. However, by adopting one of one of the quantitative measures of a warning signal pattern, such as Kendall's τ (Dakos et al., 2011a, 2009, 2008), we are able to make a side-by-side comparison of the different summary statistics and the model based

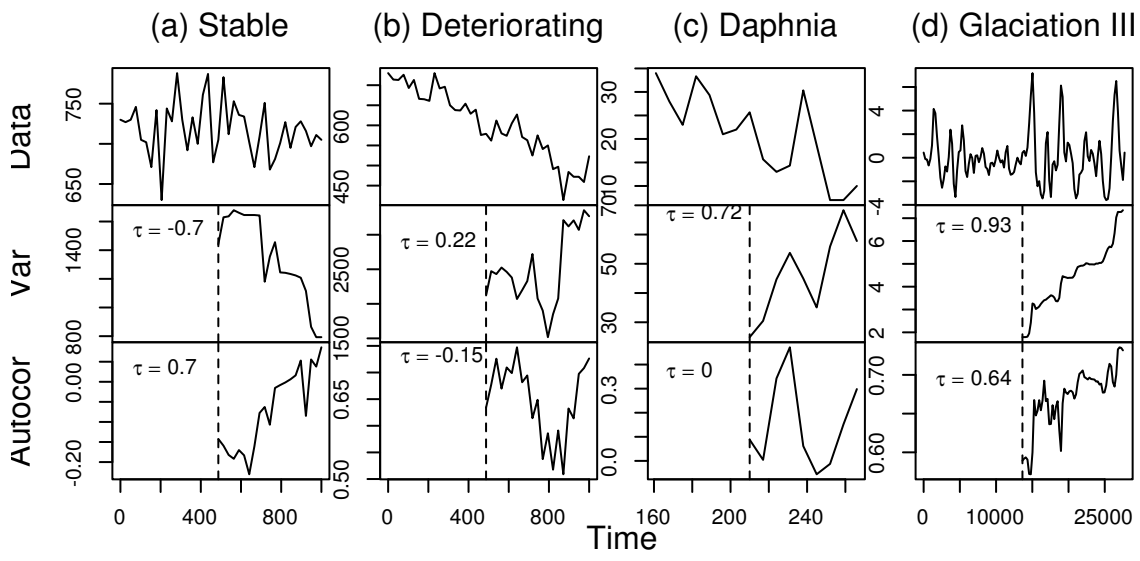


Figure 4.6: Early warning signals in simulated and empirical data sets. The first two columns are simulated data from (a) a stable system (Stable), and (b) the same system approaching a saddle-node bifurcation (Deteriorating). Empirical examples are from (c) *Daphnia magna* concentrations manipulated towards a critical transition (Daphnia), and (d) deuterium concentrations previously cited as an early warning signal of a glaciation period (Glaciation). Increases in summary statistics, computed over a moving window, have often been used to indicate if a system is moving towards a critical transition. The increase is measured by the correlation coefficient τ. Note that positive correlation does not guarantee the system is moving towards a transition, as seen in the stable system, first column.

approach in the context of false alarms and failed detections shown by the ROC curve. Values of τ near unity indicate a strongly increasing trend in the warning indicator, which is supposed to be indicative of an approaching transition. Values near zero suggest a lack of a trend, as expected in stable systems.

Figure 4.6 shows the time series for each data set in columns and the early warning indicators of variance and autocorrelation computed over a sliding window for each. Kendall's correlation coefficient τ is calculated for each warning indicator and displayed on the graphs, inset. For comparison, the left-most column includes data simulated under a stable system, which nevertheless shows a chance increasing autocorrelation with a $\tau = 0.7$ We can adapt the approach we have described above to determine how often such a strong increase would appear by chance in a stable system as follows.

By estimating the stable and critical transition models from the data, and simulating 500

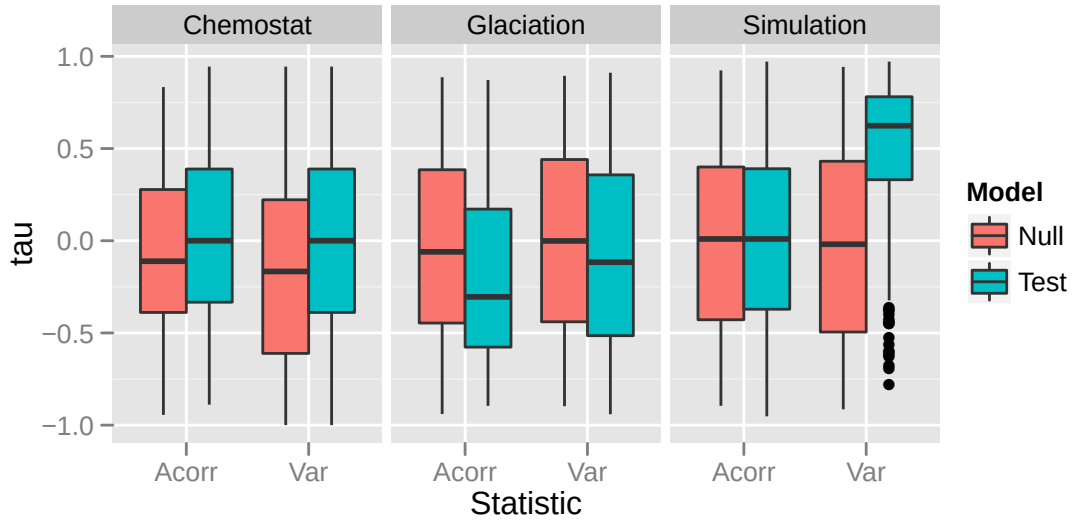


Figure 4.7: Box-plots of the distributions of Kendall's τ observed for the summary statistic methods variance and autocorrelation, applied to three different data sets (from Figures 4.2, 4.3, 4.4). The distributions show extensive overlap, suggesting that it will be difficult to distinguish early warning signals by the correlation coefficient in these summary statistics.

replicate data sets under each as in the analysis above, we can then calculate the warning signals statistic over a sliding window of size equal to one-half the length of the time series, and compute the correlation coefficient τ measuring the degree to which the statistic shows an increasing trend. This results in a distribution of τ values coming from a model of a stable system, and a corresponding distribution of τ values coming from the model with an impending transition. These distributions are shown in Figure 4.7. Contrary to the expectation that replicates of the null model (stable system, Equation (4.4)) would cluster around zero, while the test model, Equation (4.2), would cluster around larger positive τ values, the observed τ values on the replicates extend evenly across the range. This results in dramatic overlap and offers little ability to distinguish between the stable replicates and the replicates approaching a transition.

The use of box plots in Figure 4.7 provide a convenient and familiar way to visualize the overlap between more than two distributions, though they lack the resolution of the overlapping density distributions in Figures 4.2, 4.3, 4.4. The overlapping distributions are the natural representation from which to introduce the ROC curve, as in Figure 4.1.

