

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 Craig R. Allen

Lincoln, Nebraska

August, 2019



# Preface

“Ecological models are made for ecologists, by ecologists, cooked up for consumption by ecologists” — unknown

# Dedication

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

# Acknowledgements

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.

Financial support. 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

<b>Table of Definitions . . . . .</b>	<b>2</b>
<b>Chapter 1: Introduction . . . . .</b>	<b>11</b>
1.1 Dissertation Aims . . . . .	13
1.2 Dissertation Structure and Contents . . . . .	13
1.3 Appendices and Software . . . . .	16
<b>Chapter 2: Methods for Detecting Ecological Regime Shifts . . . . .</b>	<b>17</b>
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
2.4 A synthesis of the methods available for the practical ecologist . . .	35
2.4.1 Model-dependent . . . . .	36
2.4.2 Model-free . . . . .	37
2.5 Discussion . . . . .	37
<b>Chapter 3: Deconstructing the steps for calculating Fisher Information as a measure of abrupt change in ecological systems . . . . .</b>	<b>43</b>
3.1 Abstract . . . . .	43
3.2 Introduction . . . . .	44
3.2.1 Fisher Information as a Regime Detection Method . . . . .	45
3.2.2 The Sustainable Regimes hypothesis . . . . .	46
3.2.3 Fisher Information requires dimension reduction prior to calculation . . . . .	47
3.2.4 Aims . . . . .	48
3.3 Methods . . . . .	49
3.3.1 Predator-prey model system . . . . .	49
3.3.2 Inducing a Regime Shift . . . . .	49
3.3.3 Decoupling the Steps for Calculating Fisher Information . . .	50
3.4 Discussion . . . . .	54
3.5 Acknowledgements . . . . .	56

<b>Chapter 4: An application of Fisher Information to bird community data does not reveal distinct regimes in the continental United States</b>	<b>63</b>
4.1 Introduction . . . . .	63
4.2 Data and Methods . . . . .	64
4.2.1 Data: North American breeding bird communities . . . . .	64
4.2.2 Study area . . . . .	66
4.2.3 Calculating Fisher Information (FI) . . . . .	69
4.2.4 Interpreting and comparing Fisher Information across spatial transects . . . . .	72
4.3 Results . . . . .	74
4.3.1 Fisher Information across spatial transects . . . . .	74
4.3.2 Spatial correlation of Fisher Information . . . . .	78
4.4 Discussion . . . . .	79
<b>Chapter 5: Velocity (<math>v</math>): Rate of Change of System Trajectory Identifies Abrupt Changes in Ecological Systems</b>	<b>87</b>
5.1 Introduction . . . . .	87
5.1.1 Tracking ecosystem trajectory through time to explore system dynamics . . . . .	89
5.1.2 Rate of change as an indicator of abrupt change in the system trajectory . . . . .	90
5.1.3 Aims . . . . .	91
5.1.4 Analytical approach . . . . .	92
5.2 Steps for Calculating Velocity, $V$ . . . . .	92
5.2.1 Step 1: Calculate $\Delta x_i$ . . . . .	94
5.2.2 Step 2: Calculate distance traveled, $s$ . . . . .	95
5.2.3 Step 3: Calculate velocity, $V$ (or $\frac{\Delta s}{\Delta t}$ ) . . . . .	96
5.3 Velocity $v$ Performance Under a Discontinuous Transition . . . . .	97
5.4 Velocity Performance Under a Smooth Transition . . . . .	99
5.4.1 Generating the data . . . . .	100
5.4.2 Evaluating Velocity Performance Under Conditions of Changing Means and Variance . . . . .	102
5.5 Velocity Performance Under Empirical Transitions: paleolithic freshwater diatom communities . . . . .	106
5.6 Discussion . . . . .	107
5.7 Supplementary Figures . . . . .	109
<b>Chapter 6: Using Resampling Methods to Evaluate the Relative Performance of Regime Detection Measures</b>	<b>135</b>
6.1 Introduction . . . . .	135
6.2 Data and Methodology . . . . .	138
6.2.1 Study system and data . . . . .	138
6.2.2 Regime detection measures . . . . .	138
6.2.3 Simulating data quality and quantity issues using resampling techniques . . . . .	141

6.2.4	Comparing regime detection measures . . . . .	142
6.3	Results . . . . .	143
6.3.1	Velocity of the distance travelled ( $v$ ) . . . . .	143
6.3.2	Variance Index . . . . .	143
6.3.3	Fisher Information is highly sensitive to information loss . . . . .	144
6.4	Detrending the Data Prior to Calculations . . . . .	144
6.5	Conclusion . . . . .	145
6.6	Acknowledgements . . . . .	146
<b>Chapter 7: Grassland Obligates and Declining Birds Operate Near Edges of Body Mass Distributions</b>	. . . . .	<b>155</b>
7.1	Introduction . . . . .	155
7.2	Methods . . . . .	157
7.2.1	Study area . . . . .	157
7.2.2	Data . . . . .	158
7.2.3	Statistical analysis . . . . .	163
7.3	Results . . . . .	169
7.3.1	Summary statistics of censuses (NABBS data) . . . . .	171
7.3.2	Statistical analysis . . . . .	175
7.4	Discussion . . . . .	177
<b>Chapter 8: Conclusions</b>	. . . . .	<b>188</b>
8.1	Method Mining . . . . .	189
8.2	Ecological Data are Noisy . . . . .	190
8.3	Data Collection and Munging Biases Limit Inference . . . . .	190
8.4	Common Limitations of Regime Detection Measures . . . . .	191
<b>Appendix A bbsAssistant: an R package for Download and Munging Data and Information from the North American Breeding Bird Survey</b>	. . . . .	<b>192</b>
.1	Package Summary . . . . .	192
.2	Package Vignette . . . . .	193
.2.1	Installing package and loading dependencies . . . . .	193
.2.2	Downloading the BBS data from USGS FTP . . . . .	194
.2.3	Import BBS data from file into R . . . . .	196
.2.4	Subsetting the BBS count data . . . . .	198
.2.5	Retrieve BBS analysis results and data credibility measures . . . . .	199
.3	Package Manual . . . . .	204
.4	Acknowledgements . . . . .	204
<b>Appendix B regimeDetectionMeasures: an R package for calculating various regime detection measures</b>	. . . . .	<b>205</b>
.1	Example Analysis . . . . .	205
.1.1	Measures/metrics calculated . . . . .	205
.1.2	Calculate Metrics analysis . . . . .	206

.1.3	Plots . . . . .	206
<b>Appendix C bbsRDM: an R package for applying regimeDetectionMeasures functionality to spatial data . . . . .</b>	<b>208</b>	
.1	Example Analysis . . . . .	208
.1.1	Load packages & create local directories . . . . .	208
.1.2	Download the BBS data and save to file locally . . . . .	210
.1.3	Create a sampling grid . . . . .	211
.1.4	Subset the BBS data by species and/or functional traits (OPTIONAL but highly recommended) . . . . .	213
.1.5	Subset species according to AOU species codes (i.e. by family, genera, etc..) . . . . .	213
.1.6	Subset species by trait, body mass, taxonomically, etc . . . (optional)	214
.1.7	Calculate regime detection metrics across space or time . . . .	214
.1.8	Conduct analysis . . . . .	216
.1.9	Import and munge the results to prepare for visualization . . .	218
.1.10	Visualize results: one regime detection metric at a time . . . . .	220
<b>Appendix D Functions used to calculate discontinuities in avian body mass distributions. . . . .</b>	<b>222</b>	
.1	About . . . . .	222
.2	Neutral.Null function . . . . .	222
.3	Bootstrapping Function . . . . .	223
<b>References . . . . .</b>	<b>224</b>	

# List of Tables

1	A table of definitions for terms, theories, and phrases often appearing in ecological regime shift literature. . . . .	3
2.1	List of the regime detection methods identified using expert knowledge and previously published reviews of methods. . . . .	23
2.2	Potential questions for a comprehensive review of the ecological regime detection metrics literature. . . . .	41
5.1	Steps outlined for calculating system velocity, $V$ , using the 2-variable toy data as an example. . . . .	98
5.2	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. . . . .	101
7.1	Grassland obligates and species with declining trends over the period of (1966-2015) in the Central Breeding Bird Survey region in our study area. . . . .	165
7.2	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. . . . .	169
7.3	Summary statistics for annual species richness and annual turnover in all NABBS routes in study area. . . . .	171
7.4	Coefficient estimates for the linear mixed effects model predicting species' 'distance to edge' of a body mass distribution. . . . .	180

# List of Figures

2.1	Number of articles containing phrases related to ecological 'regime shifts' published per publication outlet. . . . .	21
2.2	Number of publications by year in fields 'Ecology' and 'Biodiversity Conservation' which included terms related to 'regime shift' (total = 654). . . . .	22
2.3	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. . . . .	28
2.4	The themes identified in the literature based on author-supplied keywords vary as the number of words included in the analysis increased. . . . .	30
2.5	The themes identified in the literature based on the ISI-supplied keywords vary as the number of words included in the analysis increased. . . . .	31
2.6	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.	32
2.7	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. . . . .	33
2.8	Total number of articles published per year by category as categorized by ISI. Book chapters, proceedings, editorials, and letters are excluded.	34
3.1	From top to bottom, distance traveled in phase space, speed tangential to system trajectory, acceleration tangential to system trajectory. . .	57
3.2	Carrying capacity over time with a regime shift occurring around time 200. . . . .	58
3.3	Phase space plot of system trajectories for different values of k. Colored point indicates the coordinates corresponding to the shift induced in carrying capacity. . . . .	59
3.4	Distance travelled in phase space over time. Dashed vertical line at time 200 indicates location of regime shift. . . . .	60
3.5	Speed of the system (rate of change, velocity) in phase space. Dashed vertical line at time 200 indicates location of regime shift. . . . .	61
3.6	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. . . . .	62

4.1	Locations of Breeding Bird Survey routes sampled between 1966 and 2017. . . . .	65
4.2	Locations of focal U.S. military bases, Eglin Air Force Base (AFB) and Fort Riley Military Base. . . . .	67
4.3	Three East-West running transects analyzed in this Chapter overlayed against the Omernick Ecoregion boundaries map. . . . .	68
4.4	The change in the Fisher Information values along a single, East-West-running spatial transect (Transect number 12) over time. . . . .	75
4.5	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.	76
4.6	Fisher Information of select East-West spatial transects over time. These transects are representative of the results obtained from the remaining transects. . . . .	77
4.7	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.	78
4.8	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. . . . .	84
4.9	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.	85
4.10	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. . . . .	86
5.1	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. . . . .	88
5.2	An example solution of the Lorenz ('butterfly') represented in individual system components. . . . .	89
5.3	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, $\sigma = 5$ . . . . .	93
5.4	Data used to calculate velocity at the first two time points, $t_1$ and $t_2$ .	94

5.5	Distance traveled, $s$ , for the 2-species toy system. . . . .	96
5.6	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, $\sigma = 5$ . . . . .	97
5.7	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 ( $N_{iter} = 1$ , 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 $t < 50$ . . . . .	110
5.8	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 $t < 50$ . . . . .	111
5.9	High variance of velocity ( $V$ ) in a single iteration ( $N_{iter} = 1$ , seed = 123) of simulations as we increase $\sigma_1$ at $t = 50$ . . . . .	112
5.10	The variance around the average ( $\pm 2$ SD) velocity ( $V$ ) increases as the variance of $\bar{x}_{2_{t=50(post)}}$ (post shift) increases. $\bar{x}_{1_{pre}} = 25$ , $\bar{x}_{1_{post}} = 100$ , $\bar{x}_{2_{pre}} = 25$ , $\bar{x}_{2_{post}} = 50$ , $\sigma_{1_{pre}} = 5$ , $\sigma_{2_{pre,post}} = 5$ . . . . .	113
5.11	The noise in system velocity ( $V$ ) is not obviously reduced in this system as the original data ( $x_1$ , $x_2$ ) is increasingly smoothed. . . . .	114
5.12	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: $\alpha = 1.00$ ). . . . .	115
5.13	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: $\alpha = 1.00$ ). . . . .	116
5.14	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 $\alpha \ll$ sufficed. 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 ( $\alpha_{tanh} = 1$ ) for the 'tvdiff' $\alpha = 0.50$ . Not pictured: the antidifferentiated values of $s$ (observed) is increasingly smoothed as $\alpha$ increases. . . . .	117
5.15	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). . . . .	118

5.16 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). . . . .	119
5.17 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). . . . .	120
5.18 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). . . . .	121
5.19 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). . . . .	122
5.20 The velocity signals a rapid shift in the variance of both state variables 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). . . . .	123
5.21 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). . . . .	124
5.22 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). . . . .	125
5.23 The velocity signals a shift when both variables undergo shifts in the mean and variance under a slightly abrupt transition ( $\alpha_{tanh} = 0.25$ ). True and observed values of $x_i$ (panel A), observed distance traveled ( $s$ , panel B), observed velocity (C), and the smoothed velocity (D). . . . .	126
5.24 The velocity signals a shift when both variables undergo shifts in the mean and variance under a slightly abrupt transition ( $\alpha_{tanh} = 1.00$ ). True and observed values of $x_i$ (panel A), observed distance traveled ( $s$ , panel B), observed velocity (C), and the smoothed velocity (D). . . . .	127
5.25 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. . . . .	128
5.26 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 $V$ . . . . .	129

5.27	Inter-regime (regimes A and B) trends in the velocity signal ( $V$ ) identify the fluctuating decadal and centuriel abiotic conditions discussed in Spanbauer et al. (2014). . . . .	130
5.28	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, $s$ . . . . .	131
5.29	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). . . . .	132
5.30	System change ( $s$ ) and velocity ( $V$ ) of the model system over the time period. Change in means ( $\bar{x}_{1_{pre}} = 25$ , $\bar{x}_{1_{post}} = 100$ , $\bar{x}_{2_{pre}} = 50$ , $\bar{x}_{2_{post}} = 10$ ) and an increase in variance ( $\sigma_{1_{pre}} = 2$ , $\sigma_{1_{post}} = 10$ , $\sigma_{2_{pre}} = 5$ , $\sigma_{2_{post}} = 10$ ). . . . .	133
5.31	System change ( $s$ ) and velocity ( $V$ ) of the model system over the time period. Constant means ( $\bar{x}_1 = 25$ , $\bar{x}_2 = 50$ ) and sharp change in variance for one state variable $\sigma_{1_{pre}} = 2$ , $\sigma_{1_{post}} = 12$ , $\sigma_{2_{pre,post}} = 5$ . . .	134
5.32	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 ( $/bar{x}_{1_{pre}} = 25$ , $/bar{x}_{1_{post}} = 50$ , $/bar{x}_{2_{pre}} = 15$ , $/bar{x}_{2_{post}} = 150$ ). . .	134
6.1	Relative abundances of the diatom species in Foy Lake over the time period. . . . .	147
6.2	The amount of time elapsed between observations for the Foy Lake paleodiatom data. . . . .	148
6.3	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. . . . .	149
6.4	Mean Variance Index (VI) and associated 95% confidence intervals over 10,000 iterations using the <b>observations</b> resampling method. Red line indicates the value of VI when <b>M</b> and <b>P</b> are 100%. . . . .	149
6.5	Mean Variance Index (VI) and associated 95% confidence intervals over 10,000 iterations using the <b>species</b> resampling method. Red line indicates the value of VI when <b>M</b> and <b>P</b> = 100%. . . . .	150
6.6	Mean velocity and associated 95% confidence intervals over 10,000 iterations using the observations resampling method. Red line indicates the value of velocity when <b>M</b> and <b>P</b> are 100%. . . . .	150
6.7	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 <b>M</b> and <b>P</b> = 100%. . . . .	151
6.8	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.	151

6.9	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.	152
6.10	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 <b>M</b> and <b>P</b> are 100%. A very small value was added to the mean FI prior to log transformation.	153
6.11	Local regression (loess) smoothing of a dominant species in the paleodiatom community, <i>Anomoeoneis costata</i> varies with the span parameter, making it difficult to justify smoothing the data prior to calculating various regime detection metrics. . . . .	154
7.1	My study area (red box) overlaying the terrestrial Breeding Conservation Regions (BCR) in North America. . . . .	157
7.2	North American Breeding Bird Survey routes (points), latitudinal locations (horizontal bars) of the spatial regimes identified by roberts2019shifting. . . . .	162
7.3	Number of NABBS routes analysed per year. Some routes are not sampled annually due to volunteer availability, environmental conditions, or route discontinuation. . . . .	174
7.4	Species richness increases over time across the entire study area. . . .	175
7.5	Variance in the total species turnover increases over time across the entire study area. . . . .	176
7.6	Declining species were closer to the edge of body mass aggregations in the 'northern' than were in the 'southern' regimes. . . . .	177
7.7	Declining species were closer to the edge of body mass aggregations in the 'northern' than were in the 'southern' regimes. . . . .	178
7.8	Total number of birds across the entire study area per species group per year. . . . .	179
7.9	Average coefficient of variation in the number of birds across the entire study area per species group per year. . . . .	180
7.10	Average number of birds across the entire study area per species group per year. . . . .	181
7.11	Number of body mass aggregations identified in each route unchanged across the time period (**a**) and is approximately normally distributed ( $\bar{x} = 4.7, \sigma = 1.6$ ) . . . . .	182
7.12	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. . . . .	183
7.13	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. . .	184

7.14 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. . . . .	185
7.15 The body mass distribution of declining species differ only slightly between the Southern and Northern regimes. . . . .	186
7.16 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. . . .	187

# Abstract

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.





# Table of Definitions

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

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

Term	Definition	Synonyms
Abrupt	A relative value of the speed and/or intensity of the change; the time period over which the regime shift occurs relative to the time observed (or expected to have been) in a particular state.	big, fast, quick, large
Alternative Stable State	<b>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).</b>	
Attractor	The set of values towards which a system tends regardless of its initial (starting) values.	
Basin-Boundary Collision	<b>The parameter values for a system that causes the system to shift between alternate attractors.</b>	non-local bifurcation
Catastrophe Theory	The study of abrupt changes within a dynamical system.	
Catastrophic Bifurcation	<b>A relatively abrupt jump to an alternate attractor due to initial attractor.</b>	
Change-Point	See also 'Regime Shift'. A term often used in computer science, climatology, data science; represents the point at which a state changes its configuration.	

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

Term	Definition	Synonyms
Change-Point	<b>A change point method which does not require supervision; identifies potential change points without a priori potential change points.</b>	
Detection		
Change-Point Estimation	A change point method which DOES require supervision; identifies potential change points when given a set of potential change points; well-developed in computer science, statistics, data mining, etc.; although well-developed, still lacks with giving statistical significance of change-points.	
Chaos	<b>A system with extreme sensitivity to initial conditions.</b>	
Critical Slowing Down (CSD)	When the recovery rate (time to return) of a system decreases (approaches zero) as a system approaches a critical point (possibly a threshold or tipping point). A characteristic observed in some empirical systems data (e.g. nutrient loading in shallow lakes).	
Degrees of Freedom	<b>The number of system parameters or components which vary independently.</b>	

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

Term	Definition	Synonyms
Domain of Attraction	The range of values around which a system fluctuates.	zone of fluctuation, basin of attraction, stable point, attractor
<b>Driver</b>	<b>A widespread anthropogenic source of change which leads to one or more pressures (e.g., land-use change).</b>	
Driver-Threshold Regime Shift	When a rapid change in external driver induces a rapid change in ecosystem state.	
<b>Dynamical System</b>	<b>A time-dependent system which can be described in state-space.</b>	
Dynamical Systems Theory	The study of complex systems theory; the study of time-dependent systems.	
<b>Equilibrium</b>	<b>The set of values around which a system revolves and does not change.</b>	
Exogeneous Process (Forcing, Driver)	An external process influencing the state of the dynamical system.	

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

Term	Definition	Synonyms
First-Order	<b>When the mean is constant over the observations.</b>	
Stationarity		
Fold Bifurcation	This occurs when a stable point collides with an unstable point; when crossing a tipping point induces hysteresis.	
Fractal Properties	<b>A measurement of geometrical self-similarity; when a system has similar structure regardless of the scale of observation.</b>	ergodic
Hysteresis	A system which is state-dependent (e.g. magnets); when a tipping point or threshold is crossed such that the previous state cannot be achieved by reversing the conditions.	
Leading Indicators	<b>When the statistical properties of the fluctuations (of the data) approach a critical transition.</b>	
Lyapunov Exponent (and Stability)	A value that conveys the average rate of trajectory divergence that is caused by an endogenous force; how quickly (if at all) a system will tend away from a stable point if it starts near the stable point.	
Measure Theory	<b>The study of measures and measurement (e.g. volume, mass, time).</b>	

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

Term	Definition	Synonyms
Moving (Sliding) Window Analysis	When a subsample of the data $\$X_t\$$ is used in lieu of a single observation, $\$x_t\$$ .	
<b>Noise</b>	<b>Processes manifested in data which are unaccounted for; sometimes referred to as meaningless; random variability.</b>	
Non-Stationarity of the Mean Value	Infers that a trend or a periodicity is present in the time series.	
<b>Online</b>	<b>Real-time updating of model parameters, predictions, etc. (c.f. offline).</b>	
Persistent	A relative value of the longevity of the observed change in values.	long-lasting
<b>Phase Space</b>	<b>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.</b>	
Prediction	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.	
<b>Pressure</b>	<b>A perturbation which negatively influences a system, and can be defined as pulse, press, or monotonic.</b>	

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

Term	Definition	Synonyms
Red Noise	Noise having zero mean, constant variance, and serial autocorrelation; autocorrelated random variability.	
Regime	<b>A set of system values that define a particular system state. Not necessarily stable, but some state variables or outputs of the system remain relatively constant over a defined period of time.</b>	
Regime Shift	"abrupt" and "persistent" change in a system's structure or functioning.	
Second-Order	<b>The mean is constant and the covariance is a function of a time lag,</b>	
Stationarity	<b>but not of time.</b>	
Self-Similarity	A system satisfied by power-law scaling.	
Stable Equilibrium	<b>An equilibrium is stable when small perturbations do not induce change.</b>	
State Space	The set of all possible configurations of a system.	
State-Threshold	<b>When a gradual change in external driver induces a rapid change in ecosystem state (e.g., System crosses a threshold).</b>	
Regime Shift		
Stationarity	When the probability density function of a system does not change with time.	

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

Term	Definition	Synonyms
Statistical Stationarity	A system with statistical properties unchanging over time. This concept extends to periodic stationarity for systems exhibiting periodic behavior.	
Strange Attractor	An attractor which has fractal structure (an observable fractal dimension).	
<b>Supervised Machine Learning</b>	<b>When classifiers are used to train the data a priori.</b>	
System State	The observed (current) instance of the system within a state space.	
<b>Threshold</b>	<b>A point where the system reacts to changing conditions.</b>	
Tipping Point	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.	
<b>Trajectory</b>	<b>The path of an object or system through space-time.</b>	orbit, path
Transient	A behavior or phenomenon which is responsive to initial (starting) conditions, or its effect declines over time.	
<b>Trend Smoothing</b>	<b>Local averaging of values such that the non-systematic components of the system are washed out.</b>	

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

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

# Chapter 1

## Introduction

Anthropogenic activity in the last few decades will continue to influence the 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 .1.3). 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 5.** 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 6.** 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 5) 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/trashbirdecolgy](http://www.github.com/trashbirdecolgy). A final Appendix contains the code used to conduct the discontinuity analysis used in Chapter 7

**R package:** `bbsAssistant`. Appendix 8.4 is contains a vignette for the self-authored R package, written to facilitate retrieval and munging of information and data from the North American Breeding Bird Survey (U.S. Geological Survey) FTP server and website.

**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 .1.3 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 6. 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 creating 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, 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), 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 (2005a). 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 distinct 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

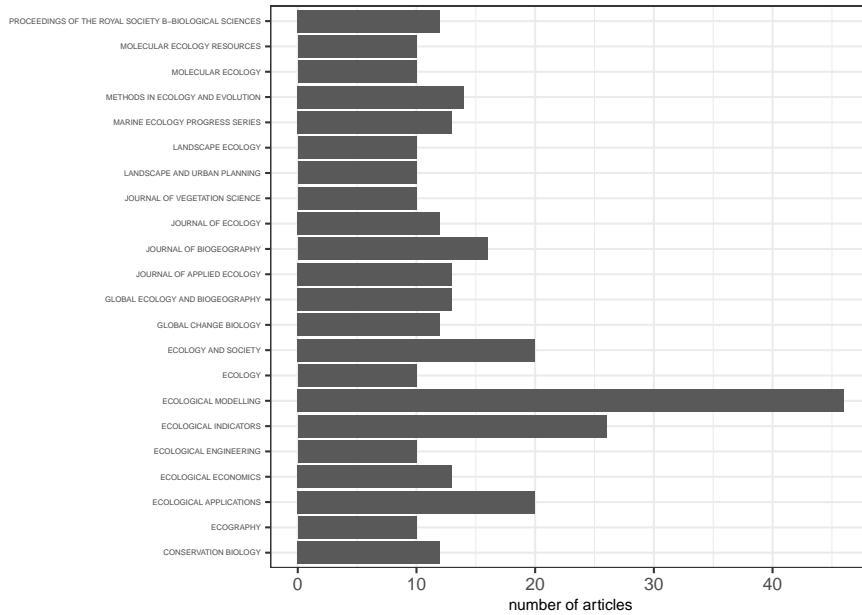
to, existing methods (Zhou & Shumway, 2008; Salehpour *et al.*, 2011; Byrski & Byrski, 2016). However Nicholls *et al.* (2011) for an addition of variable optimization to the method in Nicholls (2011) that was not included in my comprehensive list. Next, the list excludes papers which propose a combination 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*:

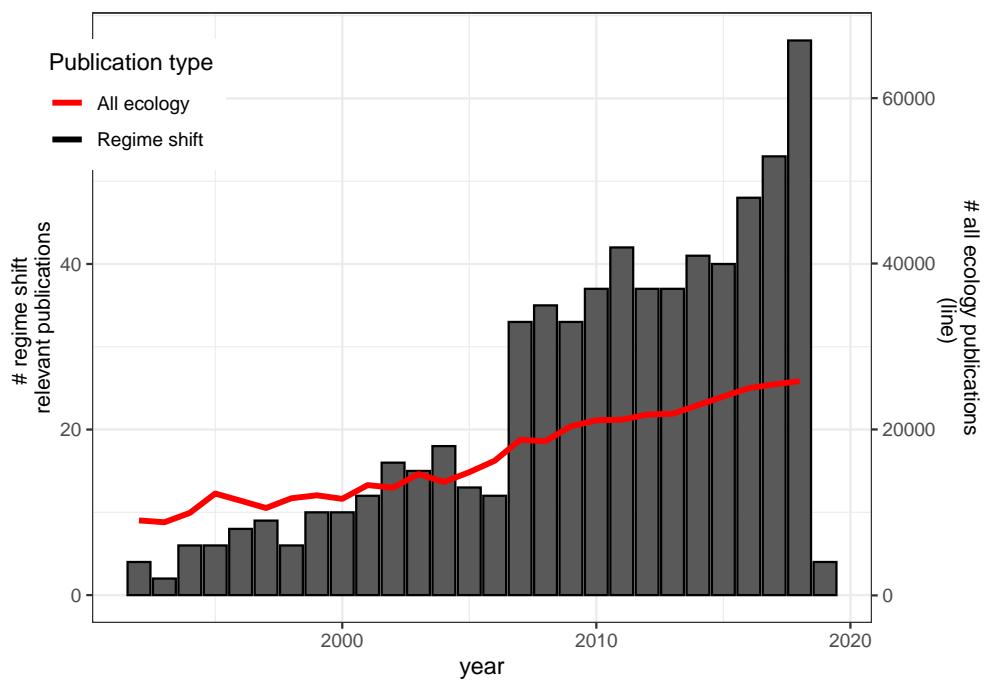
TS=(“regime shift” OR “regime shifts” OR “regime change” OR “regime changes” OR “catastrophic change” OR “catastrophic shift” OR “catastrophic changes” OR “catastrophic shifts” OR “sudden change” OR “sudden changes” OR “abrupt shift” OR “abrupt shifts” OR “abrupt change” OR “abrupt changes”) AND WC=(“Ecology” 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,



**Figure 2.1:** *Number of articles containing phrases related to ecological 'regime shifts' published per publication outlet.*

2017). This package contains function wrappers for conducting and visualizing network analyses based on keyword, authorship, and citation data. In an attempt 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 2.2:** Number of publications by year in fields 'Ecology' and 'Biodiversity Conservation' which included terms related to 'regime shift' (total = 654).

**Table 2.1:** List of the regime detection methods identified using expert knowledge and previously published reviews of methods.

Method	Metric type	Source
Average standard deviates	metric	Ebbesmeyer (1991)
BDS test	metric	Carpenter and Brock (2011)
Conditional heteroskedasticity	metric	Seekell et al. (2011)
Conditional probability analysis	metric	Zurlini et al. (2014)
Cumulative deviation test (CUSUM)	metric	Buishand (1982)
Degenerate Fingerprinting	metric	Kleinen et al. (2003)
Discontinuity analysis	metric	Peterson et al. (1998)
Downton-Katz test	metric	Karl et al. (1987)
First-order multivariate autoregressive models (MAR1)	metric	Ives et al. (2003)
Fisher Information	metric	Fath et al. (2003)
Intervention Analysis	metric	Francis (1994)
Inverse of AR(1) coefficient, variance, etc.	metric	Carpenter et al. (2008)

**Table 2.1:** List of the regime detection methods identified using expert knowledge and previously published reviews of methods. (continued)

Method	Metric type	Source
Kurtosis	metric	Biggs et al. (2009)
Lanzante method	metric	Lanzante et al. (1996)
LePage test	metric	Yonetani (1993)
Mann-Kendall test	metric	Goossens et al. (1987)
Mann-whitney U-test	metric	Mauget et al. (2003)
method-fuzzy synthetic evaluation (FSE)	metric	Wang et al. (2011)
Moving detrended fluctuation analysis (MDFA)	metric	He et al. (2008)
Nearest-neighbor statistics	metric	Pawlowski et al. (2008)
Oerleman's method	metric	Oerlemans (1978)
Pettitt test	metric	Pettit (1979)
Probability density function entropy method	metric	Pawlowski et al. (2008)

**Table 2.1:** List of the regime detection methods identified using expert knowledge and previously published reviews of methods. (continued)

Method	Metric type	Source
Quickest detection method (Shiryaev-Roberts statistic)	metric	Moustakides et al. (2009)
Rodionov method	metric	Rodionov (2005)
Sequential t-tests	metric	Rodionov (2004)
Skewness	metric	Guttal et al. (2008)
Spectral density ratio indicator	metric	Biggs et al. (2009)
Stability Index of the Ecological Units	metric	Parparov et al. (2017)
Standard deviation (rising variance)	metric	Carepenter et al. (2006)
Standard normal homogeneity	metric	Alexandersson et al. (1986)
STARS	metric	Buishand (1982)
T-test	metric	Ducre (2003)
Threshold Indicator Taxa ANalysis (TITAN)	metric	Baker et al. (2010)
Variance Index	metric	Brock et al. et al. (2006)

**Table 2.1:** List of the regime detection methods identified using expert knowledge and previously published reviews of methods. (continued)

Method	Metric type	Source
Vector-autoregressive method	metric	Mantua (2004)
Wilcoxon rank-sum	metric	Karl et al. (1987)
Zonal thresholding	metric	Yin et al. (2017)
two-phase regression	metric of a model	Easterling (1995)
Bayesian approaches	model	Jo et al. (2016)
Convex model	model	Qi et al. (2016)
Free-knot splines & piecewise linear modelling	model	Gal et al. (2010)
Generalized model	model	Lade (2012)
Multivariable autoregressive models (MAR1)	model	Ives et al. (2012)
Nonparametric drift-diffusion-jump model	model	Carpenter and Brock (2011)
Pettitt test and the Sen test	model	Vicent et al. (1998)
Potential analysis	model	Ives et al. (2012)

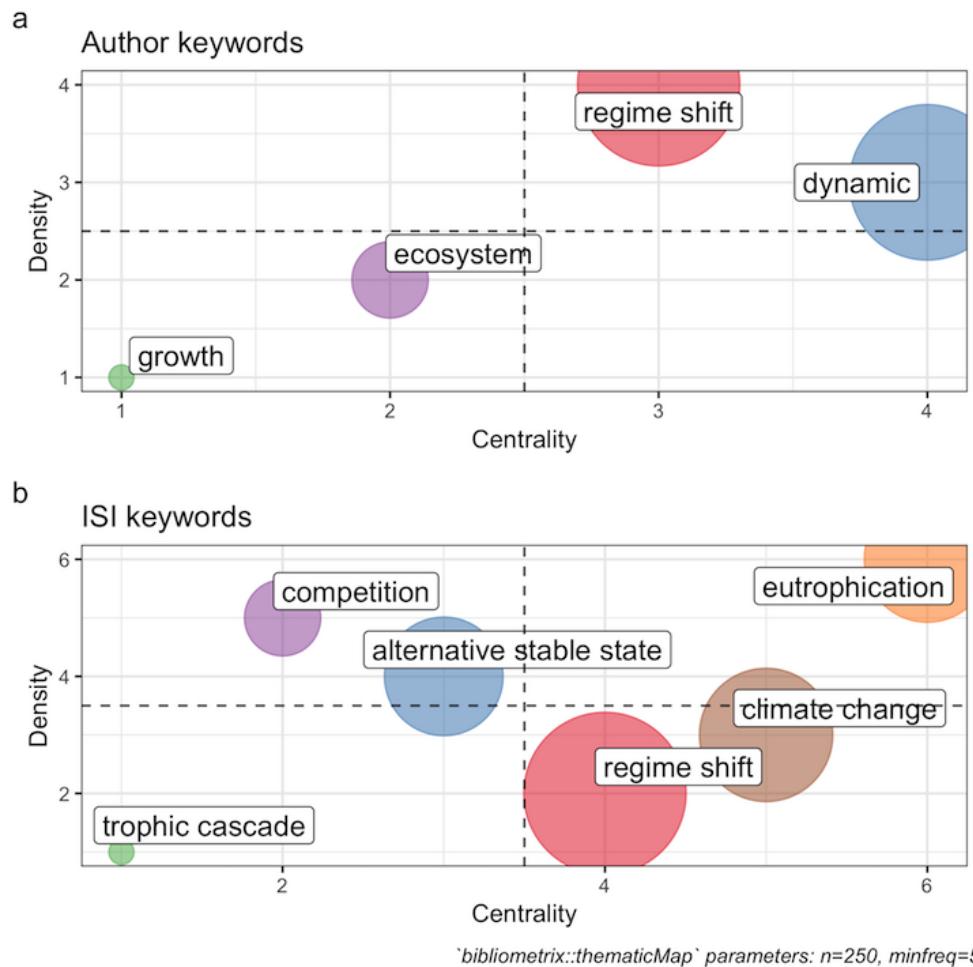
**Table 2.1:** List of the regime detection methods identified using expert knowledge and previously published reviews of methods. (continued)

Method	Metric type	Source
Regression-based models	model	Solow and Beet (1987)
Self-exciting threshold autoregressive state-space model SETARSS(p)	model	Tong (1990)
shiftogram	model	Groger et al. (2011)
Smooth transition autoregressive model	model	Gal et al. (2010)
Online dynamic linear modelling + time_varying autoregressive state_space models (TVARSS)	models	Parparov et al. (2017)
Fourier Analysis	other	Carpenter et al. (2010)
Vector-autoregressive method	other	Solow and Beet et al. (2005)
Wavelet analysis (decomposition)	other	Cazellas et al. (2008)

## 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 2.1) based on expert knowledge and previously published review articles.



**Figure 2.3:** 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

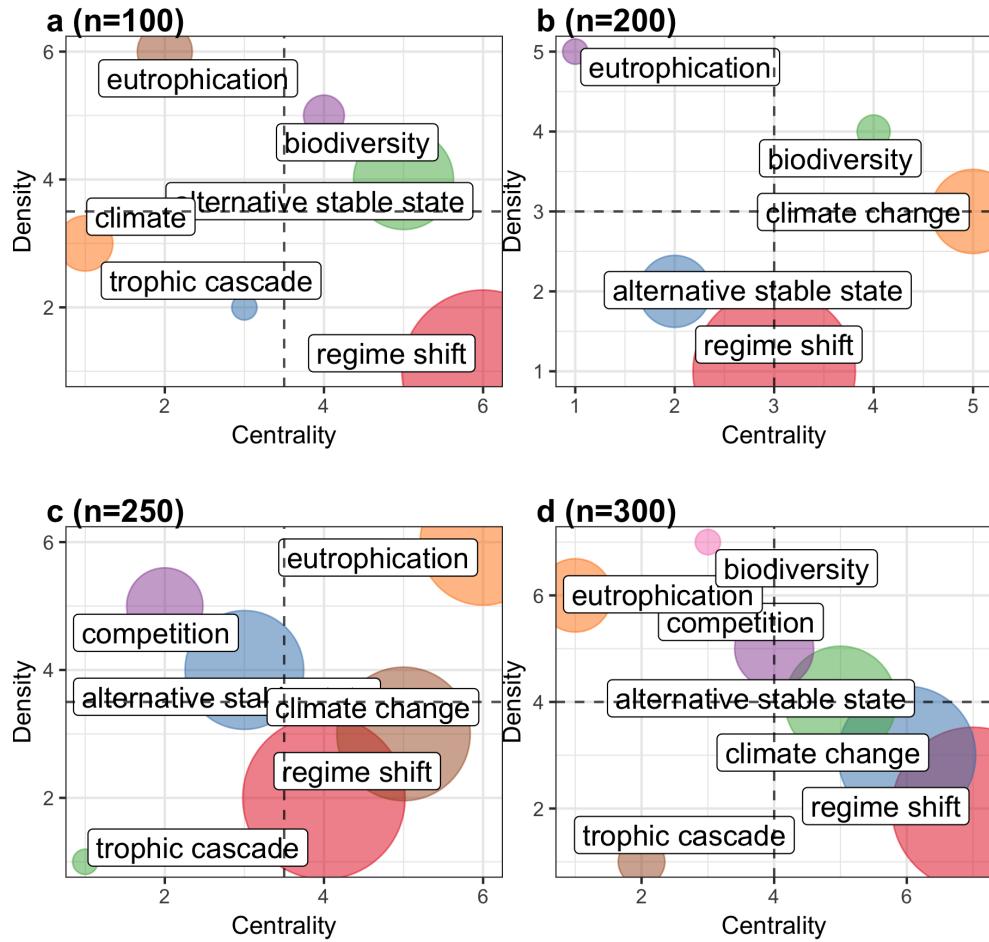
A search of Web of Science for articles in Ecology and Biodiversity Conservation containing phrases related to ‘regime shifts’ yielded 1,636 original articles. These articles were not filtered in any fashion and as such all were considered in the bibliographic analysis. I used the clustering algorithms 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).

#### 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 2.3)<sup>1</sup>. The clustering algorithm identified fewer clusters (themes) in the ISI-keywords (Figure 2.3a) than were identified among the author-supplied keywords (Figure 2.3b<sup>^</sup>; see footnote <sup>2</sup>). This pattern is not surprising given the ISI-supplied keywords are restricted to a pre-set 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 identified in the ISI keyword analysis and, interestingly, fell mostly within the

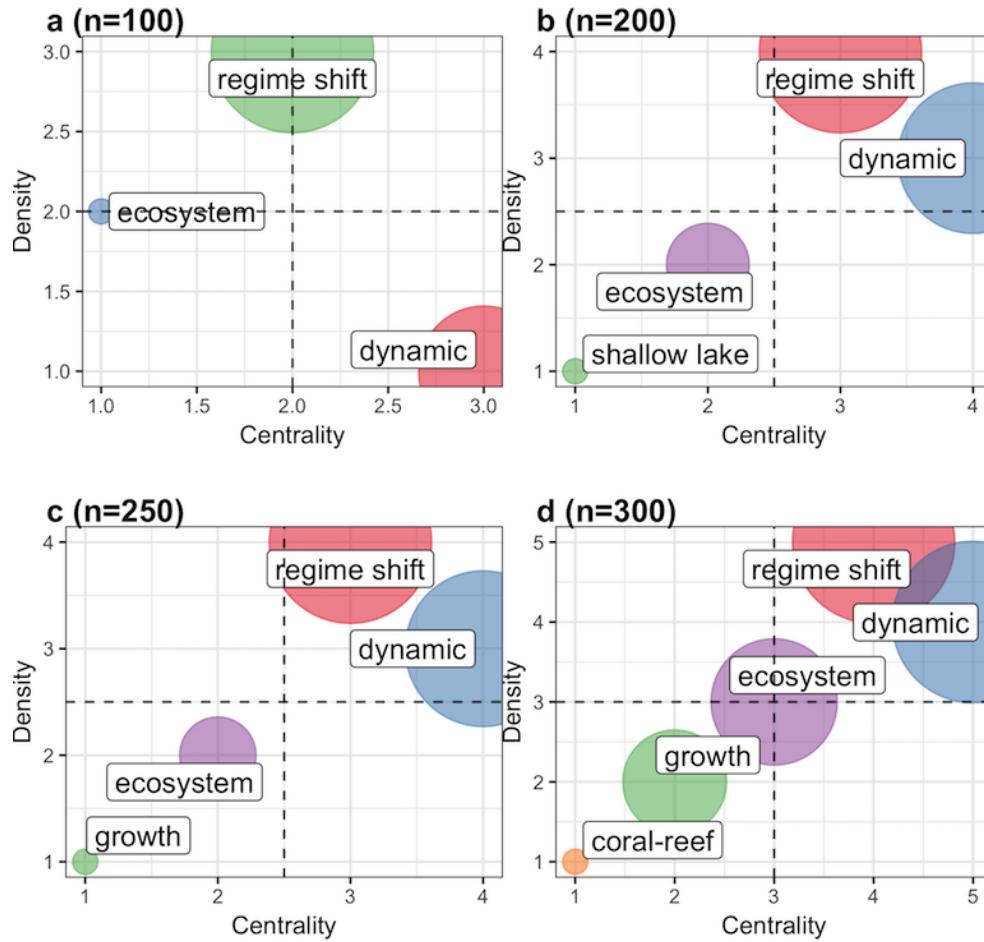
---

<sup>1</sup>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 2.4:** The themes identified in the literature based on author-supplied keywords vary as the number of words included in the analysis increased.

quadrants representing the most extreme values: the first and the third quadrants (Figure 2.5). The themes identified by the ISI-supplied keywords were much larger in scope (e.g., dynamics, ecosystems, climate; Figure 2.3a) than those identified in the analysis of author-supplied keywords (e.g., eutrophication, trophic cascade; Figure 2.3b). That is, the themes of ‘regime shifts’ and ‘ecosystem dynamics’ are highly central to and dense within the regime shift literature (Figure 2.3b-d). This suggests these two 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



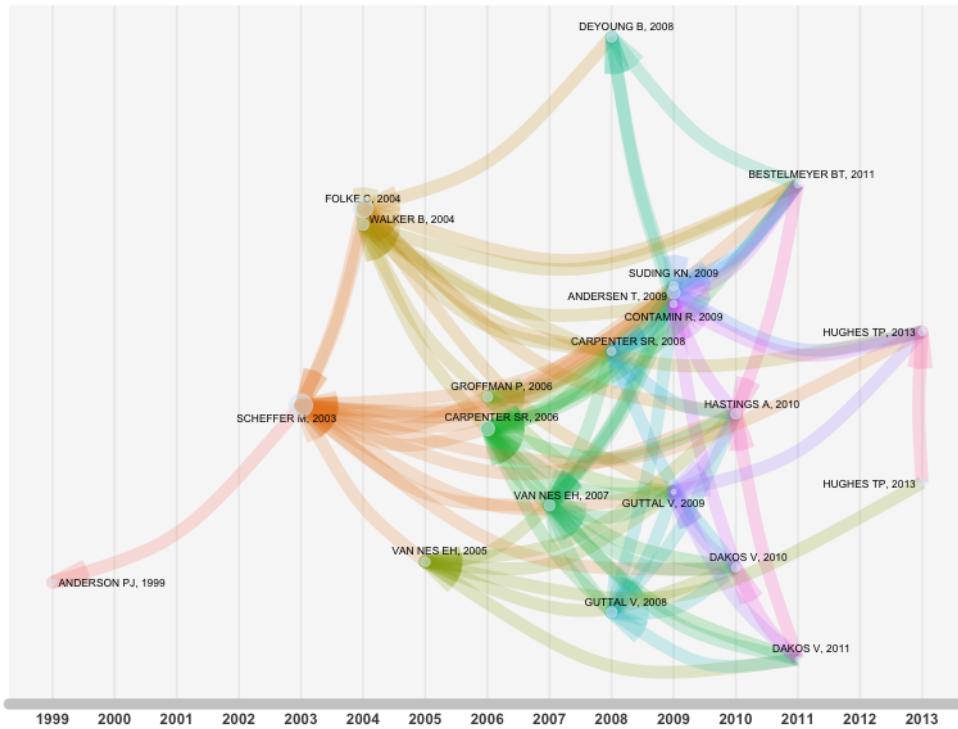
**Figure 2.5:** The themes identified in the literature based on the ISI-supplied keywords vary as the number of words included in the analysis increased.

