Regime Detection Measures for the Practical Ecologist

A Thesis
Presented to
The Division of
University of Nebraska-Lincoln

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

Jessica L. Burnett

2019

Approved for the Division
(School of Natural Resources)

Craig R. Allen

Dirac Twidwell

Table of Contents

List of Tables

List of Figures

Abstract

Identifying abrupt changes in the structure and functioning of systems, or system regime shifts, in ecological and social-ecological systems leads to an understanding of relative and absolute system resilience. Resilience is an emergent phenomenon of complex social-ecological systems, and is the ability of a system to absorb disturbance without reorganizing into a new state, or regime. Resilience science provides a framework and methodology for quantitatively assessing the capacity of a system to maintain its current trajectory (or to stay within a certain, and often desirable regime). If and when a system's resilience is exceeded, it crosses a threshold and enters into an alternate regime (or undergoes a regime shift).

I will use Fisher Information to detect regime shifts in time and space using avian community data obtained from the North American Breeding Bird Survey within the area east of the Rockies and west of the Mississippi River. Fisher Information is a technique that captures the dynamic of a system, and this metric will be calculated about a suite of bird species abundances aggregated to the route level for all possible time periods. Transmutation (aggregation error) about inclusion or exclusion of certain bird species, functional groups, and guilds will be analyzed. Efforts have been made to develop early warning indicators of regime shifts in ecosystems, however, for most ecosystems there is great uncertainty in predicting the risk of a regime shift, regarding both when and how long it will take to happen and if it can be recognized early enough to be avoided when desired. We will complement the use of Fisher

Information with multiple discontinuity analyses about body mass distributions at the route-level to achieve the aim of identifying individual species that best serve as early-warning indicators of regime shifts. For those species found on the edges of body mass aggregations, we test the hypothesis that the background variance in their abundances (on Breeding Bird Survey routes) will increase more than those not observed at the edge of discontinuity aggregations. Identification of early-warning indicators of regime shifts in ecological systems allows management efforts to focus on a single or a small number of species that inform us about ecosystem resilience and trajectory.

These methods transcend the primary objective of the Breeding Bird Survey (to monitor population trends) and use this expansive dataset in such a way that information about ecosystem order, trajectory, and resilience emerge. Here, we utilize an expansive dataset (the Breeding Bird Survey) to make broad-scale estimations and predictions about ecosystem resilience, regime status and trajectory, and ecosystem sustainability. Identification of regime shifts and early-warning indicator species may afford us the ability to predict system regime shifts in time.

Table of Definitions

Research surrounding regime shifts, threshold identification, change-point detection, bifurcation theory, etc. is muddled with jargon. Here, I provide a table of definitions (Table 1) for terms and concepts that may either be unfamiliar to the practical ecologist, or may have multiple meanings among and within ecological researchers and practitioners. With this table, I aim to both improve the clarity of this dissertation *and* highlight one potential issue associated with regime detection methods in ecology: semantics.

Table 1: A table of definitions for terms, theories, and phrases often appearing in ecological regime shift literature.

Term	**Definition**	**Synonyms**
Abrupt	A relative value of the speed and/or intensity of the change; the time period over which the regime shift occurs relative to the time observed (or expected to have been) in a particular state.	big, fast, quick, large
Alternative Stable State	**Controversially can be distilled as one of either: the number of unique stable configurations that a system can adopt (see Lewontin 1969), or the impacts that processes or pressures can have on a system's state (see May 1977).**	
Attractor	The set of values towards which a system tends regardless of its initial (starting) vaules.	

Table 1: A table of definitions for terms, theories, and phrases often appearing in ecological regime shift literature. *(continued)*

Term	Definition	Synonyms
Basin-Boundary Collision	**The parameter values for a system that causes the system to shift between alternate attractors.**	**non-local bifurcation**
Catastrophe Theory	The study of abrupt changes within a dynamical system.	
Catastrophic Bifurcation	**A relatively abrupt jump to an alternate attractor due to initial attractor.**	
Change-Point	See also 'Regime Shift'. A term often used in computer science, climatology, data science; represents the point at which a state changes its configuration.	
Change-Point Detection	**A change point method which does not require supervision; identifies potential change points without a priori potential change points.**	
Change-Point Estimation	A change point method which DOES require supervision; identifies potential change points when given a set of potential change points; well-developed in computer science, statistics, data mining, etc.; although well-developed, still lacks with giving statistical significance of change-points.	
Chaos	**A system with extreme sensitivity to initial conditions.**	
Critical Slowing Down (CSD)	When the recovery rate (time to return) of a system decreases (approaches zero) as a system approaches a critical point (possibly a threshold or tipping point). A characteristic observed in some empirical systems data (e.g. nutrient loading in shallow lakes).	
Degrees of Freedom	**The number of system parameters or components which vary independently.**	

Table 1: A table of definitions for terms, theories, and phrases often appearing in ecological regime shift literature. *(continued)*

Term	**Definition**	**Synonyms**
Domain of Attraction	The range of values around which a system fluctuates.	zone of fluctuation, basin of attraction, stable point, attractor
Driver	**A widespread anthropogenic source of change which leads to one or more pressures (e.g., land-use change).**	
Driver-Threshold Regime Shift	When a rapid change in external driver indcues a rapid change in ecosystem state.	
Dynamical System	**A time-dependent system which can be described in state-space.**	
Dynamical Systems Theory	The study of complex systems theory; the study of time-dependent systems.	
Equilibrium	**The set of values around which a system revolves and does not change.**	
Exogeneous Process (Forcing, Driver)	An external process influencing the state of the dynamical system.	
First-Order Stationarity	**When the mean is constantant over the observations.**	
Fold Bifurcation	This occurs when a stable point collides with an unstable point; when crossing a tipping point induces hysteresis.	
Fractal Properties	**A measurement of geometrical self-similarity; when a system has similar structure regardless of the scale of observation.**	**ergodic**

Table 1: A table of definitions for terms, theories, and phrases often appearing in ecological regime shift literature. *(continued)*

Term	**Definition**	**Synonyms**
Hysteresis	A system which is state-dependent (e.g. magnets); when a tipping point or threshold is crossed such that the previous state cannot be achieved by reversing the conditions.	
Leading Indicators	**When the statistical properties of the fluctuations (of the data) approach a critical transition.**	
Lyapunov Exponent (and Stability	A value that conveys the average rate of trajectory divergence that is caused by an endogenous force; how quickly (if at all) a system will tend away from a stable point if it starts near the stable point.	
Measure Theory	**The study of measures and measurement (e.g. volume, mass, time).**	
Moving (Sliding) Window Analysis	When a subsample of the data $$X_t$$ is used in lieu of a single observation, $$x_t$$.	
Noise	**Processes manifested in data which are unaccounted for; sometimes referred to as meaningless; random variability.**	
Non-Stationarity of the Mean Value	Infers that a trend or a periodicity is present in the time series.	
Online	**Real-time updating of model parameters, predictions, etc. (c.f. offline).**	
Persistent	A relative value of the longevity of the observed change in values.	long-lasting
Phase Space	**A graphical representation of two or more trajectories where one axis is not time. In this representation an equilibrium is defined as a single point in the state space.**	

Table 1: A table of definitions for terms, theories, and phrases often appearing in ecological regime shift literature. *(continued)*

Term	**Definition**	**Synonyms**
Prediction	A temporal forecast. Is intrinsic when a model and paramters are used to make forecast, is realized when the prediction becomes the actual state of the system.	
Pressure	**A perturbation which negatively influences a system, and can be defined as pulse, press, or monotonic.**	
Red Noise	Noise having zero mean, constant variance, and serial autocorrelation; autocorrelated random variability.	
Regime	**A set of system values that define a particular system state. Not necessarily stable, but some state variables or outputs of the system remain relatively constant over a defined period of time.**	
Regime Shift	"abrupt" and "persistent" change in a system's structure or functioning.	
Second-Order Stationarity	**The nean is constant and the covariance is a function of a time lag, but not of time.**	
Self-Similarity	A system satisfied by power-law scaling.	
Stable Equilibrium	**An equilibrium is stable when small perturbations do not induce change.**	
State Space	The set of all possible configurations of a system.	
State-Threshold Regime Shift	**When a graduaal change in external driver induces a rapid change in ecosystem state (e.g,. System crosses a threshold).**	
Stationarity	When the probability density function of a system does not change with time.	

Table 1: A table of definitions for terms, theories, and phrases often appearing in ecological regime shift literature. *(continued)*

Term	Definition	Synonyms
Statistical Stationarity	**A system with statistical properties unchanging over time. This concept extends to periodic stationarity for systems exhibiting periodic behavior.**	
Strange Attractor	An attractor which has fractal structure (an observable fractal dimension).	
Supervised Machine Learning	**When classifiers are used to train the data a priori.**	
System State	The observed (current) instance of the system within a state space.	
Threshold	**A point where the system reacts to changing conditions.**	
Tipping Point	A point in a system's trajectory where a small change in an endogenous force induces a large change in sytem state or values; the point where a system can flip into an alternative state.	
Trajectory	**The path of an object or system through space-time.**	**orbit, path**
Transient	A behavior or phenomenon which is responsive to intial (starting) conditions, or its effect declines over time.	
Trend Smoothing	**Local averaging of values such that the non-systematic components of the system are washed out.**	
Unstable Equilibrium	An equilibrium is unstable when small perturbations induce change.	

Table 1: A table of definitions for terms, theories, and phrases often appearing in ecological regime shift literature. *(continued)*

Term	**Definition**	**Synonyms**
Unsupervised Maain Learning	**When no prior training of the data is required (i.e. no classifications necessary a priori) to classify it.**	
White Noise	Noise having zero mean, constant variance, and is not autocorrelated; uncorrelated random variability.	

Chapter 1

Introduction

Anthropogenic activity in the last few decades will continue to influence the interations within and among ecological systems worldwide. The complexity of and drivers of changes in coupled human-natural systems is consequently altered, further limiting our ability to detect and predict change and impacts of change (J. Liu et al., 2007; Scheffer, 2009). Early warning systems are developed to detect, and in some cases predict, abrupt changes in disparate systems [e.g. cyber security [@!!!!!], infrastructure [@!!!!], banking crises (Davis & Karim, 2008), and agricultural systems]. The need to develop and improve early warning systems for natural and coupled human-natural systems is exacerbated by the consequences of climate change and globalization, especially when the human-related stakes are high.

1.1 Forecasting abrupt changes in ecology

Forecasting undesirable change is, arguably, the holy grail of ecology. Paired with an understanding of system interactions, a forecast is ideal if it provides reliable forecasts with sufficient time to prevent or mitigate unwanted systemic change. Early warning systems (or early warning signals, or early warning indicators) have been developed and tested for some ecological systems data (especially marine fisheries time series and for nutrient loading in shallow lakes). Despite the quantitative methods proposed as early warning systems for ecological data (hereafter referred to as regime detection measures, RDMs), many are currently of limited practical utility. This paradox may be a consequence of existing ecological early warning systems (or other quantitative methods for identifying systemic change) having one or more of the following characteristics:

1. not generalizable across systems or system types (especially when it requires a model or a determinsitic function to describe the system)
2. require a large number of observations
3. difficult to implement
4. difficult or to interpret
5. requires an understanding of the drivers of change
6. performs poorly under uncertainty
7. give no uncertaintiy around estimates (tying into interpretation issues)
8. cannot handle noisy data
9. ignores or does not sufficiently account for observation error
10. no baseline with which to compare results
11. no application/testing on empirical systems data
12. systems are subjectively bounded (i.e., components are chosen)
13. being overshadowed by semantics
14. are based on two observations (e.g., before-and-after)
15. cannot link the shift to potential drivers (i.e. the method reduces the dimensionality such that it is unitless and/or loses all relevant information)

Research focusing on the above areas as they relate to RDMs will contribute to the advancement and improvement of existing early warning systems, and will, hopefully, highlight methods which are useful and which are not to practitioners and decision makers.

1.2 Dissertation aims

The overarching aim of this dissertation is to advance our understanding of the utility and limitations of select early warning systems. Specifically, I focus on RDMs capable of analyzing multi-varaible data, including temporally- and spatially-explicit. Although the most widely-applied RDMs proposed in the ecological literature are those deveoped for and tested on single-variable time series (e.g., temperature or fisheries stock time series), the utility of these methods in multi-variable systems (data) is limited. Regime detection metrics for tracking and identifying changes in multivariable systems data are of greater use than single-variable RDMs in systems within which a change manifests dynamically and across multiple variables (e.g., species). Multivariable RDMs may also prove advantageous when the drivers of systemic change are unknown. Further, ecological systems are noisy, and ecological systems data are messy.

Although it's taken us many decades to produce realiable weather forecasts 5 days out (and climate is a low-number system..), ecologists produce regime detection methods with the promise of predicting high-dimensional ecosystem change in advance. Many of these RDMs are not models, like the weather forecasting models which have taken years to refine.

1.3 Dissertation structure

1.3.1 Chapter overview

The dissertation comprises a brief introduction (Chapter 1), an overview of the myriad regime detectiob measures used or proposed for use with ecological data (Chapter 2), a detailed guide to Fisher Information as a RDM written for the lay ecologist (Chapter 3), an application of Fisher Information to spatially-explicit data (Chapter 4), introduction of a new regime detection measure, velocity (v) (Chapter 5), a study of data quality and data loss on select regime detectiob measures (Chapter 6), an application of body mass discontinuity analysis to spatially explicit data (Chapter 7), and a synthesis and conclusions chapter (Chapter 3.4).

1.3.2 Accompanying software (appendices)

This dissertation is accompanied by the vignettes for two software I created, which are publicly available for use (MIT use and distribution license). The first is `regimeDetectionMeasures` (Appendix ??), is an R package for calculting multiple regime detection measures, and the second, `bbsRDM` (Appendix ??), is a package which downloads and uses the North American Breeding Bird Survey data to calculate regime detection measures (using `regimeDetectionMeasures`).

Chapter 2

A brief overview of ecological regime detection methods methods

2.1 Introduction

If a regime shift occurs and no one detects it–is it a regime shift at all?

- **No** when a regime shift is defined as a change in a system which negatively impacts humans.
- **Yes** when a regime shift is defined simply as a shift in the underlying strucutre of a system.

Long-lasting changes in the underlying structure or functioning of natural systems due to exogeneous forcings (also called regime shifts) is of interest to ecologists. The ability to identify and predict these shifts is particularly useful for systems which are actively managed, provide ecosystem services, or provide benefit to societiy. There exists a disparity among the number of methods proposed for detecting abrupt changes in ecological, oceanographic, and climatological systems and the studies evaluating these methods using empircal data. Despite the already large number of existing methods and models, new methods continue to permeate the literature. Although reviews of regime shift detection methods exist (Andersen, Carstensen, Hern??ndez-Garc??a, & Duarte, 2009; Boettiger, Ross, & Hastings, 2013; Clements & Ozgul, 2018; Dakos, Carpenter, Nes, & Scheffer, 2015a, 2015b; deYoung et al., 2008; Filatova, Polhill, & Ewijk, 2016; Kefi et al., 2014; Litzow & Hunsicker, 2016; Mac Nally, Albano, & Fleishman, 2014; Mantua, 2004; Roberts et al., 2018; S. N. Rodionov, 2005; Scheffer, Carpenter, Dakos, & Nes, 2015), the most comprehensive presentation of available methods as they is outdated (S. N. Rodionov, 2005)*[1]

There is currently not a single, current resource to which the practical ecologist can refer when identifying or researching potential regime detection measures. Previous reviews of this literature vary in both the number and detail of the methods presented. This chapter is meant to serve as an addendum, of sorts, to previous reviews. Following the style of S. N. Rodionov (2005), I present a brief, yet exhaustive, over regime detection measures in the ecological literature. I then sugest next steps for ameliorating the plethora of regime detection measures in ecology.

2.2 Methods

Methods proposed as RSDMs are not easily identified using systematic literature review techniques for a few reasons. First, the terminology associated with regime shift detection methodologies is highly variable within and among fields. For example, the terms, *regime shifts, regime changes and tipping points* are variably used in studies of ecological systems, whereas *inhomogeneities* is common in meterology and climatology and *structural change* is largely confined to econometrics. Although the definition of, e.g., a regime shift and a structural change vary across and within fields of study, some methods are shared.

Second, papers introducing a new method or approach to identifying regime

[1] I also refer the reader to Kefi et al. (2014) and Yin, Leroux, & He (2017) spatial methods, and to Ducré-Robitaille, Vincent, & Boulet (2003) select tests for homogeneity in climate data.

shifts are not often proposed in publications that focus primarily on quantitative methodologies (e.g., *Ecological Modelling*, *Methods in Ecology and Evolution*) or in general ecology journals (e.g., *Ecology*). Instead, they are often published in journals with audiences that may not necessarily overlap with typical searches of the ecological litearture (e.g., *Entropy*, *Progress in Oceanography*).

I conducted a systematic literature review to identify original papers introducing quantiative regime detection measures. Although the literature review was designed to detect as many methodological papers as possible, most methods of which I was previous aware were not identified in this search. Therefore, I filled the gaps using prior knowledge and an informal search using Google Scholar. ### Identifying candidate articles

Web of Science

I first queried the Thomson-ISI Web of Science (WoS) database (on 06 March 2019) to identify articles which mention terms related to regime shifts, or abrupt changes, using the following boolean: > TS=((‘regime shift’ OR ‘regime shifts’ OR ‘regime change’ OR ‘regime changes’ OR ‘catastrophic change’ OR ‘catastrophic shift’ OR ‘catastrophic changes’ OR ‘catastrophic shifts’ OR ‘sudden change’ OR ‘sudden changes’ OR ‘abrupt shift’ OR ‘abrupt shifts OR ’abrupt change’ OR ‘abrupt changes’ OR bistab* OR threshol* OR hystere* OR ‘phase shift’ OR ‘phase shifts’ OR ‘phase change’ OR ‘phase changes’ OR ‘step change’ OR ‘step changes’ OR ‘stepped change’ OR ‘stepped changes’ OR ‘tipping point’ OR ‘tipping points’ OR ‘stable states’ OR ‘stable state’ OR ‘state change’ OR ‘state changes’ OR ‘stark shift’ OR ‘stark change’ OR ‘stark shifts’ OR ‘stark changes’ ‘structural change’ OR ‘structural changes’ OR ‘change-point’ OR ‘change point’ OR ‘change-points’ OR ‘change point’ OR ‘break point’ OR ‘break points’ OR ‘observational inhomogeneity’ OR ‘observational inhomogeneities’) AND (‘new method’ OR ‘new approach’ OR ‘novel method’ OR ‘novel approach’)) where '*' indicates a wildcard.

Limiting the search to 'Ecology' and 'Biodiversity Conservation' (by adding AND WC=(Ecology OR 'Biodiversity Conservation') to the above boolean) excludes many climatological and does not search the data science/computer science liteartures, where change-point analyses are abundant. However, because this dissertation is focused more on multivariate methods in ecology, this is not an issue.

Next, I filtered the results to identify articles which propose a 'new' method by retaining papers which included at least one of the following phrases in the title and/or abstract: > 'new method', 'novel method', 'new approach', 'new practical method', 'new simple method', 'new multivariate method', 'new tool', 'novel tool', 'novel multivarte', 'novel approach', 'new numerical', 'novel numerical', 'new quantitative', 'novel quantitative', 'i introduce', 'we introduce'

Prior knowledge and snowball method

Next; I removed articles from the above search (WoS) results based on both prior knowledge (in my personal database) and those highlighted in previous reviews related to regime detection measures (Andersen et al., 2009; Boettiger et al., 2013; Clements & Ozgul, 2018; Dakos et al., 2015a, 2015b; deYoung et al., 2008; Filatova et al., 2016; Kefi et al., 2014; Litzow & Hunsicker, 2016; Mantua, 2004; Roberts et al., 2018; S. N. Rodionov, 2005; Scheffer et al., 2015).

Google Scholar

There was a high disparity among the number of methods of which I was previously aware and those identified in an initial Web of Science review. In an attempt to collect as many new methods as possible, I conducted an informal search of the Google Scholar database, which is notoriously broader in scope. The length of boolean for the Google Scholar database is limited by the number of characters. Unfortunately, this, coupled with the wide breadth of Google Scholar's search boundaries, limits the capacity to which Google Scholar can be used to refine the literature to a manageable number of articles. For these reasons I arbitrarily skimmed the titles of the first 25 pages of the Google Scholar results (25 pages = 250 articles). It should be noted that the order of terms appearing in the boolean are regarded as the order of desired relevancy. I used the following boolean: > ('regime shift' OR 'regime change' OR 'tipping point') AND ('new method' OR 'new approach' OR 'novel method' OR 'novel approach')

Additional filtering

In addition to using the abovementioned search booleans, I excluded the following types of articles: those which proposed a combination of previously-used methods (e.g., PCA combined with other techniques, see Kong et al., 2017; Seddon, Froyd, Witkowski, & Willis, 2014; Vasilakopoulos, Raitsos, Tzanatos, & Maravelias, 2017) as a 'novel' method; those making relatively minor methodological updates/additions to existing methods (but see K. Nicholls, Hoyle, Johannsson, & Dermott, 2011 for an addition of variable optimization to the method in K. H. Nicholls (2011) that was not included in the results; Zhou & Shumway, 2008); and articles proposing new methodologies in mathematical journals (J. Byrski & Byrski, 2016; Salehpour, Gustafsson, & Johansson, 2011) that have yet to be associated with or tested ecological data, or suggested to be useful for empirical data.

2.3 Results

2.3.1 Web of Science

The search boolean for WoS boolean *not* including restriction to fields (WC) 'Ecology' and 'Conservation Biology' yielded over 20,000 results. Restricting to the abovemen-

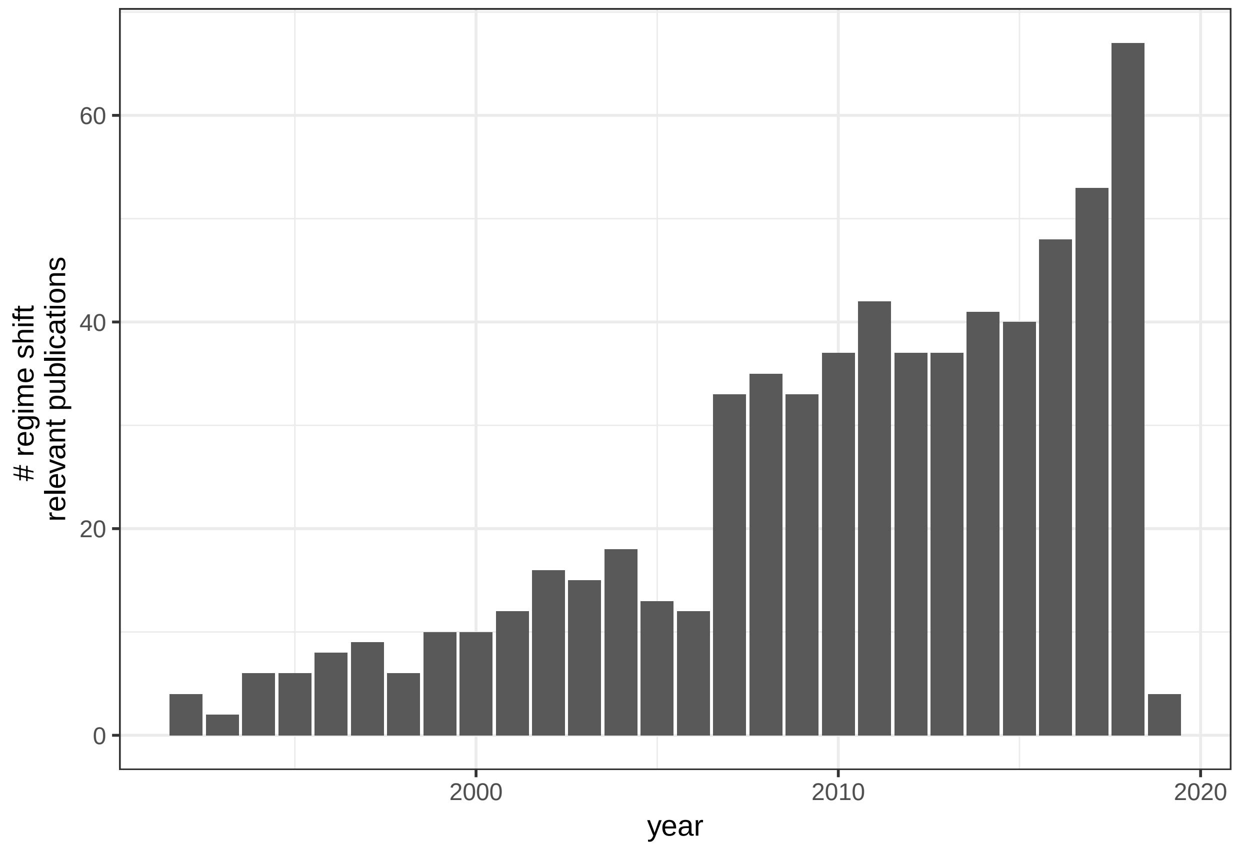


Figure 2.1: Number of publications by year in fields 'Ecology' and 'Biodiversity Conservation' which included terms related to 'regime shift' (total = 654).

tioned fields created a manageable database from which to filter. This search yielded 2,776 articles. 654 of these papers included terms relating to 'regime shifts' (Figure 2.1), many appearing in the journal *Ecological Modelling* (Figure 2.2). The rate of publication of 'regime shift' articles is not strongly correlated with the rate of papers published in 'Ecology' and 'Biodiversity Conservation' fields (Figure 2.3). Filtering this WoS results to include only articles mentioning terms related to 'new method' yielded 202 articles. After removing prior knowledge, only 93 articles remained to be reviewed 'by hand' (i.e., reading the entire paper). Only 2 'new' methods were identified from the WoS search (2.4).

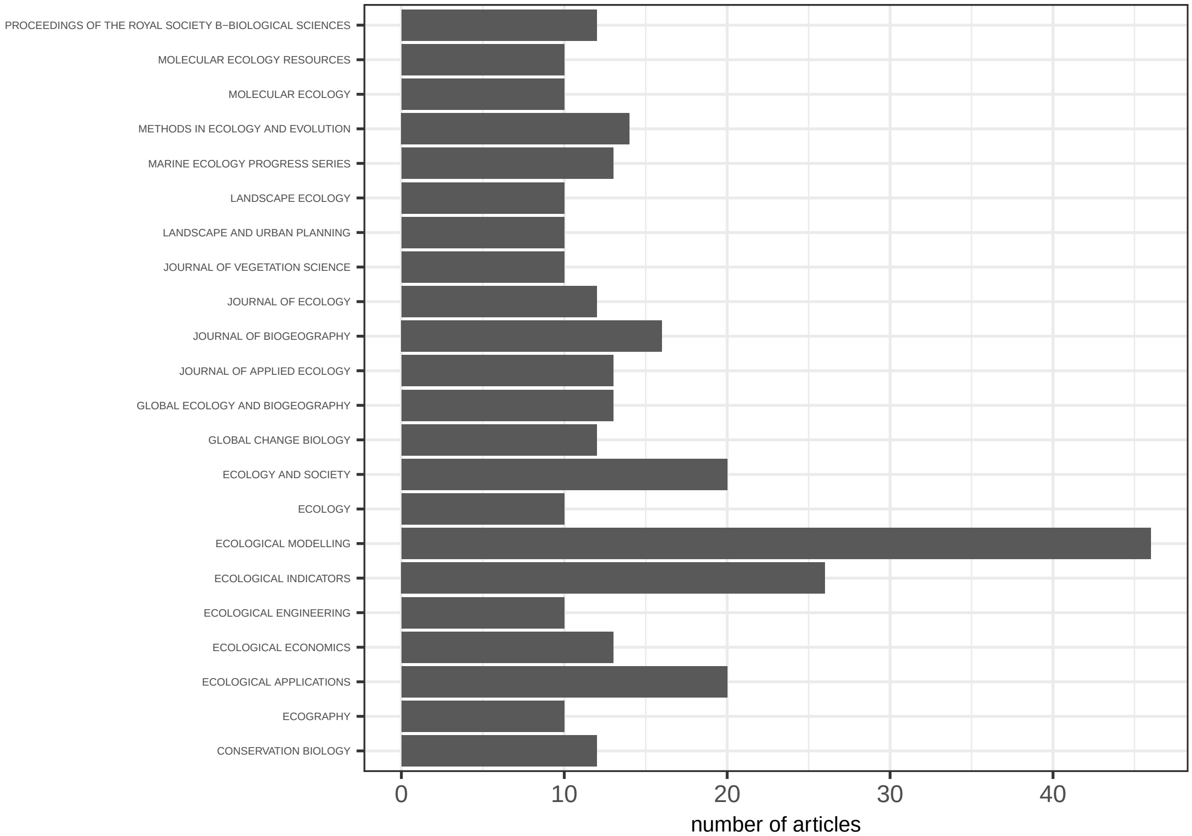


Figure 2.2: Distribution of the ‘regime shift’ articles for journals with at least 10 articles.

2.3.2 Google Scholar and prior knowledge

Of the 250 articles scanned in Google Scholar, I retained 3 methods. I was previously aware of an additional 68 articles containing ‘new’ methods (2.4).

2.3.3 List of new methods

Table 2.1: Longtable

method	type1
Autocorrelation at-lag-1	metric
Autoregressive coefficient of AR(1)	metric

Table 2.1: Longtable *(continued)*

method
Inverse of AR(1) coefficient
Detrended fluctuation analysis
Spectral density
Spectral ratio
Spectral exponent
Standard deviation
Coefficient of variation (CV)
Skewness
Kurtosis
Conditional heteroskedasticity
BDS test
Time-varying AR(p) model
Nonparametric drift-diffusion-jump model
Potential analysis
Fourier Analysis
T-test
Bayesian approaches
Mann-whitney U-test
Wilcoxon rank-sum
Pettitt test
Mann-Kendall test
LePage test

Table 2.1: Longtable *(continued)*

method	type1
Standard normal homoegeneity	metric
Regression-based models	model
Oerleman's method	metric
Cumulative deviation test (CUSUM)	metric
Signal-to-noise ratio	metric
Intervention Analysis	metric
STARS	metric
MCMC	NA
Quickest detection method (Shiryaev<d0>Roberts statistic)	metric
Variance Index	metric
Spectrum indicator	metric
Wavelet analysis (decomposition)	NA
Downton-Katz test	metric
Rodionov method	metric
Nikiforiv method	metric
Average standard deviates	metric
Fisher Information	metric
Vector-autoregressive method	NA
Lanzante method	NA
Free-knot splines & piecewise linear modelling	NA
Self-exciting threshold autoregressive state-space model SETARSS(p)	model
Smooth transition autoregressive model	model

Table 2.1: Longtable *(continued)*

method
Moving detrended fluctuation analysis (MDFA)
Nearest-neighbor statistics
Clustering, various
dimension reduction techniques (e.g., PCA)
Sequential tests/moving windows
Online dynamic linear modelling + time_varying autoregressive state_space models (TVARSS)
Stability Index of the Ecological Units
Generalized model
Threshold Indicator Taxa ANalysis (TITAN)
Convex model
Probability density function entropy method
Method 1-TBD
method-fuzzy synthetic evaluation (FSE)
Method 2-TBD
Zonal thresholding
Characteristic length scale (CLS) estimation
two-phase regression
shiftogram

Using my prior knowledge of the relevant litearture, referring to previous review articles, and searching both Web of Science and Google Scholar, I identified 64 unique regime detection measures (Figure 2.4; Table 2.1).