The ROC curves for these data (Fig. 4.8) show that the summary-statistic based indicators

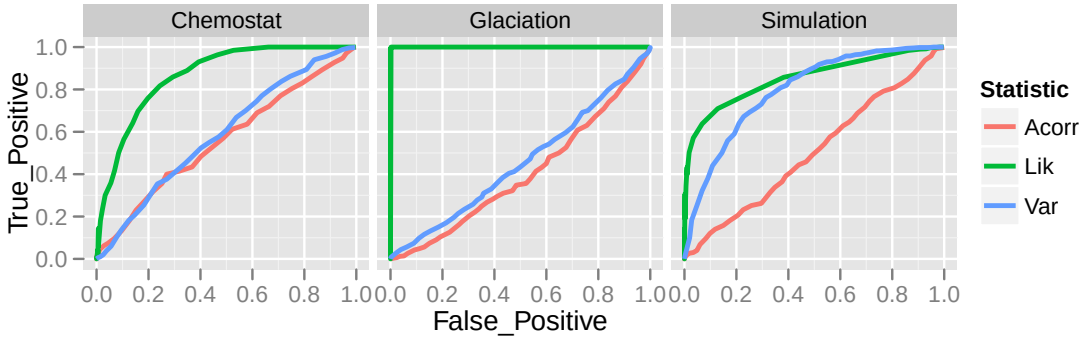


Figure 4.8: ROC curves compare the performance of the summary statistics variance and autocorrelation against the likelihood-based approach from Figure 4.5 on each of three example data sets (Figures 4.2, 4.3, 4.4).

	Variance	Likelihood
Simulation	25 %	61%
Chemostat	5.0%	34%
Glaciation	5.4%	100%

Table 4.1: Fraction of *crashes detected* when the desired false alarm rate is fixed to 5%

frequently lack the sensitivity to distinguish reliably between observed patterns from a stable or unstable system. The large correlations observed in the empirical examples (Fig. 4.6) are not uncommon in stable systems. It is notable that in both empirical examples the summary statistics approach does little better than chance in distinguishing replicates that have been simulated from models (4.2) and (4.4), despite the fact that these models correspond to the assumptions of the summary statistics approaches. On the simulated data, the variance based method approaches the true-positive rate of our likelihood method at higher levels of false positives, but performs worse when the desired level of false positives is low. The ROC curve helps us compare the performance of the different approaches at different tolerances. For instance, Table 4.1 shows the fraction of true crashes caught at a 5% false positive rate. We can instead set a desired True positive rate and read off the resulting number of false alarms, Table 4.2.

4.5 Discussion

The challenge of determining early warning signs for impending possible regime shifts requires real attention to the underlying statistical issues and other assumptions. Doing this, does, how-

	Variance	Likelihood
Simulation	49 %	55%
Chemostat	81 %	35%
Glaciation	93 %	0%

Table 4.2: Fraction of *false alarms* when the desired detection rate is fixed to 90%

ever, open up new possibilities for asking what the goal of detection should be, and for clearly identifying underlying assumptions. We consider alternative approaches based either on summary statistics or a likelihood based model choice. By assuming the underlying model corresponds to a saddle-node bifurcation, our analysis presents a "best-case scenario" for both summary statistic and likelihood-based approaches. Other literature has already begun to address the additional challenges posed when the underlying dynamics do not correspond to these models (Hastings and Wysham, 2010). Our results illustrate that even in this best-case scenario, reliable identification of warning signals from summary statistics can be difficult.

We have used three examples to illustrate the performance of this approach in data from simulation, a chemostat experiment, and paleo-atmospheric record; examples differing in sampling intensity and strength of signal of an approaching collapse. While the well-sampled geological data shows an unmistakable signal in this model-based approach, the uncertainty in the smaller simulated and experimental data forces a trade-off between errors.

As a way to clearly illustrate the choices involved in looking for warning signals while avoiding false alarms, we introduce an approach based on receiver operator curves. These curves illustrate the extent to which an potential warning signal mitigates the trade-off between missed events and false alarms. The extent of the difficulty in finding reliable indicators of impending regime shifts based on summary statistics becomes clear from the ROC curves of these statistics, where a 5% false positive rate often corresponds to only a 5% true positive rate, performing no better than the flip of a coin. By estimating the ROC curve for a given set of data, we can better avoid applying warning signals in cases of inadequate power. By taking advantage of the assumptions being made to write down a specific likelihood function, we can develop approaches that get the most information from the data available.

In any application of early warning signals, it is essential to address the question of model adequacy. Our approach formalizes the assumptions about the underlying process to match the

assumptions of the other warning signals. As the bifurcation results from the principle eigenvalue passing through zero, the warning signal is expected in linear-order dynamics; estimation of the nonlinear model is less powerful and less accurate. The performance of this approach in the simulated data – which is nonlinear in its dynamics and driven with non-Gaussian noise introduced by the Poisson demographic events – demonstrates the accuracy under violation of these assumptions.

The conclusion is not simply that likelihood approaches are more reliable, but rather more broadly that warning signals should consider the inherent trade-off between sensitivity and accuracy, and must quantify how this trade-off depends on both the indicators used and the data available. The approach developed here estimates the risk of both failed detection and false alarms; concepts which are critical to prediction-based management. Using the methods we have outlined when designing early warning strategies for natural systems can ensure that data collection has adequate power to offer a reasonable chance of detection.

4.6 Acknowledgments

We would like to thank S. Schreiber, M. Holyoak, M. Baskett, A. Perkins, N. Ross, M. Holden, and two anonymous reviewers for comments on the manuscript. This research was supported by funding from NSF Grant EF 0742674 and a Computational Sciences Graduate Fellowship from the Department of Energy grant DE-FG02-97ER25308. Data and code are available at https://github.com/cboettig/earlywarning.

APPENDIX 1

A.1 Adaptive Dynamics and the Transition Probability $w(y|x)$

In this appendix we construct the Markov process $w(y|x)$ under the assumptions of adaptive dynamics (Dieckmann and Law, 1996). The probability per unit time of making the transition in trait space from a monomorphic population with trait x to one with trait y is given by

$$w(y|x) = \mathcal{M}(y,x)\mathcal{D}(y,x). \tag{A.1}$$

In the framework presented here, a monomorphic population of residents with trait x generate mutants with trait y, some of which survive. The rate at which a mutation is generated from a population is

$$\mathcal{M}(y,x) = \mu(x)b(x)N^*(x)M(x,y), \tag{A.2}$$

where $b(x)$ is the per-capita birth rate at equilibrium, $\mu(x)$ the mutation probability per birth, $N^*(x)$ the equilibrium population size for a population with trait x, and $M(x,y)$ is the distribution from which the mutant trait is drawn. The probability of surviving accidental extinction of a branching process given the mean individual birth rate b and mean death rate d for the mutant y is $\mathcal{D}(y,x) = 1 - d(y,x)/b(y,x)$ if $d(y,x) < b(y,x)$ and $\mathcal{D}(y,x) = 0$ otherwise (Feller, 1968). The terms $b(y,x)$ and $d(y,x)$ refer to the birth and death rate, respectively, of a rare mutant with trait y in an equilibrium population of x.

