

1

Regime Detection Measures for the Practical Ecologist

2

A Thesis

3

Presented to

4

The Division of

5

University of Nebraska-Lincoln

6

In Partial Fulfillment

7

of the Requirements for the Degree

8

Doctor of Philosophy

9

Jessica Leigh Burnett

10

July XX 2019

Approved for the Division
(School of Natural Resources)

Craig R. Allen

Dirac Twidwell

¹¹ Dedication

¹² To: those not yet exposed to the great outdoors; those unaware of the possibility of
¹³ careers in ecology and research; to Mike Moulton, my first and best cheerleader; to S;
¹⁴ and myself.

¹⁵ Table of Contents

¹⁶ Abstract	1
¹⁷ Acknowledgements	3
¹⁸ Table of Definitions	6
¹⁹ Chapter 1: Introduction	15
²⁰ 1.1 Dissertation aims	17
²¹ 1.2 Dissertation structure	17
²² 1.2.1 Chapter highlights and chronology	17
²³ 1.2.2 Software Appendices	20
²⁴ Chapter 2: A Brief Overview of the Ecological Regime Detection Literature	22
²⁵ 2.1 Introduction	22
²⁶ 2.1.1 Aims	24
²⁷ 2.2 Methods	25
²⁸ 2.2.1 Identifying candidate articles and regime detection measures	25
²⁹ Bibliographic analysis of ecological regime shift literature	28
³⁰ 2.3 Results	32
³¹ 2.3.1 Quantitative methods for identifying ecological regime shifts	32
³² 2.3.2 Bibliographic analysis of ecological regime shift literature	40
³³ 2.4 A synthesis of the methods available for the practical ecologist	45
³⁴ 2.4.1 Model-dependent	46
³⁵ 2.4.2 Model-free	47
³⁶ 2.5 Discussion	47
³⁷ Chapter 3: Decoupling the Calculation of Fisher Information	54
³⁸ 3.1 Abstract	54
³⁹ 3.2 Introduction	55
⁴⁰ 3.2.1 Fisher Information as a Regime Detection Method	56
⁴¹ 3.2.2 The Sustainable Regimes Hypothesis	57
⁴² 3.2.3 Fisher Information Requires Dimension Reduction	58
⁴³ 3.2.4 Aims	59
⁴⁴ 3.3 Methods	60

46	3.3.1	Predator-Prey Model System	60
47	3.3.2	Inducing a Regime Shift	60
48	3.3.3	Decoupling the Steps for Calculating Fisher Information	61
49	3.4	Discussion	65
50	3.5	Acknowledgements	67
51	Chapter 4: An application of Fisher Information to spatially-explicit avian community data	74
53	4.1	Introduction	74
54	4.2	Data and methods	77
55	4.2.1	Data: North American breeding bird communities	77
56	4.2.2	Calculating Fisher Information (FI)	83
57	4.2.3	Interpreting and comparing Fisher Information across spatial transects	85
59	4.3	Results	88
60	4.3.1	Fisher Information across spatial transects	88
61	4.3.2	Spatial correlation of Fisher Information	93
62	4.4	Discussion	95
63	Chapter 5: Velocity (v): using rate-of-change of system trajectory to identify abrupt changes	101
65	5.1	Introduction	101
66	5.1.1	Tracking ecosystem trajectory through time to explore system dynamics	103
67	5.1.2	Rate of change as an indicator of abrupt change in the system trajectory	104
69	5.1.3	Aims	105
71	5.1.4	Analytical approach	106
72	5.2	Steps for Calculating velocity, v	106
73	5.2.1	Steps for calculating v	108
74	5.3	Velocity v performance under a discontinuous transition	111
75	5.4	Velocity performance under a smooth transition	114
76	5.4.1	Generating the data	116
77	5.4.2	Evaluating velocity performance under conditions of changing means and/or variance	119
79	5.5	Velocity performance under empirical transitions: paleolithic freshwater diatom community	125
80	5.6	Discussion	128
82	5.7	Supplementary Figures	132
83	Chapter 6: Using Resampling Methods to Evaluate the Relative Performance of Regime Detection Measures Under Varying Conditions of Data Quality and Quantity	139
85	6.1	Introduction	139
86	6.2	Data and Methodology	142

88	6.2.1	Study system and data	142
89	6.2.2	Regime detection measures	143
90	6.2.3	Simulating data quality and quantity issues using resampling techniques	146
91	6.2.4	Comparing regime detection measures	146
92	6.3	Results	150
93	6.3.1	Velocity of the distance travelled (v)	150
94	6.3.2	Variance Index	150
95	6.3.3	Fisher Information is highly sensitive to information loss	151
96	6.4	Detrending the Data Prior to Calculations	153
97	6.5	Conclusion	157
98	6.6	Acknowledgements	158
99			
100	Chapter 7: Grassland Obligates and Declining Birds Operate Near Edges of Body Mass Distributions		159
101	7.1	Introduction	159
102	7.2	Methods	161
103	7.2.1	Study area	161
104	7.2.2	Data	163
105	7.2.3	Statistical analysis	168
106	7.3	Results	173
107	7.3.1	Summary statsitics of censuses (NABBS data)	173
108	7.3.2	Statistical Analysis	174
109	7.4	Discussion	177
110			
111	Chapter 8: Conclusions		192
112	8.1	Method mining regime detection methods	193
113	8.2	Ecological data are noisy	193
114	8.3	Data collection and munging biases and limits findings	194
115	8.4	Common Limitations of Regime Detection Measures	194
116	Appendix A: R package regimeDetectionMeasures		196
117	.1	Example Analysis	196
118	.1.1	Measures/metrics calculated	196
119	.1.2	Example analysis	197
120	.1.3	Plots	198
121	Appendix B: R package bbsRDM		201
122	.1	Example Analysis	201
123	.1.1	Load packages & create local directories	201
124	.1.2	Download the BBS data and save to file locally	202
125	.1.3	Create a sampling grid	203
126	.1.4	Subset the BBS data by species and/or functional traits (OPTIONAL but highly recommended)	206
127			

128	.1.5	Subset species according to AOU species codes (i.e. by family, genera, etc..)	206
129	.1.6	Subset species by trait, body mass, taxonomically, etc... (optional)	206
130	.1.7	Calculate regime detection metrics across space or time	207
131	.1.8	Conduct analysis	208
132	.1.9	Import and munge the results to prepare for visualization . . .	210
133	.1.10	Visualize results: one regime detection metric at a time . . .	213
135	Appendix C: Functions used to calculate discontinuities in avian body mass distributions.		215
136	.1	About	215
137	.2	Neutral.Null function	215
138	.3	Bootstrapping Function	216
140	References		218

¹⁴¹ List of Tables

¹⁴² 1	A table of definitions for terms, theories, and phrases often appearing in ecological regime shift literature.	¹⁴³ 7
¹⁴⁴ 2.1	List of the regime detection methods identified in this review.	¹⁴⁵ 33
¹⁴⁵ 2.2	Potential questions for a comprehensive review of the ecological regime detection metrics literature.	¹⁴⁶ 51
¹⁴⁷ 5.1	Steps outlined for calculating system velocity, v , using the 2-variable toy data as an example.	¹⁴⁸ 111
¹⁴⁹ 5.2	Conditions for generating various scenarios of the hyperbolic tangent-induced abrupt change. σ_i represents the standard deviation of μ_{x_i} as the percent of μ_{x_i} , μ_{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.	¹⁵⁰ 118
¹⁵⁴ 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.	¹⁵⁵ 170
¹⁵⁶ 7.2	Coefficient estimates for the linear mixed effects model predicting species' 'distance to edge' of a body mass distribution.	¹⁵⁷ 179
¹⁵⁸ 7.3	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.	¹⁵⁹ 179
¹⁶² 7.4	Summary statistics for annual species richness and annual turnover in all NABBS routes in study area.	¹⁶³ 181

¹⁶⁴ **List of Figures**

165	2.1	Number of publications by year in fields 'Ecology' and 'Biodiversity 166 Conservation' which included terms related to 'regime shift' (total = 167 654).	30
168	2.2	Number of articles containing phrases related to ecological 'regime 169 shifts' published per publication outlet.	31
170	2.3	Flowchart of the litearture review process for identifying new regime 171 detection methods. Note: only the first ten pages (250 articles) of 172 Google Scholar results were examined. Node shapes: folder = unfiltered 173 articles; box = articles actively filtered; diamond = number of articles 174 with new methods.	39
175	2.4	Number of methods published over time.	40
176	2.5	Chronological direct citation newtwork suggests the intellectual struc- 177 ture can be mapped to a few papers. This historiograph identifies 178 important works explicitly in chronological, as opposed to absolute, order.	42
179	2.6	Total number of articles published and corresponding number of cita- 180 tions (for papers published that year). The most highly cited papers 181 to-date are those published in the late 2000s.	43
182	2.7	Total number of articles published per year by category as categorized 183 by ISI. Book chapters, proceedings, editorials, and letters are excluded.	44
184	2.8	Number of articles containing new methods per journal.	50
185	3.1	From top to bottom, distance traveled in phase space, speed tangential 186 to system trajectory, acceleration tangential to system trajectory.	68
187	3.2	Carrying capacity over time with a regime shift occurring around time 188 200.	69
189	3.3	Phase space plot of system trajectories for different values of k	70
190	3.4	Distance travelled in phase space over time. Dashed vertical line at 191 time 200 indicates location of regime shift.	71
192	3.5	Speed of the system (rate of change, velocity) in phase space. Dashed 193 vertical line at time 200 indicates location of regime shift.	72
194	3.6	Fisher Information calculated for non-overlapping time windows. Two 195 different window sizes were used as indicated by color. Dashed vertical 196 line at time 200 indicates approximate location of regime shift.	73

197	4.1 Locations of Breeding Bird Survey routes sampled between 1966 and 198 2017.	76
199	4.2 A single East-West transect of Breeding Bird Survey routes used to 200 calculate the Fisher Information.	78
201	4.3 Locations of Fort Riley military base in Kansas, USA.	80
202	4.4 Locations of focal U.S. military bases, Eglin Air Force Base (AFB) and 203 Fort Riley Military Base.	81
204	4.5 The three East-West running transects used to visualize results in this 205 chapter.	82
206	4.6 An example of two adjacent spatial transects (12, 13) within my sam- 207 pling grid.	87
208	4.7 Fisher Information calculated for a single transect over time.	89
209	4.8 Fisher Information of two spatially adjacent transect pairs (transects 210 12, 13) over time. Interdecadal trends in FI are very different within 211 each transect and are not highly correlated among transects over time.	91
212	4.9 Fisher Information of 5 East-West spatial transects over time.	92
213	4.10 No patterns of abrupt change detected in Fisher Information along 214 three transects in year 2010	93
215	4.11 Pairwise relationships of Fisher Information (interpolated values) of 216 spatially adjacent transects over time do not exhibit expected patterns 217 of high positive correlation.	94
218	4.12 Fisher Information (scaled and centered; point size positively correlated 219 with value) against ecoregion boundaries (EPA Level 2).	96
220	5.1 An example solution of the Lorenz ('butterfly') represented in 221 3-dimensional phase-space. Phase plots are typically used to visualize 222 stable areas within a system's trajectory but reconstruction requires 223 the difference models to be known and parameterized.	102
224	5.2 An example solution of the Lorenz ('butterfly') represented in individual 225 system components.	103
226	5.3 The 2-variable discrete time toy system used to demonstrate steps for 227 calculating system velocity. Each variable, x , is drawn from a normal 228 distribution with means that change at $t = 50$. State variables have 229 constant standard deviation, $\sigma = 5$	107
230	5.4 Data used to calculate velocity at the first two time points, t_1 and t_2	108
231	5.5 Distance traveled, s , for the 2-species toy system.	109
232	5.6 System change (s) and velocity (v) of the model system over the time 233 period. Constant means ($\bar{x}_{pre} = 25$, $\bar{x}_{post} = 10$) and sharp change in 234 variance for both state variables, $\sigma = 5$	110
235	5.7 Velocity (v) generally increases as the total change in the mean value of 236 $\bar{x}_{1_{t=50}}$ increases in a single iteration of our toy system ($N_{iter} = 1$, seed 237 = 123). This 2-variable system exhibits a regime shift at $t = 50$, where 238 variance is constant $\sigma = 5$, $\bar{x}_1 = 25$ when $t < 50$, $\bar{x}_2 = 50$ when $t \geq 50$, 239 $\bar{x}_1 = 25$ when $t < 50$	113

240	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 varoance $\sigma = 5$, $\bar{x}_2 = 25$ when $t < 50$, and changes in variable mean values, $\bar{x}_2 = 50$ when $t \geq 50$, $\bar{x}_1 = 25$ when $t < 50$	114
241		
242		
243		
244	5.9 High variance of velocity (v) in a single iteration ($N_{iter} = 1$, seed = 123) of simulations as we increase σ_1 at $t = 50$	115
245		
246	5.10 Average (± 2 SD) velocity (v) worsens as the variance of $\bar{x}_{2_{t=50(post)}}$ (post shift) increases. $\bar{x}_{1_{pre}} = 25$, $\bar{x}_{1_{post}} = 100$, $\bar{x}_{2_{pre}} = 25$, $\bar{x}_{2_{post}} = 50$, $\sigma_{1_{pre}} = 5$, $\sigma_{2_{pre,post}} = 5$	116
247		
248		
249	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.	117
250		
251	5.12 An example of the data generated by the true process model. In this example the mean values (μ_{xi}), but not the percent standard deviation (σ_{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, α (top: $\alpha = 0.25$, bottom: $\alpha = 1.00$).	119
252		
253		
254		
255		
256		
257	5.13 An example of the data generated by the true process model. In this example the mean values (μ_{xi}), but not the percent standard deviation (σ_{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, α (top: $\alpha = 0.25$, bottom: $\alpha = 1.00$).	120
258		
259		
260		
261		
262		
263	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 smooothing parameter, α . Under most conditions $\alpha \ll$ sufficed. Here, we compare the true and antidifferentiated values of s under the condition of changing μ_{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 α increases.	122
264		
265		
266		
267		
268		
269		
270		
271	5.15 The velocity signal is muted when the hyperbolic smoothing parameter, α , 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).	123
272		
273		
274		
275	5.16 The velocity signal is muted when the hyperbolic smoothing parameter, α , 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).	124
276		
277		
278		
279	5.17 The velocity signal is muted when the hyperbolic smoothing parameter, α , 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).	125
280		
281		
282		