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  $\sim 50$  of case studies actually tested or accounted for non-linear dynamics when applying early warning indicators and other regime shift measures to ecological time series.

### Historiograph, citation, and reviews

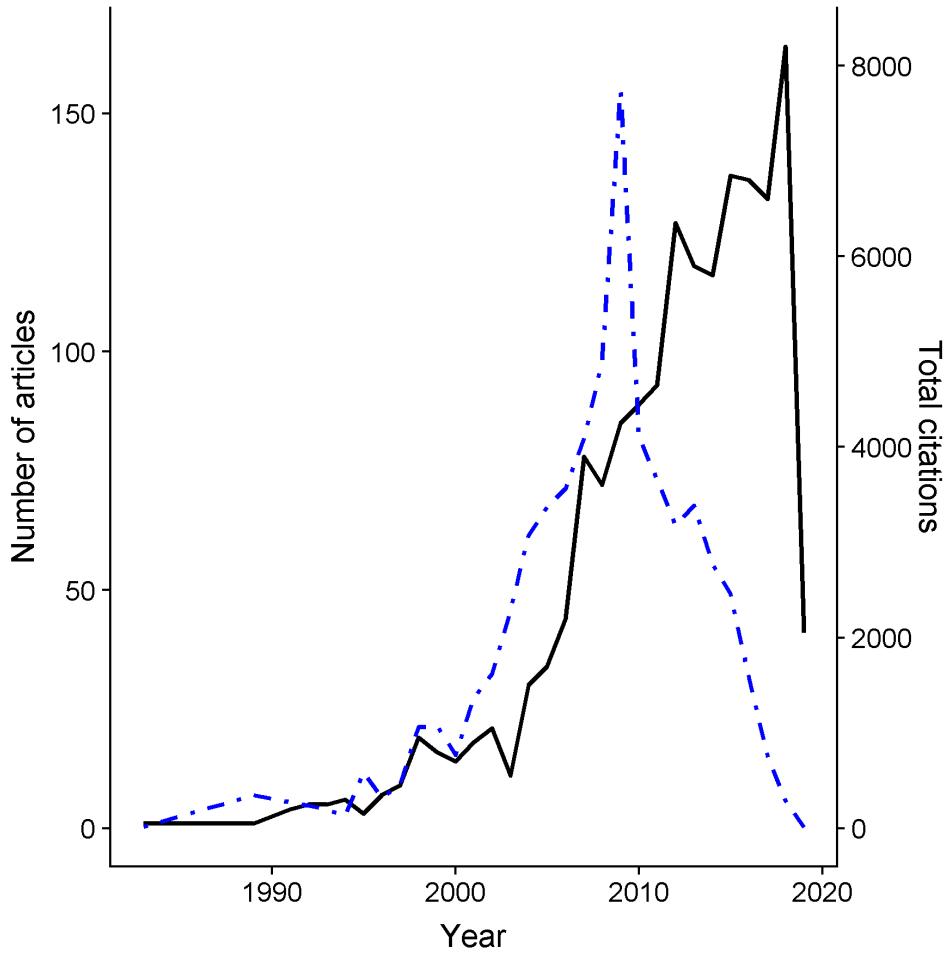
A few patterns appear in analyses of the intellectual and chronological structure of the ecological regime shift literature (Figure 2.6). First, although the concept of

Historical Direct Citation Network



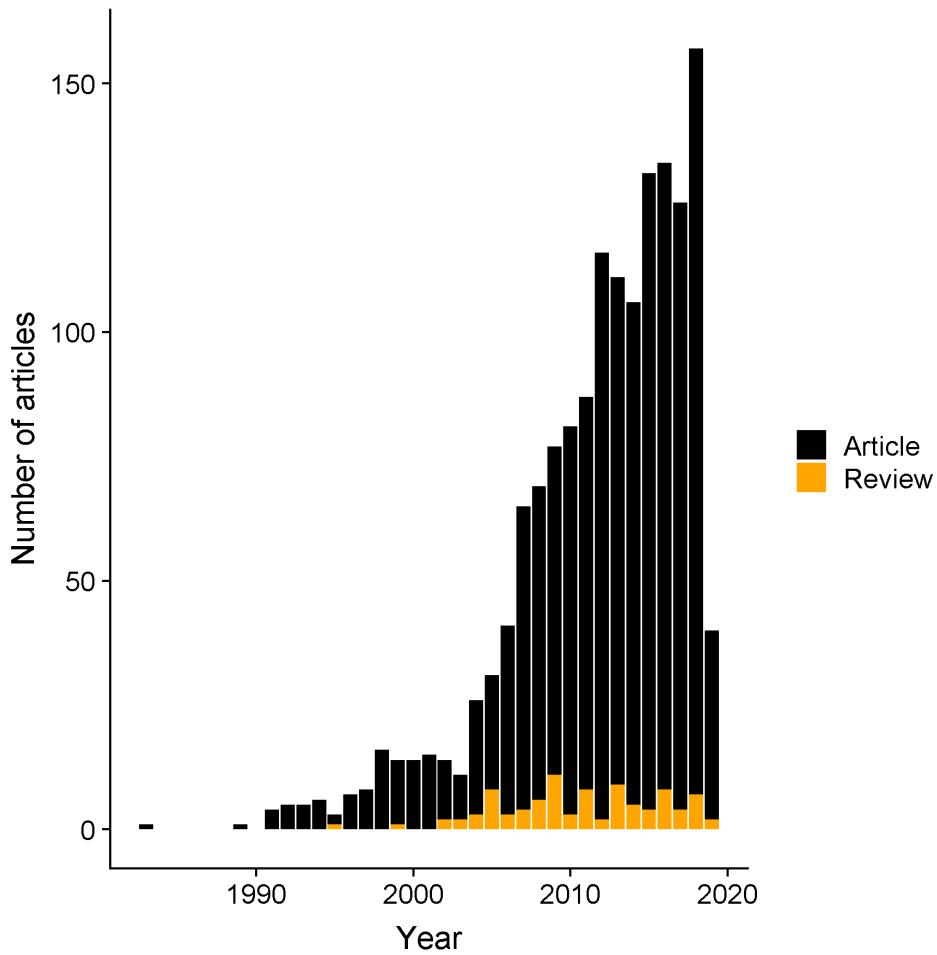
**Figure 2.6:** Chronological direct citation network suggests the intellectual structure can be mapped to a few papers. This historiograph identifies important works explicitly in chronological, as opposed to absolute, order.

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 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.



**Figure 2.7:** 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 most influential papers in the field, based solely on number of citations, are those published in the late 2000s (Figure 2.7), 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 ( $\sim 2007$ ) is when the number of regime detection measures doubled (Figure 2.2). 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 2000s),



**Figure 2.8:** Total number of articles published per year by category as categorized by ISI. Book chapters, proceedings, editorials, and letters are excluded.

Folke *et al.* (2004) and Scheffer & Carpenter (2003) (Figure 2.7).

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 2.8]. 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 2.1). 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.*, 2012a; Lindegren *et al.*, 2012; Perretti & Munch, 2012; Litzow & Hunsicker, 2016; Sommer *et al.*, 2017; Dutta *et al.*, 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 6]. 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 systems, 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 2.1), 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 2.1 have either not been applied to empirical data at all, or were tested only once, often but not always in the article introducing 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 6) 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 2.2) that would be useful in a much-needed modern and critical review of this field.

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

Type	Questions
Methodological	<p>Does the method assume smooth potential?</p> <p>Does the regime shift need to be identified <i>*a priori*</i>?</p> <p>What are the major assumptions about the distribution of the original data?</p> <p>Does the method explicitly assume the system/variable is stationary?</p> <p>Does the performance of the method change with non-stationarity?</p>
	<p>Can the method handle unstructured data?</p> <p>Can the method handle multiple regime shifts?</p> <p>What types of regime shifts can the method detect (e.g., stochastic resonance, slow-fast cycles, noise-induced transition)?</p> <p>Is it a model- or metric-based method?</p> <p>Does it have forecasting potential?</p>
	<p>Can the method handle uneven sampling?</p> <p>What are the minimum data requirements (resolution, extent, number of observations)?</p> <p>How does the method handle missing data (e.g., new invasions)?</p> <p>Does the method assume Eulerian or Lagrangian processes?</p> <p>Does the system <i>*have*</i> smooth potential?</p>
	<p>Has the method been tested on empirical data? If so, to what rigor?</p> <p>What is the impact of losing state variables on long-term predictions (e.g., species extinction)?</p>

- Can the method identify drivers?
  - What assumptions does the method make about the system?
  - What types of regime shifts are possible in the system?
- 
- Are regime shift(s) suspected *\*a priori\**?
  - What lag(s) exist in the data (system)?
  - Would a positive forecast change management action?
  - Do predictions translate to other systems?
  - Can we interpolate data if necessary? If so, what does this mean for inference?
- 
- 
- In which discipline(s) beyond ecology has the method been tested?

# **Chapter 3**

## **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 carry strict assumptions that are easily violated by dynamic systems such as ecosystems (Abadi *et al.*, 2010). Further, model mis-specification 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  $t \rightarrow \infty$  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  $t \rightarrow \infty$  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 reduction 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”:

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

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 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 **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,  $t_{shift}$  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

$$\begin{aligned} k(t) &= k_1 - 0.5(k_1 - k_2)(\tanh(\alpha(t - t^*)) + 1) \\ k'(t) &= 0.5\alpha(k_1 - k_2)(\tanh(\alpha(t - t^*))^2 + 1) \end{aligned} \quad (3.2)$$

### 3.3.3 Decoupling the Steps for Calculating Fisher Information

Two methods exist for calculating Fisher Information (FI) as applied to ecological systems 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).

$$\frac{ds}{dt} = \sqrt{\left(\frac{dx_1}{dt}\right)^2 + \left(\frac{dx_2}{dt}\right)^2} \quad (3.3)$$

**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.

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

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

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 3.1]. I simulated a regime shift in the carrying capacity of this model system, at approximately  $t = 200$  (Figure 3.2). 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 3.3). 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 3.4) changes fairly smoothly around the location of the regime shift, the system exhibits a steep decline in speed  $ds/dt$  soon after the induced regime shift (Figure 3.5).

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 ( $winsize = 13.061$  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 3.6).

## 3.4 Discussion

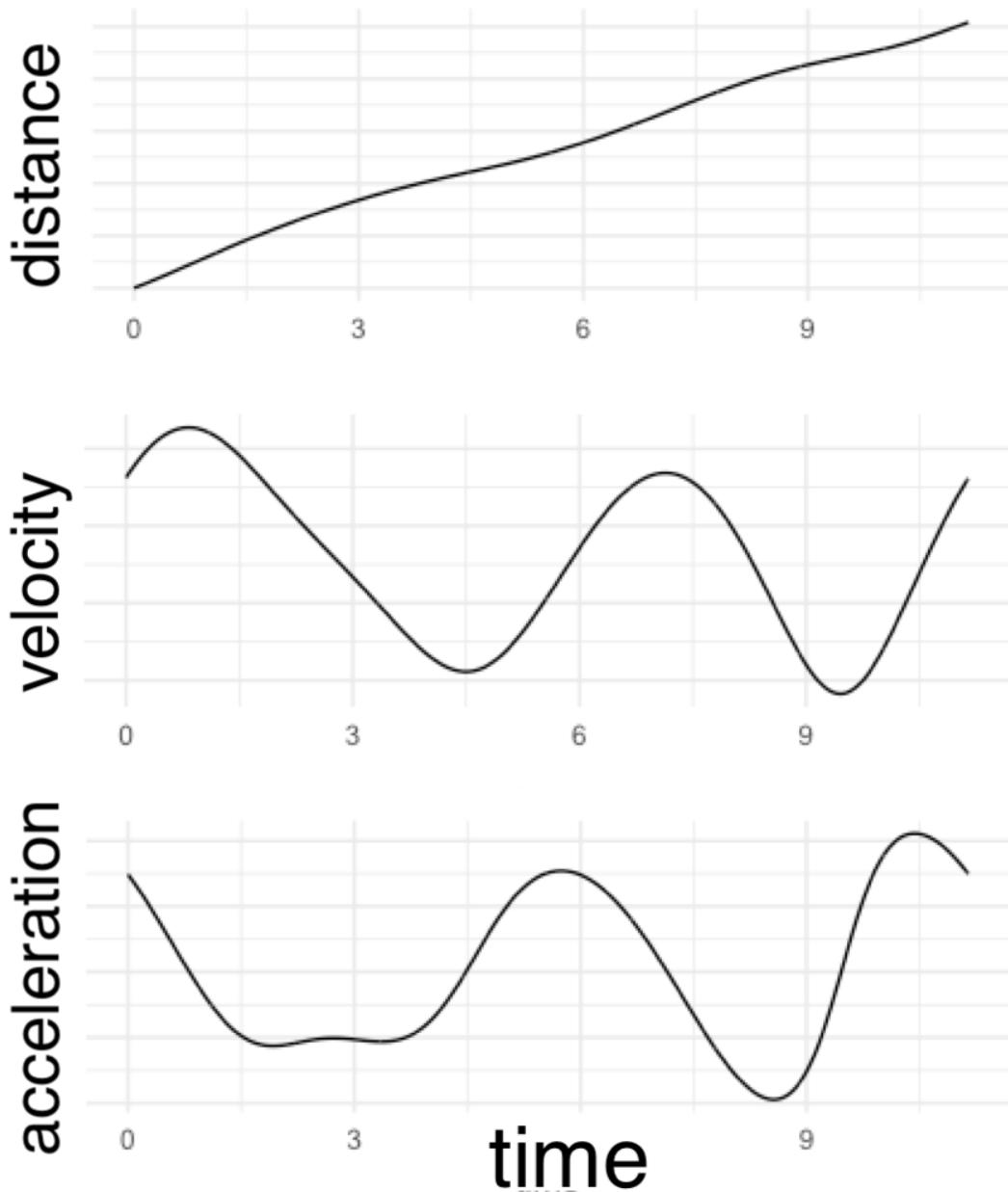
Part of the appeal of the FI method of regime shift detection is that it provides a 1-dimensional visual summary of system “orderliness”. However, I have demonstrated that, because only a single measure is required to calculate the Fisher Information, the dimensionality reduction step need not necessarily 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.

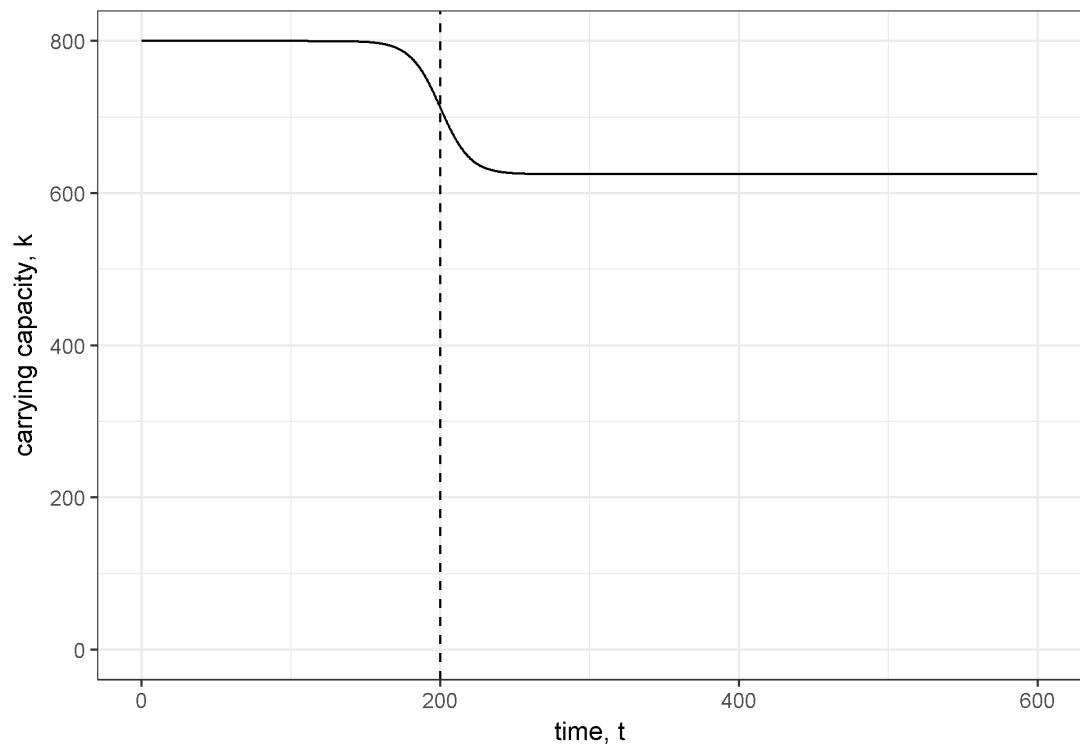
The example used in this study is unrealistic in that I assume no measurement error and therefore focus on the “derivatives-based” method of calculating FI. However, our analysis also has implications for the “binning” method of calculating FI that was later developed for high-dimension noisy data (Karunanithi *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**

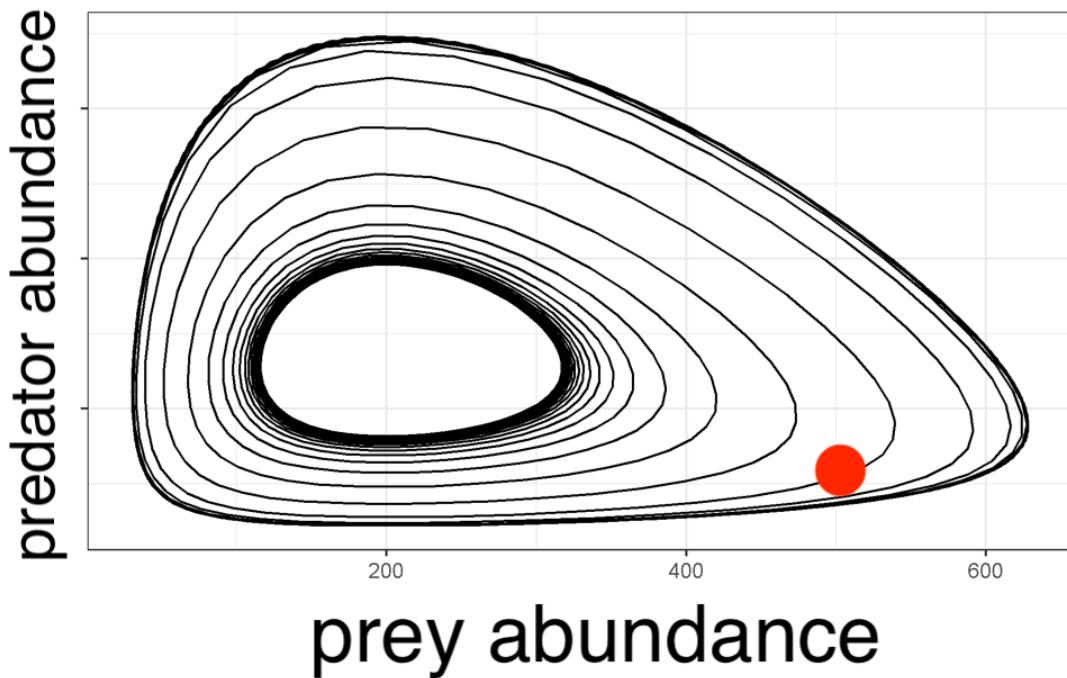
I thank H. Cabezas and B. Roy Frieden for early discussions regarding the development of Fisher Information, and T.J. Hefley for comments on an earlier draft.



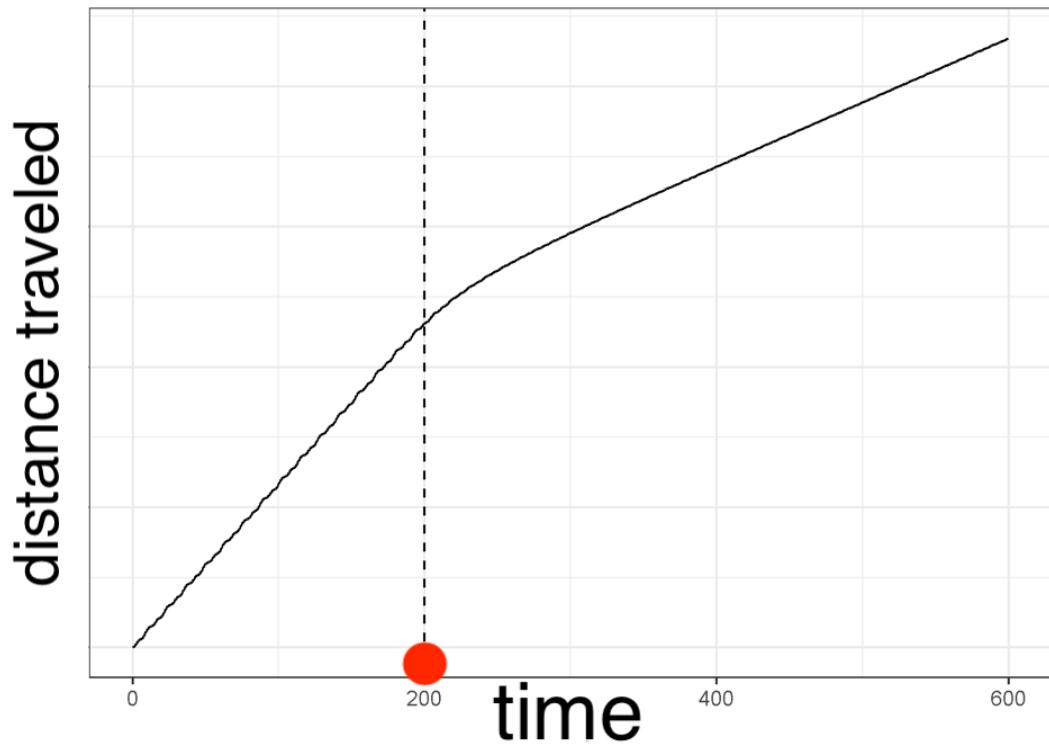
**Figure 3.1:** From top to bottom, distance traveled in phase space, speed tangential to system trajectory, acceleration tangential to system trajectory.



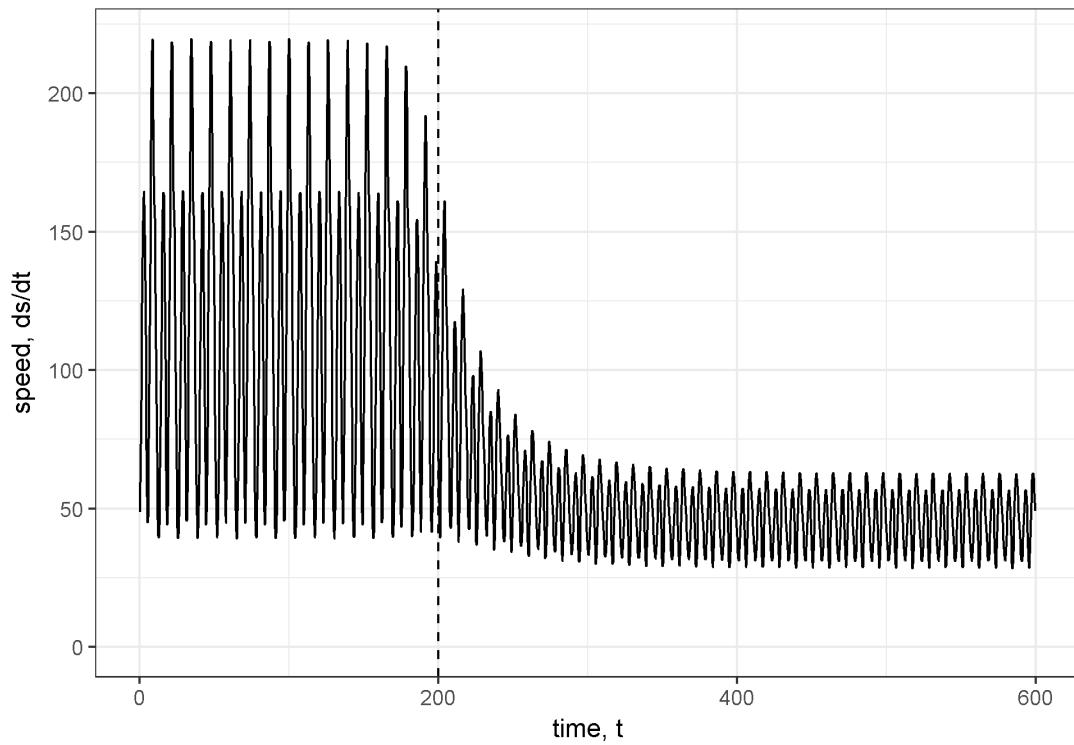
**Figure 3.2:** Carrying capacity over time with a regime shift occurring around time 200.



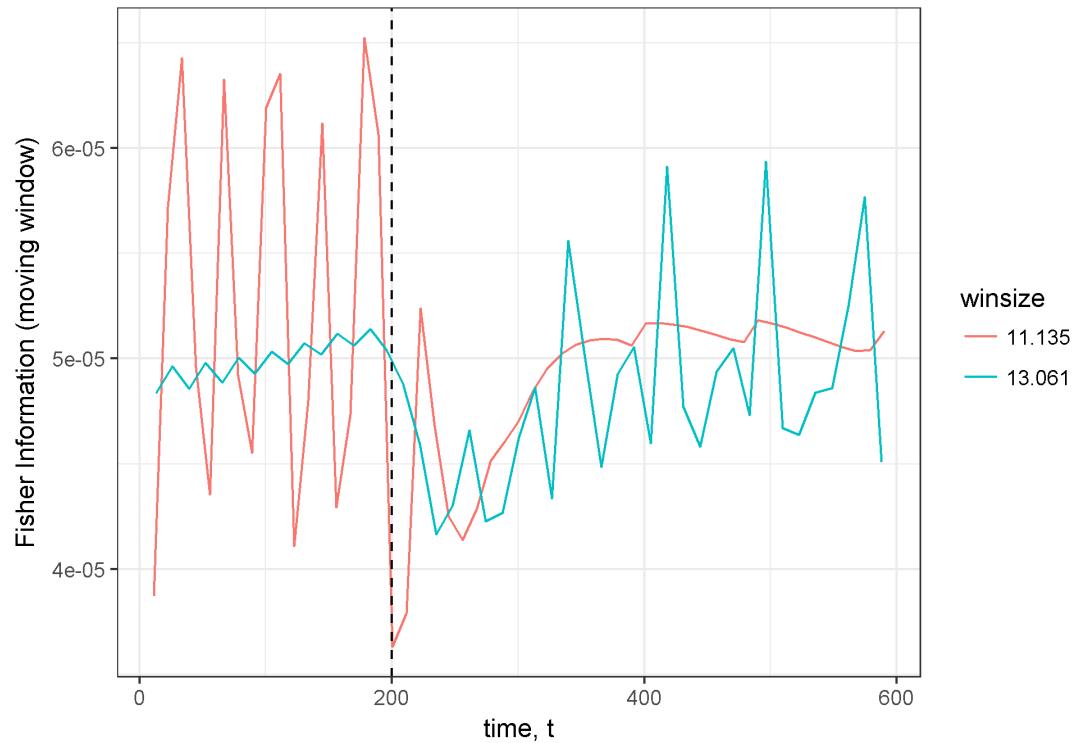
**Figure 3.3:** Phase space plot of system trajectories for different values of  $k$ . Colored point indicates the coordinates corresponding to the shift induced in carrying capacity.



**Figure 3.4:** Distance travelled in phase space over time. Dashed vertical line at time 200 indicates location of regime shift.



**Figure 3.5:** Speed of the system (rate of change, velocity) in phase space. Dashed vertical line at time 200 indicates location of regime shift.



**Figure 3.6:** 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 1) 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; Guttal & Jayaprakash, 2009; Kefi *et al.*, 2014; Butitta *et al.*, 2017;

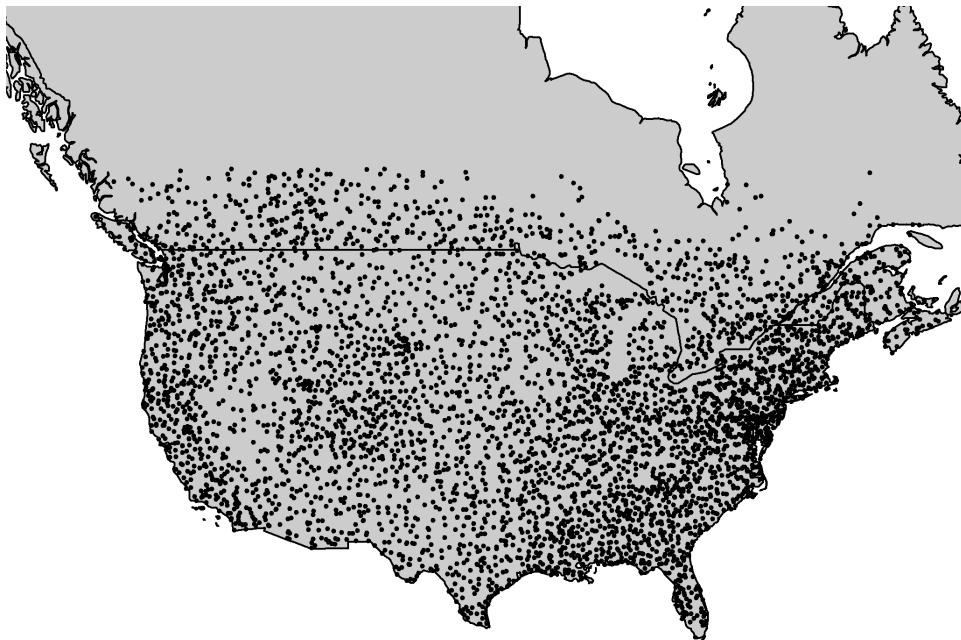
Sundstrom *et al.*, 2017). As defined in Table 1, a regime shift is largely considered an abrupt and persistent change in a system’s structure or functioning. Following this definition and without considering the pressures (Table 1) associated with the observed regime shift, it is not yet clear whether 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



**Figure 4.1:** *Locations of Breeding Bird Survey routes sampled between 1966 and 2017.*

(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 4.1) since 1966. The roadside surveys consist of 50 point counts (by sight and sound) along  $\sim 24.5$  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 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

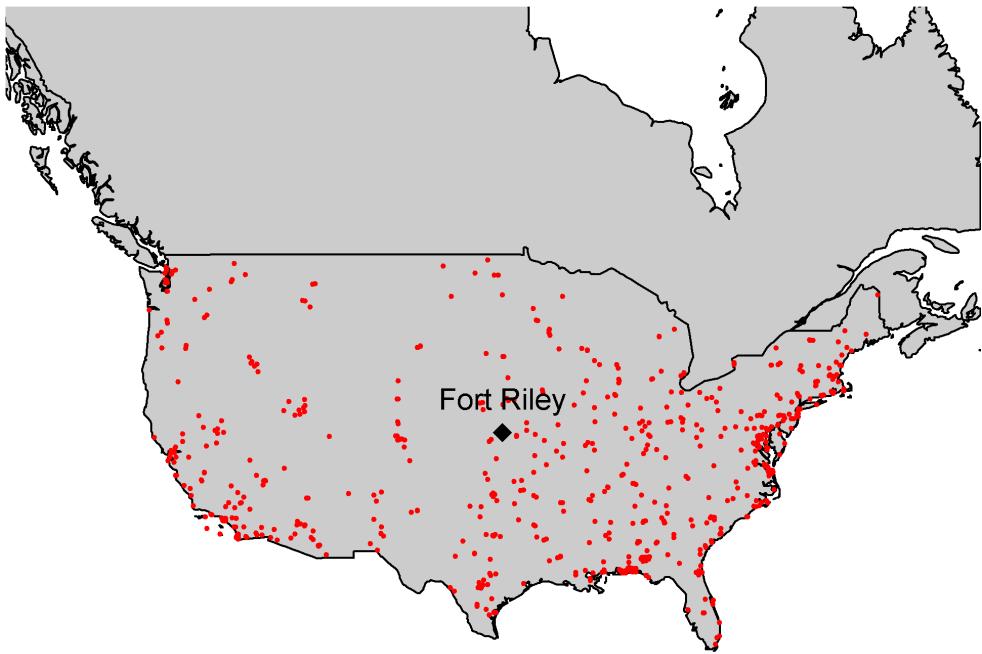
protocol. Despite a standardized survey protocol, some species are difficult to detect using these methods. For example, crepuscular species are less likely to be detected beyond the first few points of the BBS route, given they are most active at sunrise and the survey begins within 30 minutes of sunrise. 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 4.1).

##### **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 4.3). The spread of these bases (Figure 4.2), coupled with the top-down management of base-level natural resources presumably influences the inherent difficulties associated with collaborative management within and across military bases and other natural resource management groups (e.g., state management agencies,



**Figure 4.2:** Locations of focal U.S. military bases, Eglin Air Force Base (AFB) and Fort Riley Military Base.

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 primary objectives are very different. Natural resource managers of military bases, whose primary objective is to maintain military readiness, are especially concerned with if and how broad-scale external forcings might influence their lands. Prominent concerns include invasive species, wildlife disease, and federally protected species (personal communication with Department of Defense natural resource managers at Eglin Air Force and Fort Riley military bases). For these reasons, natural resource



**Figure 4.3:** Three East-West running transects analyzed in this Chapter overlaid against the Omernick Ecoregion boundaries map.

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 resource management efforts.

The NABBS routes chosen for analyses in this Chapter lie within or near Fort Riley military base (located at approximately  $39.110474^\circ$ ,  $-96.809677^\circ$ ; Kansas, USA). Fort Riley (Figure 4.2) 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.

## 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 4.3) 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,

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

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

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

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):

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

A form specific to the probability density function of distance traveled by the entire

system, which I call the ‘derivatives’ method, is defined as (Mayer *et al.*, 2007, eq. 7.12):

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

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 *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, 5 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 4.3) 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

##### 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  $(\frac{dq(s)}{ds} \rightarrow \infty)$ , we infer the system is approaching a stable state, and as  $\frac{dq(s)}{ds} \rightarrow 0$  the system is showing no preference for a single stable state and is on an unpredictable trajectory. 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 derivative, 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).

### Interpolating results across spatial transects

NABBS are not regularly spaced, and pairwise correlations of adjacent transects (see Figure 4.3) 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).

### Spatial correlation of Fisher Information

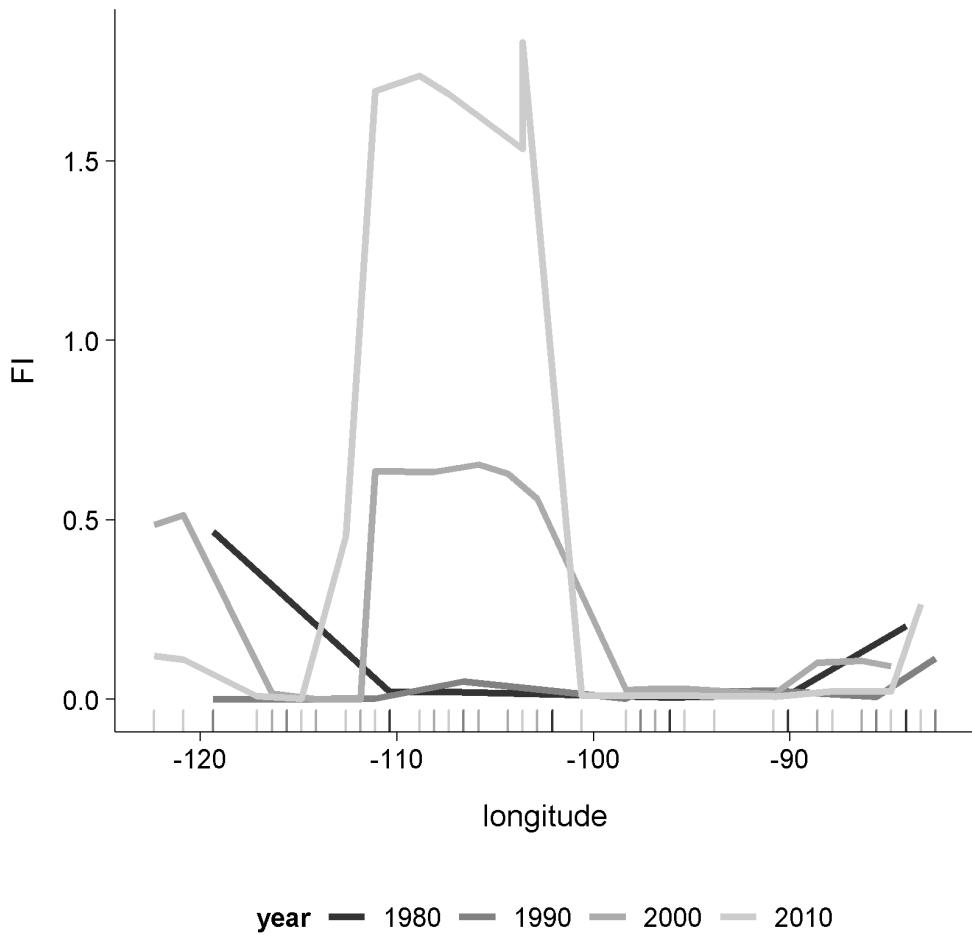
If Fisher Information captures and reduces information regarding abrupt changes in community structure across the landscape, then it follows that the values of Fisher Information should be spatially auto correlated. That is, the correlation of FI values should increase as the distance between points, 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 4.3). Fisher Information **is**, however, directly comparable within each spatial transect (e.g., 4.3). For these reasons, we can identify spatial regime shifts both within and among spatial transects by using pairwise correlations among two transects (e.g., 4.3) 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 4.3). 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.