Given a mutant strategy y such that $\mathcal{D}(y,x) > 0$ we have

$$w(y|x) = \mu(x)N^*(x)b(x)M(x,y)[b(y,x) - d(y,x)]/b(y,x). \tag{A.3}$$

Expanding the fitness, $b(y,x) - d(y,x)$, to first order the transition rate is then

$$w(y|x) \approx \mu(x)N^*(x)\partial_y s(y,x)|_{y=x}M(x,y)[y-x], \tag{A.4}$$

where $\partial_y s(y,x)|_{y=x}$ is known as the selective derivative (Geritz et al., 1997). From Eq. (A.4) one can apply a particular model by specifying expressions for the mutation rate $\mu(x)$, stationary population size, $N^*(x)$, fitness function $s(y,x)$ and mutational kernel $M(x,y)$. In the competition for a limited resource model, (Dieckmann and Doebeli, 1999) used here, these are:

$$\begin{aligned}
\mu(x) &= \mu, \\
M(y,x) &= \frac{1}{\sqrt{2\pi\sigma_\mu^2}} e^{-\frac{(y-x)^2}{2\sigma_\mu^2}}, \\
s(y,x) &= r\left(1 - \frac{N^*(y)e^{-\frac{(x-y)^2}{2\sigma_c^2}}}{N^*(x)}\right), \\
N^*(x) &= K_0 e^{-\frac{x^2}{2\sigma_k^2}}.
\end{aligned} \tag{A.5}$$

Consequently, the evolutionary transition rates in for this model are given by

$$w(y|x) = -\mu K_0 e^{-\frac{x^2}{2\sigma_k^2}} \frac{rx}{\sigma_k^2} \frac{e^{-\frac{(y-x)^2}{2\sigma_\mu^2}}}{\sqrt{2\pi\sigma_\mu^2}} [y-x]. \tag{A.6}$$

The transition rate $w(y|x)$ for the explicit resource competition model is presented along with the model details in Appendix A.3. Using the appropriate transition rate in the linear noise approximation described in Appendix A.2, we recover the equations for the curves plotted in Fig. 2.1 which are integrated to obtain the theoretical predictions of Fig. 2.2. These explicit expressions are given in Appendix A.5.

A.2 Linear Noise Approximation

A.2.1 About the approximation

The linear noise approximation is a common approach for describing Markov processes. Though often applied in discrete cases such as one-step (birth-death) processes, it can be generalized to the continuous case we consider, where a population at trait x can jump to another trait value y. The approximation transforms the Markov process specified by a master equation on the transition rates $w(y|x)$ to an approximate partial differential equation (PDE) for the probability distribution.

This PDE resembles the Fokker-Planck equation for the process[1], except that the PDE resulting from the linear noise approximation is guaranteed to be linear and its solutions Gaussian. Consequently, solving for the two moments, the mean and variance, will lead to a system of ordinary differential equations. Substituting the form of $w(y|x)$ found in Appendix A.1 into this ODE system recovers Eq. (2.3) and (2.4) in the text.

The approximation is straight-forward (involving a change of variables and a Taylor expansion), if cumbersome. The approximation is rigorously justified over any fixed time interval T in the limit of small step sizes (Kurtz, 1971), which parallels more modern justification of the Canonical equation (Champagnat et al., 2001). The original derivation of the canonical equation makes use of (unscaled) jump moments, introduced by van Kampen (2001). We review this approach first, as it provides a good intuition for the full linear noise approximation. The actual approximation relies on a change of variables which makes explicit use of the small step sizes, and *derives* rather than assumes the Gaussian character of the distribution.

A.2.2 Original jump moments

The dynamics of the average phenotypic trait are given by

$$\frac{\mathrm{d}\hat{x}(t)}{\mathrm{d}t} = \int \mathrm{d}x \; x \frac{\mathrm{d}}{\mathrm{d}t} P(x,t). \tag{A.7}$$

Using the master equation

$$\frac{\mathrm{d}}{\mathrm{d}t} P(x,t) = \int \mathrm{d}y \; \left[w(x|y)P(y,t) - w(y|x)P(x,t) \right], \tag{A.8}$$

to replace $\frac{\mathrm{d}}{\mathrm{d}t} P(x,t)$ and performing a change of variables, we find,

$$\frac{\mathrm{d}\hat{x}(t)}{\mathrm{d}t} = \int \mathrm{d}x \int \mathrm{d}y [y-x] w(y|x) P(x,t). \tag{A.9}$$

Defining the kth jump moment as $a_k = \int (y-x)^k w(y|x) \mathrm{d}y$, the dynamics can be written as,

$$\frac{\mathrm{d}\hat{x}(t)}{\mathrm{d}t} = \langle a_1(x) \rangle. \tag{A.10}$$

It is by no means obvious if or when the deterministic path approximation $\langle a_1(x) \rangle \approx a_1(\langle x \rangle)$ is valid, as a_1 will often be nonlinear. The justification lies in the linear noise approximation. Proceeding as above, we also find an expression for the second moment,

[1] Indeed, they are equivalent if transition rates are linear – in which case the PDE is also exact.

$$\frac{\mathrm{d}\langle x^2(t)\rangle}{\mathrm{d}t} = 2\langle x a_1(x)\rangle + \langle a_2(x)\rangle. \tag{A.11}$$

Which again, we will only be able to solve by means of the linear noise approximation.

A.2.3 The linear noise approximation

To justify this step we will change into variables where we can have an explicit parameter ε that relates to the step size. The trait x is approximated by an average or macroscopic value ϕ and a deviation ξ that scales with the mutational step size ε; $x = \phi + \varepsilon\xi$. Defining $r \equiv y - x$ expand the transition rate $w(y|x)$ in powers of ε,

$$w(y|x) = f(\varepsilon)\left[\Phi_0(\varepsilon x; r) + \varepsilon\Phi_1(\varepsilon x; r) + \varepsilon^2\Phi_2 + \ldots\right], \tag{A.12}$$

where the Φ terms in the expansion are functions in which ε appears only in terms of εx. The function $f(\varepsilon)$ indicates that we can rescale the entire process by some arbitrary factor of ε since it can always be absorbed into the timescale. We can then define the transformed jump moments as moments of Φ rather than w,

$$\alpha_{\nu,\lambda}(X) = \int r^\nu \Phi_\lambda(X, r)\mathrm{d}r. \tag{A.13}$$

The probability $P(x,t)$ is expressed in terms of the new variables $P(\phi(t) + \varepsilon\xi, t) = \Pi(\xi, t)$, and the master equation (2.2) becomes:

$$\begin{aligned}
&\frac{\partial}{\partial\tau}\Pi(\xi,\tau) - \varepsilon^{-1}\frac{\mathrm{d}\phi}{\mathrm{d}\tau}\frac{\partial}{\partial\xi}\Pi(\xi,\tau) \\
&\qquad = -\varepsilon^{-1}\frac{\partial}{\partial\xi}\alpha_{1,0}(\phi(\tau)+\varepsilon\xi)\cdot\Pi(\xi,\tau) \\
&\qquad +\frac{1}{2}\frac{\partial^2}{\partial\xi^2}\alpha_{2,0}(\phi(\tau)+\varepsilon\xi)\cdot\Pi(\xi,\tau) \\
&\qquad -\frac{1}{3!}\varepsilon\frac{\partial^3}{\partial\xi^3}\alpha_{3,0}(\phi(\tau)+\varepsilon\xi)\cdot\Pi(\xi,\tau) \\
&\qquad +\varepsilon\frac{\partial}{\partial\xi}\alpha_{1,1}(\phi(\tau)+\varepsilon\xi)\cdot\Pi(\xi,\tau) + \mathcal{O}(\varepsilon^2),
\end{aligned} \tag{A.14}$$

where we have rescaled time by $\varepsilon^2 f(\varepsilon)t = \tau$. Expanding the jump moments around the macro-

scopic variable ϕ,

$$\alpha_{\nu,\lambda}(\phi(t)+\varepsilon\xi) \approx \alpha_{\nu,\lambda}(\phi) + \varepsilon\xi\alpha'_{\nu,\lambda} + \frac{1}{2}\varepsilon^2\xi^2\alpha_{\nu,\lambda}(\phi)'' + \mathcal{O}\varepsilon^3,$$

