8-2019

Regime Detection Measures for the Practical Ecologist

Jessica L. Burnett
University of Nebraska - Lincoln, jburnett@huskers.unl.edu

Follow this and additional works at: https://digitalcommons.unl.edu/natresdiss

Part of the Hydrology Commons, Natural Resources and Conservation Commons, Natural Resources Management and Policy Commons, Other Environmental Sciences Commons, and the Water Resource Management Commons

Burnett, Jessica L., "Regime Detection Measures for the Practical Ecologist" (2019). Dissertations & Theses in Natural Resources. 299. https://digitalcommons.unl.edu/natresdiss/299

This Article is brought to you for free and open access by the Natural Resources, School of at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in Dissertations & Theses in Natural Resources by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.
REGIME DETECTION MEASURES FOR THE PRACTICAL ECOLOGIST

by

Jessica Leigh Burnett

A DISSERTATION

Presented to the Faculty of

The Graduate College at the University of Nebraska

In Partial Fulfilment of Requirements

For the Degree of Doctor of Philosophy

Major: Natural Resource Sciences

Under the Supervision of Professors Craig R. Allen & Dirac Twidwell

Lincoln, Nebraska
August, 2019
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 predictions in sufficient time to prevent or mitigate unwanted systemic change. Early warning systems (or early warning signals, early warning indicators) have been developed and tested for some ecological systems data, but have been mostly applied to marine fisheries time series and nutrient loadings in shallow lakes. Despite the numerous quantitative methods proposed for identifying or forecasting regime shifts in ecological data, few are used in practice. This dissertation contributes to our understanding of the utility and limitations of early warning systems for ecological regime shift detection, referred to here as ‘regime detection measures’. Using both theoretical and empirical data, I evaluate the efficacy of multivariate regime detection measures in identifying abrupt shifts in ecological communities over time and across space. I also introduce a method which I refer to as ‘velocity’ (of a system’s trajectory in phase space) as a potential regime detection measure. Using resampling techniques, I find the velocity method is more robust to data loss and data quality than are the Fisher Information and Variance Index methods which have been previously applied to empirical systems data. 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.
DEDICATION

To those not yet exposed to the great outdoors, first generation college students, Mike Moulton, S, and myself.
ACKNOWLEDGMENTS

Graduate school itself isn’t hard, but the journey is. I have a lot of people and institutions to thank for their emotional, intellectual, financial, and other support. I wish to first highlight how **great it was to be a graduate student at this university and in the School of Natural Resources**. UNL has provided tremendous support at all levels of the university. Although I am not a fan of Nebraska’s climate, I highly recommend this school to prospective students. I thank my supervisors, Craig Allen and Dirac Twidwell, for providing me with this amazing opportunity and for supporting my growth as an independent researcher. I thank my also committee members, Craig Allen, David Angeler, John De Long, Dirac Twidwell, and Drew Tyre for their support and advisement, but especially for their comprehensive examination—I found this process transformative (albeit very stress-inducing). I particularly thank Dirac for his examination questions—I never knew how much I didn’t know until I studied your recommendations. I also thank Craig and Dirac for supporting my efforts to study and conduct research in Austria. Studying at the International Institute for Applied Systems Analysis was an amazing opportunity! I thank Brian Fath and Elena Rovenskaya for their advisement, members of the Applied Systems Analysis research group for their feedback on my research, and to the postdocs and YSSPers. I owe thanks to Craig Allen and Kevin Pope for entertaining my many hours of discussion (interrogation?) regarding federal employment. I would like to especially thank some of the amazing and brilliant **female scientists** in my life for their encouragement: Jane Anderson, Karen Bailey, Hannah Birge, Mary Bomberger Brown, Tori Donovan, Brittany Dueker, Allie Schiltmeyer, Katie Sieving, Erica Stuber, Becky Wilcox, Carissa Wonkka, and Lyndsie Wszola. I thank these women and others for their contributions to my professional development: David Angeler, Christie Bahlai, Mary
Bomberger Brown, John Carroll, Jenny Dauer, John DeLong, Tarsha Eason, Brian Fath, Ahjond Garmestani, Chris Lepczyk, Frank La Sorte, Chai Molina, Zac Warren, Hao Ye. I also thank fellow graduate students with whom I forged long-lived personal and professional relationships: Hannah Birge, Tori Donovan, Caleb Roberts, Allie Schiltmeyer, and Lyndsie Wszola. It is also worth noting that I among those afflicted with mental health "disorders". I am first grateful to one friend (H) who unknowingly destigmatized mental health in my mind and without whom I may have never sought treatment. I also applaud fellow students and faculty in SNR and UNL who have been active in promoting positive mental health. I am forever indebted to my general practitioner and mental health advocate, Terry Thomas M.A., M.S.N., A.P.R.N. This research was funded by the U.S. Department of Defense’s Strategic Environmental Research and Development Program (SERDP project ID: RC-2510). The University of Nebraska-Lincoln (UNL) has been highly supportive in my doctoral studies and research. I am grateful for the generous of donors to the University of Nebraska Foundation, which provided me with two prestigious supplemental fellowships: Fling and Othmer. I also thank the Nelson Family (Nelson Memorial Fellowship, UNL) for supporting my domestic and international travel to conferences and workshops, and the Institute of Agriculture and Natural Resources, who funded large portions of my academic and research-related travel. I thank the School of Natural Resources for their financial support in my conference travel. The U.S. National Academy of Sciences generously funded part of my travel to the International Institute for Applied Systems Analysis (IIASA). This financial support provided me not only with invaluable opportunities to attend and present at national and international conferences and workshops, conduct research abroad, and network—this funding alleviated some financial pressures associated with graduate school which allowed a more refined focus on my dissertation research. The opportunities and experiences provided to me
by each funding source were amazing, thank you. Finally, to my partner of eight years–Schultzie–thank you for everything. Just kidding, thank you, Nat Price, you are amazing.
Table of Contents

List of Tables viii

List of Figures ix

Table of Definitions 1

1 Introduction 10
   1.1 Dissertation Aims . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 12
   1.2 Dissertation Structure and Contents . . . . . . . . . . . . . . . . . . 12
   1.3 Appendices and Software . . . . . . . . . . . . . . . . . . . . . . . . . 15

2 Methods for Detecting Ecological Regime Shifts 17
   2.1 Introduction . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 17
   2.2 Methods . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 19
      2.2.1 Quantitative methods for identifying ecological regime shifts . 19
      2.2.2 Bibliographic analysis of the ecological regime shift literature . 20
   2.3 Results . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 28
      2.3.1 Quantitative methods for identifying ecological regime shifts . 28
      2.3.2 Bibliographic analysis of ecological regime shift literature . . . 29
         2.3.2.1 Thematic mapping using keywords . . . . . . . . . . 29
         2.3.2.2 Historiograph, citation, and reviews . . . . . . . . 32
2.4 A synthesis of the methods available for the practical ecologist . . . . 35

2.4.1 Model-dependent . . . . . . . . . . . . . . . . . . . . . . . . 37

2.4.2 Model-free . . . . . . . . . . . . . . . . . . . . . . . . . . . . 37

2.5 Discussion . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 38

3 Deconstructing the steps for calculating Fisher Information as a measure of abrupt change in ecological systems 44

3.1 Abstract . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 44

3.2 Introduction . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 45

3.2.1 Fisher Information as a Regime Detection Method . . . . . . . . 46

3.2.2 The Sustainable Regimes hypothesis . . . . . . . . . . . . . . . 47

3.2.3 Fisher Information requires dimension reduction prior to calculation . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 48

3.2.4 Aims . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 49

3.3 Methods . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 50

3.3.1 Predator-prey model system . . . . . . . . . . . . . . . . . . . . 50

3.3.2 Inducing a Regime Shift . . . . . . . . . . . . . . . . . . . . . . 50

3.3.3 Decoupling the Steps for Calculating Fisher Information . . . . 51

3.4 Discussion . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 55

3.5 Acknowledgements . . . . . . . . . . . . . . . . . . . . . . . . . . . . 57

4 An application of Fisher Information to bird community data does not reveal distinct regimes in the continental United States 64

4.1 Introduction . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 64

4.2 Data and Methods . . . . . . . . . . . . . . . . . . . . . . . . . . . . 65

4.2.1 Data: North American breeding bird communities . . . . . . . . 65

4.2.2 Study area . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 67
4.2.1 Focal military base ........................................ 67
4.2.2 Spatial sampling grid ........................................ 69
4.2.3 Calculating Fisher Information (FI) ....................... 70
4.2.4 Interpreting and comparing Fisher Information across spatial transects ...................................................... 73
4.2.4.1 Interpreting Fisher Information values .................. 73
4.2.4.2 Interpolating results across spatial transects .......... 74
4.2.4.3 Spatial correlation of Fisher Information ............... 75
4.3 Results .................................................................. 75
4.3.1 Fisher Information across spatial transects ............... 75
4.3.2 Spatial correlation of Fisher Information ................. 79
4.4 Discussion ............................................................. 83
5 Using Resampling Methods to Evaluate the Relative Performance of Regime Detection Measures ............................................ 87
5.1 Introduction .......................................................... 87
5.2 Data and Methodology ............................................ 91
5.2.1 Study system and data ........................................ 91
5.2.2 Regime detection measures .................................. 92
5.2.2.1 Velocity (v) calculation .................................. 92
5.2.2.2 Variance Index (VI) calculation ....................... 93
5.2.2.3 Fisher Information (FI) calculation .................... 93
5.2.2.4 Using moving window analysis to calculate Fisher Information and Variance Index ....................... 94
5.2.3 Simulating data quality and quantity issues using resampling techniques ...................................................... 96
6 Velocity (v): Rate of Change of System Trajectory Identifies Abrupt Changes in Ecological Systems

6.1 Introduction .................................................. 107

6.1.1 Tracking ecosystem trajectory through time to explore system dynamics .............................................. 109

6.1.2 Rate of change as an indicator of abrupt change in the system trajectory .................................................. 110

6.1.3 Aims ................................................................. 111

6.1.4 Analytical approach ......................................... 112

6.2 Steps for Calculating Velocity, V .................................. 112

6.2.1 Step 1: Calculate $\Delta x_i$ .................................... 114

6.2.2 Step 2: Calculate distance traveled, $s$ ....................... 115

6.2.3 Step 3: Calculate velocity, $V$ (or $\frac{\Delta s}{\Delta t}$) .................. 116

6.3 Velocity v Performance Under a Discontinuous Transition .................................................. 117

6.3.0.1 Varying post-shift mean .................................. 118

6.3.0.2 Varying Post-shift Variance ............................... 119

6.3.0.3 Smoothing the data prior to calculating $v$ ............... 120
6.4 Velocity Performance Under a Smooth Transition  
   6.4.1 Generating the data  
      6.4.1.1 True process model  
      6.4.1.2 Observed process data  
   6.4.2 Evaluating Velocity Performance Under Conditions of Changing Means and Variance  
      6.4.2.1 Smooth changes in the mean  
      6.4.2.2 Smooth changes in variance  
      6.4.2.3 Smooth changes in the mean and variance  
6.5 Velocity Performance Under Empirical Transitions: paleolithic freshwater diatom communities  
6.6 Discussion  
6.7 Supplementary Figures  

7 Grassland Obligates and Declining Birds Operate Near Edges of Body Mass Distributions  
   7.1 Introduction  
   7.2 Methods  
      7.2.1 Study area  
      7.2.2 Data  
         7.2.2.1 Avian census data - North American Breeding Bird Survey  
         7.2.2.2 Identifying species of interest  
      7.2.3 Statistical analysis  
         7.2.3.1 Identifying scaling structure of avian communities using body mass distributions
Appendix C  bbsRDM: an R package for applying `regimeDetectionMeasures` functionality to spatial data 190

Appendix D Functions used to calculate discontinuities in avian body mass distributions. 191

References 193
List of Tables

01  A table of definitions for terms, theories, and phrases often appearing in ecological regime shift literature. ......................... 2

21  List of the regime detection methods identified using expert knowledge and previously published reviews of methods. ............... 23

22  Potential questions for a comprehensive review of the ecological regime detection metrics literature. ................................. 41

61  Steps outlined for calculating system velocity, $V$, using the 2-variable toy data as an example. ........................................... 117

62  Conditions for generating various scenarios of the hyperbolic tangent-induced abrupt change. $\sigma_i$ represents the standard deviation of $\mu_{x_i}$ as the percent of $\mu_{x_i}$, $\mu_{x_i}$ is the mean of the state variable, $x_i$, and pre and post represent the periods before and after the regime shift at $t = 50$, respectively. ......................................................... 124

71  Grassland obligates and species with declining trends over the period of (1966-2015) in the Central Breeding Bird Survey region in our study area. 159
The number of NABBS routes analysed in the Southern regime is smaller than those used in the Northern regime each year given the location of the regimes identified in a previous study with respect to the contiguous grasslands of Central North America.

Summary statistics for annual species richness and annual turnover in all NABBS routes in study area.

Coefficient estimates for the linear mixed effects model predicting species’ ‘distance to edge’ of a body mass distribution.
# List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>21</td>
<td>Number of articles containing phrases related to ecological ‘regime shifts’ published per publication outlet.</td>
</tr>
<tr>
<td>22</td>
<td>Number of publications by year in fields ‘Ecology’ and ‘Biodiversity Conservation’ which included terms related to ‘regime shift’ (total = 654).</td>
</tr>
<tr>
<td>23</td>
<td>A thematic map of the clusters (themes) identified among the (a) author-supplied and (b) ISI-supplied keywords for each article using a clustering algorithm.</td>
</tr>
<tr>
<td>24</td>
<td>The themes identified in the literature based on author-supplied keywords vary as the number of words included in the analysis increased.</td>
</tr>
<tr>
<td>25</td>
<td>The themes identified in the literature based on the ISI-supplied keywords vary as the number of words included in the analysis increased.</td>
</tr>
<tr>
<td>26</td>
<td>Chronological direct citation network suggests the intellectual structure can be mapped to a few papers. This historiograph identifies important works explicitly in chronological, as opposed to absolute, order.</td>
</tr>
<tr>
<td>27</td>
<td>Total number of articles published and corresponding number of citations (for papers published that year). The most highly cited papers to-date are those published in the late 2000s.</td>
</tr>
<tr>
<td>28</td>
<td>Total number of articles published per year by category as categorized by ISI. Book chapters, proceedings, editorials, and letters are excluded.</td>
</tr>
</tbody>
</table>
From top to bottom, distance traveled in phase space, speed tangential to system trajectory, acceleration tangential to system trajectory.

Carrying capacity over time with a regime shift occurring around time 200.

Phase space plot of system trajectories for different values of k. Colored point indicates the coordinates corresponding to the shift induced in carrying capacity.

Distance travelled in phase space over time. Dashed vertical line at time 200 indicates location of regime shift.

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

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.

Locations of Breeding Bird Survey routes sampled between 1966 and 2017.

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

Three East-West running transects analyzed in this Chapter overlayed against the Omernick Ecoregion boundaries map.

The change in the Fisher Information values along a single, East-West-running spatial transect (Transect number 12) over time.

Fisher Information of two spatially adjacent transect pairs (transects 12, 13) over time. Interdecadal trends in FI are very different within each transect and are not highly correlated among transects over time.
Fisher Information of select East-West spatial transects over time. These transects are representative of the results obtained from the remaining transects.

Contrary to expectations there appear no clear patterns of abrupt change in the Fisher Information results along three selected East-West running transects (Year 2010). The results are visualized here at regularly-spaced intervals such that the locations are not precise relative to the location of corresponding Breeding Bird Survey routes.

Pairwise relationships of Fisher Information (interpolated values) of spatially adjacent transects over time do not exhibit expected patterns of high positive correlation. 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.

Pairwise relationships of Fisher Information (interpolated values on the log-scale) of spatially adjacent transects over time. 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.

Fisher Information results (scaled and centered; larger circles (points) represent higher values of Fisher Information) plotted against ecoregion boundaries (EPA Level 2 boundaries) for the year 2000.

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

The amount of time elapsed between observations for the Foy Lake paleo- diatom data.
53 Density plot of the coefficient of variation (CV) as a percentage (%) of the Variance Index resampled values over 10,000 iterations. Densities are drawn based on all values of CV but values greater than 100% are not printed. ........................................... 94

54 Mean Variance Index (VI) and associated 95% confidence intervals over 10,000 iterations using the *observations* resampling method. Red line indicates the value of VI when \( M \) and \( P \) are 100%. ........................................... 95

55 Mean Variance Index (VI) and associated 95% confidence intervals over 10,000 iterations using the *species* resampling method. Red line indicates the value of VI when \( M \) and \( P = 100\% \). ........................................... 96

56 Mean velocity and associated 95% confidence intervals over 10,000 iterations using the observations resampling method. Red line indicates the value of velocity when \( M \) and \( P \) are 100%. ........................................... 97

57 Mean velocity and associated 95% confidence intervals over 10,000 iterations using the observations resampling method for a subset of the time series (the second ‘regime’ identified). Red line indicates the value of velocity when \( M \) and \( P = 100\% \). ........................................... 99

58 Density plot of the coefficient of variation (CV) as a percentage (%) of the Distance Metric (s) resampled samples (10,000 iterations). Densities are drawn based on all values of CV, but values >100% are not printed. . . . 100

59 Density plot of the coefficient of variation (CV) as a percentage (%) of the Fisher Information resampled samples (10,000 iterations). Densities are drawn based on all values of CV, but values >100% are not printed. . . . 101
Mean Fisher Information (FI; note the scale) and associated 95% confidence intervals over 10,000 iterations using the species resampling method. Red line indicates the value of FI when $M$ and $P$ are 100%. A very small value was added to the mean FI prior to log transformation.

Local regression (loess) smoothing of a dominant species in the paleodiatom community, *Anomoeoneis costata* varies with the span parameter, making it difficult to justify smoothing the data prior to calculating various regime detection metrics.

An example solution of the Lorenz ('butterfly') represented in 3-dimensional phase-space. Phase plots are typically used to visualize stable areas within a system’s trajectory but reconstruction requires the difference models to be known and parameterized.

An example solution of the Lorenz ('butterfly') represented in individual system components.

The 2-variable discrete time toy system used to demonstrate steps for calculating system velocity. Each variable, $x$, is drawn from a normal distribution with means that change at $t = 50$. State variables have constant standard deviation, $\sigma = 5$.

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

Distance traveled, $s$, for the 2-species toy system.

System change ($s$) and velocity ($V$) of the model system over the time period. Constant means ($\bar{x}_{pre} = 25$, $\bar{x}_{post} = 10$) and sharp change in variance for both state variables, $\sigma = 5$. 
Velocity \( (V) \) generally increases as the total change in the mean value of \( \bar{x}_{1_{t=50}} \) increases in a single iteration of our toy system \((N_{\text{iter}} = 1, \text{seed} = 123)\). This 2-variable system exhibits a regime shift at \( t = 50 \), where variance is constant \( \sigma = 5 \), \( \bar{x}_1 = 25 \) when \( t < 50 \), \( \bar{x}_2 = 50 \) when \( t \geq 50 \), \( \bar{x}_1 = 25 \) when \( t < 50 \).

Change in velocity \( (V) \) as the total change in the mean value of \( \bar{x}_{2_{t=50}} \) over 10,000 simulations. A regime shift was induced at \( t = 50 \) with constant variance \( \sigma = 5 \), \( \bar{x}_2 = 25 \) when \( t < 50 \), and changes in variable mean values, \( \bar{x}_2 = 50 \) when \( t \geq 50 \), \( \bar{x}_1 = 25 \) when \( t < 50 \).

High variance of velocity \( (V) \) in a single iteration \((N_{\text{iter}} = 1, \text{seed} = 123)\) of simulations as we increase \( \sigma_1 \) at \( t = 50 \).

The variance around the average \((\pm 2 \text{ SD})\) velocity \( (V) \) increases as the variance of \( \bar{x}_{2_{t=50}(\text{post})} \) (post shift) increases. \( \bar{x}_{1_{\text{pre}}}=25, \bar{x}_{1_{\text{post}}}=100 \), \( \bar{x}_{2_{\text{pre}}}=25, \bar{x}_{2_{\text{post}}}=50, \sigma_{1_{\text{pre}}}=5, \sigma_{2_{\text{pre,post}}}=5 \).

The noise in system velocity \( (V) \) is not obviously reduced in this system as the original data \((x_1, x_2)\) is increasingly smoothed.

An example of the data generated by the true process model. In this example the mean values \((\mu_{x_i})\), but not the percent standard deviation \((\sigma_{x_i})\), are varied before and after the transition point. The observed data are plotted against the true-process model for each state variable, \( x_i \). Panels represent different degrees of the smoothing parameter, \( \alpha \) (top: \( \alpha = 0.25 \), bottom: \( \alpha = 1.00 \)).
An example of the data generated by the true process model. In this example the mean values ($\mu_{xi}$), but not the percent standard deviation ($\sigma_{xi}$), are varied before and after the transition point. The observed data are plotted against the true-process model for each state variable, $x_i$. Panels represent different degrees of the smoothing parameter, $\alpha$ (top: $\alpha = 0.25$, bottom: $\alpha = 1.00$).

Antidifferentiated values (‘observed’) of the distance traveled, $s$, to the true process values of $s$ (‘true’) provides a method for identifying the best values of the smoothing parameter, $\alpha$. Under most conditions $\alpha \ll$ sufficiently. Here, we compare the true and antidifferentiated values of $s$ under the condition of changing $\mu_{x1}$ when the hyperbolic tangent function is most rapid ($\alpha_{\text{tanh}} = 1$) for the ‘tvdiff’ $\alpha = 0.50$. Not pictured: the antidifferentiated values of $s$ (observed) is increasingly smoothed as $\alpha$ increases.

The velocity signal is muted when the hyperbolic smoothing parameter, $\alpha$, is low (0.25). True and observed values of $x_i$ (panel A), observed distance traveled ($s$, panel B), observed velocity (C), and the smoothed velocity (D).

The velocity signal is muted when the hyperbolic smoothing parameter, $\alpha$, is moderate (0.50). True and observed values of $x_i$ (panel A), observed distance traveled ($s$, panel B), observed velocity (C), and the smoothed velocity (D).

The velocity signal is muted when the hyperbolic smoothing parameter, $\alpha$, is moderate (0.50). True and observed values of $x_i$ (panel A), observed distance traveled ($s$, panel B), observed velocity (C), and the smoothed velocity (D).
The velocity signal is muted when the hyperbolic smoothing parameter, \( \alpha \), is moderate (0.50). True and observed values of \( x_i \) (panel A), observed distance traveled (\( s \), panel B), observed velocity (C), and the smoothed velocity (D).

The velocity signal is regained under smooth transition (\( \alpha_{\text{tanh}} = 0.75 \)) when both state variables undergo a shift in the mean. True and observed values of \( x_i \) (panel A), observed distance traveled (\( s \), panel B), observed velocity (C), and the smoothed velocity (D).

The velocity signals a rapid shift in the variance of both state variables under a moderately abrupt transition (\( \alpha_{\text{tanh}} = 0.75 \)). True and observed values of \( x_i \) (panel A), observed distance traveled (\( s \), panel B), observed velocity (C), and the smoothed velocity (D).

The velocity does not signal shifts in the variance of a single variable (\( x_1 \)) under a moderately abrupt transition (\( \alpha_{\text{tanh}} = 0.75 \)). True and observed values of \( x_i \) (panel A), observed distance traveled (\( s \), panel B), observed velocity (C), and the smoothed velocity (D).

The velocity signal is regained under smooth transition (\( \alpha_{\text{tanh}} = 0.25 \)) when both state variables undergo a shift in the mean. True and observed values of \( x_i \) (panel A), observed distance traveled (\( s \), panel B), observed velocity (C), and the smoothed velocity (D).

The velocity signals a shift when both variables undergo shifts in the mean and variance under a slightly abrupt transition (\( \alpha_{\text{tanh}} = 0.25 \)). True and observed values of \( x_i \) (panel A), observed distance traveled (\( s \), panel B), observed velocity (C), and the smoothed velocity (D).
The velocity signals a shift when both variables undergo shifts in the mean and variance under a slightly abrupt transition ($\alpha_{\text{tanh}} = 1.00$). True and observed values of $x_i$ (panel A), observed distance traveled ($s$, panel B), observed velocity (C), and the smoothed velocity (D).

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.

Velocity $V$ and distance traveled $s$ of the paleodiatom time series. Dashed line at 1,300 years before 1950 indicates the regime shift identified in Spanbauer et al. (2014). Dotted lines indicate regime shifts as visually identified on metrics $s$ and $V$.

Inter-regime (regimes A and B) trends in the velocity signal ($V$) identify the fluctuating decadal and centurial abiotic conditions discussed in Spanbauer et al. (2014).

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

The velocity metric ($V$) signals potential periodicities in the paleo diatom time series data when the distance traveled metric, $s$, is smoothed using regularized differentiation methods (see Price and Burnett 2019).

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

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 $\sigma_{1\text{pre}} = 2$, $\sigma_{1\text{post}} = 12$, $\sigma_{2\text{pre,post}} = 5$. 

...
System change ($s$) and velocity ($V$) of the model system over the time period. Variance equal to mean ($\bar{x}_i = \sigma_i$), where means ($\bar{x}_{1\text{pre}} = 25$, $\bar{x}_{1\text{post}} = 50$, $\bar{x}_{2\text{pre}} = 15$, $\bar{x}_{2\text{post}} = 150$).

My study area (red box) overlaying the terrestrial Breeding Conservation Regions (BCR) in North America.

North American Breeding Bird Survey routes (points), latitudinal locations (horizontal bars) of the spatial regimes identified by roberts2019shifting.

Number of NABBS routes analysed per year. Some routes are not sampled annually due to volunteer availability, environmental conditions, or route discontinuation.

Species richness increases over time across the entire study area.

Variance in the total species turnover increases over time across the entire study area.

Declining species were closer to the edge of body mass aggregations in the 'northern' than were in the 'southern' regimes.

Declining species were closer to the edge of body mass aggregations in the 'northern' than were in the 'southern' regimes.

Total number of birds across the entire study area per species group per year.