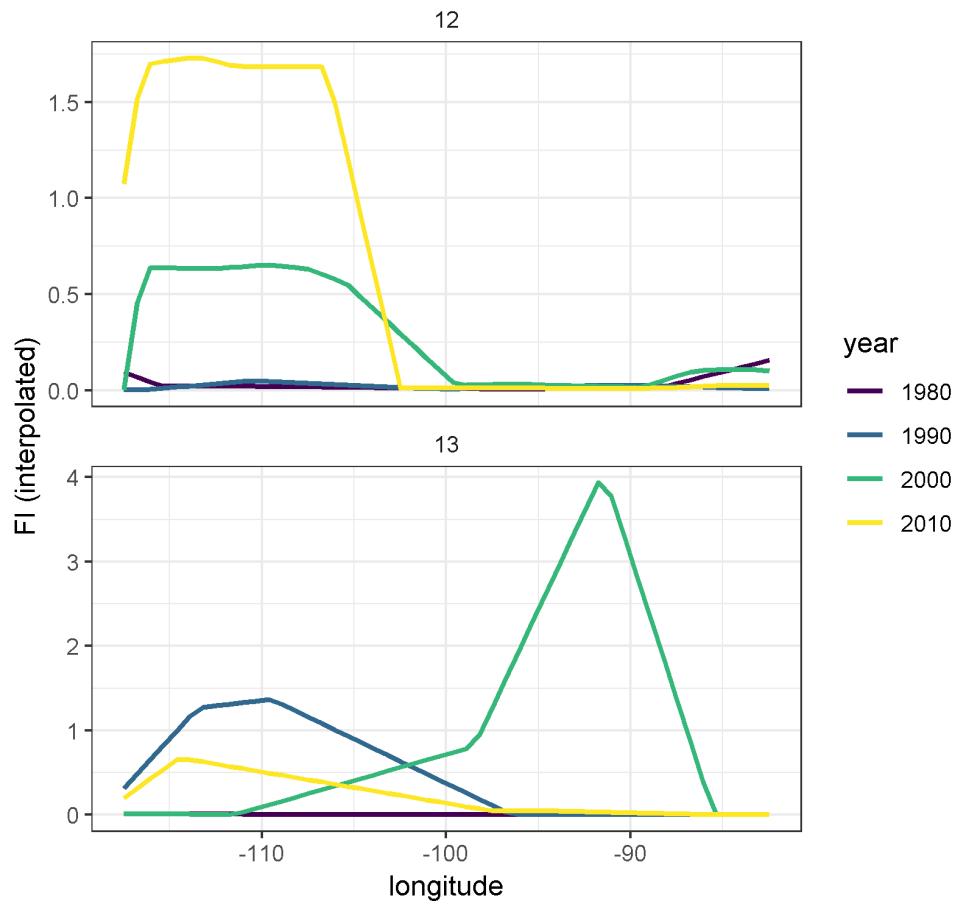
## 4.3 Results

### 4.3.1 Fisher Information across spatial transects

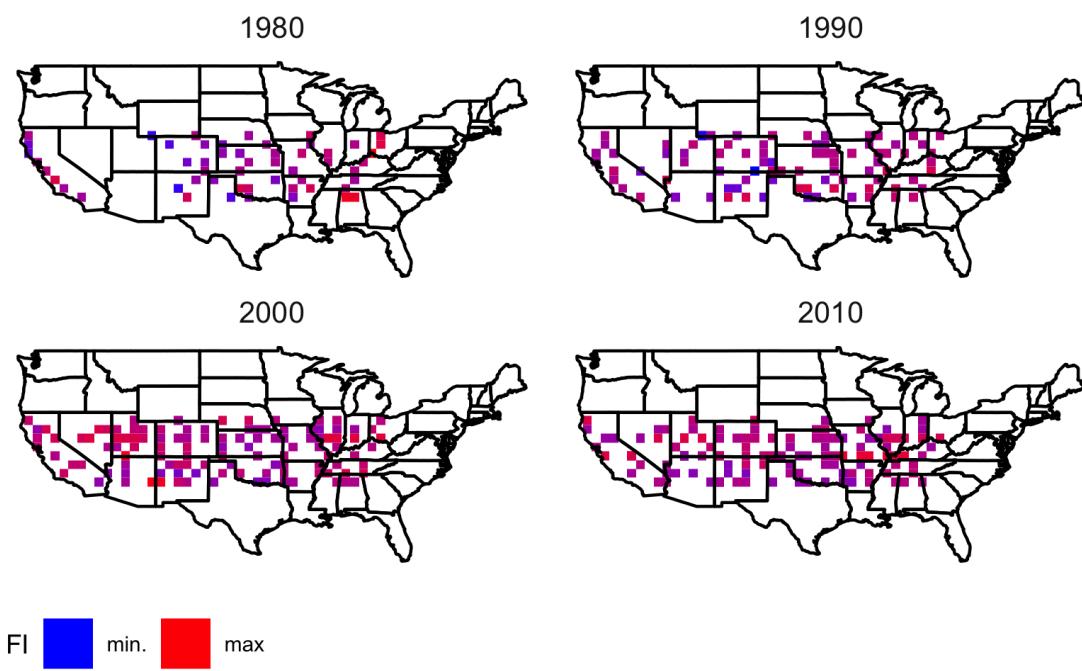
As suggested earlier, rapid increases or decreases in FI are posited indicate a change in system orderliness, potentially suggesting the location of a regime shift. Using this method yields inconclusive results regarding the location of ‘spatial regimes’ (Figure 4.4). Of the three spatial transects analyzed in this chapter (see Figure 4.3), Figure 4.4 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.



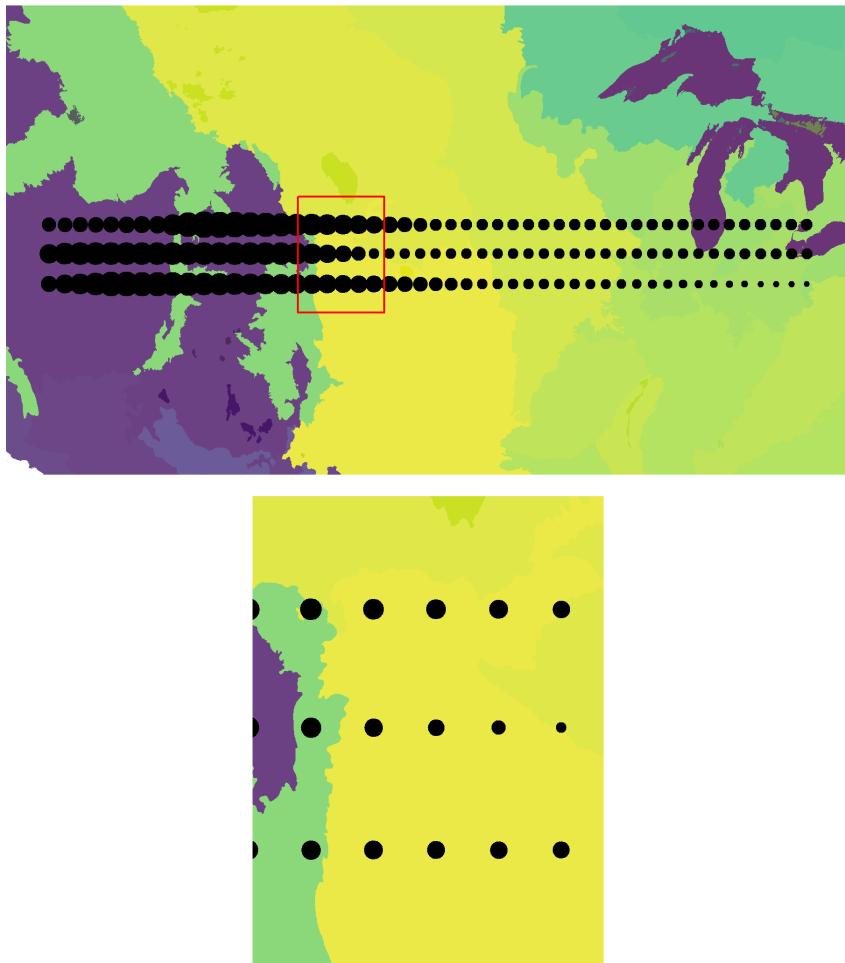
**Figure 4.4:** *The change in the Fisher Information values along a single, East-West-running spatial transect (Transect number 12) over time.*



**Figure 4.5:** 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 4.6:** Fisher Information of select East-West spatial transects over time. These transects are representative of the results obtained from the remaining transects.



**Figure 4.7:** 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 4.8 and 4.9)). For spatially-adjacent transects (e.g, transects 11 and 12, or 12 and 13 in Figures 4.8 and 4.9), we should 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 4.3), as the spatial soft-boundaries limiting the distribution and functional potential of avian communities are largely North-South (Figure 4.3). Note spatial transects in Figure 4.3 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 4.7)<sup>1</sup>. If Fisher Information is an indicator of spatial regime boundaries, we should expect to see large changes in its value (in either direction) near the edges of functional spatial boundaries (e.g., at the boundaries of ecoregions). No clear regime changes appeared in areas where we might expect rapid changes (e.g., along the 105th meridian West, where a sharp change in altitude occurs). Numerical investigation of the spatial correlation among adjacent transects also yielded no clear patterns. I did not identify any obvious correlation with changes in FI values and functional potential (using Omernick Ecoregion Level 2; see Figure 4.7). However, in years 2000 and 2010 the transects 11 and 12, and 12 and 13 were highly positively correlated (Figure 4.9). Rather than abrupt changes in Fisher Information I found gradual changes (e.g., see results for years 2000 and 2010 in Figs. 4.7 and 4.10).

## 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; Karunamithi *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 find consistent abrupt changes in the Fisher Information metric (Eq. (4.4)), which would indicate a regime shift of sorts,

---

<sup>1</sup>Here, 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.

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 6). 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 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 removes the dimensionality of these systems, which may omit information or signals integral to understanding the ecological processes at play.

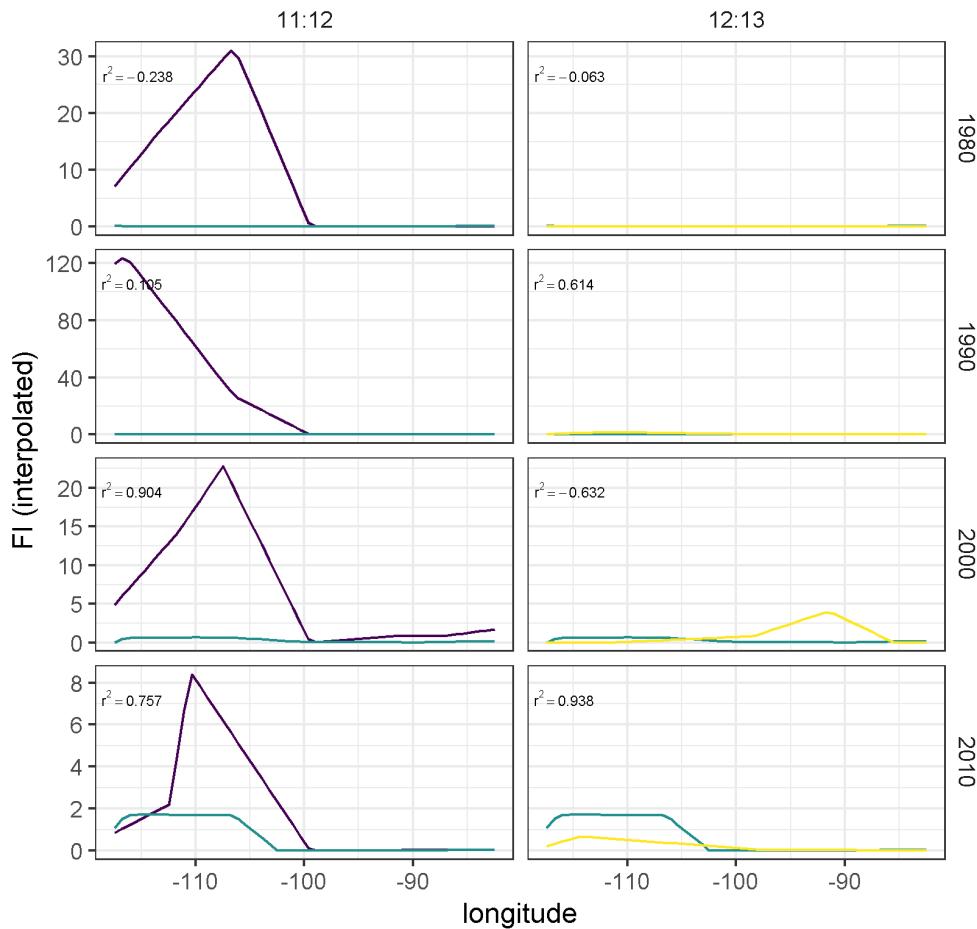
Effective regime detection measures should provide sufficient evidence of the drivers and/or pressures associated with the identified regime shifts (Mac Nally *et al.*, 2014). The Fisher Information index, while collapsing a wealth of data into a single metric, does not allow the user to relate the resulting 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 suggested a spatial regime boundary or shift, a before-and-after post-hoc analysis of the regional community dynamics might confirm the regime shift occurrence.

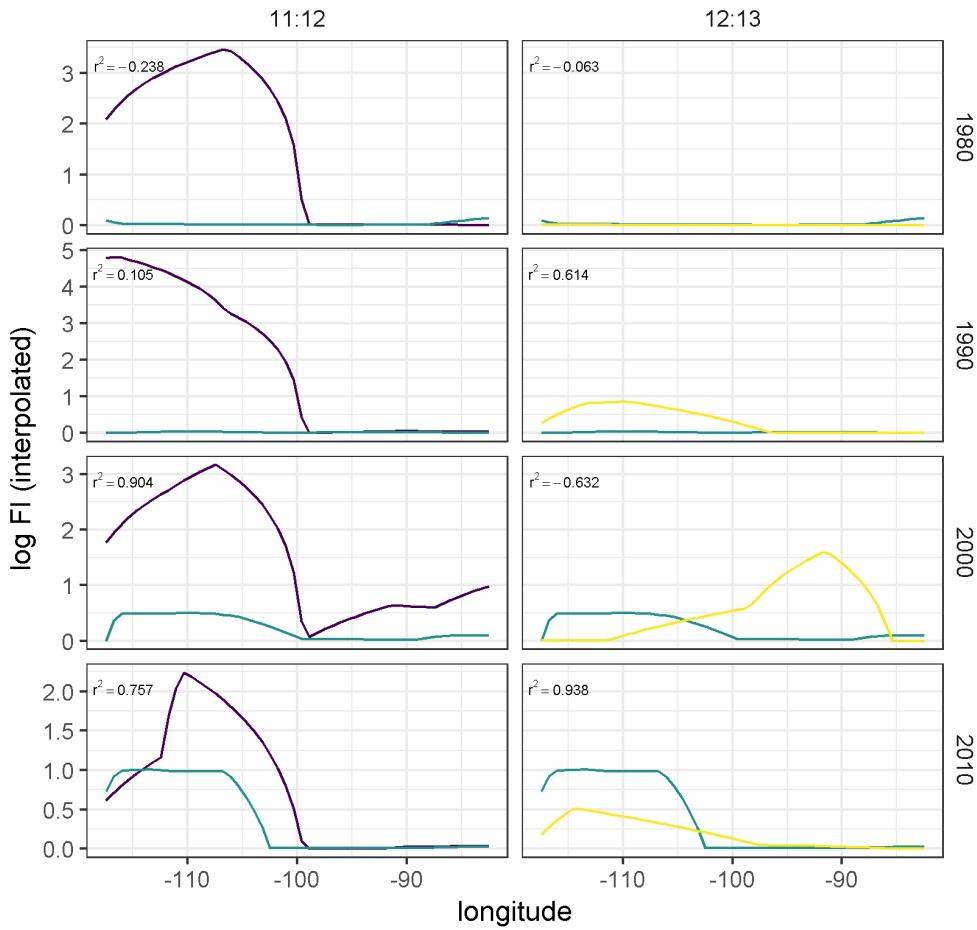
A rapid change in either direction (increase or decrease) of the Fisher Information value is proposed as an indicator of ecological regime shifts, or a change in the orderliness of a system (Mayer *et al.*, 2006; Eason & Cabezas, 2012). After calculating the Fisher Information for each spatial transect (Figure 4.3) during each sampling year in this study, I used pairwise correlation to determine whether spatial autocorrelation existed among pairs of spatial transects. If some set of points are close in space and are *not* separated by some physical or functional boundary (e.g., an ecotone, high altitude rock formations), then the Fisher 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 6).

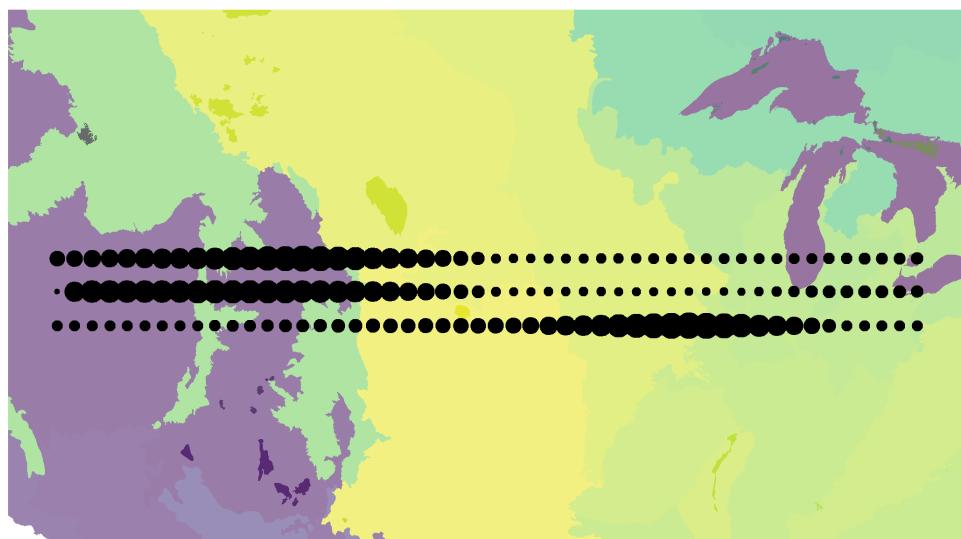
2. What, if any, advantages does FI have over other density estimation techniques?
3. Does FI provide signals in addition to or different than geophysical and vegetative (e.g. LIDAR) observations (data)?
4. Relationship of Fisher Information to likelihood ratio-based unsupervised change-point detection algorithms (e.g., ChangeFinder; Liu *et al.*, 2013).
5. How does Fisher Information perform relative to other regime detection measures (see Chapter 6)?



**Figure 4.8:** 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.



**Figure 4.9:** 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.



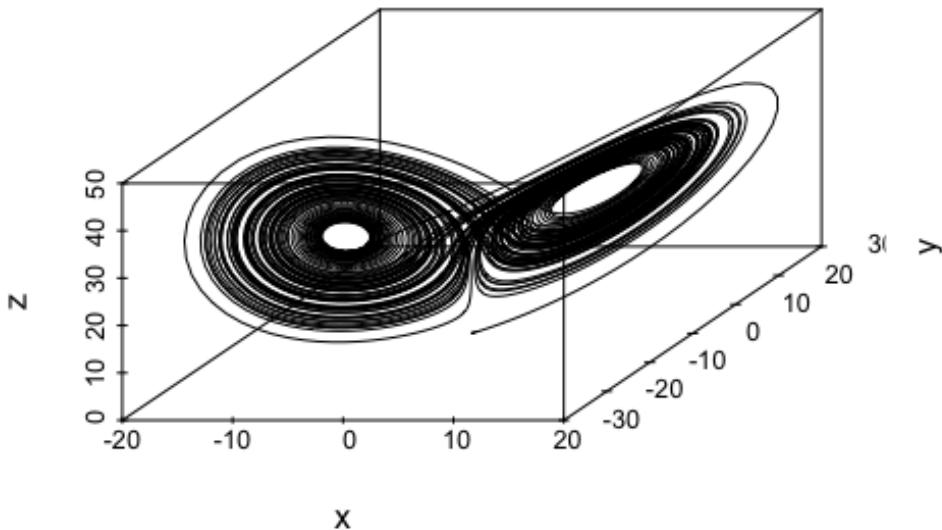
**Figure 4.10:** *Fisher Information results (scaled and centered; larger circles (points) represent higher values of Fisher Information) plotted against ecoregion boundaries (EPA Level 2 boundaries) for the year 2000.*

# Chapter 5

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

### 5.1 Introduction

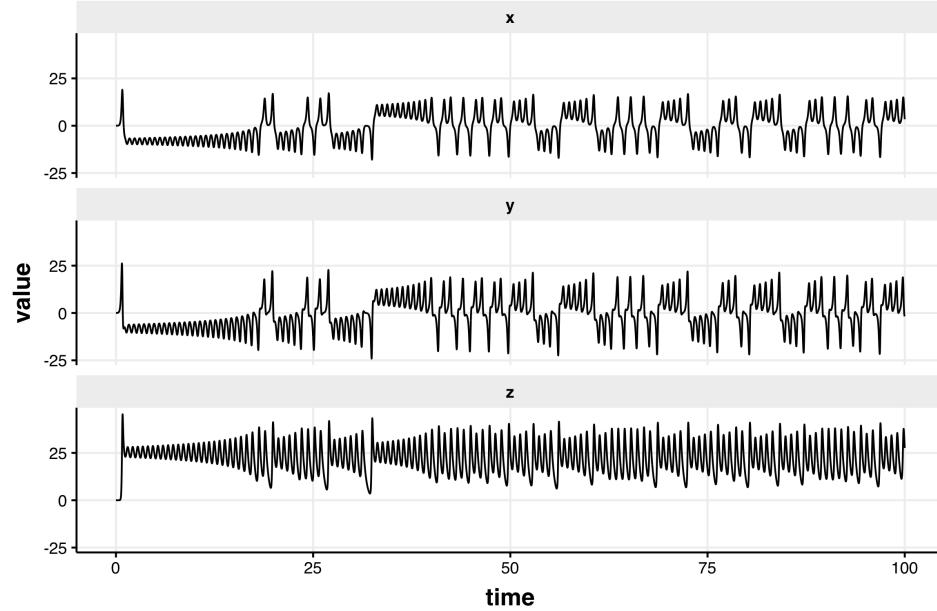
When, how and why ecological systems exhibit abrupt changes is a hallmark of modern ecological research, and changes that are unexpected and undesirable can have undesirable downstream consequences on, e.g., ecosystem services, biodiversity, and human well-being. Quantitatively detecting and forecasting 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



**Figure 5.1:** 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.

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 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.



**Figure 5.2:** An example solution of the Lorenz ('butterfly') represented in individual system components.

### 5.1.1 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 5.1; Takens, 1981). This phase plot (Figure 5.1) 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 5.2), 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.

### 5.1.2 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; 5.1] 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 variables, measuring the change in one variable relative to another. As a refresher ROC is represented as **speed (S)** or **velocity (V)**,

where ( $\mathbf{S}$ ) is the adirectional magnitude (i.e. it is a scalar) of the displacement of an object over unit time and  $\mathbf{V}$  describes both the direction and magnitude (i.e. it is a vector) of the object's movement in space-time.  $\mathbf{S}$  is a scalar taking values of  $\geq 0$  and  $\mathbf{V}$  can take any value between  $-\infty$  and  $\infty$ . For example, consider a car travelling at a constant speed of  $50 \frac{\text{km}}{\text{h}}$  around along a hilly landscape, where it is ascending and descending hills. Although  $\mathbf{S}$  is constant,  $\mathbf{V}$  changes in a sinusoidal fashion, where  $\mathbf{V}$  is  $\mathbf{V} > 0$  when ascending,  $\mathbf{V} < 0$  when descending, and  $\mathbf{V} \approx 0$  at in the valleys and at the peaks of the hills. Although  $\mathbf{S}$  is useful when estimating other scalar quantities (e.g.,  $\frac{\text{miles}}{\text{gallon}}$ ), given a starting and/or final position in space,  $\mathbf{S}$  is not informative of its the path traveled.

### 5.1.3 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.

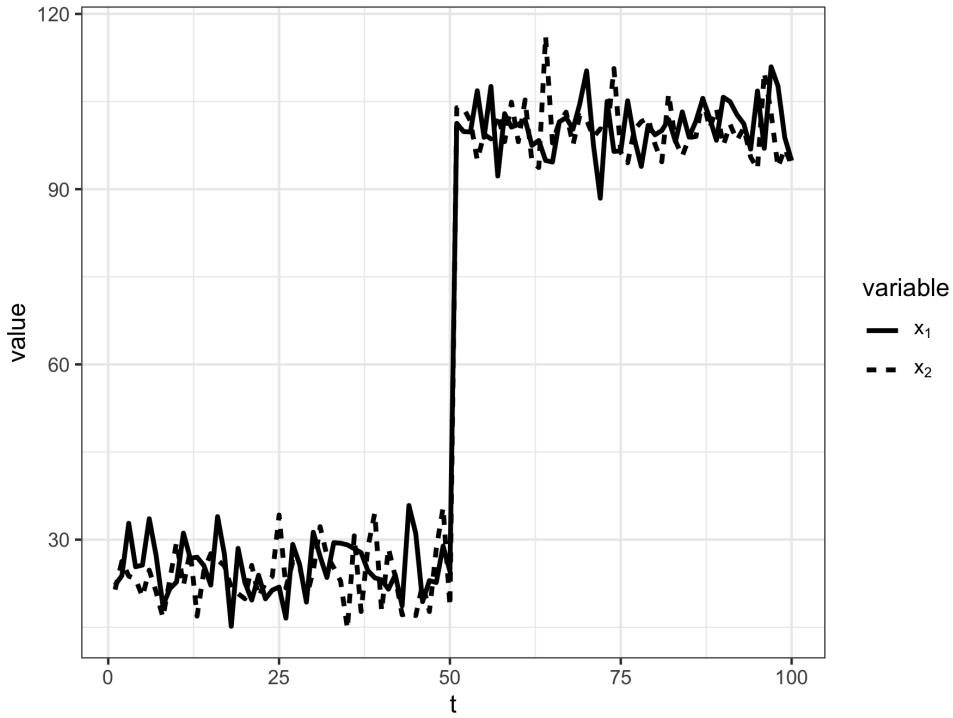
### 5.1.4 Analytical approach

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.

## 5.2 Steps for Calculating Velocity, $V$

S 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

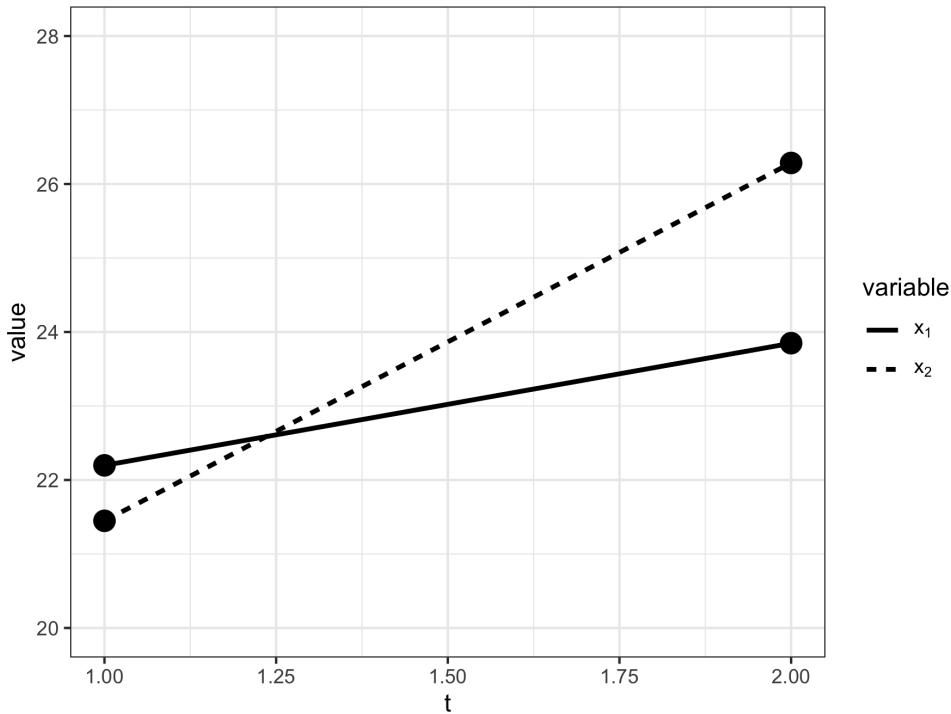
demonstrate the signals achieved with and the variability within the  $V$  calculation by exploring a number of scenarios of this simple system. For the examples in this section, observations of  $x_i$  are randomly drawn from distribution  $x_i \sim \text{Normal}(\mu, \sigma)$ , where  $\mu$  is the mean and  $\sigma$  is the standard deviation. Consider a system (Figure



**Figure 5.3:** 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,  $\sigma = 5$ .

5.3) 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,  $\{x_1 \dots x_N\}$ , 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 occurring at each time point  $t = 1, 2, 3, \dots, 100$ . 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 5.4). For all examples in this chapter, the state variables  $x_1$  and  $x_2$  were drawn from a normal distribution (using function *rnorm*), with parameters  $\bar{x}_i$  (mean) and  $\sigma_i$  (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 5.4:** Data used to calculate velocity at the first two time points,  $t_1$  and  $t_2$ .

### 5.2.1 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 5.4; Equation (5.1)]:

$$\Delta x_i = x_{i(t+1)} - x_{it} \quad (5.1)$$

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

### 5.2.2 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 (5.2)]:

$$\Delta s = \sqrt{\sum \Delta x_i^2} \quad (5.2)$$

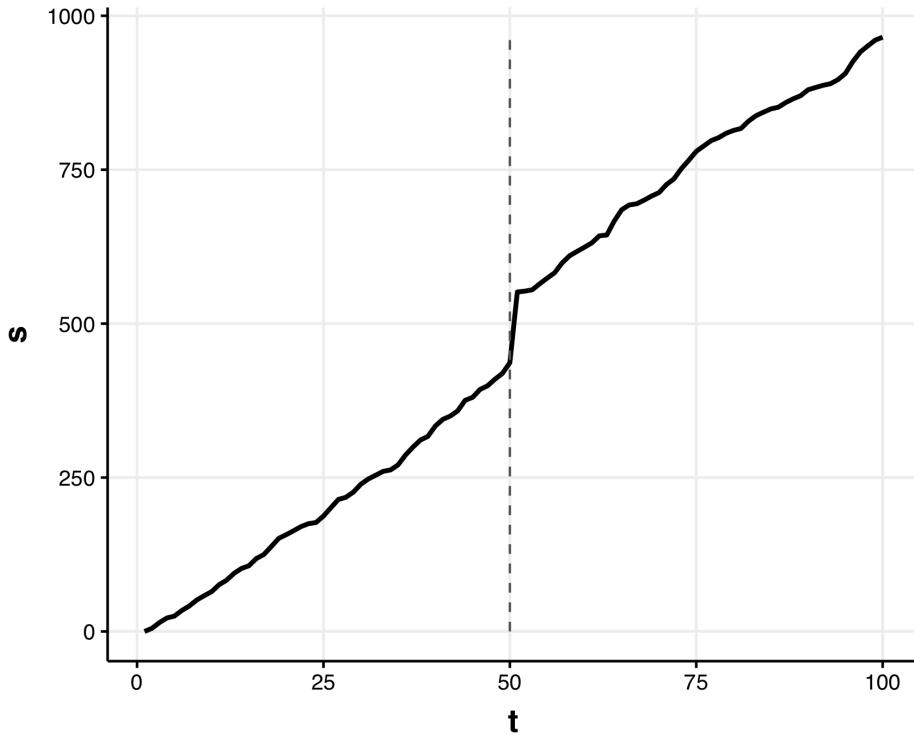
Although  $\Delta s$  represents the absolute change in the system between consecutive points in time, this measure is not yet relative along the system's trajectory. To create a relative value we next calculate the total distance traveled along the system trajectory,  $s$ , as the cumulative sum of  $\Delta s$  [Equation (5.2)] since the first observation, such that a cumulative sum is calculated for every  $t$  over the interval  $[0, T]$  [Equation (5.3)]:

$$s_T = \sum_{t=0}^T \Delta s \quad (5.3)$$

We now have a single measure,  $s_T$  [hereafter referred to as  $s$ ; Equation (5.3)] at each discrete point in time in our  $N$ -dimensional system (Figure 5.5). It should be noted that  $s$  (Figure 5.5) is monotonically increasing since the value of  $\Delta s$  [Equation (5.2)] 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.

### 5.2.3 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 (5.3)], and then divide by the total time elapsed between consecutive



**Figure 5.5:** Distance traveled,  $s$ , for the 2-species toy system.

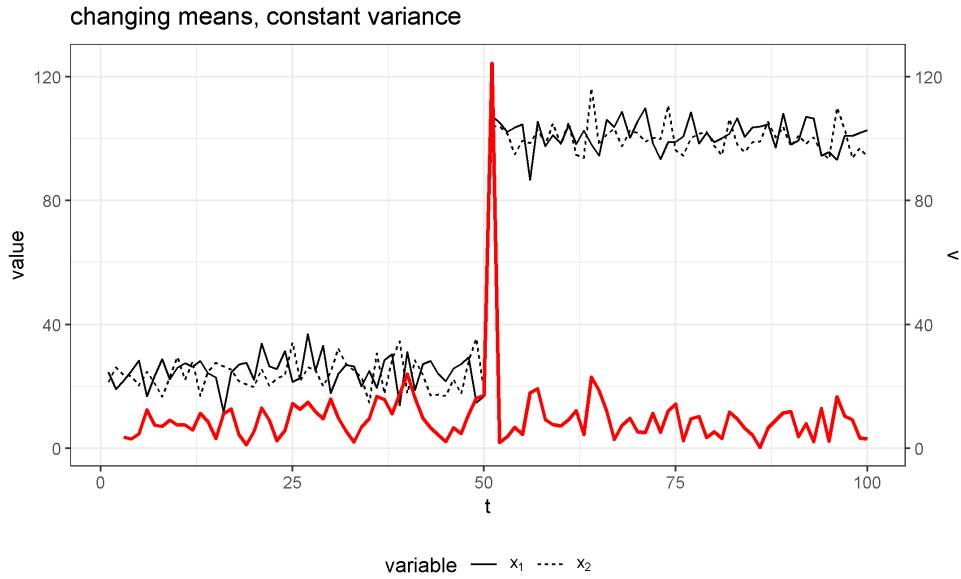
sampling points:

$$v = \frac{\Delta s}{\Delta t} \quad (5.4)$$

The numerical results for each step in the calculation of velocity [Equation (5.4)] is demonstrated using the first five time points of our toy system (Figure 5.3) in Table 5.1.

### 5.3 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 5.3). If the system undergoes a rapid and discontinuous change in one or more state



**Figure 5.6:** 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,  $\sigma = 5$ .

variables, the velocity, because it is a rate of change, has the potential to approach  $\infty$  as  $\Delta t \rightarrow 0$ . 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,  $t$ .

**Table 5.1:** Steps outlined for calculating system velocity,  $V$ , using the 2-variable toy data as an example.

t	$x_1$	$x_2$	$\Delta x_1$	$\Delta x_2$	$\Delta t$	$\sqrt{(\sum_{i=1}^N \Delta x_i^2)}$	s	$V$
1	22.198	21.448						
2	23.849	26.284	1.651	4.836	1	5.111	5.111	
3	32.794	23.767	8.944	-2.518	1	9.292	14.403	9.292
4	25.353	23.262	-7.441	-0.504	1	7.458	21.861	7.458
5	25.646	20.242	0.294	-3.020	1	3.035	24.895	3.035

### 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  $W = \{25, 30, 35, \dots, 100\}$  (Figures 5.7 and 5.8). As expected, the magnitude of  $V$  increases linearly as the total difference between  $\bar{x}_{1,pre}$  and  $\bar{x}_{1,post}$  increases (Figure 5.8). This is not surprising because  $s$  increases as the total change in abundance across the entire system increases [Equation (5.3)]. 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).

### 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  $W = \{1, 2, 3, \dots, 25\}$ . The variance for both state variables ( $x_1, x_2$ ) prior to the regime shift,  $\sigma_{x_1}$  and  $\sigma_{x_2}$ , was 5, with the change occurring in  $\sigma_{x1,post}$ .

### **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 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. (5.1) through (5.4)). 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 5.10).

## **5.4 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 damped under varying degrees of transition speeds. 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 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

**Table 5.2:** 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.

conditions	$\sigma_{x_{1pre}}$	$\sigma_{x_{1post}}$	$\sigma_{x_{2pre}}$	$\sigma_{x_{2post}}$	$\mu_{x_{1pre}}$	$\mu_{x_{1post}}$	$\mu_{x_{2pre}}$	$\mu_{x_{2post}}$
$\mu_{x_1}, \mu_{x_2}, \sigma_{x_1}, \sigma_{x_2}$	0.05	0.10	0.05	0.10	10	55	15	44
$\mu_{x_1}, \sigma_{x_1}$	0.05	0.10	0.05	0.05	10	55	15	15
$\mu_{x_1}, \mu_{x_2}$	0.05	0.05	0.05	0.05	10	55	15	44
$\mu_{x_1}$	0.05	0.05	0.05	0.05	10	55	15	15
$\sigma_{x_1}, \sigma_{x_2}$	0.05	0.10	0.05	0.10	10	10	15	15
$\sigma_{x_1}$	0.05	0.10	0.05	0.05	10	10	15	15

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.

#### 5.4.1 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 5.2 (random seed in Program R was 12345).

#### 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 (5.5); see Figure 5.13]:

$$\begin{aligned}\mu_{xipre} &\sim \text{Normal}(\mu_{xipre}, \sigma_{xipre}) \\ \mu_{xipost} &\sim \text{Normal}(\mu_{xipre}, \sigma_{xipost}) \\ \mu_{xi}(t) &= \mu_{xipre} - 0.5(\mu_{xipre} - \mu_{xipost})(\tanh(\alpha(t - t_{shift})) + 1)\end{aligned}\quad (5.5)$$

where  $\mu_{xi}(t)$  is the mean value of  $x_i$  at time  $t$  and *pre* and *post* are the periods before and after the abrupt shift ( $t_{shift}$ ), respectively. The parameter  $\alpha$  in Equation (5.5) 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 5.2), for two state variables  $x_1$  &  $x_2$  at intervals of  $t = 1$  along the temporal interval  $t = [1, 100]$ .

## Observed process data

I generated observations by imputing noise into the true process model [Equation (5.5)] through random sampling of  $\sigma_{xi}$  from a normal distribution [Equation (5.6); Figure 5.13]:

$$\begin{aligned}\mu_{xipre} &\sim \text{Normal}(\mu_{xipre}, \sigma_{xipre}) \\ \sigma_{xipre} &\sim \text{Normal}(0, \sigma_{Xipre}\mu_{Xipre}) \\ \mu_{xipost} &\sim \text{Normal}(\mu_{ipost}, \sigma_{ipost}) \\ \sigma_{xipost} &\sim \text{Normal}(0, \sigma_{Xipost}\mu_{xipost}) \\ \mu_{xi}(t) &= \mu_{xipre} - 0.5(\mu_{xipre} - \mu_{xipost})(\tanh(\alpha(t - t_{shift})) + 1)\end{aligned}\quad (5.6)$$

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 5.2). I generated the error as a percent of the mean as this scaling relationship is commonly observed in ecological data (Taylor, 1961).

## 5.4.2 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 (5.6)]. I then calculated the distance traveled,  $s$ , and the velocity of the distance traveled,  $V$  (also referred to as  $\frac{ds}{dt}$ ) using Equation (5.4)—this approach (Equation (5.4)) 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, 2019a) 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 5.14). For most conditions

and smoothness I found the tuning parameter for `tvdiff`  $\alpha = 0.50$  provided a good fit of  $s$  (Figure 5.14), however, when the hyperbolic tangent smoothing parameter,  $\alpha$  was low (i.e.  $\alpha_{tanh} = 0.25$ ) higher values of  $\alpha_{tvdiff}$  yielded more abrupt changes in the derivative.

## **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 variables underwent mean shifts (see Table 5.2), and under varying degrees of transition smoothness (i.e.  $\alpha_{tanh} = 0.25, 0.50, 0.75, 1.00$ ).

When the hyperbolic tangent smooth transition function is less steep (Figure 5.15) the observed velocity signal is damped. This signal, however, quickly recovers when the transition function becomes more abrupt (Figures 5.16, 5.17 ,5.18;  $\alpha_{tanh} = 0.5, 0.75, and 1.00$ , respectively). The signal of velocity as an indicator of abrupt change appears more rapid when the mean of both state variables (Figure 5.22), rather than a single (Figure 5.15), are shifted (assuming constant variance). Figure 5.19 is representative of the increasing signal in velocity as  $\alpha_{tanh}$  increases.

## **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 5.20), however, velocity does not signal change when only one variable exhibits a shift in variance (Figure 5.21). 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 5.2)—this percent variability

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.

### **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  $\alpha_{tanh} = 0.25$ ), velocity manifests as a signal of change (Figure 5.23). This signal is most pronounced when the shift is abrupt (Figure 5.24).

## 5.5 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 6. Briefly, the paleodiatom community comprises 109 time series over a period of approximately 6936 years (Figure 5.25). 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 5.25). 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 5.29). The results of  $V$  and  $s$  (Figure 5.26) on the relative abundance data correspond with both the large shifts in species dynamics (see Fig 5.25, and also with the regime shift identified by Spanbauer *et al.* (2014)). However, two primary results can be made from the metrics  $V$  and  $s$  that are not obvious nor identified numerically in the results of 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 5.27).

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 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  $iter = 2000$ ,  $scale = \text{small}$ ,  $ep = 1x10^{-6}$ , 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 5.29) provides a similar but smoother picture of the velocity of the system trajectory. Comparing the smoothed (Figure 5.29) to the non-smoothed velocity (Figure 5.26) 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).

## 5.6 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 5.10), 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 (Figures 5.8,5.19). For example, this method will likely identify signals of regime shifts where the shift is defined as high species turnover within a community (Figure 5.24).

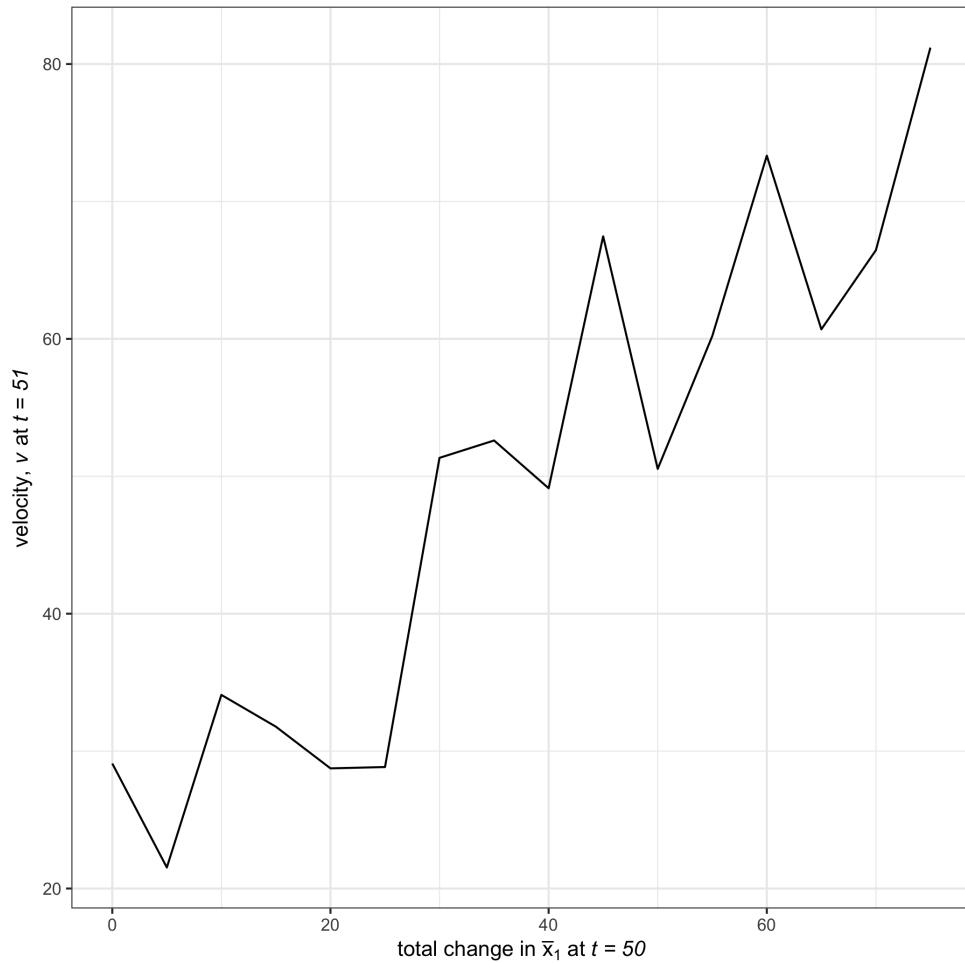
This study provides baseline expectations of the velocity of the distance traveled,  $V$ , as an indicator of abrupt change in a multivariate system. Although a useful first step, this metric should next be critiqued in a sensitivity analytical approach, where a statistical measure is used to determine 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, **regimeDetectionMeasures** (Appendix .4). In this case study, the ‘distance traveled’,  $s$  [Equation (5.1)], 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 5.8).