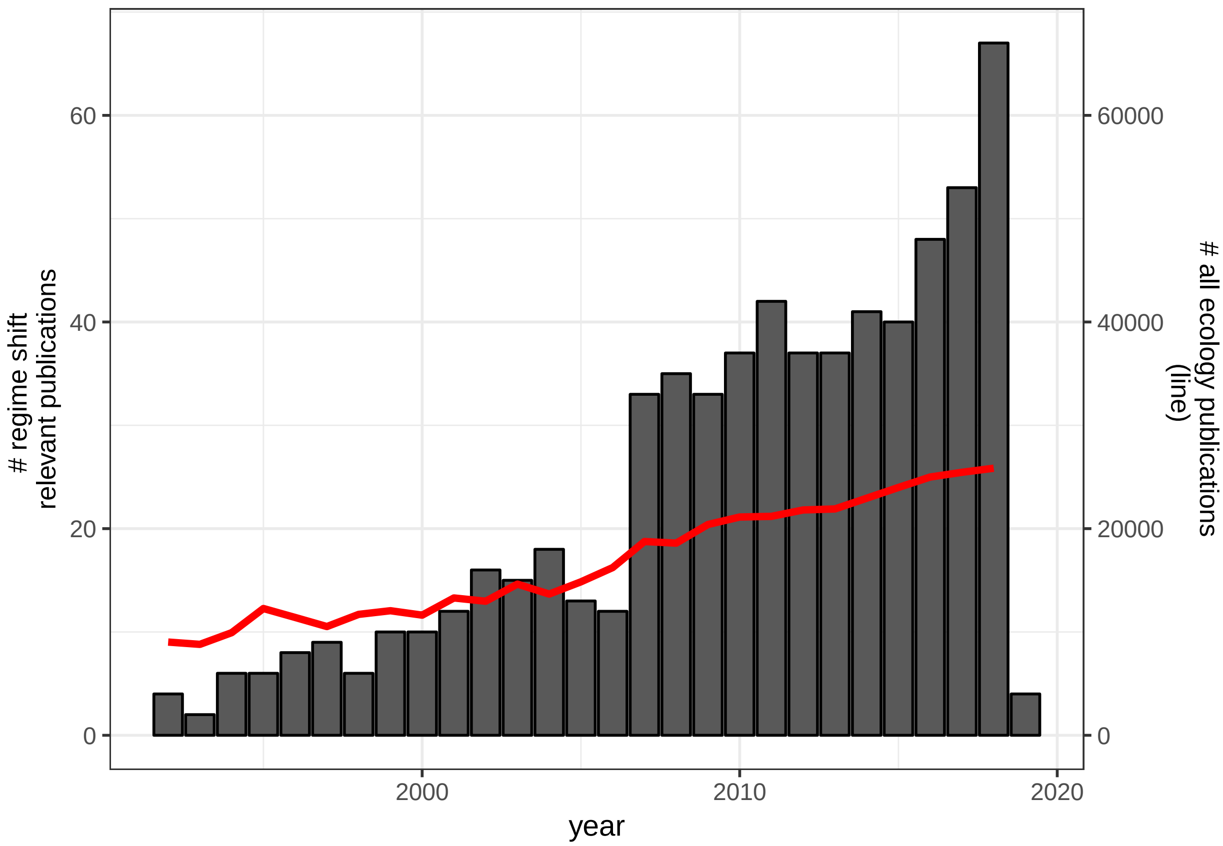


Figure 2.3: Number of publications by year in fields ‘Ecology’ and ‘Biodiversity Conservation’ which included terms related to ‘regime shift’ (total = 654).

2.4 Discussion

In this chapter I highlighted the plethora of regime detection metrics proposed in the literature for analyzing ecological data (Table 2.1). Although multiple reviews of regime detection measures exist, they are not comprehensive in their survey of the possible methods. Most reviews have summarized various aspects of regime detection measures. For example, Roberts et al. (2018) summarizes methods capable of handling multiple (c.f. a single) variable, and Dakos et al. (2015b) review only methods designed to detect the phenomenon of critical slowing down. Here, I did not discriminate–rather, I present an exaustive list of the plethora of methods proposed for detecting ecological detect regime shifts, *sensu lato*, providing a much-needed update

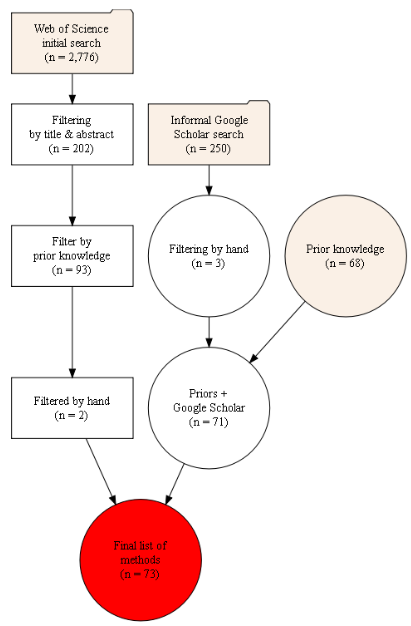


Figure 2.4: Flowchart of the litearture review process for identifying new regime detection methods. *Only the first ten pages (250 articles) of Google Scholar results were examined. Node shapes: folder = unfiltered articles; box = articles actively filtered; diamond = number of articles with new methods.

to collection provided by S. N. Rodionov (2005), and other review papers (Mac Nally et al., 2014, Scheffer et al. (2015), S. N. Rodionov (2005), Roberts et al. (2018), Dakos et al. (2015b), Mantua (2004), Litzow & Hunsicker (2016), Kefi et al. (2014), Andersen et al. (2009), Boettiger et al. (2013), Dakos et al. (2015a), Clements & Ozgul (2018), Filatova et al. (2016), deYoung et al. (2008)).

2.4.1 Barriers to identifying new regime detection measures

Clearly, as was shown in this chapter (Figure 2.4), a systematic review of the ecological literature will likely not yield anywhere near a comprehensive list of the regime detection measures proposed and/or used. This disaparity may be due to both my search methods and to the current state of regime shift research in ecology.

First, my review restricted articles to articles suggesting they were introducing a ‘new method’ as n RDM. Avoiding this potential barrier would have required I review

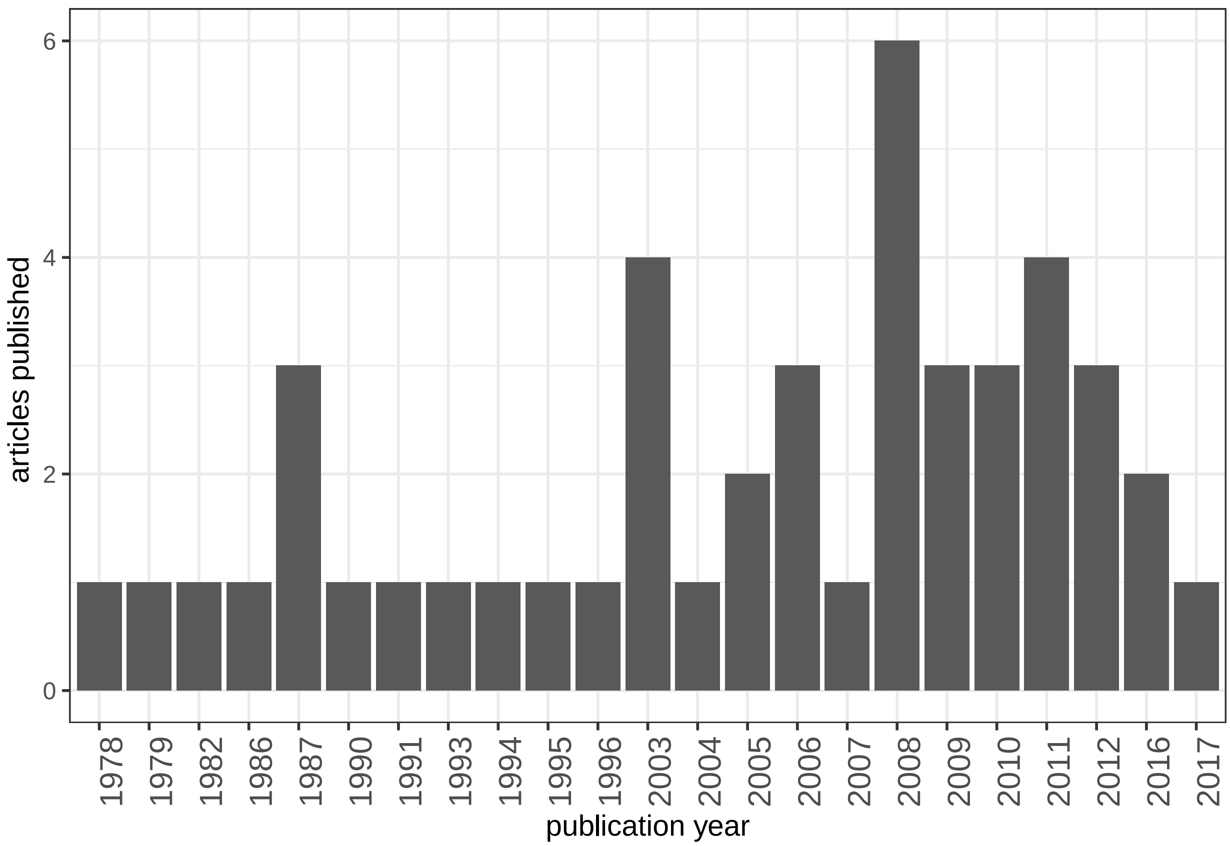


Figure 2.5: Number of methods publisheed over time.

the titles, abstracts, and bodies of over 22,000 articles (Figure 2.4). Alternatively, this may also be ameliorated by searching the relevant literature for *applications* of regime detection measures to ecological data, however, I suspect this would similarly yield a large number of articles to review.

Next, only a handful of methods have been introduced to the mainstream methodological journals in ecology (e.g., *Ecological Modelling*, *Methods in Ecology and Evolution*; Figure 2.6). Although many mainstream publications (e.g., *Science*, *Ecology Letters*) include applications of some of the methods identified in this chapter (Table 2.1), I argue that celebrity and 'new and shiny' (Steel, Kennedy, Cunningham, & Stanovick, 2013) methods may influence which methodological articles are printed in these popular journals. A critical survey of potetial and realized applications of these methods would be useful for highlighting the needs of future research and

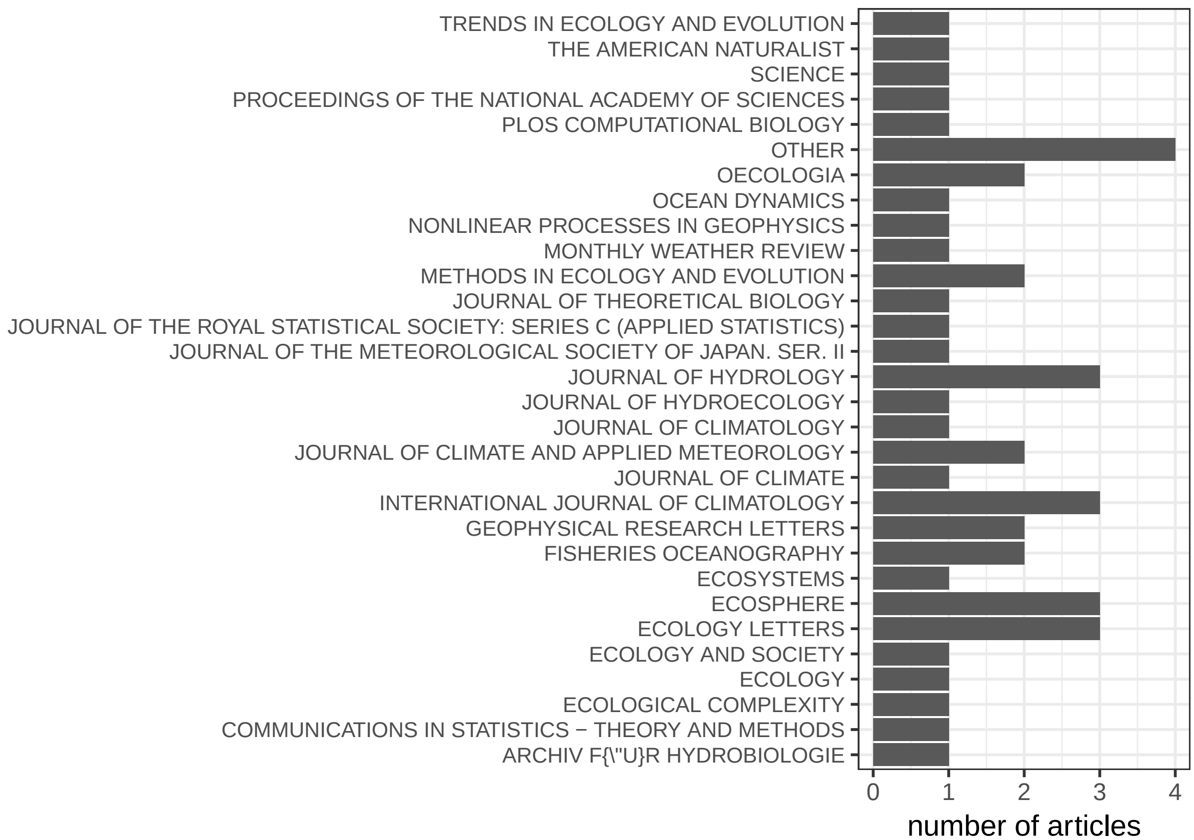


Figure 2.6: Distribution of identified methods across publications. Note: books, reports, and articles without original reference coded as ‘Other’

methodological improvements. Many of the methods presented in Table 2.1 have either not been applied to empirical data at all, or were tested only once (often but not always in the article introducing the method). Some methods, especially those dubbed ‘early warning indicators’ (variance, autoregressive model coefficients) have become relativley mainstream in their application to empirical data, however, have been shown to be less robust to noisy and nonlinear systems data (Burthe et al., 2016) and systems not exhibiting catstrophic shifts (Dutta, Sharma, & Abbott, 2018). Most other methods have yet to be rigorously tested on noisy, high dimensional, empirical data. Further, the methods which are not mainstream but have been applied to one of these data types have not any statistical indicators associated with confirming the existence and location of the regime shift.

As shown this chapter, identifying regime detection measures using traditional literature review techniques may prove difficult. Many of the methods identified in my review were not identified using Web of Science or Google Scholar–rather, I was either previously aware of most of the methods, and many others were highlighted in previous RDM reviews. To facilitate this process, an online, comprehensive database may prove useful to the practical ecologist.

2.4.2 Reducing the barriers to regime detection measures

To make the regime detection measures more available and transparent to the practical ecologist, I recommend the following: 1. consitent use of fewer methods 1. persistent collection and maintenance of baseline data (reference data) 1. an on-line database of all methods - open-sourced - linked to the original sources (in ecology and statistics or mathematics) - linked to applications 1. a critical review of the current state of methods in ecology - including methodological advancements - especially highlighting where the method fails to perform - including historical tracking of specific methods to identify which may need to be retired, rather than resuscitated 1. more empirical applications of these methods (especially of those only tested on toy and experimental data) 1. relation of RDMs in ecology to other fields (computer science, data science, climatology and oceanography)

I suggest below a suite of questions which may provide useful in a critical review of the characteristics, rigor, and promise of methods in the context of ecological data analysis.

Table 2.2: Potential questions for a comprehensive review of the ecological regime detection metrics literature.

Type	Questions
Methodological	What are the major assumptions about the distribution?

	Does the method explicitly assume stationarity? If not, can it handle non-stationary processes?
	Does the performance of the method change with non-stationarity?
	Can the method handle unstructured data (information)?
	Does the regime shift need to be identified *a priori*?
	Can the method handle multiple regime shifts?
	Does the performance of the method change with non-stationarity?
	What types of regime shifts can the method detect (e.g., stochastic resonance, slow-fast cycles, noise-induced transition)?
	Is it a model- or metric-based method?
	Does it have forecasting potential?
	Can the method handle uneven sampling?
	What are the minimum data requirements (resolution, extent, number of observations)?
	How does the method handle missing data (e.g., new invasions)?
	Does the method assume Eulerian or Lagrangian processes?
Ecological	Has the method been tested on empirical data? If so, to what rigor?
	What is the impact of losing state variables on long-term predictions (e.g., species extinction)?
	Can the method identify drivers?
	What assumptions does the method make about the system?
	What types of regime shifts are possible in the system?
	Are regime shift(s) suspected *a priori*?
	What lag(s) exist in the data (system)?
	Would a positive forecast change management action?
	Do predictions translate to other systems?

	Can we interpolate data if necessary? If so, what does this mean for inference? In which discipline(s) beyond ecology has the method been tested?

Chapter 3

A guide to Fisher Information for Ecologists

This chapter is intended for submission to the publication Methods in Ecology and Evolution.[1]

3.1 Abstract

Ecological regime shifts are increasingly prevalent in the Anthropocene. The number of methods proposed to detect these shifts are on the rise yet few are capable detecting regime shifts without a priori knowledge of the shift or are capable of handling high-dimensional and noisy data. A variation of Fisher Information (FI) in a dataset was proposed as a method for detecting changes in the orderliness of ecological systems. Although FI has been described in multiple research articles, previous presentations do not highlight a key component of FI that may make the metric easier to understand by practitioners. I used a two-species predator prey model to describe the concepts required to calculate FI. I hope this work will serve as a useful explanation of the FI metric for those seeking to understand it in the ecological systems and regime shifts.

[1] Co-authors include: N.B. Price, A.J. Tyre, C.R. Allen, T. Eason, D.G. Angeler, and D. Twidwell

3.2 Introduction

Changes in the feedback(s) governing ecosystem processes can trigger unexpected and sometimes undesirable responses in environmental conditions (Scheffer, Carpenter, Foley, Folke, & Walker, 2001; Walther et al., 2002). Ecologists often refer to such changes as regime shifts, but this term is used interchangeably in the literature with state change, state transition, or alternative state (Andersen et al., 2009). Climate change and globalization are triggering novel and unexpected changes in ecosystems, and the rapidity with which these changes occur make predictive modeling difficult. Although detecting regime shifts becomes more difficult as one increases the extent and complexity of the system in question , advances in the collection and analysis of ecological data may improve our ability to detect impending regime shifts in time for intervention (Jorgensen & Svirezhev, 2004).

Although multiple quantitative approaches are proposed as regime shift detection methods ,few are consistently applied to terrestrial ecological data. I classify a regime shift detection methods (DMs) broadly as either model-based or model-free (Boettiger & Hastings, 2012; Dakos et al., 2012; Hastings & Wysham, 2010). Model-based methods incorporate mathematical (mechanistic) representations of the system (Hefley, Tyre, & Blankenship, 2013) and carry strict assumptions, which are often violated by real systems (Abadi, Gimenez, Arlettaz, & Schaub, 2010). In addition to assumption violations nullifying parts of the model, model misspecification may yield spurious results (Charles T. Perretti, Munch, & Sugihara, 2013).

Model-free (or metric-based detectin ethods (e.g., descriptive statistics, cross-correlation mapping) require fewer assumptions to implement than do model-based DMs (Dakos et al., 2012). The most widely used model-free methods for detecting ecological regime shifts include descriptive statistics of one or a few components of a system, such as variance, skewness, and mean value (Andersen et al., 2009; Mantua, 2004; S. Rodionov & Overland, 2005) and composite measures which handle multivariable data, including principal components analysis (Petersen et al., 2008), clustering algorithms (G. Beaugrand, 2004), exergy (B. D. Fath & Cabezas, 2004), and Fisher Information (Cabezas & Fath, 2002; Karunanithi, Cabezas, Frieden, & Pawlowski, 2008).

Fisher Information, hereafter FI is a model-free composite measure of any number of variables (Fisher, 1922), and is proposed as an early warning signal for ecological regime shift detection system sustainability (D. A. L. Mayer, Pawlowski, Fath, & Cabezas, 2007, Karunanithi et al. (2008), Eason and Cabezas 2012, Eason et al. 2014a). Three definitions of FI exist: 1. A measure of the ability of the data to estimate a parameter.

1. The amount of information extracted from a set of measurements (Roy Frieden, 1998).

1. A measure representing the dynamic order/organization of a system (Cabezas & Fath, 2002).

The application of FI to complex ecological systems was posed as part of the 'Sustainable Regimes Hypothesis,' stating a system is sustainable, or is in a stable dynamic state, if over some period of time the average value of FI does not drastically change (Cabezas & Fath, 2002). This concept can be described using an ecological example. Consider the simple diffusion of a population released from a point source at $t = 0$. This process can be described by a bivariate normal distribution, $p(x, y|t)$. As the time since release (as t increases) increases the spread of the distribution, $p(x, y|t)$, becomes larger (less concentrated about the mean) because the animals have moved further from the release location. FI will decrease in value as t increases, because $p(x, y|t)$ contains less information (higher uncertainty) about where the animals will be located. As $t \to \infty$, the animals will be relatively uniformly distributed across the environment and $p(x, y|t)$ will carry no information about the location of the animals. Consequently, as $t \to \infty$, FI will approach zero. This system is not in a stable dynamic state because FI is decreasing with time.

In contrast, imagine a population varying around a carrying capacity following a simple logistic growth model. As long as the average system parameters (r and K and their variances) are stationary (not changing with time), then the logarithm of population size will have a normal distribution (check this!!!might need some different model). The FI measured over any selected window of time will be constant, indicating that the system is in a stable dynamic state. A perturbation to the population size due to disturbance will also not affect FI, as long as the disturbance does not change the distributions of r and K, and the perturbations themselves occur with some stationary probability distribution.

Although the concept of FI is firmly grounded in physics (B. R. Frieden, 1998), the concepts behind its application to ecological systems remain elusive to the average ecologist. I aim to elucidate the statistical concept of FI and the steps required to calculate it as a measure of 'ecosystem order' and as a regime shift detection method (Cabezas & Fath, 2002; B. D. Fath, Cabezas, & Pawlowski, 2003). I believe a concise and accessible synthesis of the topic, along with reproducible code, will aid the ecologists' understanding of this metric and will advance our understanding of its usefulness as an indicator of ecological regime shifts. I reproduce the analyses presented in (B. D. Fath et al., 2003) and D. A. L. Mayer et al. (2007) to fully explain these concept of and steps for calculating this form of Fisher Information. I hope this work will serve as a useful explanation of the FI metric for those seeking to understand it in the ecological regime shift context and will stimulate research using this and other multivariate, model-free, and composite measures to understand ecological regime shifts.

3.2.1 On Fisher Information

Two methods exist for calculating Fisher Information (FI) as applied to ecological systems data, which I refer to as the *derivatives-based* method, first appearing in Cabezas & Fath (2002), and the *binning* method, first appearing in Karunanithi et al. (2008). The binning method was proposed as an alternative to the derivatives-based method for handling noisy and sparse data, and requires additional calculations and system-specific decisions, and for these reasons I focus solely on the derivatives-based method. The general form of FI can be found in (B. D. Fath et al., 2003) and (D. A. L. Mayer et al., 2007), and although others can be found, I refer the reader to Cabezas & Fath (2002) for a complete derivation of FI.

3.2.2 Notation

A capital letter (e.g., A) denotes a random variable; an asterisk superscript (*) indicate a particular realization; *bold notation* indicates that the state of the system is defined in more than one dimension.

3.2.3 Steps for calculating Fisher Information (FI)

To calculate FI for a system with more than one state variable, I first estimate the probability of observing the system $p(x)$ in a given state, x, over time period T. The probability density function, $p(x)$, is then directly used to calculate the derivatives-based FI. I use bold notation to indicate that the state of the system is defined in more than one dimension (e.g., the state of a predator prey system is defined in two dimensions by the number of predators and number of prey). Here, I describe these steps and present the numerical calculation of FI using a two-species predator-prey model [B. D. Fath et al. (2003); mayer_applications_2007], hereafter referred to as the ‘model system’:

$$dx_1 = g_1 x_1 (1 - \frac{x_1}{k}) - \frac{l_{12} x_1 x_2}{1 + \beta x_1} dx_2 = \frac{g_{21} x_1 x_2}{1 + \beta x_1} - m_2 x_2) \tag{3.1}$$

The specified parameters for the model system are $g_1 = m_2 = 1$, $l_1 2 = g_1 2 = 0.01$, $k = 625$,and $\beta = 0.005$ (see B. D. Fath et al., 2003; B. R. Frieden & Gatenby, 2007; D. A. L. Mayer et al., 2007). The initial conditions (predator and prey abundances) for the model system were not provided in the original references. Using package *deSolve* in Program R (v 3.3.2) to solve the model system (3.1) I found $x_1 = 277.7815$ and $x_1 = 174.551$ provided reasonable results. I found that a complete cycle of the system corresponds to approximately 11.145 time units.

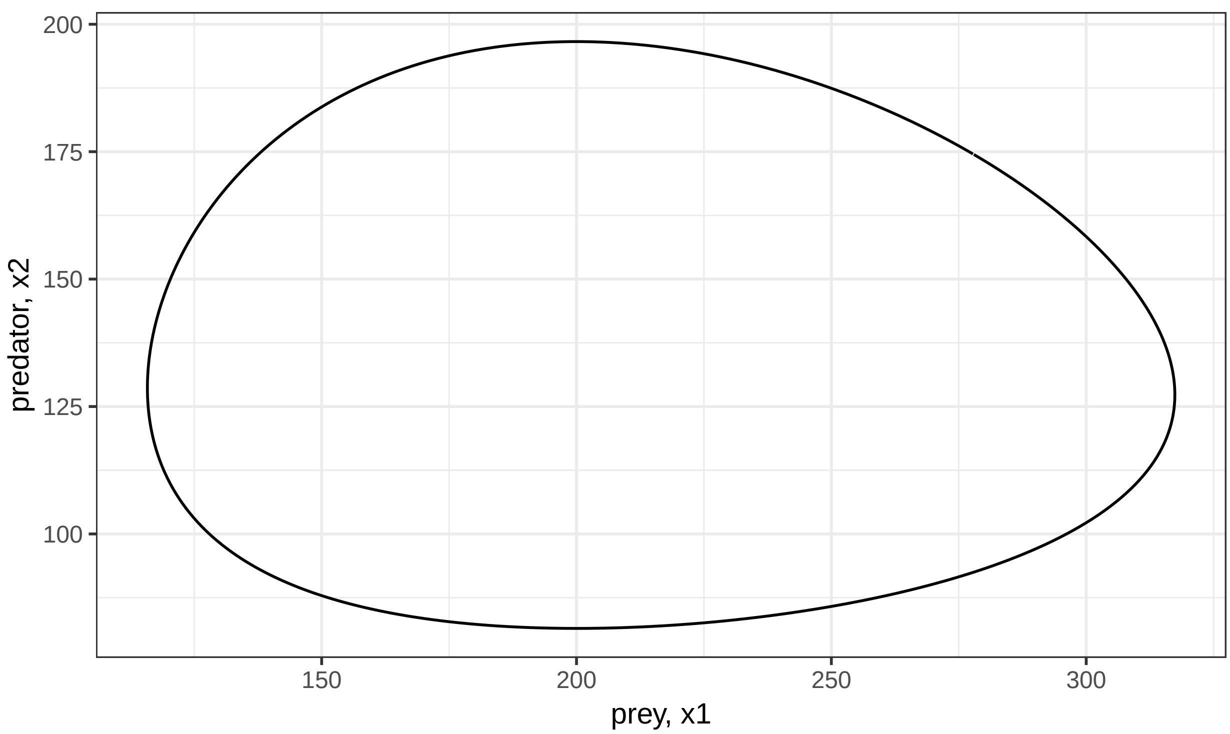


Figure 3.1: Phase space plot of two-species Lotka-Volterra predator-prey system over a single period (~11.145 time units.

3.2.4 Concepts behind the calculations

Although the numerical steps for calculating the derivatives-based FI are relatively straightforward, the concepts required to interpret the measure in the context of multiple variables is more complex. Here, I thoroughly discuss the concepts and assumptions behind FI calculation. Below, steps do not represent steps within the calculation, they represent the major concepts required

Step 1. Probability of observing the system in a particular state, $p(x)$

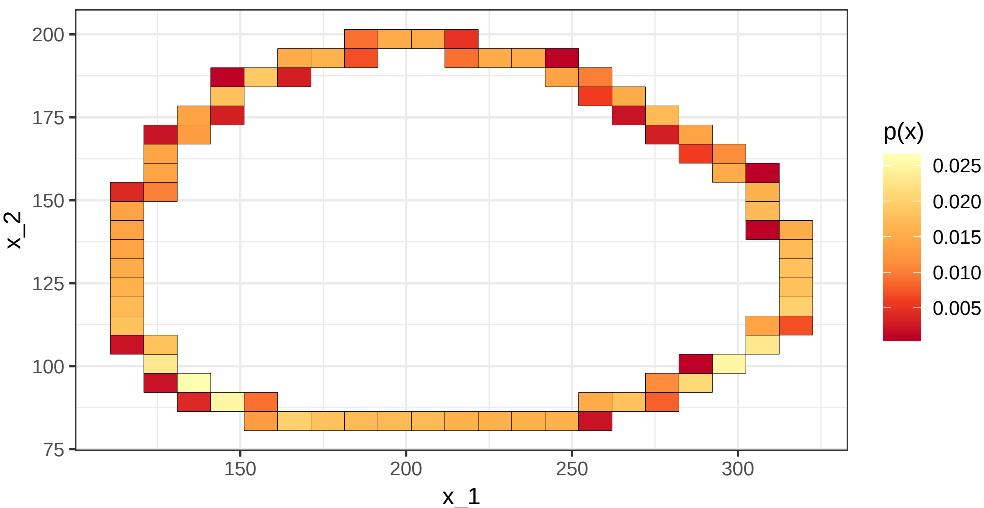


Figure 3.2: A 2-dimensional histogram of the probability of observing a system in a particular state, $p(x)$, of the 2-species Lotka-Volterra predator prey system over a single period (11.145 time units).

Fisher Information (FI) is defined with respect to a probability distribution. In the derivatives-based method, FI is calculated for a probability of observing a system (as defined by one or more state variables) in a particular state, $p(x)$, over some period of time, $(0tot_{end})$. In other words $p(x)$ is the probability that, at a specific point in time (t^*_{obs}) we will observe the system in a particular state, x^*. The time at which we observe the system is a random variable, $t_{obs}\ Uniform(0, t_{end})$. To be clear, the study system is assumed to be deterministic and we assume no observation error, however, the observed state of the system, $x(T_{obs})$, is a random variable because it is a function of the random observation time, $x^* = x(t^*_{obs})$. The state of the model system, x, is defined in two dimensions by the number of predators and the number of prey (3.1) and is easily visualized 3.1.Therefore, the probability of observing a particular state is a two-dimensional joint distribution **??**.