Average coefficient of variation in the number of birds across the entire study area per species group per year.

Average number of birds across the entire study area per species group per year.
Number of body mass aggregations identified in each route unchanged across the time period (**a**) and is approximately normally distributed ($\bar{x} = 4.7, \sigma = 1.6$)...

Relationship between species richness per route and (a) the number of aggregations identified in body mass distributions and (b) distance to the edge (units log body mass) of aggregations...

Aggregation locations of body mass distributions of the avian community at a single NABBS route (state 7 route 24) appear relatively similar across time. Panels include data for years 1=1970, 2=1985, 3=2000, 4=2015.

Red triangles indicate grassland obligate species.

Body mass distribution for species in the study area over the entire time period varies by species group. Distributions represent the species pool for each group over the entire study area and all years.

The body mass distribution of declining species differ only slightly between the Southern and Northern regimes.

Body mass distribution for species in the study area over the entire time period varies by species group. Distributions represent the species pool for each group and Bird Conservation Region over all years.
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 01) 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 01: A table of definitions for terms, theories, and phrases often appearing in ecological regime shift literature.

<table>
<thead>
<tr>
<th>Term</th>
<th>Definition</th>
<th>Synonyms</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abrupt</td>
<td>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.</td>
<td>big, fast, quick, large</td>
</tr>
<tr>
<td>Alternative Stable State</td>
<td>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).</td>
<td></td>
</tr>
<tr>
<td>Attractor</td>
<td>The set of values towards which a system tends regardless of its initial (starting) values.</td>
<td></td>
</tr>
<tr>
<td>Basin-Boundary Collision</td>
<td>The parameter values for a system that causes the system to shift between alternate attractors.</td>
<td>non-local</td>
</tr>
<tr>
<td>Catastrophe Theory</td>
<td>The study of abrupt changes within a dynamical system.</td>
<td></td>
</tr>
<tr>
<td>Catastrophic Bifurcation</td>
<td>A relatively abrupt jump to an alternate attractor due to initial attractor.</td>
<td></td>
</tr>
<tr>
<td>Change-Point</td>
<td>See also 'Regime Shift'. A term often used in computer science, climatology, data science; represents the point at which a state changes its configuration.</td>
<td></td>
</tr>
<tr>
<td>Term</td>
<td>Definition</td>
<td>Synonyms</td>
</tr>
<tr>
<td>---------------------</td>
<td>---------------------------------------------------------------------------</td>
<td>--------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Change-Point Detection</td>
<td>A change point method which does not require supervision; identifies potential change points without a priori potential change points.</td>
<td></td>
</tr>
<tr>
<td>Change-Point Estimation</td>
<td>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.</td>
<td></td>
</tr>
<tr>
<td>Chaos</td>
<td>A system with extreme sensitivity to initial conditions.</td>
<td></td>
</tr>
<tr>
<td>Critical Slowing Down (CSD)</td>
<td>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).</td>
<td></td>
</tr>
<tr>
<td>Degrees of Freedom</td>
<td>The number of system parameters or components which vary independently.</td>
<td></td>
</tr>
</tbody>
</table>
Table 01: A table of definitions for terms, theories, and phrases often appearing in ecological regime shift literature. (continued)

<table>
<thead>
<tr>
<th>Term</th>
<th>Definition</th>
<th>Synonyms</th>
</tr>
</thead>
<tbody>
<tr>
<td>Domain of Attraction</td>
<td>The range of values around which a system fluctuates.</td>
<td>zone of fluctuation, basin of attraction,</td>
</tr>
<tr>
<td></td>
<td></td>
<td>stable point, attractor</td>
</tr>
<tr>
<td>Driver</td>
<td>A widespread anthropogenic source of change which leads to one or more pressures (e.g., land-use change).</td>
<td></td>
</tr>
<tr>
<td>Driver-Threshold Regime</td>
<td>When a rapid change in external driver induces a rapid change in ecosystem state.</td>
<td></td>
</tr>
<tr>
<td>Dynamical System</td>
<td>A time-dependent system which can be described in state-space.</td>
<td></td>
</tr>
<tr>
<td>Dynamical Systems Theory</td>
<td>The study of complex systems theory; the study of time-dependent systems.</td>
<td></td>
</tr>
<tr>
<td>Equilibrium</td>
<td>The set of values around which a system revolves and does not change.</td>
<td></td>
</tr>
<tr>
<td>Exogeneous Process</td>
<td>An external process influencing the state of the dynamical system.</td>
<td></td>
</tr>
<tr>
<td>(Forcing, Driver)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### Table 01: A table of definitions for terms, theories, and phrases often appearing in ecological regime shift literature. (continued)

<table>
<thead>
<tr>
<th>Term</th>
<th>Definition</th>
<th>Synonyms</th>
</tr>
</thead>
<tbody>
<tr>
<td>First-Order Stationarity</td>
<td>When the mean is constant over the observations.</td>
<td></td>
</tr>
<tr>
<td>Fold Bifurcation</td>
<td>This occurs when a stable point collides with an unstable point; when crossing a tipping point induces hysteresis.</td>
<td></td>
</tr>
<tr>
<td>Fractal Properties</td>
<td>A measurement of geometrical self-similarity; when a system has similar structure regardless of the scale of observation.</td>
<td>ergodic</td>
</tr>
<tr>
<td>Hysteresis</td>
<td>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.</td>
<td></td>
</tr>
<tr>
<td>Leading Indicators</td>
<td>When the statistical properties of the fluctuations (of the data) approach a critical transition.</td>
<td></td>
</tr>
<tr>
<td>Lyapunov Exponent (and Stability)</td>
<td>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.</td>
<td></td>
</tr>
<tr>
<td>Measure Theory</td>
<td>The study of measures and measurement (e.g. volume, mass, time).</td>
<td></td>
</tr>
</tbody>
</table>
Table 01: A table of definitions for terms, theories, and phrases often appearing in ecological regime shift literature. (continued)

<table>
<thead>
<tr>
<th>Term</th>
<th>Definition</th>
<th>Synonyms</th>
</tr>
</thead>
<tbody>
<tr>
<td>Moving (Sliding) Window Analysis</td>
<td>When a subsample of the data $X_t$ is used in lieu of a single observation, $x_t$.</td>
<td></td>
</tr>
<tr>
<td>Noise</td>
<td>Processes manifested in data which are unaccounted for; sometimes referred to as meaningless; random variability.</td>
<td></td>
</tr>
<tr>
<td>Non-Stationarity of the Mean Value</td>
<td>Infers that a trend or a periodicity is present in the time series.</td>
<td></td>
</tr>
<tr>
<td>Online</td>
<td>Real-time updating of model parameters, predictions, etc. (c.f. offline).</td>
<td></td>
</tr>
<tr>
<td>Persistent</td>
<td>A relative value of the longevity of the observed change in values.</td>
<td>long-lasting</td>
</tr>
<tr>
<td>Phase Space</td>
<td>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.</td>
<td></td>
</tr>
<tr>
<td>Prediction</td>
<td>A temporal forecast. Is intrinsic when a model and parameters are used to make forecast, is realized when the prediction becomes the actual state of the system.</td>
<td></td>
</tr>
<tr>
<td>Pressure</td>
<td>A perturbation which negatively influences a system, and can be defined as pulse, press, or monotonic.</td>
<td></td>
</tr>
</tbody>
</table>
Table 01: A table of definitions for terms, theories, and phrases often appearing in ecological regime shift literature. (continued)

<table>
<thead>
<tr>
<th>Term</th>
<th>Definition</th>
<th>Synonyms</th>
</tr>
</thead>
<tbody>
<tr>
<td>Red Noise</td>
<td>Noise having zero mean, constant variance, and serial autocorrelation;</td>
<td></td>
</tr>
<tr>
<td></td>
<td>autocorrelated random variability.</td>
<td></td>
</tr>
<tr>
<td>Regime</td>
<td>A set of system values that define a particular system state. Not</td>
<td></td>
</tr>
<tr>
<td></td>
<td>necessarily stable, but some state variables or outputs of the system</td>
<td></td>
</tr>
<tr>
<td></td>
<td>remain relatively constant over a defined period of time.</td>
<td></td>
</tr>
<tr>
<td>Regime Shift</td>
<td>&quot;abrupt&quot; and &quot;persistent&quot; change in a system’s structure or functioning.</td>
<td></td>
</tr>
<tr>
<td>Second-Order</td>
<td>The mean is constant and the covariance is a function of a time lag,</td>
<td></td>
</tr>
<tr>
<td>Stationarity</td>
<td>but not of time.</td>
<td></td>
</tr>
<tr>
<td>Self-Similarity</td>
<td>A system satisfied by power-law scaling.</td>
<td></td>
</tr>
<tr>
<td>Stable Equilibrium</td>
<td>An equilibrium is stable when small perturbations do not induce</td>
<td></td>
</tr>
<tr>
<td></td>
<td>change.</td>
<td></td>
</tr>
<tr>
<td>State Space</td>
<td>The set of all possible configurations of a system.</td>
<td></td>
</tr>
<tr>
<td>State-Threshold</td>
<td>When a gradual change in external driver induces a rapid change in</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ecosystem state (e.g., System crosses a threshold).</td>
<td></td>
</tr>
<tr>
<td>Regime Shift</td>
<td>When a gradual change in external driver induces a rapid change in</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ecosystem state (e.g., System crosses a threshold).</td>
<td></td>
</tr>
<tr>
<td>Stationarity</td>
<td>When the probability density function of a system does not change with time.</td>
<td></td>
</tr>
</tbody>
</table>
Table 01: A table of definitions for terms, theories, and phrases often appearing in ecological regime shift literature. (continued)

<table>
<thead>
<tr>
<th>Term</th>
<th>Definition</th>
<th>Synonyms</th>
</tr>
</thead>
<tbody>
<tr>
<td>Statistical Stationarity</td>
<td>A system with statistical properties unchanging over time. This concept extends to periodic stationarity for systems exhibiting periodic behavior.</td>
<td></td>
</tr>
<tr>
<td>Strange Attractor</td>
<td>An attractor which has fractal structure (an observable fractal dimension).</td>
<td></td>
</tr>
<tr>
<td>Supervised Machine</td>
<td>When classifiers are used to train the data a priori.</td>
<td></td>
</tr>
<tr>
<td>Learning</td>
<td></td>
<td></td>
</tr>
<tr>
<td>System State</td>
<td>The observed (current) instance of the system within a state space.</td>
<td></td>
</tr>
<tr>
<td>Threshold</td>
<td>A point where the system reacts to changing conditions.</td>
<td></td>
</tr>
<tr>
<td>Tipping Point</td>
<td>A point in a system’s trajectory where a small change in an endogenous force induces a large change in system state or values; the point where a system can flip into an alternative state.</td>
<td></td>
</tr>
<tr>
<td>Trajectory</td>
<td>The path of an object or system through space-time.</td>
<td>orbit, path</td>
</tr>
<tr>
<td>Transient</td>
<td>A behavior or phenomenon which is responsive to initial (starting) conditions, or its effect declines over time.</td>
<td></td>
</tr>
<tr>
<td>Trend Smoothing</td>
<td>Local averaging of values such that the non-systematic components of the system are washed out.</td>
<td></td>
</tr>
</tbody>
</table>
**Table 01:** A table of definitions for terms, theories, and phrases often appearing in ecological regime shift literature. (continued)

<table>
<thead>
<tr>
<th>Term</th>
<th>Definition</th>
<th>Synonyms</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unstable Equilibrium</td>
<td>An equilibrium is unstable when small perturbations induce change.</td>
<td></td>
</tr>
<tr>
<td>Unsupervised Machine Learning</td>
<td>When no prior training is required (i.e. no classifications necessary a priori) to classify it.</td>
<td></td>
</tr>
<tr>
<td>White Noise</td>
<td>Noise having zero mean, constant variance, and is not autocorrelated; uncorrelated random variability.</td>
<td></td>
</tr>
</tbody>
</table>
Chapter 1

Introduction

Anthropogenic activity in the last few decades will continue to influence the interactions within and among ecological systems worldwide. The complexity 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 (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 Kaufmann et al., 2015; banking and stock markets Davis & Karim, 2008). 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. The ecological literature is inundated with quantitative methods and models with the promise of predicting abrupt change in high-dimensional ecological systems in time for intervention. The paucity of application of many of these methods by practitioners and decision makers suggests much work is to be done in advancing both our understanding of abrupt ecological change and of the methods used for detecting it.

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 predictions in sufficient time to prevent or mitigate unwanted systemic change. Early
warning systems (or early warning signals, early warning indicators) have been developed and tested for some ecological systems data, but have been mostly applied to marine fisheries time series and nutrient loadings in shallow lakes. Despite the numerous quantitative methods [see Chapter 2] proposed as early warning signals for ecological data, many are currently of limited practical utility. This paradox may be a consequence of existing early warning systems having one or more of the following characteristics:

1. Not generalizable across systems or system types (especially when it requires a model or a deterministic function to describe the system)
2. Requires 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 and in presence of noise
7. Gives no uncertainty around estimates (tying into interpretation issues)
8. Ignores or does not sufficiently account for observation error
9. Currently no baseline with which to compare results
10. Currently no application/testing on empirical systems data
11. Systems are subjectively bounded (i.e., components are chosen)
12. Being overshadowed by semantics
13. Are based on before-and-after information
14. 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)
15. Cannot handle irregular sampling
16. Cannot handle non-smooth or non-linear data
Research focusing on the above areas as they relate to regime detection measures may improve the quality of these early warning systems, further shedding light on the potential of these methods for application to place-based observations.

1.1 Dissertation Aims

This dissertation contributes to our understanding of the utility and limitations of early warning systems for ecological regime shift detection, referred to here as ‘regime detection measures’. Regime detection measures exist for handling both univariate and multi-variable data, however, it is the latter of these methods within which this dissertation focuses. Although the univariate regime detection measures are currently more widely applied and conceptually tractable than many multivariate regime detection measures, the utility of the univariate measures may be limited when change(s) in the system dynamics manifest in entire community dynamics, rather than in select indicator species, for example. Multivariable regime detection measures may also be more advantageous than analysing individual variables when the drivers of the observed systemic change are unknown. Further, ecological systems are noisy, and ecological systems data are messy, conditions which complicate the use of univariate regime detection measures.

1.2 Dissertation Structure and Contents

This dissertation comprises a glossary (preface to Chapter 1), eight distinct Chapters (Chapters 1-8), and a compendium of open-source statistical software authored during the production of this dissertation in the form of two appendices (Appendices 4 and 2). Finally, the dissertation is synthesized in Chapter 8. The terminology associated with this line of research is highly variable both within and outside the field of ecology.
For example, although many core concepts informing ecological regime shift theory stem from dynamical systems theory, the terminologies do not align (Andersen et al., 2009; Hastings & Wysham, 2010). Therefore to ensure clarity of discussion throughout this work, I provide a glossary of phrases and concepts related to this dissertation in a front-matter.

Chapter 2. There exists a staggering number of quantitative methods for identifying abrupt changes and regime shifts in ecological systems data. Despite the high number of methods proposed in the literature, few have been scrutinized against empirical data, and even fewer applied to multiple types of systems (e.g., terrestrial mammals vs. marine fisheries). Although numerous reviews of these methods are published, few are comprehensive in their presentation of the proposed methods and metrics. In this Chapter 2 I provide a comprehensive list of the regime detection measures proposed in the relevant literature. I further conduct a bibliographic analysis of the ecological regime shift relevant literature.

Chapter 3. Fisher Information is proposed as a method for identifying regime shifts in multivariable ecological time series and spatially-explicit data and has been applied to a variety of systems across at least 20 publications. Two forms of this measure exist, one of which (the ‘derivaitves-based’ method) requires fewer steps and a priori defined parameters to calculate than the other (the ‘binning method’). Chapter 3 contributes to the understanding of the ‘derivatives-based’ Fisher Information as a regime detection measure in two ways. First, I present a step-by-step overview of both the logistics and concepts required for calculating this measure. Next, I suggest that the current calculation of Fisher Information can be split into two distinct parts: a dimensionality reduction and the actual calculation of the Fisher Information. The results of this study have implications for how the method can be used in the future, and whether it will suffice as an indicator of abrupt change under certain conditions.
Chapter 4  The Fisher Information method has been applied both temporal (Cabezas et al., 2010) and spatial empirical data (Sundstrom et al., 2017; Eason et al., 2019). To demonstrate the utility of Fisher Information in identifying abrupt change in ecological communities at large spatial scales, I present an application of Fisher Information to spatially-explicit avian community data in North America.

Chapter 6  Building off of the method described in Chapter 3 I suggest a method which I refer to as the ‘velocity’ method for identifying ecological regime shifts. Previous use of the velocity metric has been embedded within larger calculations of Fisher Information, specifically in the ‘derivatives-based’ method (see Chapter 3). The velocity method is an overlooked, simple calculation that may be useful in identifying abrupt changes in high dimensional temporal or spatial series. Here, I thoroughly describe the calculations behind the velocity metric, and demonstrate its utility through application to both simulated and empirical systems data.

Chapter 5  Of the numerous regime detection measures published (see Chapter 2), few have been applied to empirical multivariate ecological data, and even fewer scrutinized as indicators of abrupt change. In this Chapter I compare the ability of select regime detection measures to identify published abrupt changes in a paleodiatom community. Further, I examine the results of these measures under various conditions of data quality and quantity (e.g., missing species, infrequent sampling) using resampling methods. This Chapter also provides a critical starting point for determining the utility of the velocity method (proposed in Chapter 6) versus other techniques.

Chapter 7  This chapter presents an application of body mass discontinuity analysis to avian community time series before and after a landscape-scale regime shift (Roberts et al., 2019). In this chapter, I test the hypothesis that species which are sensitive to grassland habitat loss and degradation should be located near the
edges of body mass aggregations. Although I find evidence suggesting declining and sensitive (grassland obligate) species operate near the edges of body mass distributions as opposed to other species, I was unable to identify clear patterns in the avian community body mass distributions using a before-and-after design with respect to the proposed spatial regime (Roberts et al., 2019). Appendix 7 contains an annotated version of the code for the functions used to identify discontinuities in avian body mass distributions (originally published in Barichievy et al., 2018).

1.3 Appendices and Software

Appendices include brief descriptions and vignettes for three self-authored R packages, and are used throughout this dissertation. Each are available for download at www.github.com/trashbirdecology. A final Appendix contains the code used to conduct the discontinuity analysis used in Chapter 7.


**R package: regimeDetectionMeasures.** Appendix 4 contains code for calculating a suite of regime detection measures, including the traditional early-warning indicators, Fisher Information, and the velocity metric.

**R package: bbsRDM.** Appendix 2 contains functions for conducting spatial and temporal analysis of the regime detection measures contained in package regimeDetectionMeasures (Appendix 4), and is closely aligned with the analyses in Chapters 4 and 5. Further, this package provides a wrapper for downloading and munging data from the U.S. Geological Survey’s Breeding Bird Survey, munging said data, and cre-
ating spatial sampling grids across North America. Minor additional functionality includes an option to download and identify U.S. military bases across the globe.
Chapter 2

Methods for Detecting Ecological Regime Shifts

2.1 Introduction

Ecological regime shifts, or persistent changes in the underlying structure or functioning of natural systems due to exogenous forcings, are increasingly relevant in ecological research and management in an era of rapid and novel change due to anthropogenic activity. Identifying and predicting these types of changes is especially important when they impact ecosystem services or society in any other way. Numerous quantitative methods are proposed for identifying ecological regime shifts yet there exists a disparity among the number of methods proposed and the number of studies evaluating these changes using empirical data (Hawkins et al., 2015). Despite the prevalence of review papers of ecological regime shift detection methods (Ducré-Robitaille et al., 2003; Mantua, 2004; Rodionov, 2005; deYoung et al., 2008; Andersen et al., 2009; Boettiger et al., 2013; Kefi et al., 2014; Mac Nally et al., 2014; Dakos et al., 2015; Scheffer et al., 2015; Filatova et al., 2016; Litzow & Hunsicker, 2016; Yin et al., 2017; Clements & Ozgul, 2018; Roberts et al., 2018), there does not currently exist a comprehensive and modern source for these quantitative methods.

Existing reviews of the ecological regime shift methods vary in both the number and detail of the methods presented (Ducré-Robitaille et al., 2003; Mantua, 2004;
Rodionov, 2005a; deYoung et al., 2008; Andersen et al., 2009; Boettiger et al., 2013; Kefi et al., 2014; Mac Nally et al., 2014; Dakos et al., 2015; Scheffer et al., 2015; Filatova et al., 2016; Litzow & Hunsicker, 2016; Yin et al., 2017; Clements & Ozgul, 2018; Roberts et al., 2018]. For example, some reviews critique the current state of the regime shift literature (e.g., Andersen et al., 2009), while others present overviews of a large number of methods with reference to a particular system type, or methodology type (e.g., Roberts et al., 2018). These and other reviews provide useful syntheses of the methods, however, do not emphasize the sheer number and breadth of the methods proposed for and applied to empirical systems data. The rapid growth of this literature and the methods proposed for identifying ecological regime shifts is contributing to the lack of comprehensiveness among existing methods review papers.

Building a comprehensive database of the proposed ecological regime shift detection methods based on a formal literature review is difficult for a few reasons. First, the terminology associated with regime shift theory and detection is highly variable within and among fields (Andersen et al., 2009). For example, the terms, regime shifts, regime changes and tipping points are variably used in studies of ecological systems, whereas inhomogeneities is common in meteorology and climatology, and structural change is largely confined to the study of economics. Although semantics vary both within and across disciplines some methods are shared or are concurrently applicable across fields. Second, papers introducing a new method or approach to identifying regime shifts are not often proposed in publication outlets with aims of disseminating new quantitative methods (e.g., Ecological Modelling, Methods in Ecology and Evolution). Rather, many new methods are published in journals with refined (e.g., Entropy, Progress in Oceanography), as opposed to publications with broader scopes (e.g., Ecology and Nature).

The primary aim of this Chapter is to provide a single and comprehensive source
of the quantitative methods proposed for identifying ecological regime shifts, akin to the most comprehensive method review by Rodionov (2005). I also conduct a bibliographic analysis of the ecological regime shift literature to identify trends in the development and current state of ecological regime shift theory. Finally, I compile a comprehensive list of the proposed regime detection measures present in the ecological regime shift literature.

2.2 Methods

This Chapter comprises two components. First, I compile a comprehensive list of the regime shift detection methods and metrics used in the ecological literature based on expert knowledge and prior reviews. Second, I conduct a bibliographic analysis of the results of a formal review of the ecological regime shift literature to identify potential themes and gaps in this area.

2.2.1 Quantitative methods for identifying ecological regime shifts

I used expert knowledge and previously published review papers to compile a comprehensive list of the quantitative methods proposed as ecological regime shift detection methods. A pilot study which used a systematic literature review failed to identify many of the methods of which I was previously aware, hence the reliance on expert knowledge and previously published reviews. Here, I identify the first instances of each method in the ecological or relevant literature.

The comprehensive list of methods makes exclusions for the following reasons. First, I do not include methods which largely recycle, or make only slight adaptations to, existing methods (Zhou & Shumway, 2008; Salehpour et al., 2011; Byrski & Byrski, 2016). However Nicholls et al. (2011) for an addition of variable optimization to the
method in Nicholls (2011) that was not included in my comprehensive list. Next, the list excludes papers which propose a combination of existing methods (e.g., Kong et al., 2017; Seddon et al., 2014; Vasilakopoulos et al., 2017).

2.2.2 Bibliographic analysis of the ecological regime shift literature

The still-vague definition of ecological regime shifts has led to a breadth of articles exploring this phenomenon. I conducted a formal literature review using the Web of Science database to identify patterns in the development and persistence of the ecological regime shift literature. I conduct a systematic literature review using ISI Web of Science, and use these results to conduct exploratory bibliographic analyses. The search was designed in an attempt to capture the development and breadth of the field of ecological regime shifts. I used the below-mentioned search boolean to identify articles related to regime shift and abrupt changes in ecological systems, restricting the search to the Web of Science ‘categories’ (‘WC’) Ecology and Biodiversity Conservation:

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”) AND WC=(“Ecology” OR “Biodiversity Conservation”)

I identified patterns and trends in the articles resulting from the above mentioned search using a bibliographic analysis (using R Package bibliometrix; Aria & Cuccurullo, 2017). This package contains function wrappers for conducting and visualizing network analyses based on keyword, authorship, and citation data. In an attempt
Figure 21: Number of articles containing phrases related to ecological 'regime shifts' published per publication outlet.

to understand the evolution of regime shift theory and relate this evolution to the quantitative methods in the ecological literature, I focus analyses on using keyword and concept themes rather than citation counts and author dominance.
Figure 22: Number of publications by year in fields 'Ecology' and 'Biodiversity Conservation' which included terms related to 'regime shift' (total = 654).
Table 21: List of the regime detection methods identified using expert knowledge and previously published reviews of methods.