Further work is required to determine the utility of system velocity as a regime detection metric, however, this chapter demonstrates that the metric may indicate clear shifts in variable means 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

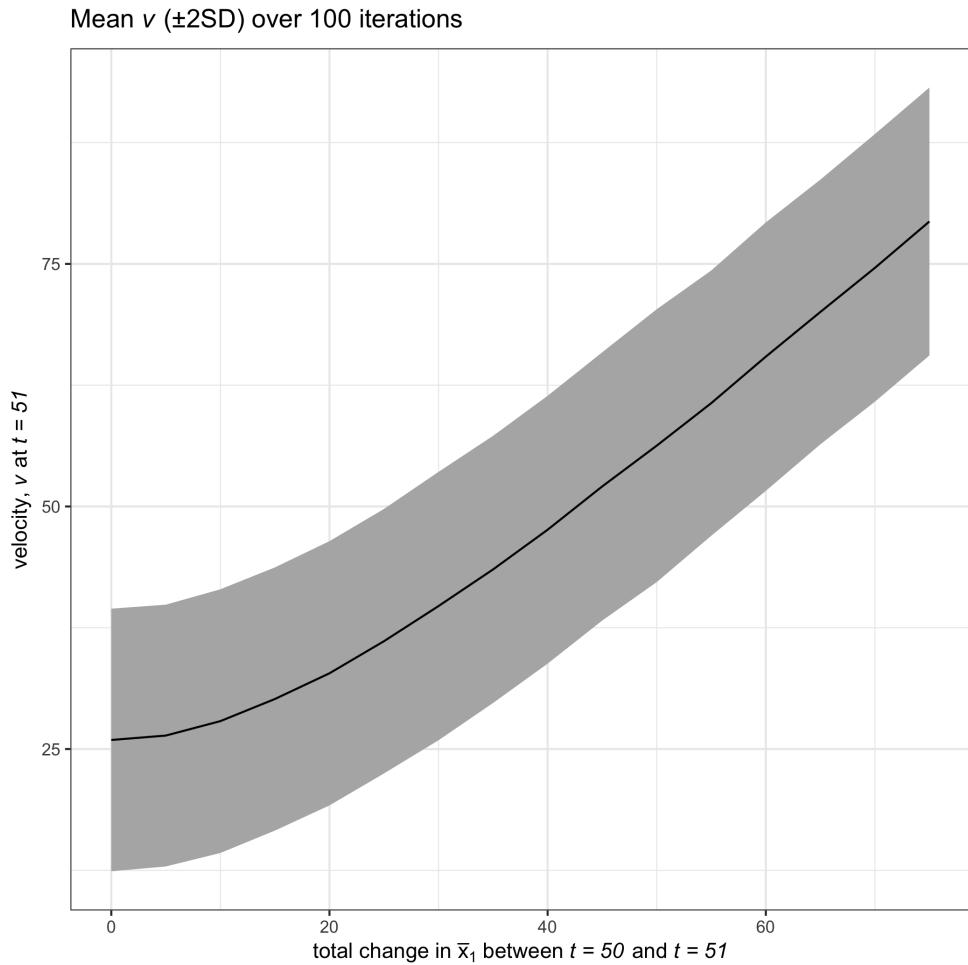
Analysis PCA, the metric is simple to calculate (even by hand), is computationally inexpensive, and is intuitive, unlike many clustering algorithms (e.g., Non-metric Multidimensional Scaling NMDS). Like system velocity, methods of the latter variety (e.g. NMDS) require post-hoc statistical analyses to 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 abrupt shifts.

## 5.7 Supplementary Figures

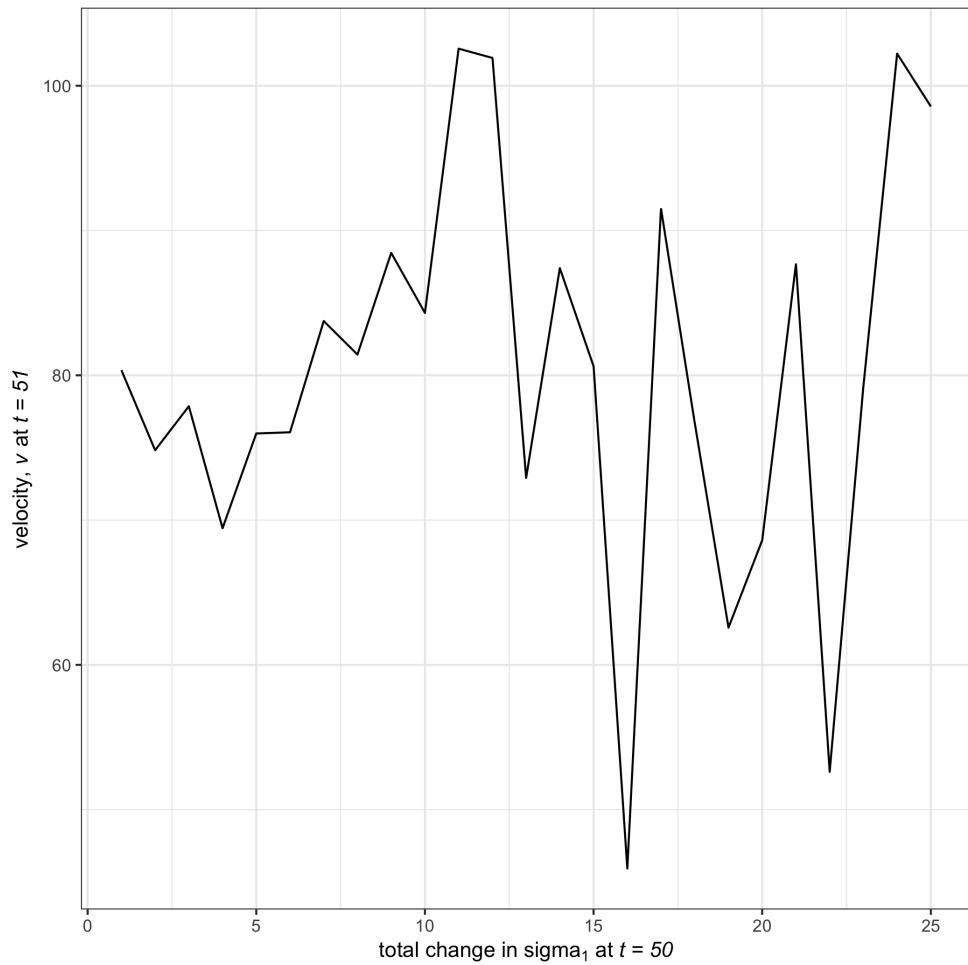
Figures 5.30, 5.31, and 5.32 provide additional examples of the behavior of velocity,  $V$  when varying the mean and/or variance prior to and/or after the induced abrupt shift in the toy system with a discontinuous transition at  $t = 50$ .



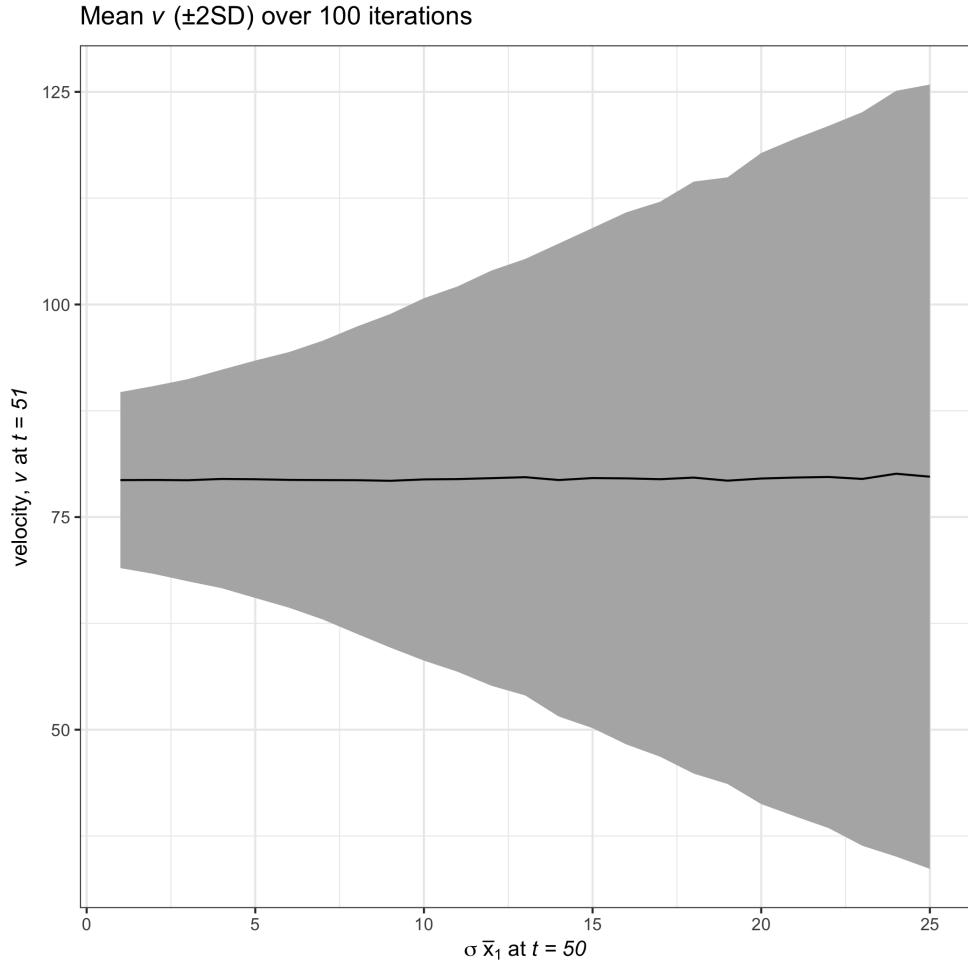
**Figure 5.7:** Velocity ( $V$ ) generally increases as the total change in the mean value of  $\bar{x}_{1_{t=50}}$  increases in a single iteration of our toy system ( $N_{iter} = 1$ , seed = 123). This 2-variable system exhibits a regime shift at  $t = 50$ , where variance is constant  $\sigma = 5$ ,  $\bar{x}_1 = 25$  when  $t < 50$ ,  $\bar{x}_2 = 50$  when  $t \geq 50$ ,  $\bar{x}_1 = 25$  when  $t < 50$ .



**Figure 5.8:** 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  $t < 50$ .

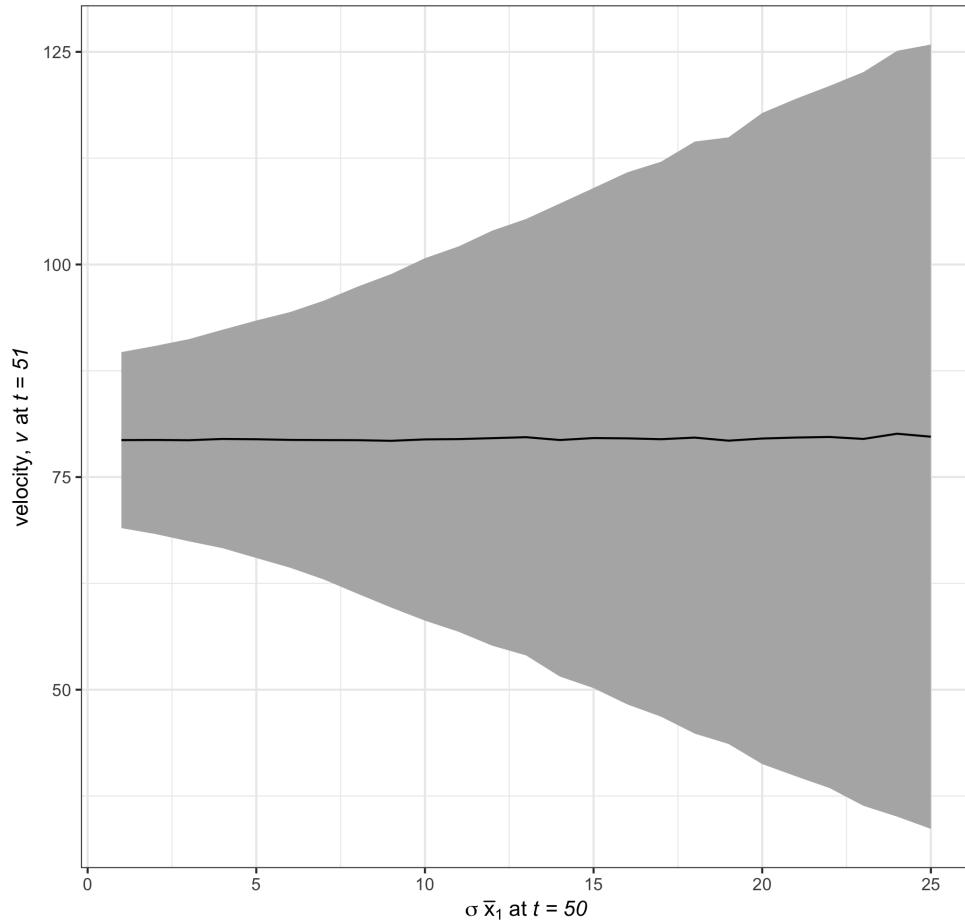


**Figure 5.9:** High variance of velocity ( $V$ ) in a single iteration ( $N_{iter} = 1$ , seed = 123) of simulations as we increase  $\sigma_1$  at  $t = 50$ .

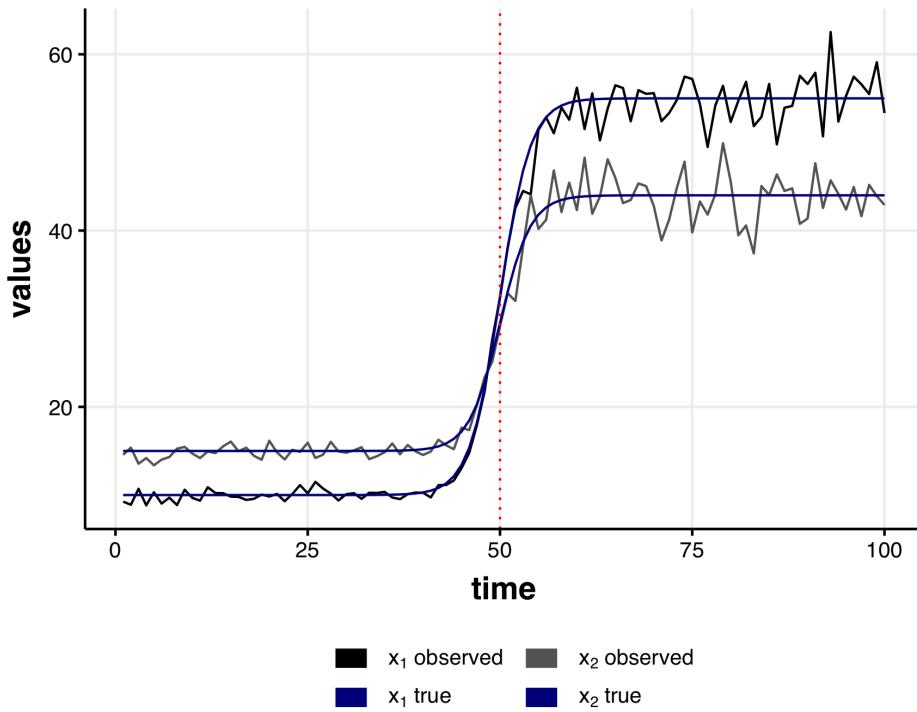


**Figure 5.10:** The variance around the average ( $\pm 2$  SD) velocity ( $V$ ) increases as the variance of  $\bar{x}_{2_{t=50(\text{post})}}$  (post shift) increases.  $\bar{x}_{1_{\text{pre}}} = 25$ ,  $\bar{x}_{1_{\text{post}}} = 100$ ,  $\bar{x}_{2_{\text{pre}}} = 25$ ,  $\bar{x}_{2_{\text{post}}} = 50$ ,  $\sigma_{1_{\text{pre}}} = 5$ ,  $\sigma_{2_{\text{pre,post}}} = 5$

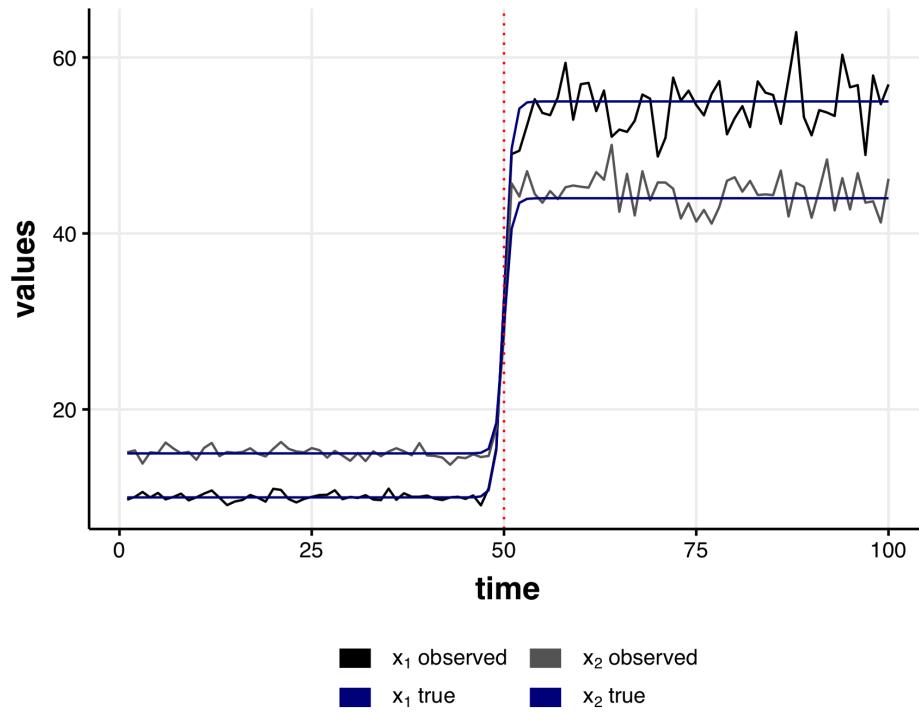
Mean  $v$  ( $\pm 2\text{SD}$ ) over 100 iterations



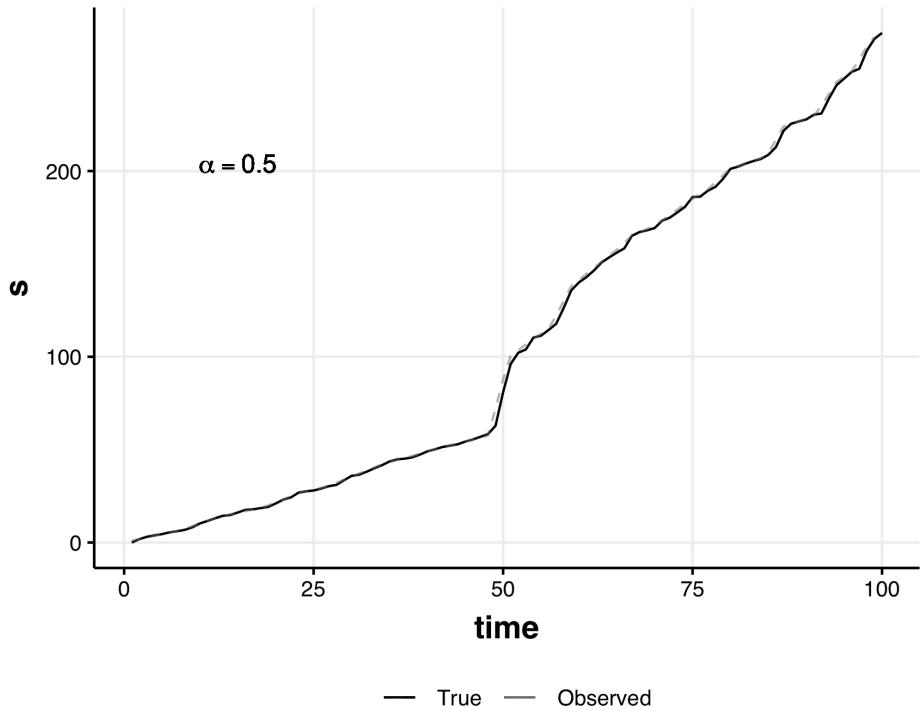
**Figure 5.11:** The noise in system velocity ( $V$ ) is not obviously reduced in this system as the original data ( $x_1, x_2$ ) is increasingly smoothed.



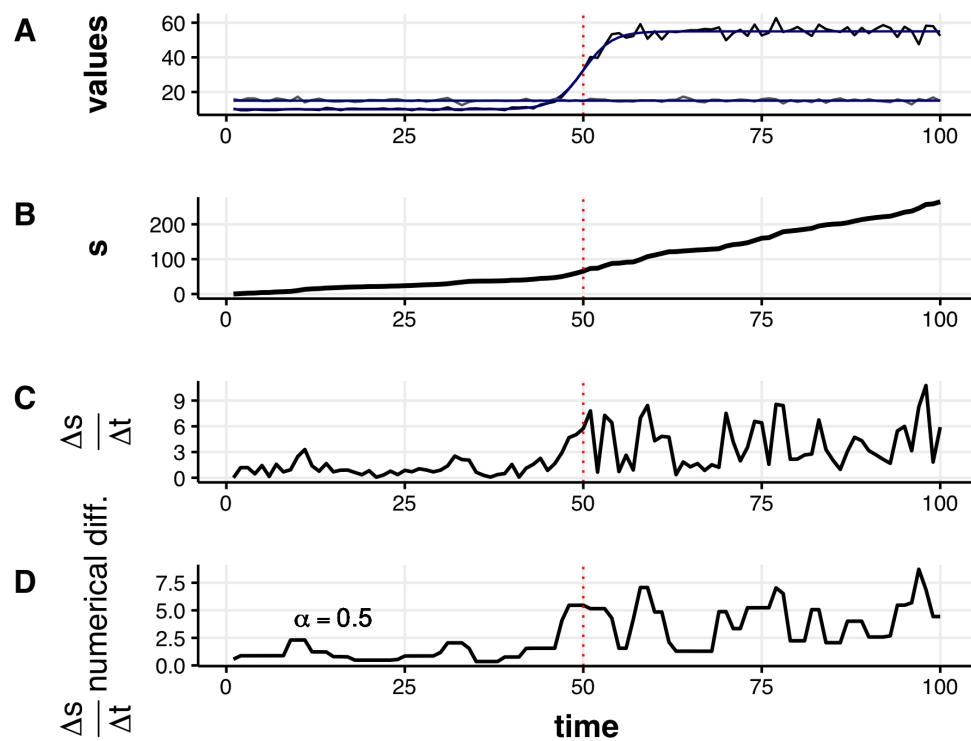
**Figure 5.12:** An example of the data generated by the true process model. In this example the mean values ( $\mu_{xi}$ ), but not the percent standard deviation ( $\sigma_{xi}$ ), are varied before and after the transition point. The observed data are plotted against the true-process model for each state variable,  $x_i$ . Panels represent different degrees of the smoothing parameter,  $\alpha$  (top:  $\alpha = 0.25$ , bottom:  $\alpha = 1.00$ ).



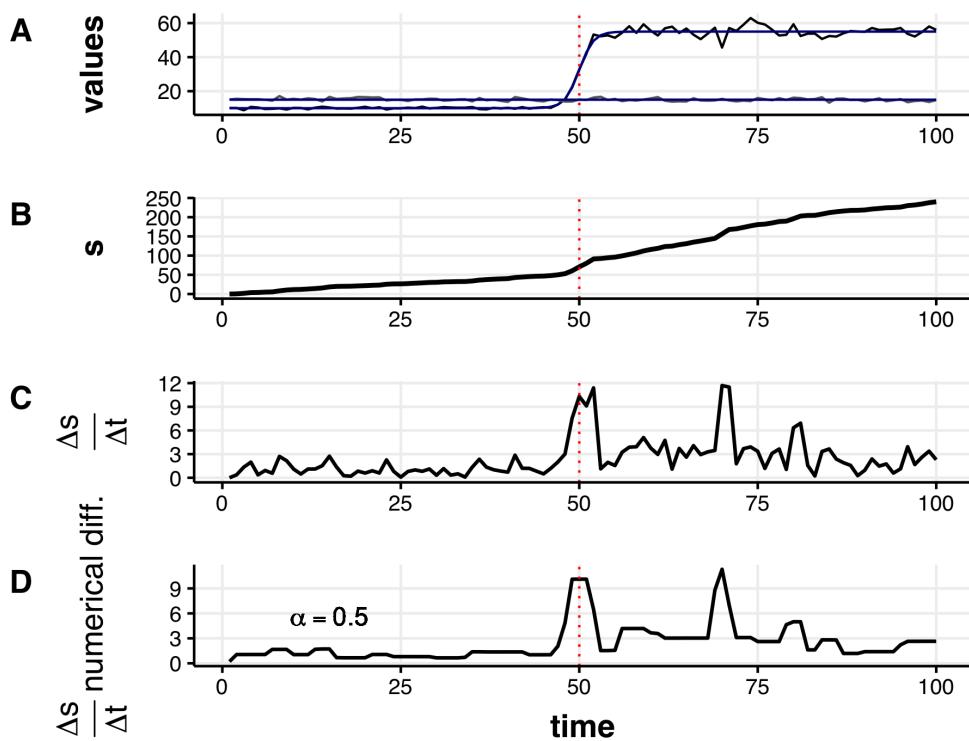
**Figure 5.13:** An example of the data generated by the true process model. In this example the mean values ( $\mu_{xi}$ ), but not the percent standard deviation ( $\sigma_{xi}$ ), are varied before and after the transition point. The observed data are plotted against the true-process model for each state variable,  $x_i$ . Panels represent different degrees of the smoothing parameter,  $\alpha$  (top:  $\alpha = 0.25$ , bottom:  $\alpha = 1.00$ ).



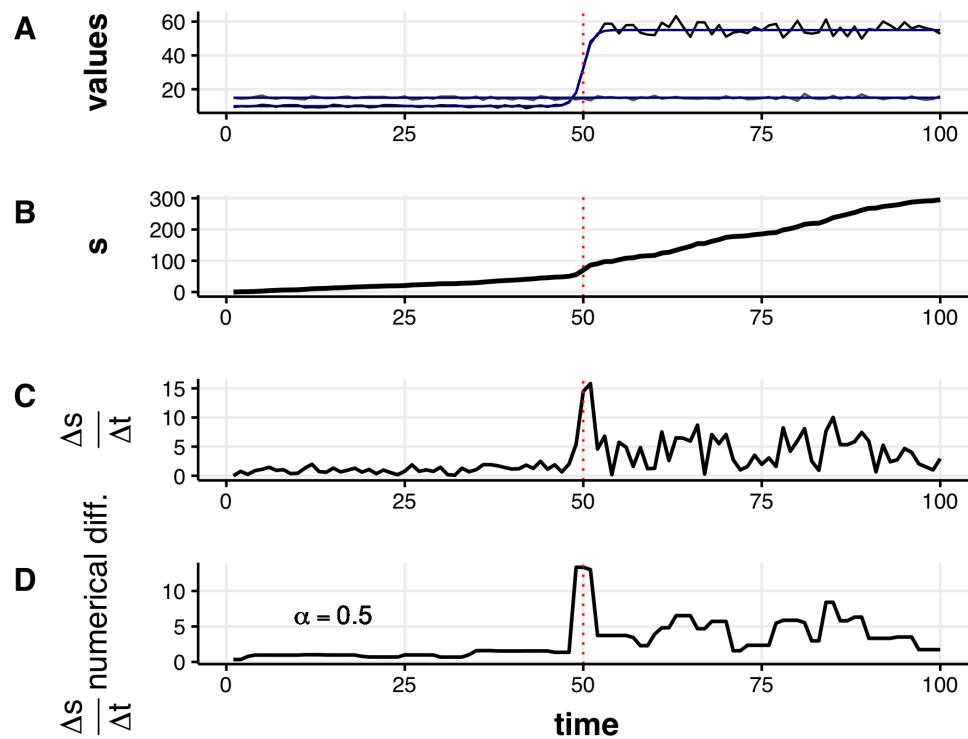
**Figure 5.14:** Antidifferentiated values ('observed') of the distance traveled,  $s$ , to the true process values of  $s$  ('true) provides a method for identifying the best values of the smoothing parameter,  $\alpha$ . Under most conditions  $\alpha \ll$  sufficed. Here, we compare the true and antidifferentiated values of  $s$  under the condition of changing  $\mu_{x1}$  when the hyperbolic tangent function is most rapid ( $\alpha_{\tanh} = 1$ ) for the 'tvdiff'  $\alpha = 0.50$ . Not pictured: the antidifferentiated values of  $s$  (observed) is increasingly smoothed as  $\alpha$  increases.



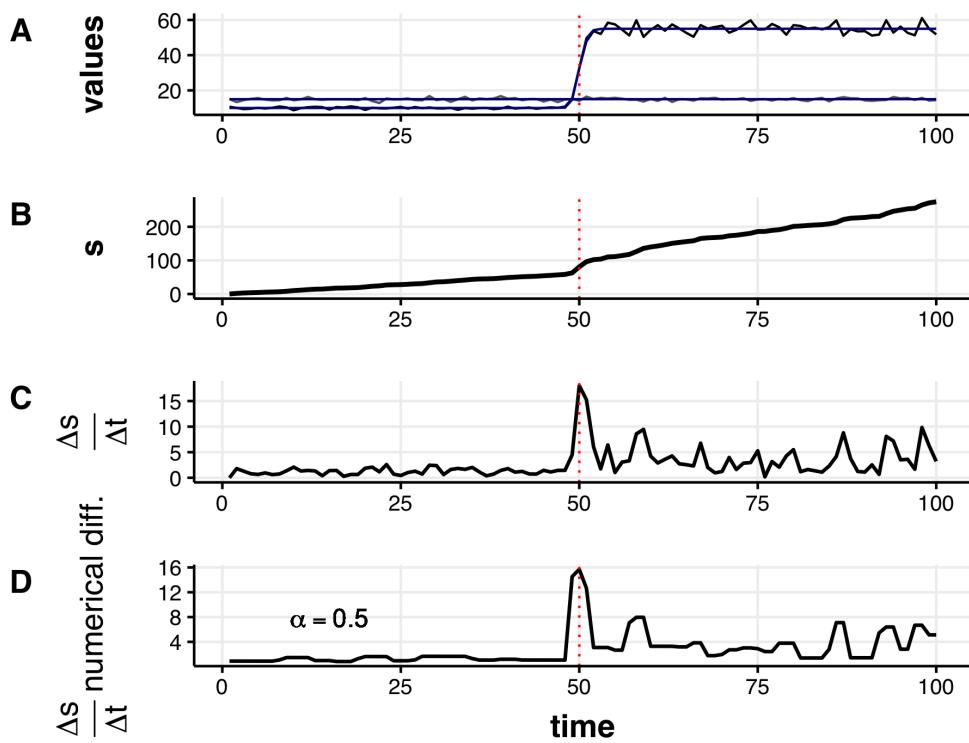
**Figure 5.15:** The velocity signal is muted when the hyperbolic smoothing parameter,  $\alpha$ , is low (0.25). True and observed values of  $x_i$  (panel A), observed distance traveled ( $s$ , panel B), observed velocity (C), and the smoothed velocity (D).



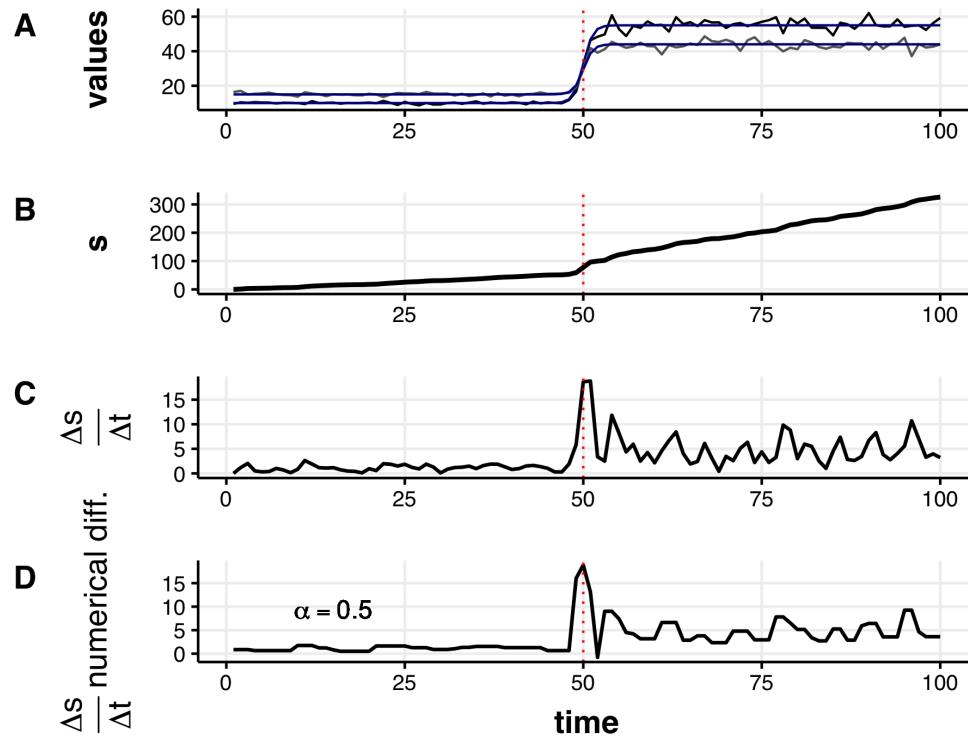
**Figure 5.16:** The velocity signal is muted when the hyperbolic smoothing parameter,  $\alpha$ , is moderate (0.50). True and observed values of  $x_i$  (panel A), observed distance traveled ( $s$ , panel B), observed velocity (C), and the smoothed velocity (D).



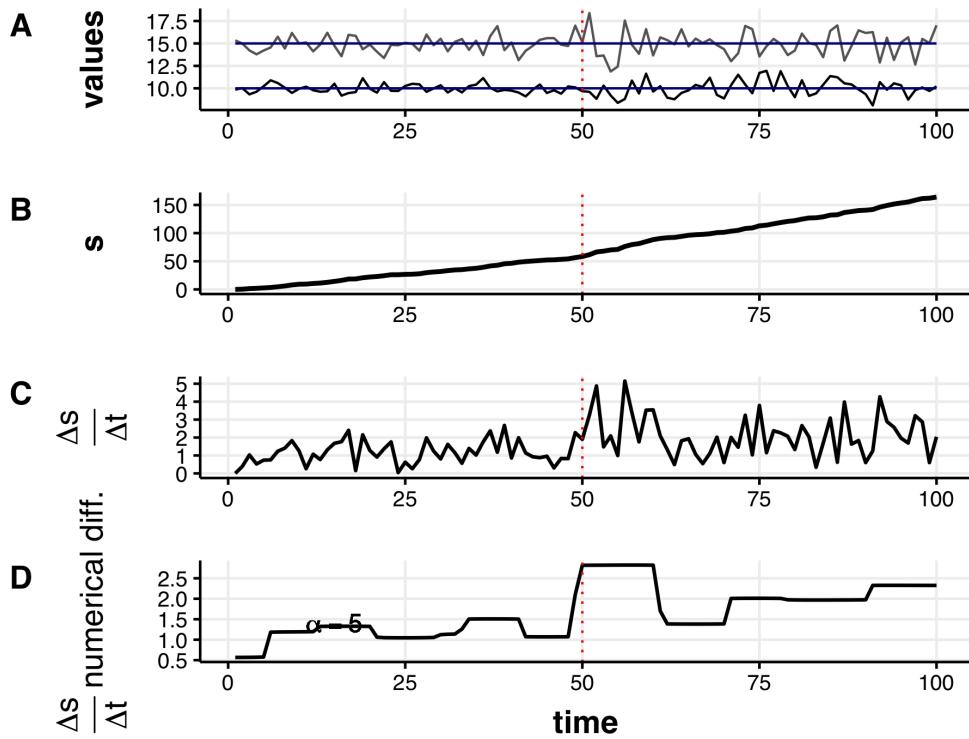
**Figure 5.17:** The velocity signal is muted when the hyperbolic smoothing parameter,  $\alpha$ , is moderate (0.50). True and observed values of  $x_i$  (panel A), observed distance traveled ( $s$ , panel B), observed velocity (C), and the smoothed velocity (D).



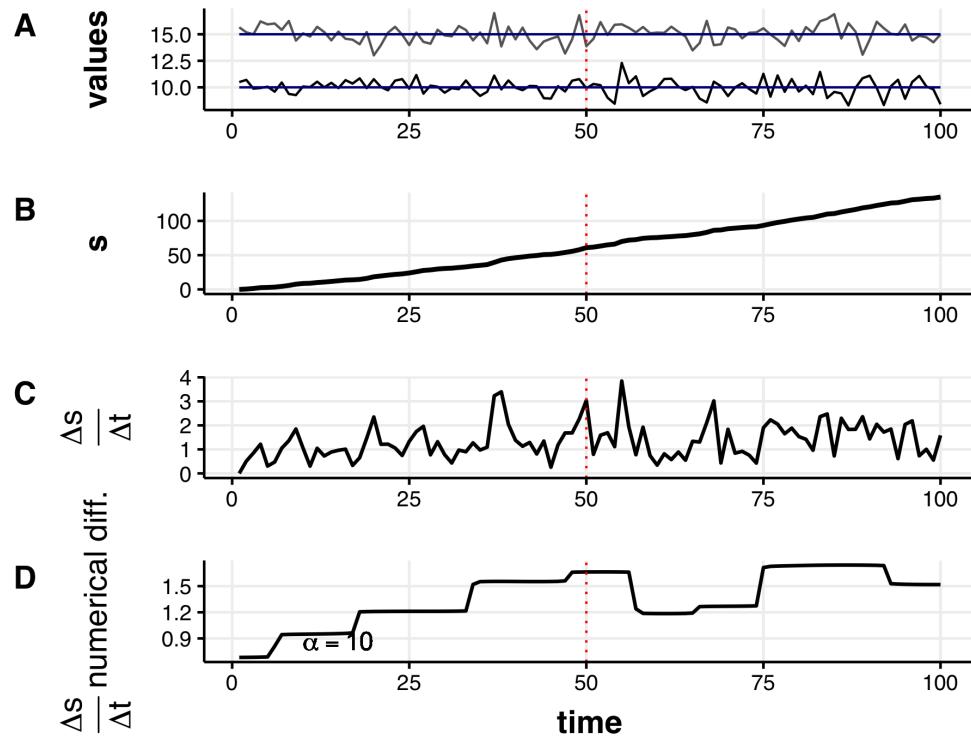
**Figure 5.18:** The velocity signal is muted when the hyperbolic smoothing parameter,  $\alpha$ , is moderate (0.50). True and observed values of  $x_i$  (panel A), observed distance traveled ( $s$ , panel B), observed velocity (C), and the smoothed velocity (D).



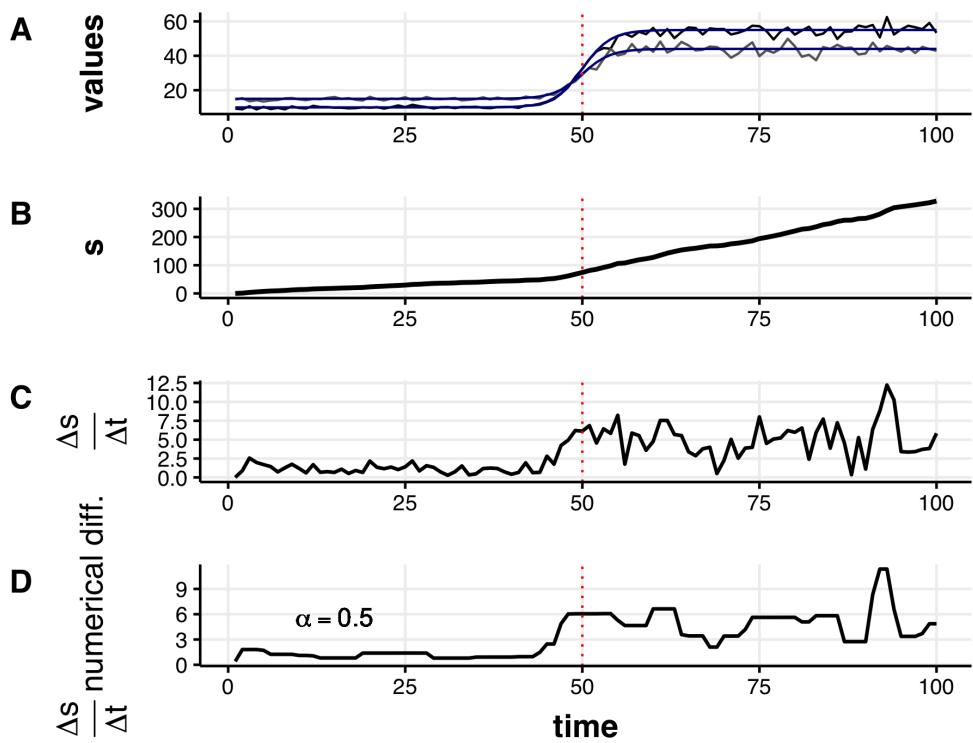
**Figure 5.19:** The velocity signal is regained under smooth transition ( $\alpha_{\text{tanh}} = 0.75$ ) when both state variables undergo a shift in the mean. True and observed values of  $x_i$  (panel A), observed distance traveled ( $s$ , panel B), observed velocity (C), and the smoothed velocity (D).