283	5.18 The velocity signal is muted when the hyperbolic smoothing parameter, 284 α , is moderate (0.50). True and observed values of x_i (panel A), 285 observed distance traveled (s , panel B), observed velocity (C), and the 286 smoothed velocity (D).	126
287	5.19 The velocity signal is regained under smooth transition ($\alpha_{tanh} = 0.25$) 288 when both state variables undergo a shift in the mean. True and 289 observed values of x_i (panel A), observed distance traveled (s , panel 290 B), observed velocity (C), and the smoothed velocity (D).	127
291	5.20 The velocity signal is regained under smooth transition ($\alpha_{tanh} = 0.75$) 292 when both state variables undergo a shift in the mean. True and 293 observed values of x_i (panel A), observed distance traveled (s , panel 294 B), observed velocity (C), and the smoothed velocity (D).	128
295	5.21 The velocity signals a rapid shift in the variance of both state variables 296 under a moderately abrupt transition ($\alpha_{tanh} = 0.75$). True and observed 297 values of x_i (panel A), observed distance traveled (s , panel B), observed 298 velocity (C), and the smoothed velocity (D).	129
299	5.22 The velocity does not signal shifts in the variance of a single variable 300 (x_1) under a moderately abrupt transition ($\alpha_{tanh} = 0.75$). True and observed 301 values of x_i (panel A), observed distance traveled (s , panel B), observed 302 velocity (C), and the smoothed velocity (D).	130
303	5.23 The velocity signals a shift when both variables undergo shifts in the 304 mean and variance under a slightly abrupt transition ($\alpha_{tanh} = 0.25$). 305 True and observed values of x_i (panel A), observed distance traveled 306 (s , panel B), observed velocity (C), and the smoothed velocity (D). .	131
307	5.24 The velocity signals a shift when both variables undergo shifts in the 308 mean and variance under a slightly abrupt transition ($\alpha_{tanh} = 1.00$). 309 True and observed values of x_i (panel A), observed distance traveled 310 (s , panel B), observed velocity (C), and the smoothed velocity (D). .	132
311	5.25 Relative abundances of the most common diatom species in the time 312 series. Few species dominate the data over the entire time series, and 313 turnover is apparent at multiple observations.	133
314	5.26 Velocity v and distance traveled s of the paleodiatom time series. 315 Dashed line at 1,300 years before 1950 indicates the regime shift identified 316 in Spanbauer et al. (2014). Dotted lines indicate regime shifts as visually 317 identified on metrics s and v	134
318	5.27 Velocity (v) indicates periodic-like conditions in the first (A) and second 319 (B) regimes.	135
320	5.28 The regularized differentiation of s was best fit using $\alpha = 100$. Higher 321 overlap of s and pred indicates a good fit of the regularized differentiated 322 metric to the non-smoothed metric, s	136
323	5.29 The velocity metric (v) signals potential periodicities in the paleo diatom 324 time series data when the distance traveled metric, s , is smoothed using 325 regularized differentiation methods [@price2019tvdiff].	137

326	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 ($\sigma_{1_{pre}} = 2$, $\sigma_{1_{post}} = 10$, $\sigma_{2_{pre}} = 5$, $\sigma_{2_{post}} = 10$).	137
330	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 $\sigma_{1_{pre}} = 2$, $\sigma_{1_{post}} = 12$, $\sigma_{2_{pre,post}} = 5$	138
333	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 ($/bar{x}_{1_{pre}} = 25$, $/bar{x}_{1_{post}} = 50$, $/bar{x}_{2_{pre}} = 15$, $/bar{x}_{2_{post}} = 150$). . . .	138
336	6.1 Relative abundances of the diatom species in Foy Lake over the time period.	154
338	6.2 The amount of time elapsed between observations.	155
339	6.3 Local regression (loess) smoothing of a dominant species in the paleodiatom community, extitAnomoeoneis costata varies with the span parameter, making it difficult to justify smoothing the data prior to calculating various regime detection metrics.	156
343	7.1 My study area (red box) overlaying the terrestrial Breeding Conservation Regions (BCR) in North America.	162
345	7.2 North American Breeding Bird Survey routes (points), latitudinal locations (horizontal bars) of the spatial regimes identified by roberts2019shifting.	166
348	7.3 Total number of birds across the entire study area per species group per year.	175
350	7.4 Average number of birds across the entire study area per species group per year.	176
352	7.5 Average number of birds across the entire study area per species group per year.	177
354	7.6 Number of body mass aggregations identified in each route unchanged across the time period (**a**) and is approximately normally distributed ($\bar{x} = 4.7$, $\sigma = 1.6$)	178
357	7.7 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.	184
360	7.8 Aggregation locations of body mass distributions of the avian community at a single NABBS route (state 7 route 24) appear relatively similar across time.	185
363	7.9 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.	186
366	7.10 The body mass distribution of declining species appears to differ slightly between the Southern and Northern regimes.	187

368	7.11 Body mass distribution for species in the study area over the entire	
369	time period varies by species group. Distributions represent the species	
370	pool for each group and Bird Conservation Region over all years. . . .	188
371	7.12 Number of NABBS routes analysed per year. Some routes are not	
372	sampled annually due to volunteer availability, environmental conditions,	
373	or route discontinuation.	189
374	7.13 Species richness increases over time across the entire study area. . . .	190
375	7.14 Variance in species turnover increases over time across the entire study	
376	area.	191

³⁷⁷ Abstract

³⁷⁸ Identifying abrupt changes in the structure and functioning of systems, or system
³⁷⁹ regime shifts, in ecological and social-ecological systems leads to an understanding
³⁸⁰ of relative and absolute system resilience. Resilience is an emergent phenomenon of
³⁸¹ complex social-ecological systems, and is the ability of a system to absorb distur-
³⁸² bance without reorganizing into a new state, or regime. Resilience science provides
³⁸³ a framework and methodology for quantitatively assessing the capacity of a system
³⁸⁴ to maintain its current trajectory (or to stay within a certain, and often desirable
³⁸⁵ regime). If and when a system's resilience is exceeded, it crosses a threshold and
³⁸⁶ enters into an alternate regime (or undergoes a regime shift).

³⁸⁷ I use Fisher Information to detect regime shifts in time and space using avian commu-
³⁸⁸ nity data obtained from the North American Breeding Bird Survey within the area
³⁸⁹ east of the Rockies and west of the Mississippi River. Fisher Information is a technique
³⁹⁰ that captures the dynamic of a system, and this metric will be calculated about a suite
³⁹¹ of bird species abundances aggregated to the route level for all possible time periods.
³⁹² Transmutation (aggregation error) about inclusion or exclusion of certain bird species,
³⁹³ functional groups, and guilds will be analyzed. Efforts have been made to develop
³⁹⁴ early warning indicators of regime shifts in ecosystems, however, for most ecosystems
³⁹⁵ there is great uncertainty in predicting the risk of a regime shift, regarding both when
³⁹⁶ and how long it will take to happen and if it can be recognized early enough to be
³⁹⁷ avoided when desired. We will complement the use of Fisher Information with multiple

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

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

⁴¹³ Acknowledgements

⁴¹⁴ Graduate school itself isn't hard, but the journey is. I have a lot of people and institu-
⁴¹⁵ tions to thank for their emotional, intellectual, financial, and other support. I wish
⁴¹⁶ to first highlight how **great it was to be a graduate student at this university**
⁴¹⁷ **and in the School of Natural Resources.** UNL has provided tremendous support
⁴¹⁸ at all levels of the university. Although I am not a fan of Nebraska's climate, I
⁴¹⁹ highly recommend this school to prospective students. I thank my supervisors, Craig
⁴²⁰ Allen and Dirac Twidwell, for providing me with this amazing opportunity and for
⁴²¹ supporting my growth as an independent researcher. I thank my also committee
⁴²² members, Craig Allen, David Angeler, John De Long, Dirac Twidwell, and Drew Tyre
⁴²³ for their support and advisement, but especially for their comprehensive examination—I
⁴²⁴ found this process transformative (albeit very stress-inducing). I particularly thank
⁴²⁵ Dirac for his examination questions—I never knew how much I didn't know until I
⁴²⁶ studied your recommendations. I also thank Craig and Dirac for supporting my efforts
⁴²⁷ to study and conduct research in Austria.

⁴²⁸ Studying at the International Institute for Applied Systems Analysis was an amazing
⁴²⁹ opportunity! I thank Brian Fath and Elena Rovenskaya for their advisement, members
⁴³⁰ of the Applied Systems Analysis research group for their feedback on my research,
⁴³¹ and to the postdocs and YSSPers. I owe thanks to Craig Allen and Kevin Pope
⁴³² for entertaining my many hours of discussion (interrogation?) regarding federal

433 employment.

434 I would like to especially thank some of the amazing and brilliant **female scientists**
435 in my life for their encouragement: Jane Anderson, Karen Bailey, Hannah Birge, Mary
436 Bomberger Brown, Tori Donovan, Brittany Dueker, Allie Schiltmeyer, Katie Sieving,
437 Erica Stuber, Becky Wilcox, Carissa Wonkka, and Lyndsie Wszola. I thank these
438 women and others for their contributions to my professional development: David
439 Angeler, Christie Bahlai, Mary Bomberger Brown, John Carroll, Jenny Dauer, John
440 DeLong, Tarsha Eason, Brian Fath, Ahjond Garmestani, Chris Lepczyk, Frank La
441 Sorte, Chai Molina, Zac Warren, Hao Ye. I also thank fellow graduate students with
442 whom I forged long-lived personal and professional relationships: Hannah Birge, Tori
443 Donovan, Caleb Roberts, Allie Schiltmeyer, and Lyndsie Wszola. It is also worth
444 noting that I among those afflicted with mental health “disorders”. I am first grafteful
445 to one friend (H) who unknowingly destigmatized mental health in my mind and
446 without whom I may have never sought treatment. I also applaud fellow students and
447 faculty in SNR and UNL who have been active in promoting positive mental health. I
448 am forever indebtted to my general practitioner and mental health advocate, Terry
449 Thomas M.A., M.S.N., A.P.R.N.

450 This research was funded by the U.S. Department of Defense’s Strategic Environmental
451 Research and Development Program (SERDP project ID: RC-2510). The University
452 of Nebraska-Lincoln (UNL) has been highy supportive in my doctoral studies and
453 reserach. I am grateful for the generous of donors to the University of Nebraska
454 Foundation, which provided me with two prestigious supplemental fellowships: Fling
455 and Othmer. I also thank the Nelson Family (Nelson Memorial Fellowship, UNL)
456 for supporting my domestic and international travel to conferences and workshops,
457 and the Institute of Agriculture and Natural Resources, who funded large portions of
458 my academic and research-related travel. I thank the School of Natural Resources

459 for their financial support in my conference travel. The U.S. National Academy
460 of Sciences generously funded part of my travel to the International Institute for
461 Applied Systems Analysis (IIASA). This financial support provided me not only with
462 invaluable opportunities to attend and present at national and international conferences
463 and workshops, conduct research abroad, and network–this funding alleviated some
464 financial pressures associated with graduate school which allowed a more refined focus
465 on my dissertation research. The opportunities and experiences provided to me by
466 each funding source were amazing, thank you.

467 Finally, to my partner of eight years–Schultzie–thank you for everything. Just kidding,
468 thank you, Nat Price, you are amazing.

⁴⁶⁹ Table of Definitions

⁴⁷⁰ Research surrounding regime shifts, threshold identification, change-point detection,
⁴⁷¹ bifurcation theory, etc. is muddled with jargon. Here, I provide a table of definitions
⁴⁷² (Table 1) for terms and concepts that may either be unfamiliar to the practical
⁴⁷³ ecologist, or may have multiple meanings among and within ecological researchers and
⁴⁷⁴ practitioners. With this table, I aim to both improve the clarity of this dissertation
⁴⁷⁵ and highlight one potential issue associated with regime detection methods in ecology:
⁴⁷⁶ semantics.

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

Term	Definition	Synonyms
Abrupt	A relative value of the speed and/or intensity of the change; the time period over which the regime shift occurs relative to the time observed (or expected to have been) in a particular state.	big, fast, quick, large
Alternative Stable State	Controversially can be distilled as one of either: the number of unique stable configurations that a system can adopt (see Lewontin 1969), or the impacts that processes or pressures can have on a system's state (see May 1977).	
Attractor	The set of values towards which a system tends regardless of its initial (starting) values.	
Basin-Boundary Collision	The parameter values for a system that causes the system to shift between alternate attractors.	non-local bifurcation
Catastrophe Theory	The study of abrupt changes within a dynamical system.	
Catastrophic Bifurcation	A relatively abrupt jump to an alternate attractor due to initial attractor.	
Change-Point	See also 'Regime Shift'. A term often used in computer science, climatology, data science; represents the point at which a state changes its configuration.	

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

Term	Definition	Synonyms
Change-Point	A change point method which does not require supervision; identifies potential change points without a priori potential change points.	
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	A system with extreme sensitivity to initial conditions.	
Critical Slowing Down (CSD)	When the recovery rate (time to return) of a system decreases (approaches zero) as a system approaches a critical point (possibly a threshold or tipping point). A characteristic observed in some empirical systems data (e.g. nutrient loading in shallow lakes).	
Degrees of Freedom	The number of system parameters or components which vary independently.	

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

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

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

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

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, $\$x_t\$$.	
Noise	Processes manifested in data which are unaccounted for; sometimes referred to as meaningless; random variability.	
Non-Stationarity of the Mean Value	Infers that a trend or a periodicity is present in the time series.	
Online	Real-time updating of model parameters, predictions, etc. (c.f. offline).	
Persistent	A relative value of the longevity of the observed change in values.	long-lasting
Phase Space	A graphical representation of two or more trajectories where one axis is not time. In this representation an equilibrium is defined as a single point in the state space.	
Prediction	A temporal forecast. Is intrinsic when a model and paramters are used to make forecast, is realized when the prediction becomes the actual state of the system.	
Pressure	A perturbation which negatively influences a system, and can be defined as pulse, press, or monotonic.	

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	A set of system values that define a particular system state. Not necessarily stable, but some state variables or outputs of the system remain relatively constant over a defined period of time.	
Regime Shift	"abrupt" and "persistent" change in a system's structure or functioning.	
Second-Order	The mean is constant and the covariance is a function of a time lag,	
Stationarity	but not of time.	
Self-Similarity	A system satisfied by power-law scaling.	
Stable Equilibrium	An equilibrium is stable when small perturbations do not induce change. <i>in the eq. system state can still vary.</i>	
State Space	The set of all possible configurations of a system.	
State-Threshold	When a gradual change in external driver induces a rapid change in ecosystem state (e.g., System crosses a threshold).	
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).	
Supervised Machine Learning	When classifiers are used to train the data a priori.	
System State	The observed (current) instance of the system within a state space.	
Threshold	A point where the system reacts to changing conditions.	
Tipping Point	A point in a system's trajectory where a small change in an endogenous force induces a large change in system state or values; the point where a system can flip into an alternative state.	
Trajectory	The path of an object or system through space-time.	orbit, path
Transient	A behavior or phenomenon which is responsive to initial (starting) conditions, or its effect declines over time.	
Trend Smoothing	Local averaging of values such that the non-systematic components of the system are washed out.	

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

⁴⁷⁷ Chapter 1

⁴⁷⁸ Introduction

⁴⁷⁹ Anthropogenic activity in the last few decades will continue to influence the interactions
⁴⁸⁰ within and among ecological systems worldwide. The complexity and drivers of changes
⁴⁸¹ in coupled human-natural systems is consequently altered, further limiting our ability
⁴⁸² to detect and predict change and impacts of change (Liu *et al.*, 2007; Scheffer, 2009).
⁴⁸³ Early warning systems are developed to detect, and in some cases predict, abrupt
⁴⁸⁴ changes in disparate systems (e.g. cyber security Kaufmann *et al.*, 2015; banking
⁴⁸⁵ and stock markets Davis & Karim, 2008). The need to develop and improve early
⁴⁸⁶ warning systems for natural and coupled human-natural systems is exacerbated by the
⁴⁸⁷ consequences of climate change and globalization, especially when the human-related
⁴⁸⁸ stakes are high. The ecological literature is inundated with quantitative methods and
⁴⁸⁹ models with the promise of predicting abrupt change in high-dimensional ecological
⁴⁹⁰ systems in time for intervention. The paucity of application of many of these methods
⁴⁹¹ by practitioners and decision makers suggests much work is to be done in advancing
⁴⁹² both our understanding of abrupt ecological change and of the methods used for
⁴⁹³ detecting it.