(where primes indicate derivatives with respect to ϕ), and collecting terms of leading order in ε we have:

$$\frac{\mathrm{d}\phi}{\mathrm{d}\tau} = \alpha_{1,0}(\phi), \tag{A.15}$$

which is a completely deterministic expression. Substituting the form of $w(y|x)$ from (A.4) recovers the canonical equation of adaptive dynamics, Eq. (2.3). Observe that the fluctuations are an order ε smaller, demonstrating that this is indeed a consistent approximation. Collecting terms of order ε^0 we have the partial differential equation

$$\frac{\partial}{\partial\tau}\Pi(\xi,\tau) = -\alpha'_{1,0}(\phi)\frac{\partial}{\partial\xi}\xi\Pi + \frac{1}{2}\alpha_{2,0}(\phi)\frac{\partial^2}{\partial\xi^2}\Pi, \tag{A.16}$$

while all other terms are order ε or smaller. This is a partial differential equation for the evolution of the probability distribution of traits. It is a linear Fokker-Planck equation, hence its solution is Gaussian and $\Pi(\xi,t)$ can be described to this order of accuracy by its first two moments,

$$\begin{aligned}\frac{\partial\langle\xi\rangle}{\partial t} &= \alpha'_{1,0}(\phi)\langle\xi\rangle,\\ \frac{\partial\langle\xi^2\rangle}{\partial t} &= 2\alpha'_{1,0}(\phi)\langle\xi^2\rangle + \alpha_{2,0}(\phi),\end{aligned}$$

where prime indicates derivative with respect to the trait x. If $\alpha'_{1,0}(\phi) < 0$ or the initial fluctuations $\langle\xi\rangle_0$ are zero, the first moment can be ignored, and the variance of the ensemble is given by transforming back into the original variables:

$$\varepsilon^2\frac{\partial\sigma^2}{\partial t} = 2\alpha'_{1,0}(\phi)\sigma^2 + \alpha_{2,0}(\phi). \tag{A.17}$$

Transforming between the scaled variables and the original variables requires the appropriate choice of ε. The assumption that mutational steps are small provides a natural choice: $\varepsilon = \sigma_\mu$. Substituting Eq. (A.4) to compute the jump moments, this recovers the fluctuation expression (2.4) in the text. Note that even before we perform this substitution that (A.17) has the same form as (2.4); fluctuations grow or diminish at a rate determined by the sign of the gradient of the deterministic equation, Eq. (A.15).

A.3 Chemostat Model

The graphical model for a second scenario is also displayed in Fig. 2.1. This scenario describes competition and evolution with explicit resource dynamics for a chemostat system. We consider a resource Q that flows in at rate D from a reservoir fixed at concentration Q_0. To retain constant volume in the chemostat, both biotic and biotic components are flushed from the system at constant rate D. The chemostat contains populations with N_i organisms, each of which take up nutrients at a rate g_i and convert these into reproductive output with efficiency η_i,

$$\dot{Q} = DQ_0 - DQ - \mathbf{g}(Q)\mathbf{N}, \tag{A.18}$$

$$\dot{N}_i = -DN_i + \eta_i g_i(Q)N_i. \tag{A.19}$$

Assume that the uptake of nutrients is governed by Michaelis-Mentin dynamics, and also that a trade-off exists between efficiency of nutrient take-up and conversion (imagining greater investment in foraging means less energy available for reproduction),

$$g(Q) = Q/(1 + hQ),$$

$$\eta(x) = x,$$

$$h(x) = x^2,$$

where the trait x may differ between populations.

The associated equilibrium population size for a population with trait x is

$$\bar{N}(x) = \frac{DQ_0 - D\bar{Q}}{g(\bar{Q}, x)} = xQ_0 - \frac{xD}{x - x^2D}.$$

Similarly, the resource uptake of a mutant with trait y in an environment at an equilibrium set by a resident type with trait x is given by

$$g_y(\bar{Q}_x) = \frac{1}{x/D - x^2 + y^2},$$

from which we find the invasion fitness function and its gradient,

$$s(y,x) = \frac{y}{x/D - x^2 + y^2} - D,$$
$$\partial_y s(y,x)|_{y=x} = \frac{D}{x} - 2D^2.$$

Assuming a Gaussian mutation kernel and constant mutation rate, from Eq. (A.4) the transition rate function $w(y|x)$ is

$$w(y|x) \approx \mu \left[xQ_0 - \frac{xD}{x - x^2 D} \right] \left[\frac{D}{x} - 2D^2 \right] \frac{e^{-\frac{(y-x)^2}{2\sigma_\mu^2}}}{\sqrt{2\pi\sigma_\mu^2}} [y - x]. \tag{A.20}$$

A.4 Branching Model

The model of implicit competition for a limiting resource, Eqs. (2.5)-(2.7), is well known to exhibit the phenomenon of evolutionary branching when the competition kernel is narrower than the resource distribution, $\sigma_c < \sigma_k$ (Dieckmann and Doebeli, 1999). Once branching occurs, the invasion fitness of a rare mutant is no longer given by $s(y,x)$ as in Eq. (A.5), but instead depends on the trait values of each of the coexisting residents x_1 and x_2, as in $s(y,x_1,x_2)$. This invasion fitness can still be calculated directly from the competition model, Eqs. (2.5)-(2.7). This mutant can either arise from the x_1 or x_2 population and replace it. If we assume $x_1 = -x_2 = x$, we have the case of a symmetrically branching population. While many realizations of branching may be close to symmetric, this is but a one-dimensional slice through a two-dimensional trait space (x_1, x_2). In this case, we can express the equilibrium density of each resident species by $K_{res}(x) = K(x)/(1 + C(x,-x))$. We then consider the initial per capita growth rate of a rare mutant with trait y:

$$s(y,x,-x) = r\left(1 - \frac{K_{\text{res}}(x)C(x,y) + K_{\text{res}}(-x)C(-x,y)}{K(y)}\right). \tag{A.21}$$

This replaces the $s(y,x)$ function for the monomorphic population, and we proceed as before to calculate $a_1(x)$ in Eq. (2.3). That is, we evaluate $\partial_y s(y,x,-x)$ at $y = x$, take $K_{\text{res}}(x) \equiv N^*(x)$ to recover a closed-form slice of the branching landscape that is depicted in Fig. 2.1(c). The expression itself is given in the Appendix A.5, Eq. (A.24).

A.5 Explicit solutions for examples

For the example logistic competition in a monomorphic population, the evolution of the mean trait x is given by:

$$\frac{\mathrm{d}x}{\mathrm{d}t} = -\frac{x}{2\sigma_k^2} r\mu\sigma_\mu^2 K_0 e^{-x/2\sigma_k^2}. \tag{A.22}$$

In Fig. 2.1, we choose parameters such that $r\mu K_0\sigma_\mu^2/2 = 1$ and $\sigma_k^2 = 1$.