<table>
<thead>
<tr>
<th>Method</th>
<th>Metric type</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average standard deviates</td>
<td>metric</td>
<td>Ebbesmeyer (1991)</td>
</tr>
<tr>
<td>BDS test</td>
<td>metric</td>
<td>Carpenter and Brock (2011)</td>
</tr>
<tr>
<td>Conditional heteroskedasticity</td>
<td>metric</td>
<td>Seekell et al. (2011)</td>
</tr>
<tr>
<td>Conditional probability analysis</td>
<td>metric</td>
<td>Zurlini et al. (2014)</td>
</tr>
<tr>
<td>Cumulative deviation test (CUSUM)</td>
<td>metric</td>
<td>Buishand (1982)</td>
</tr>
<tr>
<td>Degenerate Fingerprinting</td>
<td>metric</td>
<td>Kleinen et al. (2003)</td>
</tr>
<tr>
<td>Discontinuity analysis</td>
<td>metric</td>
<td>Peterson et al. (1998)</td>
</tr>
<tr>
<td>Downton-Katz test</td>
<td>metric</td>
<td>Karl et al. (1987)</td>
</tr>
<tr>
<td>First-order multivariate autoregressive models (MAR1)</td>
<td>metric</td>
<td>Ives et al. (2003)</td>
</tr>
<tr>
<td>Fisher Information</td>
<td>metric</td>
<td>Fath et al. (2003)</td>
</tr>
<tr>
<td>Intervention Analysis</td>
<td>metric</td>
<td>Francis (1994)</td>
</tr>
<tr>
<td>Inverse of AR(1) coefficient, variance, etc.</td>
<td>metric</td>
<td>Carpenter et al. (2008)</td>
</tr>
</tbody>
</table>
Table 21: List of the regime detection methods identified using expert knowledge and previously published reviews of methods. (continued)

<table>
<thead>
<tr>
<th>Method</th>
<th>Metric type</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kurtosis</td>
<td>metric</td>
<td>Biggs et al. (2009)</td>
</tr>
<tr>
<td>Lanzante method</td>
<td>metric</td>
<td>Lanzante et al. (1996)</td>
</tr>
<tr>
<td>LePage test</td>
<td>metric</td>
<td>Yonetani (1993)</td>
</tr>
<tr>
<td>Mann-Kendall test</td>
<td>metric</td>
<td>Goossens et al. (1987)</td>
</tr>
<tr>
<td>Mann-whitney U-test</td>
<td>metric</td>
<td>Mauget et al. (2003)</td>
</tr>
<tr>
<td>method-fuzzy synthetic evaluation (FSE)</td>
<td>metric</td>
<td>Wang et al. (2011)</td>
</tr>
<tr>
<td>Moving detrended fluctuation analysis (MDFA)</td>
<td>metric</td>
<td>He et al. (2008)</td>
</tr>
<tr>
<td>Nearest-neighbor statistics</td>
<td>metric</td>
<td>Pawlowski et al. (2008)</td>
</tr>
<tr>
<td>Oerleman’s method</td>
<td>metric</td>
<td>Oerlemans (1978)</td>
</tr>
<tr>
<td>Pettitt test</td>
<td>metric</td>
<td>Pettit (1979)</td>
</tr>
<tr>
<td>Probability density function entropy method</td>
<td>metric</td>
<td>Pawlowski et al. (2008)</td>
</tr>
</tbody>
</table>
Table 21: List of the regime detection methods identified using expert knowledge and previously published reviews of methods. (continued)

<table>
<thead>
<tr>
<th>Method</th>
<th>Metric type</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quickest detection method (Shiryaev-Roberts statistic)</td>
<td>metric</td>
<td>Moustakides et al. (2009)</td>
</tr>
<tr>
<td>Rodionov method</td>
<td>metric</td>
<td>Rodionov (2005)</td>
</tr>
<tr>
<td>Sequential t-tests</td>
<td>metric</td>
<td>Rodionov (2004)</td>
</tr>
<tr>
<td>Skewness</td>
<td>metric</td>
<td>Guttal et al. (2008)</td>
</tr>
<tr>
<td>Spectral density ratio indicator</td>
<td>metric</td>
<td>Biggs et al. (2009)</td>
</tr>
<tr>
<td>Stability Index of the Ecological Units</td>
<td>metric</td>
<td>Parparov et al. (2017)</td>
</tr>
<tr>
<td>Standard deviation (rising variance)</td>
<td>metric</td>
<td>Carepenter et al. (2006)</td>
</tr>
<tr>
<td>Standard normal homogeneity</td>
<td>metric</td>
<td>Alexandersson et al. (1986)</td>
</tr>
<tr>
<td>STARS</td>
<td>metric</td>
<td>Buishand (1982)</td>
</tr>
<tr>
<td>T-test</td>
<td>metric</td>
<td>Ducre (2003)</td>
</tr>
<tr>
<td>Threshold Indicator Taxa ANalysis (TITAN)</td>
<td>metric</td>
<td>Baker et al. (2010)</td>
</tr>
<tr>
<td>Variance Index</td>
<td>metric</td>
<td>Brock et al. et al. (2006)</td>
</tr>
</tbody>
</table>
Table 21: List of the regime detection methods identified using expert knowledge and previously published reviews of methods. (continued)

<table>
<thead>
<tr>
<th>Method</th>
<th>Metric type</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wilcoxon rank-sum</td>
<td>metric</td>
<td>Karl et al. (1987)</td>
</tr>
<tr>
<td>Zonal thresholding</td>
<td>metric</td>
<td>Yin et al. (2017)</td>
</tr>
<tr>
<td>two-phase regression</td>
<td>metric of a model</td>
<td>Easterling (1995)</td>
</tr>
<tr>
<td>Bayesian approaches</td>
<td>model</td>
<td>Jo et al. (2016)</td>
</tr>
<tr>
<td>Convex model</td>
<td>model</td>
<td>Qi et al. (2016)</td>
</tr>
<tr>
<td>Generalized model</td>
<td>model</td>
<td>Lade (2012)</td>
</tr>
<tr>
<td>Multivariable autoregressive models (MAR1)</td>
<td>model</td>
<td>Ives et al. (2012)</td>
</tr>
<tr>
<td>Nonparametric drift-diffusion-jump model</td>
<td>model</td>
<td>Carpenter and Brock (2011)</td>
</tr>
<tr>
<td>Pettitt test and the Sen test</td>
<td>model</td>
<td>Vicent et al. (1998)</td>
</tr>
<tr>
<td>Potential analysis</td>
<td>model</td>
<td>Ives et al. (2012)</td>
</tr>
<tr>
<td>Method</td>
<td>Metric type</td>
<td>Source</td>
</tr>
<tr>
<td>---------------------------------------------</td>
<td>-------------</td>
<td>-------------------------------</td>
</tr>
<tr>
<td>Regression-based models</td>
<td>model</td>
<td>Solow and Beet (1987)</td>
</tr>
<tr>
<td>Self-exciting threshold autoregressive</td>
<td>model</td>
<td>Tong (1990)</td>
</tr>
<tr>
<td>state-space model SETARSS(p)</td>
<td>model</td>
<td>Groger et al. (2011)</td>
</tr>
<tr>
<td>shiftogram</td>
<td>model</td>
<td></td>
</tr>
<tr>
<td>Smooth transition autoregressive model</td>
<td>model</td>
<td>Gal et al. (2010)</td>
</tr>
<tr>
<td>Online dynamic linear modelling + time</td>
<td>models</td>
<td>Parparov et al. (2017)</td>
</tr>
<tr>
<td>varying autoregressive state space models</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fourier Analysis</td>
<td>other</td>
<td>Carpenter et al. (2010)</td>
</tr>
<tr>
<td>Vector-autoregressive method</td>
<td>other</td>
<td>Solow and Beet et al. (2005)</td>
</tr>
<tr>
<td>Wavelet analysis (decomposition)</td>
<td>other</td>
<td>Cazellas et al. (2008)</td>
</tr>
</tbody>
</table>
2.3 Results

2.3.1 Quantitative methods for identifying ecological regime shifts

I identified 63 unique methods proposed for identifying ecological regime shifts (Table 21) based on expert knowledge and previously published review articles.

Figure 23: A thematic map of the clusters (themes) identified among the (a) author-supplied and (b) ISI-supplied keywords for each article using a clustering algorithm.
2.3.2 Bibliographic analysis of ecological regime shift literature

A search of Web of Science for articles in Ecology and Biodiversity Conservation containing phrases related to ‘regime shifts’ yielded 1,636 original articles. These articles were not filtered in any fashion and as such all were considered in the bibliographic analysis. I used the clustering algorithms of the R package bibliometrics to produce thematic maps, which use clustering algorithms to identify clusters (or themes) based on bibliography features (e.g., keywords, authors; Cobo et al., 2011).

2.3.2.1 Thematic mapping using keywords

Two types of keywords exist in the ISI Web of Science bibliography metadata: those provided to the publication by the authors (author-supplied), and those defined by the ISI Web of Science (ISI-supplied). A keyword thematic map suggests these keywords are used very differently within this literature (Figure 23). The clustering algorithm identified fewer clusters (themes) in the ISI-keywords (Figure 23a) than were identified among the author-supplied keywords (Figure 23b; see footnote 2). This pattern is not surprising given the ISI-supplied keywords are restricted to a preset number of keywords, whereas authors can and do provide synonyms, or words not used at all by ISI. The themes identified in the ISI-supplied keyword analysis were relatively consistent as the number of keywords analysed increased, but the themes varied drastically among the author-supplied keywords. For this reason I make inference on only the ISI-supplied keyword cluster analysis. Four major themes were

---

1 Axes represent (x, Callon’s centrality) the degree of interaction, or the contribution of the theme to the research field and (y, Callon’s density) the strength of the network arcs, or the importance of a theme to the field. Clusters appearing in the quadrants (from top-right moving counter-clockwise) represent the following themes: I) motor-themes (important to the field and well-developed); II) basic-themes (well-developed but marginally important to the field); III) emerging and disappearing themes (under-developed and marginal); and IV) specialized themes (important but under-developed). These themes were identified using a clustering algorithm discussed further in Cobo et al. (2011).
Figure 24: The themes identified in the literature based on author-supplied keywords vary as the number of words included in the analysis increased.

identified in the ISI keyword analysis and, interestingly, fell mostly within the quadrants representing the most extreme values: the first and the third quadrants (Figure 25). The themes identified by the ISI-supplied keywords were much larger in scope (e.g., dynamics, ecosystems, climate; Figure 23a) than those identified in the analysis of author-supplied keywords (e.g., eutrophication, trophic cascade; Figure 23b). That is, the themes of ‘regime shifts’ and ‘ecosystem dynamics’ are highly central to and dense within the regime shift literature (Figure 23c-d). This suggests these two
Figure 25: The themes identified in the literature based on the ISI-supplied keywords vary as the number of words included in the analysis increased.

themes are important to the development of the field and are still strongly influencing the evolution of this field. Although the theme ‘dynamics’ appears a central theme in the development and persistence in ecological regime shift theory based on the bibliographic analysis, it is not necessarily reflected in the many case studies of regime shifts in application (Litzow & Hunsicker, 2016). In fact, Litzow & Hunsicker (2016) found that ~50 of case studies actually tested or accounted for non-linear dynamics when applying early warning indicators and other regime shift measures to ecological
Figure 26: Chronological direct citation network suggests the intellectual structure can be mapped to a few papers. This historiograph identifies important works explicitly in chronological, as opposed to absolute, order.

time series.

2.3.2.2 Historiograph, citation, and reviews

A few patterns appear in analyses of the intellectual and chronological structure of the ecological regime shift literature (Figure 26). First, although the concept of stability, thresholds, and multiple stable states in ecological systems first appeared (and was well-received) in the literature in the 1970s (e.g., Holling, 1973; May, 1977), the most highly cited papers in this field appeared primarily in the early and mid 2000s (Scheffer & Carpenter, 2003; Folke et al., 2004; Walker et al., 2004; Nes & Scheffer, 2005; Carpenter & Brock, 2006). The most recent major contributions to
Figure 27: Total number of articles published and corresponding number of citations (for papers published that year). The most highly cited papers to-date are those published in the late 2000s.

The field were conceptual works emphasizing planetary boundaries and tipping points and the impacts of not recognizing these shifts (Hughes et al., 2013). Finally, the “rise” of resilience theory (Folke et al., 2004; Walker et al., 2004), the first efforts of a search for early warning indicators of ecological regime shifts (Carpenter & Brock, 2006) and the critiques soon following these efforts (Andersen et al., 2009; Contamin & Ellison, 2009) are highlighted in the historiograph.

The most influential papers in the field, based solely on number of citations, are
Figure 28: Total number of articles published per year by category as categorized by ISI. Book chapters, proceedings, editorials, and letters are excluded.

those published in the late 2000s (Figure 27), and include the articles which are broad in-scope and are still used today to frame studies in the context of global change, planetary boundaries, and large-scale tipping points (Bennett et al., 2009; Rockström et al., 2009; Smith & Schindler, 2009). Around this time (~2007) is when the number of regime detection measures doubled (Figure 22). Papers equally influential to the development of this field include those corresponding to the observed rapid increase in the number of total publications in ecological regime shift theory (in the early
Numerous reviews of the regime shift literature exist, ranging from conceptual reviews of the state of regime shift theory in ecology and application (e.g., Bestelmeyer et al., 2011; Andersen et al., 2009; Mac Nally et al., 2014), to studies of robustness of early warning indicators under various theoretical and practical conditions [e.g., Dutta et al. (2018); Perretti & Munch (2012); Lindegren et al. (2012); Hastings & Wysham (2010); Figure 28]. Further, comprehensive reviews of the ecological regime shift detection literature are increasingly out-dated. A permanent and open-source database containing information critical to the testing, comparison, and implementation of RDMs may prove useful to the reader who is interested in applying RDMs but is lacking the statistical or mathematical background to do so.

2.4 A synthesis of the methods available for the practical ecologist

Many of the methods identified in this review have yet to be tested on empirical data from more than a single system type (see Table 21). I categorize the regime detection methods as one of either model-free or model-dependent. Model-free and model-dependent methods are those which do and do not require a mechanistic model to describe the system, respectively. Because many of the model-dependent methods are based on auto-regressive modelling approaches, this is highlighted in the model-dependent section.

The early warning indicators that are often referred to as, “traditional early warning indicators” (variance, skewness, autocorrelation at lag-1) are fairly well-reviewed, and have been applied to a variety of conditions (Ditlevsen & Johnsen, 2010; Boettiger & Hastings, 2012; Dakos et al., 2012; Lindegren et al., 2012; Perretti & Munch, 2018).
However, many of these works apply the traditional (and other) early warning indicators to simulated data, with only some (Contamin & Ellison, 2009; Perretti & Munch, 2012; Guttal et al., 2013; Dutta et al., 2018) testing under varying conditions of noise and expected shift types. The utility and robustness of the traditional early warning indicators is not consistent across and within systems, making them of limited utility in situations where the system cannot be reliably mathematically modelled, or where we have limited data (Burthe et al., 2016; Hastings & Wysham, 2010; Perretti & Munch, 2012; see also Chapter 5). Critical reviews and comparative studies of early warning indicators suggest that no early warning indicator is reliable alone, suggesting the user apply a suite, rather than a single method, and that there remain gaps in our understanding of the conditions under which an early warning indicators might fail (deYoung et al., 2008; Kefi et al., 2014; Filatova et al., 2016; Clements & Ozgul, 2018).

Regime detection measures can be classified into one of two major groups which will largely dictate the applicability of an approach to an analyst or type of information. Some RDMs require the use of mechanistic models whereas others fall into the category of model-independent (or model-free). In most situations, the practical ecologist will have insufficient data or a limited understanding of the system with which to parameterize even the simplest mechanistic models. Further, developing an informative defining data-generating model (i.e. system of equations, differential equations) for more than a few variables (i.e. much greater than 2) is often intractable. Following the convention of Dakos et al. (2012a), I classify the 66 regime detection measure identified as a result of this review for detecting as one of model-based or model(metric)-free and generally synthesize the utility of these types of measures to the practical ecologist.
2.4.1 Model-dependent

Model-dependent require a mechanistic (mathematical) representation of the system, models which often carry strict assumptions that are easily violated by empirical systems (Abadi et al., 2010). Model-dependent methods are usefully categorized under two contexts: differentiable systems of equations or auto-regressive. The methods relying on mechanistic models include model descriptions of systems with many, dynamic and interacting components. For example, models are used to reconstruct trophic food webs where prey or predator collapse induces trophic regime shifts in freshwater lake systems (Carpenter et al., 2011).

2.4.2 Model-free

Model-free (or metric-based per Dakos et al., 2012a) methods are those which do not require a mathematical representation of the system. In fact, many require much less knowledge about the system component dynamics and their interactions to calculate a results. The utility of these methods vary with respect to the number of state variables encompassed in the method, and are therefore further categorized as either univariate (using a single dimension) or multivariate (using but not necessarily requiring multiple dimensions). The most widely used model-free univariate RDMs include descriptive statistics of individual system components (i.e. univariate), such as variance, skewness, and mean value (Mantua, 2004; Rodionov & Overland, 2005; Andersen et al., 2009). These univariate methods, often referred to as ‘traditional early-warning indicators’ require only very simple calculations of individual variables, however, their efficacy in empirical systems analysis is controversial. For example, variance (Carpenter & Brock, 2006) and skewness (of a single variable), oft referred to generally as ‘leading indicators’ or ‘early-warning indicators’ in the literature, have
been applied to both theoretical and empirical systems data with varying results.

2.5 Discussion

In this chapter I present a comprehensive list of the regime detection measures proposed for analyzing ecological data. Although multiple reviews of regime detection measures exist, they are either not comprehensive in their survey of the possible methods or are increasingly out of date with respect to the number of methods proposed in the literature (Rodionov, 2005a). Most reviews since the comprehensive list presented in Rodionov (2005a) are not comprehensive, instead focusing on a single aspect of measures that may be useful to a particular audience. For example, Roberts et al. (2018) summarizes methods capable of handling multiple variables, and Dakos et al. (2015) review only methods designed to detect the phenomenon of critical slowing down. The list presented here does not discriminate, and provides an update to the seminal methods paper by Rodionov (2005a). It is important to note that contributions of previous reviews to the understand and scrutiny of regime detection measures in ecology: Mac Nally et al. (2014); Scheffer et al. (2015); Rodionov (2005a); Roberts et al. (2018); Dakos et al. (2015); Mantua (2004); Litzow & Hunsicker (2016); Kefi et al. (2014); Andersen et al. (2009); Boettiger et al. (2013); Dakos et al. (2015); Clements & Ozgul (2018); Filatova et al. (2016); deYoung et al. (2008).

Leading indicators/regime detection measures which analyze only single variables (e.g., variance, autocorrelation at lag-1) are well-tested on both theoretical and empirical data (e.g. Burthe et al., 2016). Among the most widely used RDMs indicators applied to time-series data include an index of variance, moments around the grand mean (skewness and kurtosis), and critical slowing down (Carpenter & Brock, 2006, 2011). Although univariate indicators may provide insight into relatively simple sys-
tems, their reliability as indicators for complex systems is less certain (Bestelmeyer et al., 2011; Perretti & Munch, 2012; Burthe et al., 2016; Sommer et al., 2017; Dutta et al., 2018). Leading indicators can be a reliable warning of impending shift (Carpenter & Brock, 2011). Some methods have been applied to early-warning indicators in whole systems (Carpenter et al., 2011), however, it is uncommon to have enough information to build reliable networks or food webs. Consequently, reliably measuring the ecological system at hand is often realistically (and financially) not possible. To be useful to practitioners it may be necessary to move beyond heuristic methods, to methods which supply statistical significance or probabilities. And although critiques of some RDMs exist, the rate at which they are rigorously tested does not exceed the proliferation of new methods in the literature. For any method to gain credible traction as a pragmatic tool in ecology, studies should address the concerns of these critiques.

Only a handful of methods have been introduced to the mainstream methodological journals in ecology (e.g., *Ecological Modelling, Methods in Ecology and Evolution*). Although many mainstream publications (e.g., *Science, Ecology Letters*) include applications of some of the methods identified in this chapter (Table 21), I argue that celebrity and ‘new and shiny’ (Steel et al., 2013) methods may influence which methodological articles are printed in these popular journals. A critical survey of potential and realized applications of these methods would be useful for highlighting the needs of future research and methodological improvements. Many of the methods presented in Table 21 have either not been applied to empirical data at all, or were tested only once, often but not always in the article introducing or adapting the methodology (Hawkins et al., 2015). Some methods, especially those dubbed ‘early warning indicators’ (variance, auto-regressive model coefficients) have become relatively mainstream in their application to empirical data, despite having been shown
to be less robust in noisy and nonlinear systems (Burthe et al., 2016), in systems exhibiting lag-effects (Guttal et al., 2013), and in systems not exhibiting catastrophic shifts (Dutta et al., 2018). Unlike these early warning indicators, fewer efforts have been made to test robustness under these and more simple conditions (Dutta et al., 2018; c.f. Brock & Carpenter, 2010; Benedetti-Cecchi et al., 2015). In addition to the paucity of studies attempting to understand the limitations of these methods, this review suggests that simply identifying the suite of methods used in ecological regime shift detection may be difficult using traditional review methods. Many of the methods mentioned in this review were not easily identifiable in a pilot, systematic search process in Web of Science and Google Scholar—rather, many methods were those of which I was either previously aware and/or those previously highlighted in the existing reviews (Mantua, 2004; Rodionov, 2005a; deYoung et al., 2008; Andersen et al., 2009; Boettiger et al., 2013; Kefi et al., 2014; Dakos et al., 2015; Scheffer et al., 2015; Filatova et al., 2016; Litzow & Hunsicker, 2016; Clements & Ozgul, 2018; Roberts et al., 2018). To facilitate this process, an online, comprehensive database may prove useful to the practical ecologist.

Hastings & Wysham (2010) aptly point out an important feature of using any methods for identifying regime shifts in empirical system data: first, only have a single history exists for ecological system trajectories and second, most early warning indicators which signal critical slowing down require the system to have a small potential. Although both points are equally important, the latter infers that no ecological system that is complex, or which cannot be accurately modelled using a system of continuous equations, will be predictable via current, early warning signals. This, paired with the failure of early warning signals in some systems gives little hope for relying upon most regime detection measures and early warning signals as indicators of ecological regime shifts. Disregarding stochasticity in ecological models will often
make the model non-differentiable at the point of an abrupt shift (Graham & Tél, 1984; Hastings & Wysham, 2010).

The warnings by Hastings & Wysham (2010) have not yet been heeded—regime detection measures are still being introduced into the literature, often assuming that the system will exhibit critical slowing down prior to the shift, or having been tested on only a single trajectory of an ecological system. Given the increase in introduction and application of regime detection measures, I recommend the following practices be used in this field will aid the accessibility and transparency of the methods: consistent use of fewer methods which require fewer assumptions; persistent collection and maintenance of baseline data (reference data); an open-sourced database of methods and applications; an updated and critical review of the current state of methods in ecology (see Chapter 5) which includes a brief discussion of the methodological advancements and demonstrated failures; rigorous empirical applications of these methods (especially of those only tested on toy and experimental data); and the relationship of the RDMs used in ecology to other fields (computer science, data science, climatology and oceanography). I identify a suite of questions (Table 22) that would be useful in a much-needed modern and critical review of this field.

**Table 22: Potential questions for a comprehensive review of the ecological regime detection metrics literature.**

<table>
<thead>
<tr>
<th>Type</th>
<th>Questions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Methodological</td>
<td>Does the method assume smooth potential?</td>
</tr>
<tr>
<td></td>
<td>Does the regime shift need to be identified <em>a priori</em>?</td>
</tr>
<tr>
<td></td>
<td>What are the major assumptions about the distribution of the original data?</td>
</tr>
</tbody>
</table>
Does the method explicitly assume the system/variable is stationary?

Does the performance of the method change with non-stationarity?

Can the method handle unstructured data?

Can the method handle multiple regime shifts?

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?

Does the system *have* smooth potential?

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?
<table>
<thead>
<tr>
<th>Do predictions translate to other systems?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Can we interpolate data if necessary? If so, what does this mean for inference?</td>
</tr>
<tr>
<td>In which discipline(s) beyond ecology has the method been tested?</td>
</tr>
</tbody>
</table>
Chapter 3

Deconstructing the steps for calculating Fisher Information as a measure of abrupt change in ecological systems

This chapter is currently under review at *Ecological Modelling* with coauthors N.B. Price, A.J. Tyre, D.G. Angeler, T. Eason, D. Twidwell, and C.R. Allen.

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, and fewer are capable of handling high-dimensional, multivariate and noisy data. A variation of Fisher Information has been proposed as a method for detecting changes in the “orderliness” of ecological systems data. Although this method is described and applied in numerous published studies, its calculation and the concepts behind its calculation are not clear. Here, I succinctly describe this calculation using a two-species predator-prey model. Importantly, I demonstrate that the actual equation for calculating Fisher Information metric comprises fewer steps than was previously described, by decoupling the dimensionality-reduction component from the actual Fisher Information calculation component. I hope this work will serve as a reference for those seeking to understand
Fisher Information in the context of ecological systems and regime shifts, and will stimulate further research of the efficacy of these composite regime shift detection metrics.

3.2 Introduction

Changes in the feedback(s) governing ecosystem processes can trigger unexpected and sometimes undesirable responses in environmental conditions (Scheffer et al., 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 (Hughes, 1994; Scheffer et al., 2001; Walther et al., 2002; Parmesan, 2006), and the rapidity with which these changes occur make predictive modeling difficult. Although detecting regime shifts is increasingly difficult as we increase the extent and complexity of the system in question (Jorgensen & Svirezhev, 2004), advances in the collection and analysis of ecological data (La Sorte et al., 2018) may improve our ability to detect impending regime shifts in time for intervention (Jorgensen & Svirezhev, 2004; Groffman et al., 2006; deYoung et al., 2008; Carpenter et al., 2011; Sagarin & Pauchard, 2012; Wolkovich et al., 2014).

Numerous quantitative approaches have been proposed as regime shift detection methods (Mantua, 2004; Rodionov & Overland, 2005; Andersen et al., 2009; Clements & Ozgul, 2016), but few are consistently applied to terrestrial ecological data (deYoung et al., 2008). I broadly classify these methods as either model-based or model-free (Hastings & Wysham, 2010; Boettiger & Hastings, 2012; Dakos et al., 2012a). Model-based methods use mathematical (mechanistic) representations of the system
(Hefley et al., 2013), which often carrying strict assumptions that are easily violated by dynamic systems such as ecosystems (Abadi et al., 2010). Further, model misspecification may yield spurious results (Perretti et al., 2013). Model-free (or metric-based, per Dakos et al., 2012a) regime detection methods require fewer assumptions to implement than do model-based methods, and typically require much less knowledge (if any) about system component interactions. The most widely used model-free methods include both descriptive statistics of individual system components, such as variance, skewness, and mean value (Mantua, 2004; Rodionov & Overland, 2005; Andersen et al., 2009) and composite measures of multiple variables, notably principal components analysis (Petersen et al., 2008; Möllmann et al., 2015), clustering algorithms (Beaugrand, 2004), and the ‘variance index’ (Brock & Carpenter, 2006).