⁴⁹⁴ Forecasting undesirable change is, arguably, the holy grail of ecology. Paired with

495 an understanding of system interactions, a forecast is ideal if it provides reliable
496 forecasts with sufficient time to prevent or mitigate unwanted systemic change. Early
497 warning systems (or early warning signals, or early warning indicators) have been
498 developed and tested for some ecological systems data, but have been mostly applied
499 to marine fisheries time series and nutrient loadings in shallow lakes. Despite the
500 numerous quantitative methods [see Chapter 2] proposed as early warning systems for
501 ecological data, many are currently of limited practical utility. This paradox may be
502 a consequence of existing early warning systems having one or more of the following
503 characteristics:

- 504 1. Not generalizable across systems or system types (especially when it requires a
505 model or a deterministic function to describe the system)
- 506 2. Requires a large number of observations
- 507 3. Difficult to implement
- 508 4. Difficult or to interpret
- 509 5. Requires an understanding of the drivers of change
- 510 6. Performs poorly under uncertainty and in presence of noise
- 511 7. Gives no uncertainty around estimates (tying into interpretation issues)
- 512 8. Ignores or does not sufficiently account for observation error
- 513 9. Currently no baseline with which to compare results
- 514 10. Currently no application/testing on empirical systems data
- 515 11. Systems are subjectively bounded (i.e., components are chosen)
- 516 12. Being overshadowed by semantics
- 517 13. Are based on before-and-after information
- 518 14. Cannot link the shift to potential drivers (i.e. the method reduces the dimen-
519 sionality such that it is unitless and/or loses all relevant information)
- 520 15. Cannot handle irregular sampling
- 521 16. Cannot handle non-smooth or non-linear data

522 Research focusing on the above areas as they relate to regime detection measures will
523 can only advance and improve of these early warning systems, further shedding light
524 on the potential of these methods for application to place-based observations.

525 1.1 Dissertation aims

526 The overarching aim of this dissertation is to contribute to our understanding of
527 the utility and limitations of early warning systems for ecological regime detection,
528 which I will refer to as ‘regime detection measures’. Regime detection measures exist
529 for handling both univariable and multi-variable data, however, it is the latter of
530 these methods within which this dissertation focuses. Although the univariate regime
531 detection measures are currently more widely applied and conceptually tractable than
532 many multivariate regime detection measures, the utility of the univariable measures
533 may be limited when change(s) in the system dynamics manifest in entire community
534 dynamics, rather than in select indicator species, for example. Multivariable regime
535 detection measures may also be more advantageous than analysing individual variables
536 when the drivers of the observed systemic change are unknown. Further, ecological
537 systems are noisy, and ecological systems data are messy — conditions difficult to
538 handle using univariable regime detection measures.

539 1.2 Dissertation structure

540 1.2.1 Chapter highlights and chronology

541 This dissertation comprises a glossary (section), eight distinct Chapters (Chapters 1-8),
542 and compendium of open-source statistical software authored during the production

543 of this dissertation in the form of two appendices (Appendices 8.4- .1.3). Finally, the
544 dissertation is synthesized in Chapter 8.

545 **Section.** The terminology associated with this line of research is highly variable
546 both within and outside the field of ecology. For example, although many core
547 concepts informing ecological regime shift theory stem from dynamical systems theory,
548 the terminologies do not align (Andersen *et al.*, 2009; Hastings & Wysham, 2010).
549 Therefore to ensure clarity of discussion throughout this work, I provide a glossary of
550 phrases and concepts related to this dissertation in a front-matter glossary (section
551).

552 **Chapter 2.** There exists a staggering number of quantitative methods for identifying
553 abrupt changes and regime shifts in ecological systems data. Despite the high number
554 of methods proposed in the literature, few have been scrutinized against empirical
555 data, and even fewer applied to multiple types of systems (e.g., terrestrial mammals
556 vs. marine fisheries). Although numerous reviews of these methods are published, few
557 are comprehensive in their presentation of the proposed methods and metrics. In this
558 Chapter 2 I conducted a formal literature review to both present a comprehensive
559 source for identifying regime detection methods.

560 **Chapter 3.** Fisher Information is proposed as method for identifying regime shifts in
561 multivariable ecological time series and spatially-explicit data and has been applied
562 to a variety of systems across at least 20 publications. Two forms of this measure
563 exist, one of which (the ‘derivaitves-based’ method) requires fewer steps and *a priori*
564 defined parameters to calculate than the other (the ‘binning method’). Chapter 3
565 contributes to the understanding of the ‘derivatives-based’ Fisher Information as a
566 regime detection measure in two ways. First, I present a step-by-step overview of both
567 the logistics and concepts required for calculating this measure. Next, I suggest that
568 the current calcualtion of Fisher Information can be split into two distinct parts: a

569 dimensionality reduction and the actual calculation of the Fisher Information. This
570 results of this study has implications for how the method can be used in the future, and
571 whether it will suffice as an indicator of abrupt change under certain conditions.

572 **Chapter 4.** The Fisher Information method has been applied both temporal (Cabezas
573 *et al.*, 2010) and spatial empirical data (Sundstrom *et al.*, 2017; Eason *et al.*, 2019). To
574 demonstrate the utility of Fisher Information in identifying abrupt change in ecological
575 communities at large spatial scales, I present an application of Fisher Information to
576 spatially-explicit avian community data in North America.

577 **Chapter 5.** Building off of the method described in Chapter 3 I suggest a method
578 which I refer to as ‘velocity’ method for identifying ecological regime shifts. Previous
579 use of the velocity metric has been embedded within larger calculations of Fisher
580 Information, specifically in the ‘derivatives-based’ method (see Chapter 3). The
581 velocity method is an overlooked, simple calculation that may be useful in identifying
582 abrupt changes in high dimensional temporal or spatial series. Here, I thoroughly
583 describe the calculations behind the velocity metric, and demonstrate its utility
584 through application to both simulated and empirical systems data.

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

593 **Chapter 7.** This chapter presents an application of body mass discontinuity analysis
594 to avian community time series before and after a landscape-scale regime shift (Roberts

595 *et al.*, 2019). In this chapter, I test the hypothesis that declining and other species
596 sensitive to grassland habitat loss and degradation operate the edges of body mass
597 aggregations. Although I find evidence suggesting declining and sensitive (grassland
598 obligate) species operate near the edges of body mass distributions as opposed to other
599 species, I was unable to identify clear patterns in the avian community body mass
600 distributions using a before-and-after design with respect to the proposed spatial
601 regime (Roberts *et al.*, 2019). Appendix 7 contains an annotated version of the code
602 for the functions used to identify discontinuities in avian body mass distributions
603 (originally published in Barichievy *et al.*, 2018).

604 1.2.2 Software Appendices

605 Appendices 8.4 and .1.3 contain manuals for two, self-authored softwares used throughout
606 this dissertation. They are available for download at www.github.com/trashbirdecology.

607 R package: `regimeDetectionMeasures`

608 Appendix 8.4 contains code for calculating a suite of regime detection measures,
609 including the traditional early-warning indicators, Fisher Information, and the velocity
610 metric.

611 R package: `bbsRDM`

612 Appendix .1.3 contains functions for conducting spatial and temporal analyses of
613 the regime detection measures contained in package `regimeDetectionMeasures` (Ap-
614 pendix 8.4), and is closely aligned with the analyses in Chapters 4 and 6. Further,
615 this package provides a wrapper for downloading and munging data from the U.S.
616 Geological Survey's Breeding Bird Survey, munging said data, and creating spatial

⁶¹⁷ sampling grids across North America. Minor additional functionality includes an
⁶¹⁸ option to download and identify U.S. military bases across the globe.

619 **Chapter 2**

620 **A Brief Overview of the Ecological
621 Regime Detection Literature**

622 **2.1 Introduction**

623 Long-lasting changes in the underlying structure or functioning of natural systems due
624 to exogenous forcings (also called regime shifts) are increasingly recognized in ecology.
625 The ability to identify and predict these shifts is particularly useful for systems which
626 are actively managed, provide ecosystem services, and/or benefit society. Despite
627 the potential utility of quantitative methods for detecting ecological regime shifts or
628 abrupt changes, there exists a disparity among the number of methods proposed for
629 detecting abrupt changes in ecological, oceanographic, and climatological systems
630 and the studies evaluating these methods using empirical data (Hawkins *et al.*, 2015).
631 Further, new methods continue to permeate the literature. Although reviews of regime
632 detection measures exist and some are recent (Ducré-Robitaille *et al.*, 2003; Mantua,
633 2004; Rodionov, 2005; deYoung *et al.*, 2008; Andersen *et al.*, 2009; Boettiger *et al.*,
634 2013; Kefi *et al.*, 2014; Mac Nally *et al.*, 2014; Dakos *et al.*, 2015; Scheffer *et al.*,

635 2015; Filatova *et al.*, 2016; Litzow & Hunsicker, 2016; Yin *et al.*, 2017; Clements &
636 Ozgul, 2018; Roberts *et al.*, 2018), few are akin to the 20-year old comprehensive
637 presentation of Rodionov (2005). However, the rapid increase in methods in the
638 literature is increasingly outdated the contribution of Rodionov (2005).

639 There is currently no single source to which the uninformed ecologist can refer for
640 identifying potential regime detection measures. Existing reviews of this literature
641 vary in both the number and detail of the methods presented. For example, the focus
642 of Andersen *et al.* (2009) is on the current state and direction of regime shift theory
643 and detection, consequently not containing a comprehensive presentation of many
644 methods. Roberts *et al.* (2018) come close to presenting a comprehensive review,
645 however, the focus here is on the data required to use methods and implications for
646 rangeland ecology and management.

647 Building a comprehensive database containing published regime detection measures
648 based on formal literature review is not a simple feat for a few reasons, and this has
649 implications for the application of methods by naive First, the terminology associated
650 with regime shift theory and detection is highly variable within and among fields (An-
651 dersen *et al.*, 2009). For example, the terms, *regime shifts*, *regime changes* and *tipping*
652 *points* are variably used in studies of ecological systems, whereas *inhomogeneities* is
653 common in meterology and climatology, and *structural change* is largely confined to the
654 study of economics. Although semantics vary both within and across disciplines some
655 methods are shared or are concurrently applicable across fields. Second, papers intro-
656 ducing a new method or approach to identifying regime shifts are not often proposed
657 in publication outlets with aims of disseminating new quantitative methods (e.g.,
658 *Ecological Modelling*, *Methods in Ecology and Evolution*). Rather, many new methods
659 are published in journals with refined (e.g., *Entropy*, *Progress in Oceanography*), as
660 opposed to publications with broader scopes (e.g., *Ecology* and *Nature*).

661 Regime detection measures can be classified into one of two major groups which will
662 largely dictate the applicability of an approach to an analyst or type of information.
663 Some RDMS require the use of mechanistic models whereas others fall into the
664 category of model-independent (or model-free). In most situations, the practical
665 ecologist will have insufficient data or a limited understanding of the system with
666 which to parameterize even the simplest mechanistic models. Further, developing
667 an informative defining data-generating model (i.e. system of equations, differential
668 equations) for more than a few variables (i.e. $\gg 2$) is often intractable. Following the
669 convention of Dakos *et al.* (2012a), I classify the 66 regime detection measure identified
670 as a result of this review and a based on prior knowledge as one of model-based or
671 model(metric)-free and generally synthesize the utility of these types of measures to
672 the practical ecologist.

673 **2.1.1 Aims**

674 The primary goal of this Chapter is to provide a single and comprehensive source
675 of the quantitative methods proposed for identifying ecological regime shifts, much
676 like that of Rodionov (2005). I also conduct a biographic analysis of the relevant
677 literature to identify trends in the development and current state of ecological regime
678 shift theory. In this chapter I conduct a formal literature review in an attempt to
679 generate a comprehensive list of the available regime detection measures, however,
680 this list also required prior knowledge of available methods, as no search yielded a
681 large portion of those previously known to me.

682 2.2 Methods

683 First, I conducted a systematic literature review to identify candidate articles, or
684 those which relate to ecological regime shift theory, using these results in an attempt
685 to identify papers introducing a new or novel method to regime shift detection. These
686 methods were supplemental with prior and expert knowledge¹. I then conducted an
687 independent but similar literature review to identify trends in the development and
688 persistence of the ecological regime shift literature. These steps are described in detail
689 below.

690 **2.2.1 Identifying candidate articles and regime detection
691 measures**

692 I attempted to compose a comprehensive list of the published quantitative regime
693 detection measures by conducting a systematic literature review. Despite multiple
694 attempts at defining booleans to yield identify the greatest number of methods in the
695 literature, no single boolean yielded the majority of methods of which I was previously
696 aware. In light of this, I also relied on my prior knowledge of available regime detection
697 measures to build the database.

698 I first queried the Thomson- ISI Web of Science (WoS) database (on 06 March 2019)
699 to identify articles which mention terms related to regime shifts, or abrupt changes,
700 using the following boolean:

701 TS=((“regime shift” OR “regime shifts” OR “regime change” OR “regime
702 changes” OR “catastrophic change” OR “catastrophic shift” OR “cata-

¹This includes methods in previous reviews, including Scheffer *et al.* (2015); Rodionov (2005); Roberts *et al.* (2018); 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)

trophic changes” OR “catastrophic shifts” OR “sudden change” OR “sudden changes” OR “abrupt shift” OR “abrupt shifts” OR “abrupt change” OR “abrupt changes” OR bistab* OR threshol* OR hystere* OR “phase shift” OR “phase shifts” OR “phase change” OR “phase changes” OR “step change” OR “step changes” OR “stepped change” OR “stepped changes” OR “tipping point” OR “tipping points” OR “stable states” OR “stable state” OR “state change” OR “state changes” OR “stark shift” OR “stark change” OR “stark shifts” OR “stark changes” “structural change” OR “structural changes” OR “change-point” OR “change point” OR “change-points” OR “change point” OR “break point” OR “break points” OR “observational inhomogeneity” OR “observational inhomogeneities”) AND (“new method” OR “new approach” OR “novel method” OR “novel approach”))

where ‘*’ indicates a wildcard. A pilot study revealed that by limiting the search to the Web of Science Categories *Ecology* and *Biodiversity Conservation* [i.e. by adding boolean term ‘AND WC=(Ecology OR ’Biodiversity Conservation’)] excluded many methods that are peripheral to this study. For example, numerous univariate measures are used solely in climatology, physics, and data science/computer science literatures, where change-point analyses are abundant. Although additional methods could be identified by searching these fields, this dissertation focuses on using methods for analysing *multivariable* data. Consequently, many methods for analysing abrupt breaks in a single variable do not appear in the comprehensive list. Next, I identified articles which propose a ‘new’ method for identifying ecological regime shifts by searching within the results of the previous boolean. I restricted the search by appending the following boolean phrase required to appear in the title and/or abstract:

AND (‘new method’ OR ‘novel method’ OR ‘new approach’ OR ‘new

729 practical method' OR 'new simple method' OR 'new multivariate' OR
730 'new tool' OR 'novel tool' OR 'novel multivarte' OR 'novel approach' OR
731 'new numerical' OR 'novel numerical' OR 'new quantitative' OR 'novel
732 quantitative' OR 'i introduce' OR 'we introduce')

733 There appeared disparity among the number of methods of which I was previously
734 aware and those identified in an initial Web of Science review. In an attempt to identify
735 as many new methods as possible I conducted an informal search of the Google Scholar
736 database, a database notoriously broader in scope than other academic dataabses.
737 The length of boolean for the Google Scholar database is limited by the number of
738 characters. Unfortunately, this, coupled with the wide breadth of Google Scholar's
739 search boundaries, limits the capacity to which Google Scholar can be used to refine
740 the literature to a manageable number of articles. For these reasons I arbitrarily
741 skimmed the titles of the first 25 pages of the Google Scholar results (25 pages = 250
742 articles). It should be noted that the order of terms appearing in the boolean are
743 regarded as the order of desired relevancy. I used the following boolean to identify
744 these articles in Google Scholar:

745 ('regime shift' OR 'regime change' OR 'tipping point') AND ('new method'
746 OR 'new approach' OR 'novel method' OR 'novel approach')

747 The candidate articles identified by Web of Science and Google Scholar contained
748 numerous articles proposing a new framework for identifying regime shifts rather
749 than a new quantitative method. I exluded these from consideration as a 'new
750 method' by removing them, and also removing articles proposing a novel application
751 of a combination of previously-used methods (see Kong *et al.*, 2017; Seddon *et al.*,
752 2014; Vasilakopoulos *et al.*, 2017). I also excluded papers which made relatively
753 minor adjustments or updates to existing methods (Zhou & Shumway, 2008; but
754 see Nicholls *et al.*, 2011 for an addition of variable optimization to the method in

755 @nicholls_detection_2011 that was not included in the results) or articles proposing
756 new methodologies in mathematical journals (Salehpour *et al.*, 2011; Byrski & Byrski,
757 2016) that have yet to be associated with or tested on ecological data, or suggested to
758 be useful for empirical data.