For the chemostat:

$$\frac{\mathrm{d}x}{\mathrm{d}t} = \frac{1}{2}\mu\sigma_\mu^2 \left(Qx - \frac{D}{1-xD}\right)(D/x - 2D^2). \tag{A.23}$$

In Fig. 2.1, we choose parameters such that $\mu\sigma_\mu^2 = 1$, $D = 0.1$, and $Q = 0.1$.

And for the symmetrically dimorphic population,

$$\frac{\mathrm{d}x}{\mathrm{d}t} = \frac{-r\mu\sigma_\mu^2 K_0 e^{-\frac{x^2}{2}\left(\frac{4}{\sigma_c^2}-\frac{1}{\sigma_k^2}\right)}(\sigma_c^2 + e^{\frac{2x^2}{\sigma_c^2}}\sigma_c^2 - 2\sigma_k^2)x}{\left(1+e^{2x^2/\sigma_c^2}\right)^2 \sigma_c^2\sigma_k^2}. \tag{A.24}$$

The parameters for the dimorphic population plot of Fig. 2.1 are chosen such that $r\mu K_0\sigma_\mu^2/4 = 1$, $\sigma_k^2 = 2$ and $\sigma_c^2 = 1$.

The right-hand side of each of these expressions is the function $a_1(x)$ from the text, Eq. (2.3). This function also determines the fluctuation dynamics by way of Eq. (2.4). The respective plots of $a_1(x)$ are used in Fig. 2.1, while solving the ODEs for $\hat{x}(t)$ and $\sigma^2(t)$ gives the theoretical predictions of Fig. 2.2.

Bibliography

Abrams, P., 1993. Evolutionarily unstable fitness maxima and stable fitness minima of continuous traits. Evolutionary Ecology 7, 467–487.

Bellwood, D. R., Hughes, T. P., Folke, C., Nyström, M., Jun. 2004. Confronting the coral reef crisis. Nature 429 (6994), 827–33.
URL http://www.ncbi.nlm.nih.gov/pubmed/15215854

Berkes, F., Hughes, T. P., Steneck, R. S., Wilson, J. A., Bellwood, D. R., Crona, B., Folke, C., Gunderson, L. H., Leslie, H. M., Norberg, J., Nyström, M., Olsson, P., Osterblom, H., Scheffer, M., Worm, B., Mar. 2006. Ecology. Globalization, roving bandits, and marine resources. Science (New York, N.Y.) 311 (5767), 1557–8.
URL http://www.ncbi.nlm.nih.gov/pubmed/16543444

Bestelmeyer, B. T., Ellison, A. M., Fraser, W. R., Gorman, K. B., Holbrook, S. J., Laney, C. M., Ohman, M. D., Peters, D. P. C., Pillsbury, F. C., Rassweiler, A., Schmitt, R. J., Sharma, S., Dec. 2011. Analysis of abrupt transitions in ecological systems. Ecosphere 2 (12), art129.
URL http://www.esajournals.org/doi/abs/10.1890/ES11-00216.1

Biggs, R., Carpenter, S. R., Brock, W. A., Jan. 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), 826–31.
URL http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2630060&tool=pmcentrez&rendertype=abstract

Black, A. J., McKane, A. J., Mar. 2012. Stochastic formulation of ecological models and their applications. Trends in ecology & evolution, 1–9.
URL http://www.ncbi.nlm.nih.gov/pubmed/22406194

Blomberg, S., Garland, J. T., Ives, A., 2003. Testing for phylogenetic signal in comparative data: behavioral traits are more labile. Evolution 57 (4), 717–745.
URL http://www3.interscience.wiley.com/journal/118867878/abstract

Boettiger, C., Coop, G., Ralph, P., Jul. 2012. Is your phylogeny informative? Measuring the power of comparative methods. Evolution 66 (7), 2240–51.
URL http://doi.wiley.com/10.1111/j.1558-5646.2011.01574.xhttp://www.ncbi.nlm.nih.gov/pubmed/22759299

Boettiger, C., Dushoff, J., Weitz, J. S., 2010. Fluctuation domains in adaptive evolution. Theoretical Population Biology 77 (1), 6–13.
URL http://www.ncbi.nlm.nih.gov/pubmed/19843477

Boettiger, C., Hastings, A., May 2012. Quantifying limits to detection of early warning for critical transitions. Journal of The Royal Society Interface 9 (75), 2527 – 2539.
URL http://rsif.royalsocietypublishing.org/cgi/doi/10.1098/rsif.2012.0125

Burnham, K. P., Anderson, D. R., 2002. Model Selection and Multi-Model Inference. Springer.
URL http://www.amazon.com/Selection-Multi-Model-Inference-Kenneth-Burnham/dp/0387953647

Butler, M. A., King, A. A., Dec. 2004. Phylogenetic Comparative Analysis: A Modeling Approach for Adaptive Evolution. The American Naturalist 164 (6), 683–695.
URL http://www.jstor.org/stable/10.1086/426002

Carpenter, J., Feb. 2011. May the Best Analyst Win. Science 331 (6018), 698–699.
URL http://www.sciencemag.org/cgi/doi/10.1126/science.331.6018.698

Carpenter, S., Brock, W., Jul. 2011. Early warnings of unknown nonlinear shifts: a nonparametric approach. Ecology, 110729132431005.
URL http://www.esajournals.org/doi/abs/10.1890/11-0716.1

Carpenter, S. R., Brock, W. A., 2006. Rising variance: a leading indicator of ecological transition. Ecology letters 9 (3), 311–8.
URL http://www.ncbi.nlm.nih.gov/pubmed/16958897

Champagnat, N., Ferriere, R., G., B. A., 2001. The canonical equation of adaptive dynamics: a mathematical view. Selection 1–2, 73–83.

Champagnat, N., Ferrière, R., Meleard, S., 2006. Unifying evolutionary dynamics: From individual stochastic processes to macroscopic models. Theor. Pop. Biol. 69, 297–321.

Champagnat, N., Lambert, A., 2007. Evolution of discrete populations and the canonical equation of adaptive dynamics. Ann. Appl. Prob. 17, 102–155.

Chevin, L., Hospital, F., 2008. Selective sweep at a quantitative trait locus in the presence of background genetic variation. Genetics 180, 1645–1660.

Contamin, R., Ellison, A. M., Apr. 2009. Indicators of regime shifts in ecological systems: what do we need to know and when do we need to know it? Ecological applications : a publication of the Ecological Society of America 19 (3), 799–816.
URL http://www.ncbi.nlm.nih.gov/pubmed/19425440

Cooper, T., Rozen, D., Lenski, R., 2003. Parallel changes in gene expression after 20,0000 generations of evolution in *Escherichia coli*. Proceedings of the National Academy of Sciences, USA 100, 1072–7.

Cox, D. R., 1961. Tests of Seperate Families of Hypotheses. In: Proceedings of the 4th Berkeley Symposium, University of California Press. No. 2. pp. 105 – 123.

Cox, D. R., 1962. Further results on tests of separate families of hypotheses. Journal of the Royal Stastical Society 24 (2), 406–424.
URL http://www.jstor.org/stable/2984232

Dakos, V., Kéfi, S., Rietkerk, M., Nes, E. H. V., Scheffer, M., 2011a. Slowing Down in Spatially Patterned Ecosystems at the Brink of Collapse. The American Naturalist.