3.2.1 Fisher Information as a Regime Detection Method

A method that has been more recently applied in the analysis of ecological and social-ecological systems is Fisher Information (Cabezas & Fath, 2002; Karunanithi et al., 2008). As a multivariate, model-free method, Fisher Information integrates the information present in the entire data of a system and distills this complexity into a single metric. This allows Fisher Information to capture ecosystem dynamics with higher accuracy than uni-variate-based metrics, which frequently fail to detect regime changes (Burthe et al., 2016). However, despite the potential of this method its mathematical underpinnings – specifically its calculation and the concepts behind its calculation—may not be obvious to those without extensive training in mathematics. In this paper, I address this knowledge gap. I first provide an overview of the method and highlight the need to account for scaling properties, an inherent feature in complex systems. I then succinctly describe the decoupling of the dimensionality-reduction component from the actual Fisher Information calculation component using a two-
species predator-prey model. I finally discuss the results from a theoretical viewpoint and its practical utility for predicting regime shifts, an increasing concern motivated by current rates of fast ecological change.

3.2.2 The Sustainable Regimes Hypothesis

Fisher Information (hereafter, FI; Fisher, 1922) is a model-free, composite measure of any number of variables, and is proposed as an early warning signal for ecological regime shift detection and as a measure of system sustainability (Mayer et al., 2007; Karunanithi et al., 2008; Eason & Cabezas, 2012; Eason et al., 2014). Three definitions of FI in this context exist: (i) a measure of the ability of the data to estimate a parameter, (ii) the amount of information extracted from a set of measurements (Frieden, 1990; Roy Frieden, 1998), and (iii) a measure representing the dynamic order/organization of a system (Cabezas & Fath, 2002). Although definitions (i) and (ii) are widely applied in the statistical and physical sciences, I focus on definition (iii) as it is gaining traction as a tool used in the context of ecological system response to fast environmental change. 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 bi-variate normal distribution, $p(x, y|t)$. As the time since release, $t$, increases, the spread of the distribution, $p(x, y|t)$, disperses because the animals have moved further from the release location. As the animal moves away from the release location, the potential area within which it currently occupies will increase with time. In this example, 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. If we assume constant dispersal, 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 approaches zero (no information). Per the Sustainable Regimes Hypothesis (Cabezas & Fath, 2002), this example system is not in a stable dynamic state over the range of \( t \), since FI decreases with time.

Conversely, if a population following a simple logistic growth model, \( \frac{dN}{dt} = rN(1 - \frac{N}{K}) \), varies around some carrying capacity, \( K \), and the average system parameters (\( r \), \( K \), and their variances \( \sigma_r, \sigma_k \)) are stationary, then the logarithm of the population size should follow a normal distribution, \( N \text{ normal}(\mu, \sigma) \). In this situation, the FI measured over any selected window of time will be relatively constant and, per the Sustainable Regimes Hypothesis, indicates the system is in a stable dynamic state. Further, this Hypothesis posits that a perturbation to \( N \) will also not affect FI so long as the perturbation occurs with a stationary probability distribution and if the perturbation does not change the distributions of \( r \) and \( K \).

### 3.2.3 Fisher Information requires dimension reduction prior to calculation

An important feature of the FI method is that it requires a complete reduction in dimensionality (i.e., from > 1 to 1 system component). For example, a recent application of Fisher Information to empirical data condensed a species pool from 109 species time series into a 1-dimensional time series (Spanbauer et al., 2014). A reduction in dimensionality, i.e. condensing multivariate data into a single metric, of over two orders of magnitude likely involves a large loss of relevant information, raising the questions of what information is preserved during the dimensionality reduction step in calculating FI, what is lost, and whether this is important. Other dimension reduc-
tion techniques, e.g., principal component analysis (PCA) and redundancy analysis (RDA), attempt to preserve the variance of the data, and the number of components scales with the dimensionality of the data (i.e. they are scale explicit). In contrast, by reducing entirely the dimensionality of the data, the FI method does not identify which features of the data are preserved, and the dimensionality does not scale with the dimensionality of the original data.

3.2.4 Aims

The key contribution of this study is that I decouple the dimensionality reduction step of the FI method (Step 1) from the statistical analysis step (Step 2). By isolating the dimensionality reduction step, we can evaluate it based on its own merits and relate it to more well-known and established methods of dimensionality reduction. By isolating the statistical analysis step, one can better understand how Fisher Information is calculated on the single-dimensional data. I believe that this decoupled approach will eliminate some confusion regarding the calculation of FI, allowing interested researchers to readily evaluate the merits of this method. To facilitate our explanation of the method, I reproduce the predator-prey analysis used in (Fath et al., 2003; Mayer et al., 2007), then induce a “regime shift” into the model. 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.
3.3 Methods

3.3.1 Predator-prey model system

Our model system is a two-species predator-prey model (Equation (3.1); Fath et al., 2003; Mayer et al., 2007; Frieden & Gatenby, 2010), hereafter referred to as the “model system”:

\[
\begin{align*}
\frac{dx_1}{dt} &= g_1 x_1 (1 - \frac{x_1}{k}) - \frac{l_{12} x_1 x_2}{1 + \beta x_1} \\
\frac{dx_2}{dt} &= \frac{g_{21} x_1 x_2}{1 + \beta x_1} - m_2 x_2
\end{align*}
\]  

(3.1)

The specified parameters for the model system are \(g_1 = m_2 = 1\), \(l_{12} = g_{12} = 0.01\), \(k = 625\), and \(\beta = 0.005\) (Fath et al., 2003; Mayer et al., 2007; Frieden & Gatenby, 2010). The initial conditions (predator and prey abundances) for the model system were not provided in the original references (Fath et al., 2003; Mayer et al., 2007). I used package deSolve in Program R (version 3.3.2) to solve the model system [Equation (3.1)] finding \(x_1 = 277.781\) and \(x_2 = 174.551\) to provide reasonable results. A complete cycle of this system corresponds to 11.145 time units.

3.3.2 Inducing a Regime Shift

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 \((K)\). However, this study did not address the central question of \textbf{FI behavior during a regime shift}. As an extension of the original study, I simulated a regime shift by modeling an abrupt decline in carrying capacity, \(k\). I assume \(k\) is described by Equation (3.2) where \(k_1\) is the initial carrying capacity, \(k_2\) is the final carrying capacity, \(t_{\text{shift}}\) is the time the regime shift occurred, and \(\alpha\) is the parameter controlling the rate (slope) of the regime shift. The hyperbolic tangent function (see Equation (3.2)) simulates a smooth and continuous change in \(k\)
while still allowing for the regime shift to occur rapidly. I incorporate the change in
\(k\) into our system of differential equations by defining the rate of change in \(k\), \(k'(t)\),
given by (Equation (3.2)). I assume \(k_1 = 800\) and \(k_2 = 625\), values corresponding
to the range of carrying capacities explored by Mayer et al. (2007). I simulated a
time series of 600 time units, introducing a regime change after 200 time units, and
\(\alpha = 0.05\), where \(t^*\) indicates the

\[ k(t) = k_1 - 0.5(k_1 - k_2)(\tanh(\alpha(t - t^*)) + 1) \]

\[ k'(t) = 0.5\alpha(k_1 - k_2)(\tanh(\alpha(t - t^*))^2 + 1) \]  

(3.2)

3.3.3 Decoupling the Steps for Calculating Fisher Information

Two methods exist for calculating Fisher Information (FI) as applied to ecological sys-
tems data to which I refer the “derivatives-based” method (first appearing in Cabezas
& Fath (2002) and the binning method (first appearing in Karunanithi et al. (2008)).
Although the binning method is proposed as an alternative to the derivatives-based
method for handling noisy and sparse data, our decoupling method reveals it may be
an unnecessary method. Therefore, I focus on only the derivatives-based method for
explaining the theoretical basis for the FI method. The general form of FI can be
found in (Fath et al., 2003 Mayer et al., 2007) and I refer the reader to (Cabezas &
Fath, 2002).

\[ \frac{ds}{dt} = \sqrt{\left(\frac{dx_1}{dt}\right)^2 + \left(\frac{dx_2}{dt}\right)^2} \]  

(3.3)

Step 1: Dimensionality Reduction. The key idea of the dimensionality reduction step
is to calculate the Euclidean distance traveled in phase space. In phase space, each
coordinate axis corresponds to a system state variable (e.g., number of predators and number of prey). The state of the model system over time describes a path or trajectory through phase space. The distance traveled represents the cumulative change in state relative to an arbitrary starting point in time. For the model system, the distance traveled in phase space can be obtained by solving the differential equation given by Equation (3.3)

The original motivation for the dimensionality reduction step is that, under restrictive conditions, there is a one-to-one mapping between the state of the system \(s\), defined in a multidimensional phase space, and the distance traveled, a one-dimensional summary (Cabezas & Fath, 2002). To relate this abstract idea to a more familiar situation, we draw an analogy between the path traced by the system in phase space and the path of a car over the course of a trip. The distance traveled by the car over time is related to the position of the car. Given the route of the car, we could determine the location of the car at any point in time if we know how far it has traveled. However, the distance traveled provides no information about the proximity of locations (i.e., system states). For example, two points in phase space may be arbitrarily close, but the distance traveled would be different if these system states occur at different points in time. Moreover, if the system revisits the same state twice then the one-to-one mapping breaks down and a single state maps to potentially very different values of distance traveled.

What is preserved in the calculation of distance traveled is the rate of change of the system (e.g., the speed and acceleration of the car). The rate of change of the system is the first derivative of the distance traveled in phase space. This is an important point because the concept of a “regime shift” is often associated with the idea of a sudden change in system state. Therefore, it may not be unreasonable to employ a dimensionality reduction procedure that preserves these system dynamics.
Step 2: Statistical Analysis. The product of Step 1 is a one-dimensional time series of what I call “distance traveled”, $s$, (in phase space). The variable $s$ is referred to as “Fisher variable $s$” and “represent[s] a particular state of phase space” in the FI literature (Mayer et al., 2007). I believe distance traveled ($s$) is more descriptive than “Fisher Variable $s$” and avoids confusing the state of the system, defined in multiple dimensions by the multivariate data, with the one-dimensional summary. Using this measure, we next calculate the probability of observing the system in a particular state by assuming a one-to-one mapping between distance traveled and the system state. That is, we calculate the probability of observing the system at a particular distance, $p(s)$, along the trajectory for some period of time from 0 to $t_{end}$. The time at which we observe the system is assumed to be a random variable, $T_{obs} \sim Uniform(0, t_{end})$. This approach assumes the system is deterministic and is observed without error. However, the observed distance traveled by the system, $s$, is a random variable because it is a function of the random observation time.

\[
I = \int \frac{ds}{p(s)} \left[ \frac{dp(s)}{ds} \right]^2 \tag{3.4}
\]

\[
I = \frac{1}{T} \int_0^T dt \left[ \frac{s''}{s'^4} \right]^2 \tag{3.5}
\]

Importantly, the probability of observing the system at a particular value of $s$ increases if the system is changing slowly at that point in time. That is $p(s)$ is inversely proportional to the system rate of change, $s'$. Mathematically, the distance traveled in phase space, $s$, is a monotonically increasing function of time and we assume it is differentiable. Therefore, the probability density function of the distance traveled is $p(s) = \frac{1}{T} \frac{1}{s'}$, where $s' = \frac{ds}{dt}$ is the speed (or velocity) of $s$, and $T$ is the time interval over which the system was observed ($t_{start}$-$t_{end}$). We note that $p(s)$ is simply a constant
multiplied by the inverse of the speed of the system.

The original motivation for the FI calculation as applied to ecological systems was the hypothesis that “since Fisher Information is a measure of the variation” it is also “an indicator of system order, and thus system sustainability” (Cabezas & Fath, 2002). Equation (3.4) is a general form of FI and Equation (4.4) is the form used in the derivative-based method for FI (see Equation 7.3b and 7.12 in Mayer et al., 2007). To better understand the FI calculation, note that Equation (4.4) is, in part, a measure of the gradient content of the probability density function. As the probability density function becomes flatter, the FI value will decrease. In this way, the FI calculation is closely related to the variance. In fact, the FI value for a normal distribution calculated according to Equation (4.4) is simply one over the variance. It is also important to note that FI is zero for a uniform distribution, as the probability density function is flat. Note also that FI goes approaches $\infty$ if the system is not changing over some period of time (Equation (4.4)).

Results

Distance traveled ($s$), speed ($\frac{ds}{dt}$), and acceleration ($\frac{d^2s}{dt^2}$) capture the dynamics of the model system [Equation (3.1); Figure 31]. I simulated a regime shift in the carrying capacity of this model system, at approximately $t = 200$ (Figure 32). The location of this regime shift with respect to the system trajectory in phase space over the entire simulated time period is shown in (Figure 33). Although a slight change is captured by $s$ (Figure 4) at the location of this regime shift, it is not pronounced. Although the distance traveled, $s$ (Figure 34) changes fairly smoothly around the location of the regime shift, the system exhibits a steep decline in speed $\frac{ds}{dt}$ soon after the induced regime shift (Figure 35).

I calculated FI for the distribution of $s$ over a series of non-overlapping time windows. According to Mayer et al. (2007) the length of the time window should be equal to one system period such that FI is constant for a periodic system, however, the
system periods are not identical before, during, and after the regime shift. Therefore, I performed two separate calculations of FI using window sizes corresponding to the initial (when $t < 200$) and the final ($t > 200$) periods of the system ($\text{winsize} = 13.061$ and 11.135 time units, respectively). Using these window sizes the drop in FI at the regime shift initiation is bigger than the magnitude of the fluctuations preceding it (Figure 36).

3.4 Discussion

Part of the appeal of the FI method of regime shift detection is that it provides a 1-dimensional visual summary of system “orderliness”. However, I have demonstrated that, because only a single measure is required to calculate the Fisher Information, the dimensionality reduction step need not necessarily use the ‘distance traveled’. Rather, for example, one can calculate the Fisher Information of another reduced dimension like the components of an ordination (e.g. using Principal Components Analysis). The rate of change of the system (velocity, $\frac{ds}{dt}$), on which FI method is based, is also a 1-dimensional quantity. In the simple predator-prey example, calculating and plotting FI did not provide a clear benefit over simply plotting the system rate of change directly. I suggest that future research uncouple the dimensionality reduction step and the FI calculation step in order to better illustrate the benefits of the FI method relative to dimensionality reduction alone. In the predator-prey example, I assumed the data was free from observation error. Despite these ideal conditions, the estimated FI had high variation and the results depended on the size of the time window used in the calculation. This issue arises because the period of the cyclic system is changing during the regime shift such that it is difficult to find a single window size that works well for the entire time series. Mayer et al. (2007) describe this as a “confounding
issue” related to “sorting out the FI signal of regime change from that originating from natural cycles” and suggest using a time window that is large enough to include several periods. However, in the absence of a quantitative decision rule defining what changes in FI indicate regime shifts, it is difficult to separate the signal in the FI metric from the noise due to fluctuations in the natural cycles. Further research is needed to define quantitative decision rules for what changes in FI constitute a regime shift.

The example used in this study is unrealistic in that I assume no measurement error and therefore focus on the “derivatives-based” method of calculating FI. However, our analysis also has implications for the “binning” method of calculating FI that was later developed for high-dimension noisy data (Karunanithi et al. (2008)). Rather than attempting to estimate the rate of change of each system component (e.g., hundreds of species) and combining these estimates to get the total system rate of change, I suggest an approach where the dimensionality of the data is first reduced by calculating distance traveled in phase-space and then only a single rate of change is estimated. The advantage of this approach is that for an n-dimensional system it only requires the estimation of one derivative rather than n-derivatives. The drawback to this approach is that noisy observations will likely introduce some bias into the estimate of the system rate of change. Nonetheless, I believe this approach is worth exploring due to its simplicity relative to the “binning” method. The Fisher Information of an n-dimensional system is a vector of unit-less values which can only be compared within a data set (e.g., within a single community time series) and interpreting FI is still largely a qualitative effort (Fath et al., 2003; Mantua, 2004), not unlike most regime detection methods [Ch. 2]. When the FI of a system is increasing, the system is said to be moving toward a more orderly state, and most studies of FI propose that sharp changes in FI, regardless of the direction of the change, may
indicate a regime shift (Cabezas & Fath, 2002; Karunanithi et al., 2008; Spanbauer et al., 2014). Although the aforementioned and numerous other works interpret FI in this context (e.g., Eason et al., 2014; Eason & Cabezas, 2012), I suggest future work which clearly identifies the ecological significance of the Fisher Information metric and its significance within the ecological regime shift paradigm.

3.5 Acknowledgements

I thank H. Cabezas and B. Roy Frieden for early discussions regarding the development of Fisher Information, and T.J. Hefley for comments on an earlier draft.
Figure 31: From top to bottom, distance traveled in phase space, speed tangential to system trajectory, acceleration tangential to system trajectory.
Figure 32: Carrying capacity over time with a regime shift occurring around time 200.
Figure 33: Phase space plot of system trajectories for different values of $k$. Colored point indicates the coordinates corresponding to the shift induced in carrying capacity.
Figure 34: Distance travelled in phase space over time. Dashed vertical line at time 200 indicates location of regime shift.
Figure 35: *Speed of the system (rate of change, velocity) in phase space. Dashed vertical line at time 200 indicates location of regime shift.*
Figure 36: 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.
Chapter 4

An application of Fisher Information to bird community data does not reveal distinct regimes in the continental United States

4.1 Introduction

Ecosystems are open, dynamical systems that are most often not easily described using fully parameterized models. Some patterns have emerged in certain statistical mechanics of ecological observations, despite the complexity of most ecological systems. An uptick in recent years of studies of regime shifts (Table 01) in ecology has spurred an increase in the number of new methods for detecting ecological regime shifts (Chapter 2), some of which have been applied to spatial information (Brock & Carpenter, 2006; Gutta & Jayaprakash, 2009; Kefi et al., 2014; Butitta et al., 2017; Sundstrom et al., 2017). As defined in Table 01, a regime shift is largely considered an abrupt and persistent change in a system’s structure or functioning. Following this definition and without considering the pressures (Table 01) associated with the observed regime shift, it is not yet clear whether identifying a ‘spatial regime’ using a snapshot of a system (i.e. using a single or short period of time relative to the time scale of the system dynamics and/or pressures) is pragmatic. A concise and global
definition of the spatial regime detection measure is important since observations of non-random spatial processes (e.g., land cover) can manifest as either a rapid shift (e.g., an ecotone) or as a gradual change (e.g., slow mixing along a gradient). Consequently, and because most regime detection measures signal abrupt change, only the former may be identified as “regime shifts” using spatial regime detection measures.

Although it is suggested that statistical and pragmatic methods are advanced more rapidly by bottom-up approaches, i.e. using case studies (see DeAngelis & Yurek, 2017), there is much work to be done in the way of testing the statistical rigor of spatial regime detection measures. The objective of this chapter is to determine whether the Fisher Information as a regime detection method [Eq. (4.4)] identifies spatial regime boundaries in the bird communities of the continental United States. This chapter is also supported by original software developed for implementation in Program R, which is publicly available (see Appendix 4).

4.2 Data and Methods

4.2.1 Data: North American breeding bird communities

I used community abundance data (Sauer et al., 2014) from long-term monitoring programs to identify spatial and temporal regimes using the Fisher Information (FI) derivatives method (see Eq. (4.4)). The North American Breeding Bird Survey (NABBS) trains citizen scientist volunteers to annually collect data using a standardized roadside, single observer 3-minute point count protocol and has organized data collection annually across North America (Figure 41) since 1966. The roadside surveys consist of 50 point counts (by sight and sound) along ∼ 24.5 mile stretch of road. Due to strict reliance on volunteers, some routes are not covered every year. Additionally, some routes are moved or discontinued due to changing landscape conditions
Figure 41: Locations of Breeding Bird Survey routes sampled between 1966 and 2017.

and change in observer safety. Route-year combinations that were missing years but were not discontinued were treated as missing data. Although NABBS volunteers attempt to identify all species as possible, persistent biases exist in this protocol. Despite a standardized survey protocol, some species are difficult to detect using these methods. For example, crepuscular species are less likely to be detected beyond the first few points of the BBS route, given they are most active at sunrise and the survey begins within 30 minutes of sunrise. Further, species which congregate in large groups and are highly mobile (e.g., waterfowl) tend to have less reliable inter-annual
abundance estimates given their ability to move long-distances in a short period of time. To remove any potential influence of sampling bias on the Fisher Information result, I removed birds of these types from all analyses: waterfowl, waders, and shore species (BBS AOU numeric codes 0000 through 2880).

4.2.2 Study area

Although the NABBS conducts surveys throughout much of North America (most of the United States, Canada, and Mexico), coverage of the boreal forests of Canada are sparse in space, and many routes in Mexico have fewer than 25 years of observations. For these reasons I limited analyses largely to the continental United States and parts of Southern Canada (see Figure 41).

4.2.2.1 Focal military base

The Mission of the U.S. 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 Figure 43). The spread of these bases (Figure 42), 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 and other natural resource management groups (e.g., state management agencies, non-profit environmental groups).

Much like other actively managed landscapes, military bases are typically surrounded by non-managed lands, or lands not managed specifically for natural resource or ecological biodiversity or conservation. Natural resource managers of military bases
face environmental pressures within and surrounding their properties, yet their pri-
mary 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/ammunition
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

Figure 42: Locations of focal U.S. military bases, Eglin Air
Force Base (AFB) and Fort Riley Military Base.
resource management efforts.

The NABBS routes chosen for analyses in this Chapter lie within or near Fort Riley military base (located at approximately 39.110474°, −96.809677°; Kansas, USA). Fort Riley (Figure 42) 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., 2018) in the area surrounding Fort Riley military base (Van Auken, 2009). This phenomena should present itself as a regime boundary if Fisher Information is a reliable spatial regime detection measure.

4.2.2.2 Spatial sampling grid

Fisher Information has been applied to empirical data as a spatial regime detection measure in recent years (Sundstrom et al., 2017; Eason et al., 2019). The authors of
Sundstrom et al. (2017) used the Fisher Information binning method to demonstrate the utility of this method as an indicator of spatial regime boundaries, suggesting that rapid changes in the resulting value of Fisher Information as calculated for spatially adjacent sites should indicate ‘regime changes’. Sundstrom et al. (2017) identified sampling sites which transected multiple ecoregions, resulting in a transect which zigzagged across a region of the Midwestern United States (Sundstrom et al., 2017). I identified sites using a gridded system across the continental United States and parts of Canada to ameliorate potential bias associated with handpicking NABBS routes. The gridded system comprises East-West running transects, ameliorating potential sampling bias as the transect location and widths were designed to capture large-scale shifts in bird communities at regular intervals. This spatial sampling grid approach also allows for raster stacking, or layering data layers (e.g., vegetation, LIDAR, weather), providing an opportunity to identify potential relationships with abiotic drivers, should regime shifts be observed in the avifauna data. This spatial sampling method also provides a simple vector for visualizing changes in the Fisher Information over space-time. For brevity, I present visual results of only three, spatially-adjacent, East-West running transects (Figure 43) 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, $\theta$. Fisher Information is a measure of indeterminacy (Fisher 1922) and is defined as,

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

(4.1)
where \( p(y|\theta) \) is the probability density of obtaining the data in presence of \( \theta \). 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 (Frieden & Soffer, 1995; 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, one should 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

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

(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):

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

(4.3)

A form specific to the probability density function of distance traveled by the entire
The system, which I call the ‘derivatives’ method, is defined as (Mayer et al., 2007, eq. 7.12):

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

(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 requires multiple parameters to be defined \textit{a priori}, which have been shown to influence inference based on the metric. I therefore calculated FI using the derivatives equation [see Chapter 3].

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 & 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 multivariate 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 & Fath, 2002; Fath et al., 2003) to calculate the Fisher Information measure (see Chapters 3, 6 for detailed discussions on this topic). Briefly, 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 (Fath et al., 2003). Here, I use
the derivatives method (Eq. (4.4)) to calculate Fisher Information for all East-West transects (see Figure 43) at decadal intervals (years 1980, 1990, 2000, and 2010). Justification for using this method is provided in detail in Chapter 3.

4.2.4 Interpreting and comparing Fisher Information across spatial transects

4.2.4.1 Interpreting Fisher Information values

Interpretation of FI, like the interpretation of numerous other regime detection measures, is currently a qualitative effort. Fisher Information is proposed as an indicator of system orderliness, where periods of relatively high values of FI indicate the system is in an “orderly” state, possibly fluctuating around a single attractor. A rapid change in FI is proposed as an indicator of a change in a system’s orderliness, suggesting a potential reorganization phase. Whether Fisher Information can identify a switch among basins of attraction within a single, stable state 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 \( \frac{dq(s)}{ds} \to \infty \), we infer the system is approaching a stable state, and as \( \frac{dq(s)}{ds} \to 0 \) the system is showing no preference for a single stable state and is on an unpredictable trajectory. Eq. (4.3) bounds the potential values of Fisher Information at \([0, 8]\), whereas Eq. (4.1), Eq. (4.2), and Eq. (4.4) 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 a single datum is a spatial transect). It follows that Fisher Information should be interpreted relatively, but not absolutely.

Here I define a potential regime change as a point(s) having a non-zero deriva-
tive, and at which relatively large changes (manifested as either a sharp increase or decrease) in FI occurs. Regime shifts are identified as data changing from one state to another, thus, rapid shifts in the value of FI should indicate the locations of these shifts in the time and 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 (Sundstrom et al., 2017) within a single time period, whereas temporal analysis identifies the point in time at which the system undergoes a regime shift. I follow published recommendations for interpreting the Fisher Information results in the context of identifying regime shifts (e.g., Karunanithi et al., 2008; Fath et al., 2003; Eason & Cabezas, 2012).

