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

Chapter 1

thesisdown::thesis_gitbook: default

Placeholder

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 1.1) 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 1.1: 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 1.1: 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 1.1: 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 1.1: 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 1.1: 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 1.1: 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 1.1: 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 2

Introduction

Placeholder

2.1 Forecasting abrupt changes in ecology

2.2 Dissertation aims

2.3 Dissertation structure

2.3.1 Chapter overview

2.3.2 Accompanying software (appendices)

Chapter 3

A brief overview of ecological regime detection methods methods

Placeholder

3.1 Introduction

3.2 Methods

3.2.1 Identifying candidate articles

Web of Science

Prior knowledge and snowball method

Google Scholar

Additional filtering

3.3 Results

3.3.1 Web of Science

3.3.2 Google Scholar and prior knowledge

3.3.3 List of new methods

3.4 Discussion

3.4.1 Barriers to identifying new regime detection measures

3.4.2 Reducing the barriers to regime detection measures

Chapter 4

A guide to Fisher Information for Ecologists

Placeholder

4.1 Abstract

4.2 Introduction

4.2.1 On Fisher Information

4.2.2 Notation

4.2.3 Steps for calculating Fisher Information (FI)

4.2.4 Concepts behind the calculations

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

Step 2. Distance traveled by the system, s

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

Step 4. Calculate the derivatives-based Fisher Information

4.3 Case Study

4.4 Conclusions

4.5 Acknowledgements

Chapter 5

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

Placeholder

5.1 Introduction

5.2 Data and methods

5.2.1 Data: North American breeding bird communities

5.2.2 Study area

Focal military base

Spatial sampling grid

5.2.3 Calculating Fisher Information (FI)

5.2.4 Interpreting and comparing Fisher Information across spatial transects

Interpreting Fisher Information values

Interpolating results across spatial transects

Spatial correlation of Fisher Information

5.3 Results

5.3.1 Fisher Information across spatial transects

5.3.2 Spatial correlation of Fisher Information

5.4 Discussion

5.4.1 Efficacy of Fisher Information as a spatial RDM

Chapter 6

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

Placeholder

6.1 Introduction

6.2 Data and Methods

6.2.1 Theoretical system example: two-species time series

6.2.2 Steps for calculating system velocity, *v*

Step 1: Δx_i

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

Step 3: Use Pythagorean theorem to isolate s

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

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

Varying post-shift mean

Varying post-shift variance

Smoothing the data prior to calculating *v*

6.2.4 Performance on empirical data: paleodiatom community example

6.3 Discussion

6.4 Supplementary Materials

Chapter 7

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

7.1 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 6) 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.

7.2 Data and Methodology

7.2.1 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 [7.1]. 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 7.2). This datum was published in T. L. Spanbauer et al. (2014) and can be downloaded at the publisher's website.

7.2.2 Regime detection measures

Fewer model-free regime detection metrics exist than do model-based metrics [Chapter 3] 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 6], the Variance Index (Brock & Carpenter, 2006) and Fisher Information. These methods and the primary sources are described below.