A single state of the model system is defined by the number of predators and prey at a given point in time such that for any given point in time $x(t) = [x_1(t), x_2(t)]$. At some random time between 0 and t_{end} $[T_{obs}\ Uniform(0, t_{end})]$ we can count the number of predators and the number of prey to determine the state of the model system. We must assume the system is deterministic and there is no observation error. We can then calculate the probability of observing a particular predator and prey abundance combination, $p(x)$. Under these assumptions, the only possible states of the system are defined by the system's observed trajectory, the model parameters, and the initial conditions. Therefore, the support of the probability distribution 3.2 is the trajectory of the system.

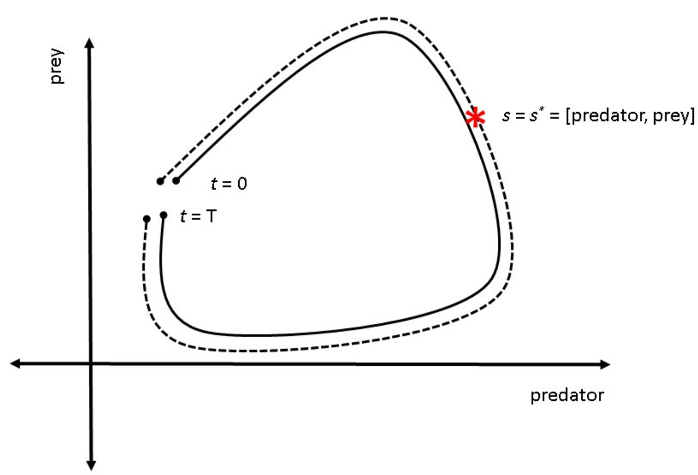


Figure 3.3: A single cycle of a hypothetical two-species system over time period $t = 0$ to $t = T$. s^* is the state of the system at some point in time. The dotted line represents the distance travelled by the system in phase space over its trajectory during time $(0, T)$.

Step 2. Distance traveled by the system, s

Distance traveled by the system, s. We can now move from an n-dimensional representation of the probability distribution to a one-dimensional representation. To better understand this, imagine placing a string over the path of the entire trajectory from $0tot_{end}$ 3.3. If we know the number of predators and prey at a particular point in time (t^*_{obs}) then we can mark that location on the string (see asterisk in 3.3. Next, imagine picking up the string and laying the string flat along a ruler. The length, s, of the entire string measures the total distance traveled by the system in phase space. The mark we made on the string (denoted $*$) lies at a distance s^* between 0 and s. We call this length the distance traveled by the system, s^*. In this context, s^* in phase space represents a measure of cumulative change in state. We note that the distance traveled in phase space increases monotonically with time. If the system never revisits the same state (i.e., the trajectory never overlaps or intersects itself), then every unique system state (i.e., point on the trajectory) is mapped to a unique value of distance traveled. Therefore, $p(x)$ (n-dimensional) is equivalent to the probability that the system is at distance s, i.e., $p(x) = p(s)$, (where $p(s)$ is one dimensional; Cabezas, Pawlowski, Mayer, & Hoagland (2005)). However, if the system revisits previous states, then a unique system state may be mapped to different values of distance traveled and the relationship between $p(x)$ and $p(s)$ is not one-to-one. We calculated the distance traveled s of the model system over a single cycle (11.145 time units; 3.4.

Step 3. $p(s)$ as a function of the rate of change of s

In previous presentations of FI, the relationship between the state of the system (n-dimensional) and the distance traveled (1-dimensional) was not always emphasized (Cabezas & Fath, 2002). Here we use x to denote the state of the system and s to denote the distance traveled to emphasize this distinction. If a system travels at a constant speed over the entire time period, then the system is equally likely to be in

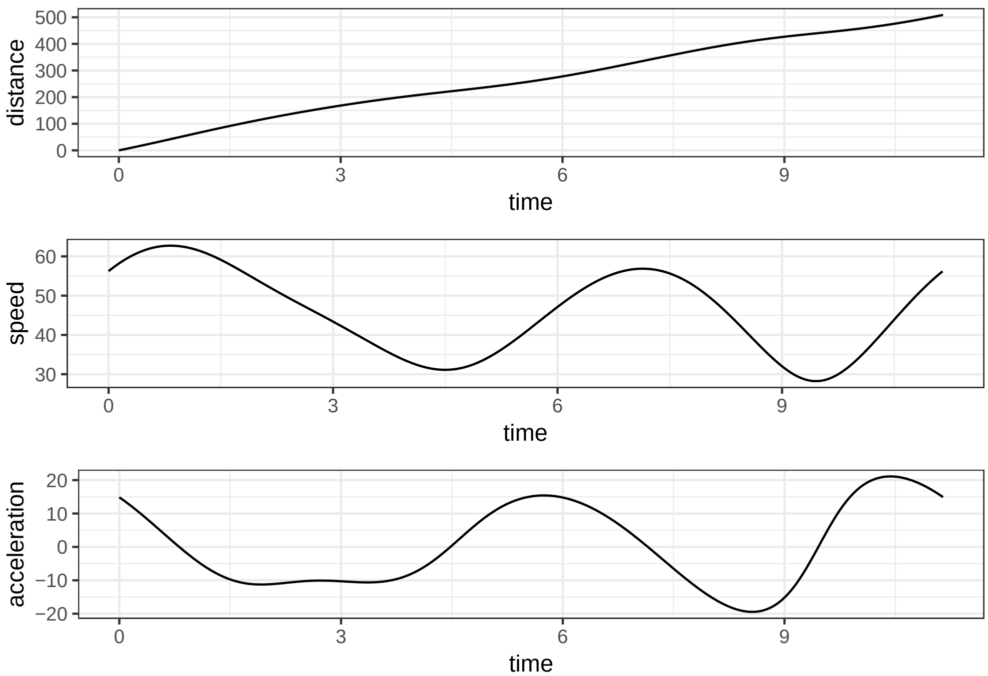


Figure 3.4: From top to bottom, distance traveled in phase space, speed tangential to system trajectory, acceleration tangential to system trajectory.

any state along the trajectory (s is linear and $p(s)$ is uniform). Referring to our model system, if the number of predators and prey are linearly related, then the speed of the system is constant. For non-linear systems, the distribution above the string will not be uniform 3.3. Rather, it will change depending on the amount of time the system spends in each state. It follows that $p(s)$ is proportional to the inverse of the rate of change of distance traveled (i.e., the speed along the path in phase space).

We will now demonstrate this using our model system as an example. Suppose the abundances of the predator and their prey in our model system predictably operate at carrying capacity. Over a relatively short period of time the prey abundance quickly declines after a severe weather event (a pulse disturbance; (Bender et al. 1984), but quickly recovers. Intuitively, the absolute rate of change at time points near the disturbance will be larger than during time periods long before or long after the disturbance. It is therefore more likely that the system will be (observed) in a state where prey and predators are operating approximately at carrying capacity than in a state with relatively low prey abundance. Mathematically, the time, $t*$, at which we calculate the abundances of prey and predators is a uniform random variable, and the distance traveled by the system, s^*, is a function of time, is differentiable, and monotonically increases. Therefore, the probability density function of the distance traveled $p(s) = \frac{1}{T}\frac{1}{s'}$, where $s' = \frac{ds}{dt}$ is the speed of the system (the speed tangential to the trajectory; the first derivative of the distance traveled; instantaneous rate of change of s). We calculated the speed (the first derivative; 3.4 and acceleration (the second derivative; 3.4 of the distance traveled s by the model system over a single cycle using function ode in package deSolve (Soetaert et al. 2010) in Program R (R Core Team 2016).

Step 4. Calculate the derivatives-based Fisher Information

Now that we understand how to calculate both the distance traveled, s, and its probability density, $p(s)$, calculating the derivatives-based FI is straightforward and computationally inexpensive (4.4). There are several comparable equations for calculating the shift-invariant FI, and some may offer numerical advantages over others. Equation (3.3) is the general form and Equation (3.4) is the amplitude form for FI (in D. A. L. Mayer et al. (2007), respectively). Although these formulations are equivalent, (3.4) is most readily calculated when the differential equations for the system are known, obviating any advantage of a model-free metric.

$$I = \frac{1}{T}\int_0^T dt \left[\frac{s''^2}{s'^4}\right]^2 \tag{3.2}$$

$$I = \int \frac{ds}{p(s)} \left[\frac{dp(s)}{ds}\right]^2 \tag{3.3}$$

$$I = 4\int ds \left[\frac{dq(s)}{ds}\right]^2 \tag{3.4}$$

This article is interested in the Fisher Information calculated for a distribution of distance traveled, s, by the entire system. We calculated the Fisher Information value using Equation (4.4) over a single period of the model system ((3.1)). We calculated Fisher Information to be 5.3 x 10^{-5} which is consistent with the results of Mayer et al. (2007).

3.3 Case Study

Mayer et al. (2007) calculated FI for a predator-prey system for several discrete values of carrying capacity of prey. The results of this study showed that FI was different for systems with different carrying capacities. However, this study did not address the central question of how FI changes during a regime shift. As an extension of the original study, we simulate a regime shift by modeling a situation where carrying capacity is abruptly decreased. To simulate an abrupt change in carrying capacity, we assume carrying capacity is described by Eq. 6 where $k1$ is the initial carrying capacity, $k2$ is the final carrying capacity, $t*$ is the time of the regime shift, and alpha is a parameter that controls how quickly the regime shift occurs. The hyperbolic tangent function simulates a smooth, continuous change in carrying capacity while still allowing for the change to occur suddenly. To incorporate the change in carrying capacity into the system differential equations we define the rate of change of carrying capacity as given by (3.5).

$$k(t) = k_1 - 0.5(k_1 - k_2)(\tanh(\alpha(t-t^*)) + 1)k'(t) = 0.5\alpha(k_1 - k_2)(\tanh(\alpha(t-t^*))^2 + 1) \tag{3.5}$$

We assumed an initial carrying capacity of 800 and a final carrying capacity of 625 which corresponds to the range of carrying capacities explored by Mayer et al. (2007). We simulated a time series of 600 time units with a regime change after 200 time units.

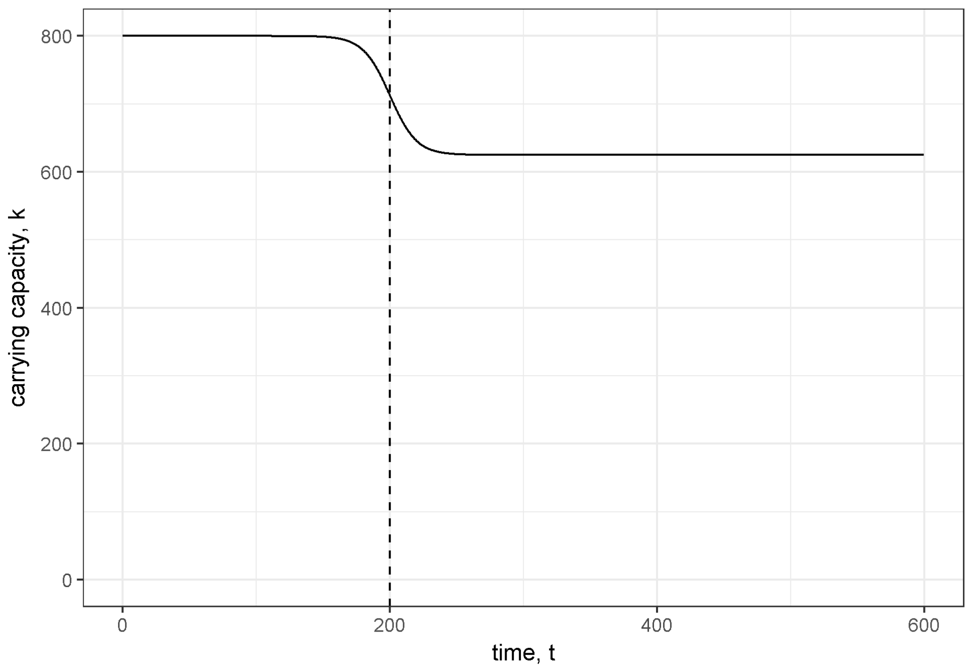


Figure 3.5: Carrying capacity over time with a regime shift occuring around time 200.

We used an alpha value of 0.05. The time series for carrying capacity is shown in 3.5 and the system trajectory in phase space is shown in 3.6. The distance travelled in phase space (i.e., cumulative change in state) is shown in **??** and the speed of the system (i.e., rate of change) is shown in 3.7. We calculated FI for the distribution of distance travelled over a series of non-overlapping time windows. Multiple sources suggest the length of the time window should be equal to one system period such that FI is constant for a periodic system (Cabezas & Fath, 2002; D. A. L. Mayer et al., 2007). However, the system period is different before, during, and after the regime shift. Therefore, we performed two separate calculations of FI using window sizes corresponding to the initial and final period of the system (13.061 and 11.135, respectively). The change in FI over time is shown in 3.8.

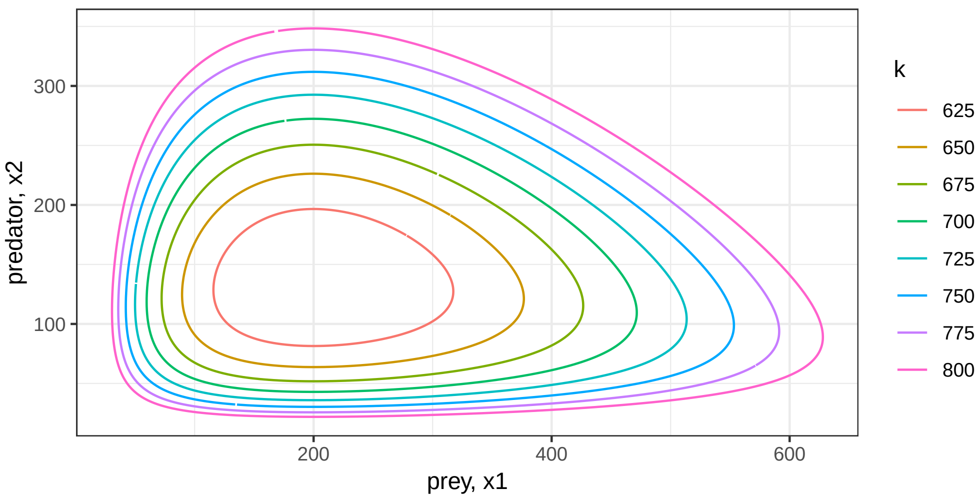


Figure 3.6: Phase space plot of system trajectories for different values of k

3.4 Conclusions

We simulated a regime shift caused by a change in carrying capacity (K) within a simulated, two-species Lotka-Volterra system. I applied the Fisher Information (FI) method for regime shift detection to the simulated time series data. The predator-prey system was modeled as deterministic and the time series data was free from measurement and observation error. Despite this, the estimated FI had high variation over time, and results were dependent on the size of the time window used (winsize) in the calculation 3.8. The FI method for regime shift detection is based on the cumulative change in the state of the system (i.e., distance traveled in phase space) and the rate of change of the system (i.e., speed tangential to trajectory in phase space). The distance travelled metric, s, and its speed, $dsdt$, appear better visual indicators of the regime shift than FI [**??**; 3.7].

In our explanation of the FI concept and calculation, I emphasize the distinction between the *state of the system* and the *distance traveled in phase space*. There are several reasons worth emphasizing this. First, there may not always be a one-to-one relationship between the probability of observing a system in a particular

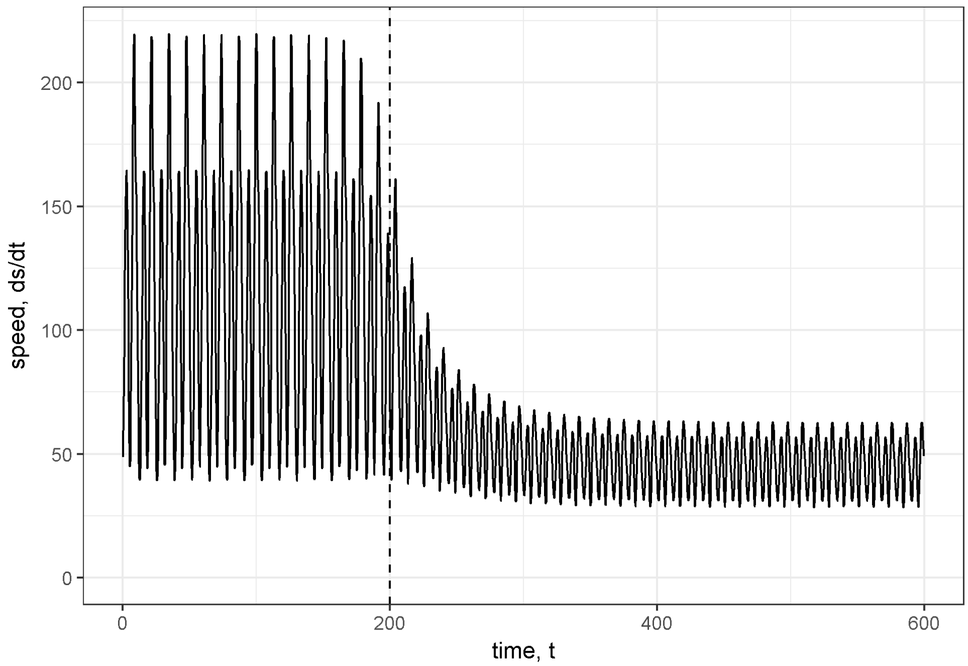


Figure 3.7: Speed of the system (rate of change) in phase space. Dashed vertical line at time 200 indicates location of regime shift.

state and the probability of observing a system at a particular distance along the trajectory. In these situations the interpretation of FI may be less clear than if a one-to-one relationship existed. Second, this distinction facilitates the separation of the dimensionality reduction step (calculating distance traveled in phase space, s) from the subsequent steps related specifically to FI. Third, the distinction suggests that the **value of FI as a regime shift detection method is related to the rate of change of the system** (i.e., velocity and acceleration tangential to system trajectory in phase space). In particular, the distribution for which FI is calculated is simply the distribution of the distance traveled in phase space, when time is assumed to be uniformly distributed over a given interval.

Our results suggest that insights can be gained directly from the calculation of distance traveled and associated rates of change. Consequently, these insights preclude the need to calculate beyond Step 3 (described above). This result also supports the use of the distance travelled metric, or the derivatives-based Fisher Information .

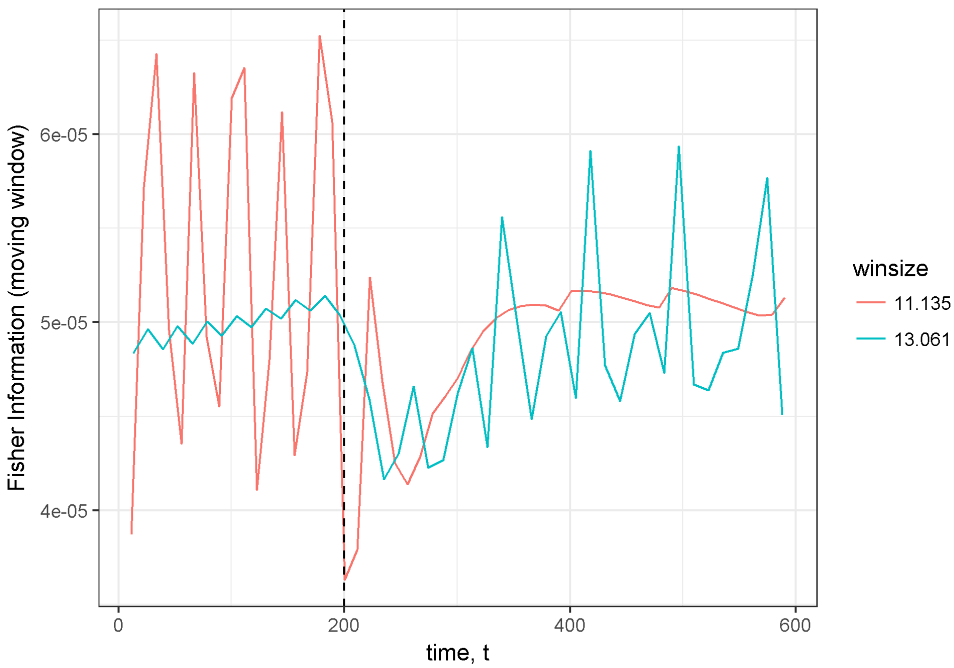


Figure 3.8: Fisher Information calculated for non-overlapping time windows. Two different window sizes were used as indicated by color. Dashed vertical line at time 200 indicates approximate location of regime shift.

One remaining issue that is prevalent across ecological field studies is the assumption that the system is observed without error. Although ecological data rarely fulfill this assumption, this does not suggest that FI is useless as a metric of system stability. The primary difficulty with noisy data, especially with observations in integer form (e.g. count data), is that the denominator in can easily be zero for some pair of observations, making FI an infinite value within windows which contain two or more adjacent zero observations. One possible solution is to smooth the multidimensional vector of observations prior to calculating the derivatives, or to treat any sequential identical value as missing, and simply use a larger time step for that portion of the window calculation.

The utility of Fisher Information in ecological studies is also stunted by its interpretability. This metric is unitless, making its values relative only within-sample (e.g., within a single time series). Further, interpreting the results within-sample is currently a qualitative effort (B. D. Fath et al., 2003; Mantua, 2004). When the FI of a system is increasing, the system is said to be moving toward a more orderly state, and most presentations of FI posit sharp changes in FI, regardless of the directionality of the change, may indicate a regime shift (Cabezas & Fath, 2002; Karunanithi et al., 2008; T. L. Spanbauer et al., 2014). Due to the qualitative nature of these interpretations of Fisher Information, intimate knowledge of the system in question and the potential driver(s) of the observed regime shift are required to confirm presence of a shift.

3.5 Acknowledgements

I thank T. Eason, H. Cabezas and B. Roy Frieden for early discussions regarding Fisher Information.

Chapter 4

An application of Fisher Information to spatially-explicit avian community data

4.1 Introduction

Ecosystems are open, dynamical systems which arguably cannot be fully represented by deterministic models. Despite the complexity of most ecological systems, some patterns have emerged in certain statistical mechanics of ecological observations. An uptick in recent years of studies of **regime shifts** (**??**) in ecology has spurred an increase in the number of 'new' methods for detecting ecological regime shifts (2), some of which are proposed as indicators of 'spatial' regime shifts (Butitta, Carpenter, Loken, Pace, & Stanley, 2017, Kefi et al. (2014), Sundstrom et al. (2017), (**???**), W. Brock & Carpenter (2006)).

As defined in **??**, a regime shift is largely considered an abrupt and persistent change in a system's structure or functioning. Following this definition and without any associated **pressures ??**, it is not yet clear whether identifying a 'spatial regime' using a snapshot of a system (a single or short period of time relative to the time scale of the pressure) is pragmatic. One spatial regime detection measure (hereafter, SRDM) is variance (W. Brock & Carpenter, 2006), despite its controversial applicability to temporal data (**???**, Dutta et al. (2018), Charles T Perretti & Munch (2012), Sommer, Benthem, Fontaneto, & Ozgul (2017), Bestelmeyer et al. (2011)).

Defining the spatial regime shift is important since observations of non-random spatial processes (e.g., land cover), could manifest as either rapid shift (e.g. an ecotone) or a gradual change (slow mixing along a gradient). Consequently, and because most RDMs signal abrupt change, only the former may be identified as "regime shifts" using SRDMs. For the concept of spatial regimes to be ecologically useful, potential pressures must be associated with system structure over space *and* time. Additionally and perhaps more importantly, the processes driving the observed information (drivers, pressures) should be such that a statistically identified regime shift will roughly correspond with the time scale on which the pressure(s) operate.

Although it is suggested that statistical and pragmatic models and methods are advanced more rapidly by bottom-up approaches, i.e. case studies (see DeAngelis & Yurek, 2017), to my knowledge no studies have yet to test the rigor of SRDMs using spatially-explicit empirical data. The objective of this chapter is to determine the utility of Fisher Information (Eq. (4.4)) as a spatial regime detection measure. This chapter is also supported by original software developed for implementation in Program R, which is publicly available [see Appendix **??**].

4.2 Data and methods

4.2.1 Data: North American breeding bird communities

I use community abundance data from long-term monitoring programs to identify spatial and temporal regimes using the Fisher Information (FI) derivatives method (see Eq. (??)). The NABBS trains citizen scientist volunteers to annually collect data using a standardized roadside, single observer point count protocol and has been collecting data regularly across North America (??) since 1966. The roadside surveys consist of 50 point counts (by sight and sound) along an approximately 24.5 mile stretch of road. Due to strict reliane on volunteers, some routes are not covered every year. Additionally, some routes are moved or discontinued, and some routes are not sampled in a given year. Route-year combinations which are missing years but are not discontinued are treated as missing data. Although NABBS volunteers identify all species as possible, persistent biases exist in this protocol. To reduce the influence of potential sampling bias, I removed waterfowl, waders, and shore species (AOU species codes 0000 through 2880).

4.2.2 Study area

Although the NABBS conducts surveys throughout much of North America, I limited analyses to the continental United States and parts of southern Canada. NABBS coverage of the boreal forests of Canada are sparse in space, and many routes in Mexico have fewer than 25 years of observations.

Focal military base

The Mission of the US Department of Defense is to provide military forces to deter war and protect the security of the country, and a primary objective of individual military bases is to maintain military readiness. To maintain readiness, military bases strictly monitor and manage their natural resources. Military bases vary in size and nature, and are heterogeneously distributed across the continental United States (See Fig. 4.1). The spread of these bases (Fig. 4.2), coupled with the top-down management of base-level natural resources presumably influences the inherent difficulties associated with collaborative management within and across military bases

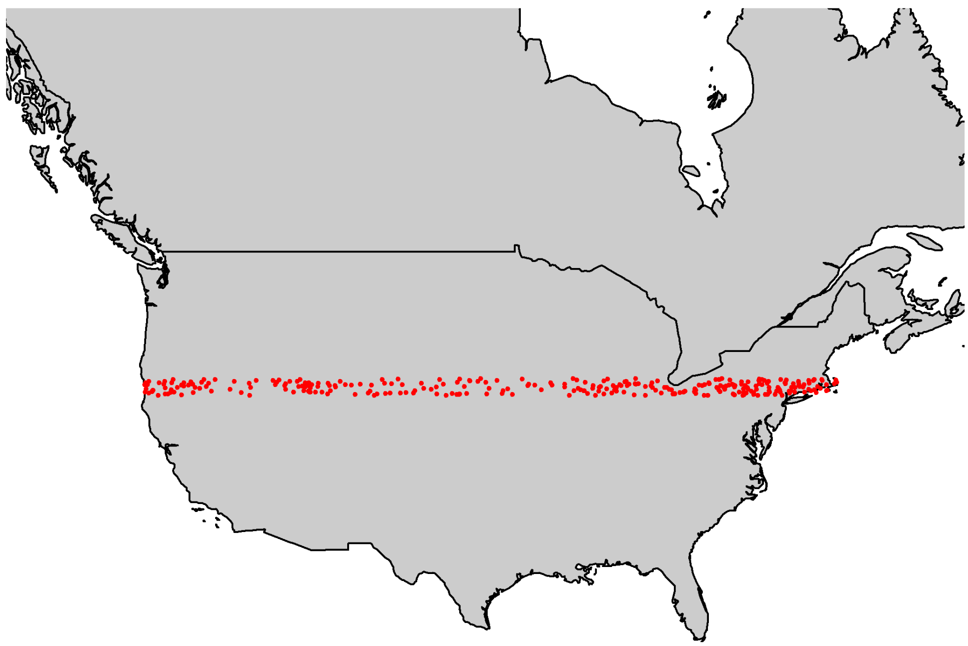


Figure 4.1: A single East-West transect of Breeding Bird Survey routes used to calculate the Fisher Information.

and other natural resource management groups (e.g., state management agencies, non-profit environmental groups.

Much like other actively managed landscapes, miltiary bases are typically surrounded by non- or improperly-managed lands. Natural resource managers of military bases face environmental pressures within and surrounding their properties, yet their primary objectives are very different. Natural resource managers of military bases, whose primary objective is to maintain military readiness, are especially concerned with if and how broad-scale external forcings might influence their lands. Prominent concerns include invasive species, wildlife disease, and federally protected species (personal communication with Department of Defense natural resource managers at Eglin Air Force and Fort Riley military bases). For these reasons, natural resource managers attempt to create buffers along their perimeters (e.g., live fire/ammunitions suppression, wide fire breaks). Identifying the proximity of military bases to historic and modern ecological shifts may provide insight into the effectiveness of their natural resource management efforts. The NABBS routes chosen for analyses in this Chapter

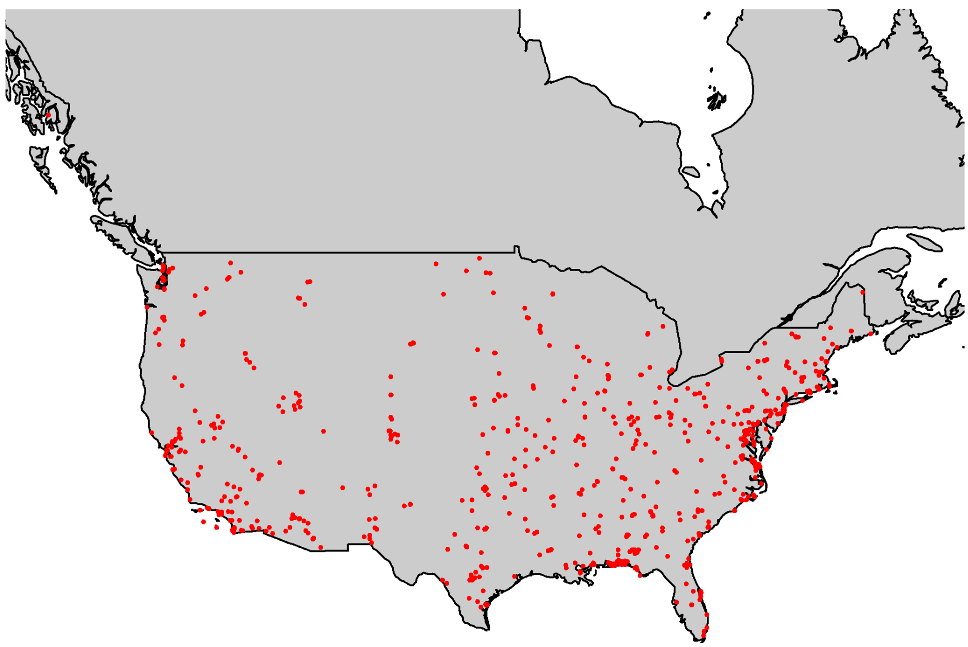


Figure 4.2: Locations of U.S. military bases in our study area.

lie within or near Fort Riley military base (located at approximately 39.110474°, −96.809677°; Kansas, USA). Fort Riley (Fig. 4.3) is a useful reference site for this study. Woody encroachment of the Central Great Plains over the last century has triggered shifts in dominant vegetative cover and diversity (Ratajczak et al. 2012) in the area surrounding Fort Riley military base (e.g., Van Auken 2009). This phenomena should present itself as a regime boundary should Fisher Information be a robust regime shift detection method.