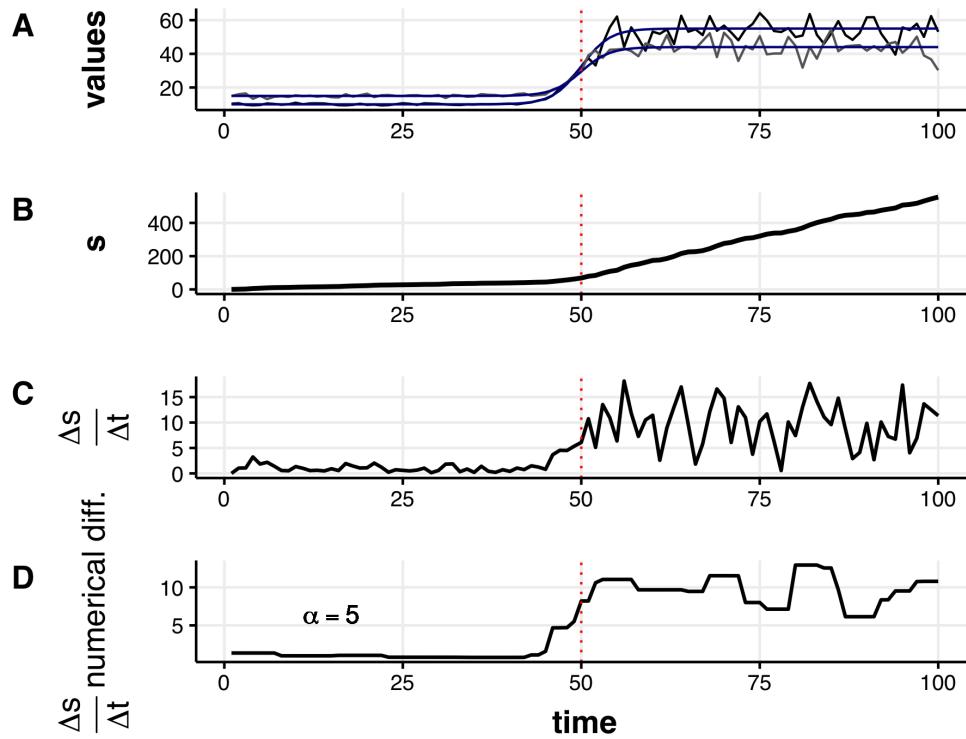
**Figure 5.20:** The velocity signals a rapid shift in the variance of both state variables under a moderately abrupt transition ( $\alpha_{\tanh} = 0.75$ ). True and observed values of  $x_i$  (panel A), observed distance traveled ( $s$ , panel B), observed velocity (C), and the smoothed velocity (D).



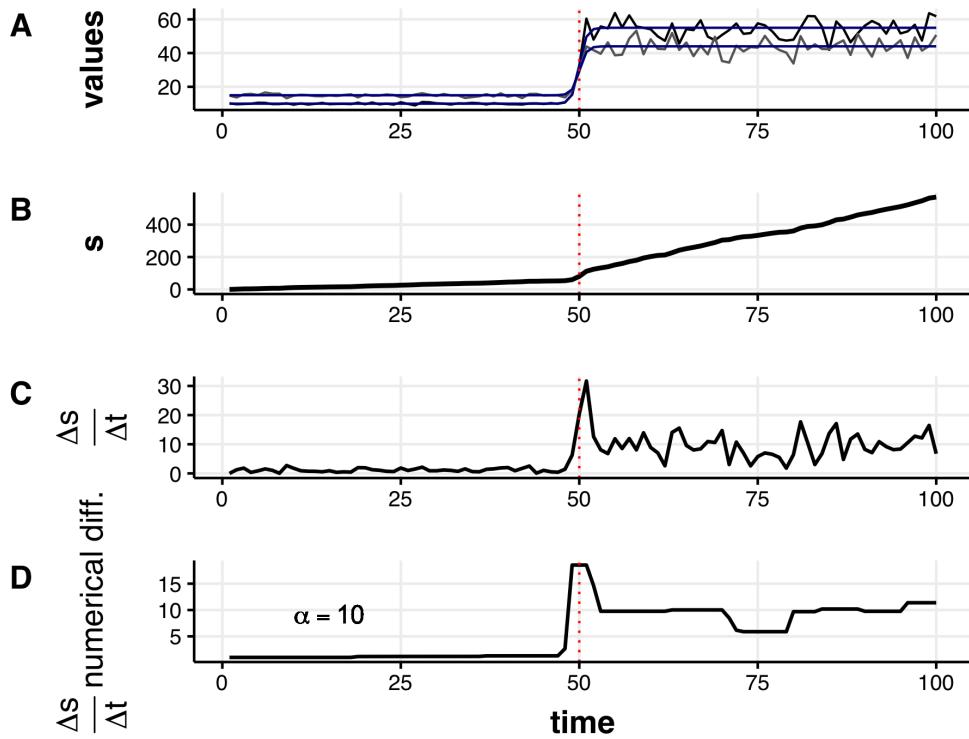
**Figure 5.21:** The velocity does not signal shifts in the variance of a single variable ( $x_1$ ) under a moderately abrupt transition ( $\alpha_{\text{tanh}} = 0.75$ ). True and observed values of  $x_i$  (panel A), observed distance traveled ( $s$ , panel B), observed velocity (C), and the smoothed velocity (D).



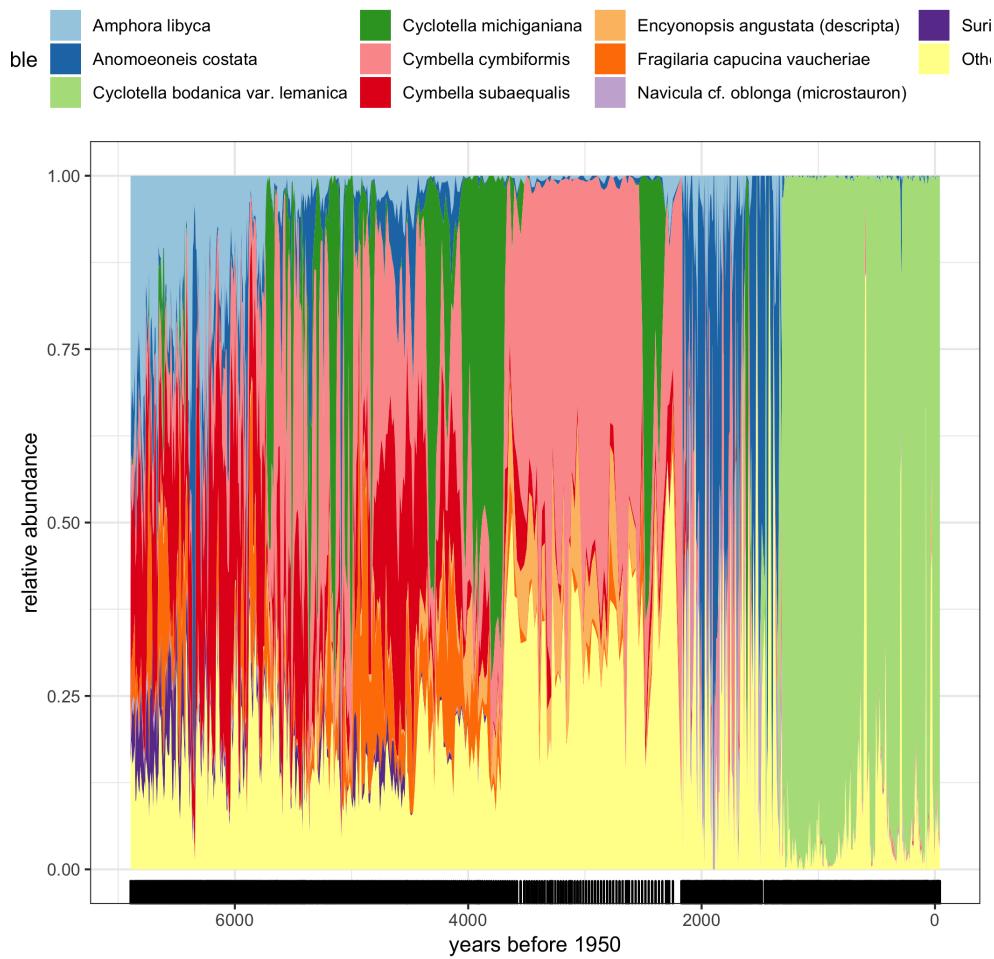
**Figure 5.22:** The velocity signal is regained under smooth transition ( $\alpha_{\text{tanh}} = 0.25$ ) when both state variables undergo a shift in the mean. True and observed values of  $x_i$  (panel A), observed distance traveled ( $s$ , panel B), observed velocity (C), and the smoothed velocity (D).



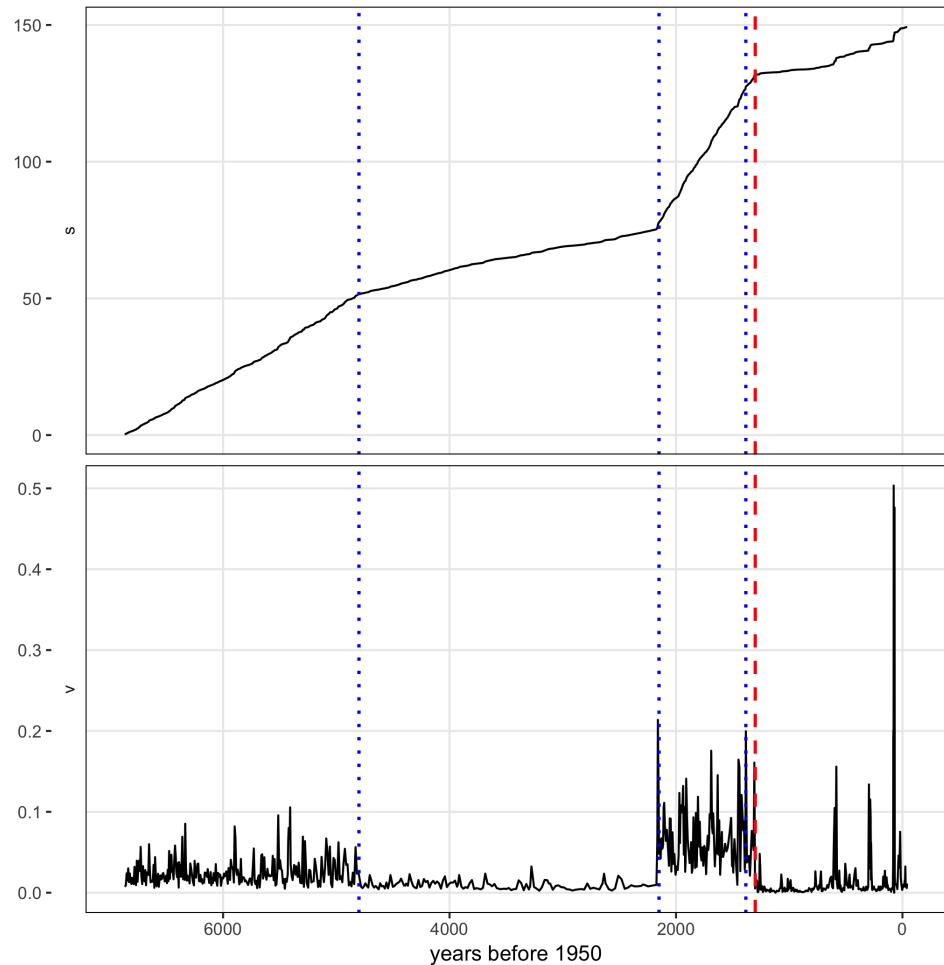
**Figure 5.23:** The velocity signals a shift when both variables undergo shifts in the mean and variance under a slightly abrupt transition ( $\alpha_{tanh} = 0.25$ ). True and observed values of  $x_i$  (panel A), observed distance traveled ( $s$ , panel B), observed velocity (C), and the smoothed velocity (D).



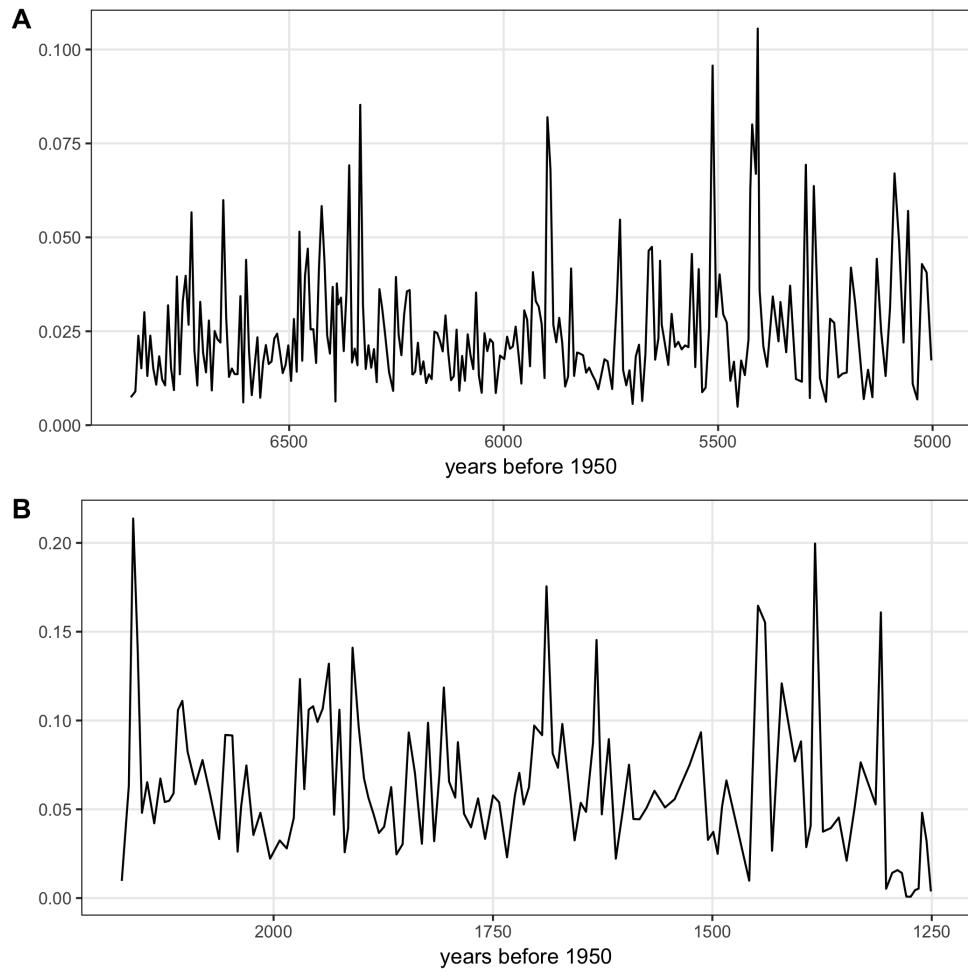
**Figure 5.24:** The velocity signals a shift when both variables undergo shifts in the mean and variance under a slightly abrupt transition ( $\alpha_{tanh} = 1.00$ ). True and observed values of  $x_i$  (panel A), observed distance traveled ( $s$ , panel B), observed velocity (C), and the smoothed velocity (D).



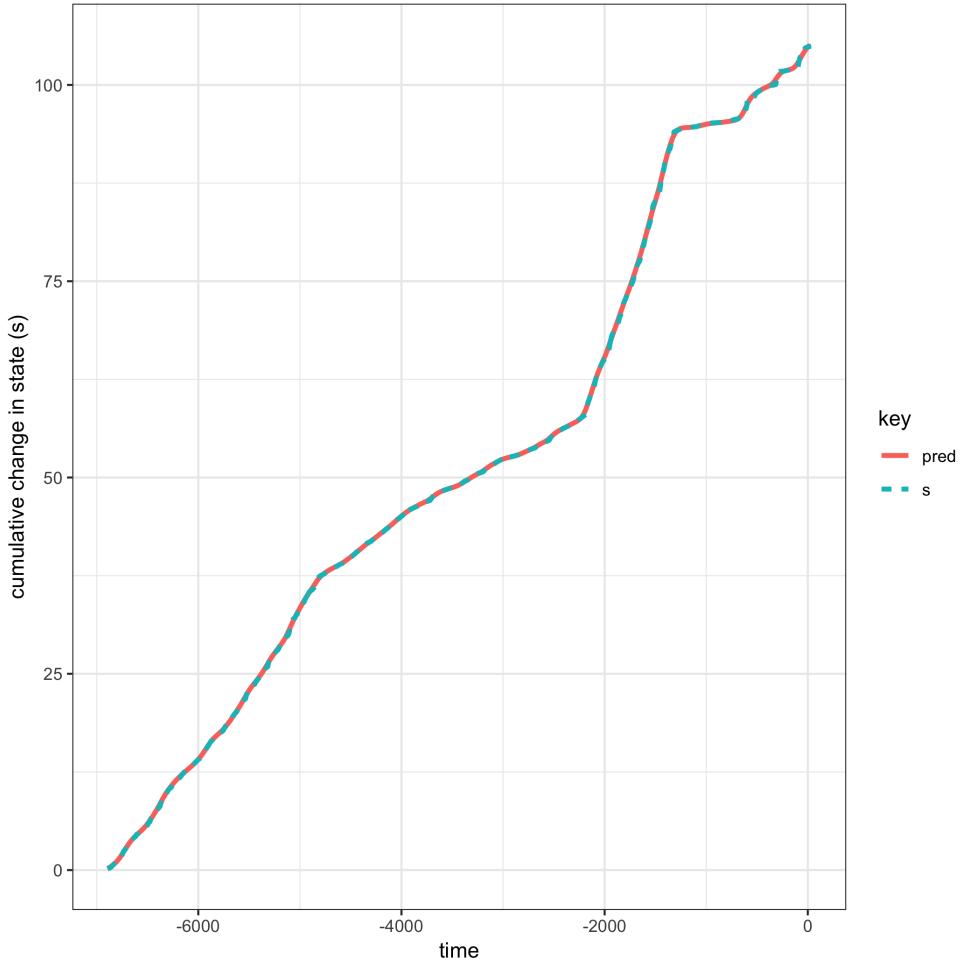
**Figure 5.25:** Relative abundances of the most common diatom species in the time series. Few species dominate the data over the entire time series, and turnover is apparent at multiple observations.



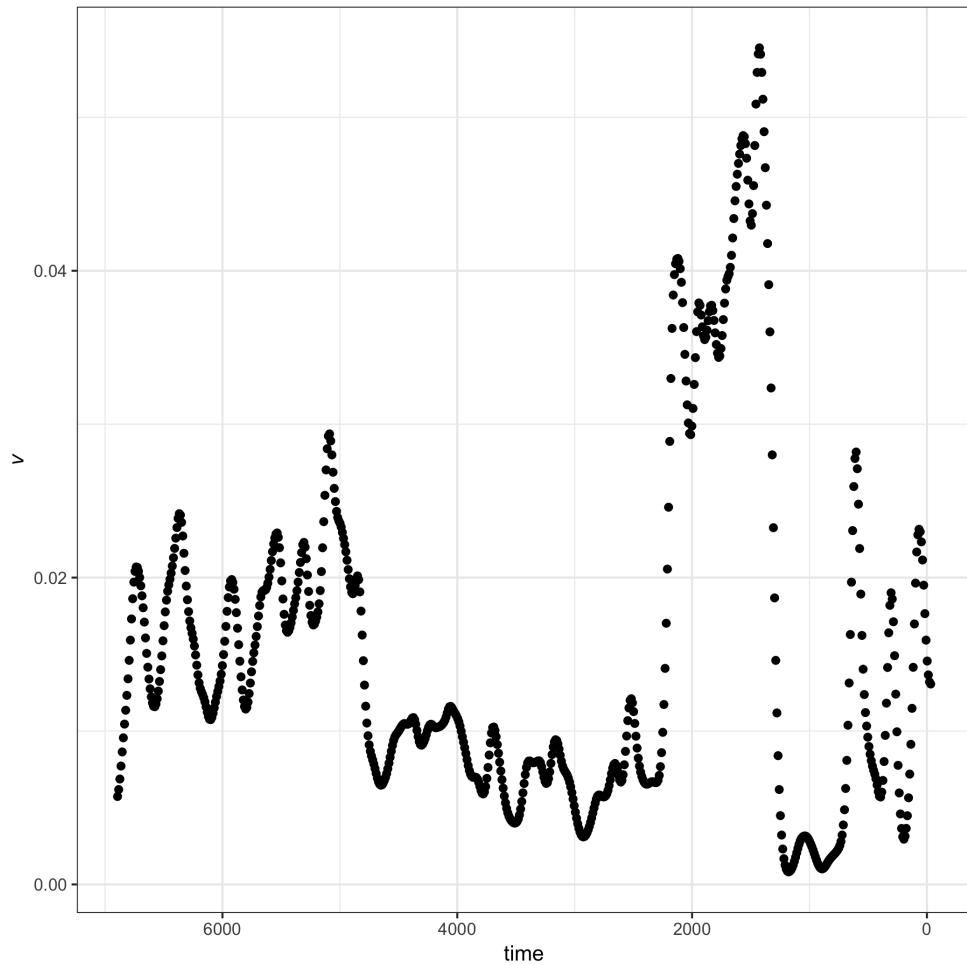
**Figure 5.26:** Velocity  $V$  and distance traveled  $s$  of the paleodiatom time series. Dashed line at 1,300 years before 1950 indicates the regime shift identified in Spanbauer et al. (2014). Dotted lines indicate regime shifts as visually identified on metrics  $s$  and  $V$ .



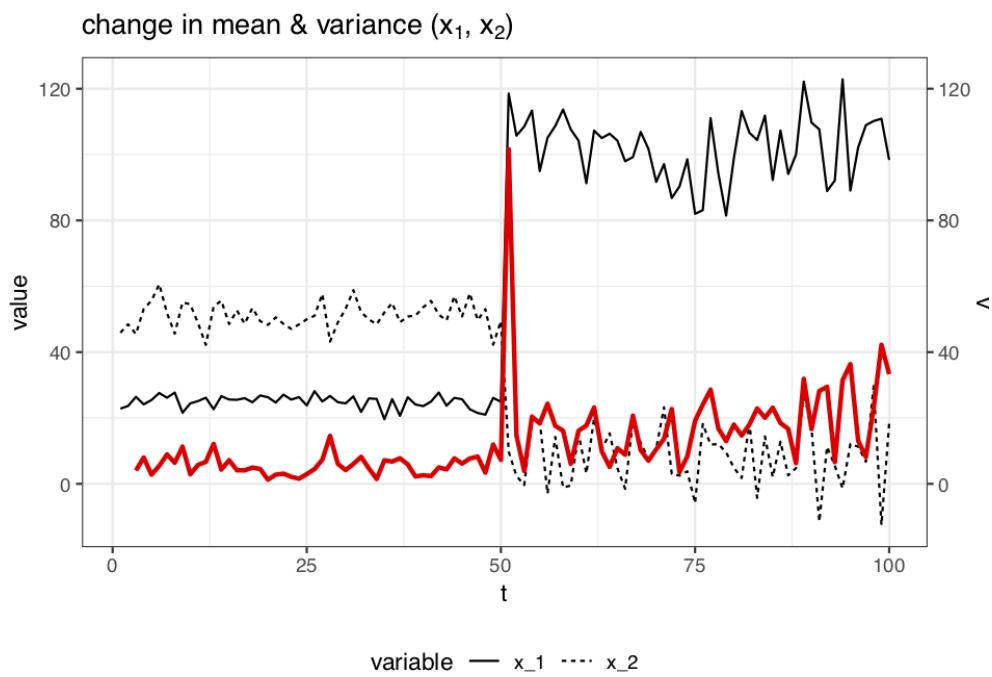
**Figure 5.27:** Inter-regime (regimes A and B) trends in the velocity signal ( $V$ ) identify the fluctuating decadal and centurial abiotic conditions discussed in Spanbauer et al. (2014).



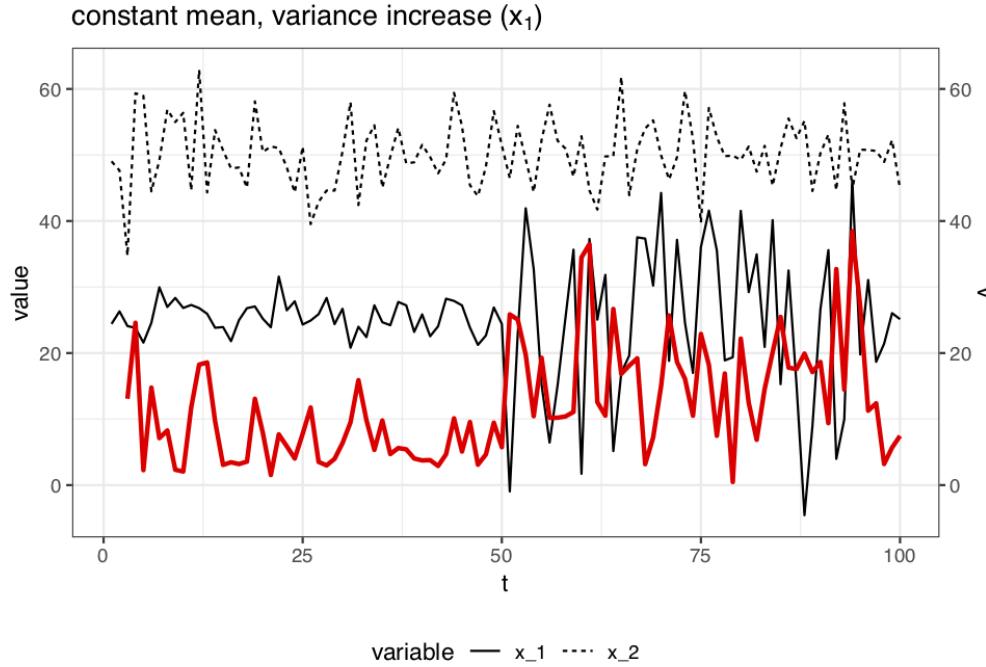
**Figure 5.28:** The regularized differentiation of  $s$  was best fit using  $\alpha = 100$ . Higher overlap of  $s$  and  $\text{pred}$  indicates a good fit of the regularized differentiated metric to the non-smoothed metric,  $s$ .



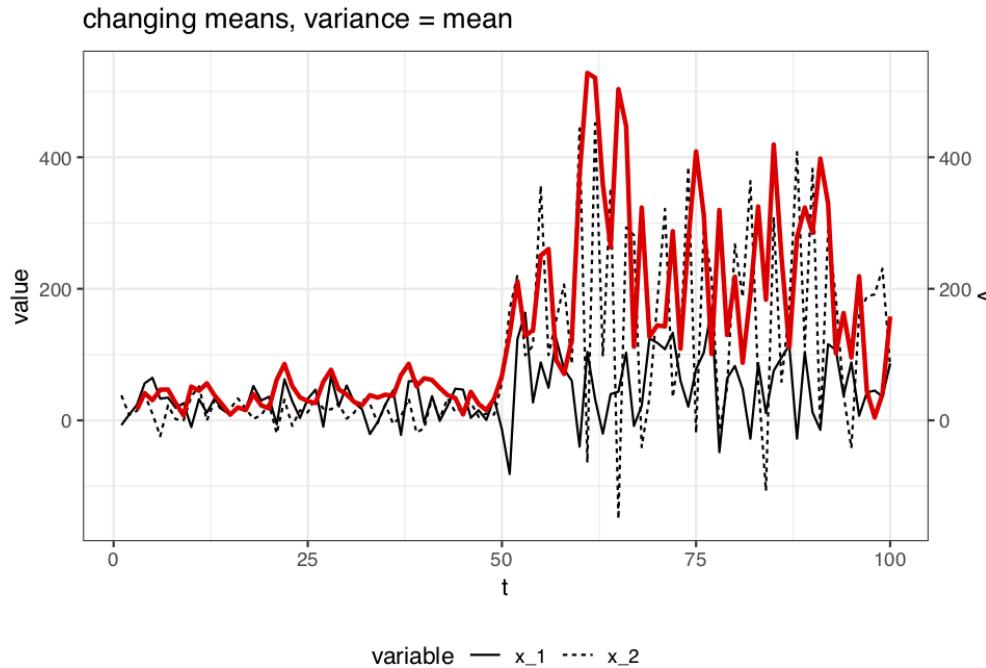
**Figure 5.29:** The velocity metric ( $V$ ) signals potential periodicities in the paleo diatom time series data when the distance traveled metric,  $s$ , is smoothed using regularized differentiation methods (see Price and Burnett 2019).



**Figure 5.30:** System change ( $s$ ) and velocity ( $V$ ) of the model system over the time period. Change in means ( $\bar{x}_{1_{pre}} = 25, \bar{x}_{1_{post}} = 100, \bar{x}_{2_{pre}} = 50, \bar{x}_{2_{post}} = 10$ ) and an increase in variance ( $\sigma_{1_{pre}} = 2, \sigma_{1_{post}} = 10, \sigma_{2_{pre}} = 5, \sigma_{2_{post}} = 10$ ).



**Figure 5.31:** System change ( $s$ ) and velocity ( $V$ ) of the model system over the time period. Constant means ( $\bar{x}_1 = 25$ ,  $\bar{x}_2 = 50$ ) and sharp change in variance for one state variable  $\sigma_{1_{pre}} = 2$ ,  $\sigma_{1_{post}} = 12$ ,  $\sigma_{2_{pre,post}} = 5$



**Figure 5.32:** 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 ( $/bar{x}_{1_{pre}} = 25$ ,  $/bar{x}_{1_{post}} = 50$ ,  $/bar{x}_{2_{pre}} = 15$ ,  $/bar{x}_{2_{post}} = 150$ ).

# Chapter 6

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

### 6.1 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; Karunamithi *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 5) 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 accompanying this Chapter allows the end user to implement these methods on this or their own system data, or on theoretical data.

## 6.2 Data and Methodology

### 6.2.1 Study system and data

I used paleodiatom time series from a freshwater system in North America (Foy Lake, present day Montana) that apparently underwent 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  $\approx 7,000$  years (Figure 6.1). 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 6.2). The data were published in Spanbauer *et al.* (2014) and can be downloaded at the publisher's website.

### 6.2.2 Regime detection measures

Fewer model-free regime detection metrics exist than do model-based metrics (Chapter 2) and of these, only a few are suggested for multivariate data. Here, I compare the results for three regime detection metrics that are model-free and can handle multivariate data: velocity (Chapter 5), 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 5 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.

## Velocity ( $v$ ) calculation

In Chapter 5, 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. (6.1)). Steps for calculating this metric are described in detail in Chapters 3 and 5.

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

## 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).

## 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 5) of the distance traveled by the system,  $s$  over some period of time ( $T$ ), and is given in Eq. (6.2):

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

I refer the reader to Chapter 3 for a complete description and to Cabezas & Fath (2002) for a complete derivation of Fisher Information.

## 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 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 poitns 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[which value] to any other point in time (e.g., the beginning or the middle) muddles the interpretation of the metric over  $T^*$ . Also note that this method has the potential to result in calculating a metric for all integers between  $0.20T^*$  and  $T^*$ .

### 6.2.3 Simulating data quality and quantity issues using re-sampling 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 information 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:  $\mathbf{P} = [0.25, 0.50, 0.75, 1.00]$ , where  $\mathbf{P}$  is the proportion of species and time points **retained** for analysis. For example, when  $\mathbf{P} = 0.25$ , a random selection of 25% of the species are retained for analysis in the species scenario. I re-sampled the data over 10,000 iterations ( $N_{samp}$ ) for each scenario and  $\mathbf{P}$  combination. Note that because when  $\mathbf{P} = 1.00$ , all data are retained. Therefore, no resampling was conducted at this level because only a single

metric (e.g. Velocity) value is possible.

#### 6.2.4 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  $CV \rightarrow 1$ , 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.

## 6.3 Results

### 6.3.1 Velocity of the distance travelled ( $v$ )

The velocity of the distance traveled,  $\frac{ds}{dt}$  or  $v$ , exhibited dispersion across the values of  $\mathbf{P}$ , however, yielded consistent results (i.e., high overlap in the densities of the CV across values of  $\mathbf{P}$  and across methodologies; e.g. Figures 6.9 and 6.3). 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 6.8), 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 6.6) and temporal observation loss (Figure 6.7) on the velocity metric. These conditions are representative of the other **P-M** combinations.

### 6.3.2 Variance Index

The Variance Index (VI) performed best under the the observations resampling method, exhibiting low values for and low dispersion in the CV density (Figure 6.3) 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 6.3). Surprisingly, the Variance Index was insensitive to temporal observation loss (Figure 6.4), exhibiting a similar amount of noise across various degrees of data loss ( $\mathbf{P}$ ). Although the signal was damped under the species method, the signals for the shifts in community composition were not lost across levels of  $\mathbf{P}$ .

(Figure 6.5). This is likely due to the high probability that the dominant species were rarely always excluded from the re-sampled observations.

### 6.3.3 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 6.10, 6.9). 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 6.10). 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 6.9). 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 6.9). The high variation in FI values across re-sampled iterations coupled with the high dispersion within each **M-P** combination (Figure 6.9) 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 6.10).

## 6.4 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 “spanning” parameter must be chosen when detrending (smoothing) non-linear time series using local regression (Loess) regression (see Figure 6.11). 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 disproportionate 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 removal, and other related decisions would be of value in understanding the efficacy of these and other regime detection measures in real-world situations.

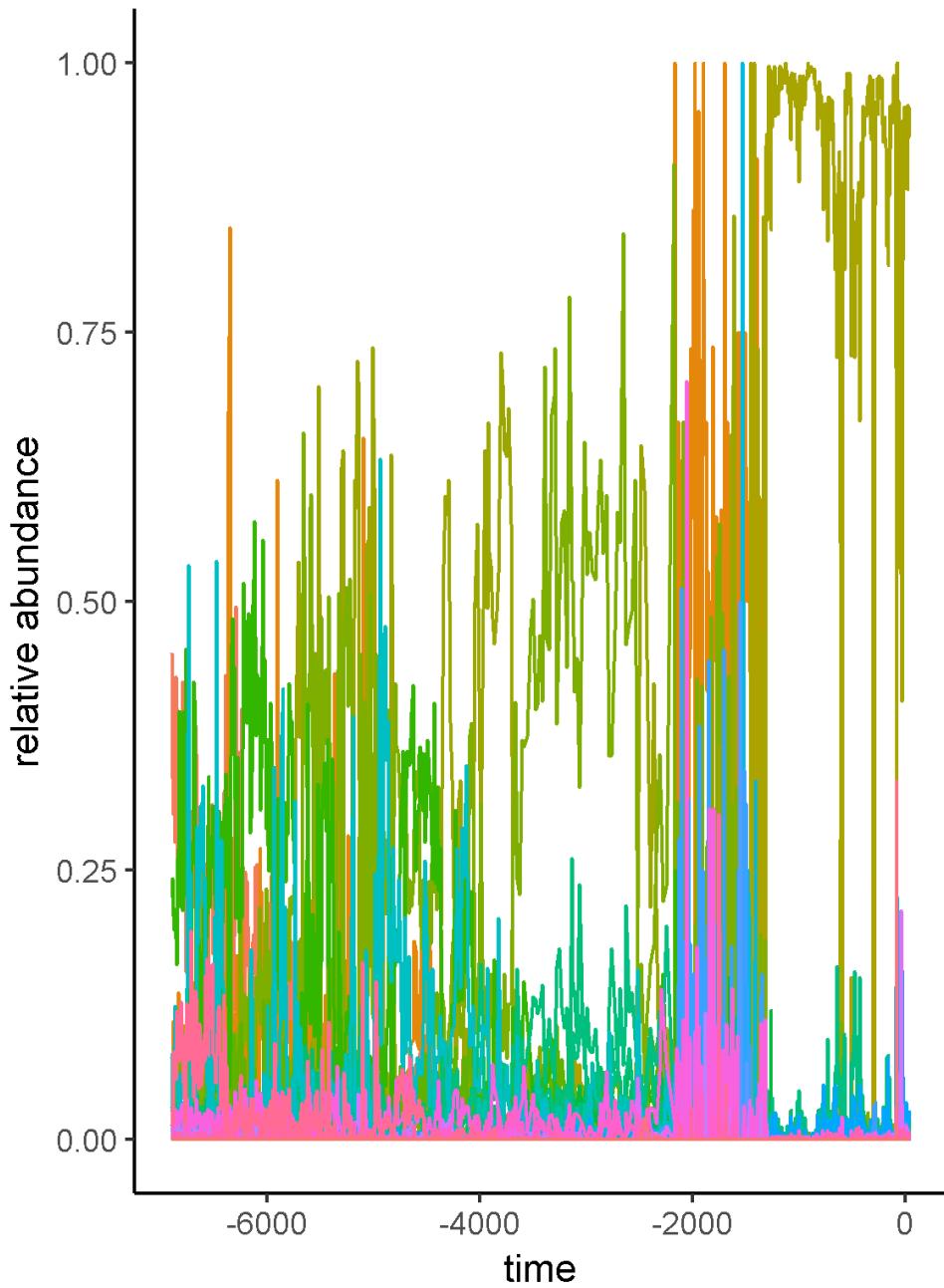
## 6.5 Conclusion

In this chapter I provide additional evidence for the sensitivity of select regime detection 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

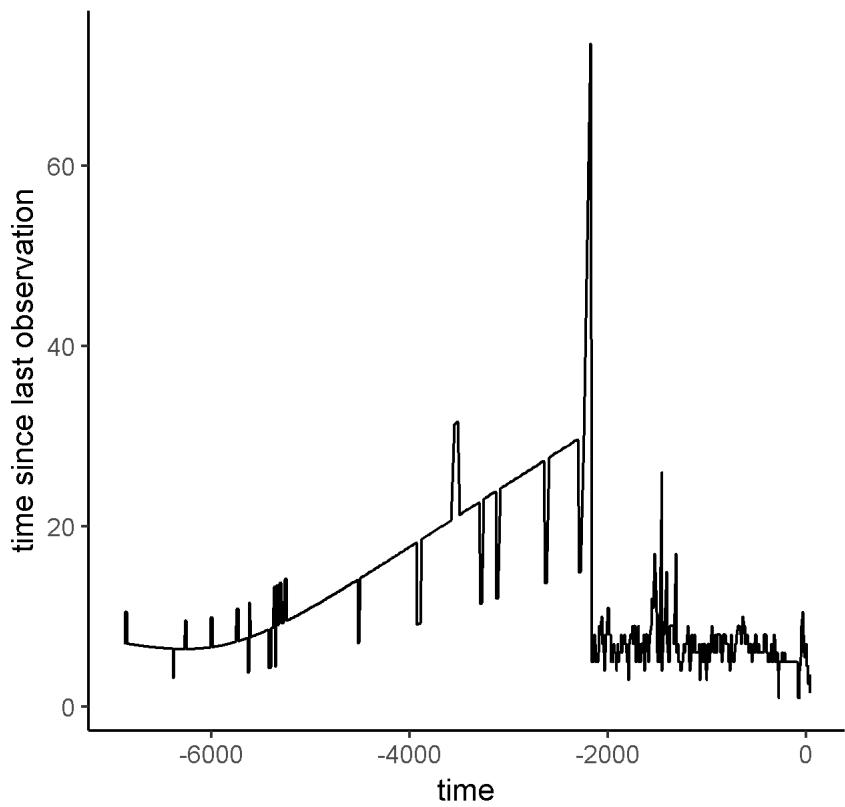
temporal observations, and the reduction in data quality manifests as a function of removing whole species from the community profile. Previous studies of the robustness of uni-variate regime detection metrics have found similar results, suggesting the measures fail in numerous real-world ecological conditions (Andersen *et al.*, 2009; Contamin & Ellison, 2009). This chapter also highlights the relative insensitivity of the new velocity metric (see Chapters 3, 5) to data and information quality and quantity (e.g., Figure 6.6) loss.

## 6.6 Acknowledgements

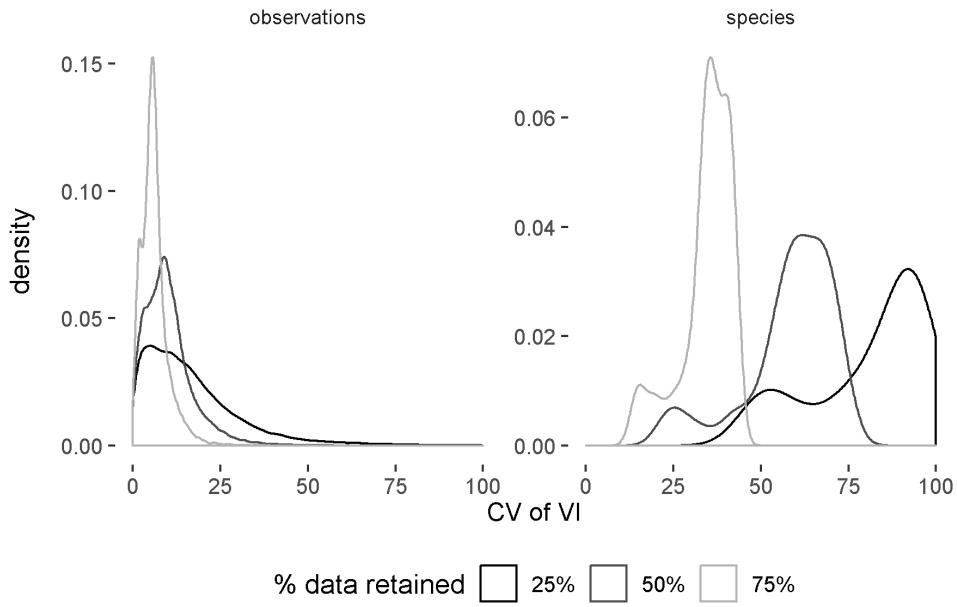
This study was conceptualized at the International Institute for Applied Systems Analysis (IIASA) as part of the Young Scholars Summer Program in 2018. I thank my IIASA program supervisors, Drs. Brian Fath and Elena Rovenskaya, for their advisement of this project and IIASA scientists Drs. Matthias Jonas, Chai Molina, Piotr Zebrowski for feedback on study design.



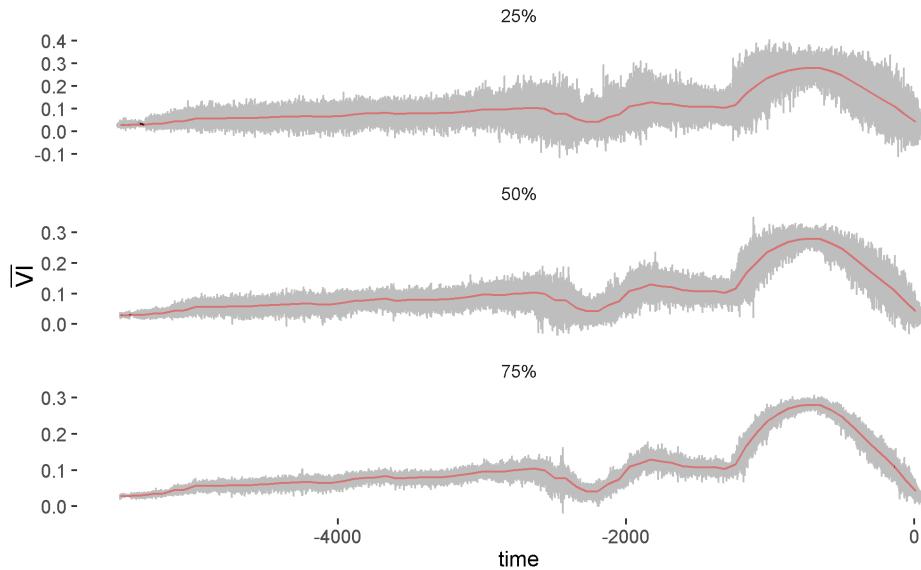
**Figure 6.1:** Relative abundances of the diatom species in Foy Lake over the time period.