4.2.4.2 Interpolating results across spatial transects

NABBS are not regularly spaced, and pairwise correlations of adjacent transects (see Figure 43) is not possible without either (1) binning the Fisher Information calculations using a moving-window analysis, or (2) interpolating the results to regularly-spaced positions in space. To avoid potential biases associated with the former option (i.e. choosing window size, location of data aggregation), I linearly interpolated the calculated Fisher Information within each spatial transect to 50, evenly-spaced points along the longitudinal dimension. The 50 longitudinal points to which I interpolated were the same across each spatial transect, while latitude varied across transects. I used the function stats::approx() (with argument rule=1) to linearly approximate the Fisher Information. I did not interpolate values beyond the longitudinal range of the original data (i.e., no extrapolation).
4.2.4.3 Spatial correlation of Fisher Information

If Fisher Information captures and reduces information regarding abrupt changes in community structure across the landscape, then it follows that the values of Fisher Information should be spatially auto correlated. That is, the correlation of FI values should increase as the distance between points, both within and among transects, decreases. Further, direct comparison of FI across routes is not possible since FI (Eq. 4.4) is a relative value with no upper limit (i.e. can take on any value between 0 and $\infty$). In other words, FI values calculated are not relatively comparable outside of a single spatial transect (Figure 43). Fisher Information is, however, directly comparable within each spatial transect (e.g., 43). For these reasons, we can identify spatial regime shifts both within and among spatial transects by using pairwise correlations among two transects (e.g., 43) to determine whether values of FI are consistent across space. Here, I calculate the pairwise correlation (Pearson’s) among each pair of adjacent spatial transects (e.g., Figure 43). I removed a pair of points if at least one point was missing an estimate for Fisher Information. This occurs when the original longitudinal range of one transect exceeded the range of the adjacent pair.

4.3 Results

4.3.1 Fisher Information across spatial transects

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’ (Figure 44). Of the three spatial transects analyzed in this chapter (see Figure 43), Figure 44 is representative of the lack of pattern observed in the Fisher Information values across all analyzed transects. I did not identify patterns of spatially contagious abrupt
changes in the Fisher Information values within or among spatial transects.
Figure 45: Fisher Information of two spatially adjacent transect pairs (transects 12, 13) over time. Interdecadal trends in FI are very different within each transect and are not highly correlated among transects over time.
Figure 46: Fisher Information of select East-West spatial transects over time. These transects are representative of the results obtained from the remaining transects.
Contrary to expectations there appear no clear patterns of abrupt change in the Fisher Information results along three selected East-West running transects (Year 2010). The results are visualized here at regularly-spaced intervals such that the locations are not precise relative to the location of corresponding Breeding Bird Survey routes.

Figure 47: Contrary to expectations there appear no clear patterns of abrupt change in the Fisher Information results along three selected East-West running transects (Year 2010). The results are visualized here at regularly-spaced intervals such that the locations are not precise relative to the location of corresponding Breeding Bird Survey routes.

4.3.2 Spatial correlation of Fisher Information

This study did not identify spatial correlation of the Fisher Information results among most of the spatially adjacent transects (Figures 48 and 49). For spatially-adjacent transects (e.g., transects 11 and 12, or 12 and 13 in Figures 48 and 49), we should
Figure 48: Pairwise relationships of Fisher Information (interpolated values) of spatially adjacent transects over time do not exhibit expected patterns of high positive correlation. 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.

I expect high and positive correlation values, and these values should 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 (Figure 43), as the spatial soft-boundaries limiting the distribution and functional potential of avian communities are largely North-South (Figure 43).
Figure 49: Pairwise relationships of Fisher Information (interpolated values on the log-scale) of spatially adjacent transects over time. 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.

Note spatial transects in Figure 43 overlap multiple, large spatial ecoregion boundaries, such that we should expect our data to identify these points (boundaries).

Upon initial investigation, there are no consistent signs of broad-scale patterns in FI across space (Figure [47]1). If Fisher Information is an indicator of spatial regime

1Here, shape size indicates the relative value of the scaled and centered Fisher Information results. Red box (top panel) indicates the extent of the results presented in the bottom panel.
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 FI values and functional potential (using Omernick Ecoregion Level 2; see Figure 47). However, in

Figure 410: Fisher Information results (scaled and centered; larger circles (points) represent higher values of Fisher Information) plotted against ecoregion boundaries (EPA Level 2 boundaries) for the year 2000.
years 2000 and 2010 the transects 11 and 12, and 12 and 13 were highly positively correlated (Figure 49). Rather than abrupt changes in Fisher Information I found gradual changes (e.g., see results for years 2000 and 2010 in Figs. 47 and 410.

4.4 Discussion

The Fisher Information measure was introduced as a method to avoid analytical issues related to complex and noisy ecological data (Fath et al., 2003; Karunanithi et al., 2008) and was recently suggested as an indicator of spatial regimes (Sundstrom et al., 2017; Eason et al., 2019). Contrary to expectations, I did not consistent abrupt changes in the Fisher Information metric (Eq. (4.4)), which would indicate a regime shift of sorts, in the avian communities. Further, there was an absence of autocorrelation among the spatially adjacent transects in my study area, suggesting that the Fisher Information may not be a suitable metric for identifying abrupt changes in bird communities at this scale.

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 empirical data. Fisher Information is capable of handling an infinite number of inputs (variables) and, given sufficiently low window size parameters, can technically calculate an index value for only two observations. It is important that the user understands the assumptions of identifying regime shifts or abrupt changes when using this method, as rigorous testing of its efficacy is necessary (but see Chapter 5). The sampling design of the North American Breeding Bird Survey data in this Chapter was designed to avoid subjective decisions present in a previous application (Sundstrom et al., 2017).

There are three primary assumptions required when using Fisher Information
to estimate relative orderliness within ecological data (Mayer et al., 2007): (i) the order or state(s) \((s)\) of the system is observable; (ii) 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, (iii) changes in \(I\) presumed to be regime shifts do not represent the peaks of cyclic (periodic) patterns. Assumption (i) is one of philosophical debate and is thus not controllable. To attempt to control for false negatives or false positives that may result from violating assumption (ii), the user of this metric should take care in their selection of state variables. In the case of a high dimensional data, relativization of state variables and/or a state variable reduction technique may be useful. 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 [assumption (iii)] in the data by ensuring integration periods capture at one full cycle of the system and, given sufficiently high number of observations, increasing the integration period may also alleviate some issues related to irreducible error, or white noise.

The lack of patterns identified using Fisher Information may be influenced by 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 these data to identify ‘spatial regime shifts’. 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 and 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 re-
moves the dimensionality of these systems, which may omit information or signals
integral to understanding the ecological processes at play.

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, while collapsing a wealth of data into a single metric,
does not allow the user to relate the resulting value to the original data, unlike other
dimension reduction techniques. For example, the loadings, or the relative influence of
variables on the ordinate axes, can be derived from a Principal Components Analysis–
this cannot be achieved using Fisher Information. If Fisher Information clearly sug-
gested a spatial regime boundary or shift, a before-and-after post-hoc analysis of the
regional community dynamics might confirm the regime shift occurrence.

A rapid change in either direction (increase or decrease) of the Fisher Information
value is proposed as an indicator of ecological regime shifts, or a change in the or-
derliness of a system (Mayer et al., 2006; Eason & Cabezas, 2012). After calculating
the Fisher Information for each spatial transect (Figure 43) during each sampling
year in this study, I used pairwise correlation to determine whether spatial autocor-
relation 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 eco-
tone, high altitude rock formations), then the Fisher Information 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 between the transects differs and/or a functional or physical boundary separates them. Contrary to these expectations, I did not find evidence of such abrupt changes within nor across the East-West running spatial transects. Several questions remain regarding the application of regime shift detection methods to spatially-explicit data. If signals of regime shifts do exist, the results of this study suggest the Fisher Information metric may not be ideal for identifying them. This study provided an objective evaluation of the Fisher Information metric as a spatial regime detection measure. Future work on the following areas may improve our understanding of if and how Fisher Information may provide insights of ecological regime shifts in spatial and/or temporal data:

1. Sensitivity of Fisher Information to data quality and quantity (this is explored in Chapter 5).
2. What, if any, advantages does FI have over other density estimation techniques?
3. Does FI provide signals in addition to or different than geophysical and vegetative (e.g. LIDAR) observations (data)?
4. Relationship of Fisher Information to likelihood ratio-based unsupervised changepoint detection algorithms (e.g., ChangeFinder; Liu et al., 2013).
5. How does Fisher Information perform relative to other regime detection measures (see Chapter 5)?
Chapter 5

Using Resampling Methods to Evaluate the Relative Performance of Regime Detection Measures

5.1 Introduction

Ecological systems have many unpredictable and variably interacting components. 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 unfeasible in ecosystem research and management (Clements & Ozgul, 2016).

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 literature, 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 system agnostic) 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 & Ozgul, 2018; c.f. Batt et al., 2013). Efforts to develop and/or improve regime detection measures that do not require such subjectivity 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 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 (Anderson & Piatt, 1999; Alheit et al., 2005; deYoung et al., 2008), climatic, marine, and Paleolithic regime shifts (Yang & Wu, 2006; Spanbauer et al., 2014; 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, which are often sparse, noisy, and characterized by irregular time intervals.

Ecological systems data are expensive to capture, often exhibiting large process
variation and observation errors. This variability reduces data quality and quantity, limiting the numerical tools for identifying trends and changes in the system (Thrush et al., 2009). Some methods, new and old, proposed as regime detection measures are purported to handle the data limitation and quality issues inherent in ecological data, and minimize subjective decisions for choosing state variables and interpreting results. For example, variable reduction techniques, e.g. principal components analysis (Rodionov & Overland, 2005; Andersen et al., 2009; Reid et al., 2016), clustering algorithms (Weijerman et al., 2005; Weissmann & Shnerb, 2016), an index of variance (Brock & Carpenter, 2006), and Fisher Information (Cabezas & Fath, 2002; Fath & 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 used 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 multivariate regime detection measures. There are three 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 above mentioned methods under multiple scenarios.

This Chapter provides baseline relative performance estimates of select, multivariate 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 accompa-
nying this Chapter allows the end user to implement these methods on this or their own system data, or on theoretical data.

**Figure 51:** Relative abundances of the diatom species in Foy Lake over the time period.
Figure 52: The amount of time elapsed between observations for the Foy Lake paleodiatom data.

5.2 Data and Methodology

5.2.1 Study system and data

I used paleodiatom time series from a freshwater system in North America (Foy Lake, present day Montana) that apparently underwent rapid shifts in algal community dynamics at multiple points in time. This data comes from a single soil core sample, from which the relative abundances of 109 diatom species were identified at 768 observations (time points) over ≈ 7,000 years (Figure 51). Although 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 52). The data were published in Spanbauer et al. (2014) and can be downloaded at the publisher’s website.

5.2.2 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 multivariate data. Here, I compare the results for three regime detection metrics that are model-free and can handle multivariate data: velocity (Chapter 6), the Variance Index (Brock & Carpenter, 2006) and Fisher Information (Fath et al., 2003). I chose the Variance Index, as this is one of the more widely applied multivariate, model-free regime detection measures, and has been shown to, in some empirical data, identify regime shifts post hoc. I introduced the velocity in Chapter 6 as a new, potential regime detection metric. As this is the first time it has been used for such a purpose, including it in this approach allows us to further identify potential flaws with the method, but also to gain some baseline estimates of its relative performance. In Chapter 3, I presented the Fisher Information metric as it is used in detecting ecological regime shifts, and discuss the situations under which it may or may not be a good metric.

5.2.2.1 Velocity (v) calculation

In Chapter 6 I describe a new method, velocity, v, as a potential dimension reduction and regime detection method. First introduced by Fath et al. (2003) as one of multiple steps in calculating their variant of Fisher Information, velocity calculates the cumulative sum of the mean root square change in all state variables over a period of time (Eq. (5.1)). Steps for calculating this metric are described in detail in
\[ \Delta s_i = \sqrt{\sum_{j=1}^{n}(x_{i,j} - x_{i-1,j})^2} s_k = \sum_{i=2}^{k} \Delta s_i 2 \leq k \leq n v = \frac{\Delta s}{\Delta t} \] (5.1)

5.2.2.2 Variance Index (VI) calculation

The Variance Index was first introduced by Brock & Carpenter (2006), and can be simple defined as the maximum eigenvalue of the covariance matrix of the system within some period, or window, of time. The Variance Index (also called Variance Indicator) was originally applied to a modelled system (Brock & Carpenter, 2006) and has since been applied to empirical systems data (Spanbauer et al., 2014; Sundstrom et al., 2017). Although rising variance has been shown to manifest prior to abrupt shifts in some empirical systems data (Nes & Scheffer, 2005; Brock & Carpenter, 2006), the Variance Index, which is intended for multivariate data, appears most useful when the system exhibits a discontinuous (non-linear) shift (Brock & Carpenter, 2006).

5.2.2.3 Fisher Information (FI) calculation

Fisher Information \((I)\) is essentially 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 3) of the distance traveled by the system, \(s\) over some period of time \((T)\), and is given in Eq. (5.2):

\[ I = \frac{1}{T} \int_0^T dt \left( \frac{s'^{2}}{s'^{4}} \right)^2 \] (5.2)

I refer the reader to Chapter 3 for a complete description and to Cabezas & Fath (2002) for a complete derivation of Fisher Information.
Figure 53: Density plot of the coefficient of variation (CV) as a percentage (%) of the Variance Index resampled values over 10,000 iterations. Densities are drawn based on all values of CV but values greater than 100% are not printed.

5.2.2.4 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 condition: $2 \leq T < (T^* - 1)$. If $T = T^* - 1$, then only a single value of the metric 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
Figure 54: Mean Variance Index (VI) and associated 95% confidence intervals over 10,000 iterations using the observations resampling method. Red line indicates the value of VI when $M$ and $P$ are 100%.

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 calculations 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 number of time points by which the window advances. In order to maximize the number of points at which results were obtained, I advanced the moving window at a rate of one time unit (rather than skipping observations). 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 calculated
values of Fisher Information and Variance Index within each moving window to the end (the last time unit) of the moving window. In 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 $T^*$.

Figure 55: Mean Variance Index (VI) and associated 95% confidence intervals over 10,000 iterations using the species resampling method. Red line indicates the value of VI when $M$ and $P = 100\%$.

5.2.3 Simulating data quality and quantity issues using resampling techniques

Using a resampling approach I calculated the regime detection measures over different scenarios simulating data quality and data quantity issues common to ecological data analysis. The scenarios 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 in-
formation about the system by removing some proportion of the species. The loss of temporal observations and the loss of species were examined at three proportions: 

\[ P = [0.25, 0.50, 0.75, 1.00] \]

where \( P \) is the proportion of species and time points retained for analysis. For example, when \( P = 0.25 \), a random selection of 25% of the species are retained for analysis in the species scenario. I resampled the data over 10,000 iterations (\( N_{\text{samp}} \)) for each scenario and \( P \) combination. Note that because when \( P = 1.00 \), all data are retained. Therefore, no resampling was conducted at this level because only a single metric (e.g. Velocity) value is possible.

![Figure 56: Mean velocity and associated 95% confidence intervals over 10,000 iterations using the observations resampling method. Red line indicates the value of velocity when \( M \) and \( P \) are 100%.

5.2.4 Comparing regime detection measures

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). It is important to not only identify the influence of data quality and quantity on the performance of individual regime detection metrics, but also to somehow relate these qualities. I visually inspect the relative performance of these metrics by comparing the coefficient of variation of the re-sampled samples for the results of resampling method (M; species, observations) and sampling percentage (P; 25%, 50%, 75%) combination for each metric (FI, VI, v). The coefficient of variation measures provides a relative measure of the variability in the estimated metric across re-sampled samples as $100\frac{\sigma}{\mu}$, where $\sigma$ is the standard deviation and $\mu$ is the mean value. I observed the distributions of the CV to identify potential flaws in the metrics should data quality or quantity (M, P) decrease. First, within a value of P a low error to mean ratio (CV) indicates that the metric value is similar across the re-sampled samples ($N_{samp} = 10,000$). The efficacy of the metric should be questioned as $CV \to 1$, and perhaps even abandoned as $CV \gg 1$. Next, we can examine how the distribution of CV changes within M and across P. As we increase P, we are increasing the volume of data we are feeding the metric. Intuitively, we can assume that as we add more data (volume), we are supplying the metric with more information, theoretically increasing the signal-to-noise ratio. Following this logic, we should expect the distribution of CV to generally decrease in mean CV value and also become less variable (less dispersion around the mean CV). A visual examination of the distribution of CV across P and within M was sufficient to achieve inference regarding the quality of these metrics upon loss of observations and species.
Figure 57: Mean velocity and associated 95% confidence intervals over 10,000 iterations using the observations resampling method for a subset of the time series (the second ‘regime’ identified). Red line indicates the value of velocity when $M$ and $P = 100\%$.

5.3 Results

5.3.1 Velocity of the distance travelled ($v$)

The velocity of the distance traveled, $\frac{ds}{dt}$ or $v$, exhibited dispersion across the values of $P$, however, yielded consistent results (i.e., high overlap in the densities of the CV across values of $P$ and across methodologies; e.g. Figures 59 and 53). Further, it should be noted that because $v$ is calculated using first differences, it will be sensitive to large changes in the state variables. By examining the density plot of the CV of the distance traveled, $s$, we notice that this measure is highly insensitive to data loss (Figure 58), suggesting that a finite differencing approach (e.g., using total variation regularized differentiation; see Chapter ) which can yield a much smoother derivative than the approach used here, may decrease the sensitivity of $v$ to data loss. This hypothesis is further supported when examining the effect of species (Figure 56) and
temporal observation loss (Figure 57) on the velocity metric. These conditions are representative of the other P-M combinations.

5.3.2 Variance Index

The Variance Index (VI) performed best under the observations resampling method, exhibiting low values for and low dispersion in the CV density (Figure 53) across iterations. However, the VI appears sensitive to high losses of species information, where the density of the CV still exhibits low dispersion but with higher overall mean values (Figure 53). Surprisingly, the Variance Index was insensitive to temporal observation loss (Figure 54), exhibiting a similar amount of noise across various degrees of data loss (P). Although the signal was dampened under the species method, the signals
for the shifts in community composition were not lost across levels of $P$ (Figure 55). This is likely due to the high probability that the dominant species were rarely always excluded from the re-sampled observations.

![Figure 59: Density plot of the coefficient of variation (CV) as a percentage (%) of the Fisher Information resampled samples (10,000 iterations). Densities are drawn based on all values of CV, but values >100% are not printed.](image)

**Figure 59:** Density plot of the coefficient of variation (CV) as a percentage (%) of the Fisher Information resampled samples (10,000 iterations). Densities are drawn based on all values of CV, but values >100% are not printed.

### 5.3.3 Fisher Information is highly sensitive to information loss

The Fisher Information method did not yield conclusive results regarding the abrupt shifts in the paleodiatom community composition. Further, this method appears highly sensitive to varying quality and quantities of data (Figures 510, 59). Although the Fisher Information identifies the shift in community composition at $\sim 1,300$ years before present, it fails to identify shifts outside this period. Further, it is difficult to visually analyze any value of the Fisher Information on the original scale as the values...
range from $\approx 0$ to $10^{15}$ (Figure 510). In addition to failing to identify the shifts in community composition, the standard deviation of Fisher Information far exceeded the mean value of Fisher Information under all M-P scenarios (Figure 59). When I re-sampled the data using 25% and 50% of the species the ratio of mean Fisher Information to standard deviation (CV) of Fisher Information is always $\gg 1$ (i.e., not pictured in Figure 59). The high variation in FI values across re-sampled iterations coupled with the high dispersion within each M-P combination (Figure 59) suggests Fisher Information will not produce similar trends when we lose or distort the data collected. This is also suggested by the high confidence intervals surrounding each M-P combination (Figure 510).

5.4 Detrending the Data Prior to Calculations

If and how to manipulate the original data prior to calculating various regime detection methods is an important consideration, and a line of research that has not yet been fully explored. Although most of the multivariate methods identified in the literature review do not require data that conforms to a specific distribution, how the results of these methods can vary as we change the quality and characteristics of the original data (Michener & Jones, 2012). In fact, since many of the methods for regime shift detection are specifically looking for changes in variance structure and autocorrelation, standardizing variances is not counter-intuitive.

Some studies detrend the original time series prior to data aggregation and calculation of regime detection metrics. I did not detrend the original data for two reasons. First, the authors of the original paper analyzing this data set (Spanbauer et al., 2014) did not detrend species time series. Like Spanbauer et al. (2014) I only scaled the original data, rather than detrending. Second, detrending a time series
requires yet another subjective decision by the data analyst. For example, a “span-
ning” parameter must be chosen when detrending (smoothing) non-linear time series
using local regression (Loess) regression (see Figure 511). Other smoothing methods
are being explored for both detrending (e.g., PcR; Beck et al., 2018) and regime shift
identification (e.g., generalized additive modelling; Beck et al., 2018). Finally, this
data exhibits rapid and drastic shifts in community composition and contains a dis-
proportionate amount of dominant versus non-dominant species. Consequently, most
species contain more zero than non-zero observations, which makes loess smoothing
difficult. Future work studying the impact of detrending, data scaling, outlier re-
moval, and other related decisions would be of value in understanding the efficacy of
these and other regime detection measures in real-world situations.

5.5 Conclusion

In this chapter I provide additional evidence for the sensitivity of select regime de-
tection measures to information (data) quality and quantity loss. The loss of data
quantity was simulated by randomly sampling subsets of both the species and the
temporal observations, and the reduction in data quality manifests as a function of
removing whole species from the community profile. Previous studies of the robust-
ness of uni-variate regime detection metrics have found similar results, suggesting
the measures fail in numerous real-world ecological conditions (Andersen et al., 2009
Contamin & Ellison, 2009). This chapter also highlights the relative insensitivity
of the new velocity metric (see Chapters 3, 6) to data and information quality and
quantity (e.g., Figure 56) loss.
5.6 Acknowledgements

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 their advisement of this project and IIASA scientists Drs. Matthias Jonas, Chai Molina, Piotr Zebrowski for feedback on study design.
Figure 510: Mean Fisher Information (FI; note the scale) and associated 95% confidence intervals over 10,000 iterations using the species resampling method. Red line indicates the value of FI when $M$ and $P$ are 100%. A very small value was added to the mean FI prior to log transformation.
Figure 511: Local regression (loess) smoothing of a dominant species in the paleodiatom community, Anomoeoneis costata varies with the span parameter, making it difficult to justify smoothing the data prior to calculating various regime detection metrics.
Chapter 6

Velocity ($v$): Rate of Change of System Trajectory Identifies Abrupt Changes in Ecological Systems

6.1 Introduction

When, how and why ecological systems exhibit abrupt changes is a hallmark of modern ecological research, and changes that are unexpected and undesirable can have undesirable downstream consequences on, e.g., ecosystem services, biodiversity, and human well-being. Quantitatively detecting and forecasting these changes, however, has yet to be accomplished for most ecological systems (Chapter 2 Ratajczak et al., 2018). Moving from abrupt change methods requiring highly descriptive models and a priori assumptions of the state variable responses to drivers to methods requiring few, if any, a priori assumptions or knowledge is increasingly necessary for forecasting and managing complex ecosystems under an era of intensifying anthropogenic pressures. A few broad classes of quantitative approaches exist for quantitatively identifying abrupt changes in complex ecosystems. First, one can use simple mathematical models to describe the system and statistically test for discontinuities in the observed variables (e.g., in coral reefs, Mumby et al., 2013). Although mathematical representations are ideal, very rarely are ecological systems easily and well-described by them and often fail to meet the assumptions of the model. Second, we can track changes
Figure 61: An example solution of the Lorenz ('butterfly') represented in 3-dimensional phase-space. Phase plots are typically used to visualize stable areas within a system’s trajectory but reconstruction requires the difference models to be known and parameterized.

in the mean or variance of state variables to identify departures from the norm (e.g., early-warning indicators such as variance and variance index, Brock & Carpenter, 2006). Much like the mathematical modelling approach, these early-warning indicators have shown to be useful in some simple driver-response systems (e.g., lake eutrophication Carpenter et al., 2008b), but are unreliable in other empirical systems (e.g., Perretti & Munch, 2012; Dakos et al., 2012b; Dutta et al., 2018). The last type of approach is the model-free approach (Dakos et al., 2012a; Chapter 2). This group of abrupt change indicators can incorporate multiple state variables, and ideally requires no a priori assumptions about the expected driver-response relationships, or even about the drivers at all. It is this class of abrupt change indicators to that this chapter contributes.
6.1.1 Tracking ecosystem trajectory through time to explore system dynamics

A classic example of state-switching by a system is demonstrated in the Lorenz (‘butterfly’) attractor (Figure 61; Takens, 1981). This phase plot (Figure 61) provides an informative visual of the behavior of a chaotic system manifesting two attractors. Although the periodic, attractor behaviors are made clear when examining the time series of each dimension (Fig 62), identifying such behaviors in additional dimensions becomes increasingly difficult.

System trajectory in phase space are used often in dynamical systems theory and systems ecology to make inference regarding system behavior and dynamics, but phase space (trajectory) dynamics are not commonly applied outside theoretical studies as a tool for ecological data analysis (c.f. Sugihara et al., 2012 for an example of phase-space reconstruction using Taken’s theorem applied to ecological time series).
Some methods of attractor reconstruction have been applied to environmental data (e.g., individual time series of fisheries stocks, climate, stock market; Sugihara et al., 2012; Ye et al., 2015), yet they do not incorporate the dynamics of whole systems. Model-free methods for exploring and describing the dynamics of whole ecological systems are largely restricted to the commonly-applied dimension reduction techniques and clustering algorithms (e.g., Principal Components Analysis, K-means clustering). In fact, this is true of many abrupt change and regime shift indicators.

6.1.2 Rate of change as an indicator of abrupt change in the system trajectory

How quickly a system switches states [e.g., moving from attractor to another; 61] may yield insights into the responses of ecological systems to perturbations (e.g., anthropogenically induced pressures such as climate change, urbanization) and community shifts (e.g., species introductions or extinctions, shifts in dominance). For example, Beck et al. (2018) tracked rate of change using chord distances—a data transformation for positive values and that is suitable prior to ordination analysis—to capture abrupt changes in community composition of a temperate, paleodiatom community. Chord distance, however, is greatest when the observations among data rows (e.g., time, location) have no species in common. In other words, this measurement may be most useful in high community turnover conditions. Alternative numerical methods for estimating system rates of change may be useful when the system does not exhibit, for example, high degrees of turnover or changes other in simple, biodiversity metrics.