759 **Bibliographic analysis of ecological regime shift literature**

760 The still-vague definition of ecological regime shifts has led to a breadth of articles
761 exploring this phenomenon. I again used the Web of Science database to identify
762 patterns in the development and persistence of the ecological regime shift literature. I
763 conduct a systematic literature review using ISI Web of Science, and use these results
764 to conduct exploratory bibliographic analyses. This literature search differs from the
765 search conducted in Section 2.2.1 in that it was conducted to capture the overall
766 development and breadth of the field of ecological regime shifts, rather than capturing
767 the *quantitative methods* for identifying shifts. I used the following boolean to identify
768 articles related to regime shift and abrupt change, and restricted this search to the
769 ecological literature by including specifying which Web of Science categories *Ecology*
770 and *Biodiversity Conservation* to search:

771 TS=(“regime shift” OR “regime shifts” OR “regime change” OR “regime
772 changes” OR “catastrophic change” OR “catastrophic shift” OR “cata-
773 strophic changes” OR “catastrophic shifts” OR “sudden change” OR “sud-
774 den changes” OR “abrupt shift” OR “abrupt shifts” OR “abrupt change”
775 OR “abrupt changes”) AND WC=(“Ecology” OR “Biodiversity Conserva-
776 tion”)

777 To explore the patterns and trends in the development of this field, I used clustering
778 algorithms available in the R Package **bibliometrix** (Aria & Cuccurullo, 2017).
779 Among other things, this package contains function wrappers for conducting and

780 visualizing network analyses based on keyword, authorship, and citation data. In an
781 attempt to understand the evolution of regime shift theory and relate this evolution
782 to the quantitative methods in the ecological literature, I focus analyses on using
783 keyword and concept themes rather than citation counts and author dominance.

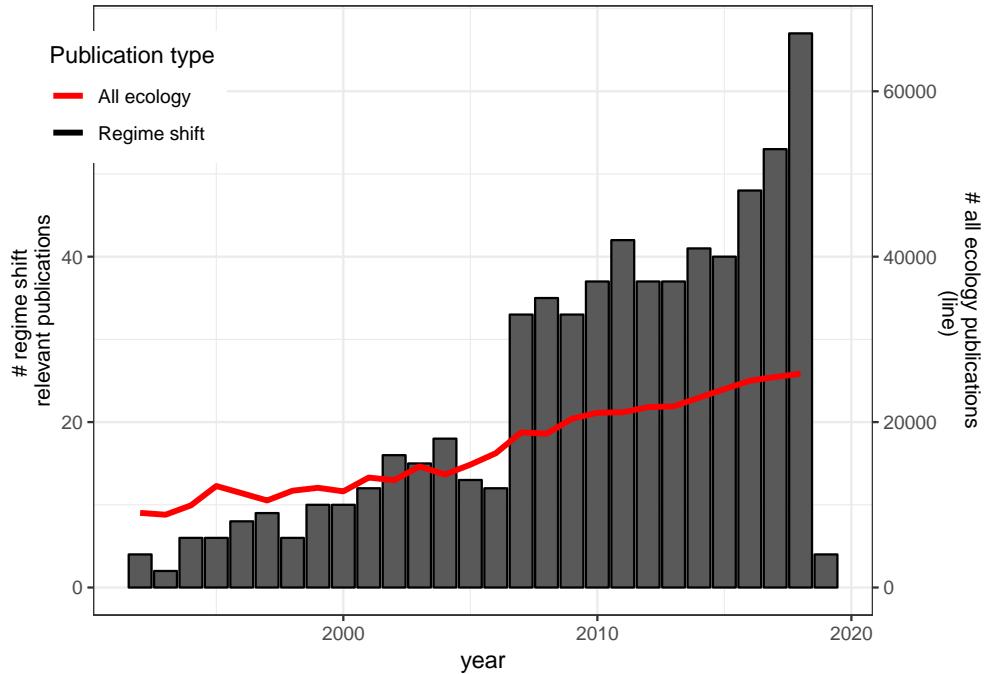


Figure 2.1: Number of publications by year in fields 'Ecology' and 'Biodiversity Conservation' which included terms related to 'regime shift' (total = 654).

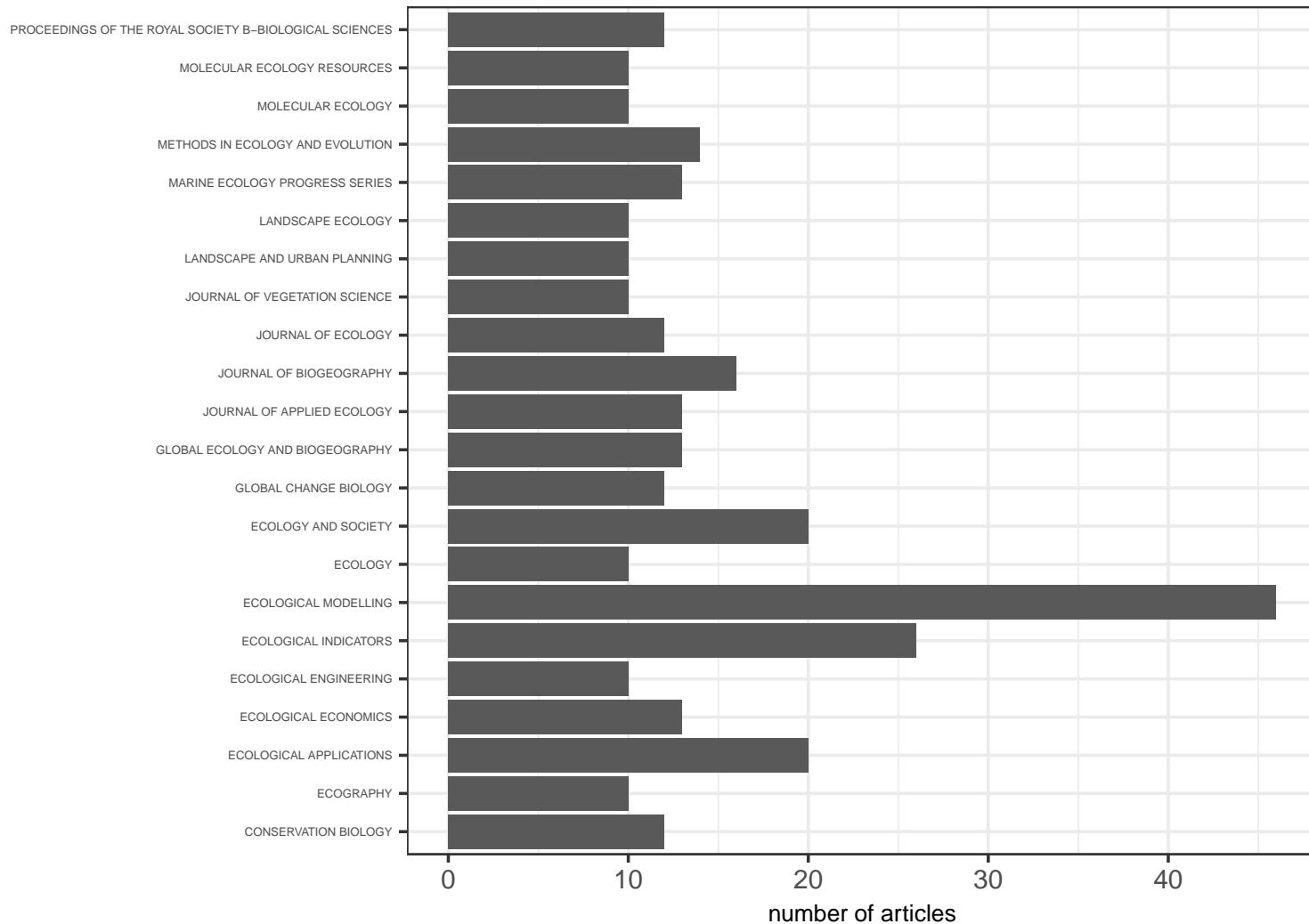


Figure 2.2: Number of articles containing phrases related to ecological 'regime shifts' published per publication outlet.

784 2.3 Results**785 2.3.1 Quantitative methods for identifying ecological regime
786 shifts**

787 I restricted the ISI Web of Science to the Web of Science categories (WC) ‘Ecology’
788 and ‘Conservation Biology’ as the original search yielded > 20,000 results. This
789 filtered the results down to a manageable amount of papers to review. Of the 2,776
790 articles, 654 included the terms relating to ‘regime shifts’ [see Section 2.2.1; Figure 2.1],
791 many of which were published in a predominately methodological journal, *Ecological
792 Modelling* (Figure 2.2). The rapid rate of change publication of articles containing
793 terms related to ‘regime shifts’ does not correspond with the linear change in the total
794 number of ecological articles published over time (Figure 2.1).

Table 2.1: List of the regime detection methods identified in this review.

Method	Metric type	Source
Characteristic length scale (CLS) estimation	attractor reconstruction	\@NA
Average standard deviates	metric	\@ebbesmeyer19911976
BDS test	metric	\@carpenterBrock2011early
Coefficient of variation (CV)	metric	\@NA
Conditional heteroskedasticity	metric	\@seekell2011conditional
Cumulative deviation test (CUSUM)	metric	\@buishand1982some
Degenerate Fingerprinting	metric	\@held2004detection
Degenerate Fingerprinting	metric	\@livina2007modified
Downton-Katz test	metric	\@karl1987approach
First-order multivariate autoregressive models (MAR1)	metric	\@ives2003estimating
Fisher Information	metric	\@fath_regime_2003
Intervention Analysis	metric	\@francis1994decadal

Table 2.1: List of the regime detection methods identified in this review. (continued)

Method	Metric type	Source
Inverse of AR(1) coefficient, variance, etc.	metric	\@carpenter2008leading
Kurtosis	metric	\@biggs2009turning
Lanzante method	metric	\@lanzante1996resistant
LePage test	metric	\@yonetani1993detection
Mann-Kendall test	metric	\@goossens1987recognize
Mann-whitney U-test	metric	\@mauguet2003multidecadal
Moving detrended fluctuation analysis (MDFA)	metric	\@he2008new
Nearest-neighbor statistics	metric	\@pawlowski_identification_2008
Nikiforiv method	metric	\@NA
Oerleman's method	metric	\@oerlemans1978objective
Pettitt test	metric	\@pettitt1979non
Probability density function entropy method	metric	\@pawlowski_identification_2008

Table 2.1: List of the regime detection methods identified in this review. (continued)

Method	Metric type	Source
Quickest detection method (Shiryayevδ_0>Roberts statistic)	metric	\@moustakides2009numerical
Rodionov method	metric	\@rodionov_sequential_2005
STARS	metric	\@buishand1982some
Sequential t-tests	metric	\@rodionov2004sequential
Signal-to-noise ratio	metric	\@NA
Skewness	metric	\@guttal2008changing
Spectral density ratio indicator	metric	\@biggs2009turning
Spectrum indicator	metric	\@NA
Stability Index of the Ecological Units	metric	\@parparov2015quantifying
Standard deviation (rising variance)	metric	\@carpenter2006rising
Standard normal homogeneity	metric	\@alexandersson1986homogeneity
T-test	metric	\@ducre2003comparison

Table 2.1: List of the regime detection methods identified in this review. (continued)

Method	Metric type	Source
Threshold Indicator Taxa ANalysis (TITAN)	metric	\@baker2010new
Variance Index	metric	\@brock_variance_2006
Wilcoxon rank-sum	metric	\@karl1987approach
dimension reduction techniques (e.g., PCA)	metric	\@NA
two-phase regression	metric of a model	\@easterling1995new
Zonal thresholding	metric*	\@yin2017methods
Bayesian approaches	model	\@jo2016bayesian
Convex model	model	\@qi2016resilience
Free-knot splines & piecewise linear modelling	model	\@gal2010novel
Generalized model	model	\@lade2012early
Multivariable autoregressive models (MAR1)	model	\@ives2012detecting
Nonparametric drift-diffusion-jump model	model	\@carpenter2011early
Potential analysis	model	\@ives2012detecting

Table 2.1: List of the regime detection methods identified in this review. (continued)

Method	Metric type	Source
Regression-based models	model	\@solow1987testing
Self-exciting threshold autoregressive state-space model SETARSS(p)	model	\@tong1990nonlinear
Smooth transition autoregressive model shiftogram	model	\@see gal2010novel
Pettitt test and the Sen test	model-based	\@vincent1998technique
Online dynamic linear modelling + time_varying autoregressive state_space models (TVARSS)	models	\@parparov2017quantifying
Clustering, various	other	\@NA
Fourier Analysis	other	\@carpenter2010early
Vector-autoregressive method	other	\@solow_test_2005
Wavelet analysis (decomposition)	other	\@cazelles2008wavelet
Degenerate Fingerprinting	NA	\@kleinen2003potential

Table 2.1: List of the regime detection methods identified in this review. (continued)

Method	Metric type	Source
MCMC	NA	\@NA
method-fuzzy synthetic evaluation (FSE)	NA	\@wang2011application

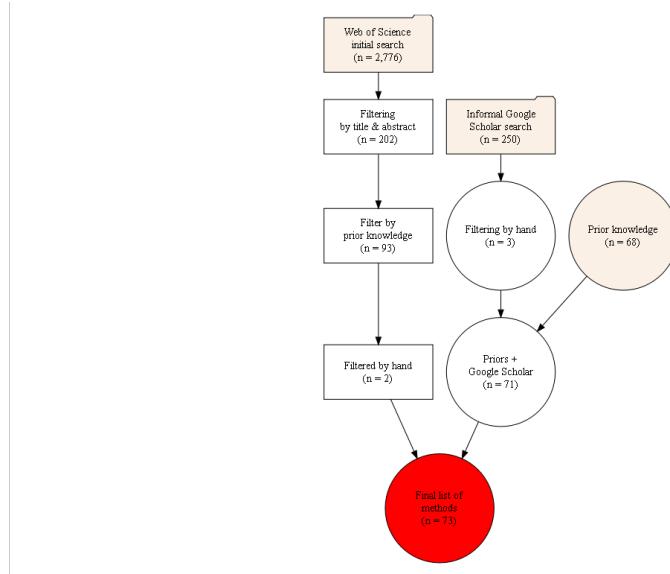


Figure 2.3: Flowchart of the literature review process for identifying new regime detection methods. Note: only the first ten pages (250 articles) of Google Scholar results were examined. Node shapes: folder = unfiltered articles; box = articles actively filtered; diamond = number of articles with new methods.