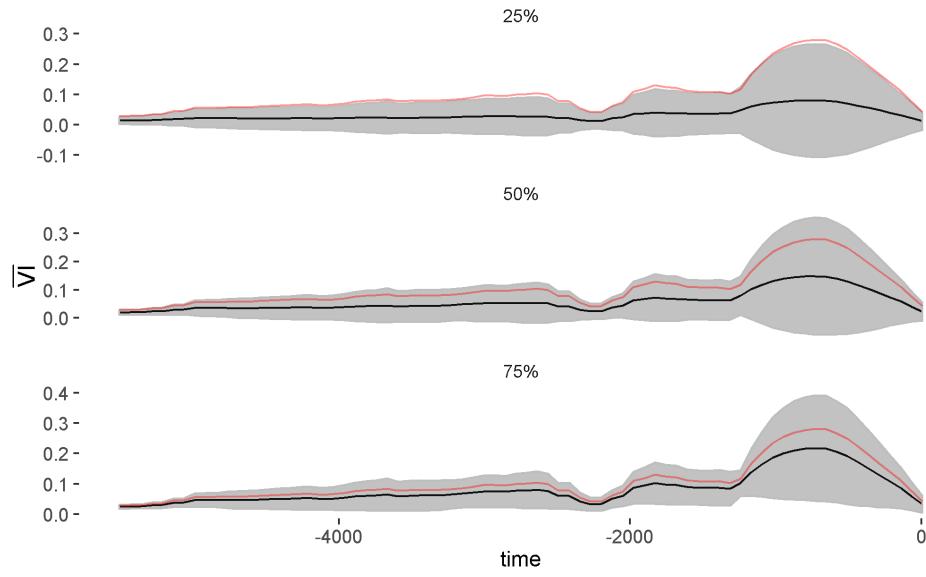
**Figure 6.2:** *The amount of time elapsed between observations for the Foy Lake paleodiatom data.*



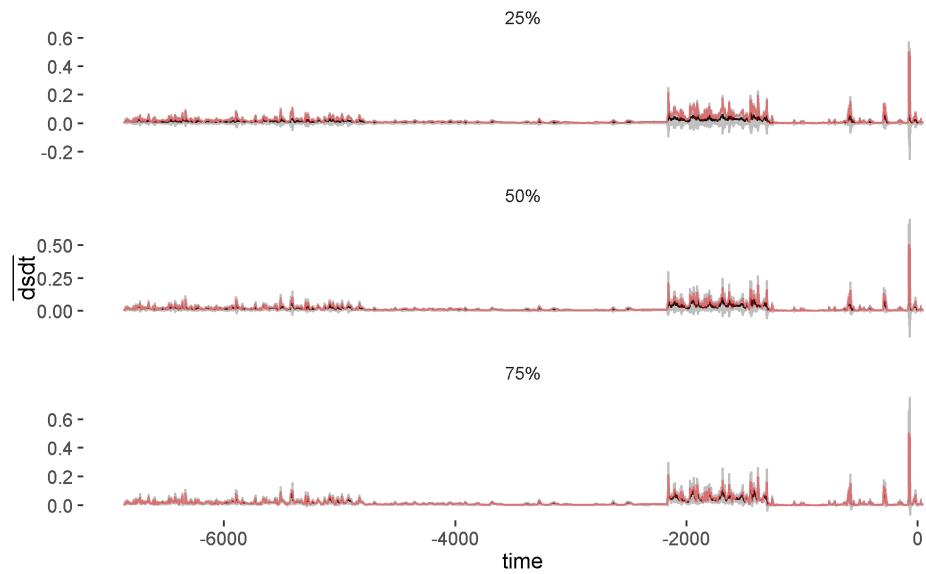
**Figure 6.3:** Density plot of the coefficient of variation ( $CV$ ) as a percentage (%) of the Variance Index resampled values over 10,000 iterations. Densities are drawn based on all values of  $CV$  but values greater than 100% are not printed.



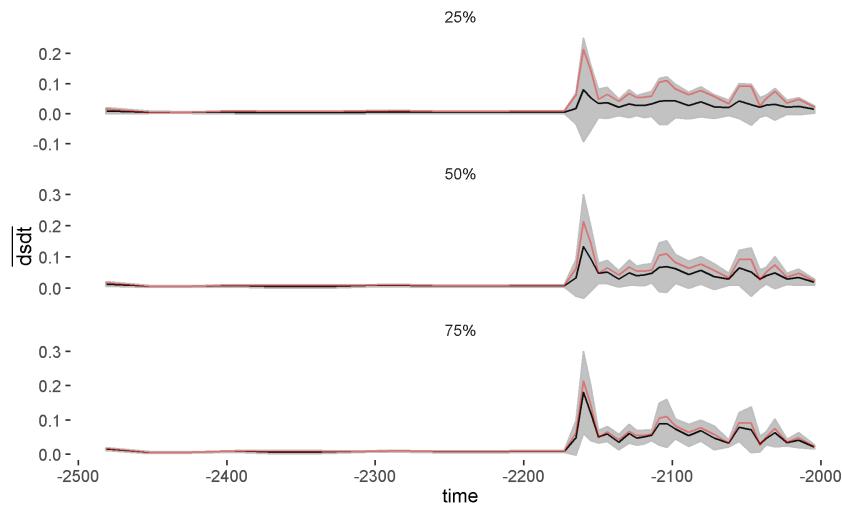
**Figure 6.4:** Mean Variance Index ( $VI$ ) and associated 95% confidence intervals over 10,000 iterations using the **observations** resampling method. Red line indicates the value of  $VI$  when  $\mathbf{M}$  and  $\mathbf{P}$  are 100%.



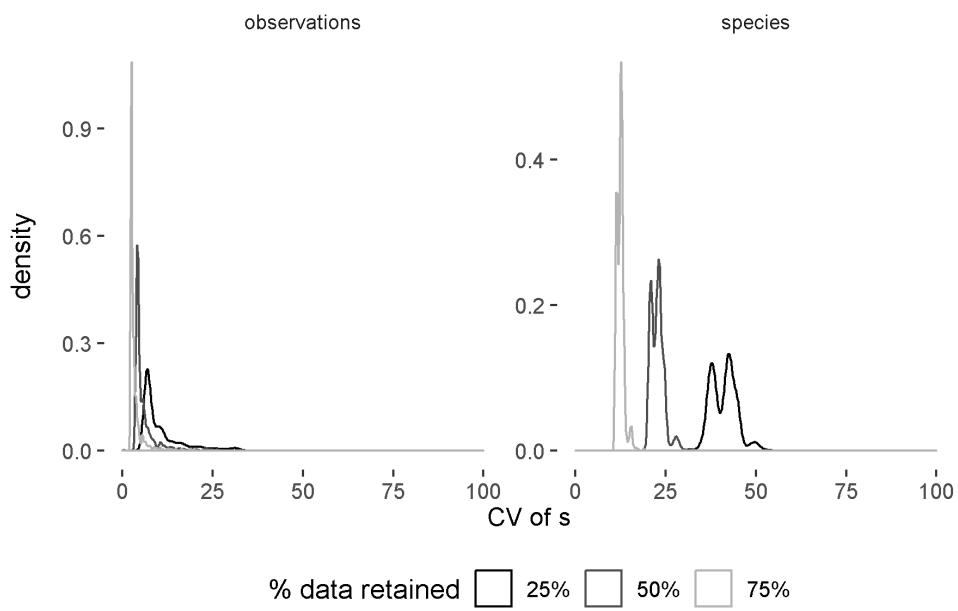
**Figure 6.5:** Mean Variance Index (VI) and associated 95% confidence intervals over 10,000 iterations using the *species* resampling method. Red line indicates the value of VI when  $M$  and  $P = 100\%$ .



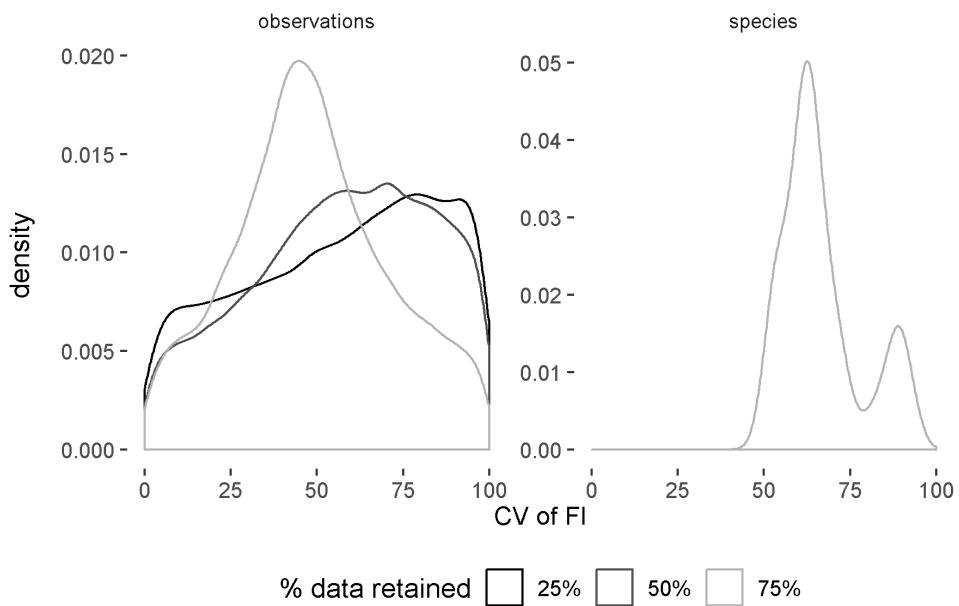
**Figure 6.6:** Mean velocity and associated 95% confidence intervals over 10,000 iterations using the observations resampling method. Red line indicates the value of velocity when  $M$  and  $P$  are 100%.



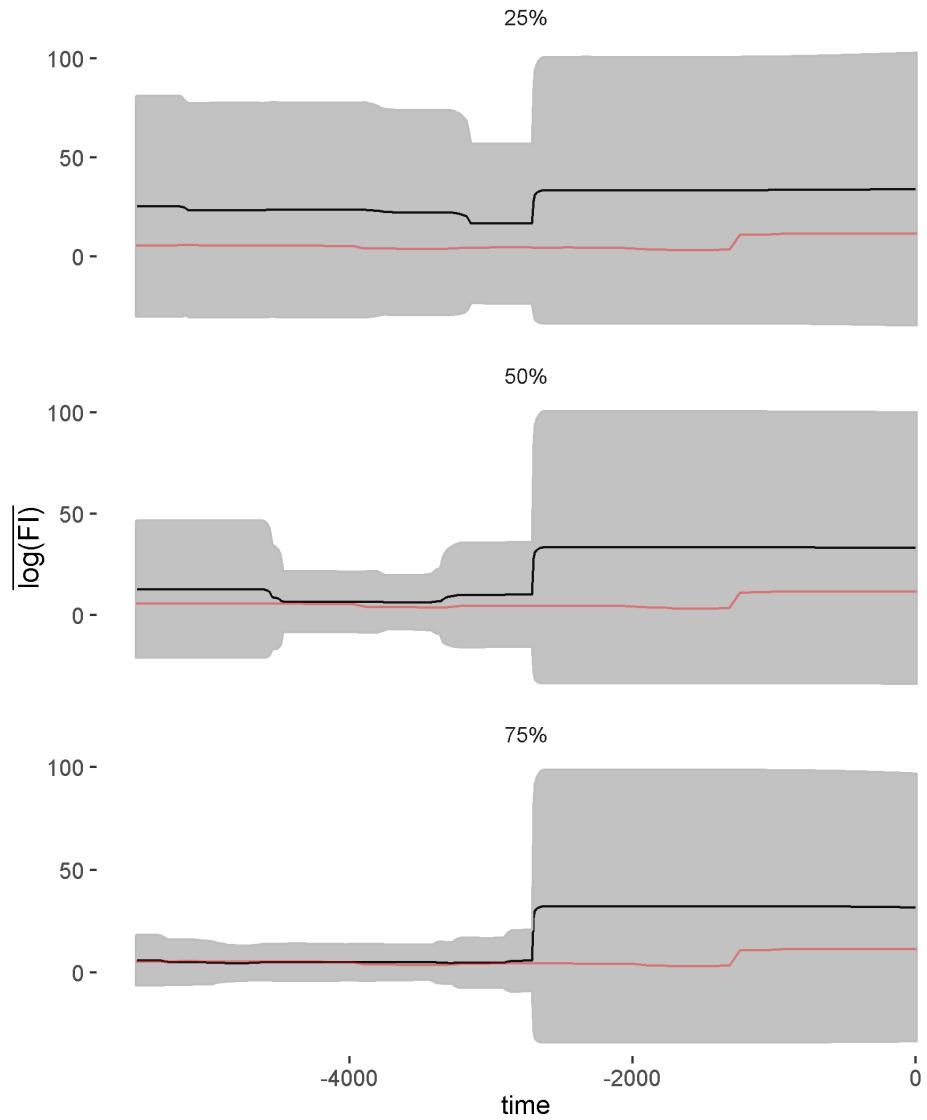
**Figure 6.7:** Mean velocity and associated 95% confidence intervals over 10,000 iterations using the observations resampling method for a subset of the time series (the second ‘regime’ identified). Red line indicates the value of velocity when  $M$  and  $P = 100\%$ .



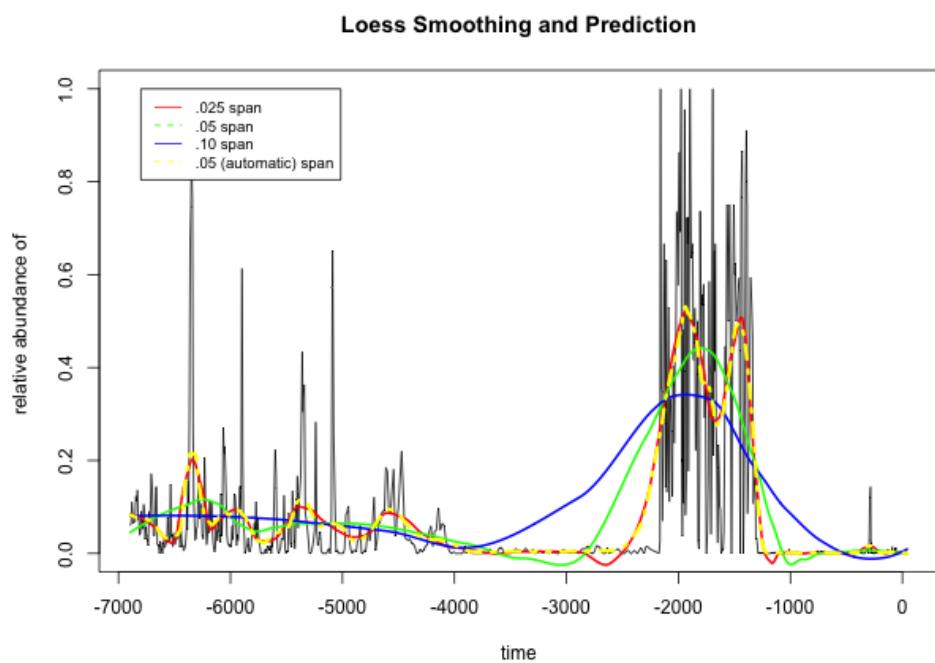
**Figure 6.8:** Density plot of the coefficient of variation (CV) as a percentage (%) of the Distance Metric ( $s$ ) resampled samples (10,000 iterations). Densities are drawn based on all values of CV, but values  $>100\%$  are not printed.



**Figure 6.9:** 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.



**Figure 6.10:** 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  $\mathbf{M}$  and  $\mathbf{P}$  are 100%. A very small value was added to the mean  $FI$  prior to log transformation.



**Figure 6.11:** 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 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*

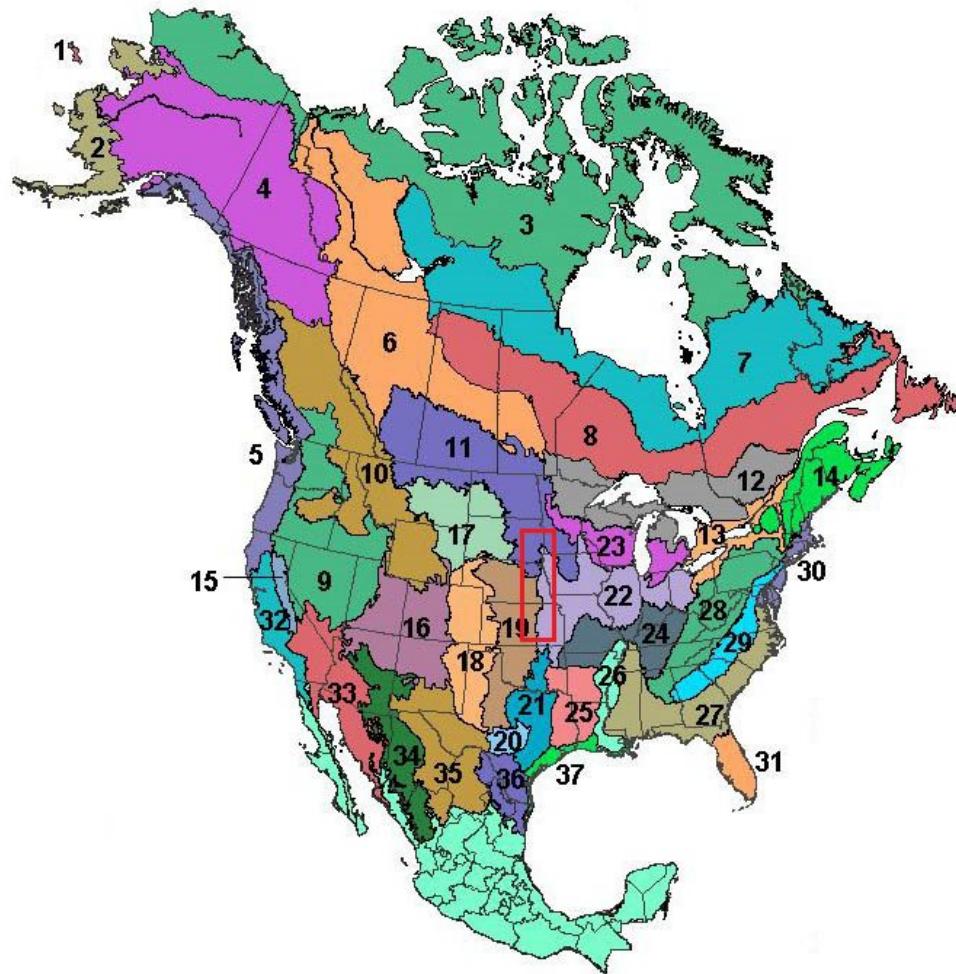
*al.*, 2006).

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 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



**Figure 7.1:** My study area (red box) overlaying the terrestrial Breeding Conservation Regions (BCR) in North America.

### 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 7.1) using inference

gained from discontinuity analyses of avian body mass distributions. The authors 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 7.2). 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 7.1). 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

#### 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 4.1) since 1966. Each roadside survey consists

of 50, 3-minute point counts (data collected using sight and sound) along  $\sim$  24.5 mile stretch of road. Although the point counts are designed to collect estimates of relative abundance, 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  $\pm 1$  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 structure, I restricted analysis to the grassland habitat of the Central Great Plains. All routes falling within a rectangular area bounded by coordinates  $37.8^\circ$  and  $44.5^\circ$  latitude, and  $-101^\circ$  and  $-95.5^\circ$  longitude [see Figure 7.2]. 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 identification 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 for download at CRC press).

**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.

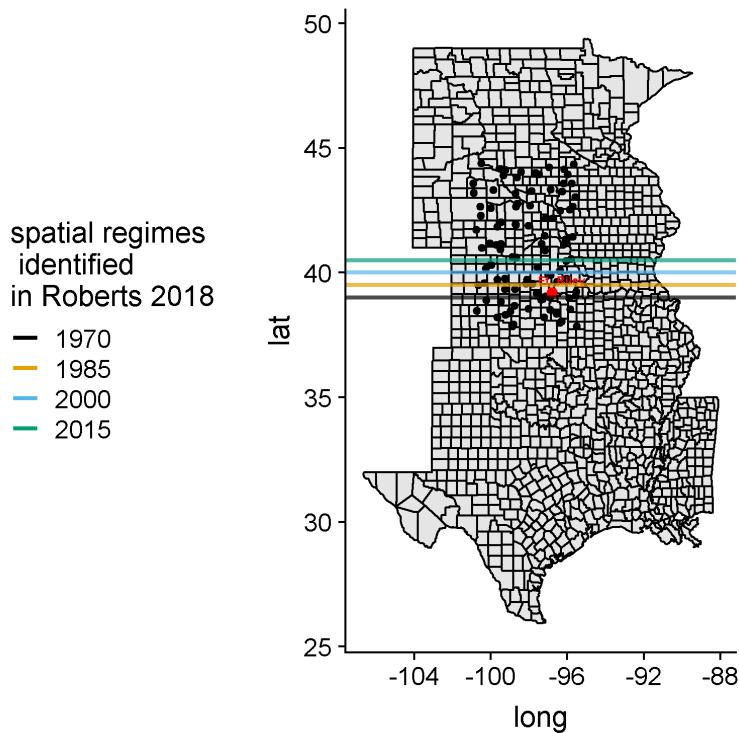
I made decisions regarding species-specific classification based on the North American breeding range maps provided by the Cornell Lab of Ornithology. Many unidentified species were easily categorized given the lack of overlap in species' ranges in our study area. For example, *Baeolophus bicolor* is nearly indistinguishable from *Baeolophus atricristatus*, however *B. atricristatus* is not known to occur in our study

area (Figure 7.1)—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*, *Coccyzus erythrophthalmus / 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 *Chlidonias 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.

### 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.



**Figure 7.2:** *North American Breeding Bird Survey routes (points), latitudinal locations (horizontal bars) of the spatial regimes identified by roberts2019shifting.*

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 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

#### 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.*, 2014a; 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 .1.10. 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 .1.10 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 mas 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.

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

Common Name	Species Group
Bobolink	Grassland Obligates
Lark Sparrow	Grassland Obligates
Chipping Sparrow	Grassland Obligates
Henslow’s Sparrow	Grassland Obligates
Ferruginous Hawk	Grassland Obligates

Upland Sandpiper	Declining Grassland Obligates
Ring-necked Pheasant	Declining Grassland Obligates
Horned Lark	Declining Grassland Obligates
Eastern Meadowlark	Declining Grassland Obligates
Western Meadowlark	Declining Grassland Obligates
Vesper Sparrow	Declining Grassland Obligates
Grasshopper Sparrow	Declining Grassland Obligates
Field Sparrow	Declining Grassland Obligates
Dickcissel	Declining Grassland Obligates
Savannah Sparrow	Declining Grassland Obligates
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
Brewer's Blackbird	Declining
Clay-colored Sparrow	Declining
Willet	Declining
Marbled Godwit	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 7.2], 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 7.12), 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 ( $\beta_{1i}$ ) and regime location ( $\beta_{2i}$ ), and an interaction among regime ( $\beta_{2i}$ ) 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 7.2:** *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.*

Year	North	South
1967	38	11
1968	43	12
1969	53	11
1970	45	17
1971	43	13
1972	38	17
1973	39	15
1974	29	17
1975	41	16
1976	46	18
1977	47	16
1978	45	18
1979	46	17
1980	43	16
1981	40	18
1982	38	16
1983	37	16
1984	30	14
1985	30	24
1986	31	21

1987	36	21
1988	34	24
1989	33	25
1990	40	21
1991	31	22
1992	33	23
1993	39	23
1994	37	23
1995	37	22
1996	39	22
1997	36	24
1998	34	21
1999	32	23
2000	37	37
2001	31	34
2002	32	36
2003	36	38
2004	35	35
2005	38	36
2006	39	35
2007	39	36
2008	35	34
2009	39	37
2010	37	30
2011	38	33
2012	33	35

	2013	39	37
	2014	42	38
	2015	36	33
	2016	37	30
	2017	35	32

### 7.3.1 Summary statistics of censuses (NABBS data)

Given the location of the study area (Figure 7.1) 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 7.2). 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 7.4). Annual turnover rates were relatively low but became more dispersed over time (Figure 7.5; Table 7.3).

**Table 7.3:** *Summary statistics for annual species richness and annual turnover in all NABBS routes in study area.*

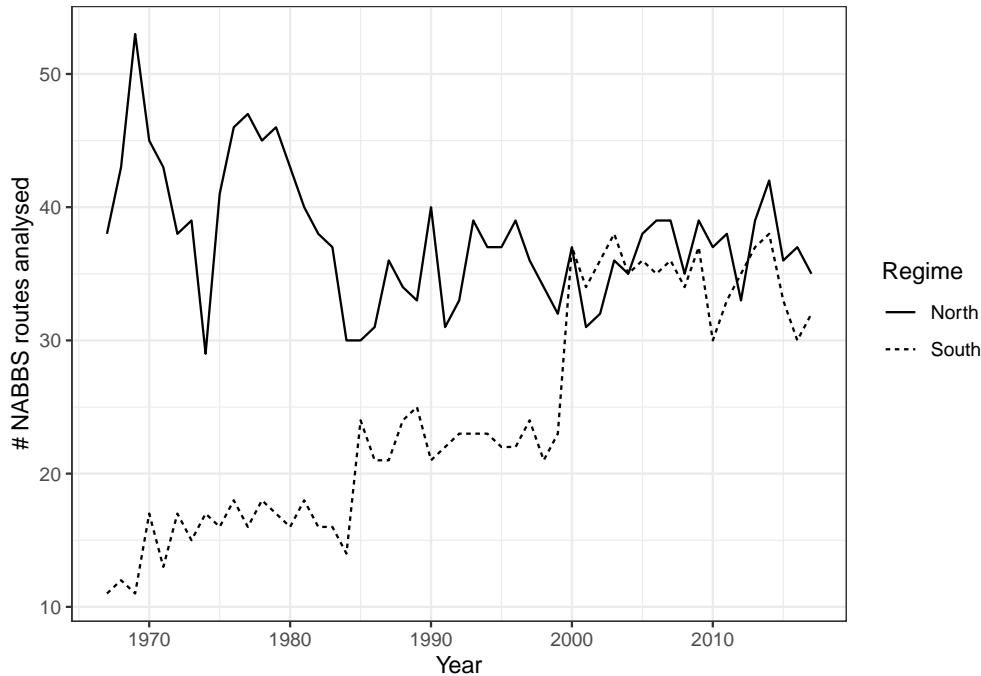
Year	Annual Richness			Annual Turnover		
	$\bar{x}$	$\sigma$	N	$\bar{x}$	$\sigma$	N
1967	39.4	8.14	27	-0.1	8.13	27
1968	41.4	8.11	37	1.6	10.25	37
1969	39.9	7.44	47	0.6	10.12	47
1970	41.3	7.73	43	0.9	9.74	43
1971	39.4	7.50	36	0.7	10.62	36
1972	41.0	9.80	32	-1.2	12.94	32
1973	40.1	8.33	33	-2.1	10.42	33

1974	38.8	7.67	27	-2.2	10.53	27
1975	38.3	7.37	32	-1.4	9.57	32
1976	41.2	7.53	45	0.6	10.31	45
1977	39.5	8.48	43	-1.7	9.69	43
1978	39.5	8.88	42	-1.9	10.09	42
1979	39.7	8.06	43	-1.7	10.03	43
1980	40.4	7.18	37	-1.1	10.16	37
1981	41.2	8.76	39	-1.2	9.26	39
1982	42.3	6.65	32	-1.6	8.28	32
1983	41.3	8.47	33	-1.9	8.91	33
1984	41.7	7.85	22	-4.1	6.12	22
1985	40.8	8.54	34	-1.1	11.30	34
1986	42.2	9.17	32	-2.7	10.42	32
1987	43.4	8.94	38	-1.3	9.82	38
1988	41.7	9.35	37	-1.4	12.19	37
1989	42.3	8.67	40	-1.8	9.98	40
1990	42.3	9.47	38	-1.6	9.69	38
1991	43.3	9.69	30	-2.1	10.14	30
1992	44.6	7.52	31	-0.8	9.61	31
1993	42.7	9.35	41	-1.0	12.87	41
1994	43.4	9.76	38	0.4	12.50	38
1995	42.2	9.98	39	-0.8	12.75	39
1996	44.1	7.44	36	1.0	10.68	36
1997	44.6	9.45	39	1.0	12.65	39
1998	43.8	9.00	32	-1.9	12.25	32
1999	45.9	9.52	28	-1.8	11.66	28

	2000	45.8	9.30	56	0.5	11.86	56
	2001	46.5	8.51	46	0.7	11.95	46
	2002	46.8	8.94	50	0.5	13.91	50
	2003	48.1	9.84	57	-0.2	13.11	57
	2004	45.8	10.88	49	0.1	13.97	49
	2005	46.3	9.13	52	-0.6	11.20	52
	2006	46.2	10.46	52	-0.5	11.94	52
	2007	46.7	9.59	57	0.3	11.29	57
	2008	48.2	11.00	47	1.0	14.84	47
	2009	47.8	9.92	56	-0.4	12.11	56
	2010	47.4	9.62	45	-0.9	11.93	45
	2011	47.0	10.53	47	-1.0	14.61	47
	2012	47.7	10.33	46	0.7	13.46	46
	2013	47.0	10.03	57	1.1	13.50	57
	2014	47.8	9.35	59	-0.2	11.22	59
	2015	48.5	9.33	49	0.4	10.79	49
	2016	50.8	9.38	44	2.0	12.73	44
	2017	47.6	9.84	45	0.2	11.84	45

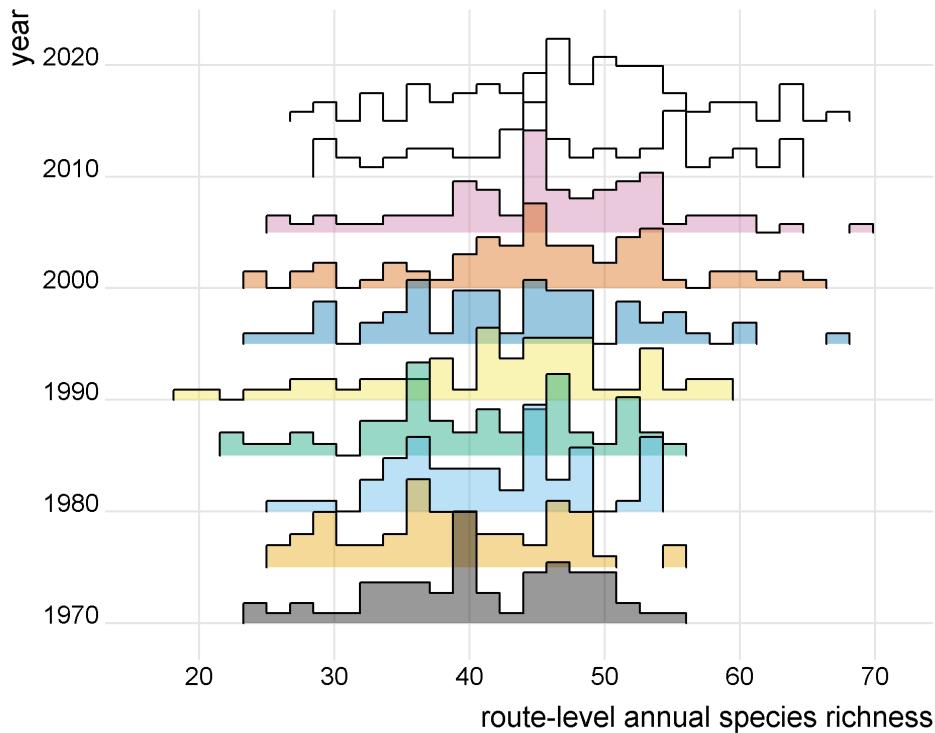
### Species of interest

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 7.1).



**Figure 7.3:** Number of NABBS routes analysed per year. Some routes are not sampled annually due to volunteer availability, environmental conditions, or route discontinuation.

The total (Figure 7.8) and mean (Figure 7.9) 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 7.9, 7.8)). The latter BCR comprises a much smaller portion of the study area (Figure 7.1) and accordingly the coefficient of variation (CV; ratio of deviation to the mean) around the stop totals (Figure 7.10) was highest in this region. It is worth noting the high CV (CV is considered low when  $< \sim 40$ ) in all regions.

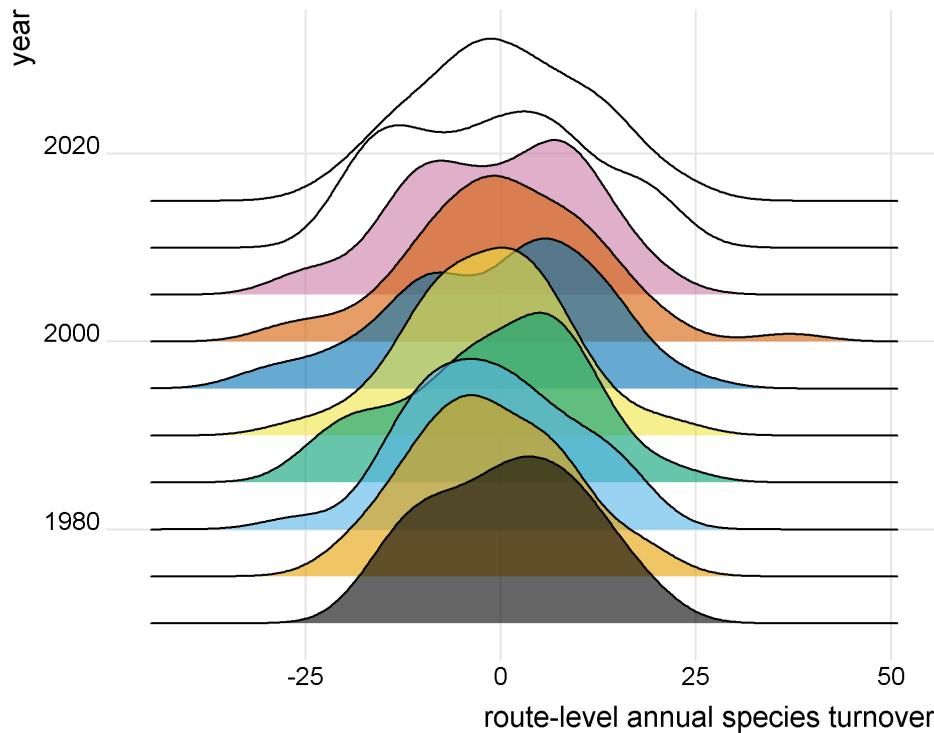


**Figure 7.4:** *Species richness increases over time across the entire study area.*

### 7.3.2 Statistical analysis

#### 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 7.1)). 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 (Barichievy *et al.*, 2018) was similar was similar across time (Figure 7.11a) and was approximately normally distributed across all survey-year combinations ( $\bar{x} = 4.7, \sigma = 1.6$ ; Figure 7.11a). Species richness at the route level was strongly positively correlated with the number of aggregations (Figure 7.12). The distance to edge variable was statistically, but not strongly, correlated with body mass  $r = -0.02, p = < 0.01$ , and



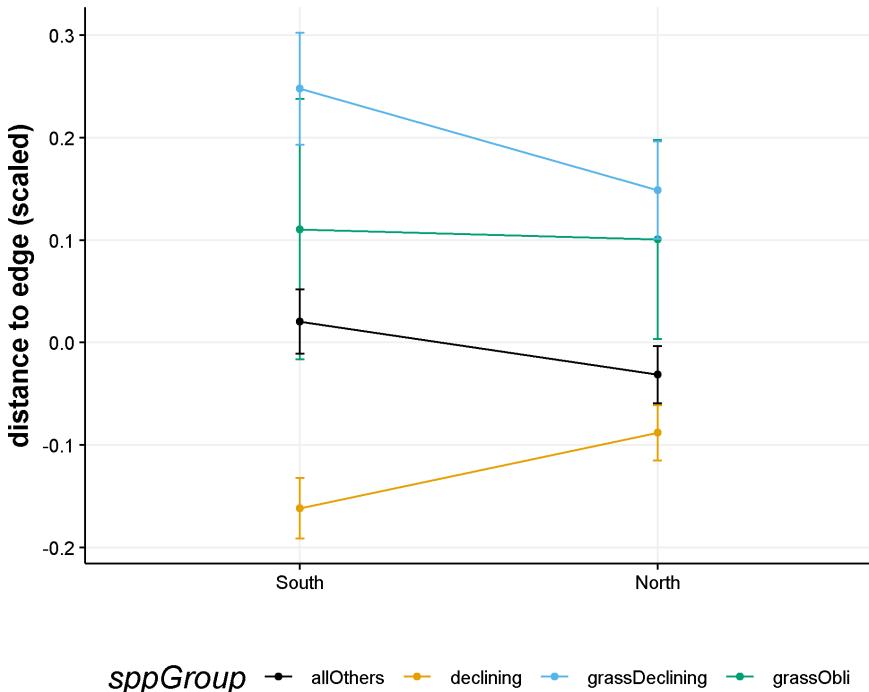
**Figure 7.5:** Variance in the total species turnover increases over time across the entire study area.

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 7.13). If the shifting spatial regimes proposed in an earlier study [Roberts *et al.* (2019); Figure 7.2], 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 7.13 is representative of the ~5 NABBS locations falling in this area of expected changes shifting).

### 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 7.4). Similar trends held for declining grassland obligates

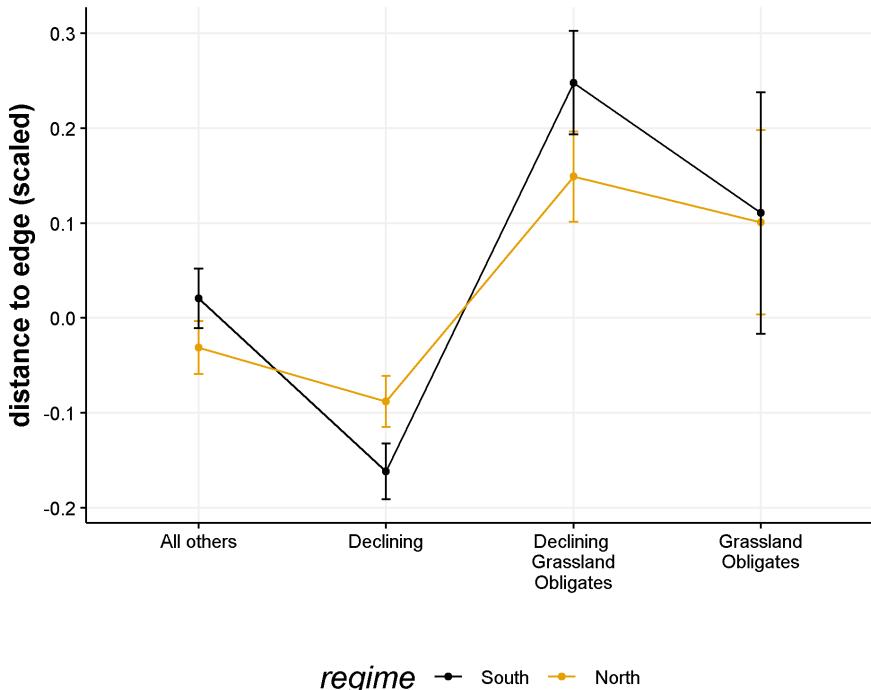


**Figure 7.6:** Declining species were closer to the edge of body mass aggregations in the 'northern' than were in the 'southern' regimes.

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 7.4).

## 7.4 Discussion

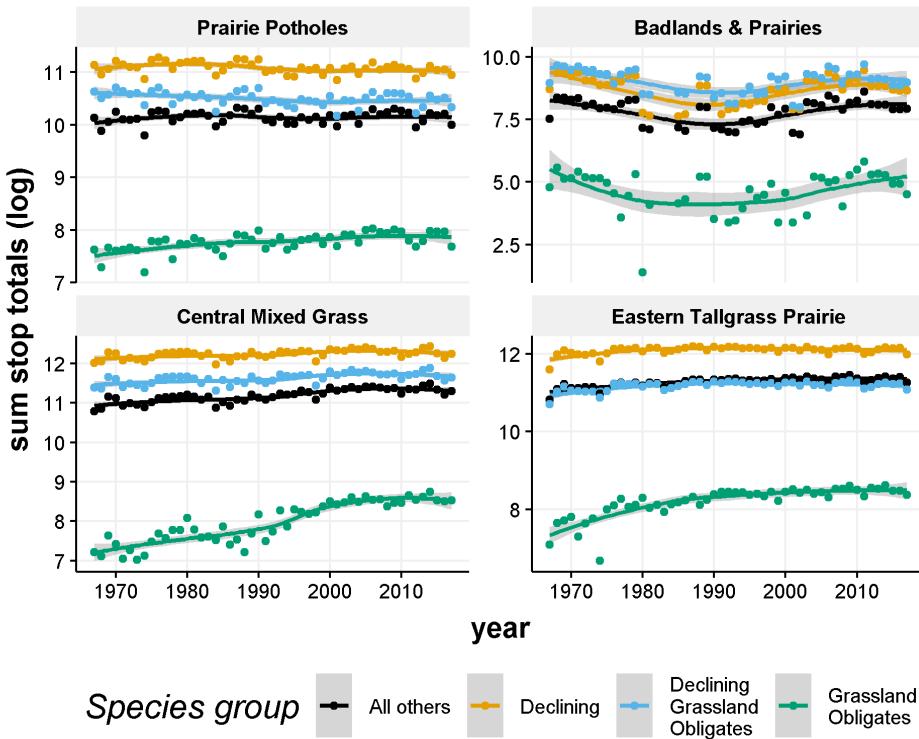
South-North shifts in the past 50+ years have been demonstrated in large scale processes, including bird populations and ranges (Sorte & III, 2007) and plant hardening zones (McKenney *et al.*, 2014). A recent application of discontinuities in body mass distributions (Holling, 1992) corroborated these 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



**Figure 7.7:** Declining species were closer to the edge of body mass aggregations in the 'northern' than were in the 'southern' regimes.

proposed spatial regime boundary (Roberts *et al.*, 2019) proposed in impact of these proposed spatial regime shifts on local bird communities

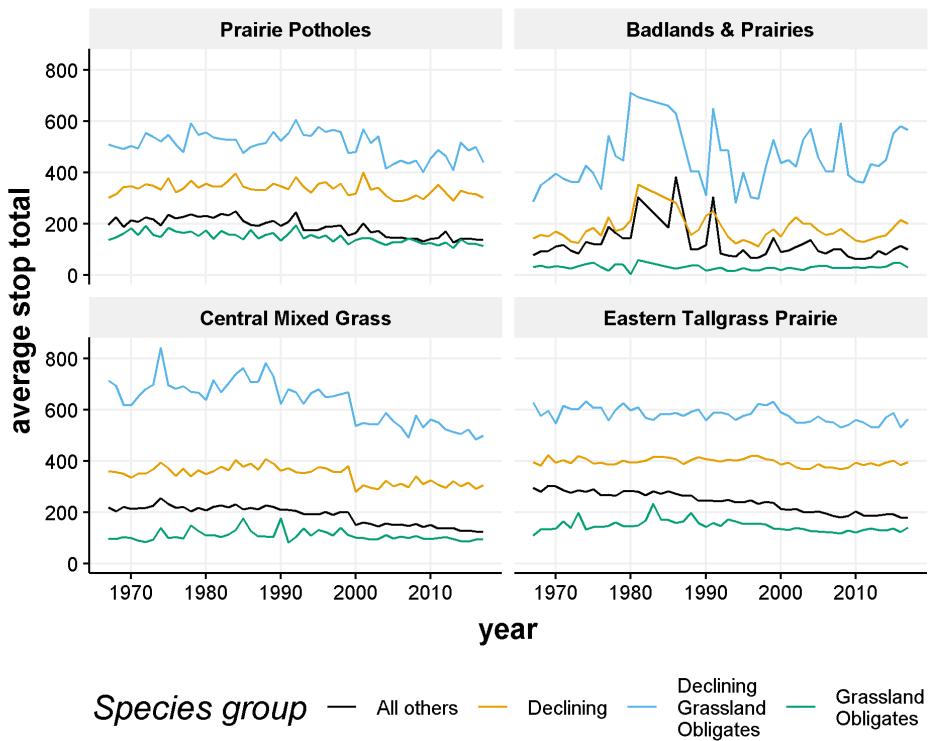
The lack of additive effects in the presence of multiplicative effects of the regime location (Table 7.4) strengthens the support for the differences in grassland obligates and declining grassland obligates with respect to all other species, however, the confidence intervals around the estimates of declining grassland obligates, grassland obligates may suggest that the evidence for such an effect is relatively weak (Figure 7.7). This is unsurprising given the distribution of grassland obligate body masses is highly skewed right relative to the remaining species (Figures 7.14), 7.15)). 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



**Figure 7.8:** Total number of birds across the entire study area per species group per year.

small portion of our study area, Shortgrass Prairie BCR (Figures 7.1, 7.16).