Rate of change (ROC, often represented as \( \Delta \)) is a term used for various measures that describe the relationship among to variables, measuring the change in one variable relative to another. As a refresher ROC is represented as speed (S) or velocity (V), where (S) is the adirectional magnitude (i.e. it is a scalar) of the displacement
of an object over unit time and \( V \) describes both the direction and magnitude (i.e. it is a vector) of the object’s movement in space-time. \( S \) is a scalar taking values of \( \geq 0 \) and \( V \) can take any value between \(-\infty\) and \( \infty \). For example, consider a car travelling at a constant speed of \( 50 \frac{km}{h} \) around along a hilly landscape, where it is ascending and descending hills. Although \( S \) is constant, \( V \) changes in a sinusoidal fashion, where \( V \) is \( V > 0 \) when ascending, \( V < 0 \) when descending, and \( V \approx 0 \) at in the valleys and at the peaks of the hills. Although \( S \) is useful when estimating other scalar quantities (e.g., \( \frac{miles}{gallon} \)), given a starting and/or final position in space, \( S \) is not informative of its the path traveled.

### 6.1.3 Aims

Here, I propose a method that simply describes the rate of change behavior of system dynamics in phase space: **velocity**, \( V \). An alternative to other model-free approaches (e.g., Fisher Information; Cabezas & Fath, 2002), the velocity metric allows one to examine the behavior of an entire system along its trajectory (through space or time) without having to reconstruct the phase space. The ability to handle noisy and high-dimensional data and the lack of subjective parameters in calculating the metric makes this method an alternative to existing early warning indicators and phase-space reconstruction methods.

I first describe the steps for calculating this new metric \( (V) \), as both a dimension reduction technique and abrupt change indicator. 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 Chapter 3, first introduced in Fath et al. (2003)]. I use a simple, two-variable system theoretical system to present baseline estimates of the expected behavior of \( V \) under various scenarios. I induce abrupt shifts under varying conditions of changing means
and variance of each of these two variables and discuss the contexts under which this metric (V) may signal or may fail to signal abrupt changes. Finally, I explore the utility of this metric in identifying known regime shifts in an empirical, high-dimensional paleoecological community time series.

6.1.4 Analytical approach

I first describe the steps for calculating velocity by constructing a simple, two-variable system that exhibits only a rapid, discontinuous change in the means of the state variables. I next vary the mean and variance of the state variables of this system to demonstrate baseline expectations for the behavior of velocity under a simple rapid shift scenario. Next, I construct a second model system similar to the first, but one that exhibits a continuous rapid change in the state variables. The purpose of this section is three-fold. First, I demonstrate how velocity behaves when the system undergoes varying degrees of change (e.g., slow change versus nearly discontinuous, rapid). Second, I concurrently identify baseline expectations of velocity under varying conditions of mean and variability of the state variables before and after a shift. Third, by introducing a smoothing function to the rapid shift, we gain an understanding of how process variability (noise) impacts the shift detectability by the velocity metric. Finally, I calculate the velocity of an empirical, paleolithic freshwater diatom community time series to demonstrate the utility of the velocity metric in highly noisy, high dimensional, and irregularly-sampled data.

6.2 Steps for Calculating Velocity, V

In this section, I first demonstrate the calculations of velocity using a very simple, two-variable toy system. The first system exhibits a rapid shift at a single point
in time, where mean and variance are constant before and after the shift point. I demonstrate the signals achieved with and the variability within the $V$ calculation by exploring a number of scenarios of this simple system. For the examples in this section, observations of $x_i$ are randomly drawn from distribution $x_i \sim \text{Normal}(\mu, \sigma)$, where $\mu$ is the mean and $\sigma$ is the standard deviation. Consider a system (Figure 63) 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 this simple, two-variable system to demonstrate how velocity is calculated. The system comprises variables $x_1$ and $x_2$, with observations

**Figure 63:** The 2-variable discrete time toy system used to demonstrate steps for calculating system velocity. Each variable, $x$, is drawn from a normal distribution with means that change at $t = 50$. State variables have constant standard deviation, $\sigma = 5$. 

[Image: graph of two variables $x_1$ and $x_2$ over time $t$]
occurring at each time point \( t = 1, 2, 3, \ldots 100 \). 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 \( t = 1 \) & \( t = 2 \) (Figure 64). For all examples in this chapter, the state variables \( x_1 \) and \( x_2 \) were drawn from a normal distribution (using function \texttt{rnorm}) , with parameters \( \bar{x}_i \) (mean) and \( \sigma_i \) (sd) for 100 time steps, \( t \). Under each scenario I induce a regime shift at time \( t = 50 \), where either or both a change in \( \bar{x}_i \) or \( \sigma_i \) occurs.

![Figure 64: Data used to calculate velocity at the first two time points, \( t_1 \) and \( t_2 \).](image)

6.2.1 Step 1: Calculate \( \Delta x_i \)

The first step is to calculate the change in values for each state variables, \( x_i \), between two consecutive time points [e.g., from time \( t \) to \( t + 1 \) for the discrete-time system;
Figure 64: Equation (6.1):

$$\Delta x_i = x_{i(t+1)} - x_{it}$$  \hspace{1cm} (6.1)

Note that $\Delta x_i$ can take any value between $-\infty$ and $\infty$.

### 6.2.2 Step 2: Calculate distance traveled, $s$

Next, we calculate the total change in the multivariate system as a function of the change in all state variables $x_i$. First, we calculate $\Delta s$ as the square root of the sum of squares of the changes in all state variables per Pythagoras’s theorem [Equation (6.2)]:

$$\Delta s = \sqrt{\sum \Delta x_i^2}$$ \hspace{1cm} (6.2)

Although $\Delta s$ represents the absolute change in the system between consecutive points in time, this measure is not yet relative along the system’s trajectory. To create a relative value we next calculate the total distance traveled along the system trajectory, $s$, as the cumulative sum of $\Delta s$ [Equation (6.2)] since the first observation, such that a cumulative sum is calculated for every $t$ over the interval $[0, T]$ [Equation (6.3)]:

$$s_T = \Sigma_{t=0}^{T} \Delta s$$ \hspace{1cm} (6.3)

We now have a single measure, $s_T$ [hereafter referred to as $s$; Equation (6.3)] at each discrete point in time in our $N$-dimensional system (Figure 65). It should be noted that $s$ (Figure 65) is monotonically increasing since the value of $\Delta s$ [Equation (6.2)] is a sum of squares. Although discussed in a later section, it is important to note that $s$ is not unitless—that is, $s$ has units of the state variables, $x_i$. For example, if our 2-variable toy system represents biomass, then the units of $s$ represents the cumulative absolute change in biomass of the entire system.
6.2.3 Step 3: 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 $s$ [Equation (6.3)], and then divide by the total time elapsed between consecutive sampling points:

$$v = \frac{\Delta s}{\Delta t} \quad (6.4)$$

The numerical results for each step in the calculation of velocity [Equation (6.4)] is demonstrated using the first five time points of our toy system (Figure 63) in Table 61.
Figure 66: System change ($s$) and velocity ($V$) of the model system over the time period. Constant means ($\bar{x}_{pre} = 25$, $\bar{x}_{post} = 10$) and sharp change in variance for both state variables, $\sigma = 5$.

Table 61: Steps outlined for calculating system velocity, $V$, using the 2-variable toy data as an example.

<table>
<thead>
<tr>
<th>$t$</th>
<th>$x_1$</th>
<th>$x_2$</th>
<th>$\Delta x_1$</th>
<th>$\Delta x_2$</th>
<th>$\Delta t$</th>
<th>$\sqrt{\sum_{i=1}^{N} \Delta x_i^2}$</th>
<th>$s$</th>
<th>$V$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>22.198</td>
<td>21.448</td>
<td></td>
<td></td>
<td></td>
<td>5.111</td>
<td>5.111</td>
<td>5.111</td>
</tr>
<tr>
<td>4</td>
<td>25.353</td>
<td>23.262</td>
<td>-7.441</td>
<td>-0.504</td>
<td>1</td>
<td>3.035</td>
<td>24.895</td>
<td>3.035</td>
</tr>
<tr>
<td>5</td>
<td>25.646</td>
<td>20.242</td>
<td>0.294</td>
<td>-3.020</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

6.3 Velocity $v$ Performance Under a Discontinuous Transition

I used simulation techniques to determine the baseline expectations of the performance of velocity $V$ under varying degrees of rapid shifts in the mean and variance of the toy system. The toy system in this section undergoes a discontinuous shift at
$t = 50$ (see 63). If the system undergoes a rapid and discontinuous change in one or more state variables, the velocity, because it is a rate of change, has the potential to approach $\infty$ as $\Delta t \to 0$. Therefore, it is important to understand the degree to that velocity can detect very sudden changes in mean values, despite effect sizes. Here, I varied each of the following system parameters at the regime shift location ($t = 50$): $\bar{x}_1$, increase in the mean value of $x_1$ and $\sigma_1$, the change in variance of $x_1$.

Simulations consisted of 10,000 random samples drawn from the normal distribution for each parameter, I randomly drew the toy system samples 10,000 times under increasing values of $\bar{x}_1$ and $\sigma_1$. To identify patterns in the influence of parameter values on velocity, I present the mean values of $V$ across all simulations, with confidence intervals of $\pm 2$ standard deviations. As mentioned 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 $\sigma_i$ (sd) for 50 time steps, $t$.

### 6.3.0.1 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 $W = \{25, 30, 35, ... 100\}$ (Figures 67 and 68). As expected, the magnitude of $V$ increases linearly as the total difference between $\bar{x}_{1_{\text{pre}}}$ and $\bar{x}_{1_{\text{post}}}$ increases (Figure 68). This is not surprising because $s$ increases as the total change in abundance across the entire system increases [Equation (6.3)]. Consequently the potential of $V$ also increases with total state variable values (e.g. abundance, biomass). The linear relationship among $V$ and total state variable values indicates that while $V$ is capable of identifying large shifts in data structure, it may fail to identify subtle changes (i.e. lower effect sizes).
6.3.0.2 Varying Post-shift Variance

In the previous example, variance was constant before and after the abrupt shift at $t = 50$. To determine whether the signal emitted by $V$ at the regime shift is lost or dampened when increasing variance I varied the variance parameter, $\sigma_1$ along the set $W = \{1, 2, 3, \ldots, 25\}$. The variance for both state variables ($x_1, x_2$) prior to the regime
Figure 68: Change in velocity ($V$) as the total change in the mean value of $\bar{x}_{2t=50}$ over 10,000 simulations. A regime shift was induced at $t = 50$ with constant variance $\sigma = 5$, $\bar{x}_2 = 25$ when $t < 50$, and changes in variable mean values, $\bar{x}_2 = 50$ when $t \geq 50$, $\bar{x}_1 = 25$ when $t < 50$.

shift, $\sigma_{x_1}$ and $\sigma_{x_2}$, was 5, with the change occurring in $\sigma_{x_{1\text{post}}}$.

6.3.0.3 Smoothing the data prior to calculating $v$

To determine whether process or observational noise influences the signal in $V$, I used linear approximation techniques to smooth the data prior to calculating the
Figure 69: High variance of velocity ($V$) in a single iteration ($N_{iter} = 1$, seed = 123) of simulations as we increase $\sigma_1$ at $t = 50$.

derivatives. I used the function `stats::approx` that linearly interpolates the original data, $x_1$ and $x_2$, to regularly-spaced time points along the set $t = \{1 : 100\}$. I then calculated $V$ as described in (Eqs. \[(6.1)\] through \[(6.4)\]). Increasing the number of points ($t$) at that the original state variables were smoothed (i.e., $t$) did not influence the amount of noise surrounding the signal of the regime shift (at $t = 50$) in system velocity, $V$ (Figure 610).
6.4 Velocity Performance Under a Smooth Transition

In the previous section I presented expectations for velocity signals under a discontinuous transition in a discrete-time system. Given velocity is a measure of the rate of a change in a system and the range of transition speeds ecological systems exhibit (e.g., slow driver-response or threshold dynamics), it is important to understand if and when the velocity signal is dampened under varying degrees of transition speeds.

Figure 6.10: The variance around the average (±2 SD) velocity ($V$) increases as the variance of $\bar{x}_{2_{t=50}}$ (post shift) increases. $\bar{x}_{1_{pre}} = 25$, $\bar{x}_{1_{post}} = 100$, $\bar{x}_{2_{pre}} = 25$, $\bar{x}_{2_{post}} = 50$, $\sigma_{1_{pre}} = 5$, $\sigma_{2_{pre,post}} = 5$
Figure 611: The noise in system velocity ($V$) is not obviously reduced in this system as the original data ($x_1, x_2$) is increasingly smoothed.

In this section I use a similar toy system, to demonstrate the expectations of velocity under a smooth shift and under varying degrees of rapidity.

Although the data constructed in this section are similar to that used in the previous section in that we are manipulating the mean and variance of two state variables before and/or after an abrupt shift, this section introduces a component of process noise into the shift itself. This is important because the derivative of a nearly discontinuous function is infinite. Although we are interested in identifying
Table 62: Conditions for generating various scenarios of the hyperbolic tangent-induced abrupt change. $\sigma_i$ represents the standard deviation of $\mu_{x_i}$ as the percent of $\mu_{x_i}$, $\mu_{x_i}$ is the mean of the state variable, $x_i$, and pre and post represent the periods before and after the regime shift at $t = 50$, respectively.

<table>
<thead>
<tr>
<th>conditions</th>
<th>$\sigma_{x1}$pre</th>
<th>$\sigma_{x1}$post</th>
<th>$\sigma_{x2}$pre</th>
<th>$\sigma_{x2}$post</th>
<th>$\mu_{x1}$pre</th>
<th>$\mu_{x1}$post</th>
<th>$\mu_{x2}$pre</th>
<th>$\mu_{x2}$post</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu_{x1}, \mu_{x2}, \sigma_{x1}, \sigma_{x2}$</td>
<td>0.05</td>
<td>0.10</td>
<td>0.05</td>
<td>0.10</td>
<td>10</td>
<td>55</td>
<td>15</td>
<td>44</td>
</tr>
<tr>
<td>$\mu_{x1}, \sigma_{x1}$</td>
<td>0.05</td>
<td>0.10</td>
<td>0.05</td>
<td>0.05</td>
<td>10</td>
<td>55</td>
<td>15</td>
<td>15</td>
</tr>
<tr>
<td>$\mu_{x1}, \mu_{x2}$</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
<td>10</td>
<td>55</td>
<td>15</td>
<td>44</td>
</tr>
<tr>
<td>$\mu_{x1}$</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
<td>10</td>
<td>55</td>
<td>15</td>
<td>15</td>
</tr>
<tr>
<td>$\sigma_{x1}, \sigma_{x2}$</td>
<td>0.05</td>
<td>0.10</td>
<td>0.05</td>
<td>0.10</td>
<td>10</td>
<td>10</td>
<td>15</td>
<td>15</td>
</tr>
<tr>
<td>$\sigma_{x1}$</td>
<td>0.05</td>
<td>0.10</td>
<td>0.05</td>
<td>0.05</td>
<td>10</td>
<td>10</td>
<td>15</td>
<td>15</td>
</tr>
</tbody>
</table>

For rapid shifts in systems, velocity will approach infinity as the rate of change in the shift increases and the sampling intervals decrease. In other words, if the system exhibits turnover in e.g. 25% of the state variables, we expect the value of velocity to be similar to that of a turnover in e.g. 75% of the variables. Removing the possibility of infinite values provides more relative measures within the community time series.

6.4.1 Generating the data

Here we consider a two-variable system over the time interval $[1, 100]$ with state variables $x_1$ and $x_2$ that exhibits abrupt shifts in mean and/or variance of one or both variables at time $t = 50$. I generated species observations for the true process and the true process with process variability. The true process data were created using the parameters for $\mu$ and $\sigma$ for each of the conditions in described in Table 62 (random seed in Program R was 12345).
6.4.1.1 True process model

Data were generated from a normal distribution and an abrupt shift in the mean was incorporated using a hyperbolic tangent function. The true process for each state variable, \( x_i \), was generated from [Equation (6.5); see Figure 613]:

\[
\begin{align*}
\mu_{x_{ipre}} &\sim \text{Normal}(\mu_{x_{ipre}}, \sigma_{x_{ipre}}) \\
\mu_{x_{ipost}} &\sim \text{Normal}(\mu_{x_{ipre}}, \sigma_{x_{ipost}}) \\
\mu_{xi}(t) &= \mu_{x_{ipre}} - 0.5(\mu_{x_{ipre}} - \mu_{x_{ipost}})(\tanh(\alpha(t - t_{shift})) + 1)
\end{align*}
\] (6.5)

where \( \mu_{xi}(t) \) is the mean value of \( x_i \) at time \( t \) and \( \text{pre} \) and \( \text{post} \) are the periods before and after the abrupt shift \( t_{shift} \), respectively. The parameter \( \alpha \) in Equation (6.5) controls for the rate of change at the point of the abrupt change, \( t_{shift} \), where higher values of \( \alpha \) correspond with a higher slope at \( t_{shift} \). I simulated a single iteration (data set) for various conditions of changing \( \mu_{xi} \) and \( \sigma_{xi} \) (see Table 62), for two state variables \( x_1 \) & \( x_2 \) at intervals of \( t = 1 \) along the temporal interval \( t = [1, 100] \).

6.4.1.2 Observed process data

I generated observations by imputing noise into the true process model [Equation (6.5)] through random sampling of \( \sigma_{xi} \) from a normal distribution [Equation (6.6); Figure 613]:

\[
\begin{align*}
\mu_{x_{ipre}} &\sim \text{Normal}(\mu_{x_{ipre}}, \sigma_{x_{ipre}}) \\
\sigma_{x_{ipre}} &\sim \text{Normal}(0, \sigma_{X_{ipre}}\mu_{X_{ipre}}) \\
\mu_{x_{ipost}} &\sim \text{Normal}(\mu_{x_{ipost}}, \sigma_{x_{ipost}}) \\
\sigma_{x_{ipost}} &\sim \text{Normal}(0, \sigma_{X_{ipost}}\mu_{X_{ipost}}) \\
\mu_{xi}(t) &= \mu_{x_{ipre}} - 0.5(\mu_{x_{ipre}} - \mu_{x_{ipost}})(\tanh(\alpha(t - t_{shift})) + 1)
\end{align*}
\] (6.6)
where $\sigma_{xi}$ is the observed error around $\mu_{xi}$, and $\sigma_{Xi}$ is $X\%$ of $\mu_{xi}$ under various sampling conditions (as described in Table 62). I generated the error as a percent of the mean as this scaling relationship is commonly observed in ecological data (Taylor, 1961).
Figure 6.13: An example of the data generated by the true process model. In this example the mean values ($\mu_{xi}$), but not the percent standard deviation ($\sigma_{xi}$), are varied before and after the transition point. The observed data are plotted against the true-process model for each state variable, $x_i$. Panels represent different degrees of the smoothing parameter, $\alpha$ (top: $\alpha = 0.25$, bottom: $\alpha = 1.00$).

6.4.2 Evaluating Velocity Performance Under Conditions of Changing Means and Variance

I simulated a single data set (using in Program R) by randomly drawing a single realisation (observed data) of the hyperbolic tangent process model with additive noise process [Equation (6.6)]. I then calculated the distance traveled, $s$, and the velocity of the distance traveled, $V$ (also referred to as $\frac{ds}{dt}$) using Equation (6.4)—this approach (Equation (6.4)) is a simple alternative to numerical integration techniques, requiring only simple algebraic techniques. This method is ideal for discrete time data, or
where computational power would not suffice for numerical integration. When using the first differences method, however, \( V \) will demonstrate high variability, depending on the amount of time between samples (i.e. as the intervals of \( t - t + 1 \) increase).

I also calculated \( V \) using a numerical integration method for non-smooth, noisy data, called total variation regularized differentiation (Chartrand, 2011). I used the R package `tvdiff` (Price & Burnett, 2019) to perform numerically integrate the distance traveled, \( s \). The regularized differentiation method in this package (function `tbdiff::tvRegDiff`; described fully in Chartrand, 2011) provides a numerical solution for calculating non-noisy derivatives of noisy, non-smooth data. Using this smooth-derivative estimation technique may be an ideal supplement to the velocity method in cases where process and observational error generate noisy observational data. Although not possible in most ecological systems data, here we can compare the fit of the smooth-derivative to the derivative of the true process, allowing us to determine the usefulness of calculating a smooth-derivative. There are two tuning parameters required to be chosen by the analyst when implementing the total-variation regularized differentiation, each of that influence the amount of noise smoothed out in the resulting derivative: \( \alpha \) and the number of iterations. I implemented this numerical differentiation over 1,000 iterations, and selected \( \alpha \) by comparing the anti-differentiated distance traveled, \( s \), to the true process values of \( s \) (e.g., see Figure 614). For most conditions and smoothness I found the tuning parameter for `tvdiff` \( \alpha = 0.50 \) provided a good fit of \( s \) (Figure 614), however, when the hyperbolic tangent smoothing parameter, \( \alpha \) was low (i.e. \( \alpha_{\text{tanh}} = 0.25 \)) higher values of \( \alpha_{\text{tvdiff}} \) yielded more abrupt changes in the derivative.
6.4.2.1 Smooth changes in the mean

As discussed earlier, the velocity of the distance traveled, $V$, is a measure of how quickly the sum of the squared system variables change between observations (i.e. time). Consequently, as the total change in state variables grows, so will the maximum potential of the velocity, $V$. Following this logic, we should expect to see a spike in the derivative of the distance traveled when the system changes quickly. I tested this hypothesis under two conditions of changing means, where either one or both vari-
Figure 615: The velocity signal is muted when the hyperbolic smoothing parameter, $\alpha$, is low (0.25). True and observed values of $x_i$ (panel A), observed distance traveled ($s$, panel B), observed velocity (C), and the smoothed velocity (D).

Variables underwent mean shifts (see Table 62), and under varying degrees of transition smoothness (i.e. $\alpha_{\text{tanh}} = 0.25, 0.50, 0.75, 1.00$).

When the hyperbolic tangent smooth transition function is less steep (Figure 615), the observed velocity signal is dampened. This signal, however, quickly recovers when the transition function becomes more abrupt (Figures 616, 617, 618; $\alpha_{\text{tanh}} = 0.5, 0.75, and 1.00$, respectively). The signal of velocity as an indicator of abrupt change appears more rapid when the mean of both state variables (Figure 622), rather than a single (Figure 615), are shifted (assuming constant variance). Figure 619 is representative of the increasing signal in velocity as $\alpha_{\text{tanh}}$ increases.
Figure 616: *The velocity signal is muted when the hyperbolic smoothing parameter, \( \alpha \), is moderate (0.50). True and observed values of \( x_i \) (panel A), observed distance traveled \( (s, \text{panel B}) \), observed velocity \( (C) \), and the smoothed velocity \( (D) \).*

6.4.2.2 Smooth changes in variance

Abrupt changes sometimes manifest first as a change in the variability, rather than the mean value, of the state variables. This condition manifests in the velocity signal when both variables experience a shift in relative variance (Figure 620), however, velocity does not signal change when only one variable exhibits a shift in variance (Figure 621). Again, given the total magnitude of change influences the distance traveled, \( s \), and the derivative of \( s, V \), it is not surprising that the velocity signal is greater around the transition point when both, compared to a single, state variable exhibits increased variability about the mean. In these scenarios I shifted the variability in the state variables \( x_i \) from only \( \sim 5\% \) to \( \sim 10\% \) (see Table 62)—this percent variability
The velocity signal is muted when the hyperbolic smoothing parameter, $\alpha$, is moderate (0.50). True and observed values of $x_i$ (panel A), observed distance traveled ($s$, panel B), observed velocity (C), and the smoothed velocity (D) is low relative to most empirical observational ecological data sets. As such, I expect the velocity signal to be more pronounced when empirical systems undergo shifts in variance in at least one state variable.

6.4.2.3 Smooth changes in the mean and variance

Given the signals identified in the velocity when one or both state variables exhibits a shift in mean and/or variance, it is unsurprising that even under smooth transitions (when $\alpha_{\text{tanh}} = 0.25$), velocity manifests as a signal of change (Figure 623). This signal is most pronounced when the shift is abrupt (Figure 624).
Figure 618: The velocity signal is muted when the hyperbolic smoothing parameter, $\alpha$, is moderate (0.50). True and observed values of $x_i$ (panel A), observed distance traveled ($s$, panel B), observed velocity (C), and the smoothed velocity (D).

6.5 Velocity Performance Under Empirical Transitions: paleolithic freshwater diatom communities

To gather baseline information on the use of velocity in empirical systems data, I calculated velocity for the paleodiatom system described in Chapter 5. Briefly, the paleodiatom community comprises 109 time series over a period of approximately 6936 years (Figure 625). As elaborated in 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 abundances over time, as there are extreme levels of species turnover at multiple points in the data (Figure 625).
Figure 619: The velocity signal is regained under smooth transition ($\alpha_{\tanh} = 0.75$) when both state variables undergo a shift in the mean. True and observed values of $x_i$ (panel A), observed distance traveled ($s$, panel B), observed velocity ($C$), and the smoothed velocity ($D$).

Using Fisher Information and climatological records, Spanbauer et al. (2014) suggest that regime shifts in this system at approximately 1,300 years before present (where years before present is equal to year 1950; Spanbauer et al., 2014). Spanbauer et 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 traveled ($s$) and velocity ($V$; Figure 629). The results of $V$ and $s$ (Figure 626) on the relative abundance data correspond with both the large shifts in species dynamics (see Fig 625) and also with the regime shift identified by Spanbauer et al.
Figure 620: The velocity signals a rapid shift in the variance of both state variables under a moderately abrupt transition (\(\alpha_{\text{tanh}} = 0.75\)). True and observed values of \(x_i\) (panel A), observed distance traveled (\(s\), panel B), observed velocity (C), and the smoothed velocity (D).