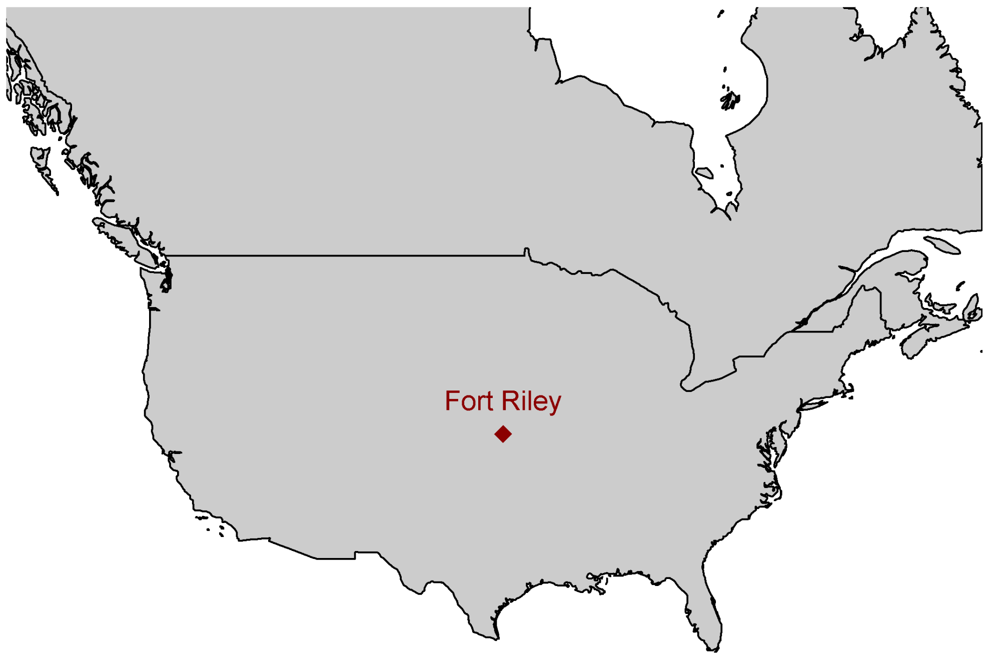


Figure 4.3: Locations of focal U.S. military bases, Eglin Air Force Base (AFB) and Fort Riley Military Base.

Spatial sampling grid

To my knowledge, (**???**) is the only study to use the Fisher Information on spatially-referenced data. The authors of this study hand-picked NABBS routes to be included in their samples such that their metrics should detect 'regime changes' when adjacent sampling points represented different ecoregions (broad-scale vegetation classification system). The authors also suggest each ecoregion is similarly represented, having a similar number of NABBS routes within each ecoregion in the analysis. However, this method of handpicking routes resulted in a transect which was neither North-South nor East-West running (see (**???**)), but rather zigzagged across a midwestern region. I

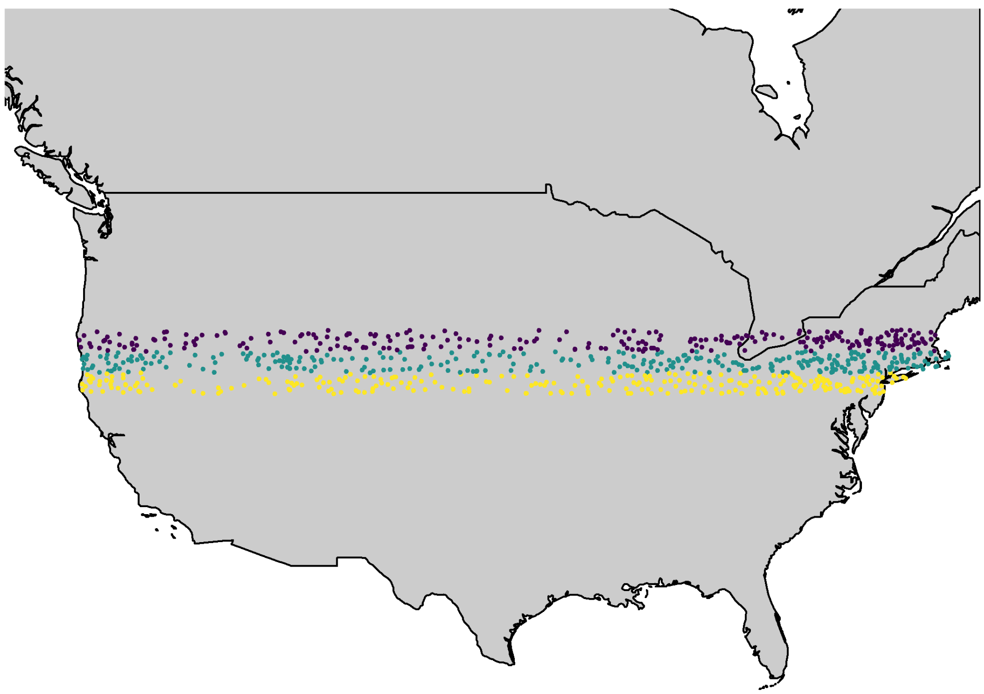


Figure 4.4: The three East-West running transects used to visualize results in this chapter.

constructed a gridded system across the continental United States and parts of Canada. The gridded system comprises East-West running transects transects running in either North-South or East-West directions. This method ameliorates some sampling bias, as I have arbitrarily defined sampling transects, rather than hand-picking sites to include in the analysis. Additionally, this approach allows for raster stacking, or layering data layers (e.g., vegetation, LIDAR, weather) on top of the sampling grid and results, allowing one to identify potential relationships with large-scale drivers. This method also provides a simple vector for visualizing changes in the Fisher Information over space-time, using animations and still figures. For brevity, I present visual results of only three, spatially-adjacent, East-West running transects (Fig. 4.4) at multiple time periods.

4.2.3 Calculating Fisher Information (FI)

Fisher Information, $I(\theta)$, was developed in 1922 by Ronald Fisher as a measure of the amount of information that an observable variable, X, reveals about an unknown parameter, θ. Fisher Information is a measure of indeterminacy (Fisher 1922) and is defined as,

$$I(\theta) = \int \frac{dy}{p(y|\theta)} \left[\frac{dp(y|\theta)}{d\theta} \right]^2 \tag{4.1}$$

where $p(y|\theta)$ is the probability density of obtaining the data in presence of θ. The Fisher Information measure (FIM) is used to calculate the covariance matrix associated with the likelihood, $p(y|\theta)$. Fisher Information is described as Extreme Physical Information (EPI; Frieden and Soffer 1995, Kibble 1999, Frieden et al. 2002), a measure that has been used to track the complexity of systems in many scientific disciplines including, physics, cancer research, electrical engineering, and, recently, complex systems theory and ecology

Fisher Information as gathered from observational data provides insight as to the dynamic order of a system, where an orderly system is one with constant (i.e., unchanging) observation points, and one whose nature is highly predictable. A disorderly system is just the opposite, where each next data point is statistically unpredictable. In ecological systems, patterns are assumed to be a realization of ecosystem order; therefore, oneshould expect orderliness in a system with relatively stable processes and feedbacks. Orderliness, however, does not necessarily infer long-term predictability. Equation (4.1) is next adapted to estimate the dynamic order of an entire system, s, as

$$I = \int \frac{ds}{p(s)} \left[\frac{dp(s)}{ds} \right]^2 \tag{4.2}$$

where $p(s)$ is the probability density for s. Here, a relatively high Fisher Information value (I) infers higher dynamic order, whereas a lower value (approaching zero) infers less orderliness. To limit the potential values of I in real data, we can calculate the amount of Fisher Information by re-expressing it in terms of a probability amplitude function $q(s)$ (Fath et al. 2003, Mayer et al. 2007, eq. 7.3):

$$I = 4 \int ds \left[\frac{dq(s)}{ds} \right]^2 \tag{4.3}$$

A form specific to the pdf of distance travelled by the entire system, which I call the 'derivatives' method, is defined as (D. A. L. Mayer et al., 2007, eq. 7.12):

$$I = \frac{1}{T} \int_0^T dt \left[\frac{s''^2}{s'^4} \right]^2 \tag{4.4}$$

where T is the number of equally spaced time points over which the data are integrated. Numerical calculation of I using the binning method (Eq. (4.3) and (4.4)) each incorporate a moving-window procedure for calculating the probability of the system, $p(s)$, as being in one of an unidentified number of states (s). Although previously applied to spatially-explicit terrestrial community data,the binning method (Eq. **??**) requires multiple parameters to be defined *a priori*, which have been shown to influence inference based on the metric. I therefore calculated FI using the derivatives equation (Eq. **??**).

The binning procedure allows for a single point in time or space to be categorized into more than one state, which violating the properties of alternative stable states theory. The size of states (see Eason and Cabezas 2012) measure is required to construct p(s). In the case of high dimensional data, a univariate binning procedure of p(s) is not intuitive (i.e., reducing a multivariable system to a single probability distribution rather than constructing a multivariate probability distribution). Importantly, when using community or abundance data, rare or highly abundant species can influence the size of states criterion, thus influencing the assignment of each point into states. Finally, Eq. (4.3) assumes equal spacing (in space or time) between sampling points. Each of these violations can be avoided by using Eq. (4.4); Cabezas and Fath 2002, Fath et al. 2003) to calculate the Fisher Information measure. The derivatives method (Eq. (4.4)) estimates the trajectory of the system's state by calculating the integral of the ratio of the system's acceleration and speed in state space (B. D. Fath et al., 2003). I calculated Fisher Information using Equation (4.4) for all East-West transect (see Fig. **??**) for years 1980, 1990, 2000, and 2010.

4.2.4 Interpreting and comparing Fisher Information across spatial transects

Interpreting Fisher Information values

Here I define a potential regime change as a point(s) having a non-zero derivative, and at which relatively large changes (sharp increase or decrease) in the Fisher Information measure occur. Regime shifts are identified as data changing from one state to another, thus, rapid shifts in the value of FI should indicate the points, in time or space, at which the system undergoes reorganization. Spatial and temporal Fisher Information calculation does not vary, but interpretation of either differ in that a spatial analysis will identify a spatial regime boundary (**???**) in space within a single time period, whereas analysis of temporal data will identify a point(s) in time at which a system in a specific location undergoes a regime shift. I follow the methods outlined in the relevant literature for interpreting the Fisher Information (e.g., Karunanithi et al., 2008, Eason & Cabezas (2012)).

Increases in FI is proposed as an indicator of system orderliness, where periods of relatively high values of FI indicate the system is in an "orderly" state, or is fluctuating around a single attractor. A rapid change in FI is supposed to indicated the system is no longer orderly and may be undergoing a reorganization phase. Whether Fisher Information can identify a switch among basins of attraction within a single, stable state (or around a single attractor) remains unknown, as does the number of states which a system can occupy. When a system occurs within any number of states equally, i.e., p(s) is equal for each state, both the derivative, $(\frac{dq(s)}{ds}$, and I are zero. As $(\frac{dq(s)}{ds} \rightarrow \infty)$, we infer the system is approaching a stable state, and as $\frac{dq(s)}{ds} \rightarrow 0$ the system is showing no preference for a single stable state and is on an unpredictable trajectory. (4.3) bounds the potential values of Fisher Information at $[0, 8]$, whereas (4.1), (3.4), and (4.4) have are positively unbounded $[0, \infty)$. If the Fisher Information is assumed to represent the probability of the system being observed in some state, s, then the absolute value of the Fisher Information index is relative within a single datum (here, transect). It follows that Fisher Information should be interpreted relatively, but not absolutely.

Interpolating results across spatial transects

Because the BBS routes are not regularly spaced, pairwise correlations of adjacent transects are not possible without either binning the Fisher Information calculations using a moving-window analysis, or interpolating the results to regularly-spaced positions in space. To avoid potential biases associated with the former option, I linearally interpolated Fisher Information within each spatial transect (Fig. 4.4) at 50 points along the longitudinal axis. The 50 longitudinal points at which I interpolated were the same across each spatial transect. I used the function *stats::approx()* to linearly approximate the Fisher Information. I did not interpolate values beyond the longitudinal range of the original data (using argument *rule=1* in package *approx*).

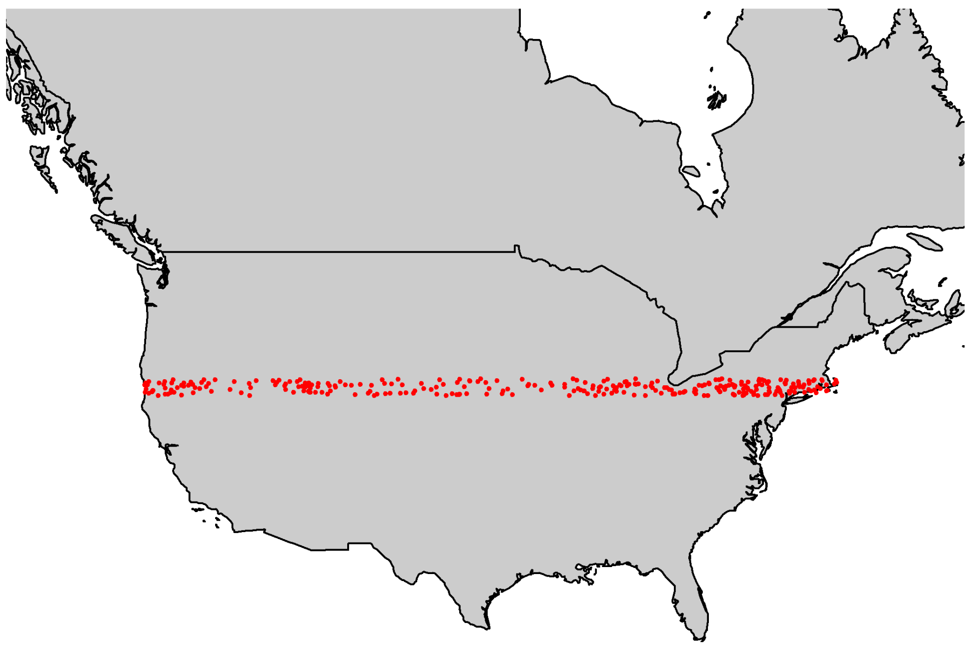


Figure 4.5: An example of two adjacent spatial transects within my sampling grid.

Spatial correlation of Fisher Information

If Fisher Information captures and reduces information regarding abrupt changes in community structure across the landscape, then the values of FI should be spatially autocorrelated. That is, the correlation of FI values should increase as the distance between points decreases. Fisher Information values calculated using Eq. (4.4) are **not** relatively comparable outside of our spatial transects, because the possible values

Figure 4.6: An example of two adjacent spatial transects (12, 13) within my sampling grid.

are unbounded (can take on any value between $-\infty$ and ∞. However, because FI is directly comparabe **within** each spatial transect (e.g., 4.5), we can use use pairwise correlations among two transects (e.g., 4.5) to determine whether values of FI are consistent across space. I calculate the pairwise correlation (Pearson's) among each pair of adjacent spatial transects (e.g., Fig. 4.6). I removed a pair of points if at least one point was missing an estimate for Fisher Information. This occurred when the original longitudinal range of one transect exceeded its pair's range, since I did not interpolate beyond the original longitudinal range.

4.3 Results

4.3.1 Fisher Information across spatial transects

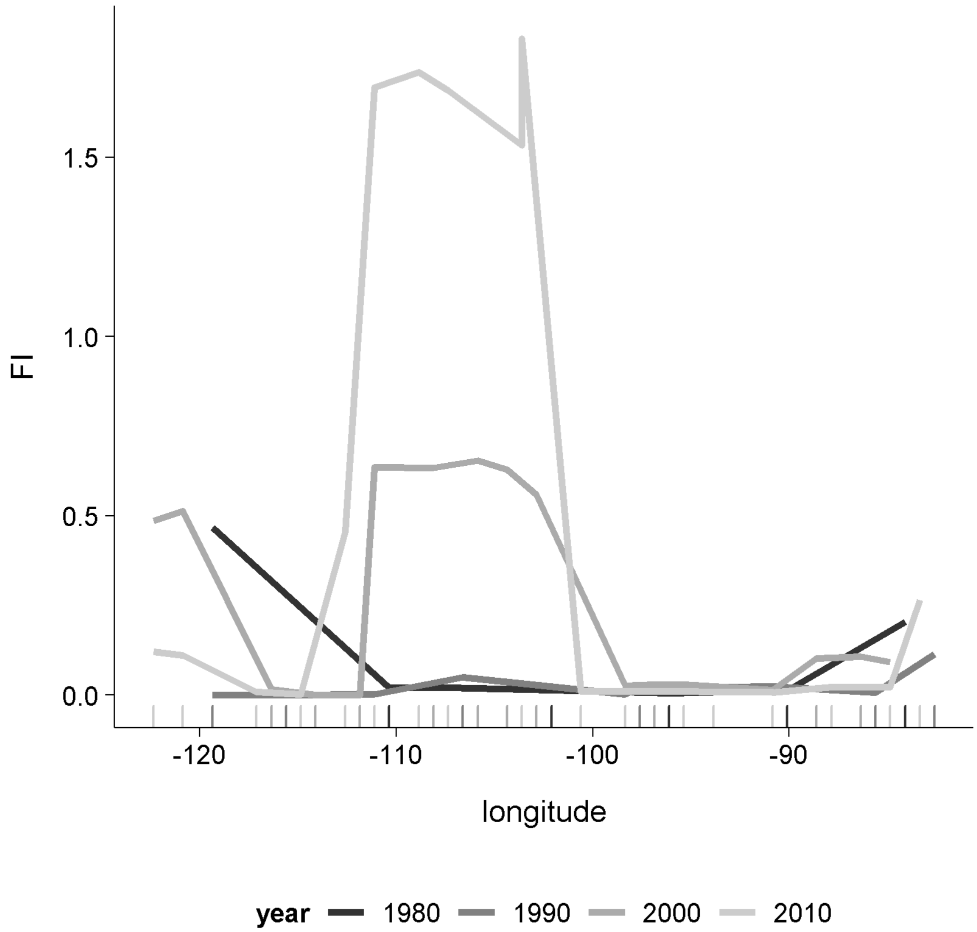


Figure 4.7: Fisher Information calculated for a single transect over time.

Interpreting the Fisher Information is currently a qualitative effort. As suggested earlier, rapid increases or decreases in FI are posited indicate a change in system orderliness, potentially suggesting the location of a regime shift. Using this method yields inconclusive results regarding the location of 'spatial regimes' (Fig. 4.7). Of the three spatial transects analyzed in this chapter (Fig. 4.4), Fig.4.7 is representative of the lack of pattern observed in the Fisher Information values aross transects. I identified no clear pattern within or among spatial transects. Log-transforming the

Fisher Information metric suppresses some of the extreme values, but still does not clearly identify sharp changes in the Fisher Information values.

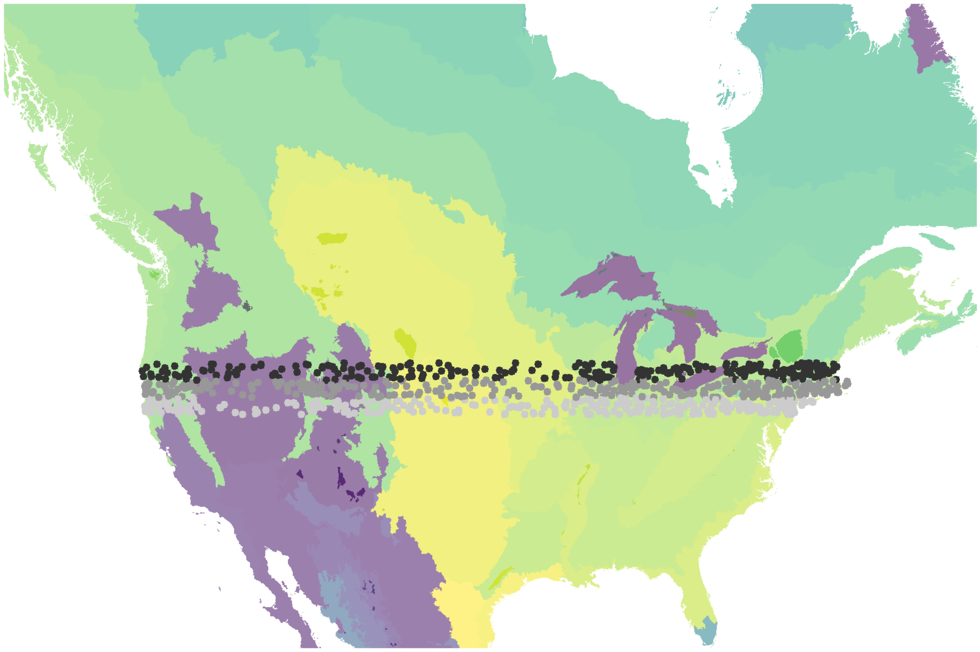
4.3.2 Spatial correlation of Fisher Information

In addition to failing to identfiy clear geological boundaries across large swaths of our study area, (Fig **??**) I also did not identify spatial correlation of Fisher Information among adjacent spatial transects (Fig. 4.8)[1]. For spatially-adjacent transects (e.g, transects 11 and 12, or 12 and 13 in Fig. 4.8), we should expect high and positive correlation values, and these values shoudl stay consistent across time *unless* the spatial transects were separated by an East-West running physical or functional boundary. This is not, however, what I expect in our East-West running transects (Fig. **??**), as the spatial soft-boundaries limiting the distribution and functional potential of avian communities are largely North-South (Fig. @ref(ewRoutes_ecoRegions)). Note spatial transects in Fig. @ref(fig:ewRoutes_ecoRegions) overlap multiple, large spatial ecoregion boundaries, such that we should expect our data to identify these points (boundaries). Upon initial investigation, there are no obvious signs of broad-scale patterns in FI across space (Fig. 4.10)[2]. If Fisher Information is an indicator of spatial regime boundaries, we should expect to see large changes in its value (in either direction) near the edges of functional spatial boundaries (e.g., at the boundaries of ecoregions). No clear regime changes appeared in areas where we might expect rapid changes (e.g., along the 105th meridian West, where a sharp change in altitude occurs).

Numerical investigation of the spatial correlation among adjacent transects also yielded no clear patterns. I did not identify any obvious correlation with changes in

[1] Pairs were compared (column) at select sampling years (rows), and pair-wise correlations among paired transects are presented. Large, positive correlations indicate Fisher Information signals similarly at adjacent spatial transects.

[2] Size indicates value of Fisher Information (values are scaled and centered within transects). Red box (in top panel) indicates extent of bottom panel.



FI values and functional potential (using Omernick Ecoregion Level 2; see Fig. 4.10). Rather than abrupt changes in Fisher Information I found graduaal changes (e.g., see results for years 2000 and 2010 in Figs. 4.11,4.10).

4.4 Discussion

The Fisher Information measure was introduced as a method to avoid some analytical issues related to complex and noisy ecological data (Karunanithi et al., 2008), and has also been suggested as an indicator of *spatial* regimes (**???**). I found no evidence suggesting Fisher Information [Eq. (4.4)] can identify ‘spatial regimes’. Further, the absence of autocorrelation among spatially adjacent transects suggests Fisher Information may not be a realiable indicator of changes in bird community structure.

Although the Fisher Information equation [Eq. (4.4)] used in this study is a relatively straightforward and fairly inexpensive computational calculation, extreme care should be taken when applying this index to ecological data. Fisher Information is capable of handling an infinite number of inputs (variables), and given sufficiently

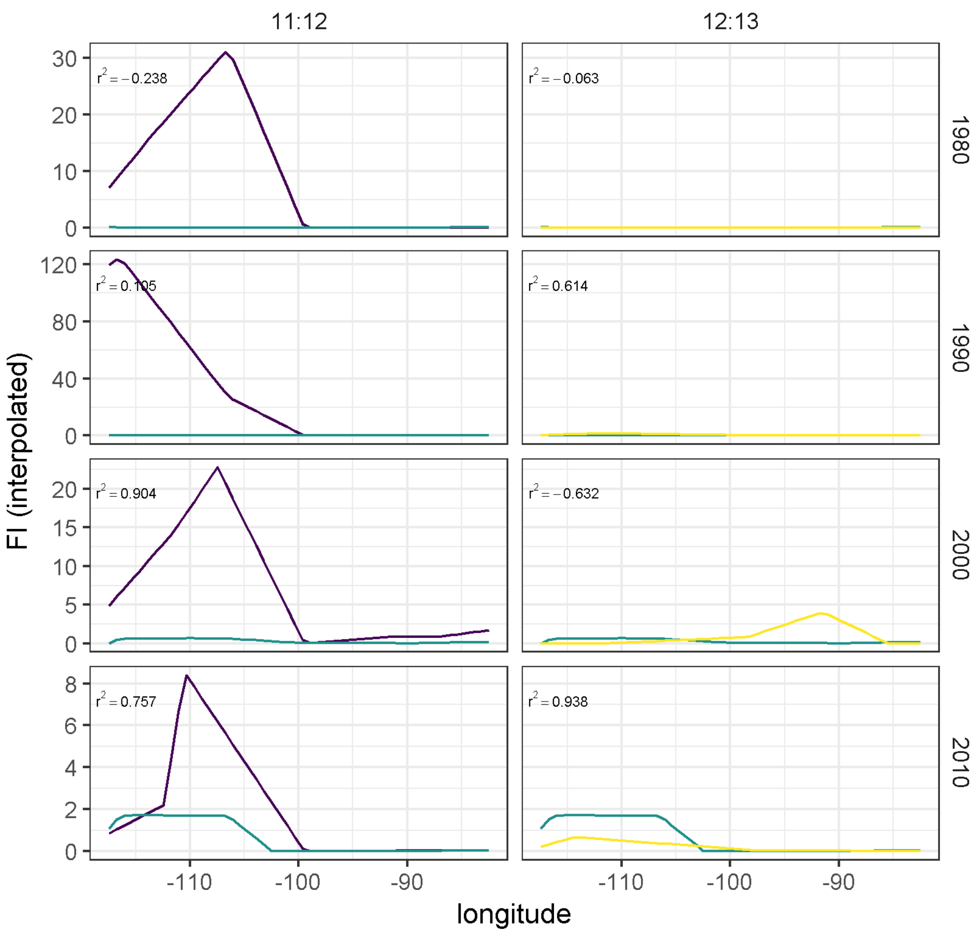


Figure 4.8: Pairwise relationships of Fisher Information (interpolated values) of spatially adjacent transects over time.

low window size paramters, can technically calculate an index value for only two observations. It is important that the user understands the assumptions of identifying 'regime shifts; using Fisher Information, since the efficacy of this method has not been yet subjected to rigorous tests (but see 6). There are three primary assumptions required when using Fisher Information to estimate relative orderliness within ecological data (D. A. L. Mayer et al., 2007):

1. the order or state(s) (s) of the system is observable, 1. any observable change in the information observed in the data represents reality and the variables used in the analyses will not produce false negatives, and 1. changes in I presumed to be regime shifts do not represent the peaks of cyclic (periodic) patterns.

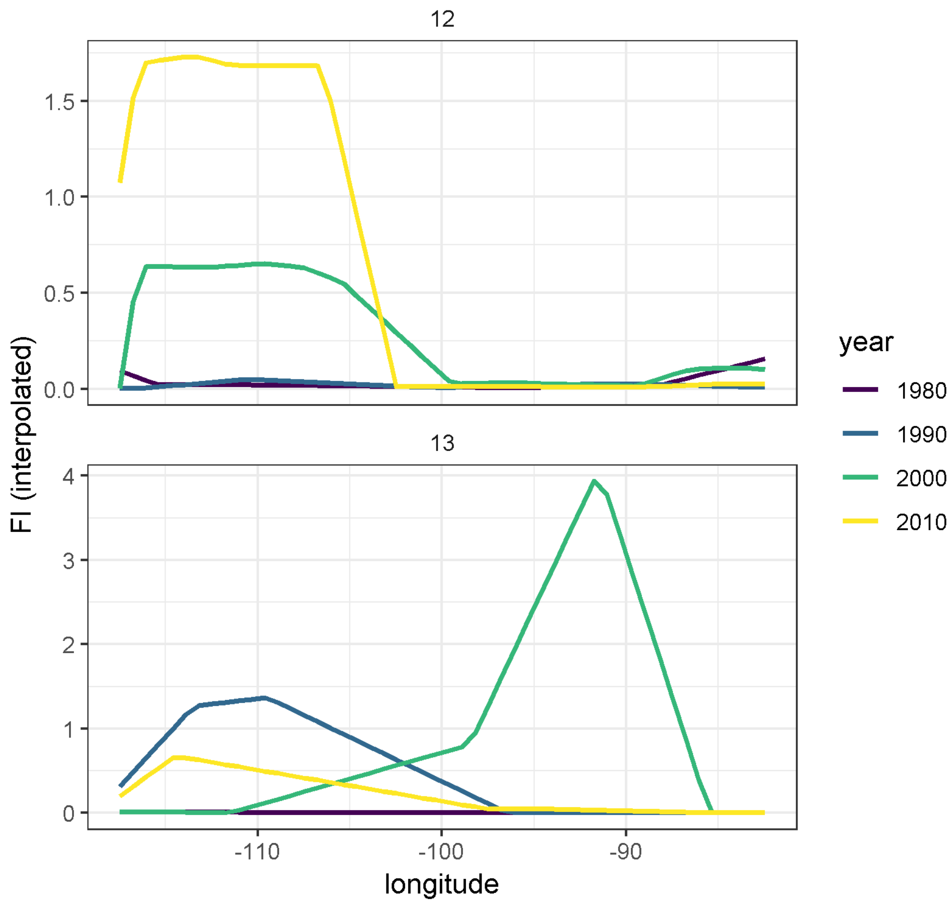


Figure 4.9: Fisher Information of two transect pairs over time.

The first assumption is one of philosophical debate and is thus not controllable. To attempt to control for false negatives, the user should take caution in her choice of input variables. In the the case of a high dimensional data, relativization and/or variable reduction measures may be useful (Rodionov 2005). However, Fisher Information does not convey information on how specific variables relate to the calculated index. Finally, we can take measures to account for cyclic behavior in the data by ensuring integration periods capture at one full cycle of the system and, given sufficently high number of observations, increasing the integration period may also alleviate some issues related to irreducible error (white noise).