795 Using my prior knowledge of the relevant literature and by systematically searching the
 796 Web of Science and Google Scholar databases, I identified 60 unique regime detection
 797 measures [Figure 2.3; Table 2.1]. A total of 202 of the 2,776 articles identified by
 798 Web of Science contained terms in the title, keywords, or abstract related to a ‘new
 799 method’ (see Section 2.2.1). Based on my prior knowledge of the literature and based
 800 on reviews of prior regime detection and regime shift review papers, I was previously
 801 aware of 68 articles containing ‘new’ methods (Figure 2.3). Approximately half were
 802 successfully identified in the Web of Science and Google Scholar systematic review.
 803 After removing the papers containing the methods of which I was already aware prior
 804 knowledge, only 93 articles remained to be reviewed ‘by hand’. A review of these 93
 805 articles resulted in the identification of 5 new methods [2 from Web of Science and 3
 806 from Google Scholar; Figure 2.3].

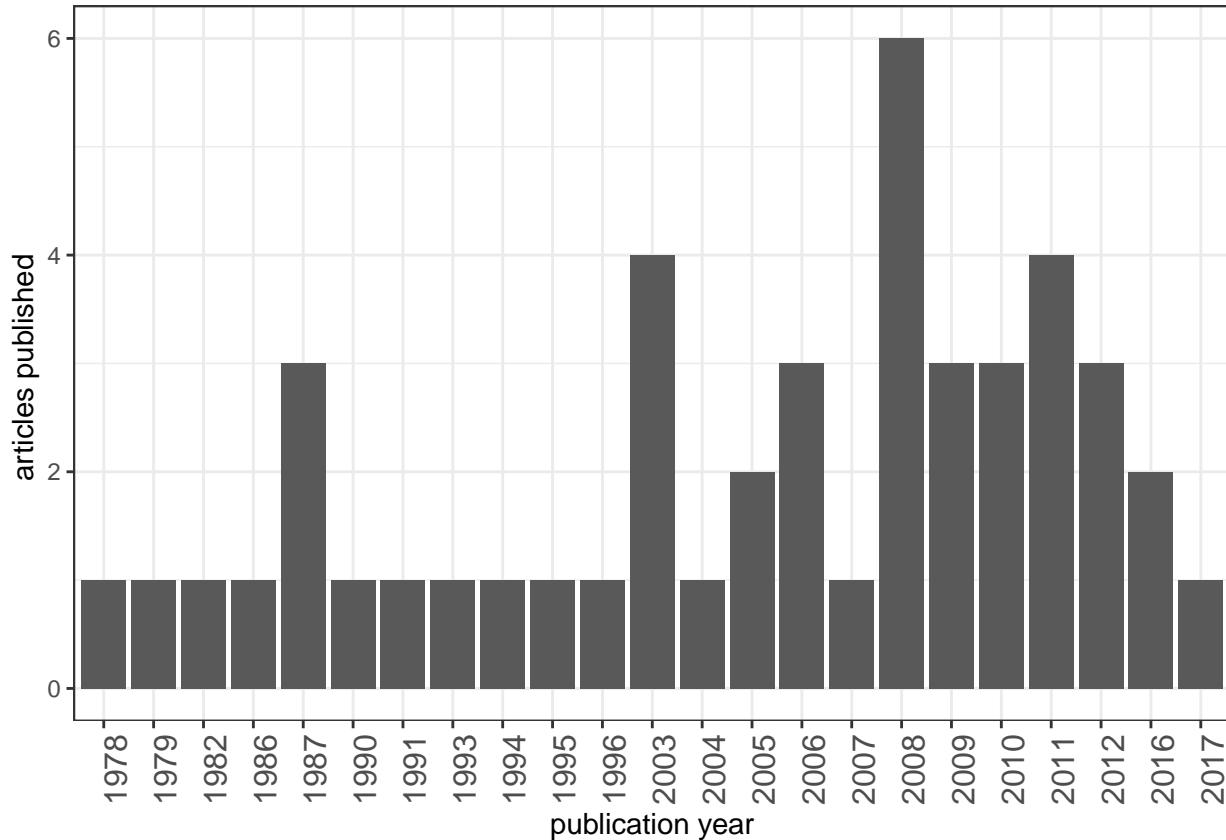


Figure 2.4: Number of methods published over time.

⁸⁰⁷ **2.3.2 Bibliographic analysis of ecological regime shift litera-**
⁸⁰⁸ **ture**

⁸⁰⁹ A search of Web of Science for articles in Ecology and Biodiversity Conservation con-
⁸¹⁰ taining phrases related to ‘regime shifts’ yielded 1,636 original articles. These articles
⁸¹¹ were not filtered in any fashion and as such all were considered in the bibliographic
⁸¹² analysis. I used the clustering algorithms of the R package **bibliometRICS** to produce
⁸¹³ thematic maps, which use clustering algorithms to identify clusters (or themes) based
⁸¹⁴ on bibliography features [e.g., keywords, **authors**; @cobo2011approach].

815 Thematic mapping using keywords

816 Two types of keywords exist in the ISI Web of Science bibliography metadata: those
817 provided to the publication by the authors (author-supplied), and those defined by the
818 ISI Web of Science (ISI-supplied). A keyword thematic map suggests these keywords are
819 used very differently within this literature (Figure @ref(fig:thematicMaps_keyword))².
820 The clustering algorithm identified fewer clusters (themes) in the ISI-keywords (Figure
821 @ref(fig:thematicMaps_keyword)a) than were identified among the author-supplied
822 keywords (Figure @ref(fig:thematicMaps_keyword)b); see footnote ²). This pat-
823 tern is not surprising given the ISI-supplied keywords are restricted to pre-set
824 number of keywords, whereas authors can and do provide synonyms, or words
825 not used at all by ISI. The themes identified in the ISI-supplied keyword analysis
826 were relatively consistent as the number of keywords analysed increased (Figure
827 @ref(fig:thematicMaps_keyword_isi)), but the themes varied drastically among the
828 author-supplied keywords (Figure @ref(fig:thematicMaps_keyword_author)). For this
829 reason I make inference on only the ISI-supplied keyword cluster analysis.

830 Four major themes were identified in the ISI keyword analysis and, interestingly,
831 fell mostly within the quadrants representing the most extreme values: the first
832 and the third quadrants (Figure @ref(fig:thematicMaps_keyword_isi)). The themes
833 identified by the ISI-supplied keywords were much larger in scope (e.g., dynamics,
834 ecosystems, climate; Figure @ref(fig:thematicMaps_keyword)a) than those identified
835 in the analysis of author-supplied keywords (e.g., eutrophication, trophic cascade;
836 Figure @ref(fig:thematicMaps_keyword)b). That is, the themes of ‘regime shifts’

²Axes 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).

and ‘ecosystem dynamics’ are highly central to and dense within the regime shift literature (Figure @ref(fig:thematicMaps_keyword)b-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 dynamics (i.e. the study of non-linearity and stochasticity in ecological systems) plays a central role in the theory of ecological systems theory, the theme of dynamics is not reflected in the many case studies of regime shifts in application (Litzow & Hunsicker, 2016). In fact, Litzow & Hunsicker (2016) found that ~ 50 of case studies actually tested or accounted for non-linear dynamics when applying early warning indicators and other regime shift measures to ecological time series.

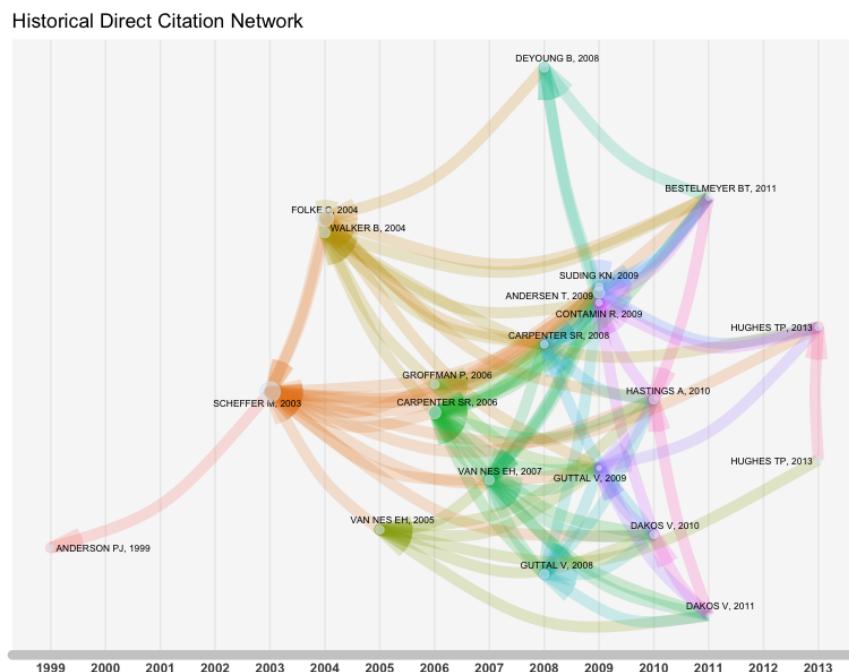


Figure 2.5: Chronological direct citation network suggests the intellectual structure can be mapped to a few papers. This historiograph identifies important works explicitly in chronological, as opposed to absolute, order.

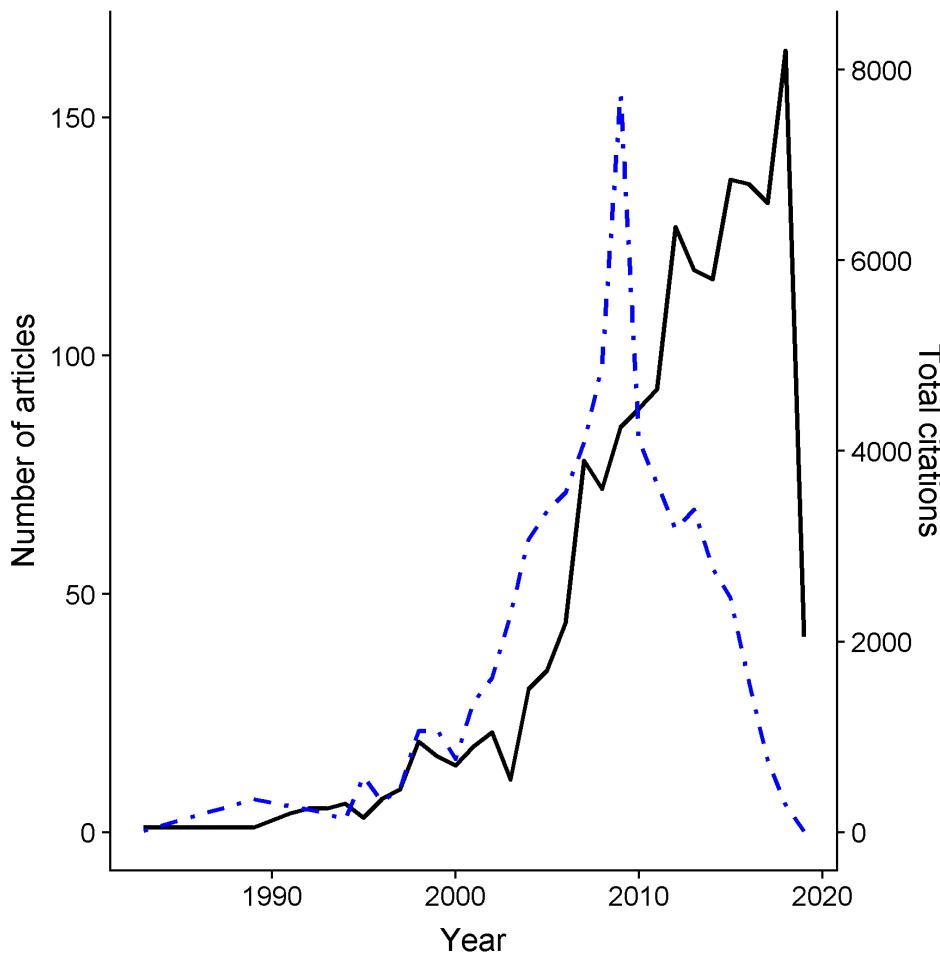


Figure 2.6: Total number of articles published and corresponding number of citations (for papers published that year). The most highly cited papers to-date are those published in the late 2000s.

⁸⁴⁷ **Historiograph, citation, and reviews**

⁸⁴⁸ A few patterns appear in analyses of the intellectual and chronological structure
⁸⁴⁹ of the ecological regime shift literature (Figure 2.5). First, although the concept of
⁸⁵⁰ stability, thresholds, and multiple stable states in ecological systems first appeared
⁸⁵¹ (and was well-received) in the literature in the 1970s (e.g., Holling, 1973; May, 1977),
⁸⁵² the most important papers in this field appeared primarily in the early and mid 2000s
⁸⁵³ (Scheffer & Carpenter, 2003; Folke *et al.*, 2004; Walker *et al.*, 2004; Nes & Scheffer,

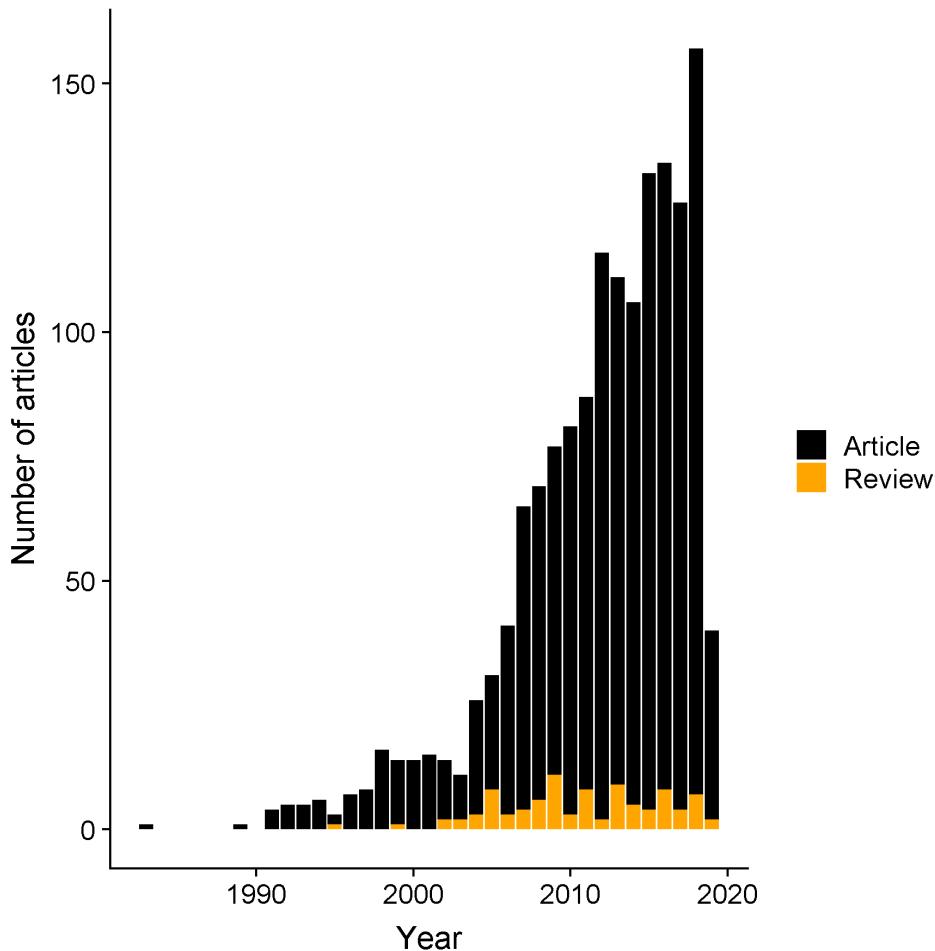


Figure 2.7: Total number of articles published per year by category as categorized by ISI. Book chapters, proceedings, editorials, and letters are excluded.