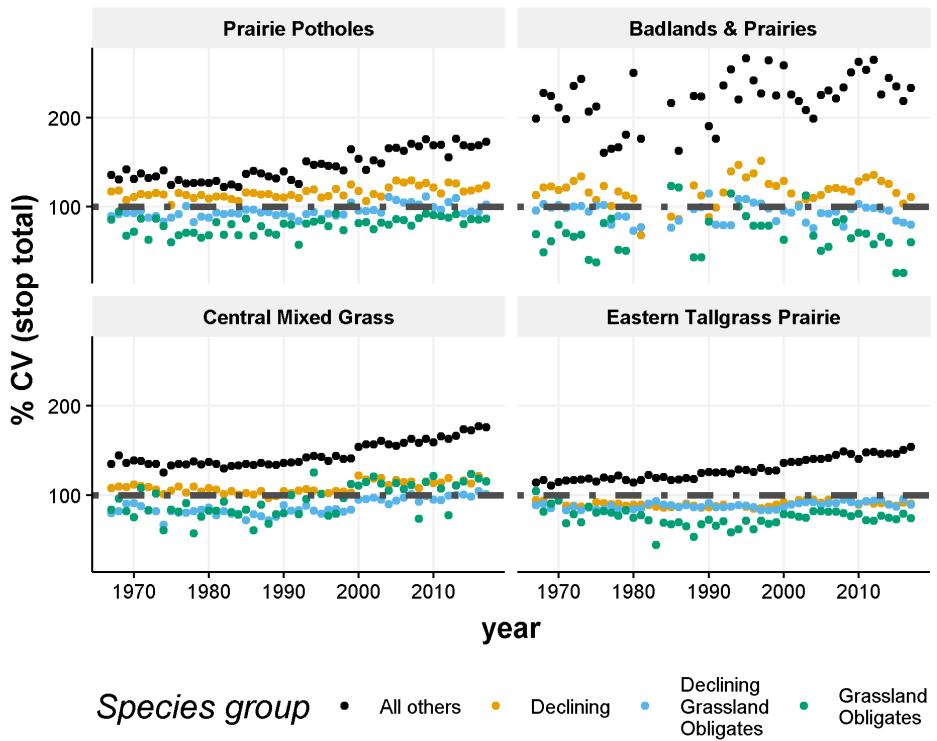
Although the body mass distributions of terrestrial communities at small scales tends to differ from those at larger spatial scales (Blackburn & Gaston, 1994), numerous studies confirm the evidence for ‘discontinuities’ in these distributions across multiple spatial scales for a variety of phenomenon (Allen *et al.*, 1999; Wardwell & Allen, 2009; Nash *et al.*, 2014a; 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.



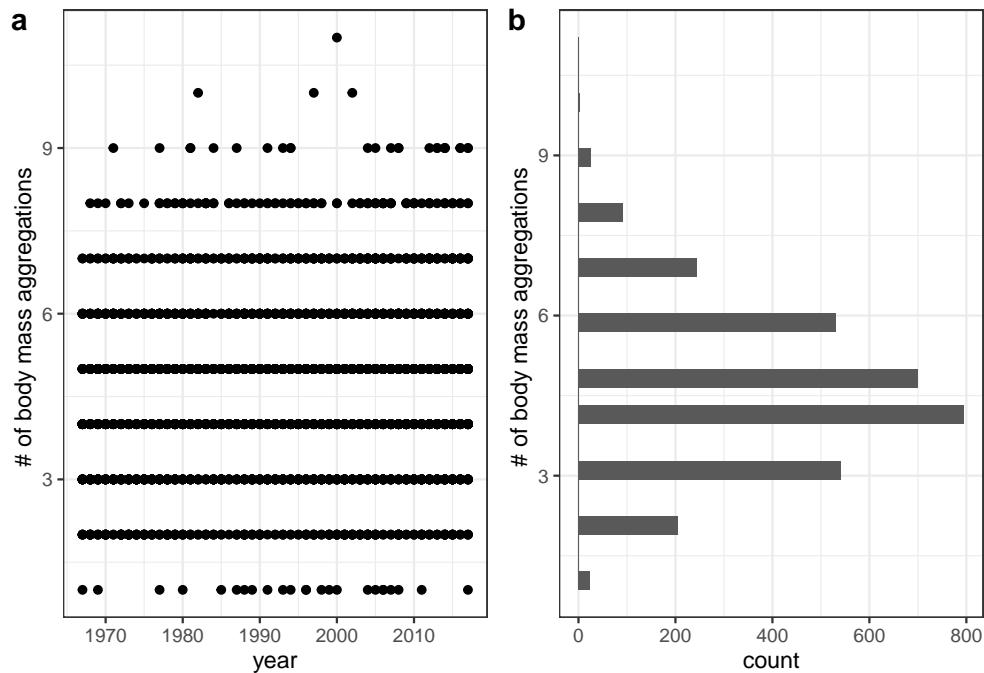
**Figure 7.9:** Average coefficient of variation in the number of birds across the entire study area per species group per year.

**Table 7.4:** Coefficient estimates for the linear mixed effects model predicting species' 'distance to edge' of a body mass distribution.

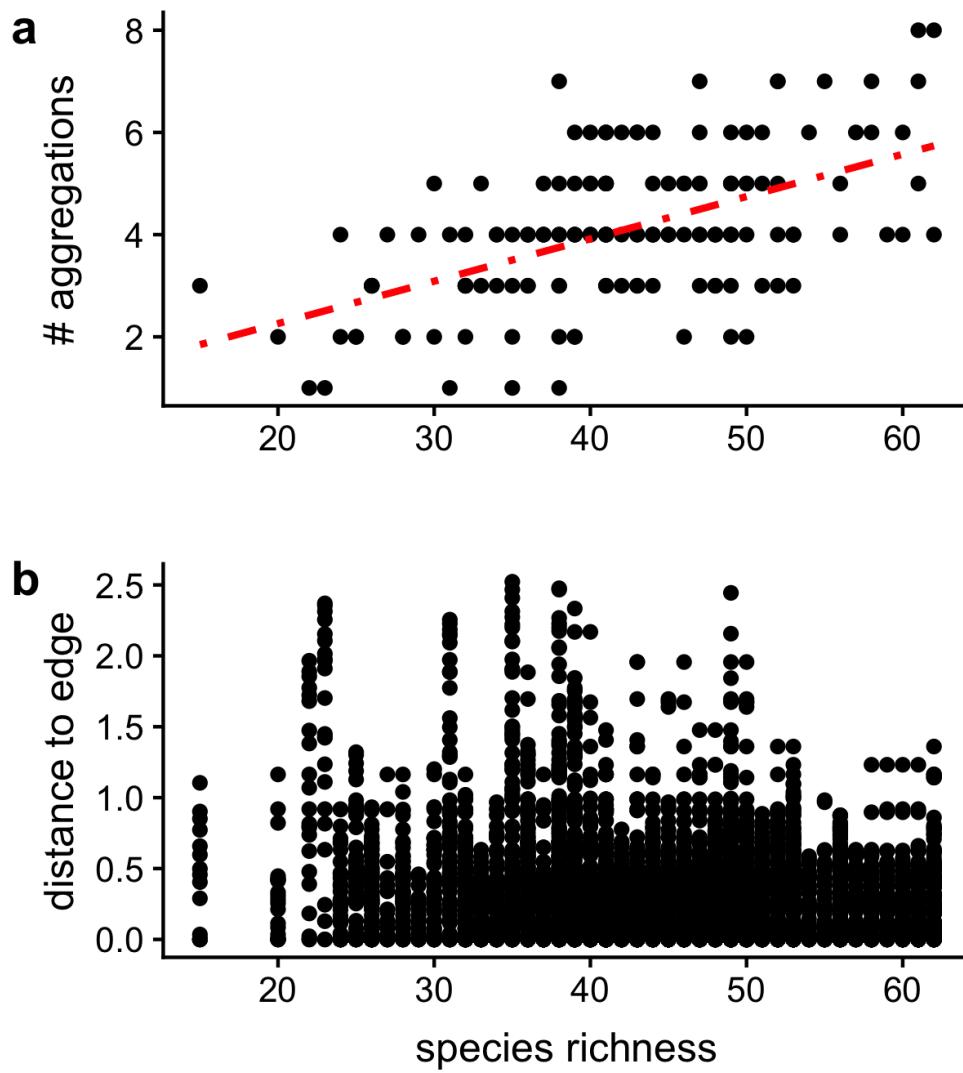
	Estimate	CI
(Intercept)	0.04	(0.02, 0.07)
Year	0.00	(0, 0)
North	0.01	(-0.01, 0.03)
Declining	-0.05	(-0.08, -0.02)
Declining Grassland Obligates	-0.17	(-0.2, -0.14)
Grassland Obligates	0.15	(0.11, 0.2)
Year x Declining	0.00	(0, 0)
North x Declining	0.02	(-0.01, 0.04)
North x Declining Grassland Obligates	-0.05	(-0.07, -0.03)
North x Grassland Obligates	0.04	(0.01, 0.07)



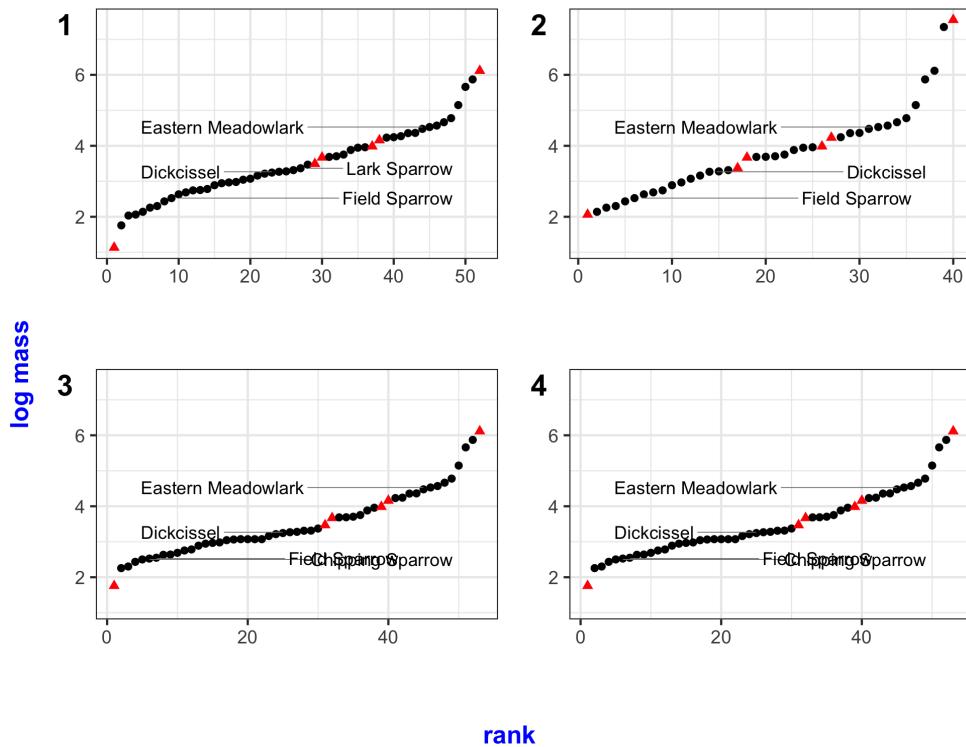
**Figure 7.10:** Average number of birds across the entire study area per species group per year.



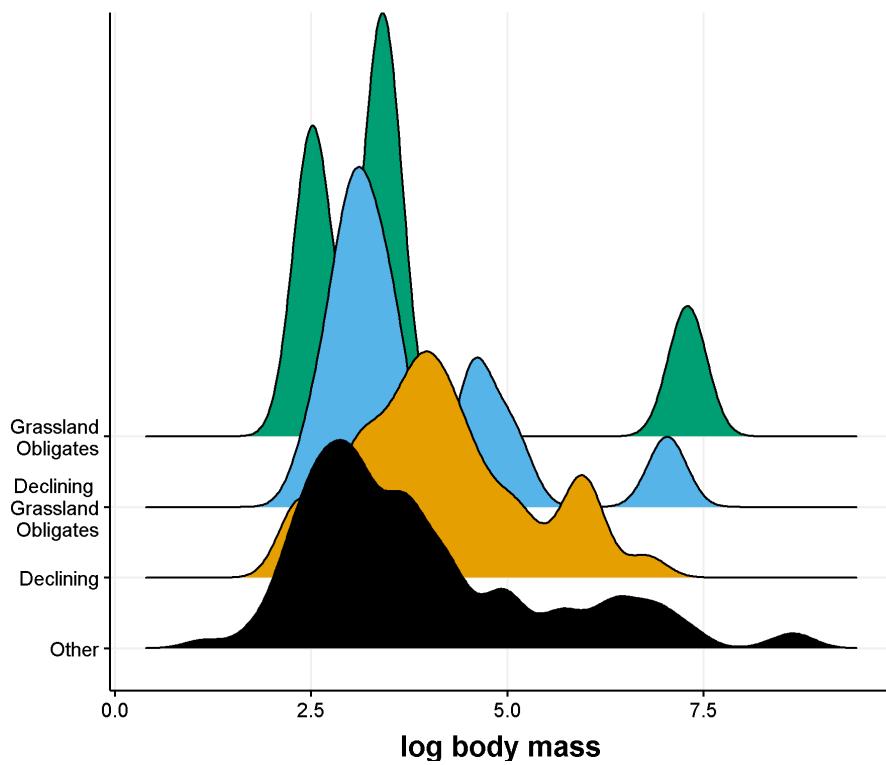
**Figure 7.11:** Number of body mass aggregations identified in each route unchanged across the time period (\*\*a\*\*) and is approximately normally distributed ( $\bar{x} = 4.7, \sigma = 1.6$ )



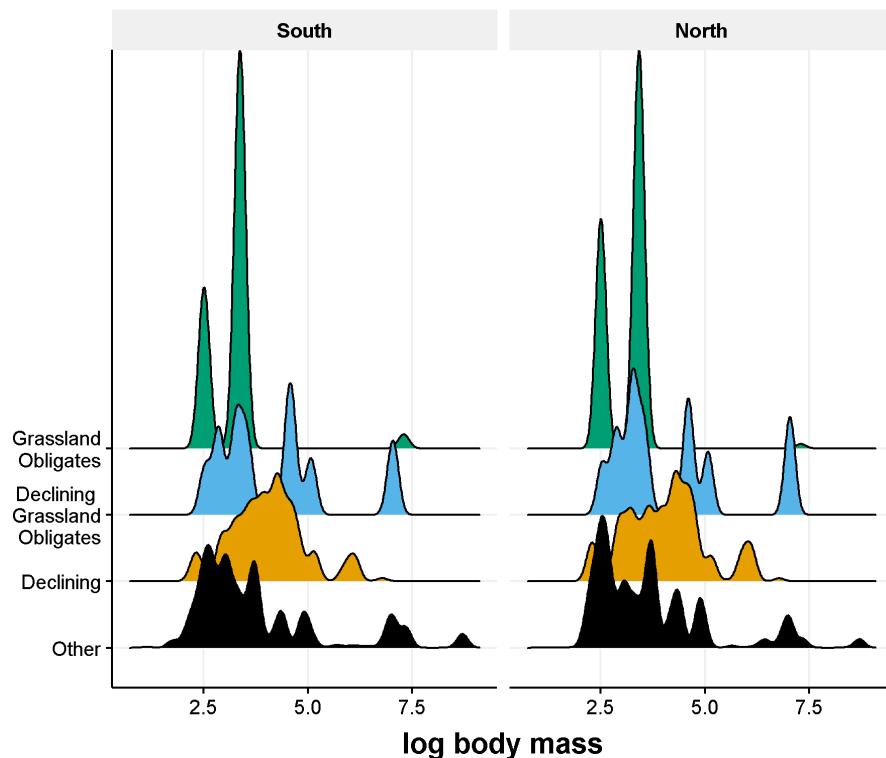
**Figure 7.12:** Relationship between species richness per route and (a) the number of aggregations identified in body mass distributions and (b) distance to the edge (units log body mass) of aggregations.



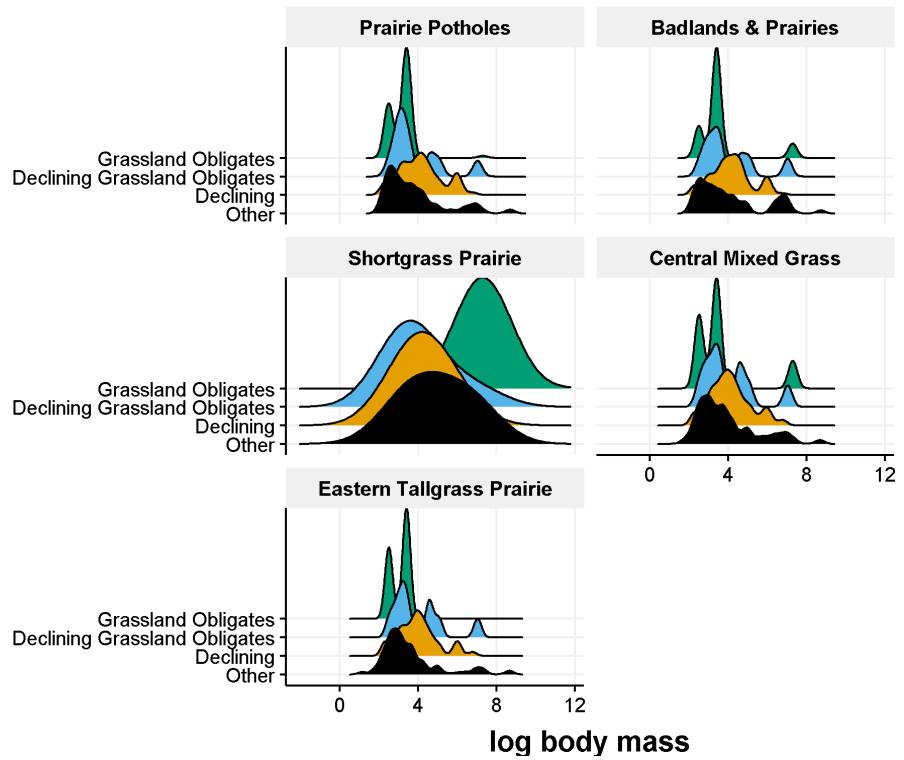
**Figure 7.13:** Aggregation locations of body mass distributions of the avian community at a single NABBS route (state 7 route 24) appear relatively similar across time. Panels include data for years 1=1970, 2=1985, 3=2000, 4=2015. Red triangles indicate grassland obligate species.



**Figure 7.14:** 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 7.15:** *The body mass distribution of declining species differ only slightly between the Southern and Northern regimes.*



**Figure 7.16:** *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.

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

## 8.1 Method Mining

Many regime detection measures have yet to be statistically (or numerically) scrutinized for robustness or sensitivity to data quality, data quantity, 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<sup>1</sup>. 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.

---

<sup>1</sup>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)

## 8.2 Ecological Data are Noisy

Regime detection metrics appear more reliable when the signal-to-noise ratio is high (Chapter 2, Chapter 5, Taranu *et al.*, 2018). Ecological systems are noisy, and the observational data we are collecting at large scales (e.g., the North American Breeding Bird survey), is noisy. Using methods incapable of identifying meaningful signals in noisy data appears futile, 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 5), in that the resulting measure is composite and carries no information regarding the influence of state variables on the metric result. The actual limitations to the system should be, theoretically, known as a result of bounding the system. Inference beyond this system is extrapolation, and should be

treated as speculation, especially when not accompanied by a measure of uncertainty around one's predictions.

## 8.4 Common Limitations of Regime Detection Measures

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

1. Signals in the indicators are restricted to the ecological processes captured by the input data. Extrapolation occurs when processes manifest at scales different than the data collected [resolution; Chapter 4]
2. Normalization and weighting techniques often impact results (whether ecological or numerical) (Appendices .1.3 and .4)
3. Data aggregation techniques often impact results (Chapter 6)
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

*Under 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, `get_credibility_trends()`; and a function for the current and archived population trends estimates for *all* species and regions, `get_analysis_results()`.

## .2 Package Vignette

### .2.1 Installing package and loading dependencies

```
knitr::opts_chunk$set(eval=FALSE)

# devtools::install_github("trashbirdecology/bbsAssistant",
#   dependencies = TRUE, force=FALSE)

library(bbsAssistant)

library(rvest)

library(gdata)

library(feather)

library(here)

library(dplyr)
```

```
library(tidyverse)
```

## .2.2 Downloading the BBS data from USGS FTP

This section reviews functionality for downloading and writing to file the BBS data from the FTP server.

### Define and/or create local directories

This function will create, if it does not already exist, folder `./bbsData/` within which to locally store BBS data and results. **NOTE:** If the directory exists, it will not overwrite files. If the bbs data already exists inside `bbsDir`, then we will create a logical to NOT download the data (see below). If you wish to download more, or overwrite existing data, please specify `downloadBBSData=TRUE` or remove .zip files from `/bbsData/`.

```
# Create a directory to store and/or load the BBS data as feathers
bbsDir <- here::here("bbsData/")
dir.create(bbsDir)
```

### Retrieve and import BBS data

If necessary, download all or some of the BBS state-level data. Note: Downloading all the data to file takes 10-15 minutes, so only run if you have not recently downloaded the BBS data. First, let's retrieve the regions of data that are available. The function `get_regions` retrieves the .zip filenames of all U.S. states and Canadian provinces, including their reference numbers and region codes.

```
# a. Load the regional .txt file from  
# Patuxent's FTP server (you must be connected to the  
# internet to perform this step)  
regions <- get_regions()
```

Let's restrict our data download to **Florida** data:

```
regionFileName.use <- "Florida.zip"  
  
if (!exists("regionFileName.use")) {  
  regionFileName <- regions$zipFileName %>%  
    na.omit()  
  (regionFileName.use <-  
    regionFileName[which(str_detect(regionFileName, "Flori") == TRUE)])  
}
```

Once we have one or more region filenames, we can use function `get_bbsData` to download the .zip file to a temporary folder (unless otherwise specified), and *import* the temp file to R object. The R object, `f1BBS`, contains the raw BBS count data.

```
f1BBS <- get_bbsData(file=regionFileName.use)
```

Next, we can download the BBS route-level geographic information and metadata, and append this to the original data.

```
# retrieve route-level data  
routes <- get_routeInfo()  
  
# merge route-level data to bird count data  
f1BBS <- left_join(f1BBS, routes)  
glimpse(f1BBS %>% dplyr::select(aou, year, route, statenum, countrynum, stoptotal,
```

If we wish to save these data to file, we can do so by saving as `**.feather`s, a compressed file formatted for use in R.

```
export_bbsFeathers(dataIn = flBBS,
                     newDir  = bbsDir,
                     filename = regionFileName.use)
```

### .2.3 Import BBS data from file into R

If the BBS data was downloaded previously and saved as `.feather`, we can import it using `import_bbsFeathers`. The code below is particularly useful if you are importing multiple files (e.g., multiple states)

```
(featherNames <- list.files(bbsDir, pattern = ".feather"))

featherNames <- str_c("/", featherNames) #add separator

feather <- import_bbsFeathers(newDir  = bbsDir,
                               filename = featherNames)

glimpse(feather)

# Notice that the data imported from disk (feathers)
# differs from the original BBS data in that the
# of columns is fewer (9 and 12 columns, respectively).
```

### Option for downloading ALL BBS data

If you wish to download and/or import ALL the data, you might choose to do so in a loop. Note: this is expensive! The following are not run in this example.

```

# Throw a warning if files exist

  if(length(list.files(bbsDir, pattern = "*.feather")) > 0 ){

    downloadBBSData = FALSE

  }else(

    {dir.create(bbsDir)

    downloadBBSData = TRUE}

  )

## Download ALL the regions of BBS data

if(downloadBBSData==TRUE){

  for(i in 1:length(regionFileName)){

    bbsData <- import_bbsData(

      # arguments for get_bbsData()

      file = regionFileName[i] ,

      # arguments for get_routeInfo():

      routesFile = "routes.zip" ,
      routesDir = "ftp://ftpext.usgs.gov/pub/er/md/laurel/BBS/DataFiles/" ,
      RouteTypeID = 1

    )

# d. Save the unzipped files to disk.

export_bbsFeathers(dataIn = bbsData,

  newDir = bbsDir,

  filename = regionFileName[i])

# e. Clear object from memory

rm(bbsData)

} # end section I. loop

} # end if-else to download the data

} # end if-else to download the data

```

Use the same code as above to import *multiple* feathers from file:

```
feathers <- NULL

featherNames <- list.files(bbsDir, pattern = ".feather")

featherNames <- str_c("/", featherNames) #add separator

for (i in 1:length(featherNames)) {

  feather <- NULL

  feather <- import_bbsFeathers(newDir = bbsDir,
                                 filename = featherNames[i])

  feathers <- rbind(feathers, feather)

  rm(feather)

}

}
```

## .2.4 Subsetting the BBS count data

### Subset BBS data by taxonomic groups

First, retrieve the species list from the BBS FTP server.

```
spp <- get_speciesList()

glimpse(spp)
```

Subset by species AOU # (e.g. House Sparrow aou = 06882)

```
subset_speciesList(myData = flBBS, aou.ind = 06882) %>% glimpse()
```

We could merge the bbs count data with the species list to avoid having to refer to AOU, then just subset using species name (e.g. 'House Sparrow).

```
f1BBS <- left_join(f1BBS, spp)
hospBBS <- f1BBS %>% filter(commonName=="House Sparrow")
```

We can also use the `subset_SpeciesList` as a convenient way to `remove` taxonomic groups from the BBS data.

```
f1BBS.subset <- subset_speciesList(f1BBS, fam.ind = "Passeridae")
f1BBS.subset <- subset_speciesList(f1BBS, fam.ind = c("Passeridae", "Parulidae"))
```

## .2.5 Retrieve BBS analysis results and data credibility measures

There are a few options for obtaining species trends estimates and credibility measures: 1) download the entire region-species csvs for various analyses or 2) provide a URL to species- or region-specific estimates for the 1966-2015 trend estimates.

### Option 1: Download CSV for all species-region combinations

The function `get_analysis_results` allows you to specify an analysis type, and upload all species-regions combination estimates or annual indices to object. Let's look at Florida House Sparrow trend estimates for Florida:

```
results <- get_analysis_results(analysis = "trend.est") # default here is to obt
results.flHOSP <- results %>% filter(Species.Name=="House Sparrow", Region.Code=="F")
```

Get annual trend estimates for Florida 1966-2016 analysis:

```
results <- get_analysis_results(analysis = "annual.inds.2016") # default here is
```

```

library(ggplot2)

ggplot(data = results %>% filter(AOU.Number=='s06882',
                                    Region.Code=="S05"),
       aes(Year, Annual.Index)) +
  geom_point() +
  labs(y="HOSP annual population \ntrend index in region S05") +
  theme_bw()

```

## Options 2: Retrieve region-specific estimates using web-scraping

Another useful feature of this package is the ability to retrieve data credibility and species trend estimates from the BBS results using the function `get_credibility_trends`. This function allows the user to input a url to the region- or species-specific results page (see instructions below), as opposed to using function `get_analysis_results`. As an example, we retrieve the credibility scores and species trend estimates for **House Sparrows in Florida**.

```

cred <- get_credibility_trends() # default here is Florida House Sparrows.

# credibility colors correspond with the color scheme used on the BBS results page
cred %>% distinct(credibilityNumber, credibilityColor, credibilityClass)

```

Trend estimates are also listed in `cred` for Florida House Sparrow data:

```

cred %>%
  filter(Species == "House Sparrow") %>%
  glimpse()

```

Steps for obtaining argument “url” in function `get_credibility_trends`:

First, visit the USGS Patuxent Wildlife Research Center’s website for BBS results Online <https://www.mbr-pwrc.usgs.gov/>. Next, enable the drop-down **Survey Results**, and choose **Trend Estimates (by region)** (left) and choose the desired region



Patuxent Wildlife Research Center ([link](#))

## North American Breeding Bird Survey



Patuxent Wildlife Research Center ([PWRC](#)) PWRC Breeding Bird Survey



Dedicated to [Chandler S. Robbins](#), originator of the survey.

### Analytical Tools

--- select ---

### Survey Results

✓ --- select ---

Species Group Summaries

Trend Estimates (by species)

Trend Estimates (by region)

Results from Prior Years

See the [BBS Summary and Analysis](#)

Please select the region of interest:

✓ ===== select region ===

- Alabama
- Arizona
- Arkansas
- British Columbia
- California
- Colorado
- Connecticut
- Delaware
- Florida**
- Georgia
- Idaho
- Illinois
- Indiana
- Iowa
- Kansas
- Kentucky
- Louisiana
- Maine
- Manitoba
- Maryland
- Massachusetts
- Michigan
- Minnesota
- Mississippi
- Missouri
- Montana

Finally, copy the URL address for the page including the results and credibility mea-



## Florida

### North American Breeding Bird Survey Trend Resu

#### [Help file](#)

Check out our [Regional Credibility](#) measure

Species	N	Trend
<u><a href="#">Pied-billed Grebe</a></u>	35	-6.13
<u><a href="#">Herring Gull</a></u>	8	-2.74
<u><a href="#">Ring-billed Gull</a></u>	18	2.17
<u><a href="#">Laughing Gull</a></u>	48	-1.96
<u><a href="#">Gull-billed Tern</a></u>	6	6.52
<u><a href="#">Caspian Tern</a></u>	17	4.28
<u><a href="#">Royal Tern</a></u>	20	-4.27
<u><a href="#">Forster's Tern</a></u>	12	6.95

sures (e.g. Florida):

### **.3 Package Manual**

For functions and descriptions please see the manual.

### **.4 Acknowledgements**

We 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 rBBS package and are mentioned in function source code as applicable.

# Appendix B

## regimeDetectionMeasures: an R package for calculating various regime detection measures

This appendix contains example analysis associated with the R Package, `regimeDetectionMeasures`. Development source code for this package is available on GitHub as a compressed file at <https://github.com/TrashBirdEcology/regimeDetectionMeasures/archive/master.zip> or at <https://github.com/TrashBirdEcology/regimeDetectionMeasures>.

### .1 Example Analysis

#### .1.1 Measures/metrics calculated

This package will calculate a various regime detection metrics that have been used to ‘detect ecological regime shifts’. A ‘new’ metric, **distance travelled** is also calculated (Burnett and others, *in prep*).

**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`)

## .1.2 Calculate Metrics analysis

```
knitr::opts_chunk$set(echo=TRUE, eval=FALSE, message=FALSE, warning=FALSE, out.width = "100%")

library(regimeDetectionMeasures)

# Munge original data -----
origData = munge_orig_dat(example = T)
```

## .1.3 Plots

Plot the original time series as relative abundances.

```
# Plot the original time series abundances

plot_orig_data(data = origData,
               example = F,
               print = T)
```

Plot the time elapsed among observations.

```
# Plot the distance between observations  
  
plot_timeDiff(data = origData,  
               example = F,  
               print = T)
```

Plot species richness over time.

```
# Plot species richness over time  
  
plot_richness(data = origData,  
               example = T,  
               print = T)
```

Plot the distance travelled ( $s$ ) and velocity of distance travelled ( $v$ )

```
# Calculate distance travelled and derivatives of this -----  
  
distances <- calculate_distanceTravelled(origData, derivs = T)
```

Plot the moving-window metrics (FI, VI, and the univariate EWSs)

# Appendix C bbsRDM: an R package for applying regimeDetectionMeasures functionality to spatial data

This appendix contains a vignette associated with the R Package, `bbsRDM`. Development source code for this package is available on GitHub as a compressed file, <https://github.com/TrashBirdEcology/bbsRDM/archive/master.zip> or at <https://github.com/TrashBirdEcology/rRDM>.

This vignette runs through the capabilities of the `bbsRDM` package, which relies on the package `trashbirdecology::regimeDetectionMeasures`. Although this package can be used to calculate and visualize BBS data using time series, the example at hand runs presents an application to spatial transects.

## .1 Example Analysis

### .1.1 Load packages & create local directories

There are a lot of dependencies to load.

```

# Create directories to locally store BBS data and results.

# a. Create a directory to store and/or load the BBS data as feathers

bbsDir <- "./chapterFiles/appendix_bbsRDM/bbs_raw_data"

# If the bbs data already exists inside bbsDir, then we will create a logical copy of it

if(length(list.files(bbsDir, pattern = "*.feather")) > 0 ){

  downloadBBSData = FALSE

}else{

  dir.create(bbsDir)

  downloadBBSData = TRUE}

}

# If this returns a warning, proceed with caution as directory already exists, as it will be overwritten

# b. Create a directory to store and/or load the BBS data as feathers

resultsDir <- "./chapterFiles/appendix_bbsRDM/myResults"

dir.create(resultsDir)

# ditto

# c. Create directory for storing early warning signal results

resultsDirEWS <- "./chapterFiles/appendix_bbsRDM/myResults/ews"

dir.create(resultsDirEWS)

# d. Create directory for storing distance travelled results

resultsDirDIST <- "./chapterFiles/appendix_bbsRDM/myResults/distances"

dir.create(resultsDirDIST)

```

## .1.2 Download the BBS data and save to file locally

If necessary, download all the state data. This takes 10-15 minutes, so only run if you have not recently downloaded the BBS data or are missing data.

```
# a. Load the regional .txt file from Patuxent's FTP server (you must be connected to
regions <- GetRegions()

# b. Create a series of one filenames for states, regions
regionFileName <- regions$zipFileName %>% na.omit()

# c. Download and unzip the BBS data.

if(downloadBBSData==TRUE){

for(i in 1:length(regionFileName)){

    bbsData <- importDataBBS(
        # arguments for getDataBBS()
        file = regionFileName[i],
        dir = "ftp://ftpext.usgs.gov/pub/er/md/laurel/BBS/DataFiles/States/",
        year = NULL,
        aou = NULL,
        countrynum = NULL,
        states = NULL,
        # arguments for getRouteInfo():
        routesFile = "routes.zip",
        routesDir = "ftp://ftpext.usgs.gov/pub/er/md/laurel/BBS/DataFiles/",
        RouteTypeID = 1,
        # one or more of c(1,2,3)
        Stratum = NULL,
```

```

BCR = NULL
)

# d. Save the unzipped files to disk.

birdsToFeathers(dataIn = bbsData,
                 newDir = bbsDir,
                 filename = regionFileName[i])

# e. Clear object from memory
rm(bbsData)
} # end section I. loop
} else{print(paste0("NOT DOWNLOADING BBS DATA. If you wish to download the BBS data

```

### .1.3 Create a sampling grid

Next, build a sampling grid to force route information onto a regular gridded area.

This allows us to compare the results across space-time.

```

# III: Build sampling grid -----
# Define the grid's cell size (lat, long; unit:degrees)
    ## 1 deg latitude ~= 69 miles
    ## 1 deg longitude ~= 55 miles
cs <-
  c(0.5, 0.5) # default is cell size 0.5 deg lat x 0.5 deg long

# Create the grid
routes_gridList <- createSamplingGrid(cs = cs)

```

```
# Define the components of the sampling grid as individual objects

routes_grid <- routes_gridList$routes_grid

sp_grd <- routes_gridList$sp_grd

rm(cs)
```

See if the results are already saved to file. This will save us some computational time when knitting the document. Otherwise, will take about a minute to calculate, depending on the size of birdsData

```
tempResultsDir <- here::here("/tempFiles_bbsRDM/")

dir.create(tempResultsDir)

fns <- list.files(tempResultsDir)

import.ind = ifelse(length(fns) != 0 %in% fns, TRUE, FALSE)

import.feathers <- ifelse("feathers.RDS" %in% list.files(tempResultsDir), FALSE, TRUE)
```

Now we load in the BBS data from the feathers we created and align with the sampling grid. This requires a bit of memory, proceed with caution. If it's already saved in the tempResultsDir, then it will merely upload the data from file.

```
if(import.feathers){

  feathers <- NULL

  featherNames <- list.files(bbsDir, pattern = ".feather")

  featherNames <- str_c("/", featherNames) #add separator

  for (i in 1:length(featherNames)) {

    feather <- NULL

    feather <- loadBirdFeathers(newDir = bbsDir,
                                filename = featherNames[i])

    feather <- feather %>%
```

```

dplyr::rename(lat = latitude,
             long = longitude) %>%
left_join(routes_grid, by = c("countrynum", "statenum", "route", "lat", "long"))

feathers <- rbind(feathers, feather)
rm(feather)
if(i == length(featherNames)) saveRDS(feathers, paste0(tempResultsDir, "feathers.RDS"))
} } else (feathers <- readRDS(paste0(tempResultsDir, "feathers.RDS")))

```

## .1.4 Subset the BBS data by species and/or functional traits (OPTIONAL but highly recommended)

Although subsetting the speices is optional, I recommend removing waterfowl, wading birds, and shorebirds from analyses, especially as the spatial extent of the analysis increases.