The lack of patterns identified using Fisher Information may be influenced by one or more of the following: (1) the Breeding Bird Survey data collection scheme was

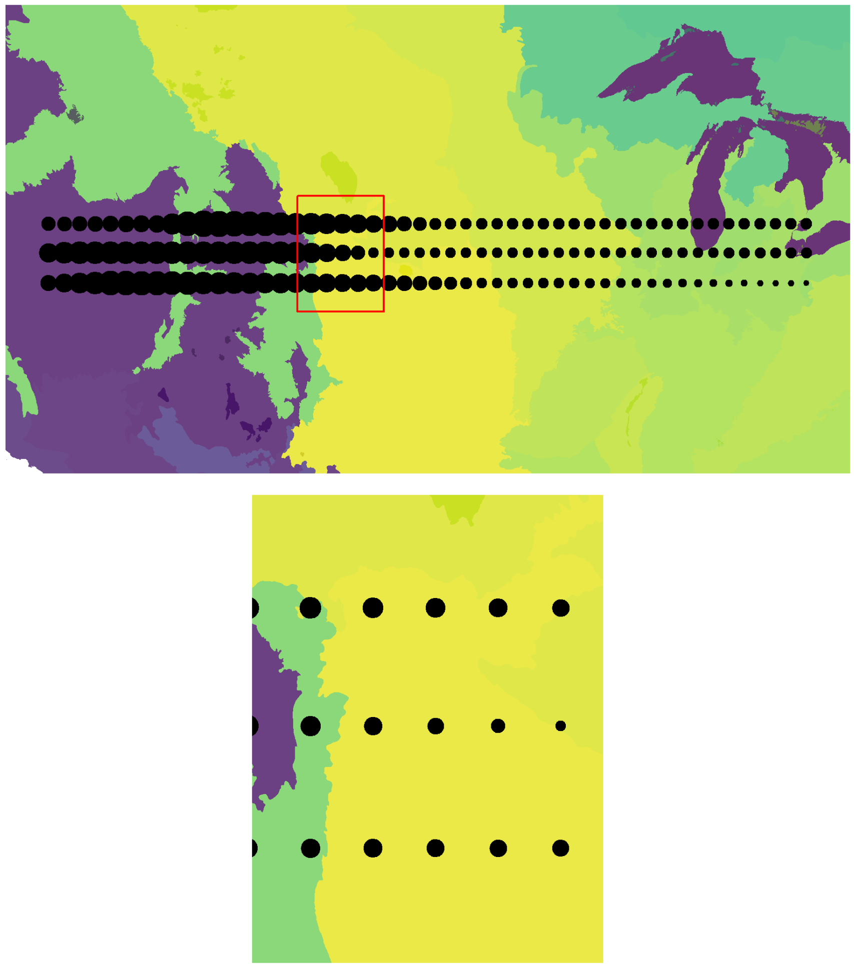


Figure 4.10: No patterns of abrupt change detected in Fisher Information along three transects in year 2010

designed to estimate and track **species** trends and not changes in entire communities; (2) these data consist of < 50 time points, and for some BBS routes much fewer. Ecological processes affecting large regions in this study area (e.g., the Central Great Plains) operate on larger time scales (i.e., $>> 50$ points). A mismatch among the ecologically relevant scales and the temporal resolution and extent of our data may influence the ability of this index to capture large-scale changes in whole bird communities.

Aside from the typical biases associated with the BBS data (e.g., species detection probability, observer bias), there are additional considerations to be made when using

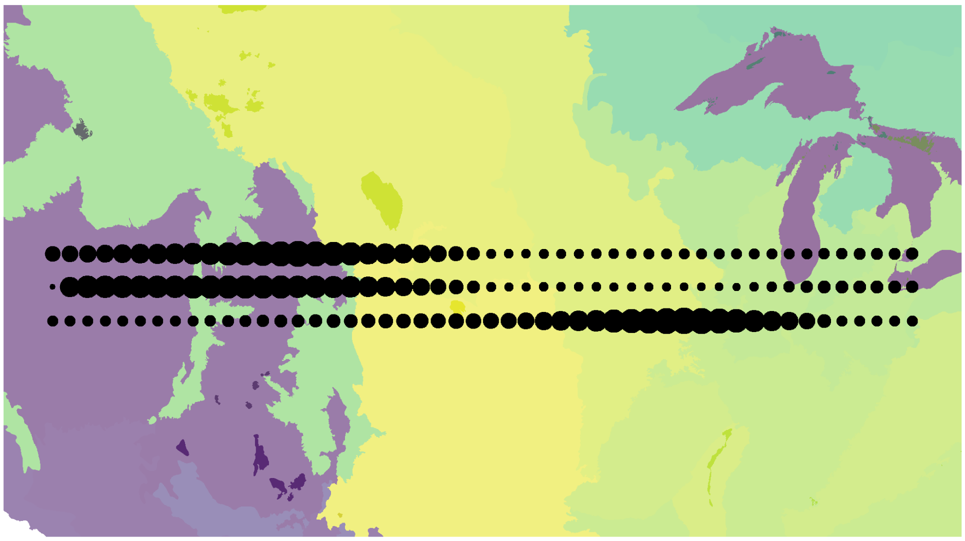


Figure 4.11: Fisher Information (scaled and centered; point size positively correlated with value) against ecoregion boundaries (EPA Level 2).

these data to identify 'spatial regimes'. Breeding Bird Survey routes are spaced apart so as to reduce the probability of observing the same individuals, but birds which fly (especially in large flocks) overhead to foraging or roosting sites have a higher probability of being detected on multiple routes. We have, however, removed these species (waders, shorebirds, waterfowl, herons) from analysis. Regardless, this study assumes there is potential for each unique BBS route to represent its own state. If routes were closer together, it is more probable that the same type adn number of species would be identified on adjacent routes. Therefore, if this method does not detect slight changes in nearby routes which occupy the same 'regime', then it follows that the method is sensitive to loss or inclusion of new species, which are spatially bounded by geological and vegetative characteristics. What new information does this give us about the system? Fisher Information reduces and removes the dimensionality of these middle-numbered systems, which omits critical information.

Effective regime detection measures should provide sufficient evidence of the drivers and/or pressures associated with the identified regime shifts (Mac Nally et al., 2014). The Fisher Information index collapses a wealth of data into a single metric, thereby foregoing the ability to relate state variables to the observed changes in Fisher Information, unlike other dimension reduction techniques. For example, loadings, or the relative influence of variables on the ordinated axes, can be derived from a Principal Components Analysis–this cannot be achieved using Fisher Information. If Fisher Information clearly suggested a spatial regime boundary or shift, a before-and-after post-hoc analysis of the regional community dynamics might confirm the regime shift occurrence.

4.4.1 Efficacy of Fisher Information as a spatial RDM

This study found no evidence suggesting Fisher Information accurately and consistently detects spatial boundaries of avian communities. Rapid changes in either direction of Fisher Information is suggested to indicate of a regime shift (Mayer, Pawlowski, & Cabezas, 2006, (**???**) 2012). Although this interpretation has been applied to multiple case studies of Fisher Information, there is yet a statistical indicator to objectively identify these abrupt changes. After calculating the Fisher Information for each spatial transect (Fig. 4.4) during each sampling year, I used pairwise correlation to determine whether spatial autocorrelation existed among pairs of spatial transects. If some set of points are close in space and are *not* separated by some physical or functional boundary (e.g., an ecotone, high altitude rock formations), then the Fisher Infomration calculate should exhibit a relatively high degree of spatial autocorrelation that is consistent over time. It follows that the correlation coefficient of spatially adjacent transects should be similar, diverging only as the distance beteween the transects differs and/or a functional or physical boundary separates them.

Several questions remain regarding the efficacy of Fisher Information as a regime detection measure in both spatial and temporal data. If signals of regime shifts do exist, it is clearly not possible to identify them using visual interpretation. I also did not find evidence to suggest spatial autocorrelation of the calculations. I suggest future studies of Fisher Infomration focuses on temporal, rather than spatial data. Potential areas of research and questions include:

1. Relationship of Fisher Information to likelihood ratio-based unsupervised change-point detection algorithms (e.g., ChangeFinder (S. Liu, Yamada, Collier, & Sugiyama, 2013)).

2. Sensitivity of Fisher Information to data quality and quantity [this is explored in Chapter 6].

3. What, if any, advantages does FI have over other density estimation techniques?

4. Does FI provide signals in addition to or different than geophysical and vegetative (e.g. LIDAR) observations (data)?

Chapter 5

Velocity (v): using rate-of-change of a system's trajectory to identify abrupt changes

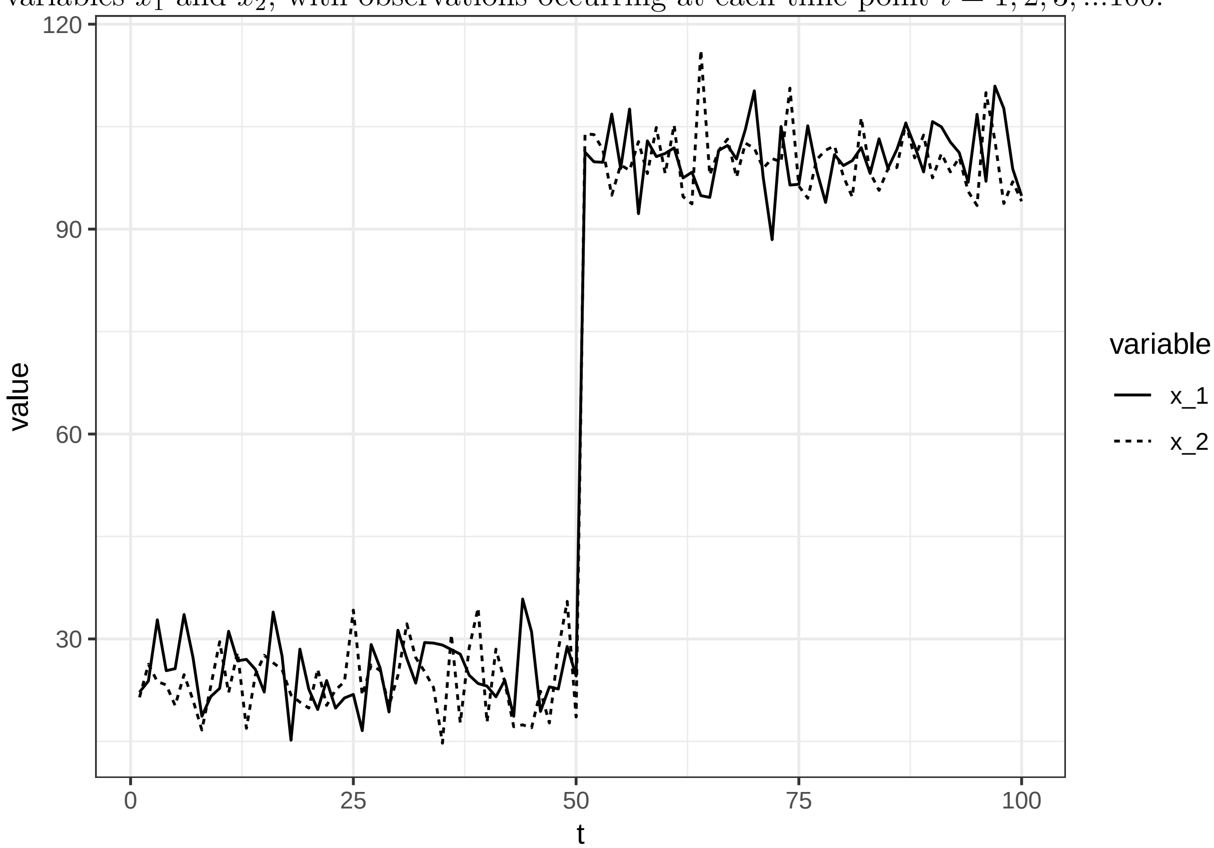
5.1 Introduction

In this Chapter I describe the steps for calculating a 'new' metric, **system velocity**, for reducing the dimensionality and identifying abrupt shifts in high dimensional data. Although this is the first instance of this calculation to, alone, be suggested as a regime detection metric, it has been used as part of a larger series of calculations of the Fisher Information metric [see 3], first introduced in B. D. Fath et al. (2003). Below, I describe the steps for calculating system velocity, simply defined as the cumulative sum of the squared change in all state variables over a period of time.

5.2 Data and Methods

5.2.1 Theoretical system example: two-species time series

Consider a system (Fig. ??) with N state variables (x_i), with observations taken at time points, t. System velocity is calculated as the cumulative sum over time period t_0 to t_j, as the total change in all state variables, $\{x_1...x_N\}$, between two adjacent time points, e.g., t_j and t_{j+1}, denoted $t_{j,j+1}$. I use a simple, two-variable system to demonstrate the calculation of each step below. The system comprises variables x_1 and x_2, with observations occurring at each time point $t = 1, 2, 3, ...100$.



5.2.2 Steps for calculating system velocity, v

First, we calculate the change in each state variable, x_i, between two adjacent points in time, t_j and t_{j+1}, such that the difference, $x_{t_{j+1}} - x_{t_j}$ is assigned to the latter time point, t_{j+1}. For example, in our toy data, we use observations at time points $t = 1$ &

$t = 2$ (Fig. 5.1). For all examples in this chapter, the state variables x_1 and x_2 were drawn from a normal distribution (using function *rnorm*), with parameters $\bar{x}_i$ (mean) and σ_i (sd) for 100 time steps, t. The regime shift occurs at $t = 50$, where a shift in either or both $\bar{x}_i$ or σ_i.

Step 1: Calculate Δx_i

The first step in calculating v is to obtain the change in values for each state variables, x_1 and x_2 between two consecutive time points (e.g., from $t = 1$ to $t = 2$:

$$\begin{aligned}\Delta x_1 &= x_{1_{t=2}} - x_{1_{t=1}} \\ \Delta x_2 &= x_{2_{t=2}} - x_{1_{t=1}}\end{aligned} \tag{5.1}$$

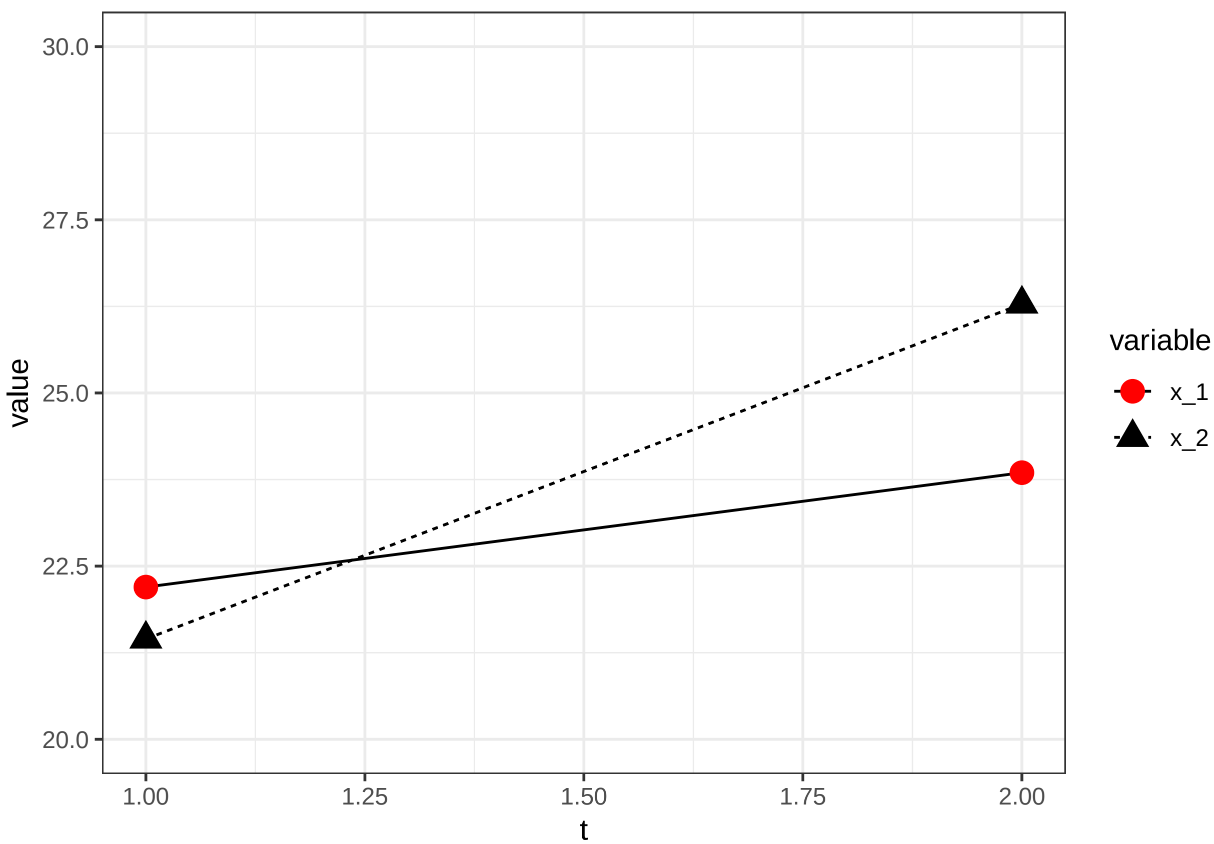


Figure 5.1: Data used to calculate velocity at the first two time points, t_1 and t_2.

Step 2: Calculate $\sqrt{(\sum_i^N \Delta x_1^2)}$

After calculating the differences for each state variable, we will next calculate the total change in the system over the time elapsed, following Pythagora's theorem,

$$X_1^2 + X_2^2 = s^2 \tag{5.2}$$

where s represents the total change in the system, and X_1 and X_2 represent the changes in all state variables $(x_{i_{t=2}} - x_{i_{t=1}})$. We achieve this by first squaring the differences obtained in Eq. (5.1):

$$\begin{aligned} &(x_{1_{t=2}} - x_{1_{t=1}})^2 \\ &(x_{2_{t=2}} - x_{2_{t=1}})^2 \end{aligned} \tag{5.3}$$

Step 3: Use Pythagorean theorem to isolate s

Next, we isolate s in Eq. (5.2), capturing the total change in all state variables into a single measure by taking the 2nd root of the squared sums of all x:

$$\begin{aligned} \textstyle\sum_{i=1}^N \Delta x_i = \textstyle\sum_{i=1}^N (x_{t_{i+1}} - x_{t_i})^2 \\ = \Delta s \\ = \sqrt{([x_{1_{t=2}} - x_{1_{t=1}}]^2 + [x_{2_{t=2}} - x_{2_{t=1}}]^2)} \end{aligned} \tag{5.4}$$

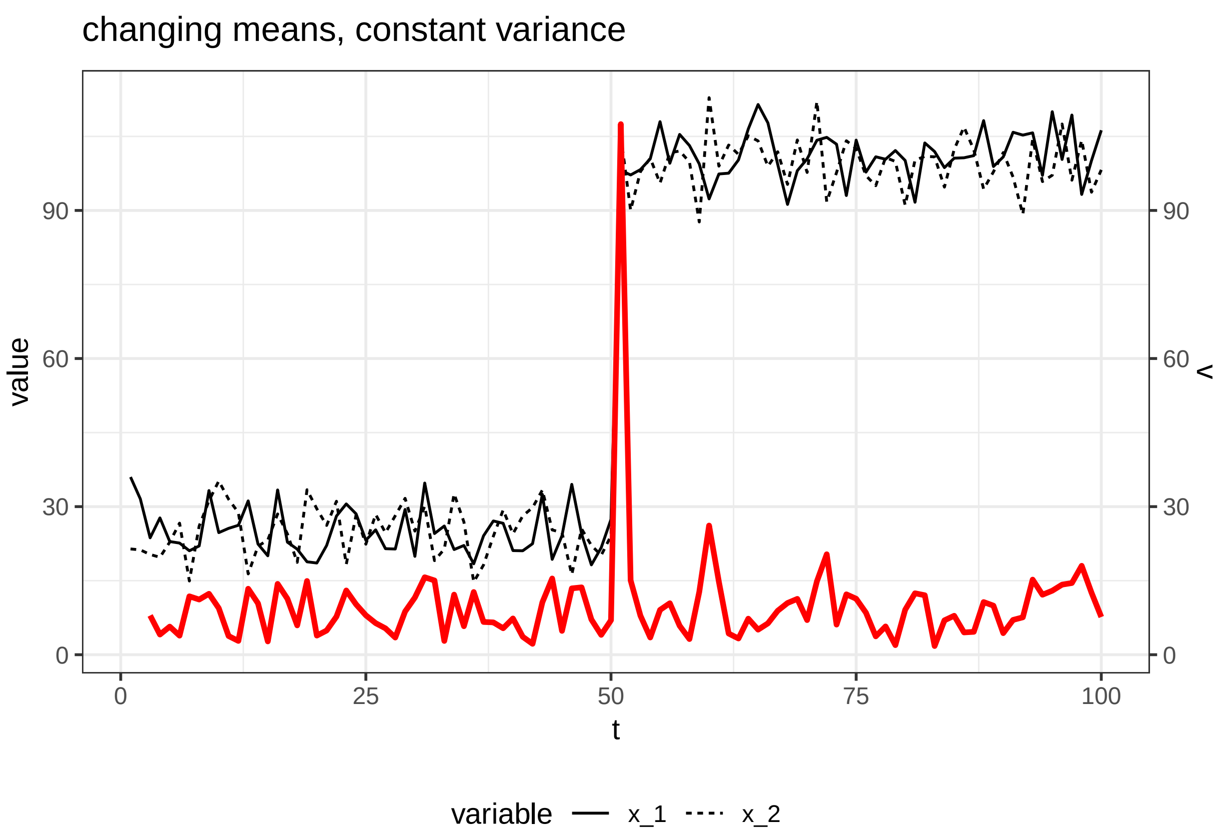
We now have a single measure, Δs (Eq. (5.4)), for each pair of time points in our N-dimensional system. It is obvious that Δs will always be a positive value, since we took the 2nd root of a squared value. Although discussed in a later section, it is important to note that this value is not unitless–that is, our example system takes on the units of our state variables, x_1 and x_2. Because we are interested in identifying abrupt changes in the entire system, we calculate the cumulative sum of Δs at every time point, such that:

$$s = \sum_{t=1}^{T} \Delta s \tag{5.5}$$

Step 4: Calculate velocity, v (or $\frac{\Delta s}{\Delta t}$)

Finally, we calculate the **system velocity**, v (or $\frac{\Delta s}{\Delta t}$), by first calculating the change in s (Eq. (5.5)), and then divide by the total time elapsed between consecutive sampling points:

$$v = \frac{s_{t+1} - s_t}{\Delta t} \tag{5.6}$$



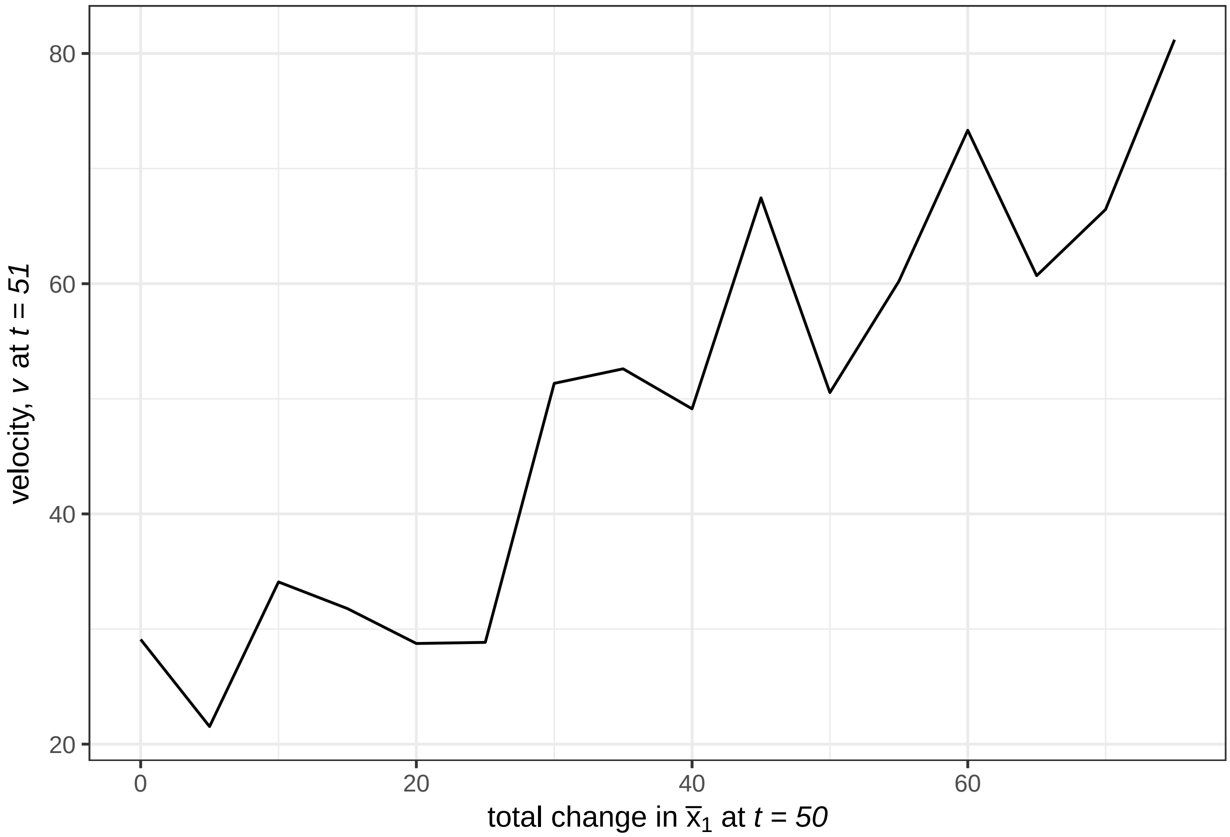
The steps for calculating velocity [Eq. (5.6)] are demonstrated using the first five time points of our toy system (Fig. **??**) in Table **??**.

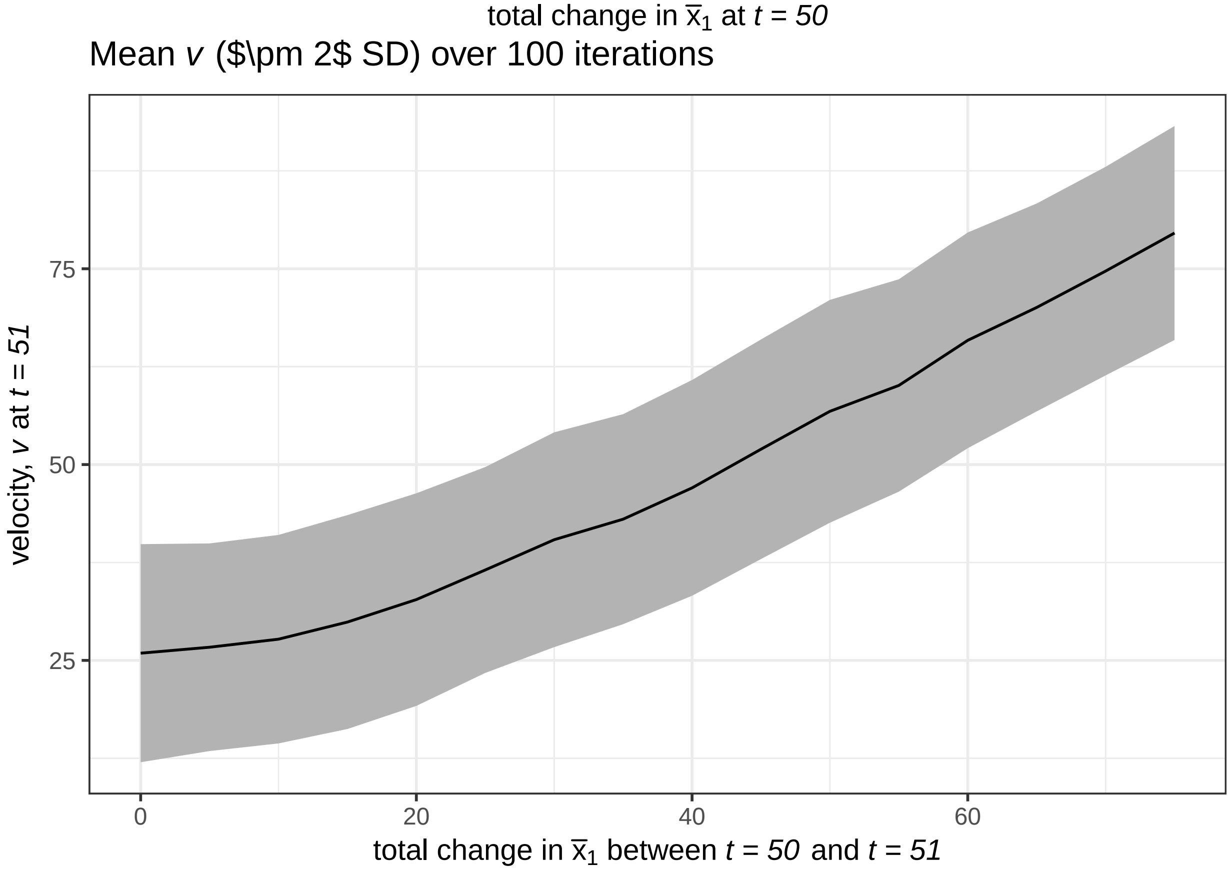
5.2.3 Velocity v performance under varying mean and variance in the toy system

I simulated 10,000 random draws of the toy system, which experiences a rapid shift at $t = 50$, while varying two each of the following system paramters at the regime shift: $\bar{x}_1$, increased the mean value of x_1 σ_1, change in variance of x_1 Simulations consisted of 10,000 random samples drawn from the normal distribution for each paramter, I randomly drew the toy system samples 10,000 times under increasing values of $\bar{x}_1$ and σ_1. To identify patterns in the influence of paramter values on velocity, I present the mean values of v across all simulations, with confidence intervals of ± 2 standard deviations. As mentione above, the state variables x_1 and x_2 were drawn from a normal distribution (using function *rnorm*), with parameters $\bar{x}_i$ (mean) and σ_i (sd) for 50 time steps, t.

Varying post-shift mean

I examined the influence of the magnitude of change in x_1 in the period before (pre; $t < 50$) and after (post; $t \geq 50$) by varying the mean parameter, $\bar{x}_1$ in the set $W = \{25, 30, 35, ...100\}$ (Figs. **??**,**??**). As expected, the magnitude of v increased linearly as the total difference between $\bar{x}_{1_{pre}}$ and $\bar{x}_{1_{post}}$ increased (**??**). This is not surprising, as s increases as the total change in abundance across the entire sytem increases (Eq. (5.5)), therefore, the potential maximum of v also increases. This may indicate that v, while capable of identifying large shifts in data structure, may not pick up subtle changes (i.e. lower effect sizes).


velocity, v at t = 51
80
60
40
20
0
20
40
60
total change in $\overline{x}_1$ at t = 50


Mean v (± 2 SD) over 100 iterations
velocity, v at t = 51
75
50
25
0
20
40
60
total change in $\overline{x}_1$ between t = 50 and t = 51

Varying post-shift variance

In the previous example, variance was constant before and after the shift at $t = 50$. To determine whether the signal emitted by v at the regime shift is lost with increasing variance, I varied the variance parameter, σ_1 in the set $W = \{1, 2, 3, ...25\}$. The variance for both state variables prior to the regime shift, σ_1 and σ_2, was 5, with the change occurring in σ_{1post}. Sytem velocity v appears senstive to increases in the variance at the point of the regime shift (Figs. **??**, **??**). This extreme sensitivity of v to σ_{post} (Fig. **??**) is unsurprising, given the fact that, without smoothing the derivatives, the tangential speed of a 'noisy' variable will always be noisy itself (see Figs. **??**, **??**, **??**, **??**).