However, two primary results can be made from the metrics \(V\) and \(s\) that are not obvious nor identified numerically in the results of Spanbauer et al. (2014):

1. Two additional large shifts occurred at approximately 2,500, 4,800 and years before 1950

2. The periods before the first and after the second large shifts appear oscillatory (Figure 627).

To determine the effect of dampening the noise in the data on the velocity signal, I interpolated the each time series using function `stats::approx` to 700 time
Figure 621: The velocity does not signal shifts in the variance of a single variable ($x_1$) under a moderately abrupt transition ($\alpha_{tanh} = 0.75$). True and observed values of $x_i$ (panel A), observed distance traveled ($s$, panel B), observed velocity (C), and the smoothed velocity (D).

points. Next, I calculated the distance traveled 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 $\alpha = 100$). This method of regularized differentiation is an ideal approach to smoothing $s$ because it assumes the data are non-smooth and incorporates finite differencing.

The total variation regularized differentiation is described in Chartrand (2011), Price & Burnett (2019a), and in the previous first-level section. The smoothed velocity (Figure 629) provides a similar but smoother picture of the velocity of the system trajectory. Comparing the smoothed (Figure 629) to the non-smoothed velocity (Fig-
Figure 622: The velocity signal is regained under smooth transition ($\alpha_{tanh} = 0.25$) when both state variables undergo a shift in the mean. True and observed values of $x_i$ (panel A), observed distance traveled ($s$, panel B), observed velocity (C), and the smoothed velocity (D).

Figure 626 yields similar inference regarding the location of the regime shifts at 2,200 and 1,300 years before present, however, it more clearly demonstrates potential inter-regime dynamics (e.g., between 7,000 and 4,800 years before present), that were not identified in previous study of this system (Spanbauer et al., 2014).

6.6 Discussion

Here, I described the steps for calculating a novel regime detection metric, system velocity ($V$). First described in Fath et al. (2003), $V$ is used as a single step for calculating 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 (Figure 610), and smoothing the original data does not reduce the noise surrounding this metric when variance is moderate. Variance is a commonly-used indicator of ecological regime shifts (Brock & Carpenter [2006]), however, is difficult to interpret when the number of variables is much greater than a few. System velocity, $V$, may be useful in situations where the number of state variables is much greater than few, and appears especially useful when the magnitude of change in one or more state variables is high.

Figure 623: The velocity signals a shift when both variables undergo shifts in the mean and variance under a slightly abrupt transition ($\alpha_{\text{tanh}} = 0.25$). True and observed values of $x_i$ (panel A), observed distance traveled ($s$, panel B), observed velocity (C), and the smoothed velocity (D).
Figure 624: The velocity signals a shift when both variables undergo shifts in the mean and variance under a slightly abrupt transition \( \alpha_{\text{tanh}} = 1.00 \). True and observed values of \( x_i \) (panel A), observed distance traveled (\( s \), panel B), observed velocity (C), and the smoothed velocity (D).

(Figures 68, 619). For example, this method will likely identify signals of regime shifts where the shift is defined as high species turnover within a community (Figure 624).

This study provides baseline expectations of the velocity of the distance traveled, \( V \), as an indicator of abrupt change in a multivariate system. Although a useful first step, this metric should next be critiqued in a sensitivity analytical approach, where a statistical measure is used to determine whether \( V \) indicates abrupt shifts prior to occurrence (c.f. during or after), particularly with respect to its performance in community-level empirical data. The paleolithic diatom data used in the last section of this chapter is also presented in the documentation for my R Package,
Figure 625: 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.

regimeDetectionMeasures (Appendix 4). In this case study, the ‘distance traveled’, \( s \) [Equation (6.1)], 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 package materials) indicates a large shift at only the most predominant shift in the time series, perhaps due to the sensitivity of the metric to variance (Figure 68).

Further work is required to determine the utility of system velocity as a regime
Figure 626: Velocity $V$ and distance traveled $s$ of the paleoatom time series. Dashed line at 1,300 years before 1950 indicates the regime shift identified in Spanbauer et al. (2014). Dotted lines indicate regime shifts as visually identified on metrics $s$ and $V$.

detection metric, however, this chapter demonstrates that the metric may indicate clear shifts in variable means and variability about the means. 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
Figure 628: 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, $s$.

6.7 Supplementary Figures

Figures 630, 631, and 632 provide additional examples of the behavior of velocity, $V$ when varying the mean and/or variance prior to and/or after the induced abrupt shift in the toy system with a discontinuous transition at $t = 50$. 

abrupt shifts.
Figure 629: The velocity metric ($V$) signals potential periodicities in the paleo diatom time series data when the distance traveled metric, $s$, is smoothed using regularized differentiation methods (see Price and Burnett 2019).
Figure 630: System change ($s$) and velocity ($V$) of the model system over the time period. Change in means ($\bar{x}_{1\text{pre}} = 25$, $\bar{x}_{1\text{post}} = 100$, $\bar{x}_{2\text{pre}} = 50$, $\bar{x}_{2\text{post}} = 10$) and an increase in variance ($\sigma_{1\text{pre}} = 2$, $\sigma_{1\text{post}} = 10$, $\sigma_{2\text{pre}} = 5$, $\sigma_{2\text{post}} = 10$).
Figure 631: 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 $\sigma_{1\text{pre}} = 2$, $\sigma_{1\text{post}} = 12$, $\sigma_{2\text{pre,post}} = 5$. 

constant mean, variance increase ($x_1$)
changing means, variance = mean

**Figure 632:** System change ($s$) and velocity ($V$) of the model system over the time period. Variance equal to mean ($\bar{x}_i = \sigma_i$), where means ($\bar{x}_{1\text{pre}} = 25$, $\bar{x}_{1\text{post}} = 50$, $\bar{x}_{2\text{pre}} = 15$, $\bar{x}_{2\text{post}} = 150$).
Chapter 7

Grassland Obligates and Declining Birds Operate Near Edges of Body Mass Distributions

7.1 Introduction

Animal body mass distributions have been used to identify scaling structures of ecological communities (Holling, 1992, Allen & Holling, 2002, Allen, 2006). Using statistical methods to identify gaps, or discontinuities, in body mass distributions, some patterns are observed within and across taxonomic groups and biomes. Given the ubiquity of discontinuities identified in body mass distributions of fauna and social systems (Allen, 2006), the ecological significance of these patterns may prove useful in understanding ecosystem structure and functioning (Angeler et al., 2016).

Various hypotheses are posited as drivers of the observed discontinuities in animal body mass distributions, including those related to resource use (the Energetic and Textural Discontinuity hypotheses), community interactions, biogeography, and evolution/phylogenetics (Holling, 1992, Blackburn & Gaston, 1994, Allen, 2006, Allen et al., 2006).

Body size influences the frequency and intensity of inter- and intraspecific competition for resources, territory, and mates, thereby dictating the spatial and temporal scales at which a species of a distinct body size operates (Peters & Wassenberg, 1983).
 Silva & Downing, 1995; Allen et al., 2006. The scaling structure of terrestrial communities are have been found to have ‘lumpy’ distributions; that is, they are not well-described using parametric statistical descriptions. If the scaling structure of a community manifests in the body mass distribution of the community, it is considered reflective of the discontinuous and heterogeneous nature of resource use. Specifically, Holling (1992) suggests that the body mass distribution of a community or group of species reflects the discontinuous nature of environmental structures and processes. Quantitative analyses of animal body sizes (Allen et al., 2006; Nash et al., 2014b) and other similar distributions has revealed the ubiquity of the discontinuous nature of distributions of animal body masses (Havlicek & Carpenter, 2001; Skillen & Maurer, 2008), plant biomass (Spanbauer et al., 2016), city population sizes (Garmestani et al., 2005), and animal home range sizes (Restrepo & Arango, 2008).

Avian distribution and presence data are abundant, easily accessible and, more importantly, provide insights into resource availability and structure at the local and landscape scales. In this Chapter I used discontinuity analysis of avian body mass distributions to identify the scaling structures of local avian communities in the Prairie Potholes, Central Mixed Grass, and Eastern Tall Grass regions of the central Great Plains of North America. I then use linear mixed modelling to determine whether the body mass distributions of these avian communities exhibit change corresponding to the spatial boundaries recently proposed in (Roberts et al., 2019). Although I found evidence supporting the hypothesis that declining grassland birds operate near the ‘edges’ of body mass aggregations, which represent the spatiotemporal scaling of resource use and availability, I did not identify shifts in the body mass distributions corresponding with the previously proposed pole ward, spatial shifts.
7.2 Methods

7.2.1 Study area

A recent study (Roberts et al., 2019) identified what they refer to as spatial regimes across a large portion of the central United States (see Figure 71) using inference gained from discontinuity analyses of avian body mass distributions. The authors

Figure 71: My study area (red box) overlaying the terrestrial Breeding Conservation Regions (BCR) in North America.
hypothesize that a spatial regime boundary exists in the Central Great Plains and suggests it has exhibited a Northward shift at a rate of $\sim \frac{0.05^\circ \text{latitude}}{\text{year}}$. The authors used discontinuity analysis to identify these ‘spatial regimes’, using the body masses of breeding birds. Their hypothesized spatial regime boundary occurs at $39^\circ$ latitude in year 1970, $39.5^\circ$ latitude in year 1985, $40^\circ$ latitude in year 2000, and $40.5^\circ$ latitude in year 2015 (see Figure 72). Sampling sites were classified each year as belonging to either the Southern or Northern regime according to whether the location was below (Southern) or above (Northern) the regimes identified by Roberts et al. (2019).

My study area is designed such that there is minimal crossing of very different BCR, or habitat types (Figure 71). In other words, this study area largely falls within BCRs which can be generally classified as grassland habitat (BCR 11, Prairie Potholes; BCR 19, Central Mixed Grass ; BCR 22, Eastern Tall Grass). Using this design we should expect that the functional groups within our avian communities should be similar across BCR boundaries, despite a potential turnover in species identity. Accounting for change in habitat across space allows us to assume that any observed change in the scaling structure of the avian community is due to changes in habitat and resource availability such that similar species are or are not included in the community.

7.2.2 Data

7.2.2.1 Avian census data - North American Breeding Bird Survey

I constructed body mass distributions using route-level data from the North American Breeding Bird Survey hereafter (NABBS; Sauer et al., 2014). The NABBS uses citizen scientist volunteers to annually collect data using a standardized roadside, single observer, 3-minute point count protocol and has organized data collection annually.
across North America (Figure 41) since 1966. Each roadside survey consists of 50, 3-
minute point counts (data collected using sight and sound) along ~ 24.5 mile stretch
of road. Although the point counts are designed to collect estimates of relative abun-
dance, the method for building body mass distributions used in this chapter require
only presence absence data. I therefore converted abundances to presence-absence
data. I considered a species as ‘present’ if it was detected on the year in question or
the ±1 year to account for potential false negative observations (i.e., a species was
not detected in the NABBS route despite its presence in the local community).

**Identifying avian census locations** To determine whether the spatial regime
shifts identified in Roberts *et al.* (2019) manifested in local avian community struc-
ture, I restricted analysis to the grassland habitat of the Central Great Plains. All
routes falling within a rectangular area bounded by coordinates 37.8° and 44.5° lati-
tude, and −101° and −95.5° longitude [see Figure 72]. I retained all NABBS routes
which used the sampling protocol ‘101’, which is the standard method for conducting
NABBS point count surveys.

**Avian body mass data** Species operating at similar spatial and temporal scales
are those which are close in body size as identified using statistical aggregation identi-
fication techniques (Allen *et al.*, 1999). The interactions among species within a single
body mass aggregation are presumed to experience a higher frequency and intensity of
interspecific interactions with each other as opposed to those in different aggregations
(Peterson *et al.*, 1998). Although some species of birds are sexually size dimorphic, I
am unaware of any sexually size-dimorphic birds that would likely operate at different
spatial and temporal. Therefore, I constructed body mass distributions of each avian
census using the sex-averaged body masses published in Dunning Jr (2007) (available
Removing species from analysis  Due to strict reliance on volunteers, some routes are not covered every year. Although NABBS volunteers attempt identify all species in the point-count area, biases exist in data collection. Rather than retain observations of cryptic or species with low detection rates, I removed select species from the censuses (see Methods section in Chapter 4 for further discussion of this topic). I analyzed species of the following taxonomic families: Accipitriformes, Apodiformes, Cathartiformes, Charadriiformes, Columbiformes, Coraciiformes, Cuculiformes, Galliformes, Gruiformes, Passeriformes and Piciformes. Although removing cryptic, nocturnal, and some crepuscular species (e.g. Caprimulgiformes) from the analysis may yield a more conservative body mass distribution, including them may result in correctly identifying additional scaling structures (or body mass aggregations) in some routes but not in others. This method of exclusion also results in a loss of some medium- and larger-bodied Ciconiiformes (Podicipediformes, Phoenicopteriformes, Ciconiiformes; e.g. grebes, pelicans).

Taxonomic munging of the census data  Although the NABBS survey reports species-specific abundances, some birds are only classified to genera or order. Common examples of these species are those which are nearly indistinguishable from each other (e.g., Glossy Ibis and White-faced Ibis), birds which are difficult to see under certain conditions (e.g., hummingbirds, fast-moving hawks or accipiters), or species whose songs are similar. Numerous species were presented as identified to family or genus (e.g., Accipiter sp., Buteo sp., and Trochilids sp.) and others are categorized as hybrid.

I made decisions regarding species-specific classification based on the North Amer-
ican breeding range maps provided by the Cornell Lab of Ornithology. Many unidentified species were easily categorized given the lack of overlap in species’ ranges in our study area. For example, *Baeolophus bicolor* is nearly indistinguishable from *Baeolophus atricristatus*, however *B. atricristatus* is not known to occur in our study area (Figure 71)—therefore all accounts classified as either *B. bicolor* or *B. atricristatus* were classified as the former. Informed decisions of this nature were made regarding the following unidentified species, where the second name in the binomial was assigned as the species preceding the “/”:*Passerina cyanea / amoena*, *Corvus brachyrhynchos / ossifragus*, *Petrochelidon pyrrhonota / fulva*, *Corvus brachyrhynchos*, *Quiscalus major / mexicanus*, *Pipilo maculatus / erythrophthalmus*, *Sturnella magna / neglect*, *Plegadis chihi / falcinellus*, *Coceyzus erythropthalmus / americanus*, *Empidonax traillii / alnorum*, *Icterus galbula / bullockii*, *Nyctanassa nycticorax / violacea*, and *Poecile atricapillus / carolinensis* were all classified according to their known distributions. I classified unidentified hummingbirds (*Trochilid* sp.) as *Selasphorus rufus*, and unidentified Terns (Tern sp.) as *Childonias niger*. All unidentified Accipiters (Accipiter sp.), Buteos (Buteo sp.), and Gulls (Gull sp.) were removed from analysis entirely as there are no clear differences in the probability of occurrence in our study area.

7.2.2.2 Identifying species of interest

Allen *et al.* (2006) propose alternative hypotheses for the relative locations of species within the body mass aggregation distributions as a function of ‘distance-to-edge’, a measure indicating the distance (in log-mass units) of each species to the edge of a body mass aggregation. This distance-to-edge measure is zero when the species falls at the edge of a statistically identified body mass aggregation. This species is often referred to as an ‘edge species’. To determine the effect of spatial regime shifts on
edge species, I identified three types of species of interest: (1) grassland obligates species, (2) species with widespread population declines in the study area, and (3) a combination of these groups. All remaining species were classified as ‘other’.

**Grassland obligate species**  The spatial regimes identified in Roberts et al. (2019) are attributed to large-scale changes in the landscape, including woody plant invasion. The loss of native grassland in our study area due to land conversion is largely attributed to anthropogenic land use change (e.g., development) and fire suppression. Numerous species have been negatively impacted by this widespread habitat loss, but grassland obligates are particularly at risk. Grassland obligates should be strong indicators of the large-scale spatial regime shifts identified in Roberts et al. (2019), given

**Figure 72:** North American Breeding Bird Survey routes (points), latitudinal locations (horizontal bars) of the spatial regimes identified by roberts2019shifting.
their high sensitivity to grassland habitat loss (Herkert, 1994). I identified North American grassland obligate species from the grey literature (Shriver et al., 2005; Initiative et al., 2009) and white literature (Peterjohn & Sauer, 1999). Although some grassland obligates were positively impacted by the Conservation Reserve Program (CRP; Peterjohn & Sauer, 1999), this group of birds exhibited strong declines in North America until approximately 2003 (Initiative et al., 2009).

**Declining species** I classified a species as ‘declining’ based on the results of the North American Breeding Bird Survey (Sauer et al., 2014). The Patuxent Wildlife Research Center uses hierarchical modelling techniques to estimate the trends of species using the NABBS data at various spatial scales. Sauer et al. (2014) also provides estimates of data credibility according to data availability, number of routes used to build the population trend estimate, abundance, and probability of detecting a small change in population trend. These credibility scores are generated for multiple spatial extents: state-level, BCR-level, and across the three regions of the United States (Western, Central, and Eastern). Given the extent of this study, I considered the data credibility estimates using the Central Breeding Bird Survey Region, the Prairie Pothole BCR (BCR 11) and Eastern Tallgrass Prairie (BCR 22). A species was considered as declining only if the trend estimate was categorized as having moderate precision and abundance (blue) or having a deficiency (yellow). I considered the population trend estimates provided for the period of 1966 - 2015.
7.2.3 Statistical analysis

7.2.3.1 Identifying scaling structure of avian communities using body mass distributions

Discontinuities in body mass distributions been quantified using various methods (e.g., multivariate time series models, regression trees, and gap rarity index) which are collectively referred to as ‘discontinuity analyses’ (Allen, 2006; Stow et al., 2007; Nash et al., 2014; Barichievy et al., 2018). Using various methods, the discontinuous nature of body masses of ecological communities has been observed in various taxa of both terrestrial (Allen et al., 2006) and aquatic (Spanbauer et al., 2016) communities. Multiple methods are proposed for identifying discontinuities in body mass aggregations (Allen & Holling, 2001), including clustering algorithms (Stow et al., 2007), body mass difference indices (Holling, 1992), gap rarity index (Restrepo & Arango, 2008), and more recently the discontinuity detector (Barichievy et al., 2018), an extension of the gap rarity index (Restrepo & Arango, 2008).

I used the discontinuity detector described in Barichievy et al. (2018), which uses likelihood to determine whether the observed data contains multiple modes as compared to that of a Gaussian (uni-modal) distribution. This method requires multiple user-defined parameters, including an imputation resolution (1000) and a bootstrap sample size (1000) over which the null distribution is randomly sampled. I provided a slightly altered and annotated version of the functions `Neutral.Null` and `Bootstrap.Gaps` (first printed in Barichievy et al. (2018)) used to identify discontinuities in a continuous variable in Appendix 4. Two criterion have been used to determining the exact location of discontinuities within a rank-ordered continuous variable: using a constant significance/threshold level (Barichievy et al., 2018) and a power constant table for varying sample sizes (Roberts et al., 2019). It should be
noted because the power-constant method identifies a larger proportion of “significant” discontinuities, or edges, identifying aggregations requires subjective measures regarding actual aggregation locations. Using the percentile method avoids this subjectivity, however, the aggregation number and locations are sensitive to choice of percentile or threshold value. Following the methods of Barichievy et al. (2018), I considered a value to be a discontinuity if the gap percentile (see Appendix [4] was $\geq 90$. I built route-level body mass distributions for each route-year combination by using presence absence data from the current, previous, and following year to account for observational and process errors impacting the detectability of a species within a single route. Using this method reduces the amount of species-specific, and consequently specific-body mass turnover within a route over time. This also assumes that an unobserved species is truly absent, an assumption which is difficult to avoid without a sophisticated occupancy modelling approach for each species in the community.

7.2.3.2 Determining Effects of a Spatial Regime Boundary on Grassland Birds

If the spatial regime shift occurred in the bird community, it should manifest in the local community scaling structure through one or both of species turnover and a shift in the number of body mass aggregations. I used linear mixed modelling to determine whether the local scaling structure and the location of grassland obligates and declining species within these scaling structures are impacted by the spatial regime boundaries proposed by Roberts et al. (2019).

I used a linear mixed model to determine whether the proposed moving spatial regime boundaries influenced the location of species of interest (grassland obligates, declining species) within their respective body mass aggregation. Each species was assigned a ‘distance to edge’, which served as a proxy of the proximity of a species
to the nearest edge of its respective body mass aggregation. Previous studies suggest that this distance to edge measure can be used to identify zones of transitions, as invasive and threatened species tend to be located at the edges of aggregations (Allen et al., 1999). Following this hypothesis, one should expect to see changes in the locations of sensitive and declining species change in the areas undergoing so called spatial-regime shifts (Roberts et al., 2019).

**Table 71:** Grassland obligates and species with declining trends over the period of (1966-2015) in the Central Breeding Bird Survey region in our study area.

<table>
<thead>
<tr>
<th>Common Name</th>
<th>Species Group</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bobolink</td>
<td>Grassland Obligates</td>
</tr>
<tr>
<td>Lark Sparrow</td>
<td>Grassland Obligates</td>
</tr>
<tr>
<td>Chipping Sparrow</td>
<td>Grassland Obligates</td>
</tr>
<tr>
<td>Henslow’s Sparrow</td>
<td>Grassland Obligates</td>
</tr>
<tr>
<td>Ferruginous Hawk</td>
<td>Grassland Obligates</td>
</tr>
<tr>
<td>Upland Sandpiper</td>
<td>Declining Grassland Obligates</td>
</tr>
<tr>
<td>Ring-necked Pheasant</td>
<td>Declining Grassland Obligates</td>
</tr>
<tr>
<td>Horned Lark</td>
<td>Declining Grassland Obligates</td>
</tr>
<tr>
<td>Eastern Meadowlark</td>
<td>Declining Grassland Obligates</td>
</tr>
<tr>
<td>Western Meadowlark</td>
<td>Declining Grassland Obligates</td>
</tr>
<tr>
<td>Vesper Sparrow</td>
<td>Declining Grassland Obligates</td>
</tr>
<tr>
<td>Grasshopper Sparrow</td>
<td>Declining Grassland Obligates</td>
</tr>
<tr>
<td>Field Sparrow</td>
<td>Declining Grassland Obligates</td>
</tr>
<tr>
<td>Dickcissel</td>
<td>Declining Grassland Obligates</td>
</tr>
<tr>
<td>Savannah Sparrow</td>
<td>Declining Grassland Obligates</td>
</tr>
</tbody>
</table>
Lark Bunting  Declining Grassland Obligates
Cassin's Sparrow  Declining Grassland Obligates
Chestnut-collared Longspur  Declining Grassland Obligates
Killdeer  Declining
Northern Bobwhite  Declining
Rock Pigeon  Declining
Mourning Dove  Declining
Yellow-billed Cuckoo  Declining
Black-billed Cuckoo  Declining
Belted Kingfisher  Declining
Downy Woodpecker  Declining
Red-headed Woodpecker  Declining
Red-bellied Woodpecker  Declining
Chimney Swift  Declining
Eastern Kingbird  Declining
Western Kingbird  Declining
Great Crested Flycatcher  Declining
Blue Jay  Declining
American Crow  Declining
European Starling  Declining
Red-winged Blackbird  Declining
Orchard Oriole  Declining
Baltimore Oriole  Declining
Common Grackle  Declining
Song Sparrow  Declining
Purple Martin  Declining
Barn Swallow  Declining
Loggerhead Shrike  Declining
Common Yellowthroat  Declining
House Sparrow  Declining
Brown Thrasher  Declining
Black-capped Chickadee  Declining
Gray Partridge  Declining
Wood Thrush  Declining
Northern Harrier  Declining
Greater Prairie-Chicken  Declining
Scissor-tailed Flycatcher  Declining
Black-billed Magpie  Declining
Northern Mockingbird  Declining
Bewick’s Wren  Declining
Kentucky Warbler  Declining
Carolina Chickadee  Declining
King Rail  Declining
Bullock’s Oriole  Declining
Rock Wren  Declining
Ovenbird  Declining
Prothonotary Warbler  Declining
Curve-billed Thrasher  Declining
I modelled the ‘distance to edge’ for each species as a function of time, whether the NABBS route is in the ‘Southern’ or ‘Northern’ spatial regime [South or North of the boundaries proposed in Roberts et al. (2019); see Figure 72], time, route identity, and a descriptive group (one of grassland obligate, declining, declining grassland obligate, other). The ‘distance to edge’ response variable was not strongly correlated with body mass (Figure 712), however, I scaled and centered the response values to avoid obtaining predictions lower than zero for the unscaled response.

Fixed effects included an interaction among year ($\beta_1$) and regime location ($\beta_2$), and an interaction among regime ($\beta_2$) and species group ($\beta_3$). Random intercepts were estimated for each species ($b_1$) within each NABBS route ($b_2$). That is, species was nested within route. An auto-regressive lag-1 correlation structure was assumed for the random intercept estimates. The model was fitted using restricted maximum likelihood. The model was fitted using `nlme::lme` and was coded as: > `nlme::lme(distEdge.scaled ~ year.center * regime + regime * sppGroup, random = ~ 1 | loc/aou, correlation = corAR1(form = ~ 1 | loc / aou ), method = “REML”)`

### 7.3 Results
Table 72: The number of NABBS routes analysed in the Southern regime is smaller than those used in the Northern regime each year given the location of the regimes identified in a previous study with respect to the contiguous grasslands of Central North America.