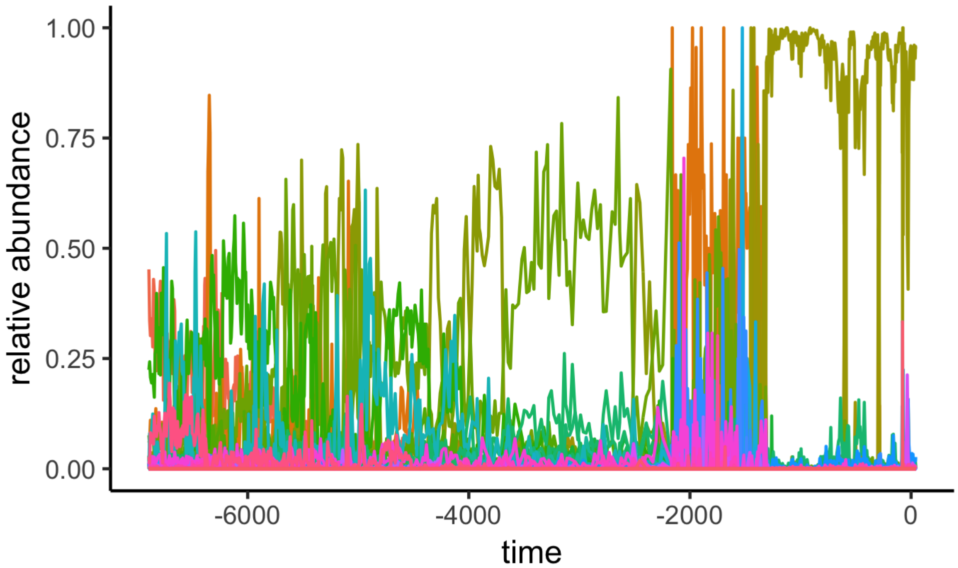

Figure 7.1: Relative abundances of the diatom species in Foy Lake over the time period.

Velocity (v)

In Chapter 6, I describe a new method, **velocity**, v, as a potential dimension reduction and regime detection method. First introduced in by Fath, Cabezas, & Pawlowski (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. (7.1)]. Steps for calculating this metric are described in detail in Chapters 4 and 6.

$$\Delta s_i = \sqrt{\sum_{j=1}^{n}(x_{i,j} - x_{i-1,j})^2} s_k = \sum_{i=2}^{k} \Delta s_i 2 \leq k \leq n v = \frac{\Delta s}{\Delta t} \tag{7.1}$$

Variance Index

The Variance Index was introduced by 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 (Brock & Carpenter, 2006), and has since

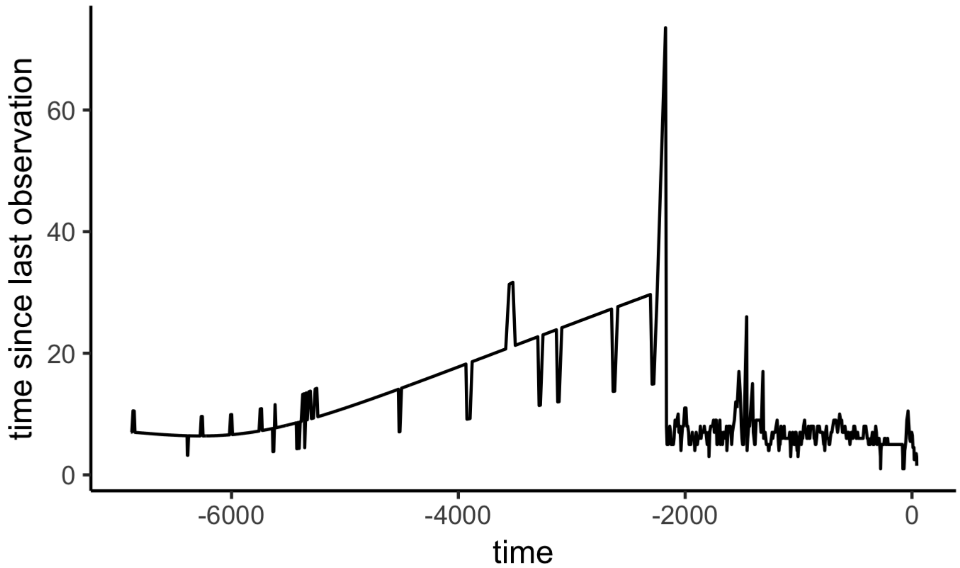

Figure 7.2: The amount of time elapsed between observations.

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 (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 6 and in the next section] of the distance travelled by the system, s over some period of time (T), and is given in Eq. (7.2):

$$I = \frac{1}{T}\int_0^T dt \left[\frac{s''^2}{s'^4}\right]^2 \tag{7.2}$$

I describe this method in detail in Chapter 4.

Calculating Fisher Information and Variance Index using moving window analysis

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^*.

7.2.3 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).

7.3 Results

7.4 Discussion

7.5 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 8

Discontinuity chapter under construction

8.1 Introduction

8.2 Data and Methods

8.3 Results

8.4 Conclusions

Chapter 9

Conclusions

Placeholder

9.1 Method mining regime detection methods

9.2 Ecological data are noisy

9.3 Data collection and munging biases and limits findings

9.4 Common Limitations of Regime Detection Measures

9.5 Specific synthesis of chapter results

Appendix A: R package regimeDetectionMeasures

Placeholder

9.6 Measures/metrics calculated

9.7 Example analysis

Appendix B: R package bbsRDM

Placeholder

References

Placeholder

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

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

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.

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