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 a spur of critique of regime shift detection methods (Andersen *et al.*, 2009; Contamin & Ellison, 2009) are highlighted in the historiograph. The most influential papers in the field, based solely on number of citations, are those published in the late 2000s (Figure 2.6), and include the articles which are broad in-scope and are still used today

863 to frame studies in the context of global change, planetary boundaries, and large-scale
864 tipping points (Bennett *et al.*, 2009; Rockström *et al.*, 2009; Smith & Schindler, 2009).
865 Around this time (\sim 2007) is when the number of regime detection measures doubled
866 (Figure 2.1). Papers equally influential to the development of this field include those
867 corresponding to the observed rapid increase in the number of total publications in
868 ecological regime shift theory (in the early 2000s), Folke *et al.* (2004) and Scheffer &
869 Carpenter (2003) (Figure 2.6).

870 Numerous reviews of the regime shift literature exist, ranging from conceptual reviews
871 of the state of regime shift theory in ecology and application (e.g., Bestelmeyer *et al.*,
872 2011; Andersen *et al.*, 2009; Mac Nally *et al.*, 2014), to studies of robustness of early
873 warning indicators under various theoretical and practical conditions [e.g., Dutta *et*
874 *al.* (2018); Perretti & Munch (2012); Lindegren *et al.* (2012); Hastings & Wysham
875 (2010); Figure 2.7]. Further, comprehensive reviews of the ecological regime shift
876 detection literature are increasingly out-dated. A permanent and open-source database
877 containing information critical to the testing, comparison, and implementation of
878 RDMs may prove useful to the reader who is interested in applying RDMs but is
879 lacking the statistical or mathematical background to do so.

880 **2.4 A synthesis of the methods available for the**
881 **practical ecologist**

882 Many of the methods identified in this review have yet to be tested on multiple, em-
883 pirical data (see Table 2.1). I categorize the regime detection methods as one of either
884 model-free or model-dependent. Model-free and model-dependent methods are those
885 which do and do not require a mechanistic model to describe the system, respectively.
886 Because many of the model-dependent methods are based on autoregressive modelling

887 apporaches, this is highlighted in the model-dependent section.

888 The early warning indicators that are often referred to as, “traditional early warning
889 indicators” (variance, skewness, autocorrelation at lag-1) are fairly well-reviewed, and
890 have been tested under a variety of conditions (Ditlevsen & Johnsen, 2010; Boettiger
891 & Hastings, 2012; Dakos *et al.*, 2012a; Lindegren *et al.*, 2012; Perretti & Munch, 2012;
892 Litzow & Hunsicker, 2016; Sommer *et al.*, 2017; Dutta *et al.*, 2018). However, many
893 of these works apply the traditional (and other) early warning indicators to simulated
894 data, with only some (Contamin & Ellison, 2009; Perretti & Munch, 2012; Guttal *et*
895 *al.*, 2013; Dutta *et al.*, 2018) testing under varying conditions of noise and expected
896 shift types. The utility and robustness of the traditional early warning indicators is
897 not consistent across and within sytems, making them of limited utility in situations
898 where the system cannot be reliably mathematically modelled, or where we have
899 limited data [see also Ch. 6]. The authors of most reviews and comparative studies
900 of early warning indicators suggest that no early warning indicator is reliable alone,
901 or that work is needed to understand under what empirical conditions early warning
902 indicators might fail (deYoung *et al.*, 2008; Kefi *et al.*, 2014; Filatova *et al.*, 2016;
903 Clements & Ozgul, 2018).

904 **2.4.1 Model-depdendent**

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

912 shifts in freshwater lake systems (Carpenter *et al.*, 2011).

913 **2.4.2 Model-free**

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

929 **2.5 Discussion**

930 In this chapter I present a comprehensive list of the regime detection measures proposed
931 for analyzing ecological data. Although multiple reviews of regime detection measures
932 exist, they are either not comprehensive in their survey of the possible methods or
933 are increasingly out of date with respect to the number of methods proposed in the
934 literature [Rodionov (2005); Figure 2.4]. Most reviews since the comprehensive list

presented in Rodionov (2005) 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 (c.f. a single) variable, 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 (2005). 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 (2005); 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, 2011, @carpenter2006rising). Although univariate indicators may provide insight into relatively simple systems, their reliability as indicators for complex systems is less certain (Burthe *et al.*, 2016, @dutta2018robustness, @perretti2012regime, @sommer2017generic, @bestelmeyer_analysis_2011). 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 significances or probabilities. And although critiques of some RDMs exist, the rate at

which they are rigorously tested do 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. These can be addressed using, e.g., bootstrapping, simulations.

In this review I restricted articles to those implying they introduced a ‘new method’. Avoiding this potential barrier would have required I review the titles, abstracts, and bodies of over 22,000 articles (Figure 2.3). Alternatively, this may also be ameliorated by searching the relevant literature for *applications* of regime detection measures to ecological data, however, I suspect this would similarly yield a large number of articles to review. Also, only a handful of methods have been introduced to the mainstream methodological journals in ecology (e.g., *Ecological Modelling*, *Methods in Ecology and Evolution*; Figure 2.8). 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, autoregressive 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

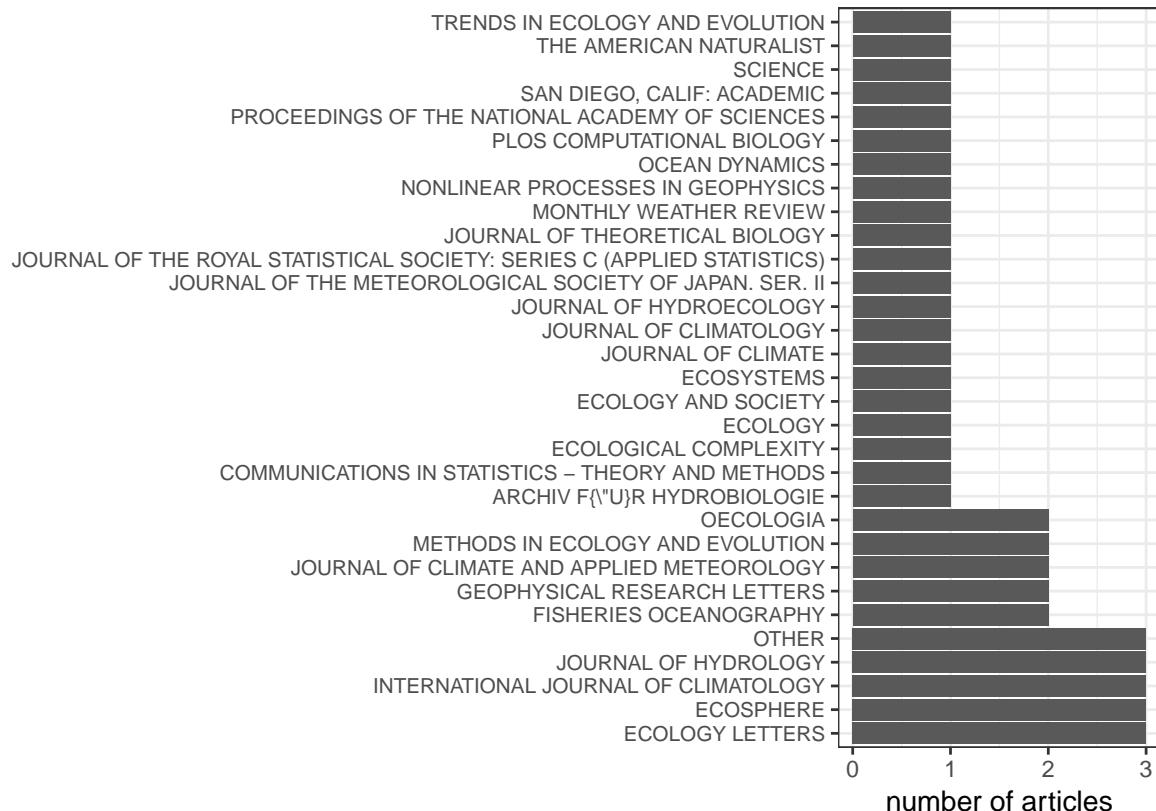


Figure 2.8: Number of articles containing new methods per journal.

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 detections may be difficult using traditional review methods. Many of the methods mentioned in this review were not identified using a systematic search process in Web of Science and Google Scholar—rather, they were methods of which I was either previously aware and/or highlighted in the few methods reviews (Mantua, 2004; Rodionov, 2005; 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) point out an important feature of using any methods for

1001 identifying regime shifts in empirical system data: we only have a single history within
1002 which we can compare AND these metrics which depend on the system exhibiting a
1003 change in variance or skewness around a mean value before and after a regime shift
1004 require the system to have a smooth potential, rather than one which can manifest
1005 complex dynamics (i.e. non-smooth potential). If we are using RDMs to attempt to
1006 forecast and prevent non-smooth or abrupt changes, then there is little justification
1007 for relying upon these early warning indicators. Specifically, these early-warning
1008 indicators may be most useful when the system is expected to undergo a transcritical
1009 or critical bifurcation before exiting a regime (Lenton, 2011). Hastings & Wysham
1010 (2010) aptly point out that any realistic ecological model should incorporate some
1011 degree of stochasticity, and when this stochasticity is introduced into the function,
1012 the function will likely not be differentiable at the point of the regime shift (Graham
1013 & Tél, 1984). In other words, most (if not all) ecological systems have non-smooth
1014 potentials, and many of the current methods for identifying regime shifts assume
1015 otherwise, often failing if the assumption is violated.

1016 To make the regime detection measures more available and transparent to the practical
1017 ecologist, I recommend: consistent use of fewer methods; persistent collection and
1018 maintenance of baseline data (reference data); an open-sourced database of methods
1019 and applications; an updated and critical review of the current state of methods in
1020 ecology (see Chapter 6) which includes methodological advancements and application
1021 failures; rigorous empirical applications of these methods (especially of those only
1022 tested on toy and experimental data); and the relationship of the RDMs used in
1023 ecology to other fields (computer science, data science, climatology and oceanography).
1024 I identify a suite of questions (Table 2.2) that would be useful in a much-needed
1025 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> <p>Can the method identify drivers?</p> <p>What assumptions does the method make about the system?</p> <p>What types of regime shifts are possible in the system?</p>
	<p>Are regime shift(s) suspected <i>*a priori*</i>?</p>

- | |
|---|
| 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? |
-

1026 **Chapter 3**

1027 **Decoupling the Calculation of**
1028 **Fisher Information**

1029 This chapter will be submitted for review with authors Burnett, J.L., N.B. Price, A.J.
1030 Tyre, D.G. Angeler, T. Eason, D. Twidwell, and C.R. Allen].

1031 **3.1 Abstract**

1032 Ecological regime shifts are increasingly prevalent in the Anthropocene. The number
1033 of methods proposed to detect these shifts are on the rise, yet few are capable
1034 detecting regime shifts without a priori knowledge of the shift, and fewer are capable
1035 of handling high-dimensional, multivariate and noisy data. A variation of Fisher
1036 Information has been proposed as a method for detecting changes in the “orderliness”
1037 of ecological systems data. Although this method is described and applied in numerous
1038 published studies, its calculation and the concepts behind its calculation are not
1039 clear. Here, I succinctly describe this calculation using a two-species predator-prey
1040 model. Importantly, I demonstrate that the actual equation for calculating Fisher

1041 Information metric comprises fewer steps than was previously described, by decoupling
1042 the dimensionality-reduction component from the actual Fisher Information calculation
1043 component. I hope this work will serve as a reference for those seeking to understand
1044 Fisher Information in the context of ecological systems and regime shifts, and will
1045 stimulate further research of the efficacy of these composite regime shift detection
1046 metrics..

1047 **3.2 Introduction**

1048 Changes in the feedback(s) governing ecosystem processes can trigger unexpected and
1049 sometimes undesirable responses in environmental conditions (Scheffer *et al.*, 2001;
1050 Walther *et al.*, 2002). Ecologists often refer to such changes as regime shifts, but
1051 this term is used interchangeably in the literature with state change, state transition,
1052 or alternative state (Andersen *et al.*, 2009). Climate change and globalization are
1053 triggering novel and unexpected changes in ecosystems (Hughes, 1994; Scheffer *et*
1054 *al.*, 2001; Walther *et al.*, 2002; Parmesan, 2006), and the rapidity with which these
1055 changes occur make predictive modeling difficult. Although detecting regime shifts
1056 is increasingly difficult as we increase the extent and complexity of the system in
1057 question (Jorgensen & Svirezhev, 2004), advances in the collection and analysis of
1058 ecological data (La Sorte et al. 2018) may improve our ability to detect impending
1059 regime shifts in time for intervention (Jorgensen & Svirezhev, 2004; Groffman *et*
1060 *al.*, 2006; deYoung *et al.*, 2008; Carpenter *et al.*, 2011; Sagarin & Pauchard, 2012;
1061 Wolkovich *et al.*, 2014).

1062 Numerous quantitative approaches have been proposed as regime shift detection
1063 methods (Mantua, 2004; Rodionov & Overland, 2005, @andersen_ecological_2009;
1064 Clements & Ozgul, 2016), but few are consistently applied to terrestrial ecological

1065 data (deYoung *et al.*, 2008). I broadly classify these methods as either model-based
1066 or model-free (Hastings & Wysham, 2010; Boettiger & Hastings, 2012; Dakos *et al.*,
1067 2012a). Model-based methods use mathematical (mechanistic) representations of
1068 the system (Hefley *et al.*, 2013), which often carrying strict assumptions that are
1069 easily violated by dynamic systems such as ecosystems (Abadi *et al.*, 2010). Further,
1070 model misspecification may yield spurious results (Perretti *et al.*, 2013). Model-free
1071 (or metric-based, per Dakos *et al.*, 2012a) regime detection methods require fewer
1072 assumptions to implement than do model-based methods, and typically require much
1073 less knowledge (if any) about system component interactions. The most widely used
1074 model-free methods include both descriptive statistics of individual system components,
1075 such as variance, skewness, and mean value (Mantua, 2004; Rodionov & Overland,
1076 2005; Andersen *et al.*, 2009) and composite measures of multiple variables, notably
1077 principal components analysis (Petersen *et al.*, 2008; Möllmann *et al.*, 2015), clustering
1078 algorithms (Beaugrand, 2004), and variance index (Brock & Carpenter, 2006).

1079 3.2.1 Fisher Information as a Regime Detection Method

1080 A method which has been more recently applied in the analysis of ecological and
1081 social-ecological systems is Fisher Information (Cabezas & Fath, 2002; Karunanithi
1082 *et al.*, 2008). As a multivariate, model-free method, Fisher Information integrates
1083 the information present in the entire data of a system and distills this complexity
1084 into a single metric. This allows Fisher Information to capture ecosystem dynamics
1085 with higher accuracy than univariate-based metrics, which frequently fail to detect
1086 regime changes (Burthe *et al.*, 2016). However, despite the potential of this method its
1087 mathematical underpinnings – specifically its calculation and the concepts behind its
1088 calculation – are not clear. In this paper, I address this knowledge gap. I first provide
1089 an overview of the method and highlight the need to account for scaling properties, an

1090 inherent feature in complex systems. I then succinctly describe the decoupling of the
1091 dimensionality-reduction component from the actual Fisher Information calculation
1092 component using a two-species predator-prey model. I finally discuss the results from a
1093 theoretical viewpoint and its practical utility for predicting regime shifts, an increasing
1094 concern motivated by current rates of fast ecological change.