Dakos, V., Nes, E. H., Donangelo, R., Fort, H., Scheffer, M., Nov. 2009. Spatial correlation as leading indicator of catastrophic shifts. Theoretical Ecology, 163–174.
URL http://www.springerlink.com/index/10.1007/s12080-009-0060-6

Dakos, V., Scheffer, M., van Nes, E. H., Brovkin, V., Petoukhov, V., Held, H., Sep. 2008. Slowing down as an early warning signal for abrupt climate change. Proceedings of the National Academy of Sciences 105 (38), 14308–12.
URL http://www.ncbi.nlm.nih.gov/pubmed/18787119

Dakos, V., van Nes, E. H., D'Odorico, P., Scheffer, M., Oct. 2011b. Robustness of variance and autocorrelation as indicators of critical slowing down. Ecology, 111018130520007.
URL http://www.esajournals.org/doi/pdf/10.1890/11-0889.1http://www.esajournals.org/doi/abs/10.1890/11-0889.1

Diciccio, T. J., Efron, B., 1996. Bootstrap Confidence Intervals. Statistical Science 11 (3), 189–212.

Dieckmann, U., Doebeli, M., 1999. On the origin of species by sympatric speciation. Nature 400, 354–357.

Dieckmann, U., Law, R., 1996. The dynamical theory of coevolution: a derivation from stochastic ecological processes. J. Math. Biol. 34, 579–612.

Ditlevsen, P. D., Johnsen, S. J., Oct. 2010. Tipping points: Early warning and wishful thinking. Geophysical Research Letters 37 (19), 2–5.
URL http://www.agu.org/pubs/crossref/2010/2010GL044486.shtml

Drake, J. M., Griffen, B. D., Sep. 2010. Early warning signals of extinction in deteriorating environments. Nature 467 (7314), 456–459.
URL http://www.nature.com/doifinder/10.1038/nature09389

Duffy, M., Sivars-Becker, L., 2007. Rapid evolution and ecological host-parasite dynamics. Ecology Letters 10, 44–53.

Easterling, D. R., Peterson, T. C., Apr. 1995. A new method for detecting undocumented discontinuities in climatological time series. International Journal of Climatology 15 (4), 369–377.
URL http://doi.wiley.com/10.1002/joc.3370150403

Edwards, A. W., Cavalli-Sforza, L. L., 1964. Reconstruction of evolutionary trees. In: V. H. Heywood, McNeill, J. (Eds.), Phenetic and Phylogenetic Classification. Systematists Association, London, pp. 67–76.

Efron, B., 1987. Better bootstrap confidence intervals. Journal of the American Statistical Association 82 (397), 171– 185.
URL http://www.jstor.org/stable/2289144

Feller, W., 1968. An Introduction to Probability Theory and Its Applications. Vol. 1. Wiley, New York.

Felsenstein, J., Jan. 1985. Phylogenies and the Comparative Method. The American Naturalist 125 (1), 1–15.
URL http://www.journals.uchicago.edu/doi/abs/10.1086/284325

Folke, C., Carpenter, S. R., Walker, B., Scheffer, M., Elmqvist, T., Gunderson, L., Holling, C., Dec. 2004. REGIME SHIFTS, RESILIENCE, AND BIODIVERSITY IN ECOSYSTEM MANAGEMENT. Annual Review of Ecology, Evolution, and Systematics 35 (1), 557–581.
URL http://www.annualreviews.org/doi/abs/10.1146/annurev.ecolsys.35.021103.105711

Freckleton, R. P., Pagel, M., Harvey, P. H., Dec. 2002. Phylogenetic analysis and comparative data: a test and review of evidence. The American Naturalist 160 (6), 712–26.
URL http://www.ncbi.nlm.nih.gov/pubmed/18707460

Gardiner, C., 2009. Stochastic Methods: A Handbook for the Natural and Social Sciences (Springer Series in Synergetics). Springer.
URL http://www.amazon.com/Stochastic-Methods-Handbook-Sciences-Synergetics/dp/3540707123

Garland, T., Dickerman, a. W., Janis, C. M., Jones, J. a., Sep. 1993. Phylogenetic Analysis of Covariance by Computer Simulation. Systematic Biology 42 (3), 265–292.
URL http://sysbio.oxfordjournals.org/cgi/doi/10.1093/sysbio/42.3.265

Gentleman, R., Lang, D. T., 2007. Statistical analyses and reproducible research.

Geritz, S., Gyllenberg, M., Jacobs, F., Parvinen, K., 2002. Invasion dynamics and attractor inheritance. Journal of Mathematical Biology 44, 548–560.

Geritz, S. A. H., 2004. Resident-invader dynamics and the coexistence of similar strategies. J. Math. Biol. 50, 67–82.

Geritz, S. A. H., Kisdi, E., Meszéna, G., Metz, J. A. J., 1998. Evolutionarily singular strategies and the adaptive growth and branching of the evolutionary tree. Evol. Ecol. 12, 35–57.

Geritz, S. A. H., Metz, J. A. J., Kisdi, E., Meszéna, G., 1997. The dynamics of adaptation and evolutionary branching. Phys. Rev. Lett. 78, 2024–2027.

Gillespie, D., 1977. Exact stochastic simulation of coupled chemical reactions. Journal of Physical Chemistry 81.25, 2340–2361.

Gittleman, J. L., Kot, M., 1990. Adaptation : Statistics and a Null Model for Estimating Phylogenetic Effects. Society of Systematic Biologists 39 (3), 227–241.

Goldman, N., Feb. 1993. Statistical tests of models of DNA substitution. Journal of Molecular Evolution 36 (2), 182–198.
URL http://www.springerlink.com/index/10.1007/BF00166252

Gorman, G. C., Kim, Y. J., Mar. 1976. Anolis Lizards of the Eastern Caribbean: A Case Study in Evolution. II. Genetic Relationships and Genetic Variation of the Bimaculatus Group. Systematic Zoology 25 (1), 62–77.
URL http://sysbio.oxfordjournals.org/cgi/content/abstract/25/1/62http://www.jstor.org/stable/2412779?origin=crossref

Green, D. M., Swets, J. A., 1989. Signal Detection Theory and Psychophysics. Peninsula Pub.
URL http://www.amazon.com/Signal-Detection-Theory-Psychophysics-Marvin/dp/0932146236

Green, P. J., 1995. Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. Biometrika 82 (4), 711–732.
URL http://biomet.oxfordjournals.org/cgi/doi/10.1093/biomet/82.4.711

Guckenheimer, J., Holmes, P., 1983. Nonlinear Oscillations, Dynamical Systems, and Bifurcations of Vector Fields (Applied Mathematical Sciences Vol. 42). Springer.
URL http://www.amazon.com/Nonlinear-Oscillations-Dynamical-Bifurcations-Mathematical/dp/0387908196

Guttal, V., Jayaprakash, C., May 2008a. Changing skewness: an early warning signal of regime shifts in ecosystems. Ecology letters 11 (5), 450–60.
URL http://www.ncbi.nlm.nih.gov/pubmed/18279354

Guttal, V., Jayaprakash, C., Dec. 2008b. Spatial variance and spatial skewness: leading indicators of regime shifts in spatial ecological systems. Theoretical Ecology 2 (1), 3–12.
URL http://www.springerlink.com/index/10.1007/s12080-008-0033-1

Hall, P., Wilson, S., 1991. Two guidelines for bootstrap hypothesis testing. Biometrics 47 (2), 757–762.
URL http://www.jstor.org/stable/2532163

Hansen, T. F., 1997. Stabilizing selection and the comparative analysis of adaptation. Evolution 51 (5), 1341– 1351.
URL http://www.jstor.org/stable/2411186

Hansen, T. F., Martins, E. P., 1996. Translating between microevolutionary process and macroevolutionary patterns: the correlation structure of interspecific data. Evolution 50, 1404–1417.