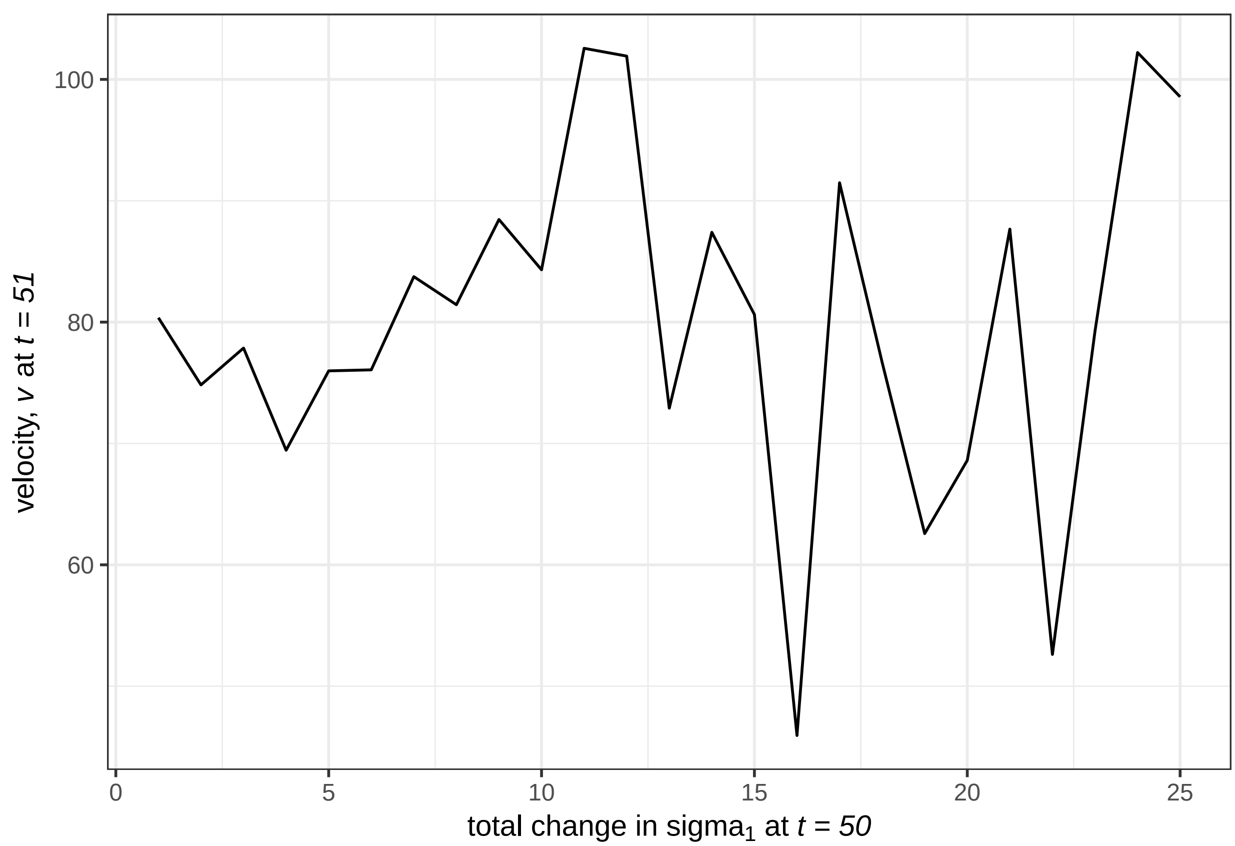


Figure 5.2: High variance of velocity (v) in a single iteration ($N_{iter} = 1$, seed $= 123$) of simulations as we increase σ_1 at $t = 50$.

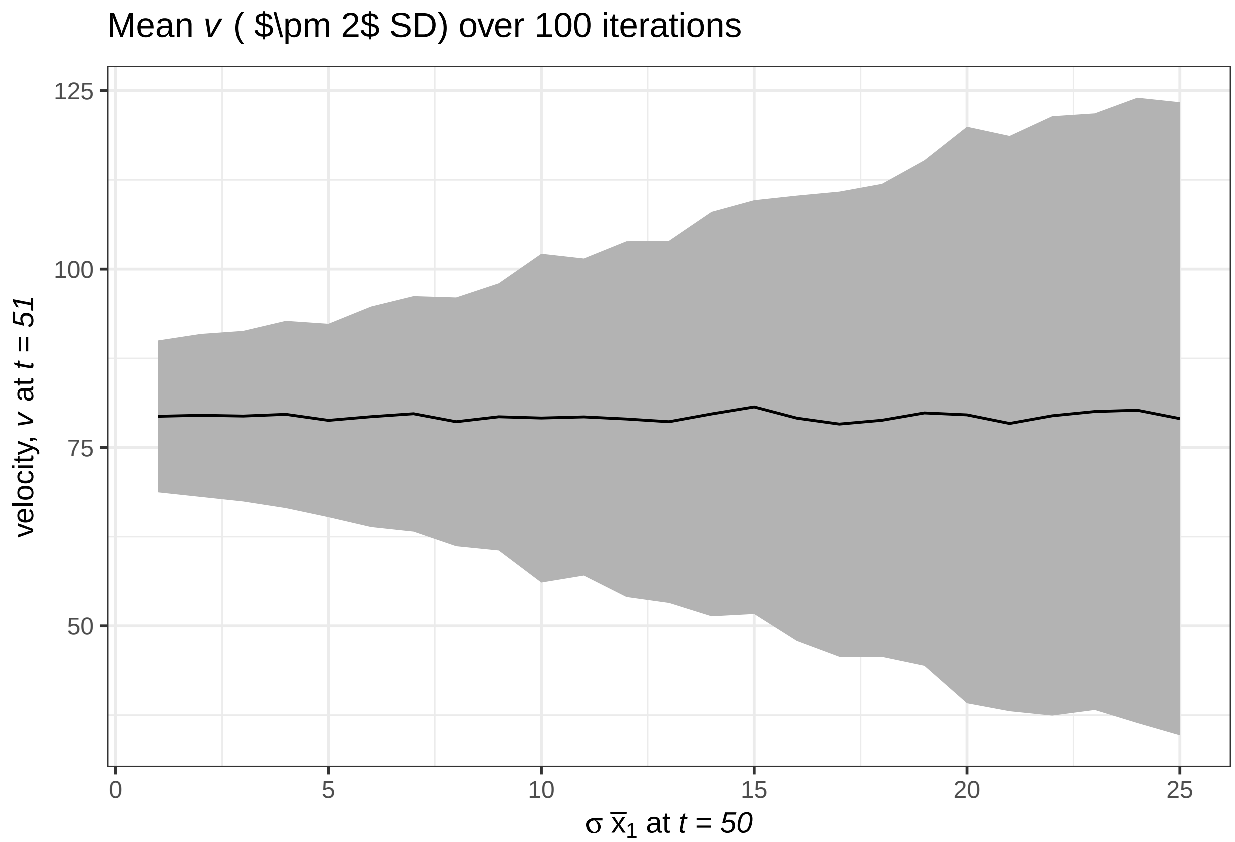
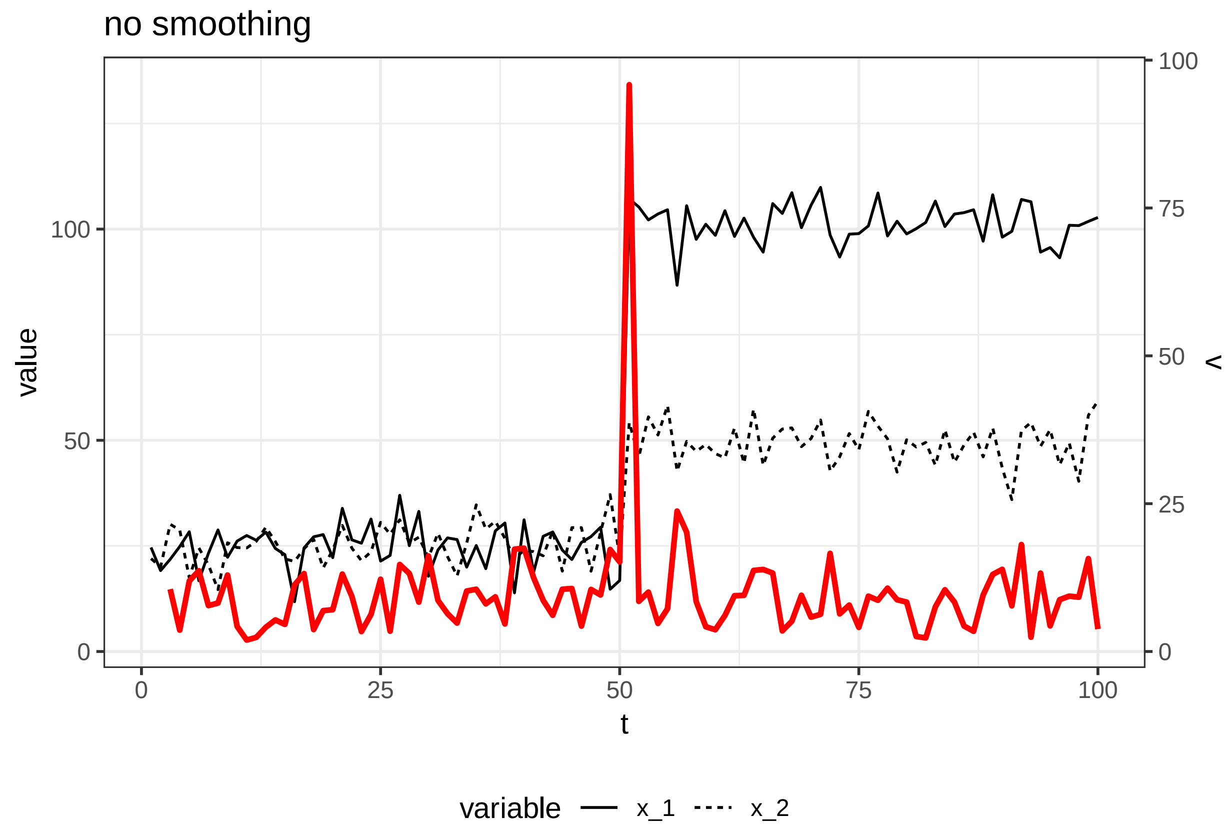
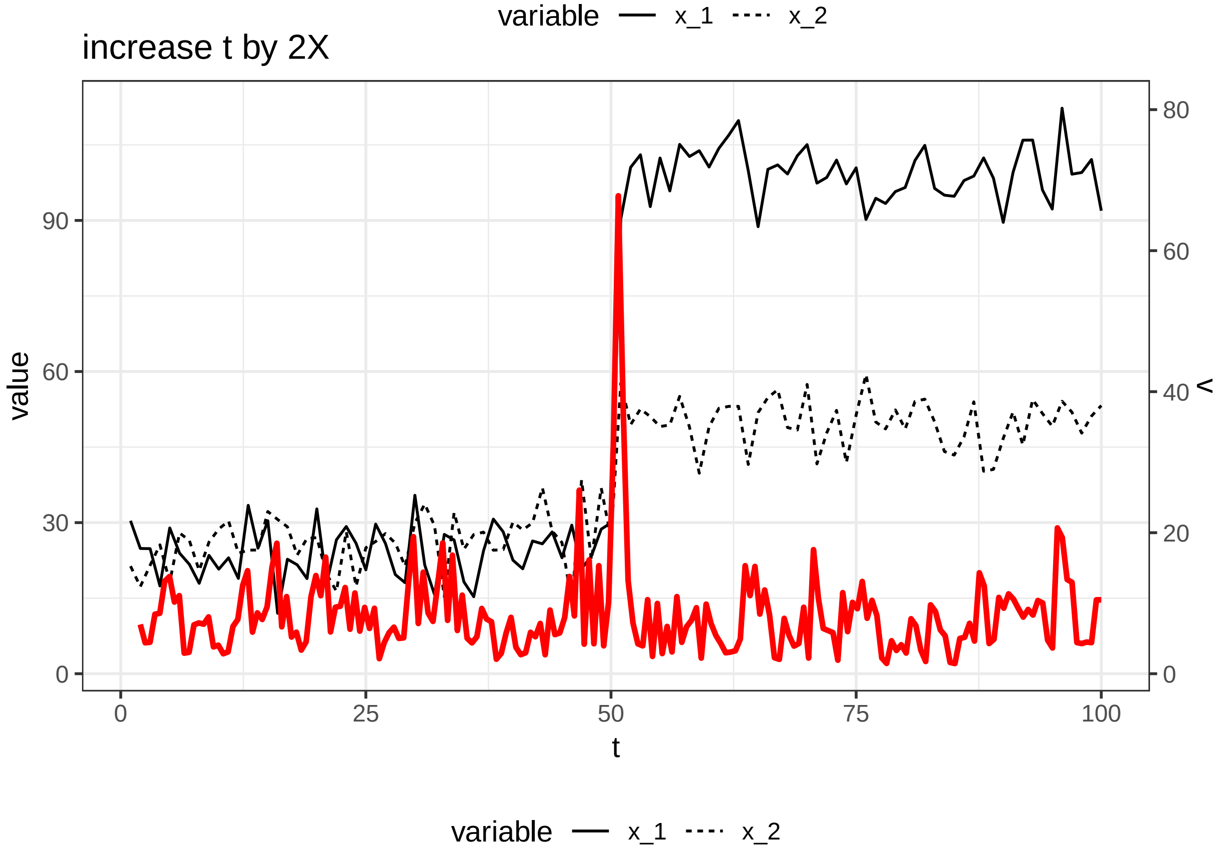


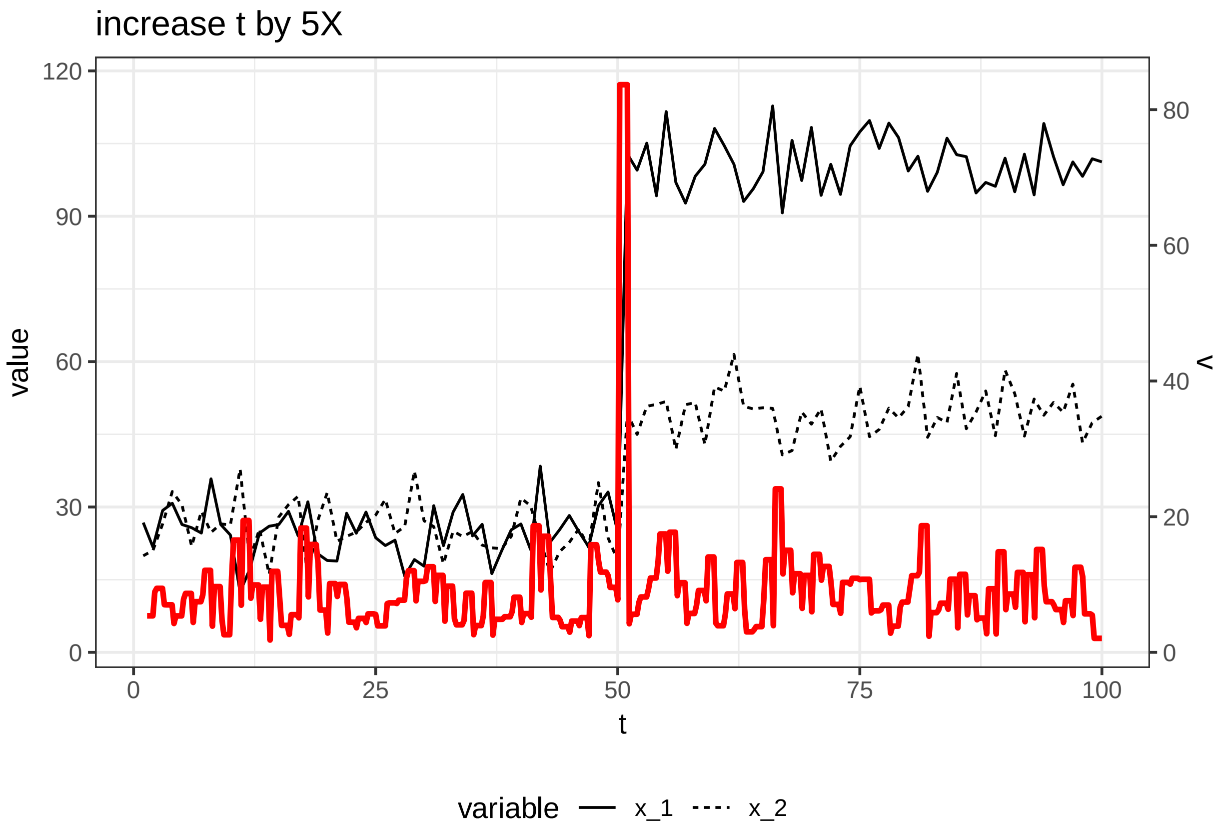
Figure 5.3: Average ($\pm$2 SD) velocity (v) worsens as the variance of $\bar{x}_{2_{t=50(post)}}$ (post shift) increases. $\bar{x}_{1_{pre}} = 25$, $\bar{x}_{1_{post}} = 100$, $\bar{x}_{2_{pre}} = 25$, $\bar{x}_{2_{post}} = 50$, $\sigma_{1_{pre}} = 5$, $\sigma_{2_{pre,post}} = 5$

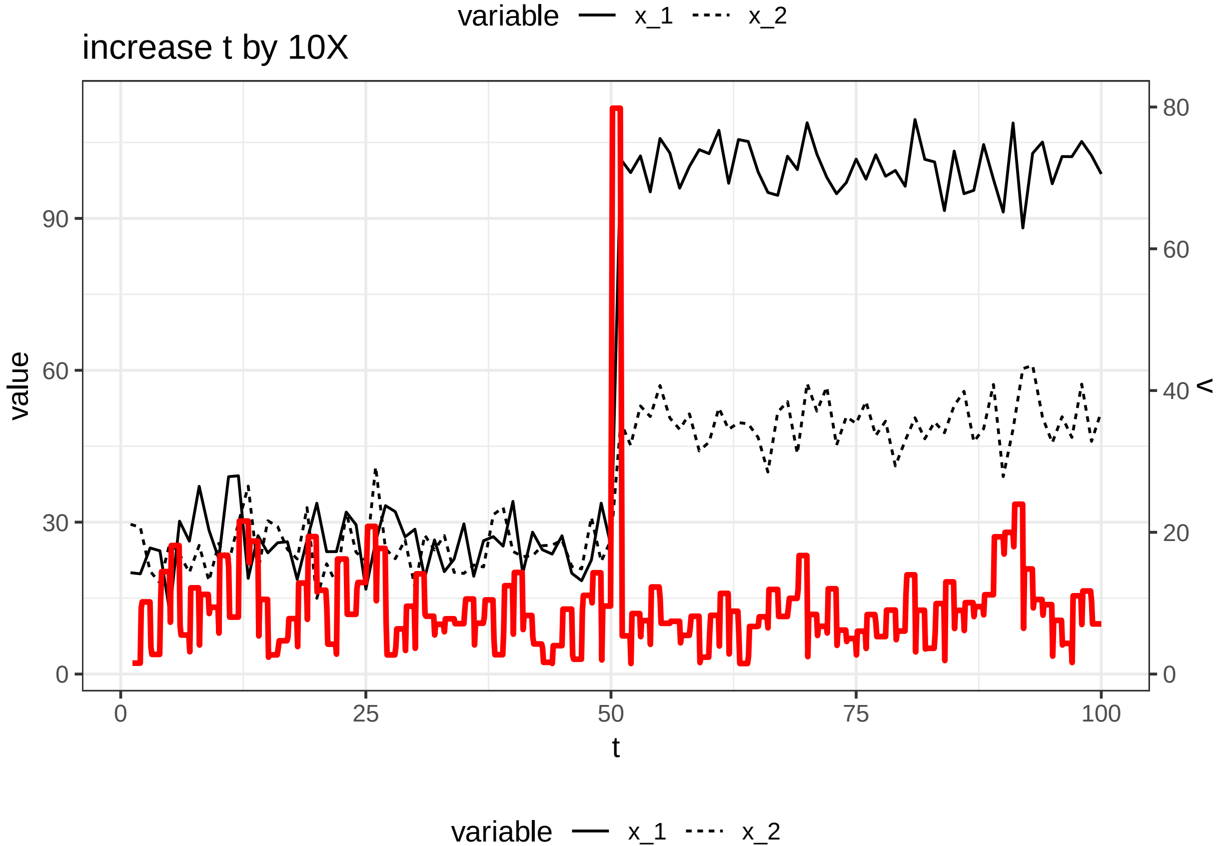
Smoothing the data prior to calculating v

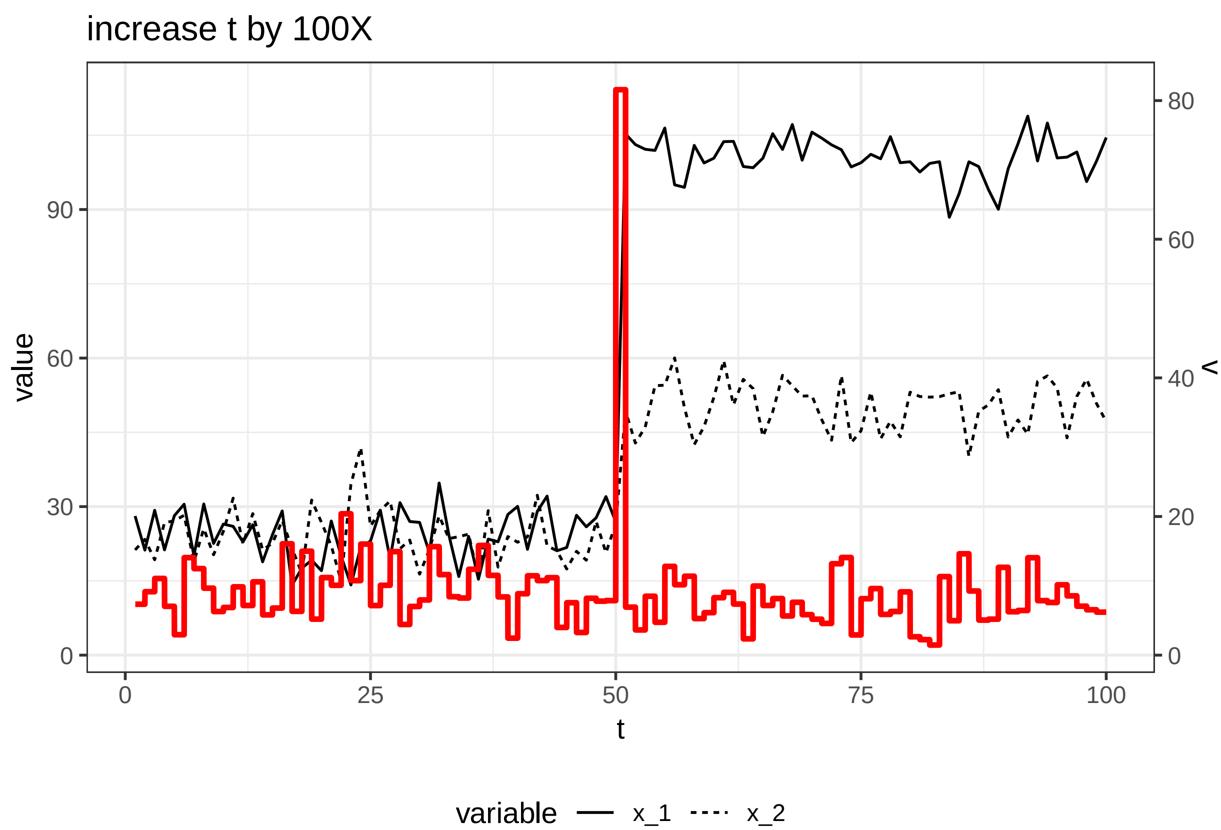
To ameliorate the influence of noise (e.g. Fig. **??**) on the regime shift signal in v, I used linear approximation techniques in attempt to smooth the velocity (derivatives). I used the function *stats::approx* to interpolate values of x_1 and x_2 to regularly-spaced time points in the set $t = \{1 : 100\}$, and then calculated v as described in the steps above (Eqs. (5.1):(5.6)). Increasing the number of points (t) at which the original state variables were smoothed did not influence the amount of noise surrounding the signal of the regime shift (at $t = 50$) in system velocity, v (Fig. **??**).


no smoothing
value
100
50
0
t
0
25
50
75
100
v
75
25
variable
x_1
x_2


increase t by 2X
value
90
60
30
0
t
0
25
50
75
100
v
80
40
20
variable
x_1
x_2


increase t by 5X
value
120
90
60
30
0
80
60
40
20
0
v
0
25
50
75
100
t
variable
x_1
x_2


increase t by 10X
value
90
60
30
0
80
60
40
20
0
v
0
25
50
75
100
t
variable
x_1
x_2



5.2.4 Performance of velocity using empirical data: paleodiatom community example

To gather baseline information on the use of velocity in empirical systems data, I calculated velocity for the paleodiatom system described in Chapter 6 (see also Appendix ??. Briefly, the paleodiatom community comprises 109 time series over a period of approximately 6936 years (Fig. 5.4). As elaborated in T. L. Spanbauer et al. (2014), the paleodiatom community is suggested to have undergone regime shifts at multiple points. These abrupt changes are apparent when exploring the relative abundaces over time, as there are extreme levels of species turnover at multiple points in the data (Fig. 5.4). Using Fisher Information and climatological records, T. L. Spanbauer et al. (2014) suggest that regime shifts in this system at approximately 1,300 years before present (where present is equal to year 1950). T. L. Spanbauer et

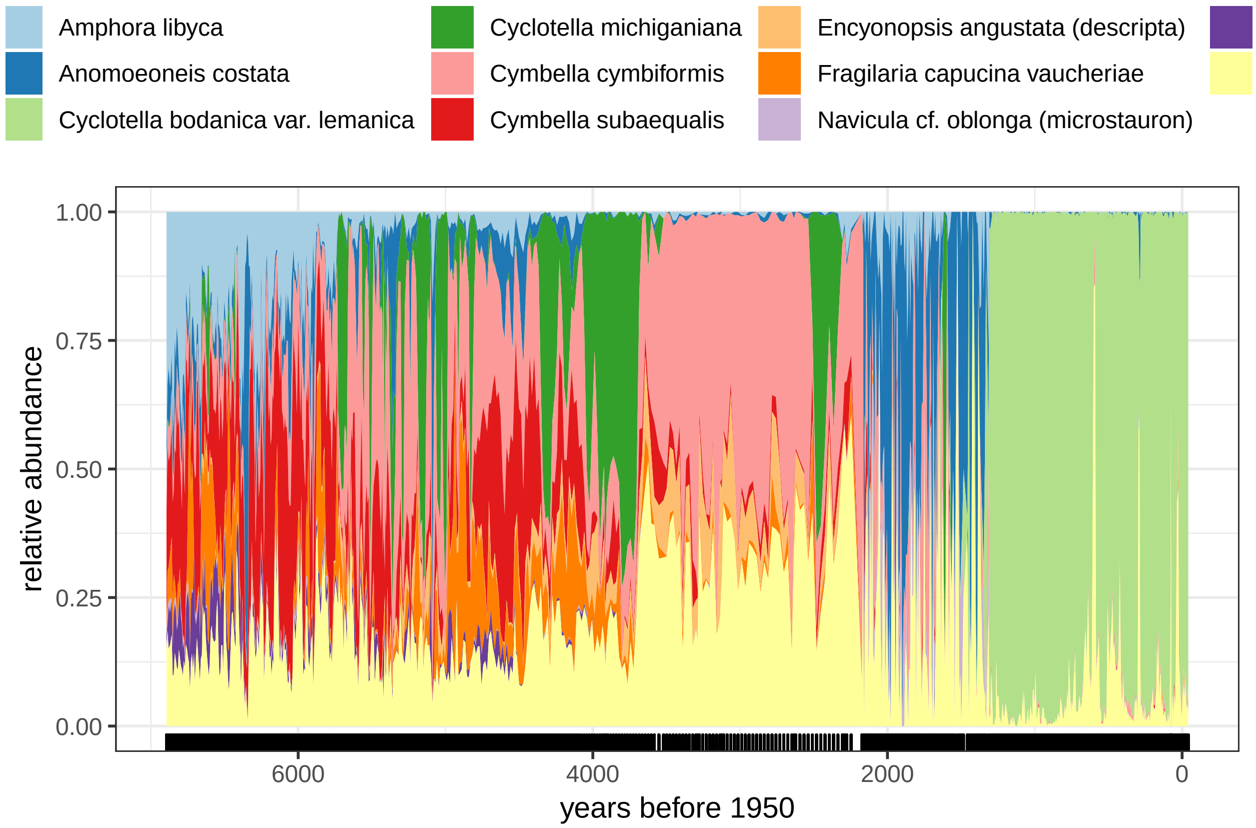
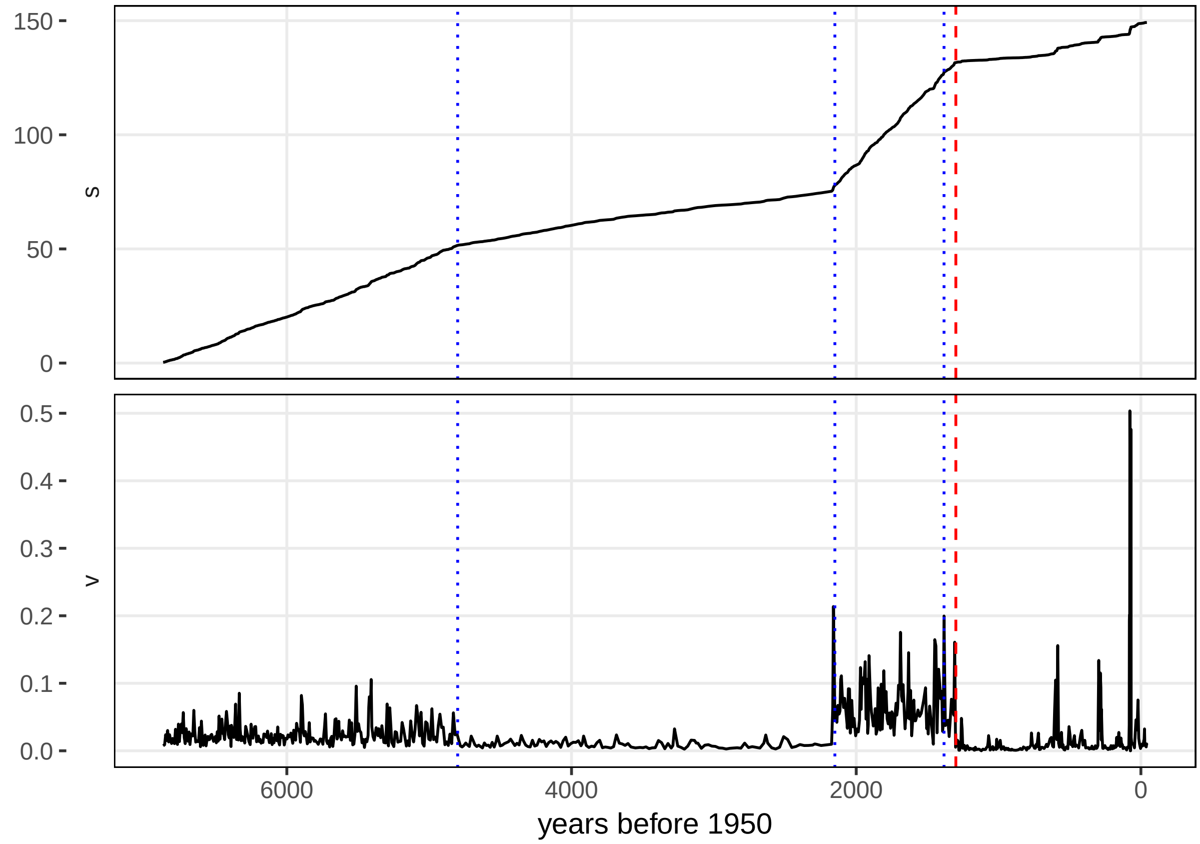
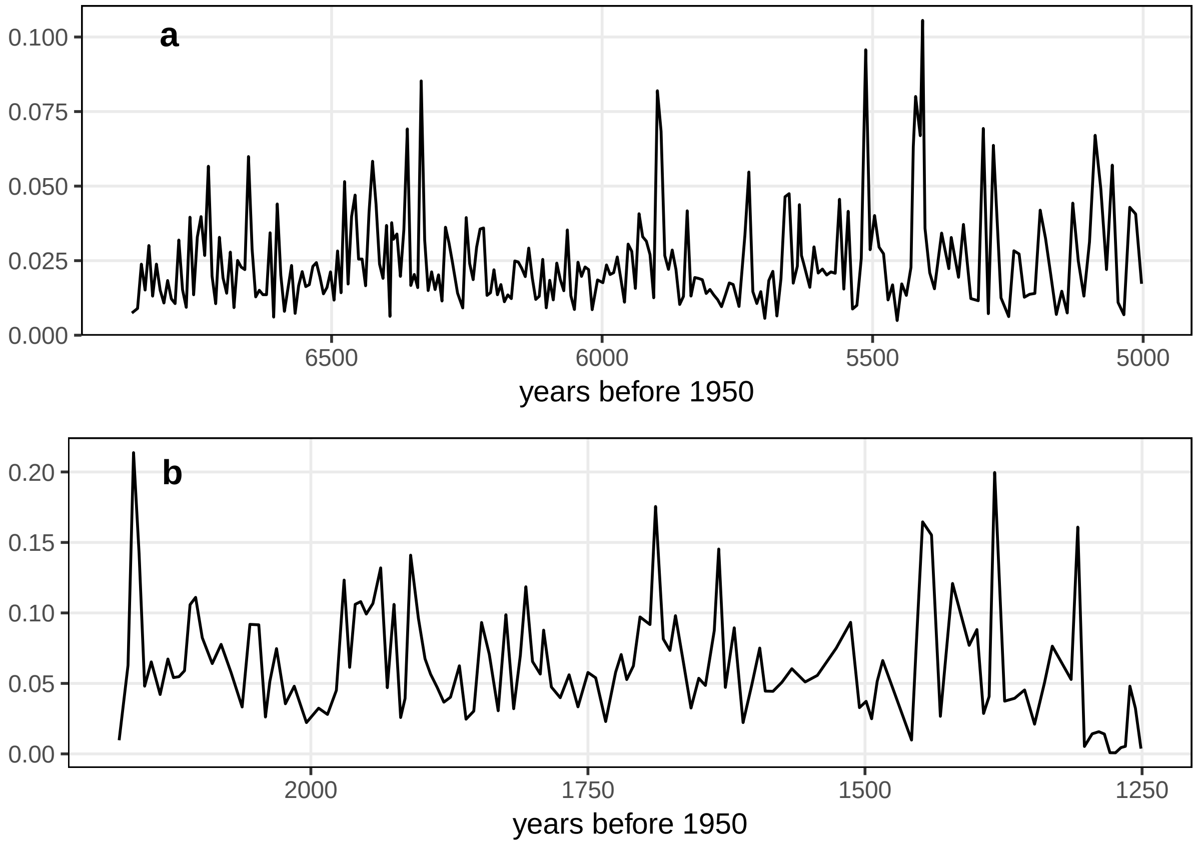