1095 **3.2.2 The Sustainable Regimes Hypothesis**

1096 Fisher Information (hereafter, FI; Fisher, 1922) is a model-free, composite measure
1097 of any number of variables, and is proposed as an early warning signal for ecological
1098 regime shift detection and as a measure of system sustainability (Mayer *et al.*, 2007;
1099 Karunanihi *et al.*, 2008; Eason & Cabezas, 2012; Eason *et al.*, 2014). Three definitions
1100 of FI in this context exist: (i) a measure of the ability of the data to estimate a
1101 parameter, (ii) the amount of information extracted from a set of measurements
1102 (Frieden, 1990; Roy Frieden, 1998), and (iii) a measure representing the dynamic
1103 order/organization of a system (Cabezas & Fath, 2002). Although definitions (i) and
1104 (ii) are widely applied in the statistical and physical sciences, I focus on definition
1105 (iii) as it is gaining traction as a tool to analyze used in the context of eco ecological
1106 systems analysisresponses to fast environmental change. The application of FI to
1107 complex ecological systems was posed as part of the “Sustainable Regimes Hypothesis,”
1108 stating a system is sustainable, or is in a stable dynamic state, if over some period
1109 of time the average value of FI does not drastically change (Cabezas & Fath, 2002).
1110 This concept can be described using an ecological example. Consider the simple
1111 diffusion of a population released from a point source at $t = 0$. This process can
1112 be described by a bivariate normal distribution, $p(x, y|t)$. As the time since release,
1113 t , increases, the spread of the distribution, $p(x, y|t)$, disperses because the animals
1114 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 σ_r , σ_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

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

1140 step in calculating FI, what is lost, and whether this is important. Other dimension
1141 reduction techniques, e.g., principal component analysis (PCA) and redundancy
1142 analysis (RDA), attempt to preserve the variance of the data, and the number of
1143 components scales with the dimensionality of the data (i.e. they are scale explicit).
1144 In contrast, by reducing entirely the dimensionality of the data, the FI method does
1145 not identify which features of the data are preserved, and the dimensionality does not
1146 scale with the dimensionality of the original data.

1147 **3.2.4 Aims**

1148 The key contribution of this study is that I decouple the dimensionality reduction step
1149 of the FI method (Step 1) from the statistical analysis step (Step 2). By isolating the
1150 dimensionality reduction step, we can evaluate it based on its own merits and relate it
1151 to more well-known and established methods of dimensionality reduction. By isolating
1152 the statistical analysis step, one can better understand how Fisher Information is
1153 calculated on the single-dimensional data. I believe that this decoupled approach
1154 will eliminate some confusion regarding the calculation of FI, allowing interested
1155 researchers to readily evaluate the merits of this method. To facilitate our explanation
1156 of the method, I reproduce the predator-prey analysis used in (Fath *et al.*, 2003;
1157 Mayer *et al.*, 2007), then induce a “regime shift” into the model. I hope this work will
1158 serve as a useful explanation of the FI metric for those seeking to understand it in
1159 the ecological regime shift context and will stimulate research using this and other
1160 multivariate, model-free, and composite measures to understand ecological regime
1161 shifts.

1162 3.3 Methods

1163 3.3.1 Predator-Prey Model System

1164 Our model system is a two-species predator-prey model (Equation (3.1); Fath *et al.*,
1165 2003; Mayer *et al.*, 2007; Frieden & Gatenby, 2010), hereafter referred to as the “model
1166 system”:

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

1167 The specified parameters for the model system are $g_1 = m_2 = 1, l_{12} = g_{12} = 0.01$
1168, $k = 625$, and $\beta = 0.005$ (Fath *et al.*, 2003; Mayer *et al.*, 2007; Frieden & Gatenby,
1169 2010). The initial conditions (predator and prey abundances,) for the model system
1170 were not provided in the original references (Fath *et al.*, 2003; Mayer *et al.*, 2007).
1171 I used package `deSolve` in Program R (version 3.3.2) to solve the model system
1172 [Equation Equation (3.1)] finding $x_1 = 277.781$ and $x_2 = 174.551$ to provide
1173 reasonable results. A complete cycle of this system corresponds to 11.145 time
1174 units.

1175 3.3.2 Inducing a Regime Shift

1176 Mayer *et al.* (2007) calculated FI for a predator-prey system for several discrete values
1177 of carrying capacity of prey. The results of this study showed that FI was different for
1178 systems with different carrying capacities (K). However, this study did not address
1179 the central question of **FI behavior during a regime shift**. As an extension of the
1180 original study, I simulated a regime shift by modeling an abrupt decline in carrying
1181 capacity, k . I assume k is described by Equation (3.2) where k_1 is the initial carrying
1182 capacity, k_2 is the final carrying capacity, t_{shift} is the time the regime shift occurred,
1183 and α is the parameter controlling the rate (slope) of the regime shift. The hyperbolic

1184 tangent function (see Equation (3.2)) simulates a smooth and continuous change in k
 1185 while still allowing for the regime shift to occur rapidly. I incorporate the change in
 1186 k into our system of differential equations by defining the rate of change in k , $k'(t)$,
 1187 given by (Equation (3.2)). I assume $k_1 = 800$ and $k_2 = 625$, values corresponding
 1188 to the range of carrying capacities explored by Mayer *et al.* (2007). I simulated a
 1189 time series of 600 time units, introducing a regime change after 200 time units, and
 1190 $\alpha = 0.05$.

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

1191

1192 3.3.3 Decoupling the Steps for Calculating Fisher Information 1193

1194 Two methods exist for calculating Fisher Information (FI) as applied to ecological
 1195 systems data to which I refer the “derivatives-based” method (first appearing in
 1196 Cabezas & Fath (2002) and the binning” method (first appearing in Karunanithi *et al.*
 1197 (2008)). Although the binning method is proposed as an alternative to the derivatives-
 1198 based method for handling noisy and sparse data, our decoupling method reveals
 1199 it may be an unnecessary method. Therefore, I focus on only the derivatives-based
 1200 method for explaining the theoretical basis for the FI method. The general form of
 1201 FI can be found in (Fath *et al.*, 2003; Mayer *et al.*, 2007) and I refer the reader to
 1202 (Cabezas & Fath, 2002).

1203 **Step 1:** Dimensionality Reduction. The key idea of the dimensionality reduction step
 1204 is to calculate the Euclidean distance travelled in phase space. In phase space, each
 1205 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 travelled represents the cumulative change in state relative to an arbitrary starting point in time. For the model system, the distance travelled in phase space can be obtained by solving the differential equation given by Equation (5.3)

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

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 travelled, 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 travelled 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 travelled. However, the distance travelled 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 travelled 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 travelled.

What is preserved in the calculation of distance travelled 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 travelled 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.

1231 **Step 2: Statistical Analysis.** The product of **Step 1** is a one-dimensional time
 1232 series of what I call “distance travelled”, s , (in phase space). The variable s is referred
 1233 to as “Fisher variable s” and “represent[s] a particular state of phase space” in the
 1234 FI literature (Mayer *et al.*, 2007). I believe distance travelled (s) is more descriptive
 1235 than “Fisher Variable s” and avoids confusing the state of the system, defined in
 1236 multiple dimensions by the multivariate data , with the one-dimensional summary.
 1237 Using this measure, we next calculate the probability of observing the system in a
 1238 particular state by assuming a one-to-one mapping between distance travelled and
 1239 the system state. That is, we calculate the probability of observing the system at
 1240 a particular distance, $p(s)$, along the trajectory for some period of time from 0 to
 1241 t_{end} . The time at which we observe the system is assumed to be a random variable,
 1242 $T_{obs} \sim Uniform(0, t_{end})$. This approach assumes the system is deterministic and is
 1243 observed without error. However, the observed distance travelled by the system, s , is
 1244 a random variable because it is a function of the random observation time.

1245 Importantly, the probability of observing the system at a particular value of s increases
 1246 if the system is changing slowly at that point in time. That is $p(s)$ is inversely
 1247 proportional to the system rate of change, s' . Mathematically, the distance travelled
 1248 in phase space, s , is a monotonically increasing function of time and we assume it is
 1249 differentiable. Therefore, the probability density function of the distance travelled is
 1250 $p(s) = \frac{1}{T} \frac{1}{s'}$, where $s' = \frac{ds}{dt}$ is the speed (or velocity) of s , and T is the time interval over
 1251 which the system was observed ($t_{start}-t_{end}$). We note that $p(s)$ is simply a constant
 1252 multiplied by the inverse of the speed of the system.

1253 The original motivation for the FI calculation as applied to ecological systems was the
 1254 hypothesis that “since Fisher Information is a measure of the variation” it is also “an
 1255 indicator of system order, and thus system sustainability” (Cabezas & Fath, 2002).
 1256 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 inf if the system is not changing over some period of time (Equation (4.4)).

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

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

##Results Distance travelled (s), speed ($\frac{ds}{dt}$), and acceleration ($\frac{d^2s}{dt^2}$) capture the dynamics of the model system [Equation (3.1); Figure @ref(fig:distSpeedAccel)]. 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 travelled, 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,

1280 I performed two separate calculations of FI using window sizes corresponding to the
1281 initial (when $t < 200$) and the final ($t > 200$) periods of the system ($winsize = 13.061$
1282 and 11.135 time units, respectively). Using these window sizes the drop in FI at the
1283 regime shift initiation is bigger than the magnitude of the fluctuations preceding it
1284 (Figure 3.6).

1285 3.4 Discussion

1286 Part of the appeal of the FI method of regime shift detection is that it provides a
1287 1-dimensional visual summary of system “orderliness”. However, I have demonstrated
1288 that the dimensionality reduction step can be performed separately from the calculation
1289 of FI. The rate of change of the system (velocity, $\frac{ds}{dt}$), on which FI method is based,
1290 is also a 1-dimensional quantity. In the simple predator-prey example, calculating and
1291 plotting FI did not provide a clear benefit over simply plotting the system rate of
1292 change directly. I suggest that future research uncouple the dimensionality reduction
1293 step and the FI calculation step in order to better illustrate the benefits of the FI
1294 method relative to dimensionality reduction alone. In the predator-prey example, I
1295 assumed the data was free from observation error. Despite these ideal conditions,
1296 the estimated FI had high variation and the results depended on the size of the time
1297 window used in the calculation. This issue arises because the period of the cyclic
1298 system is changing during the regime shift such that it is difficult to find a single
1299 window size that works well for the entire time series. Mayer *et al.* (2007) describe this
1300 as a “confounding issue” related to “sorting out the FI signal of regime change from
1301 that originating from natural cycles” and suggest using a time window that is large
1302 enough to include several periods. However, in the absence of a quantitative decision
1303 rule defining what changes in FI indicate regime shifts, it is difficult to separate
1304 the signal in the FI metric from the noise due to fluctuations in the natural cycles.

₁₃₀₅ Further research is needed to define quantitative decision rules for what changes in FI
₁₃₀₆ constitute a regime shift.

₁₃₀₇ The example used in this study is unrealistic in that I assume no measurement error
₁₃₀₈ and therefore focus on the “derivatives-based” method of calculating FI. However, our
₁₃₀₉ analysis also has implications for the “binning” method of calculating FI that was later
₁₃₁₀ developed for high-dimension noisy data (Karunanihi *et al.* (2008)). Rather than
₁₃₁₁ attempting to estimate the rate of change of each system component (e.g., hundreds
₁₃₁₂ of species) and combining these estimates to get the total system rate of change, I
₁₃₁₃ suggest an approach where the dimensionality of the data is first reduced by calculating
₁₃₁₄ distance travelled in phases-pace and then only a single rate of change is estimated.
₁₃₁₅ The advantage of this approach is that for an n-dimensional system it only requires
₁₃₁₆ the estimation of one derivative rather than n-derivatives . The drawback to this
₁₃₁₇ approach is that noisy observations will likely introduce some bias into the estimate
₁₃₁₈ of the system rate of change. Nonetheless, I believe this approach is worth exploring
₁₃₁₉ due to its simplicity relative to the “binning” method. The Fisher Information of
₁₃₂₀ an *n*-dimensional system is a vector of unitless values which can only be compared
₁₃₂₁ within a dataset (e.g., within a single community time series) and interpreting FI is
₁₃₂₂ still largely a qualitative effort (Fath *et al.*, 2003; Mantua, 2004), not unlike most
₁₃₂₃ regime detection methods [Ch. 2]. When the FI of a system is increasing, the system
₁₃₂₄ is said to be moving toward a more orderly state, and most studies of FI propose
₁₃₂₅ that sharp changes in FI, regardless of the directionality of the change, may indicate
₁₃₂₆ a regime shift (Cabezas & Fath, 2002; Karunanihi *et al.*, 2008; Spanbauer *et al.*,
₁₃₂₇ 2014). Although the aforementioned and numerous other works interpret FI in this
₁₃₂₈ context (e.g., Eason *et al.*, 2014; Eason & Cabezas, 2012), I suggest future work which
₁₃₂₉ clearly identifies the ecological significance of the Fisher Information metric and its
₁₃₃₀ significance within the ecological regime shift paradigm.

1331 3.5 Acknowledgements

1332 I thank H. Cabezas and B. Roy Frieden for early discussions regarding the development
1333 of Fisher Information, and T.J. Hefley for comments on an earlier draft. This work
1334 was funded by the U.S. Department of Defense's Strategic Environmental Research
1335 and Development Program (project ID: RC-2510).

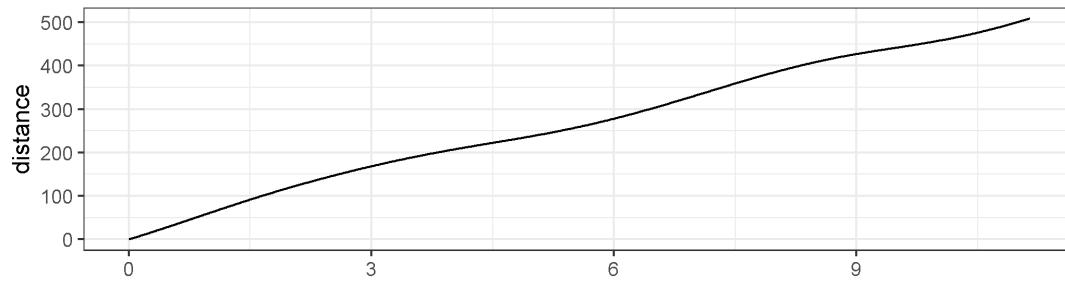
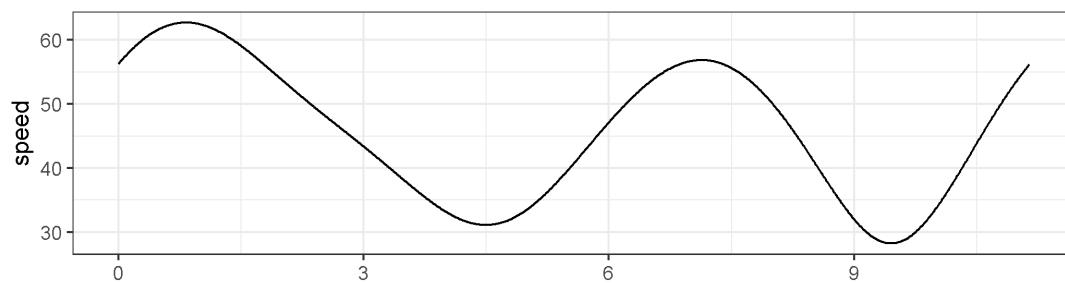
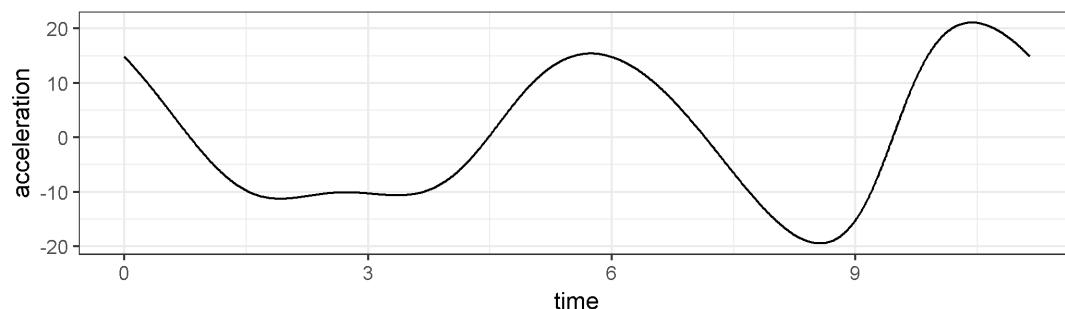
a**b****c**