Harmon, L. J., Losos, J. B., Davies, J. T., Gillespie, R. G., Gittleman, J. L., Bryan Jennings, W., Kozak, K. H., McPeek, M. A., Moreno-Roark, F., Near, T. J., Purvis, A., Ricklefs, R. E., Schluter, D., Schulte Ii, J. A., Seehausen, O., Sidlauskas, B. L., Torres-Carvajal, O., Weir, J. T., Mooers, A. O., Aug. 2010. Early bursts of body size and shape evolution are rare in comparative data. Evolution 64 (8), 2385–96.
URL http://www3.interscience.wiley.com/journal/123397103/abstracthttp://www.ncbi.nlm.nih.gov/pubmed/20455932

Harmon, L. J., Weir, J. T., Brock, C. D., Glor, R. E., Challenger, W., 2008. Geiger: investigating evolutionary radiations. Bioinformatics 24 (1), 129–131.

Hastings, A., Jan. 1991. Structured models of metapopulation dynamics. Biological Journal of the Linnean Society 42 (1-2), 57–71.
URL http://doi.wiley.com/10.1111/j.1095-8312.1991.tb00551.x

Hastings, A., Wysham, D. B., 2010. Regime shifts in ecological systems can occur with no warning. Ecology letters.
URL http://www.ncbi.nlm.nih.gov/pubmed/20148928

Held, H., 2004. Detection of climate system bifurcations by degenerate fingerprinting. Geophysical Research Letters 31 (23), 1–4.
URL http://www.agu.org/pubs/crossref/2004/2004GL020972.shtml

Hofbauer, J., Sigmund, K., 1998. Evolutionary Games and Population Dynamics. Cambridge University Press, Cambridge, U.K.

Holling, C. S., Nov. 1973. Resilience and Stability of Ecological Systems. Annual Review of Ecology and Systematics 4 (1), 1–23.
URL http://arjournals.annualreviews.org/doi/abs/10.1146/annurev.es.04.110173.000245

Huelsenbeck, J. P., Bull, J. J., Mar. 1996. A Likelihood Ratio Test to Detect Conflicting Phylogenetic Signal. Systematic Biology 45 (1), 92–98.
URL http://sysbio.oxfordjournals.org/cgi/content/abstract/45/1/92

Inman, M., Jun. 2011. Sending out an SOS. Nature Climate Change, 1–4.
URL http://www.nature.com/doifinder/10.1038/nclimate1146

Kampen, N. V., 2007. Stochastic Processes in Physics and Chemistry, Third Edition (North-Holland Personal Library). North Holland.
URL http://www.amazon.com/Stochastic-Processes-Chemistry-North-Holland-Personal/dp/0444529659

Kéfi, S., Rietkerk, M., Alados, C. L., Pueyo, Y., Papanastasis, V. P., Elaich, A., de Ruiter, P. C., Sep. 2007. Spatial vegetation patterns and imminent desertification in Mediterranean arid ecosystems. Nature 449 (7159), 213–7.
URL http://www.ncbi.nlm.nih.gov/pubmed/17851524

Keller, R. P., Lodge, D. M., Lewis, M. A., Shogren, J. F. (Eds.), 2009. Bioeconomics of Invasive Species: Integrating Ecology, Economics, Policy, and Management. Oxford University Press, USA.
URL http://www.amazon.com/Bioeconomics-Invasive-Species-Integrating-Management/dp/0195367979

Kimura, M., 1968. Evolutionary rate at the molecular level. Nature 217, 624–626.

Kimura, M., 1984. The neutral theory of molecular evolution. Cambridge University Press, Cambridge, UK.

Kuehn, C., Jan. 2011. A mathematical framework for critical transitions: normal forms, variance and applications. arxiv, 55.
URL http://arxiv.org/abs/1101.2908

Kurtz, T., 1971. Limit theorems for sequences of jump Markov processes approximating ordinary differential equations. Journal of Applied Probability 8, 344–356.

Lade, S. J., Gross, T., Feb. 2012. Early Warning Signals for Critical Transitions: A Generalized Modeling Approach. PLoS Computational Biology 8 (2), e1002360.
URL http://dx.plos.org/10.1371/journal.pcbi.1002360

Lande, R., 1979. quantitative genetic analysis of multivariate evolution, applied to brain: body size allometry. Evolution 33, 402–416.

Lande, R., Arnold, S. J., 1983. The measurement of selection on correlated characters. Evolution 37, 1210–1226.

Lazell, J. D., 1972. The Anoles (Sauria, Iguanidae) of the Lesser Antilles. Vol. 143. Harvard University, Cambridge.

Lenton, T. M., Livina, V., Dakos, V., 2012. Early warning of climate tipping points from critical slowing down: comparing methods to improve robustness. Philosophical Transactions of The Royal Society A (in press).
URL http://sites.google.com/site/noraclpreprints/contributions/lenton-et-al/Early_warning_methods_revised_with_figs_and_appendix2.pdf

Lenton, T. M., Myerscough, R. J., Marsh, R., Livina, V. N., Price, A. R., Cox, S. J., Genie Team, Mar. 2009. Using GENIE to study a tipping point in the climate system. Philosophical transactions. Series A, Mathematical, physical, and engineering sciences 367 (1890), 871–84.
URL http://www.ncbi.nlm.nih.gov/pubmed/19087945

Levins, R., 1962. Theory of fitness in a heterogeneous environment. i. the fitness set and adaptive function. The American Naturalist 96 (891), 361–373.

Levins, R., 1964. Theory of fitness in a heterogeneous environment. iii. the response to selection. Journal of Theoretical Biology 7, 224–240.

Levins, R., 1966. The strategy of model building in population biology. American Scientist 54 (4), 421–431.
URL http://earthweb.ess.washington.edu/roe/Knowability_590/Week3/Levins_1966.pdf

Livina, V., Ditlevsen, P., Lenton, T., Feb. 2012. An independent test of methods of detecting system states and bifurcations in time-series data. Physica A: Statistical Mechanics and its Applications 391 (3), 485–496.
URL http://linkinghub.elsevier.com/retrieve/pii/S0378437111006534

Losos, J. B., 1990. A Phylogenetic Analysis of Character Displacement in Caribbean Anolis Lizards. Evolution 44 (3), 558–569.