Figure 5.4: Relative abundances of the most common diatom species in the time series. Few species dominate the data over the entire time series, and turnover is apparent at multiple observations.

al. (2014) used different regime detection metrics coupled with regional climatological events to identify regime shifts in the system, suggest that a regime shift occurred at ~1,300 years before present. Using the methods outlined above, I calculated the distance travelled (s) and velocity (v; Fig. 5.6). The results of v and s (**??**) on the relative abundance data correspond with both the large shifts in species dynamics (see Fig 5.4, and also with the regime shift identified by T. L. Spanbauer et al. (2014). However, two primary results can be made from the metrics v and s that are not obvious nor identified numerically in the results of T. L. Spanbauer et al. (2014) ():



1. Two additional large shifts occurred at approximately 2,500, 4,800 and years before 1950
1. The periods before the first and after the second large shifts appear oscillatory (Fig. ??).



To determine whether removing the noise in the data, I interpolated the each time series using function `stats::approx` to 700 time points. Next, I calculated the distance travelled of the entire system, s. Finally, I obtained the derivative of s by using a regularized differentiation (using function `tvdiff::TVRegDiffR`; parameters were $iter = 2000$, scale = small, $ep = 1x10^{-}6$, and $\alpha = 100)$[1].. This method of regularized differentiation is an ideal approach to smoothing s because it assumes the data are non-smooth, unlike other popular smoothing techniques e.g., Generalized Additive Models. The smoothed velocity (5.6) provides a similar but smoother picture of the velocity of the system trajectory. Comparing the smoothed (5.6) to the non-smoothed velocity (**??**) yields similar inference regarding the location of the regime shifts at 2,200 and 1,300 years before present, but more clearly identifies the inter-regime dynamics (e.g., between 7,000 and 4,800 years before present).

[1] *We created the R-wrapper `tvdiff` as a Python wrapper for the tvdiff MatLab package (**???**)

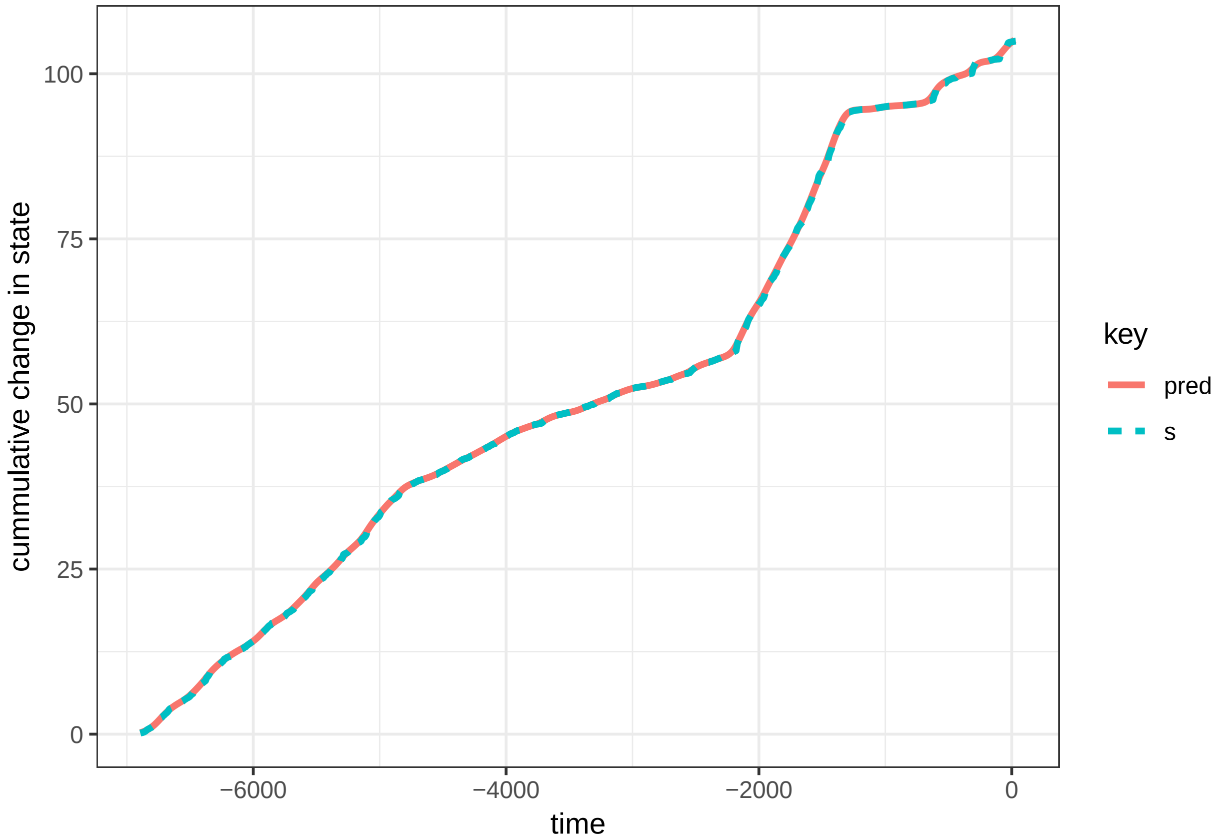


Figure 5.5: The regularized differentiation of s was best fit using $\alpha = 100$. Higher overlap of s and pred indicates a good fit of the regularized differentiated metric to the non-smoothed metric, s.

5.3 Discussion

In this chapter, I described the steps for calculating a novel regime detection metric, system velocity (v). First described in B. D. Fath et al. (2003), v is used as a single step for caclulating a more complicated regime detection metric, Fisher Information (see also Chapter 3). System velocity is arguably simple to calculate, as shown in this chapter, captures the total change in system variables under a variety of mean and variance conditions. The metric does not, however, perform well as variance increases (Fig. **??**), and smooothing the original data does not reduce the noise surrounding this metric when variance is moderate (Fig. **??**).

Variance is a commonly-used indicator of ecological regime shifts (W. Brock &

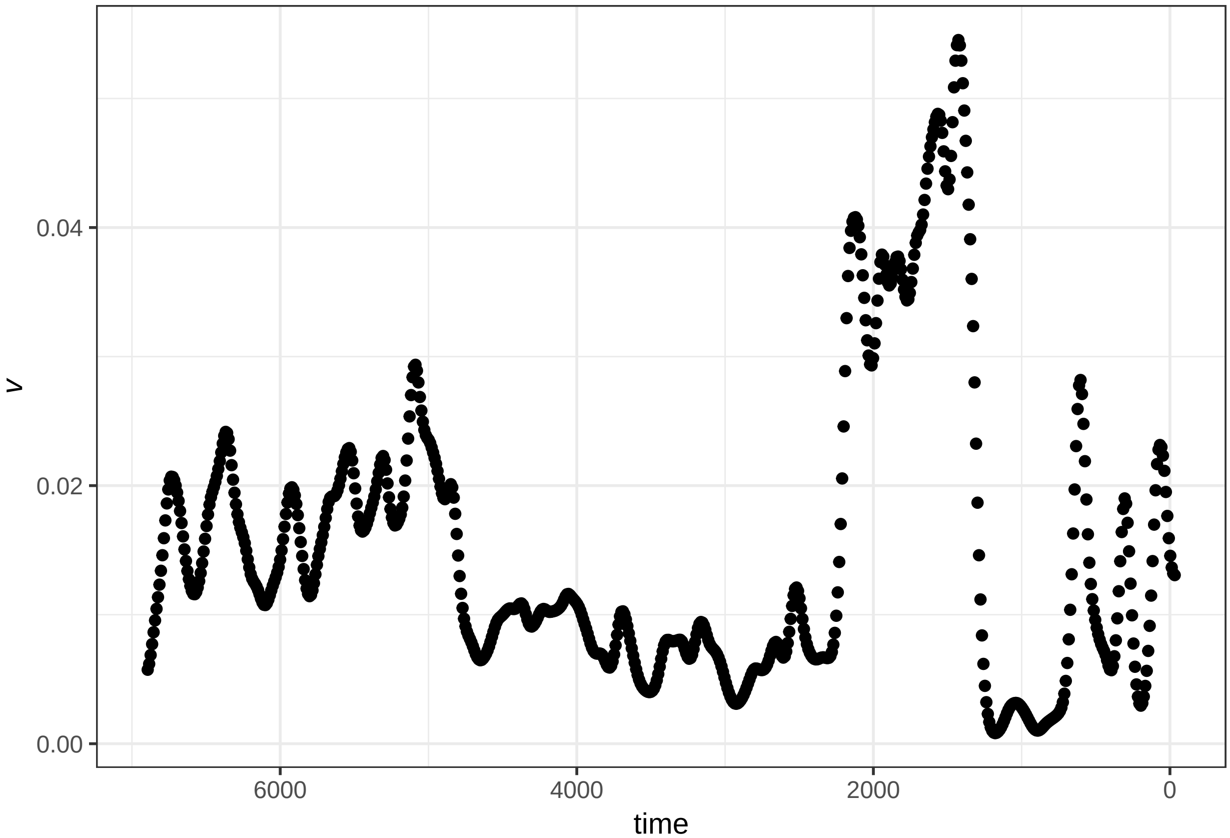


Figure 5.6: Need a cpation here!!!

Carpenter (2006)), however, fails to perform when the number of variables is >> a few. System velocity, v, may be useful in situations where the number of state variables is much greater than a few, and appears especially useful when the magnitude of change in one or more state variables is high (Fig. **??**). For example, this method will likely identify signals of regime shifts where the shift is defined as high species turnover within a community.

I tested the efficacy of this metric as an indicator of abrupt change in a two-variable system. Although a useful first step, this metric should be considered in a multi-species context, and particularly in community-level empirical data which is difficult to simulate. I demonstrate a compelling case study in materials associated with my R Package, **regimeDetectionMeasures**, and in Appendix **??** in which multiple species turnover events are apparent in a paleodiatom community time series. In this case study, the ‘distance travelled’, s (Eq. (5.4)), clearly exhibits shifts at points where expert opinion and species turnover (in species dominance) agree that a large change occurred. Further, velocity, v (see *dsdt* in the package materials) indicates a large shift at only the most predonimnant shift in the time series, perhaps due to the metric’s sensitivity to variance (Fig. **??**.

Further work is required to determine the utility of system velocity as a regime detection metric, however, this chapter demonstrates that the metric may indicate clear shifts in variable means. For multispecies data you will typically need to reduce dimensionality before you can proceed with analyses, for example using some sort of ordination. In addition to examining high-dimensional and noisy data, a study of the performance of v under conditions where few variables exhibit large changes while many variables are relatively constant may also prove useful. Additionally, this metric may be a useful tool for reducing the dimensionality of high dimensional data. Although the metric loses much information, as opposed to some dimension reduction techniques, e.g. Principal Components Analysis PCA, the metric is simple to calculate (even by hand), is computationally inexpensive, and is intuitive, unlike many clustering algorithms (e.g., Non-metric Multidimensional Scaling NMDS). Like system velocity, methods of the latter variety (e.g. NMDS) require post-hoc statistical analyses to confirm the location of clusters (or abrupt change, regime shifts), while methods of the former variety (e.g. PCA) retain loadings but do not necessarily identify the locations of abrupt shifts.

5.4 Supplementary Materials

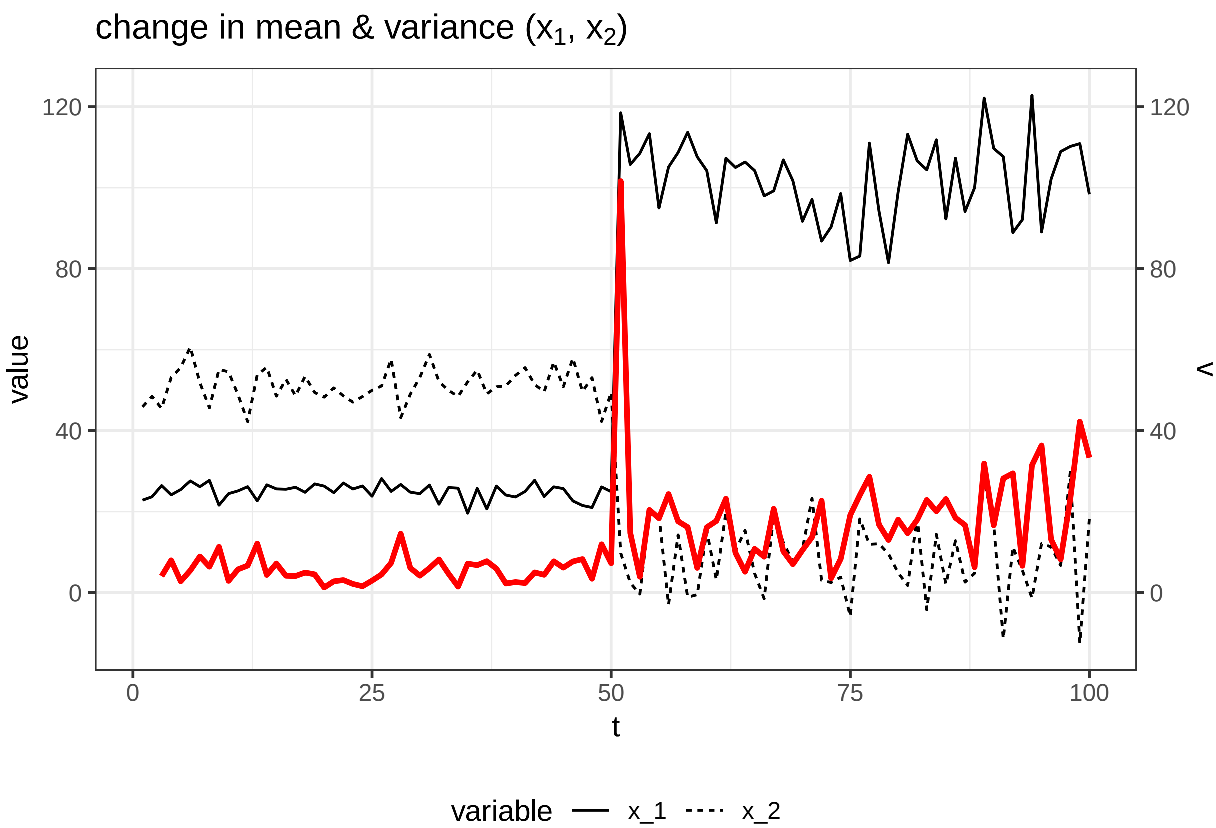


Figure 5.7: System change (s) and velocity (v) of the model system over the time period. Change in means ($\bar{x}_{1_{pre}} = 25$, $\bar{x}_{1_{post}} = 100$, $\bar{x}_{2_{pre}} = 50$, $\bar{x}_{2_{post}} = 10$) and an increase in variance ($\sigma_{1_{pre}} = 2$, $\sigma_{1_{post}} = 10$, $\sigma_{2_{pre}} = 5$, $\sigma_{2_{post}} = 10$).

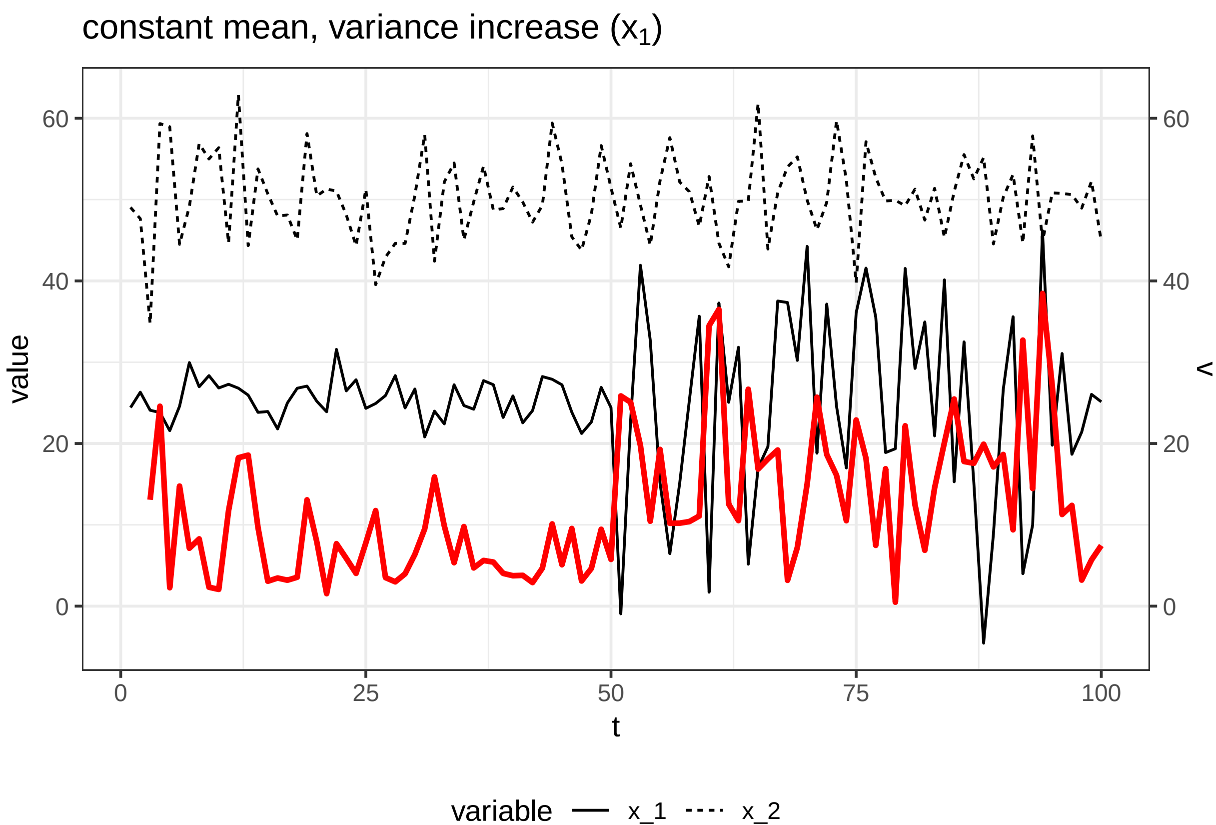


Figure 5.8: System change (s) and velocity (v) of the model system over the time period. Constant means ($\bar{x}_1 = 25$, $\bar{x}_2 = 50$) and sharp change in variance for one state variable $\sigma_{1_{pre}} = 2$, $\sigma_{1_{post}} = 12$, $\sigma_{2_{pre,post}} = 5$

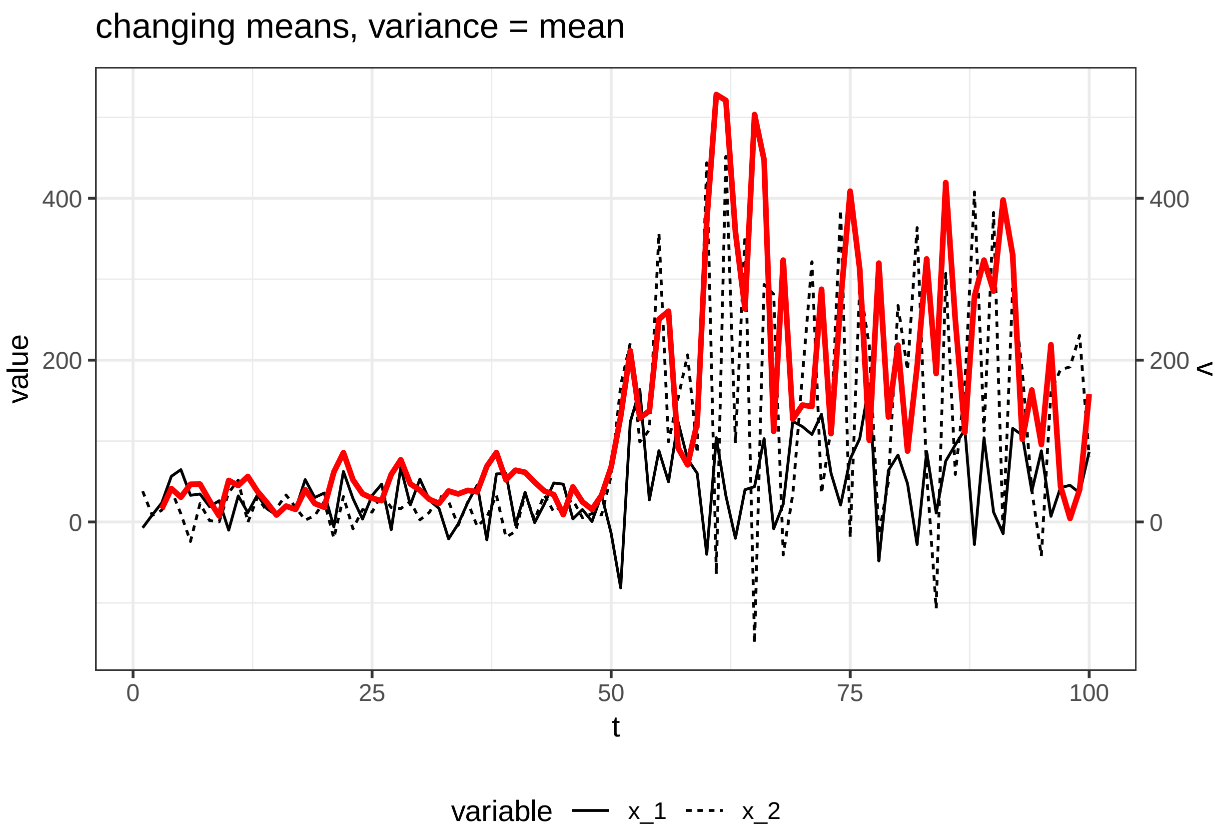


Figure 5.9: System change (s) and velocity (v) of the model system over the time period. Variance equal to mean ($/barx_i = /sigma_i$), where means ($/barx_{1_{pre}} = 25$, $/barx_{1_{post}} = 50$, $/barx_{2_{pre}} = 15$, $/barx_{2_{post}} = 150$).

Chapter 6

Robustness of Multivariate Regime Detection Measures to Varying Data Quality and Quantity

6.1 Introduction

Ecological systems have many unpredictable and variably interacting components (Jørgensen et al. 2011). Methods for analyzing these complex systems, e.g. Dynamic Bayesian Networks, network models, and food webs are designed to handle these complexities, yet require data- and knowledge-intensive models. Although ecological data collection and data management techniques are improving (La Sorte et al. 2018), the aforementioned approaches to modeling and understanding complex system are often infeasible in ecosystem research and management (Clements et al. 2015).

A growing concern with anthropogenic impacts on the environment has increased the demand for mathematical and statistical techniques that capture these dynamics. These often undesirable changes in the structure or functioning of ecological systems are often referred to as *regime shifts*, *regime changes*, *state change*, *abrupt change*, etc.

(Andersen et al. 2009) . A yet-unattained goal of ecological research and management is to reach a point where these methods can predict impending regime shifts in real-time and with high confidence. Ideally, ecological regime shift detection methods (hereafter, regime detection measures) would require little knowledge of the intrinsic drivers of the system, and the users of the method would not be required to know if and where a regime shift occurred in the data.

Despite the suite of regime detection measures in the environmental and ecological research literatures, they are not used in ecological management. We can describe the current state of regime detection measures as being either system specific (i.e., the method is not widely applicable or generalizable across systems) or not. Methods of the latter type are convenient in that they can be applied across various system and data types, but the results of these analyses require some degree of subjective interpretation (Clements and Ozgul 2018; c.f. Batt et al. 2013). Efforts to develop and/or improve regime detection measures that can handle these biases will aid the advance of regime detection measures research and application.

Current efforts to improve regime detection measures may be stunted by the lack of application beyond simple and/or theoretical (toy) systems data. Like most statistical and mathematical approaches, the evolution of many regime detection measures begins with application to theoretical data, followed by application to empirical data. Current applications of regime detection measures to empirical, ecological data are largely limited to data describing populations (e.g., Anderson and Piatt 1999, Alheit et al. 2005, deYoung et al. 2008), climatic, marine (e.g., Lipizer et al. n.d., Nicholls 2011), and Paleolithic regime shifts (Spanbauer et al. 2014, Yang et al. 2017, Kong et al. 2017), with few applications to terrestrial data (*c.f.* Bahlai et al. 2015; Sundstrom et al., 2017). Although testing the performance and inference boundaries of theoretical and simple systems is important, they are of little use to ecosystem managers if they are not proven to be easily and reliably applicable to their system. Additionally, regime detection measures should be capable of handling empirical ecological data are often sparse and noisy.

Ecological systems data is not only expensive to capture, but are often difficult to perfectly capture due to the large process and observation errors. The variability resulting from imperfect observation influences data quality and quantity, sometimes limiting the potential numerical tools used to identify trends and changes in the system in question (Thrush et al. 2009). Some methods, new and old, are proposed in the literature as regime detection measures which are capable of handling data limitation and quality issues inherent in ecological data and require few subjective decisions for choosing state variables and interpreting results. For example, variable reduction techniques, e.g. principal components analysis (Rodionov 2005, Andersen et al. 2009, Reid et al. 2016) and clustering algorithms (Weijerman et al. 2005, Weissmann and Shnerb 2016), an index of variance (Brock and Carpenter 2006) and Fisher Information (Cabezas and Fath 2002, Fath and Cabezas 2004, Karunanithi et al. 2008) were introduced as methods which collapse the system into a single indicator of ecological regime shifts. Although these methods have been tested on empirical ecological systems data, their robustness to empirical data quality and quantity have yet to be examined.

In this Chapter I examine the influence of observation and process errors on the inference obtained from select multivariable regime detection measures. There are two major objectives:

1. Identify the effects of data quality on regime detection measure inference.
2. Identify the effects of data quantity on regime detection measure inference.
3. Explore the relative performance of velocity (described in Chapter 5) to the abovementioned methods under multiple scenarios.

This Chapter provides baseline relative performance estimates of select, multivariable regime detection measures under various scenarios of data quality and quantity. The results from this Chapter inform the practical ecologist of the potential limitations to consider when applying these regime detection measures to their data, and has potential to inform the data collection process. Additionally, the software accompanying this Chapter allows the end user to implement these methods on this diatom system, a toy system, or their own data.

6.2 Data and Methodology

6.2.1 Study system and data

I used paleodiatom time series from a freshwater system in North America (Foy Lake, present day Montana) that apparently underwent a rapid shift in algal community dynamics at multiple periods in time. This datum comprises a single soil core sample, from which the relative abundances of 109 diatom species were identified at 768 observations (time points) over $\approx 7,000$ years (Figure 6.1. Althouh the soil core was sampled at regular distances, the soil accumulation process is not necessarily linear over time, resulting in irregularly-sampled observations (i.e., time elapsed between sampling points differs varies; see Figure 6.2). This datum was published in T. L. Spanbauer et al. (2014) and can be downloaded at the publisher's website.