## .1.5 Subset species according to AOU species codes (i.e. by family, genera, etc..)

For this example we will remove shorebirds, wading birds, and waterfowl (i.e., AOU species' codes 0000:2880). \*See R/subsetByAOU.R source code or documentation for options (see: `subset.by`)

```

# Subset the species
feathers <- subsetByAOU(myData = feathers, subset.by= 'remove.shoreWaderFowl')

```

## .1.6 Subset species by trait, body mass, taxonomically, etc... (optional)

```
# Load the functional trait and mass data, and munge/merge
require(dplyr)

funMass <-

funcMass(dataWD = paste0(getwd(), "/data"),
          fxn = T, # get functional trait data?
          mass = F) # get body mass data?

# Combine the functional traits and/or body mass
bbsData <-

mergeFunMassBBS(bbsData = feather, funMass = funMass)

rm(funMass)
```

## .1.7 Calculate regime detection metrics across space or time

First, define the parameters required to calculate the metrics.

```
# Which metrics do you want to calculate?
metrics.to.calc <- c("distances", "ews")

# If calculating "EWSs, you can calculate select metrics.
## Default = all early-warning signals, FI, and VI
to.calc = c("EWS", "FI", "VI")

# Choose spatial or temporal analysis
```

```

direction <-

  "South-North" # choose one of : 'South-North', 'East-West', or 'temporal'

# Choose the fill value for species present in the entire time series (or sampling)
fill = 0

# Minimum number of sites (if spatial) or years (if temporal) required to be in
min.samp.sites = 8

# Minimum number of sites (if spatial) or years (if temporal) required to be within
min.window.dat = 3

# Which Equation of Fisher Information to use (default = 7.12)
fi.equation = "7.12"

# By what % of the entire data should the window move?
winMove = 0.25

# Define some filtering and labeling parameters based on direction of spatial and temporal
# analysis

if (direction == "South-North") {

  dir.use = unique(feathers$colID) %>% na.omit(colID) %>% sort()

if (direction == "East-West") {

  dir.use = unique(feathers$rowID) %>% na.omit(rowID) %>% sort()
}

```

Define the years we want to analyze. For this (spatial) example, we will analyze only every fifth year in the available data.

```

# Get all possible years

years.use = unique(feathers$year)

# Keep only the years which are divisible by T

T = 10

years.use <- years.use[which(years.use %% T == 0 & years.use > 1975)] %>% sort()

```

## .1.8 Conduct analysis

This section will loop through `years.use` and `dir.use`, running each BBS route (temooral analysis) or spatial transect by year (spatial analysis) at a time. Results are saved in directories created earlier.

```

# **Please note: depending on the # of years and spatial transects, this could take a
if(import.ind==FALSE) for (j in 1:length(dir.use)) {

  # For east-west analysis

  if (direction == "East-West"){

    birdsData <- feathers %>%
      filter(rowID == dir.use[j]) %>%
      mutate(direction = direction,
             dirID = dir.use[j])

  }

  # For south-north analysis

  if (direction == "South-North"){

    birdsData <- feathers %>%
      filter(colid == dir.use[j]) %>%
      mutate(direction = direction,
             dirID = dir.use[j])
  }
}
```

```

}

if (nrow(birdsData) < min.samp.sites) {

  next(print(paste0("Not enough data to analyze. Skipping j-loop ", dir.use[

} 

# VX. Analyze the data -----
#-----#
for (i in 1:length(years.use)) {

  # a. Subset the data according to year, colID, rowID, state, country, etc.
  birdData <- birdsData %>%
    filter(year == years.use[i]) %>%
    dplyr::rename(variable = aou,
                  value = stoptotal)

  if (nrow(birdData) == 0){

    next

  }

  # b. Munge the data further
  birdData <- mungeSubsetData(birdData)

# X. Calculate the metrics -----
## This function analyzes the data and writes results to file (in subdirec
# browser()

  calculateMetrics(dataIn = birdData, metrics.to.calc, direction = direction)
}

```

```

    print(paste0("End i-loop (years) ", i, " of ", length(years.use)))

}

} # end i-loop

print(paste0("End j-loop (transects) ", j, " of ", length(dir.use)))
} # end j-loop

```

## .1.9 Import and munge the results to prepare for visualization

First, import and combine the results as generated in the previous section. This following chunk imports the EWS results and the distance results separately, combining each into their own data frames.

```

# Import EWS results

if(!"results_ews.RDS" %in% fns){results_ews <-
  importResults(resultsDir = resultsDir, myPattern = 'ews',
                subset.by = direction) %>%
  # assign the end of the window as the cellID
  mutate(cellID = cellID_max)
  saveRDS(results_ews, file = paste0(tempResultsDir, "results_ews.RDS"))}

## FYI: rows will likely be absent when metricTypes = c(FI, VI), because FI and VI are

# Import distance results

if(!"results_dist.RDS" %in% fns){ results_dist <-
  importResults(resultsDir = resultsDir, myPattern = 'distances', subset.by = direction)

  saveRDS(results_dist, file = paste0(tempResultsDir, "results_dist.RDS"))}

```

```

if(import.ind){

  fn = list.files(tempResultsDir, "results_ews", full.names = T)
  results_ews <- readRDS(fn)

  fn = list.files(tempResultsDir, "results_dist", full.names = T)
  results_dist <- readRDS(fn)

  rm(fn)

}

```

Next, get the results to align with our sampling grid for visualizing results across space.

```

# Get the spatial sampling grid coordinates

coords_grd <-
  cbind(routes_gridList$sp_grd@data,
        coordinates(routes_gridList$sp_grd)) %>%
  rename(lat = s2,
         long = s1,
         cellID = id)

# Join coords_grd with results

# note: a full join will likely produce many cells with NO results data..
# but NO lat or long should == NA!

distResults <-
  full_join(coords_grd,
            results_dist) %>%
  na.omit(metricType)

```

```

ewsResults <-
  full_join(coords_grd,
             results_ews) %>%
  na.omit(metricType) %>%
  dplyr::select(-cellID_min, -cellID_max, -winStart , -winStop)

# d. Set coordinate system and projection
coordinates(distResults) <-
  coordinates(ewsResults) <- c("long", "lat")
sp::proj4string(distResults) <-
  sp::proj4string(ewsResults) <-
  sp::CRS("+proj=longlat +datum=WGS84")

```

## .1.10 Visualize results: one regime detection metric at a time

First, specify plotting parameters. We can visualize either the distance results (`distResults`) or the early-warning signal results (`ewsResults`). Define the results we want to visualize:

```

# Specify results to visualize
plotResults <- distResults

# Which metric do we want to visualize
metric.ind <- "s"

# Sort the years
year.ind <- unique(plotResults@data$year) %>% sort()

```

```
# Create a label for plotting, depending on direction
sortVar.lab <-
  ifelse(unique(plotResults@data$direction) == "South-North",
        "latitude",
        "longitude")
```

Plot the individual transects. Note: please specify dirID.ind as desired spatial transect number, and dirInd as direction (E-W or N-S)

```
# Specify the transect # we want to see
dirID.ind <- 13

metric.ind <- "s"

pl1 <- sort.year.line(plotResults, metric.ind, year.ind, dirID.ind, dirInd, scale = 1,
# +
#      # place a v-line at the
#      geom_vline(aes(xintercept=-96.8), color = "grey", linetype = 2)
pl1
```

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

## .1 About

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)

## .2 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 <- den-
sity(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 = min(which(bws == 1)) hnull = ks[r] return(hnull) }

```

### .3 Bootstrapping Function

```

DD <- function(log10.data, hnull, Sample.N = 1000) {
  NNull <- density(log10.data,
  bw = hnull, "gaussian", adjust = 1)
  N <- length(log10.data)

  # generate matrix null.samples <- matrix(data = 0, ncol = Sample.N, nrow = N)
  for (i in 1:Sample.N) {
    #sample the null model
    rand.N <- sample(NNull$x, N, replace = TRUE, prob = NNull$y)
    #calculate the gaps
    null.samples[, i] <- sort(rand.N, decreasing = FALSE) #put into the matrix
  }

  # generate gaps
  gaps.log10.data <- diff(log10.data)
  gaps.null.samples <- diff(null.samples, decreasing = FALSE) # difference between random samples and 1st
  diff orig dat
  gap.percentile <- matrix(data = 0, nrow = length(gaps.log10.data), ncol = 1)
  for (i in 1:length(gaps.log10.data)) {
    # generate distribution of gaps per row
    # (per gap rank)
    gap.percentile[i] <- ecdf(gaps.null.samples[i, ])(gaps.log10.data[i]) # returns the percentile at each observation
  }

  Bootstrap.gaps <- rbind(gap.percentile, 0)
  Bootstrap.gaps <- cbind(log10.data,
  Bootstrap.gaps)
  return(Bootstrap.gaps)
}

```

# References

- Abadi, F., Gimenez, O., Arlettaz, R. & Schaub, M. (2010) An assessment of integrated population models: Bias, accuracy, and violation of the assumption of independence. *Ecology*, **91**, 7–14.
- Alexandersson, H. (1986) A homogeneity test applied to precipitation data. *Journal of climatology*, **6**, 661–675.
- Alheit, J., Möllmann, C., Dutz, J., Kornilovs, G., Loewe, P., Mohrholz, V. & Wasmund, N. (2005) Synchronous ecological regime shifts in the central Baltic and the North Sea in the late 1980s. *ICES Journal of Marine Science: Journal du Conseil*, **62**, 1205–1215.
- Allaire, J., Xie, Y., McPherson, J., Luraschi, J., Ushey, K., Atkins, A., Wickham, H., Cheng, J., Chang, W. & Iannone, R. (2019) *Rmarkdown: Dynamic documents for r*,
- Allen, C.R. (2006) Discontinuities in ecological data. *Proceedings of the National Academy of Sciences*, **103**, 6083–6084.
- Allen, C.R., Forys, E.A. & Holling, C. (1999) Body mass patterns predict invasions and extinctions in transforming landscapes. *Ecosystems*, **2**, 114–121.
- Allen, C.R., Garmestani, A., Havlicek, T., Marquet, P.A., Peterson, G., Restrepo, C., Stow, C. & Weeks, B. (2006) Patterns in body mass distributions: Sifting among

- alternative hypotheses. *Ecology Letters*, **9**, 630–643.
- Allen, C.R. & Holling, C.S. (2001) Cross-scale morphology. *Encyclopedia of Environments and metrics*.
- Allen, C.R. & Holling, C.S. (2002) Cross-scale structure and scale breaks in ecosystems and other complex systems. *Ecosystems*, **5**, 315–318.
- Andersen, T., Carstensen, J., Hernandez-Garcia, E. & Duarte, C.M. (2009) Ecological thresholds and regime shifts: Approaches to identification. *Trends in Ecology & Evolution*, **24**, 49–57.
- Anderson, P.J. & Piatt, J.F. (1999) Community reorganization in the Gulf of Alaska following ocean climate regime shift. *Marine Ecology Progress Series*, 117–123.
- Angeler, D.G., Allen, C.R., Barichievy, C., Eason, T., Garmestani, A.S., Graham, N.A., Granholm, D., Gunderson, L.H., Knutson, M., Nash, K.L. & others (2016) Management applications of discontinuity theory. *Journal of Applied Ecology*, **53**, 688–698.
- Aria, M. & Cuccurullo, C. (2017) Bibliometrix: An r-tool for comprehensive science mapping analysis. *Journal of Informetrics*, **11**, 959–975.
- Arnold, J.B. (2019) *Ggthemes: Extra themes, scales and geoms for 'ggplot2'*,
- Auguie, B. (2017) *GridExtra: Miscellaneous functions for "grid" graphics*,
- Bahlai, C.A., Werf, W. vander, O'Neal, M., Hemerik, L. & Landis, D.A. (2015) Shifts in dynamic regime of an invasive lady beetle are linked to the invasion and insecticidal management of its prey. *Ecological Applications*, **25**, 1807–1818.
- Baker, M.E. & King, R.S. (2010) A new method for detecting and interpreting biodiversity and ecological community thresholds. *Methods in Ecology and Evolution*, **1**, 25–37.

Barichievy, C., Angeler, D.G., Eason, T., Garmestani, A.S., Nash, K.L., Stow, C.A., Sundstrom, S. & Allen, C.R. (2018) A method to detect discontinuities in census data. *Ecology and Evolution*, **8**, 9614–9623.

Batt, R.D., Carpenter, S.R., Cole, J.J., Pace, M.L. & Johnson, R.A. (2013) Changes in ecosystem resilience detected in automated measures of ecosystem metabolism during a whole-lake manipulation. *Proceedings of the National Academy of Sciences*, **110**, 17398–17403.

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

Beck, K.K., Fletcher, M.-S., Gadd, P.S., Heijnis, H., Saunders, K.M., Simpson, G.L. & Zawadzki, A. (2018) Variance and rate-of-change as early warning signals for a critical transition in an aquatic ecosystem state: A test case from Tasmania, Australia. *Journal of Geophysical Research: Biogeosciences*.

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

Benedetti-Cecchi, L., Tamburello, L., Maggi, E. & Bulleri, F. (2015) Experimental perturbations modify the performance of early warning indicators of regime shift. *Current biology*, **25**, 1867–1872.

Bennett, E.M., Peterson, G.D. & Gordon, L.J. (2009) Understanding relationships among multiple ecosystem services. *Ecology letters*, **12**, 1394–1404.

Bestelmeyer, B.T., Ellison, A.M., Fraser, W.R., Gorman, K.B., Holbrook, S.J., Laney, C.M., Ohman, M.D., Peters, D.P., Pillsbury, F.C., Rassweiler, A. & others (2011) Analysis of abrupt transitions in ecological systems. *Ecosphere*, **2**, 1–26.

Bibi, H., Raffaelli, D., Iqbal, M., Sharif, M. & Khattak, M.J.K. (2019) Body size pat-

- ters in stream communities: A test of holling?S textural discontinuity hypothesis. *Journal of the Entomological Research Society*, **21**, 17–35.
- Biggs, R., Carpenter, S.R. & Brock, W.A. (2009) Turning back from the brink: Detecting an impending regime shift in time to avert it. *Proceedings of the National Academy of Sciences*, **106**, 826–831.
- Bivand, R.S., Pebesma, E. & Gomez-Rubio, V. (2013) *Applied spatial data analysis with R, second edition*, Springer, NY.
- Blackburn, T.M. & Gaston, K.J. (1994) Animal body size distributions: Patterns, mechanisms and implications. *Trends in Ecology & Evolution*, **9**, 471–474.
- Boettiger, C. & Hastings, A. (2012) Quantifying limits to detection of early warning for critical transitions. *Journal of the Royal Society Interface*, **9**, 2527–39.
- Boettiger, C., Ross, N. & Hastings, A. (2013) Early warning signals: The charted and uncharted territories. *Theoretical ecology*, **6**, 255–264.
- Brock, W.A. & Carpenter, S.R. (2010) Interacting regime shifts in ecosystems: Implication for early warnings. *Ecological Monographs*, **80**, 353–367.
- Brock, W. & Carpenter, S. (2006) Variance as a Leading Indicator of Regime Shift in Ecosystem Services. *Ecology and Society*, **11**.
- Bryan, J. (2017) *Gapminder: Data from gapminder*,
- Buishand, T.A. (1982) Some methods for testing the homogeneity of rainfall records. *Journal of hydrology*, **58**, 11–27.
- Burnett, J. (2019) *BbsRDM: Calculate regime detection metrics using bird communities*,
- Burnett, J. & Price, N. (2019) *RegimeDetectionMeasures: Calculate measures for*

*detecting ecological regimes,*

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

Butitta, V.L., Carpenter, S.R., Loken, L.C., Pace, M.L. & Stanley, E.H. (2017) Spatial early warning signals in a lake manipulation. *Ecosphere*, **8**, n/a–n/a.

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

Cabezas, H., Campbell, D., Eason, T., Garmestani, A., Heberling, M., Hopton, M., Templeton, J., White, D., Zanowick, M. & Sparks, R. (2010) San luis basin sustainability metrics project: A methodology for evaluating regional sustainability. USEPA. USA.

Cabezas, H. & Fath, B.D. (2002) Towards a theory of sustainable systems. *Fluid Phase Equilibria*, **194**, 3–14.

Carpenter, S. & Brock, W. (2010) Early warnings of regime shifts in spatial dynamics using the discrete fourier transform. *Ecosphere*, **1**, 1–15.

Carpenter, S. & Brock, W. (2011) Early warnings of unknown nonlinear shifts: A nonparametric approach. *Ecology*, **92**, 2196–2201.

Carpenter, S., Brock, W., Cole, J., Kitchell, J. & Pace, M. (2008a) Leading indicators of trophic cascades. *Ecology letters*, **11**, 128–138.

Carpenter, S.R. & Brock, W.A. (2006) Rising variance: A leading indicator of ecological transition. *Ecology letters*, **9**, 311–318.

Carpenter, S.R., Brock, W.A., Cole, J.J., Kitchell, J.F. & Pace, M.L. (2008b) Leading indicators of trophic cascades. *Ecology Letters*, **11**, 128–138.

Carpenter, S.R., Cole, J.J., Pace, M.L., Batt, R., Brock, W., Cline, T., Coloso, J., Hodgson, J.R., Kitchell, J.F., Seekell, D.A. & others (2011) Early warnings of regime shifts: A whole-ecosystem experiment. *Science*, **332**, 1079–1082.

Cazelles, B., Chavez, M., Berteaux, D., Ménard, F., Vik, J.O., Jenouvrier, S. & Stenseth, N.C. (2008) Wavelet analysis of ecological time series. *Oecologia*, **156**, 287–304.

Chang, W. (2018) *Webshot: Take screenshots of web pages*,

Chartrand, R. (2011) Numerical differentiation of noisy, nonsmooth data. *ISRN Applied Mathematics*, **2011**.

Clements, C.F. & Ozgul, A. (2016) Including trait-based early warning signals helps predict population collapse. *Nature Communications*, **7**.

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

Cobo, M.J., López-Herrera, A.G., Herrera-Viedma, E. & Herrera, F. (2011) An approach 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.

Contamin, R. & Ellison, A.M. (2009) Indicators of regime shifts in ecological systems: What do we need to know and when do we need to know it. *Ecological Applications*, **19**, 799–816.

Dakos, V., Carpenter, S.R., Brock, W.A., Ellison, A.M., Guttal, V., Ives, A.R., Kefi, S., Livina, V., Seekell, D.A. & Nes, E.H. van (2012a) Methods for detecting early

warnings of critical transitions in time series illustrated using simulated ecological data. *PloS one*, **7**, e41010.

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

Dakos, V., Van Nes, E.H., D'Odorico, P. & Scheffer, M. (2012b) Robustness of variance and autocorrelation as indicators of critical slowing down. *Ecology*, **93**, 264–271.

Daróczi, G. & Tsegelskyi, R. (2018) *Pander: An r 'pandoc' writer*,

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

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

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

Ditlevsen, P.D. & Johnsen, S.J. (2010) Tipping points: Early warning and wishful thinking. *Geophysical Research Letters*, **37**.

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

Dunning Jr, J.B. (2007) *CRC handbook of avian body masses*, CRC press.

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

- Eason, T. & Cabezas, H. (2012) Evaluating the sustainability of a regional system using Fisher information in the San Luis Basin, Colorado. *Journal of environmental management*, **94**, 41–49.
- Eason, T., Ching-Chuang, W., Sundstrom, S. & Cabezas, H. (2019) An information theory-based approach to assessing spatial patterns in complex systems. *Entropy*, **21**, 182.
- Eason, T., Garmestani, A.S. & Cabezas, H. (2014) Managing for resilience: Early detection of regime shifts in complex systems. *Clean Technologies and Environmental Policy*, **16**, 773–783.
- Easterling, D.R. & Peterson, T.C. (1995) A new method for detecting undocumented discontinuities in climatological time series. *International journal of climatology*, **15**, 369–377.
- Ebbesmeyer, C.C., Cayan, D.R., McLain, D.R., Nichols, F.H., Peterson, D.H. & Redmond, K.T. (1991) 1976 step in the pacific climate: Forty environmental changes between 1968-1975 and 1977-1984.
- Fath, B.D. & Cabezas, H. (2004) Exergy and Fisher Information as ecological indices. *Ecological Modelling*, **174**, 25–35.
- Fath, B.D., Cabezas, H. & Pawlowski, C.W. (2003) Regime changes in ecological systems: An information theory approach. *Journal of theoretical biology*, **222**, 517–530.
- Filatova, T., Polhill, J.G. & Ewijk, S. van (2016) Regime shifts in coupled socio-environmental systems: Review of modelling challenges and approaches. *Environmental modelling & software*, **75**, 333–347.
- Fisher, R.A. (1922) On the Mathematical Foundations of Theoretical Statistics.

*Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*, **222**, 309–368.

Folke, C., Carpenter, S., Walker, B., Scheffer, M., Elmquist, T., Gunderson, L. & Holling, C.S. (2004) Regime shifts, resilience, and biodiversity in ecosystem management. *Annu. Rev. Ecol. Evol. Syst.*, **35**, 557–581.

Francis, R.C. & Hare, S.R. (1994) Decadal-scale regime shifts in the large marine ecosystems of the north-east pacific: A case for historical science. *Fisheries oceanography*, **3**, 279–291.

Frieden, B.R. (1990) Fisher information, disorder, and the equilibrium distributions of physics. *Physical Review A*, **41**, 4265–4276.

Frieden, B.R., Plastino, A., Plastino, A.R. & Soffer, B.H. (2002) Non-equilibrium thermodynamics and Fisher information: An illustrative example. *Physics Letters A*, **304**, 73–78.

Frieden, B.R. & Soffer, B.H. (1995) Lagrangians of physics and the game of fisher-information transfer. *Physical Review E*, **52**, 2274.

Frieden, R. & Gatenby, R.A. (2010) *Exploratory data analysis using fisher information*, Springer Science & Business Media.

Gal, G. & Anderson, W. (2010) A novel approach to detecting a regime shift in a lake ecosystem. *Methods in Ecology and Evolution*, **1**, 45–52.

Garmestani, A.S., Allen, C.R. & Bessey, K.M. (2005) Time-series analysis of clusters in city size distributions. *Urban Studies*, **42**, 1507–1515.

Goossens, C. & Berger, A. (1987) How to recognize an abrupt climatic change? 31–45.

Graham, R. & Tél, T. (1984) Existence of a potential for dissipative dynamical systems. *Physical review letters*, **52**, 9.

- Groffman, P.M., Baron, J.S., Blett, T., Gold, A.J., Goodman, I., Gunderson, L.H., Levinson, B.M., Palmer, M.A., Paerl, H.W., Peterson, G.D., Poff, N.L., Rejeski, D.W., Reynolds, J.F., Turner, M.G., Weathers, K.C. & Wiens, J. (2006) Ecological Thresholds: The Key to Successful Environmental Management or an Important Concept with No Practical Application? *Ecosystems*, **9**, 7:e01614.
- Gröger, J.P., Missong, M. & Rountree, R.A. (2011) Analyses of interventions and structural breaks in marine and fisheries time series: Detection of shifts using iterative methods. *Ecological indicators*, **11**, 1084–1092.
- Guidoum, A.C. (2015) *Kedd: Kernel estimator and bandwidth selection for density and its derivatives.*,
- Guttal, V. & Jayaprakash, C. (2008) Changing skewness: An early warning signal of regime shifts in ecosystems. *Ecology letters*, **11**, 450–460.
- Guttal, V. & Jayaprakash, C. (2009) Spatial variance and spatial skewness: Leading indicators of regime shifts in spatial ecological systems. *Theoretical Ecology*, **2**, 3–12.
- Guttal, V., Jayaprakash, C. & Tabbaa, O.P. (2013) Robustness of early warning signals of regime shifts in time-delayed ecological models. *Theoretical ecology*, **6**, 271–283.
- Hastings, A. & Wysham, D.B. (2010) Regime shifts in ecological systems can occur with no warning. *Ecology letters*, **13**, 464–472.
- Havlicek, T.D. & Carpenter, S.R. (2001) Pelagic species size distributions in lakes: Are they discontinuous? *Limnology and Oceanography*, **46**, 1021–1033.
- Hawkins, S.J., Bohn, K. & Doncaster, C.P. (2015) Ecosystems: The rocky road to regime-shift indicators. *Current Biology*, **25**, R666–R669.

He, W., Feng, G., Wu, Q., Wan, S. & Chou, J. (2008) A new method for abrupt change detection in dynamic structures. *Nonlinear Processes in Geophysics*, **15**, 601–606.

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

Held, H. & Kleinen, T. (2004) Detection of climate system bifurcations by degenerate fingerprinting. *Geophysical Research Letters*, **31**.

Herkert, J.R. (1994) The effects of habitat fragmentation on midwestern grassland bird communities. *Ecological applications*, **4**, 461–471.

Hijmans, R.J. (2019) *Raster: Geographic data analysis and modeling*,  
Holling, C.S. (1992) Cross-scale morphology, geometry, and dynamics of ecosystems. *Ecological monographs*, **62**, 447–502.

Holling, C.S. (1973) Resilience and stability of ecological systems. *Annual review of ecology and systematics*, **4**, 1–23.

Hughes, T.P. (1994) Catastrophes, phase shifts, and large-scale degradation of a caribbean coral reef. *Science*, **265**, 1547–1551.

Hughes, T.P., Carpenter, S., Rockström, J., Scheffer, M. & Walker, B. (2013) Multiscale regime shifts and planetary boundaries. *Trends in ecology & evolution*, **28**, 389–395.

Iannone, R. (2018) *DiagrammeR: Graph/network visualization*,  
Initiative, N.A.B.C., Committee, U. & others (2009) The state of the birds, united states of america, 2009. *US Department of Interior: Washington, DC*, **5**.

Ives, A., Dennis, B., Cottingham, K. & Carpenter, S. (2003) Estimating community

stability and ecological interactions from time-series data. *Ecological monographs*, **73**, 301–330.

Ives, A.R. & Dakos, V. (2012) Detecting dynamical changes in nonlinear time series using locally linear state-space models. *Ecosphere*, **3**, 1–15.

Jo, S., Kim, G. & Jeon, J.-J. (2016) Bayesian analysis to detect abrupt changes in extreme hydrological processes. *Journal of Hydrology*, **538**, 63–70.

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

Karl, T.R. & Williams Jr, C.N. (1987) An approach to adjusting climatological time series for discontinuous inhomogeneities. *Journal of Climate and Applied Meteorology*, **26**, 1744–1763.

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

Kaufmann, H., Hutter, R., Skopik, F. & Mantere, M. (2015) A structural design for a pan-european early warning system for critical infrastructures. *e & i Elektrotechnik und Informationstechnik*, **132**, 117–121.

Kefi, S., Guttal, V., Brock, W.A., Carpenter, S.R., Ellison, A.M., Livina, V.N., Seekell, D.A., Scheffer, M., Nes, E.H. van & Dakos, V. (2014) Early warning signals of ecological transitions: Methods for spatial patterns. *PLoS one*, **9**, e92097.

Kleinen, T., Held, H. & Petschel-Held, G. (2003) The potential role of spectral properties in detecting thresholds in the earth system: Application to the thermohaline circulation. *Ocean Dynamics*, **53**, 53–63.

Kong, X., He, Q., Yang, B., He, W., Xu, F., Janssen, A.B., Kuiper, J.J., Van Gerven,

- L.P., Qin, N., Jiang, Y. & others (2017) Hydrological regulation drives regime shifts: Evidence from paleolimnology and ecosystem modeling of a large shallow chinese lake. *Global change biology*, **23**, 737–754.
- Lade, S.J. & Gross, T. (2012) Early warning signals for critical transitions: A generalized modeling approach. *PLoS computational biology*, **8**, e1002360.
- 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 Royal Meteorological Society*, **16**, 1197–1226.
- La Sorte, F.A., Lepczyk, C.A., Burnett, J.L., Hurlbert, A.H., Tingley, M.W. & Zuckerberg, B. (2018) Opportunities and challenges for big data ornithology. *The Condor: Ornithological Applications*, **120**, 414–426.
- Lindegren, M., Dakos, V., Gröger, J.P., Gårdmark, A., Kornilovs, G., Otto, S.A. & Möllmann, C. (2012) Early detection of ecosystem regime shifts: A multiple method evaluation for management application. *PLoS One*, **7**, e38410.
- Litzow, M.A. & Hunsicker, M.E. (2016) Early warning signals, nonlinearity, and signs of hysteresis in real ecosystems. *Ecosphere*, **7**, –.
- Liu, J., Dietz, T., Carpenter, S.R., Alberti, M., Folke, C., Moran, E., Pell, A.N., Deadman, P., Kratz, T., Lubchenco, J. & others (2007) Complexity of coupled human and natural systems. *science*, **317**, 1513–1516.
- Liu, S., Yamada, M., Collier, N. & Sugiyama, M. (2013) Change-point detection in time-series data by relative density-ratio estimation. *Neural Networks*, **43**, 72–83.
- Livina, V.N. & Lenton, T.M. (2007) A modified method for detecting incipient bifurcations in a dynamical system. *Geophysical Research Letters*, **34**.

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

Manly, B.F. (1996) Are there clumps in body-size distributions? *Ecology*, **77**, 81–86.

Mantua, N. (2004) Methods for detecting regime shifts in large marine ecosystems: A review with approaches applied to North Pacific data. *Progress in Oceanography*, **60**, 165–182.

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 stable states. *Nature*, **269**, 471.

Mayer, A.L., Pawlowski, C., Fath, B.D. & Cabezas, H. (2007) *Applications of fisher information to the management of sustainable environmental systems. Exploratory data analysis using fisher information*, pp. 217–244. Springer.

Mayer, A.L., Pawlowski, C.W. & Cabezas, H. (2006) Fisher information and dynamic regime changes in ecological systems. *Ecological modelling*, **195**, 72–82.

McKenney, D.W., Pedlar, J.H., Lawrence, K., Papadopol, P., Campbell, K. & Hutchinson, M.F. (2014) Change and evolution in the plant hardiness zones of canada. *BioScience*, **64**, 341–350.

Michener, W.K. & Jones, M.B. (2012) Ecoinformatics: Supporting ecology as a data-intensive science. *Trends in ecology & evolution*, **27**, 85–93.

Moustakides, G.V., Polunchenko, A.S. & Tartakovsky, A.G. (2009) Numerical comparison of cusum and shiryaev–roberts procedures for detecting changes in distributions. *Communications in Statistics - Theory and Methods*, **38**, 3225–3239.

Möllmann, C., Folke, C., Edwards, M. & Conversi, A. (2015) Marine regime shifts around the globe: Theory, drivers and impacts.

Mumby, P.J., Steneck, R.S. & Hastings, A. (2013) Evidence for and against the existence of alternate attractors on coral reefs. *Oikos*, **122**, 481–491.

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

Nash, K.L., Allen, C.R., Angeler, D.G., Barichievy, C., Eason, T., Garmestani, A.S., Graham, N.A., Granholm, D., Knutson, M., Nelson, R.J. & others (2014a) Discontinuities, cross-scale patterns, and the organization of ecosystems. *Ecology*, **95**, 654–667.

Nash, K.L., Allen, C.R., Barichievy, C., Nyström, M., Sundstrom, S. & Graham, N.A. (2014b) Habitat structure and body size distributions: Cross-ecosystem comparison for taxa with determinate and indeterminate growth. *Oikos*, **123**, 971–983.

Nes, E.H. van & Scheffer, M. (2005) Implications of spatial heterogeneity for catastrophic regime shifts in ecosystems. *Ecology*, **86**, 1797–1807.

Nicholls, K.H. (2011) Detection of regime shifts in multi-species communities: The bay of quinte phytoplankton example. *Methods in Ecology and Evolution*, **2**, 416–426.

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

Oerlemans, J. (1978) An objective approach to breaks in the weather. *Monthly weather review*, **106**, 1672–1679.

Pardieck, K., Ziolkowski Jr, D., Lutmerding, M. & Hudson, M. (2018) North american breeding bird survey dataset 1966–2017, version 2017.0. *US Geological Survey, Patuxent Wildlife Research Center, Laurel, Maryland, USA.* [online] URL:

<https://www.pwrc.usgs.gov/BBS/RawData>.

Parmesan, C. (2006) Ecological and evolutionary responses to recent climate change. *Annu. Rev. Ecol. Evol. Syst.*, **37**, 637–669.

Parparov, A. & Gal, G. (2017) Quantifying ecological stability: From community to the lake ecosystem. *Ecosystems*, **20**, 1015–1028.

Parparov, A., Gal, G. & Zohary, T. (2015) Quantifying the ecological stability of a phytoplankton community: The lake kinneret case study. *Ecological indicators*, **56**, 134–144.

Pawlowski, C.W. & Cabezas, H. (2008) Identification of regime shifts in time series using neighborhood statistics. *Ecological Complexity*, **5**, 30–36.

Pebesma, E.J. & Bivand, R.S. (2005) Sp:Classes and methods for spatial data in R. *R News*, **5**, 9–13.

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

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

Peterjohn, B.G. & Sauer, J.R. (1999) Population status of north american grassland birds from the north american breeding bird survey. *Studies in Avian Biology*, **19**, 27–44.

Peters, R.H. & Wassenberg, K. (1983) The effect of body size on animal abundance. *Oecologia*, **60**, 89–96.

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

**18**, 497–510.

Peterson, G., Allen, C.R. & Holling, C.S. (1998) Ecological resilience, biodiversity, and scale. *Ecosystems*, **1**, 6–18.

Pettitt, A. (1979) A non-parametric approach to the change-point problem. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, **28**, 126–135.

Price, Nathaniel B. & Burnett, J.L. (2019a) Tvdiff: An r package for numerical differentiation of noisy, nonsmooth data.

Price, N. & Burnett, J. (2019b) *Tvdiff: Numerical differentiation of noisy, nonsmooth data*,

Qi, M., Feng, M., Sun, T. & Yang, W. (2016) Resilience changes in watershed systems: A new perspective to quantify long-term hydrological shifts under perturbations. *Journal of Hydrology*, **539**, 281–289.

Ratajczak, Z., Carpenter, S.R., Ives, A.R., Kucharik, C.J., Ramiadantsoa, T., Stegner, M.A., Williams, J.W., Zhang, J. & Turner, M.G. (2018) Abrupt change in ecological systems: Inference and diagnosis. *Trends in ecology & evolution*.

R Core Team (2018) *R: A language and environment for statistical computing*, R Foundation for Statistical Computing, Vienna, Austria.

Reid, P.C., Hari, R.E., Beaugrand, G., Livingstone, D.M., Marty, C., Straile, D., Barichivich, J., Goberville, E., Adrian, R., Aono, Y., Brown, R., Foster, J., Groisman, P., Hélaouët, P., Hsu, H.-H., Kirby, R., Knight, J., Kraberg, A., Li, J., Lo, T.-T., Myneni, R.B., North, R.P., Pounds, J.A., Sparks, T., Stübi, R., Tian, Y., Wiltshire, K.H., Xiao, D. & Zhu, Z. (2016) Global impacts of the 1980s regime shift. *Global Change Biology*, **22**, 682–703.

Restrepo, C. & Arango, N. (2008) Discontinuities in the geographical range size of

north american birds and butterflies. *Discontinuities in ecosystems and other complex systems*. Columbia University Press, New York, New York, USA, 101–135.

Richard A. Becker, O.S. code by, Ray Brownrigg. Enhancements by Thomas P Minka, A.R.W.R. version by & Deckmyn., A. (2018) *Maps: Draw geographical maps*,

Roberts, C.P., Twidwell, D., Burnett, J.L., Donovan, V.M., Wonkka, C.L., Bielski, C.L., Garmestani, A.S., Angeler, D.G., Eason, T., Allred, B.W. & others (2018) Early warnings for state transitions. *Rangeland Ecology & Management*, **71**, 659–670.

Roberts, C.R., Allen, C.R., Angeler, D.G. & Twidwell, D. (2019) Shifting avian spatial regimes in a changing climate. *xxxx*, **xx**, xx–xx.

Rockström, J., Steffen, W.L., Noone, K., Persson, Å., Chapin III, F.S., Lambin, E., Lenton, T.M., Scheffer, M., Folke, C., Schellnhuber, H.J. & others (2009) Planetary boundaries: Exploring the safe operating space for humanity. *Ecology and society*.

Rodionov, S.N. (2005a) A brief overview of the regime shift detection methods. *Large-Scale Disturbances (Regime Shifts) and Recovery in Aquatic Ecosystems: Challenges for Management Toward Sustainability*, 17–24.

Rodionov, S.N. (2004) A sequential algorithm for testing climate regime shifts. *Geophysical Research Letters*, **31**.

Rodionov, S.N. (2005b) A sequential method for detecting regime shifts in the mean and variance. *Large-Scale Disturbances (Regime Shifts) and Recovery in Aquatic Ecosystems: Challenges for Management Toward Sustainability*, 68–72.

Rodionov, S. & Overland, J.E. (2005) Application of a sequential regime shift detection method to the Bering Sea ecosystem. *ICES Journal of Marine Science*, **62**, 328–332.

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

Cambridge.

Sagarin, R. & Pauchard, A. (2012) *Observation and ecology: Broadening the scope of science to understand a complex world*, Island Press.

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

Sauer, J.R., Niven, D.K., Hines, J.E., Ziolkowski, D J, Pardieck, K.L., Fallon, J.E. & Link, W.A. (2014) The north american breeding bird survey, results and analysis 1966 - 2015. Version 2.07.2017.

Sauer, J.R., Pardieck, K.L., Ziolkowski Jr, D.J., Smith, A.C., Hudson, M.-A.R., Rodriguez, V., Berlanga, H., Niven, D.K. & Link, W.A. (2017) The first 50 years of the north american breeding bird survey. *The Condor: Ornithological Applications*, **119**, 576–593.

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

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

Scheffer, M. & Carpenter, S.R. (2003) Catastrophic regime shifts in ecosystems: Linking theory to observation. *Trends in ecology & evolution*, **18**, 648–656.

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

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

3046–3055.

Seekell, D.A., Carpenter, S.R. & Pace, M.L. (2011) Conditional heteroscedasticity as a leading indicator of ecological regime shifts. *The American Naturalist*, **178**, 442–451.

Shriver, W.G., Jones, A.L., Vickery, P.D., Weik, A. & Wells, J. (2005) *The distribution and abundance of obligate grassland birds breeding in new england and new york. Bird conservation implementation and integration in the americas: Proceedings of the third international partners in flight conference*. Eds: Ralph, c. John; rich, terrell d., US Dept. of Agriculture, Forest Service, Pacific Southwest Research Station.

Siemann, E. & Brown, J.H. (1999) Gaps in mammalian body size distributions reexamined. *Ecology*, **80**, 2788–2792.

Silva, M. & Downing, J.A. (1995) The allometric scaling of density and body mass: A nonlinear relationship for terrestrial mammals. *The American Naturalist*, **145**, 704–727.

Skilled, J.J. & Maurer, B.A. (2008) The ecological significance of discontinuities in body-mass distributions. *Discontinuities in ecosystems and other complex systems*, 193–218.

Smith, V.H. & Schindler, D.W. (2009) Eutrophication science: Where do we go from here? *Trends in ecology & evolution*, **24**, 201–207.

Soetaert, K., Petzoldt, T. & Setzer, R.W. (2010) Solving differential equations in r: Package deSolve. *Journal of Statistical Software*, **33**, 1–25.

Solow, A.R. (1987) Testing for climate change: An application of the two-phase regression model. *Journal of Climate and Applied Meteorology*, **26**, 1401–1405.

Solow, A.R. & Beet, A.R. (2005) A test for a regime shift. *Fisheries Oceanography*, **14**, 236–240.

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

Sorte, F.A.L. & III, F.R.T. (2007) Poleward shifts in winter ranges of north american birds. *Ecology*, **88**, 1803–1812.

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

Spanbauer, T.L., Allen, C.R., Angeler, D.G., Eason, T., Fritz, S.C., Garmestani, A.S., Nash, K.L., Stone, J.R., Stow, C.A. & Sundstrom, S.M. (2016) Body size distributions signal a regime shift in a lake ecosystem. *Proceedings of the Royal Society B: Biological Sciences*, **283**, 20160249.

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

Stow, C., Allen, C.R. & Garmestani, A.S. (2007) Evaluating discontinuities in complex systems: Toward quantitative measures of resilience.

Sugihara, G., May, R., Ye, H., Hsieh, C.-h., Deyle, E., Fogarty, M. & Munch, S. (2012) Detecting causality in complex ecosystems. *science*, **338**, 496–500.

Sundstrom, S.M., Eason, T., Nelson, R.J., Angeler, D.G., Barichievy, C., Garmestani, A.S., Graham, N.A., Granholm, D., Gunderson, L., Knutson, M. & others (2017) Detecting spatial regimes in ecosystems. *Ecology letters*, **20**, 19–32.

Takens, F. (1981) *Detecting strange attractors in turbulence. Dynamical systems and*

- turbulence, warwick 1980*, pp. 366–381. Springer.
- Taranu, Z.E., Carpenter, S.R., Frossard, V., Jenny, J.-P., Thomas, Z., Vermaire, J.C. & Perga, M.-E. (2018) Can we detect ecosystem critical transitions and signals of changing resilience from paleo-ecological records? *Ecosphere*, **9**.
- Taylor, L. (1961) Aggregation, variance and the mean. *Nature*, **189**, 732.
- Thrush, S.F., Hewitt, J.E., Dayton, P.K., Coco, G., Lohrer, A.M., Norkko, A., Norkko, J. & Chiantore, M. (2009) Forecasting the limits of resilience: Integrating empirical research with theory. *Proceedings of the Royal Society B: Biological Sciences*, **276**, 3209–3217.
- Tong, H. (1990) Nonlinear time series analysis: A dynamical systems perspective. *San Diego, Calif: Academic*.
- Tuszynski, J. (2019) *CaTools: Tools: Moving window statistics, gif, base64, roc auc, etc.*,
- Van Auken, O. (2009) Causes and consequences of woody plant encroachment into western north american grasslands. *Journal of environmental management*, **90**, 2931–2942.
- Vasilakopoulos, P., Raitsos, D.E., Tzanatos, E. & Maravelias, C.D. (2017) Resilience and regime shifts in a marine biodiversity hotspot. *Scientific reports*, **7**, 13647.
- Vincent, L.A. (1998) A technique for the identification of inhomogeneities in canadian temperature series. *Journal of Climate*, **11**, 1094–1104.
- Walker, B., Holling, C.S., Carpenter, S. & Kinzig, A. (2004) Resilience, adaptability and transformability in social–ecological systems. *Ecology and society*, **9**.
- Walther, G.-R., Post, E., Convey, P., Menzel, A., Parmesan, C., Beebee, T.J., Fronmentin, J.-M., Hoegh-Guldberg, O. & Bairlein, F. (2002) Ecological responses to

- recent climate change. *Nature*, **416**, 389.
- Wang, Z., Li, G., Wang, G., Shen, Y. & Liu, Y. (2011) Application of fuzzy synthetic evaluation on the regime shift phases of lake erhai. *Journal of Hydroecology*, **32**, 53–59.
- Wardwell, D. & Allen, C.R. (2009) Variability in population abundance is associated with thresholds between scaling regimes.
- Weijerman, M., Lindeboom, H. & Zuur, A.F. (2005) Regime shifts in marine ecosystems of the north sea and wadden sea. *Marine Ecology Progress Series*, **298**, 21–39.
- Weissmann, H. & Shnerb, N.M. (2016) Predicting catastrophic shifts. *Journal of theoretical biology*, **397**, 128–134.
- Westgate, M.J. (2018a) Revtools: Bibliographic data visualization for evidence synthesis in r. *bioArXiv*.
- Westgate, M.J. (2018b) *revtools: Tools to support evidence synthesis*,
- Wickham, H. (2016) *Ggplot2: Elegant graphics for data analysis*, Springer-Verlag New York.
- Wickham, H. (2017) *Tidyverse: Easily install and load the 'tidyverse'*,
- Wickham, H., François, R., Henry, L. & Müller, K. (2019a) *Dplyr: A grammar of data manipulation*,
- Wickham, H., Hester, J. & Chang, W. (2019b) *Devtools: Tools to make developing r packages easier*,
- Wolkovich, E.M., Cook, B.I., McLauchlan, K.K. & Davies, T.J. (2014) Temporal ecology in the Anthropocene. *Ecology Letters*, **17**, 1365–1379.
- Xie, Y. (2018) *Bookdown: Authoring books and technical documents with r markdown*,

- Yang, Q. & Wu, X. (2006) 10 challenging problems in data mining research. *International Journal of Information Technology & Decision Making*, **05**, 597–604.
- Ye, H., Beamish, R.J., Glaser, S.M., Grant, S.C., Hsieh, C.-h., Richards, L.J., Schnute, J.T. & Sugihara, G. (2015) Equation-free mechanistic ecosystem forecasting using empirical dynamic modeling. *Proceedings of the National Academy of Sciences*, **112**, E1569–E1576.
- Yin, D., Leroux, S.J. & He, F. (2017) Methods and models for identifying thresholds of habitat loss. *Ecography*, **40**, 131–143.
- Yonetani, T. (1993) Detection of long term trend, cyclic variation and step-like change by the lepage test. *Journal of the Meteorological Society of Japan. Ser. II*, **71**, 415–418.
- Zhou, T. & Shumway, R. (2008) One-step approximations for detecting regime changes in the state space model with application to the influenza data. *Computational Statistics & Data Analysis*, **52**, 2277–2291.
- Zhu, H. (2019) *KableExtra: Construct complex table with 'kable' and pipe syntax*, Zurlini, G., Jones, K.B., Riitters, K.H., Li, B.-L. & Petrosillo, I. (2014) Early warning signals of regime shifts from cross-scale connectivity of land-cover patterns. *Ecological indicators*, **45**, 549–560.