Losos, J. B., Jun. 2011. Seeing the forest for the trees: the limitations of phylogenies in comparative biology (american society of naturalists address) *. The American naturalist 177 (6), 709–27.
URL http://www.ncbi.nlm.nih.gov/pubmed/21597249

May, R. M., Oct. 1977. Thresholds and breakpoints in ecosystems with a multiplicity of stable states. Nature 269 (5628), 471–477.
URL http://www.nature.com/doifinder/10.1038/269471a0

McLachlan, G. J., 1987. On Bootstrapping the Likelihood Ratio Test Stastistic for the Number of Components in a Normal Mixture. Applied Statistics 36 (3), 318.
URL http://www.jstor.org/stable/2347790

Mumby, P. J., Hastings, A., Edwards, H. J., Nov. 2007. Thresholds and the resilience of Caribbean coral reefs. Nature 450 (7166), 98–101.
URL http://www.ncbi.nlm.nih.gov/pubmed/17972885

Neyman, J., Pearson, E., 1933. On the problem of the most efficient tests of statistical hypotheses. Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character 231 (694-706), 289–337.
URL http://rsta.royalsocietypublishing.org/content/231/694-706/289.full.pdf

Nowak, M. A., Sigmund, K., 2004. Evolutionary dynamics of biological games. Science 303, 793–9.

Noy-Meir, I., 1975. Stability of grazing systems: an application of predator-prey graphs. The Journal of Ecology 63 (2), 459–481.
URL http://www.jstor.org/stable/10.2307/2258730

Ohta, T., 2002. Near-neutrality in revolution of genes and gene regulation. Proc. Natl. Acad. Sci. 99, 16134–7.

Olsen, E., Lilly, G., Heino, M., Morgan, M., Brattey, J., Dieckmann, U., 2005. Assessing changes in age and size at maturation in collapsing populations of atlantic cod (*Gadus morhua*). Canadian Journal of Fisheries and Aquatic Sciences 62, 811–823.

O'Meara, B. C., Ané, C., Sanderson, M. J., Wainwright, P. C., May 2006. Testing for different rates of continuous trait evolution using likelihood. Evolution 60 (5), 922–33.
URL http://www.ncbi.nlm.nih.gov/pubmed/16817533

Pagel, M., 1994. Detecting Correlated Evolution on Phylogenies: A General Method for the Comparative Analysis of Discrete Characters. Proceedings of The Royal Society B 255, 37–45.

Pagel, M., Oct. 1999. Inferring the historical patterns of biological evolution. Nature 401 (6756), 877–84.
URL http://www.ncbi.nlm.nih.gov/pubmed/10553904

Petit, J. R., Jouzel, J., Raynaud, D., Barkov, N. I., Barnola, J., Basile, I., Bender, M., Chappellaz, J., Davis, M., Delaygue, G., Delmotte, M., Kotlyakov, V. M., Legrand, M., Lipenkov, V. Y., Lorius, C., Pepin, L., Ritz, C., Saltzman, E., Stievenard, M., 1999. Climate and atmospheric history of the past 420,000 years from the Vostok ice core, Antarctica. Nature 399 (6735), 429–436.
URL http://www.daycreek.com/dc/images/1999.pdf

Price, T., 1997. Correlated evolution and independent contrasts. Proceedings of The Royal Society B 352 (1352), 519–529.
URL http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1691942

Revell, L. J., Dec. 2010. Phylogenetic signal and linear regression on species data. Methods in Ecology and Evolution 1 (4), 319–329.
URL http://blackwell-synergy.com/doi/abs/10.1111/j.2041-210X.2010.00044.xhttp://doi.wiley.com/10.1111/j.2041-210X.2010.00044.x

Revell, L. J., Collar, D. C., 2009. Phylogenetic analysis of the evolutionary correlation using likelihood. Evolution 63, 1090–1100.

Revell, L. J., Harmon, L. J., 2008. Testing quantitative genetic hypotheses about the evolutionary rate matrix for continuous characters. Evolutionary Ecology Research 10 (3), 311–331.
URL http://anolis.oeb.harvard.edu/~liam/pdfs/Revell_and_Harmon_2008.EER.pdf

Rodionov, S. N., 2004. A sequential algorithm for testing climate regime shifts. Geophysical Research Letters 31 (9), 2–5.
URL http://www.agu.org/pubs/crossref/2004/2004GL019448.shtml

Rueffler, C., Dooren, T. V., Metz, J., 2004. Adaptive walks on changing landscapes: Levins' approach extended. Theoretical Population Biology 65, 165–178.

Scheffer, M., Sep. 2010. Complex systems: Foreseeing tipping points. Nature 467 (7314), 411–2.
URL http://www.ncbi.nlm.nih.gov/pubmed/20864992

Scheffer, M., Bascompte, J., Brock, W. A., Brovkin, V., Carpenter, S. R., Dakos, V., Held, H., van Nes, E. H., Rietkerk, M., Sugihara, G., 2009. Early-warning signals for critical transitions. Nature 461 (7260), 53–9.
URL http://www.ncbi.nlm.nih.gov/pubmed/19727193

Scheffer, M., Carpenter, S. R., Foley, J. A., Folke, C., Walker, B., Oct. 2001. Catastrophic shifts in ecosystems. Nature 413 (6856), 591–6.
URL http://www.ncbi.nlm.nih.gov/pubmed/11595939

Schneider, C. J., Losos, J. B., Queiroz, K. D. E., Journal, S., Mar, N., 2001. Evolutionary Relationships of the Anolis bimaculatus Group from the Northern Lesser Antilles. Journal of Herpetology 35 (1), 1–12.

Schreiber, S., 2003. Allee effects, extinctions, and chaotic transients in simple population models. Theoretical population biology 64 (2), 201–209.
URL http://linkinghub.elsevier.com/retrieve/pii/S0040580903000728

Schreiber, S. J., Rudolf, V. H. W., Jun. 2008. Crossing habitat boundaries: coupling dynamics of ecosystems through complex life cycles. Ecology letters 11 (6), 576–87.
URL http://www.ncbi.nlm.nih.gov/pubmed/18371091

Schwab, M., Karrenbach, M., Claerbout, J., 2000. Making scientific computations reproducible. Science and Engineering 2 (6), 61–67.

Seekell, D. a., Carpenter, S. R., Pace, M. L., Oct. 2011. Conditional heteroscedasticity as a leading indicator of ecological regime shifts. The American naturalist 178 (4), 442–51.
URL http://www.ncbi.nlm.nih.gov/pubmed/21956023

Stenson, A. G., Thorpe, R. S., Malhotra, A., Jul. 2004. Evolutionary differentiation of bimaculatus group anoles based on analyses of mtDNA and microsatellite data. Molecular phylogenetics and evolution 32 (1), 1–10.
URL http://www.ncbi.nlm.nih.gov/pubmed/15186792

Stodden, V., 2009. The legal framework for reproducible scientific research: Licensing and copyright. Computing in Science and Engineering 11 (1), 35–40.

van Kampen, N., 2001. Stochastic Processes in Physics and Chemistry. Elsevier Science.

Wichman, H., Badgett, M., Scott, L., Boulianne, C., Bull, J., 1999. Different trajectories of parallel evolution during viral adaptation. Science 285, 422–424.

Wissel, C., Dec. 1984. A universal law of the characteristic return time near thresholds. Oecologia 65 (1), 101–107.
URL http://www.springerlink.com/index/10.1007/BF00384470

Wright, S., 1931. Evolution in mendelian populations. Genetics 16, 97âĂŞ159.

Wright, S., 1932. The roles of mutation, inbreeding, crossbreeding and selection in evolution. In: Proceedings of the 6th International Congress of Genetics. Vol. 1. pp. 356–66.