6.2.2 Regime detection measures

Fewer model-free regime detection metrics exist than do model-based metrics (Chapter 2) and of these, only a few are suggested for handling multivariable data. Here, I examine the regime detection metrics that are model-free and can handle multivariable data: velocity (Chapter 5), the Variance Index (W. Brock & Carpenter, 2006) and Fisher Information. These methods and the primary sources are described below.

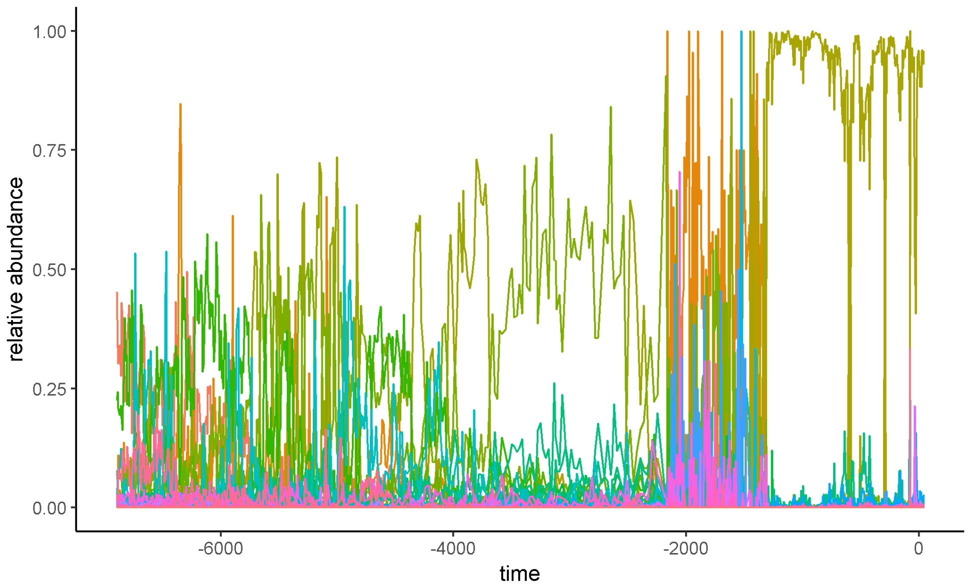


Figure 6.1: Relative abundances of the diatom species in Foy Lake over the time period.

Velocity (v)

In Chapter 5, I describe a new method, **velocity**, v, as a potential dimension reduction and regime detection method. First introduced in by B. D. Fath et al. (2003) as one of multiple steps in calculating their variant of Fisher Information, velocity calculates the cumulative sum of the square root of the sum of the squared change in all state variables over a period of time (Eq. (6.1)). Steps for calculating this metric are described in detail in Chapters 3 and 5.

$$\Delta s_i = \sqrt{\sum_{j=1}^{n}(x_{i,j} - x_{i-1,j})^2} s_k = \sum_{i=2}^{k} \Delta s_i 2 \leq k \leq n v = \frac{\Delta s}{\Delta t} \tag{6.1}$$

Variance Index

The Variance Index was introduced by W. Brock & Carpenter (2006), and is simply defined asthe maximum eigenvalue of the covariance matrix of the system over some period (window) of time. The Variance Index (also called Variance Indicator) was originally applied to a modelled system (W. Brock & Carpenter, 2006), and has since

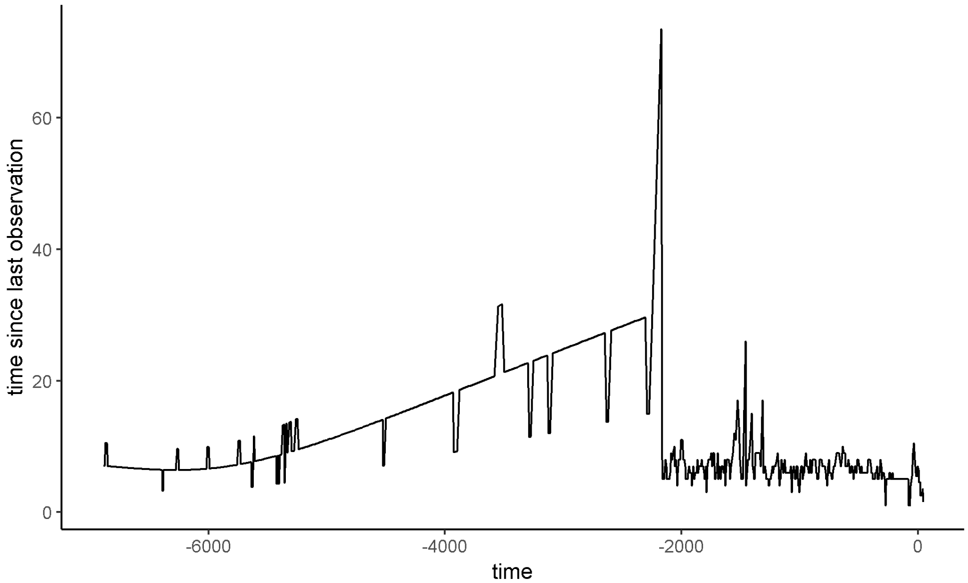


Figure 6.2: The amount of time elapsed between observations.

been applied to empirical data (T. L. Spanbauer et al., 2014; Sundstrom et al., 2017). Although rising variance has been useful in many real systems (van Nes and Scheffer 2003, Brock et al. 2006, Carpenter and Brock 2006), the Variance Index, which is intended for multivariate data, appears most useful when the system exhibits a discontinuous regime shift (W. Brock & Carpenter, 2006).

Fisher Information

Fisher Information (I) is essentially calculted as the area under the curve of the acceleration to the fourth degree (s''^4) divided by the squared velocity (s'^2; also referred to as v in Chapter 5) of the distance travelled by the system, s over some period of time (T), and is given in Eq. (6.2):

$$I = \frac{1}{T}\int_0^T dt \left[\frac{s''^2}{s'^4}\right]^2 \tag{6.2}$$

I describe this method in detail in Chapter 3.

Using moving window analysis to calculate Fisher Information and Variance Index

Unlike *velocity*, the Variance Index and Fisher Information are calculated using moving window analysis. That is, over the entire time series, T^*, these metrics are calculated within multiple windows of time, T. In this approach, all state variables, x_i, are used to inform the calculations (of Variance Index and Fisher Information) over a time interval, T, where T is the length in [time] units of the time interval and satisfies the following conditions: $T < T^*$ and $2 \leq T < (T^* - 1)$. If $T = T^* - 1$, then only a single value of the metrics will be calculated for entire time series, which does not allow for any estimate of change.

When using these metrics in the context of identifying abrupt changes in ecological systems data across $T*$, it is ideal the value of T meets the following conditions: $3 < T \ll T^* - 1$. The length of a time window dictates the number of calculations one can obtain over T^*, such that the number of potential metric calulations increases as $\frac{T}{T^*}$ decreases. Previous applications of moving window analyses to calculate Fisher Information found that at least eight observations (time points) should be used.

An additional parameter is required when conducting moving window analyses: the amount of time points by which the window advances. In order to maximize the data, I force the window to advance at a rate of one time unit. However, it is important to note that because these data are not sampled annually and the because the window always advances by a single time unit, the number of observations included in each calculation will not be the same. If fewer than 5 observations are in a window, I did not calculate metrics, advancing the window forward. I assigned the calcuated values of Fisher Information and Variance Index within each moving window to the **end** (the last time unit) of the moving window. I temporal analyses, assigning the value to any other point in time (e.g., the beginning or the middle) muddles the interpretation of the metric over T^*. Also note that this method has the potential to result in calculating a metric for all integers between $0.20T^*$ and T^*.

6.2.3 Resampling Techniques for Simulating Data Quality and Quantity Issues

Using a bootstrap approach I calculated the regime detection measures over varying degrees of scenarios to simulate data quality and data quantity issues that are common to ecological data anlaysis. The scnearios are categorized as *observations* and *species*. The observations scenario simulates a loss of temporal observations (decreasing the number of times the system was observed), and the species scenario simulates a loss of information about the system by removing a larger proportion of the species. The loss of temporal observations and the loss of species were examined at three proportions: $\mathbf{P} = [0.25, 0.50, 0.75, 1.00]$, where $\mathbf{P}$ is the proportion of species and time points **retained** for analysis. For example, when $\mathbf{P} = 0.25$, a random selection of 25% of the species are retained for analysis in the species scenario. I bootstrapped the datum over 10,000 iterations for each scenario and $\mathbf{P}$ combination. Note that because when $\mathbf{P} = 1.00$, all data are retained. Therefore, no resampling was conducted at this level because only a single metric (e.g. Velocity) value is possible.

Interpretation of the regime detection measures used in this analysis are currently limited to visual inspection. Therefore, I limit inference in this study largely to the impact of data loss on the variability with a regime detection measure (i.e. how robust is the measure to data loss).

6.3 Results

In many cases the standard deviation of FI far exceeded the mean value of FI. I calculated the coefficient of variation, $\frac{\sigma}{\mu}$, for each ????!!!!??? as the proportion of data (and observations) used decreased. For example, when when we retained

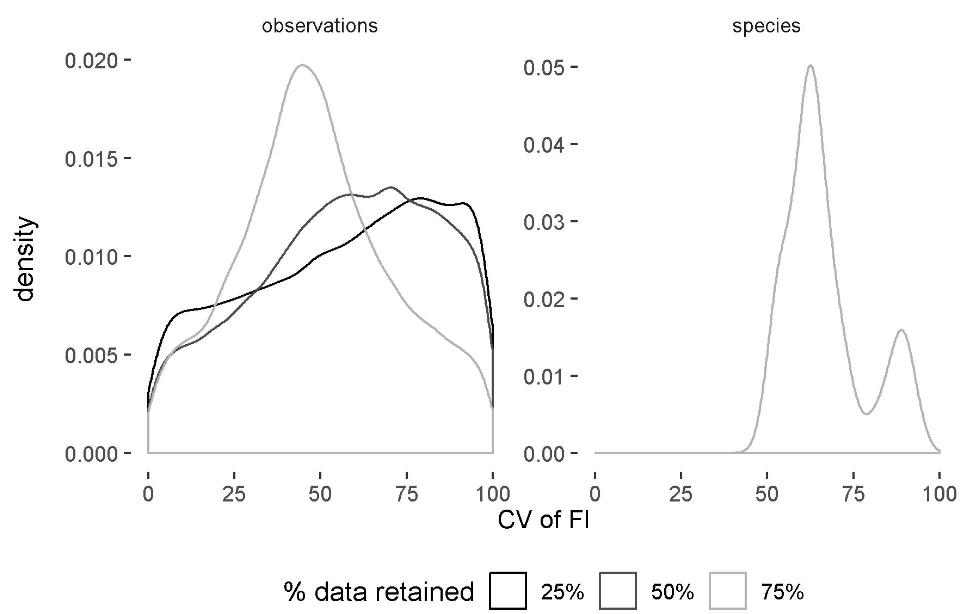
6.3.1 Velocity of the distance travelled produces similar results with information loss

Ad lorem ipsum blahblahlhba

6.3.2 Variance Index produces

6.3.3 Fisher Information is highly sensitive to information loss

When we bootstrap 25% of the species, the ratio of mean Fisher Information to standard deviation of Fisher Information (over $10,000$ iterations) is always < 1, suggesting Fisher Information does not produce fidel results when information is lost about the system.



\begin{figure}

\caption{Density plot of the coefficient of variation (CV) as a percentage (%) of the Fisher Information bootstrapped samples (10,000 iterations). Densiies based on all values of CV, but values >100% are not printed.} \end{figure}

6.4 Discussion

6.5 Ackowledgements

This study was conceptualized at the International Institute for Applied Systems Analysis (IIASA) as part of the Young Scholars Summer Program in 2018. I thank my IIASA program supervisors, Drs. Brian Fath and Elena Rovenskaya, for advisement during this period.

Chapter 7

Discontinuity chapter under construction

7.1 Introduction

7.2 Data and Methods

7.3 Results

7.4 Conclusions

Chapter 8

Conclusions

$$
\begin{aligned}
Data &= Information \\
&= Signal \\
&= Process + Noise
\end{aligned}
\tag{8.1}
$$

Climate change is expected to induce an increase in both the intensity and frequency of rapid ecological change or disturbance, impacting social systems, potentially to the detriment of human communities most vulnerable. Identifying and forecasting these changes is critical for community and ecological planning, management, and disaster mitigation. Because ecological and social systems are tightly coupled, it is commonplace to use ecological indicators to identify change and potential changes that may impact these systems. Many papers introducing or discussing regime detection measures suggest the ecologist uses multiple lines of evidence, ranging from historical observations to ecological modelling results, for identifying an ecological regime shift (Lindegren et al., 2012). Although valid, comparing results of multiple methods or lines of evidence within a single system has yielded inconsistent results, and inconsistent results can result in either improper conclusions, or in what I am calling **method mining**. That is, a dataset is analyzed using until a sufficient number of methods yield affirmative results.

8.1 Method mining regime detection methods

Many regime detection measures have yet to be properly and statistically (or numerically) scrutinized. However, it should be noted that, in part due to both (i) the popularity and (ii) the sheer number of 'new' methods a handful of authors[1].

Managing systems using quantitative methods that yield different results may yield improper management techniques and objectives. –>

Ecological indicators (a.k.a. indices, metrics) have been suggested as 'early-warning indicators' of ecological regime shifts or abrupt change (Chapters 1 and 2). Ecological indicators (or indices) are methods of measurement which are designed to provide inference about one or more unobserved or latent processes, are inherently biased. Regardless of the state of the theory supporting *regime shifts* in ecology, ecological indicators and the methods for calculating them should be heavily scrutinized prior to being used in an ecological management or policy-making setting. Rather, new methods (indices, metrics, etc.) are being introduced into the literature at a rate exceeding that at which they are scrutinized (Chapter 2). This dissertation demonstrates that, while potentially useful, regime detection metrics are inconsistent, not generalizable, and are currently not validated using probabilities or other statistical measurements of certainty.

8.2 Ecological data are noisy

Regime detection metrics appear more reliable when the signal-to-noise ratio is high (Ch. 2, Ch. 5, **???**). Ecological systems are noisy, and the observational data we are collecting at large scales (e.g., the North American Breeding Bird survey), is noisy. Using methods incapable of identifying meaningful signals in noisy data appears futile,

[1] S.R. Carpenter is one example of an author who has relative infamy in the field and has, as primary author or otherwise, introduced a relatively large number of new methods (e.g., rising variance, the variance index, Fourier transform, online dynamic linear modelling, TVARSS, to name a few)

yet, methods for doing so are increasingly introduced in the scientific literature (Ch. 2).

8.3 Data collection and munging biases and limits findings

Regime detection measures and other ecological indicators can signal (see (8.1)) various changes in the data, however, understanding what processes are embedded in the signals (i.e., removing the noise) requires expert judgement. And because a consequence of data collection and data analysis limits the extent to which we can identify and infer processes and change within an ecological system, **I suggest the practical ecologist scrutinizes her data prior to identifying and conducting analyses**, including those that are purely exploratory. By collecting and analysing data, the ecologist has defined the bounaries of the system *a priori*ˆ+(+ Beisner, Haydon, & Cuddington, 2003 states this eloquently as, "The number and choice of variables selected to characterize the community will be determined by what we wish to learn from the model"). The influence of state variable selection is ignored by some metrics (e.g. Fisher Information Eason, Garmestani, & Cabezas, 2014 and *v* Chapter 5), in that the resulting measure is composite and carries no information regarding the influence of state variables on the metric result.

The actual limitations to the system should be, theoretically, known as a result of bounding the system. Inference beyond this system is extrapolation, and should be treated as speculation, especially when not accompanied by a measure of uncertainty around one's predictions.

8.4 Common Limitations of Regime Detection Measures

Limitations of the findings in this dissertation and of the regime detection methods used herein are largely influenced by the **data collection**, **data munging** processes. Although the below mentioned points may seem logical to many, these assumptions are overlooked by many composite indicators, including regime detection measures.

1. Signals in the indicators are restricted to the ecological processes captured by the input data. Extrapolation occurs when processes manifest at scales different than the data collected. (resolution; Chapter **??**)
1. normalization and weighting techniques often impact results (whether ecological or numerical) (Appendices **??** and **??**)
1. data aggregation techniques often impact results (Chapter 6)
1. some indices fail to generalize across systems or taxa (see Chapters 1 and **??**)

8.5 Specific synthesis of chapter results

Abadi, F., Gimenez, O., Arlettaz, R., & Schaub, M. (2010). An assessment of integrated population models: Bias, accuracy, and violation of the assumption of independence. *Ecology*, *91*(1), 7–14.

Andersen, T., Carstensen, J., Hern??ndez-Garc??a, E., & Duarte, C. M. (2009). Ecological thresholds and regime shifts: Approaches to identification. *Trends in Ecology & Evolution*, *24*(1), 49–57. http://doi.org/10.1016/j.tree.2008.07.014

Beaugrand, G. (2004). The north sea regime shift: Evidence, causes, mechanisms and consequences. *Progress in Oceanography*, *60*(2-4), 245–262.

Beisner, B. E., Haydon, D. T., & Cuddington, K. (2003). Alternative stable states in ecology. *Frontiers in Ecology and the Environment*, *1*(7), 376–382.

Bestelmeyer, B. T., Ellison, A. M., Fraser, W. R., Gorman, K. B., Holbrook, S. J., Laney, C. M., … others. (2011). Analysis of abrupt transitions in ecological systems. *Ecosphere*, *2*(12), 1–26.

Boettiger, C., & Hastings, A. (2012). Quantifying limits to detection of early warning for critical transitions. *Journal of the Royal Society Interface*, *9*, 2527–39.

Boettiger, C., Ross, N., & Hastings, A. (2013). Early warning signals: The charted and uncharted territories. *Theoretical Ecology*, *6*(3), 255–264.

Brock, W., & Carpenter, S. (2006). Variance as a Leading Indicator of Regime Shift in Ecosystem Services. *Ecology and Society*, *11*(2). `http://doi.org/10.5751/ES-01777-110209`

Burthe, S. J., Henrys, P. A., Mackay, E. B., Spears, B. M., Campbell, R., Carvalho, L., … others. (2016). Do early warning indicators consistently predict nonlinear change in long-term ecological data? *Journal of Applied Ecology*, *53*(3), 666–676.

Butitta, V. L., Carpenter, S. R., Loken, L. C., Pace, M. L., & Stanley, E. H. (2017). Spatial early warning signals in a lake manipulation. *Ecosphere*, *8*(10), n/a–n/a. `http://doi.org/10.1002/ecs2.1941`

Byrski, J., & Byrski, W. (2016). A double window state observer for detection and isolation of abrupt changes in parameters. *International Journal of Applied Mathematics and Computer Science*, *26*(3), 585–602.

Cabezas, H., & Fath, B. D. (2002). Towards a theory of sustainable systems. *Fluid Phase Equilibria*, *194*(Supplement C), 3–14. `http://doi.org/10.1016/S0378-3812(01)00677-X`

Cabezas, H., Pawlowski, C. W., Mayer, A. L., & Hoagland, N. T. (2005). Simulated experiments with complex sustainable systems: Ecology and technology. *Resources, Conservation and Recycling*, *44*(3), 279–291. Retrieved from `http:`

//www.sciencedirect.com/science/article/pii/S092134490500025X

Clements, C. F., & Ozgul, A. (2018). Indicators of transitions in biological systems. *Ecology Letters*, *21*(6), 905–919.

Dakos, V., Carpenter, S. R., Brock, W. A., Ellison, A. M., Guttal, V., Ives, A. R., … Nes, E. H. van. (2012). Methods for detecting early warnings of critical transitions in time series illustrated using simulated ecological data. *PloS One*, *7*(7), e41010.

Dakos, V., Carpenter, S. R., Nes, E. H. van, & Scheffer, M. (2015a). Resilience indicators: Prospects and limitations for early warnings of regime shifts. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *370*(1659), 20130263. http://doi.org/10.1098/rstb.2013.0263

Dakos, V., Carpenter, S. R., Nes, E. H. van, & Scheffer, M. (2015b). Resilience indicators: Prospects and limitations for early warnings of regime shifts. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *370*(1659), 20130263.

Davis, E. P., & Karim, D. (2008). Comparing early warning systems for banking crises. *Journal of Financial Stability*, *4*(2), 89–120.

DeAngelis, D. L., & Yurek, S. (2017). Spatially explicit modeling in ecology: A review. *Ecosystems*, *20*(2), 284–300.

deYoung, B., Barange, M., Beaugrand, G., Harris, R., Perry, R. I., Scheffer, M., & Werner, F. (2008). Regime shifts in marine ecosystems: Detection, prediction and management. *Trends in Ecology & Evolution*, *23*(7), 402–409. http://doi.org/10.1016/j.tree.2008.03.008

Ducré-Robitaille, J.-F., Vincent, L. A., & Boulet, G. (2003). Comparison of techniques for detection of discontinuities in temperature series. *International Journal of Climatology: A Journal of the Royal Meteorological Society*, *23*(9), 1087–1101.

Dutta, P. S., Sharma, Y., & Abbott, K. C. (2018). Robustness of early warning signals for catastrophic and non-catastrophic transitions. *Oikos*, *127*(9), 1251–1263.

Eason, T., & Cabezas, H. (2012). Evaluating the sustainability of a regional system using Fisher information in the San Luis Basin, Colorado. *Journal of Environmental Management*, *94*(1), 41–49.

Eason, T., Garmestani, A. S., & Cabezas, H. (2014). Managing for resilience: Early detection of regime shifts in complex systems. *Clean Technologies and Environmental Policy*, *16*(4), 773–783.

Fath, B. D., & Cabezas, H. (2004). Exergy and Fisher Information as ecological indices. *Ecological Modelling*, *174*(1), 25–35. http://doi.org/10.1016/j.ecolmodel.2003.12.045

Fath, B. D., Cabezas, H., & Pawlowski, C. W. (2003). Regime changes in ecological systems: An information theory approach. *Journal of Theoretical Biology*, *222*(4), 517–530. http://doi.org/10.1016/S0022-5193(03)00067-5

Filatova, T., Polhill, J. G., & Ewijk, S. van. (2016). Regime shifts in coupled socio-environmental systems: Review of modelling challenges and approaches. *Environmental Modelling & Software*, *75*, 333–347.

Fisher, R. A. (1922). On the Mathematical Foundations of Theoretical Statistics. *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*, *222*, 309–368. Retrieved from http://www.jstor.org/stable/91208

Frieden, B. R. (1998). *Physics from Fisher information: A unification.* New York, NY: Cambridge University Press.

Frieden, B. R., & Gatenby, R. (Eds.). (2007). *Exploratory Data Analysis Using Fisher Information*. Springer. Retrieved from http://www.springer.com/us/book/9781846285066

Hastings, A., & Wysham, D. B. (2010). Regime shifts in ecological systems can occur with no warning. *Ecology Letters*, *13*(4), 464–472. http://doi.org/10.1111/j.1461-0248.2010.01439.x

Hefley, T. J., Tyre, A. J., & Blankenship, E. E. (2013). Statistical indicators and state–space population models predict extinction in a population of bobwhite quail. *Theoretical Ecology*, *6*(3), 319–331.

Jorgensen, S. E., & Svirezhev, Y. M. (2004). *Towards a Thermodynamic Theory for Ecological Systems*. Elsevier.

Karunanithi, A. T., Cabezas, H., Frieden, B. R., & Pawlowski, C. W. (2008). Detection and assessment of ecosystem regime shifts from fisher information. *Ecology and Society*, *13*(1).

Kefi, S., Guttal, V., Brock, W. A., Carpenter, S. R., Ellison, A. M., Livina, V. N., … Dakos, V. (2014). Early warning signals of ecological transitions: Methods for spatial patterns. *PloS One*, *9*(3), e92097.

Kong, X., He, Q., Yang, B., He, W., Xu, F., Janssen, A. B., … others. (2017). Hydrological regulation drives regime shifts: Evidence from paleolimnology and ecosystem modeling of a large shallow chinese lake. *Global Change Biology*, *23*(2), 737–754.

Lindegren, M., Dakos, V., Gr??ger, J. P., G??rdmark, A., Kornilovs, G., Otto, S. A., & M??llmann, C. (2012). Early Detection of Ecosystem Regime Shifts: A Multiple Method Evaluation for Management Application. *PLoS ONE*, *7*(7), e38410. http://doi.org/10.1371/journal.pone.0038410

Litzow, M. A., & Hunsicker, M. E. (2016). Early warning signals, nonlinearity, and signs of hysteresis in real ecosystems. *Ecosphere*, *7*(12), n/a–n/a. http://doi.org/10.1002/ecs2.1614

Liu, J., Dietz, T., Carpenter, S. R., Alberti, M., Folke, C., Moran, E., … others. (2007). Complexity of coupled human and natural systems. *Science*, *317*(5844), 1513–1516.

Liu, S., Yamada, M., Collier, N., & Sugiyama, M. (2013). Change-point detection in time-series data by relative density-ratio estimation. *Neural Networks*, *43*, 72–83.

Mac Nally, R., Albano, C., & Fleishman, E. (2014). A scrutiny of the evidence for pressure-induced state shifts in estuarine and nearshore ecosystems. *Austral Ecology*, *39*(8), 898–906.

Mantua, N. (2004). Methods for detecting regime shifts in large marine ecosystems: A review with approaches applied to North Pacific data. *Progress in Oceanography*, *60*(2), 165–182. http://doi.org/10.1016/j.pocean.2004.02.016

Mayer, A. L., Pawlowski, C. W., & Cabezas, H. (2006). Fisher Information and dynamic regime changes in ecological systems. *Ecological Modelling*, *195*(1???2), 72–82. http://doi.org/10.1016/j.ecolmodel.2005.11.011

Mayer, D. A. L., Pawlowski, D. C., Fath, P. B. D., & Cabezas, D. H. (2007). Applications of Fisher Information to the Management of Sustainable Environmental Systems. In B. R. F. B. MS & R. A. G. B. MD (Eds.), *Exploratory Data Analysis Using Fisher Information* (pp. 217–244). Springer London. http://doi.org/10.1007/978-1-84628-777-0_7

Nicholls, K. H. (2011). Detection of regime shifts in multi-species communities: The Bay of Quinte phytoplankton example. *Methods in Ecology and Evolution*, *2*(4), 416–426. http://doi.org/10.1111/j.2041-210X.2011.00093.x

Nicholls, K., Hoyle, J., Johannsson, O., & Dermott, R. (2011). A biological regime shift in the bay of quinte ecosystem (lake ontario) associated with the establishment of invasive dreissenid mussels. *Journal of Great Lakes Research*, *37*(2), 310–317.

Perretti, C. T., & Munch, S. B. (2012). Regime shift indicators fail under noise levels commonly observed in ecological systems. *Ecological Applications*, *22*(6), 1772–1779.

Perretti, C. T., Munch, S. B., & Sugihara, G. (2013). Model-free forecasting outperforms the correct mechanistic model for simulated and experimental data. *Proceedings of the National Academy of Sciences*, *110*(13), 5253–5257.

Petersen, J. K., Hansen, J. W., Laursen, M. B., Clausen, P., Carstensen, J., & Conley, D. J. (2008). Regime shift in a coastal marine ecosystem. *Ecological*

Applications, *18*(2), 497–510.

Roberts, C. P., Twidwell, D., Burnett, J. L., Donovan, V. M., Wonkka, C. L., Bielski, C. L., … others. (2018). Early warnings for state transitions. *Rangeland Ecology & Management*, *71*(6), 659–670.

Rodionov, S. N. (2005). A brief overview of the regime shift detection methods. *Large-Scale Disturbances (Regime Shifts) and Recovery in Aquatic Ecosystems: Challenges for Management Toward Sustainability*, 17–24. Retrieved from `http://www.beringclimate.noaa.gov/regimes/rodionov_overview.pdf`

Rodionov, S., & Overland, J. E. (2005). Application of a sequential regime shift detection method to the Bering Sea ecosystem. *ICES Journal of Marine Science*, *62*(3), 328–332. `http://doi.org/10.1016/j.icesjms.2005.01.013`

Roy Frieden, B. (1998). *Physics from Fisher information.* Cambridge: Cambridge University Press.

Salehpour, S., Gustafsson, T., & Johansson, A. (2011). An on-line method for estimation of piecewise constant parameters in linear regression models. *IFAC Proceedings Volumes*, *44*(1), 3171–3176.

Scheffer, M. (2009). *Critical transitions in nature and society.* Princeton University Press.

Scheffer, M., Carpenter, S. R., Dakos, V., & Nes, E. H. van. (2015). Generic indicators of ecological resilience: Inferring the chance of a critical transition. *Annual Review of Ecology, Evolution, and Systematics*, *46*, 145–167.

Scheffer, M., Carpenter, S., Foley, J. A., Folke, C., & Walker, B. (2001). Catastrophic shifts in ecosystems. *Nature*, *413*(6856), 591–596.

Seddon, A. W., Froyd, C. A., Witkowski, A., & Willis, K. J. (2014). A quantitative framework for analysis of regime shifts in a galápagos coastal lagoon. *Ecology*, *95*(11), 3046–3055.

Sommer, S., Benthem, K. J. van, Fontaneto, D., & Ozgul, A. (2017). Are generic early-warning signals reliable indicators of population collapse in rotifers? *Hydrobiologia*, *796*(1), 111–120.

Spanbauer, T. L., Allen, C. R., Angeler, D. G., Eason, T., Fritz, S. C., Garmestani, A. S., … Stone, J. R. (2014). Prolonged instability prior to a regime shift. *PLoS One*, *9*(10), e108936.

Steel, E. A., Kennedy, M. C., Cunningham, P. G., & Stanovick, J. S. (2013). Applied statistics in ecology: Common pitfalls and simple solutions. *Ecosphere*, *4*(9), 1–13.

Sundstrom, S. M., Eason, T., Nelson, R. J., Angeler, D. G., Barichievy, C., Garmestani, A. S., … Allen, C. R. (2017). Detecting spatial regimes in ecosystems. *Ecology Letters*, *20*(1), 19–32. http://doi.org/10.1111/ele.12709

Vasilakopoulos, P., Raitsos, D. E., Tzanatos, E., & Maravelias, C. D. (2017). Resilience and regime shifts in a marine biodiversity hotspot. *Scientific Reports*, *7*(1), 13647.

Walther, G.-R., Post, E., Convey, P., Menzel, A., Parmesan, C., Beebee, T. J., … Bairlein, F. (2002). Ecological responses to recent climate change. *Nature*, *416*(6879), 389.

Yin, D., Leroux, S. J., & He, F. (2017). Methods and models for identifying thresholds of habitat loss. *Ecography*, *40*(1), 131–143.

Zhou, T., & Shumway, R. (2008). One-step approximations for detecting regime changes in the state space model with application to the influenza data. *Computational Statistics & Data Analysis*, *52*(5), 2277–2291.