<table>
<thead>
<tr>
<th>Year</th>
<th>North</th>
<th>South</th>
</tr>
</thead>
<tbody>
<tr>
<td>1967</td>
<td>38</td>
<td>11</td>
</tr>
<tr>
<td>1968</td>
<td>43</td>
<td>12</td>
</tr>
<tr>
<td>1969</td>
<td>53</td>
<td>11</td>
</tr>
<tr>
<td>1970</td>
<td>45</td>
<td>17</td>
</tr>
<tr>
<td>1971</td>
<td>43</td>
<td>13</td>
</tr>
<tr>
<td>1972</td>
<td>38</td>
<td>17</td>
</tr>
<tr>
<td>1973</td>
<td>39</td>
<td>15</td>
</tr>
<tr>
<td>1974</td>
<td>29</td>
<td>17</td>
</tr>
<tr>
<td>1975</td>
<td>41</td>
<td>16</td>
</tr>
<tr>
<td>1976</td>
<td>46</td>
<td>18</td>
</tr>
<tr>
<td>1977</td>
<td>47</td>
<td>16</td>
</tr>
<tr>
<td>1978</td>
<td>45</td>
<td>18</td>
</tr>
<tr>
<td>1979</td>
<td>46</td>
<td>17</td>
</tr>
<tr>
<td>1980</td>
<td>43</td>
<td>16</td>
</tr>
<tr>
<td>1981</td>
<td>40</td>
<td>18</td>
</tr>
<tr>
<td>1982</td>
<td>38</td>
<td>16</td>
</tr>
<tr>
<td>1983</td>
<td>37</td>
<td>16</td>
</tr>
<tr>
<td>1984</td>
<td>30</td>
<td>14</td>
</tr>
<tr>
<td>1985</td>
<td>30</td>
<td>24</td>
</tr>
<tr>
<td>1986</td>
<td>31</td>
<td>21</td>
</tr>
<tr>
<td>Year</td>
<td>Page</td>
<td>Column 1</td>
</tr>
<tr>
<td>------</td>
<td>------</td>
<td>----------</td>
</tr>
<tr>
<td>1987</td>
<td>36</td>
<td>21</td>
</tr>
<tr>
<td>1988</td>
<td>34</td>
<td>24</td>
</tr>
<tr>
<td>1989</td>
<td>33</td>
<td>25</td>
</tr>
<tr>
<td>1990</td>
<td>40</td>
<td>21</td>
</tr>
<tr>
<td>1991</td>
<td>31</td>
<td>22</td>
</tr>
<tr>
<td>1992</td>
<td>33</td>
<td>23</td>
</tr>
<tr>
<td>1993</td>
<td>39</td>
<td>23</td>
</tr>
<tr>
<td>1994</td>
<td>37</td>
<td>23</td>
</tr>
<tr>
<td>1995</td>
<td>37</td>
<td>22</td>
</tr>
<tr>
<td>1996</td>
<td>39</td>
<td>22</td>
</tr>
<tr>
<td>1997</td>
<td>36</td>
<td>24</td>
</tr>
<tr>
<td>1998</td>
<td>34</td>
<td>21</td>
</tr>
<tr>
<td>1999</td>
<td>32</td>
<td>23</td>
</tr>
<tr>
<td>2000</td>
<td>37</td>
<td>37</td>
</tr>
<tr>
<td>2001</td>
<td>31</td>
<td>34</td>
</tr>
<tr>
<td>2002</td>
<td>32</td>
<td>36</td>
</tr>
<tr>
<td>2003</td>
<td>36</td>
<td>38</td>
</tr>
<tr>
<td>2004</td>
<td>35</td>
<td>35</td>
</tr>
<tr>
<td>2005</td>
<td>38</td>
<td>36</td>
</tr>
<tr>
<td>2006</td>
<td>39</td>
<td>35</td>
</tr>
<tr>
<td>2007</td>
<td>39</td>
<td>36</td>
</tr>
<tr>
<td>2008</td>
<td>35</td>
<td>34</td>
</tr>
<tr>
<td>2009</td>
<td>39</td>
<td>37</td>
</tr>
<tr>
<td>2010</td>
<td>37</td>
<td>30</td>
</tr>
</tbody>
</table>
7.3.1 Summary statistics of censuses (NABBS data)

Given the location of the study area (Figure 71) with respect to the location of the contiguous Central Great Plains, fewer NABBS routes falling into the Southern regime were analysed than those in the Northern (Table 72). Likely due to the increase in the total number of routes surveyed over time across the entire North American Breeding Bird Survey region, species richness increased over time within our study area (Figure 74). Annual turnover rates were relatively low but became more dispersed over time (Figure 75; Table 73).

Table 73: Summary statistics for annual species richness and annual turnover in all NABBS routes in study area.

<table>
<thead>
<tr>
<th>Year</th>
<th>Annual Richness</th>
<th>Annual Turnover</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( \bar{x} )</td>
<td>( \sigma )</td>
</tr>
<tr>
<td>1967</td>
<td>39.4</td>
<td>8.14</td>
</tr>
<tr>
<td>1968</td>
<td>41.4</td>
<td>8.11</td>
</tr>
<tr>
<td>1969</td>
<td>39.9</td>
<td>7.44</td>
</tr>
<tr>
<td>1970</td>
<td>41.3</td>
<td>7.73</td>
</tr>
<tr>
<td>Year</td>
<td>Value 1</td>
<td>Value 2</td>
</tr>
<tr>
<td>------</td>
<td>---------</td>
<td>---------</td>
</tr>
<tr>
<td>1971</td>
<td>39.4</td>
<td>7.50</td>
</tr>
<tr>
<td>1972</td>
<td>41.0</td>
<td>9.80</td>
</tr>
<tr>
<td>1973</td>
<td>40.1</td>
<td>8.33</td>
</tr>
<tr>
<td>1974</td>
<td>38.8</td>
<td>7.67</td>
</tr>
<tr>
<td>1975</td>
<td>38.3</td>
<td>7.37</td>
</tr>
<tr>
<td>1976</td>
<td>41.2</td>
<td>7.53</td>
</tr>
<tr>
<td>1977</td>
<td>39.5</td>
<td>8.48</td>
</tr>
<tr>
<td>1978</td>
<td>39.5</td>
<td>8.88</td>
</tr>
<tr>
<td>1979</td>
<td>39.7</td>
<td>8.06</td>
</tr>
<tr>
<td>1980</td>
<td>40.4</td>
<td>7.18</td>
</tr>
<tr>
<td>1981</td>
<td>41.2</td>
<td>8.76</td>
</tr>
<tr>
<td>1982</td>
<td>42.3</td>
<td>6.65</td>
</tr>
<tr>
<td>1983</td>
<td>41.3</td>
<td>8.47</td>
</tr>
<tr>
<td>1984</td>
<td>41.7</td>
<td>7.85</td>
</tr>
<tr>
<td>1985</td>
<td>40.8</td>
<td>8.54</td>
</tr>
<tr>
<td>1986</td>
<td>42.2</td>
<td>9.17</td>
</tr>
<tr>
<td>1987</td>
<td>43.4</td>
<td>8.94</td>
</tr>
<tr>
<td>1988</td>
<td>41.7</td>
<td>9.35</td>
</tr>
<tr>
<td>1989</td>
<td>42.3</td>
<td>8.67</td>
</tr>
<tr>
<td>1990</td>
<td>42.3</td>
<td>9.47</td>
</tr>
<tr>
<td>1991</td>
<td>43.3</td>
<td>9.69</td>
</tr>
<tr>
<td>1992</td>
<td>44.6</td>
<td>7.52</td>
</tr>
<tr>
<td>1993</td>
<td>42.7</td>
<td>9.35</td>
</tr>
<tr>
<td>1994</td>
<td>43.4</td>
<td>9.76</td>
</tr>
<tr>
<td>Year</td>
<td>Value 1</td>
<td>Value 2</td>
</tr>
<tr>
<td>------</td>
<td>---------</td>
<td>---------</td>
</tr>
<tr>
<td>1995</td>
<td>42.2</td>
<td>9.98</td>
</tr>
<tr>
<td>1996</td>
<td>44.1</td>
<td>7.44</td>
</tr>
<tr>
<td>1997</td>
<td>44.6</td>
<td>9.45</td>
</tr>
<tr>
<td>1998</td>
<td>43.8</td>
<td>9.00</td>
</tr>
<tr>
<td>1999</td>
<td>45.9</td>
<td>9.52</td>
</tr>
<tr>
<td>2000</td>
<td>45.8</td>
<td>9.30</td>
</tr>
<tr>
<td>2001</td>
<td>46.5</td>
<td>8.51</td>
</tr>
<tr>
<td>2002</td>
<td>46.8</td>
<td>8.94</td>
</tr>
<tr>
<td>2003</td>
<td>48.1</td>
<td>9.84</td>
</tr>
<tr>
<td>2004</td>
<td>45.8</td>
<td>10.88</td>
</tr>
<tr>
<td>2005</td>
<td>46.3</td>
<td>9.13</td>
</tr>
<tr>
<td>2006</td>
<td>46.2</td>
<td>10.46</td>
</tr>
<tr>
<td>2007</td>
<td>46.7</td>
<td>9.59</td>
</tr>
<tr>
<td>2008</td>
<td>48.2</td>
<td>11.00</td>
</tr>
<tr>
<td>2009</td>
<td>47.8</td>
<td>9.92</td>
</tr>
<tr>
<td>2010</td>
<td>47.4</td>
<td>9.62</td>
</tr>
<tr>
<td>2011</td>
<td>47.0</td>
<td>10.53</td>
</tr>
<tr>
<td>2012</td>
<td>47.7</td>
<td>10.33</td>
</tr>
<tr>
<td>2013</td>
<td>47.0</td>
<td>10.03</td>
</tr>
<tr>
<td>2014</td>
<td>47.8</td>
<td>9.35</td>
</tr>
<tr>
<td>2015</td>
<td>48.5</td>
<td>9.33</td>
</tr>
<tr>
<td>2016</td>
<td>50.8</td>
<td>9.38</td>
</tr>
<tr>
<td>2017</td>
<td>47.6</td>
<td>9.84</td>
</tr>
</tbody>
</table>
Figure 73: Number of NABBS routes analysed per year. Some routes are not sampled annually due to volunteer availability, environmental conditions, or route discontinuation.

7.3.1.1 Species of interest

A total of 163 species were considered for analysis across the entire study area. 18 were classified as grassland obligate species, or species deemed highly sensitive to changes in amount and quality of grassland habitat. Of the grassland obligates, 13 were considered as declining species. An additional 49 species were classified as non-grassland obligate and declining (Table 71).

The total (Figure 78) and mean (Figure 79) number of birds counted within each species group was relatively constant across time in the Prairie Potholes and Eastern Tallgrass Prairie bird conservation regions, but fluctuation in stop totals appeared greater in the Badlands and Prairies BCR (Figures 79, 78). The latter BCR comprises a much smaller portion of the study area (Figure 71) and accordingly the coeffi-
Figure 74: Species richness increases over time across the entire study area.

Coefficient of variation (CV; ratio of deviation to the mean) around the stop totals (Figure 710) was highest in this region. It is worth noting the high CV (CV is considered low when $< \sim 40$) in all regions.

7.3.2 Statistical analysis

7.3.2.1 Identifying scaling structure in body mass distributions

Discontinuity analysis was conducted to identify the aggregations in the body mass distributions of 103 routes over a 50 period across the Central Great Plains (Figure 711). Discontinuity analysis suggested discontinuities existed in all routes analysed, and were relatively similar within NABBS routes over time. The number of body mass aggregations identified within each NABBS route using the discontinuity detector
Figure 75: Variance in the total species turnover increases over time across the entire study area.

(Barichevy et al., 2018) was similar across time (Figure 711a) and was approximately normally distributed across all survey-year combinations ($\bar{x} = 4.7, \sigma = 1.6$; Figure 711a). Species richness at the route level was strongly positively correlated with the number of aggregations (Figure 712). The distance to edge variable was statistically, but not strongly, correlated with body mass $r = -0.02, p =< 0.01$, and this relationship was similar across species groups except grassland obligates.

The discontinuities in the body mass distribution identified appeared relatively similar over time at most NABBS routes (Figure 713). If the shifting spatial regimes proposed in an earlier study [Roberts et al. (2019); Figure 72], then we should expect changes in the body mass distribution of NABBS falling within or near the regime boundary. This was not observed on the routes falling within this zone (Figure 713).
Figure 76: Declining species were closer to the edge of body mass aggregations in the ‘northern’ than were in the ‘southern’ regimes.

is representative of the ~5 NABBS locations falling in this area of expected changes shifting).

7.3.2.2 Linear mixed effects analysis of distance to edge

Declining species and declining grassland obligate species were located closer to the edge than the ‘other’ species, while grassland obligate species were further from the edge than ‘other’ species (Table 74). Similar trends held for declining grassland obligates and grassland obligates in the Northern regime location. There was, however, no evidence to suggest additive effects of the regime location or year, or their interactions (Table 74).
Figure 77: Declining species were closer to the edge of body mass aggregations in the 'northern' than were in the 'southern' regimes.

7.4 Discussion

South-North shifts in the past 50+ years have been demonstrated in large scale processes, including bird populations and ranges (Sorte & III, 2007) and plant hardening zones (Mckenney et al., 2014). A recent application of discontinuities in body mass distributions (Holling, 1992) corroborated these previous observations of large-scale changes as manifested in the avian community (Roberts et al., 2019). Support for the hypothesis that declining species operate at the edges of body mass aggregations, but did not identify changes in the body mass distributions of avian communities near the proposed spatial regime boundary (Roberts et al., 2019) proposed in impact of these proposed spatial regime shifts on local bird communities.

The lack of additive effects in the presence of multiplicative effects of the regime
location (Table 74) strengthens the support for the differences in grassland obligates and declining grassland obligates with respect to all other species, however, the confidence intervals around the estimates of declining grassland obligates, grassland obligates may suggest that the evidence for such an effect is relatively weak (Figure 77). This is unsurprising given the distribution of grassland obligate body masses is highly skewed right relative to the remaining species (Figures 714, 715). Many grassland obligate species have small body masses, reducing the probability that they will appear in different body mass aggregations, however, depending on the local bird community identities, local grassland obligates may all have similar body masses, removing this effect. The larger, non-declining grassland obligates occur in a relatively small portion of our study area, Shortgrass Prairie BCR (Figures 71, 716).
Although the body mass distributions of terrestrial communities at small scales tends to differ from those at larger spatial scales (Blackburn & Gaston, 1994), numerous studies confirm the evidence for ‘discontinuities’ in these distributions across multiple spatial scales for a variety of phenomenon (Allen et al., 1999; Wardwell & Allen, 2009; Nash et al., 2014; Spanbauer et al., 2016; but see Siemann & Brown, 1999; Manly, 1996 for applications yielding negative results; Bibi et al., 2019). This study found evidence supporting the hypothesis that scaling structures manifest in the body masses of fauna communities, adding to the growing number of case studies which suggest this phenomena is widespread across systems and scales.
Figure 710: Average number of birds across the entire study area per species group per year.

Table 74: Coefficient estimates for the linear mixed effects model predicting species’ ‘distance to edge’ of a body mass distribution.

<table>
<thead>
<tr>
<th></th>
<th>Estimate</th>
<th>CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>0.04</td>
<td>(0.02, 0.07)</td>
</tr>
<tr>
<td>Year</td>
<td>0.00</td>
<td>(0, 0)</td>
</tr>
<tr>
<td>North</td>
<td>0.01</td>
<td>(-0.01, 0.03)</td>
</tr>
<tr>
<td>Declining</td>
<td>-0.05</td>
<td>(-0.08, -0.02)</td>
</tr>
<tr>
<td>Declining Grassland Obligates</td>
<td>-0.17</td>
<td>(-0.2, -0.14)</td>
</tr>
<tr>
<td>Grassland Obligates</td>
<td>0.15</td>
<td>(0.11, 0.2)</td>
</tr>
<tr>
<td>Year x Declining</td>
<td>0.00</td>
<td>(0, 0)</td>
</tr>
<tr>
<td>North x Declining</td>
<td>0.02</td>
<td>(-0.01, 0.04)</td>
</tr>
<tr>
<td>North x Declining Grassland Obligates</td>
<td>-0.05</td>
<td>(-0.07, -0.03)</td>
</tr>
<tr>
<td>North x Grassland Obligates</td>
<td>0.04</td>
<td>(0.01, 0.07)</td>
</tr>
</tbody>
</table>
Figure 711: Number of body mass aggregations identified in each route unchanged across the time period (**a**) and is approximately normally distributed ($\bar{x} = 4.7, \sigma = 1.6$)
Figure 712: Relationship between species richness per route and (a) the number of aggregations identified in body mass distributions and (b) distance to the edge (units log body mass) of aggregations.
Figure 713: Aggregation locations of body mass distributions of the avian community at a single NABBS route (state 7 route 24) appear relatively similar across time. Panels include data for years 1=1970, 2=1985, 3=2000, 4=2015. Red triangles indicate grassland obligate species.
Figure 714: Body mass distribution for species in the study area over the entire time period varies by species group. Distributions represent the species pool for each group over the entire study area and all years.
Figure 715: The body mass distribution of declining species differ only slightly between the Southern and Northern regimes.
Figure 716: Body mass distribution for species in the study area over the entire time period varies by species group. Distributions represent the species pool for each group and Bird Conservation Region over all years.
Chapter 8

Conclusions

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 data set is analyzed using until a sufficient number of methods yield affirmative results.

\[
\text{Data} = \text{Information} = \text{Signal} = \text{Process} + \text{Noise}
\]


8.1 Method Mining

Many regime detection measures have yet to be statistically (or numerically) scrutinized for robustness or sensitivity to data quality, data quantitaty, and shift types. However, it should be noted that, in part due to both (i) the popularity and (ii) the sheer number of ‘new’ methods introduced by only a handful of authors\(^1\) 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) and are methods of measurement 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 (Chapter 2, Chapter 6, Taranu \textit{et al.}, 2018). Ecological systems are noisy, and the observational data we are collecting at large scales (e.g., the North American Breeding

\(^1\)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)
Bird survey), is noisy. Using methods incapable of identifying meaningful signals in noisy data appears futile, yet, methods for doing so are increasingly introduced in the scientific literature (Chapter 2).

8.3 Data Collection and Munging Biases Limit Inference

Regime detection measures and other ecological indicators can signal 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 analyzing data, the ecologist has defined the boundaries of the system *a priori* (Beisner *et al.*, 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 *et al.*, 2014 and velocity, Chapter 6), 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 4]

2. Normalization and weighting techniques often impact results (whether ecological or numerical) (Appendices 2 and 4)

3. Data aggregation techniques often impact results (Chapter 5)

4. Some indices fail to generalize across systems or taxa (see Chapters 1 and 2)
Appendix A bbsAssistant: an R package for Download and Munging Data and Information from the North American Breeding Bird Survey

This package is awaiting review at the Journal for Open Source Software. JLB is the creator and maintainer of the package. G. Palomo-Muñoz and Lyndsie Wszola are co-authors. Tentative doi: 10.21105/joss.01550

.1 Package Summary

This package contains functions for downloading and munging data from the North American Breeding Bird Survey (BBS) FTP server (Sauer et al., 2017; Pardieck et al., 2018). Although the FTP server provides a public interface for retrieving data and analysis results, this package consolidates the efforts of the data user by automating downloading and decompression of .zip data files, downloading route-level information, and saving them as .feather files for speedy import from disk. The data subsetting features of this package also allow the user to readily import and save to file only the data necessary for her purposes. Although the primary audience is for those wishing to use BBS data in Program R for visualization or analysis, this package can be used to quickly download the BBS data to file for use elsewhere.

The BBS team uses hierarchical modelling techniques to generate population trend estimates (Sauer et al., 2014) at various spatial scales see the BBS results webpage.
Given the variability in data availability, the BBS team also provides data credibility scores for species-regions combinations. This package contains two functions for retrieving the population trend estimates produced by Sauer et al. (2014) and the associated data credibility scores: a web-scraping function for obtaining current region and/or species-specific population trend estimates and data credibility scores via a supplied url, \texttt{get_credibility_trends()}; and a function for the current and archived population trends estimates for \textit{all} species and regions, \texttt{get_analysis_results()}.

\section{Package Vignette}

Package vignette is available at \url{https://github.com/TrashBirdEcology/bbsAssistant/tree/master/vignettes}

\section{Package Manual}

For functions and descriptions please see the manual at \url{github.com/bbsAssistant/tree/master/man}

\section{Acknowledgements}

The authors of this package thank the participatory scientists who collect data annually for the North American Breeding Bird Survey, and the Patuxent Wildlife Research Center for making these data publicly and easily accessible. Some functions in this package were adapted from the \texttt{rBBS} package and are mentioned in function source code as applicable.
Appendix B regimeDetectionMeasures: an R package for calculating various regime detection measures

.1 Package Summary

This package provides functionality for calculating numerous regime shift detection measures (see below). Additionaly, a ‘new’ metric, distance travelled and velocity are also calculated:

**Composite measures:**
1. Distance travelled -see also package `distanceTravelled`
1. Fisher Information
1. Variance Index

**Single-variable measures:**
1. Skewness (mean and mode versions)
1. Kurtosis
1. Variance
1. Mean
1. Mode
1. Coefficient of variation, CV
1. Autocorrelation lag-1 (using `stats::acf`)

.2 Package Manual

For functions and descriptions please see the manual at `github.com/bbsAssistant/tree/master/man`
Appendix C bbsRDM: an R package for applying regimeDetectionMeasures functionality to spatial data

The R package bbsRDM provides a template for using the North American Breeding Bird Survey to calculate various regime detection measures across space. This document provides a guide to to package functionality and results visualization. This package is not actively developed. Development source code for this package is available on GitHub as a compressed file, [https://github.com/TrashBirdEcology/bbsRDM/archive/master.zip](https://github.com/TrashBirdEcology/bbsRDM/archive/master.zip) or at [https://github.com/TrashBirdEcology/rRDM](https://github.com/TrashBirdEcology/rRDM).

.3 Package Vignette

Package vignette is available at [https://github.com/TrashBirdEcology/bbsRDM/tree/master/vignettes](https://github.com/TrashBirdEcology/bbsRDM/tree/master/vignettes). This example provides source code for analyzing the North American Breeding Bird Survey across regularly-sampled spatial transects at one or more points in time.

.4 Package Manual

For functions and descriptions please see the manual at [github.com/bbsRDM/tree/master/man](https://github.com/bbsRDM/tree/master/man).
Appendix D Functions used to calculate discontinuities in avian body mass distributions.

This code was first published in Barichievy et al. (2018) and has been slightly modified and annotated for the purposes of this dissertation. This code was used to produce body mass discontinuities in Chapter #ref(discontinuity)

.1 Neutral_Null function

Neutral_Null <- function(log10.data, resolution = 4000) { Dmax = max(log10.data, na.rm = FALSE) Dmin = min(log10.data, na.rm = FALSE) ds = (Dmax - Dmin) / resolution MaxK = (Dmax - Dmin) / 2 MinK = ds * 2

#define h’s to analyze ks = seq(MinK, MaxK, by = 1 / resolution)

# generate matrix bws = matrix(data = NA, nrow = length(ks), ncol = 1)

for (i in c(1:length(ks))) { # Calculate KS density estimate KSDens <- density(log10.data, bw = ks[i], “gaussian”, adjust = 1)

# Test if the ksdensity is unimodal

TF <- which(diff(sign(diff(KSDens$y))) == 2) + 1

if (length(TF) == 0)

  bws[i] = 1

else

  bws[i] = 0

}
# Define the neutral Null

```r
r = min(which(bws == 1))
null = ks[r]
return(null)
```

## 2 Bootstrapping Function

```r
DD <- function(log10.data, null, Sample.N = 1000) {
  NNull <- density(log10.data, bw = null, “gaussian”, adjust = 1)
  N <- length(log10.data)
  null.samples <- matrix(data = 0, ncol = Sample.N, nrow = N)
  for (i in 1:Sample.N) {
    rand.N <- sample(NNull, N, replace = TRUE, prob = NNull)
    null.samples[, i] <- sort(rand.N, decreasing = FALSE)
  }
  gaps.log10.data <- diff(log10.data)
  gaps.null.samples <- diff(null.samples, decreasing = FALSE)
  gaps.percentile <- matrix(data = 0, nrow = length(gaps.log10.data), ncol = 1)
  for (i in 1:length(gaps.log10.data)) {
    gap.percentile[i] <- ecdf(gaps.null.samples[i, ])(gaps.log10.data[i])
  }
  Bootstrap.gaps <- rbind(gap.percentile, 0)
  return(Bootstrap.gaps)
}```
References


Allen, C.R., Garmestani, A., Havlicek, T., Marquet, P.A., Peterson, G., Restrepo, C.,


Auguie, B. (2017) *GridExtra: Miscellaneous functions for ”grid” graphics*,


Bryan, J. (2017) *Gapminder: Data from gapminder*,

Burnett, J. (2019) *BbsRDM: Calculate regime detection metrics using bird communities,*


proach for detecting, quantifying, and visualizing the evolution of a research field: A practical application to the fuzzy sets theory field. *Journal of Informetrics*, **5**, 146–166.


deYoung, B., Barange, M., Beaugrand, G., Harris, R., Perry, R.I., Scheffer, M. & Werner, F. (2008) Regime shifts in marine ecosystems: Detection, prediction and


Hijmans, R.J. (2019) *Raster: Geographic data analysis and modeling,*


Iannone, R. (2018) *DiagrammeR: Graph/network visualization*,


Lanzante, J.R. (1996) Resistant, robust and non-parametric techniques for the analysis of climate data: Theory and examples, including applications to historical radiosonde station data. *International Journal of Climatology: A Journal of the*


Litzow, M.A. & Hunsicker, M.E. (2016) Early warning signals, nonlinearity, and signs of hysteresis in real ecosystems. Ecosphere, 7, –.


Mauget, S.A. (2003) Multidecadal regime shifts in us streamflow, precipitation, and
temperature at the end of the twentieth century. *Journal of Climate*, **16**, 3905–
3916.

May, R.M. (1977) Thresholds and breakpoints in ecosystems with a multiplicity of

information to the management of sustainable environmental systems. Exploratory
data analysis using fisher information*, pp. 217–244. Springer.


Mckenney, D.W., Pedlar, J.H., Lawrence, K., Papadopol, P., Campbell, K. & Hutchin-


parison of cusum and shiryaev–roberts procedures for detecting changes in distrib-

around the globe: Theory, drivers and impacts.

Mumby, P.J., Steneck, R.S. & Hastings, A. (2013) Evidence for and against the

Müller, K. (2017) *Here: A simpler way to find your files,*


and scale. *Ecosystems, 1*, 6–18.


Tuszynski, J. (2019) *CaTools: Tools: Moving window statistics, gif, base64, roc auc, etc.*,.


Westgate, M.J. (2018b) *revtools: Tools to support evidence synthesis*,


Wickham, H. (2017) *Tidyverse: Easily install and load the 'tidyverse'*,