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

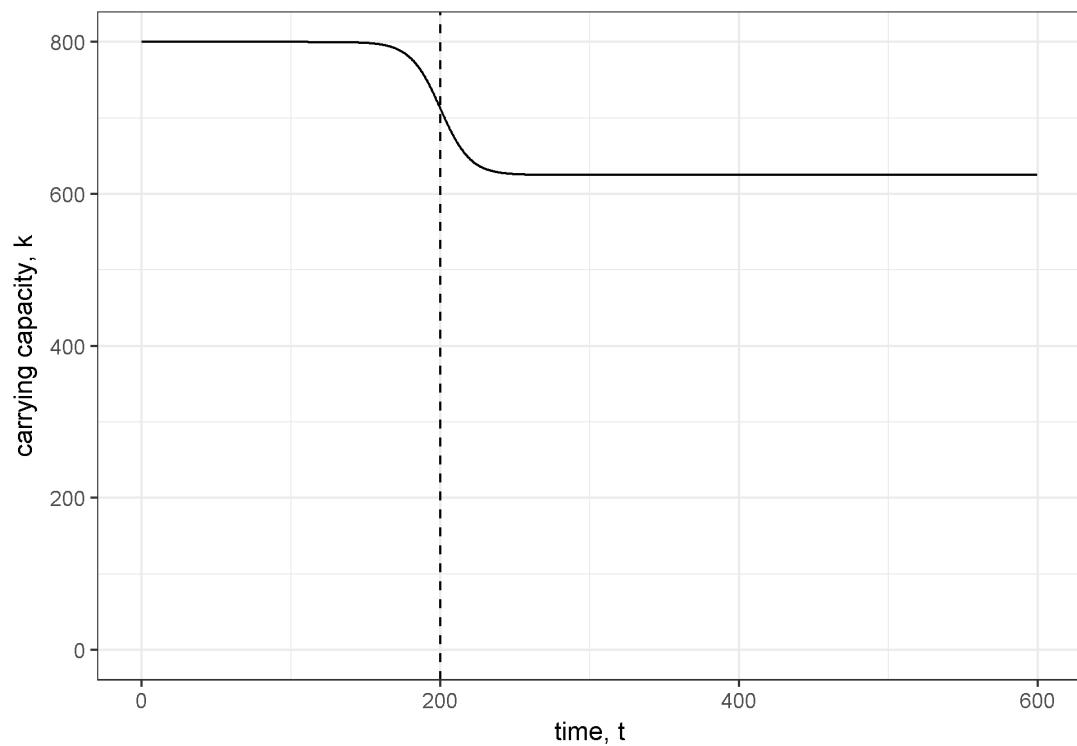


Figure 3.2: Carrying capacity over time with a regime shift occurring around time 200.

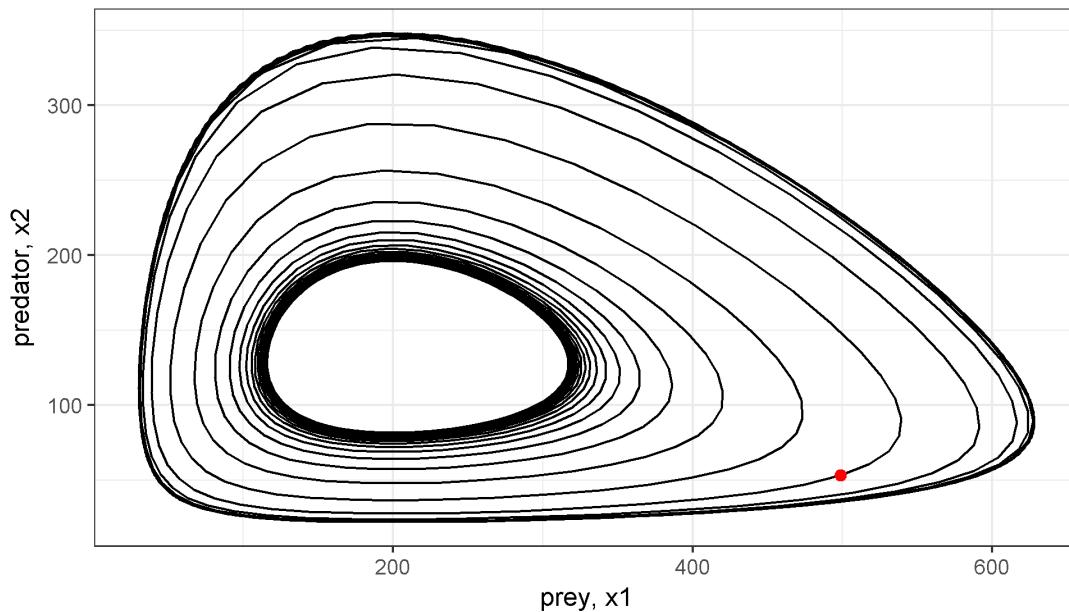


Figure 3.3: Phase space plot of system trajectories for different values of k

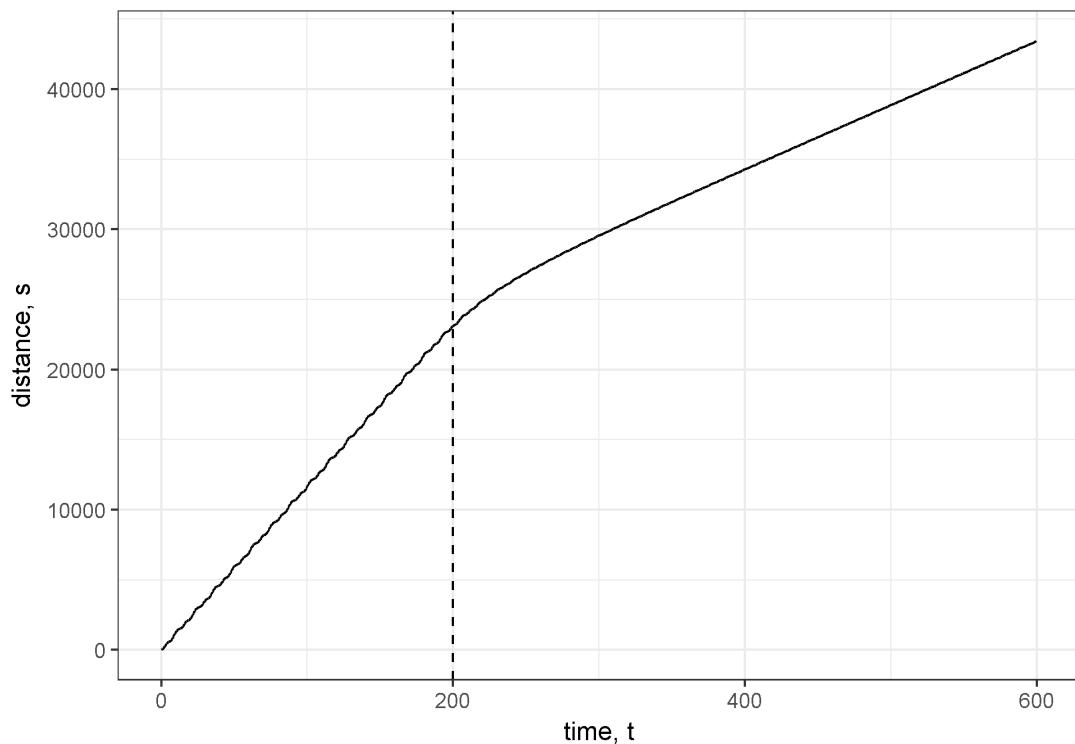


Figure 3.4: Distance travelled in phase space over time. Dashed vertical line at time 200 indicates location of regime shift.

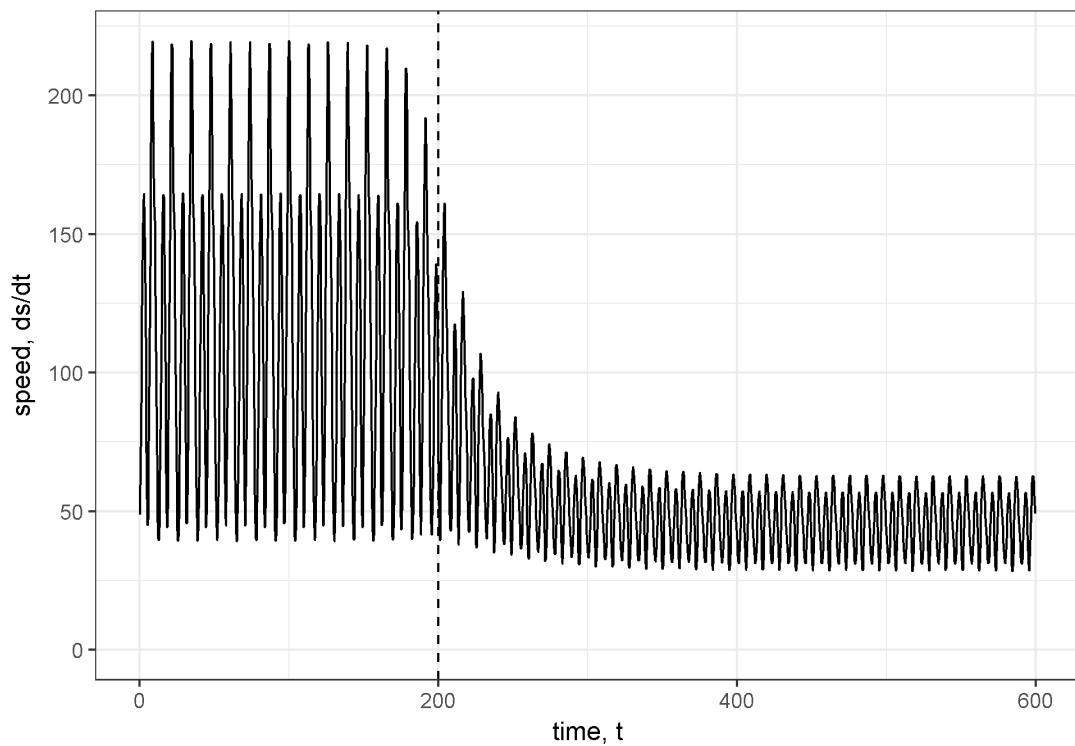


Figure 3.5: Speed of the system (rate of change, velocity) in phase space. Dashed vertical line at time 200 indicates location of regime shift.

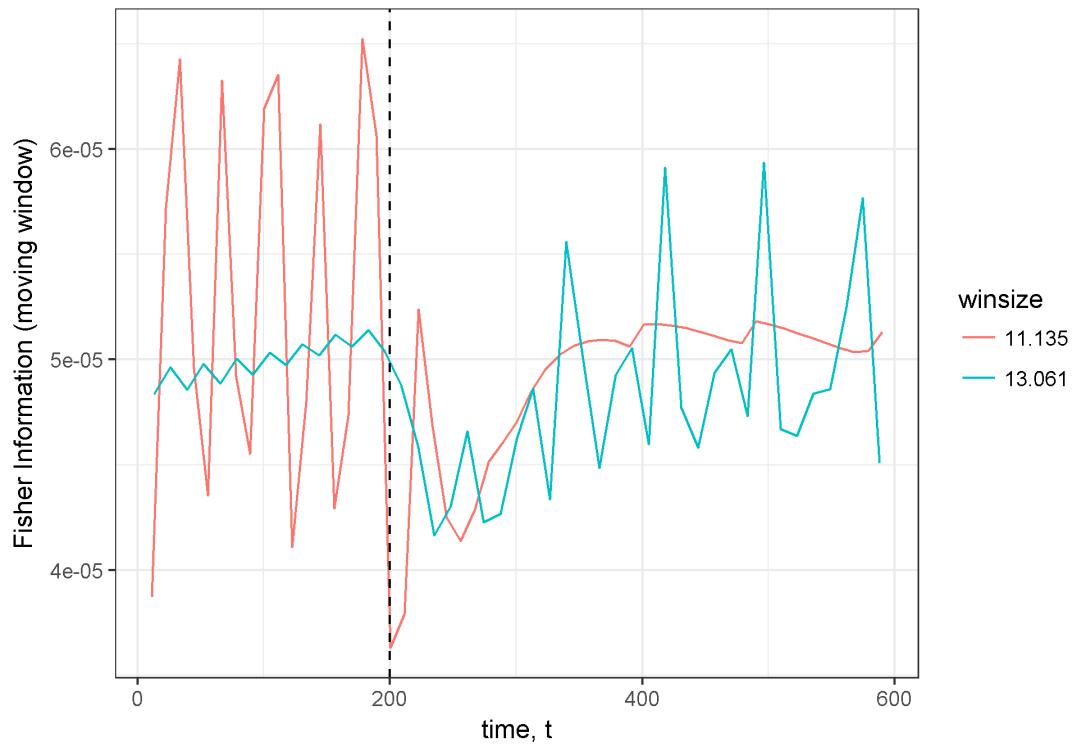


Figure 3.6: Fisher Information calculated for non-overlapping time windows. Two different window sizes were used as indicated by color. Dashed vertical line at time 200 indicates approximate location of regime shift.

1336 **Chapter 4**

1337 **An application of Fisher**

1338 **Information to spatially-explicit
1339 avian community data**

1340 **4.1 Introduction**

1341 Ecosystems are open, dynamical systems which in most cases cannot be fully rep-
1342 resented by fully parameterized models. Some patterns have emerged in certain
1343 statistical mechanics of ecological observations. despite the complexity of most eco-
1344 logical systems. An uptick in recent years of studies of **regime shifts** (Table 1)
1345 in ecology has spurred an increase in the number of ‘new’ methods for detecting
1346 ecological regime shifts (Chapter 2), some of which are proposed as indicators of
1347 ‘spatial’ regime shifts (Brock & Carpenter, 2006; Guttal & Jayaprakash, 2009; Kefi *et*
1348 *al.*, 2014, @sundstrom2017detecting; Butitta *et al.*, 2017).

1349 As defined in Table 1, a regime shift is largely considered an abrupt and persistent
1350 change in a system’s structure or functioning. Following this definition and without

1351 considering the **pressures** (Table 1) associated with the observed regime shift, it
1352 is not yet clear whether identifying a ‘spatial regime’ using a snapshot of a system
1353 (i.e. using a single or short period of time relative to the time scale of the system
1354 dynamics and/or pressures) is pragmatic. A concise and global definition of the SRDM
1355 is important since observations of non-random spatial processes (e.g., land cover)
1356 can manifest as either a rapid shift (e.g. an ecotone) or as a gradual change (e.g.,
1357 slow mixing along a gradient). Consequently, and because most RDMs signal abrupt
1358 change, only the former may be identified as “regime shifts” using SRDMs. For the
1359 concept of SRDMs to be practical to the ecological management and conservation and
1360 given the controversy in many methods used to detect temporal and spatial regime
1361 shifts, multiple measures should concur the existence and location of a regime shift in
1362 space *and* time. Identifying the potential pressures associated any observed shift can
1363 lend credibility to the results obtained using temporal and spatial RDMs. Additionally
1364 and perhaps more importantly, the processes driving the observed information (drivers)
1365 should be such that a statistically identified regime shift will roughly correspond with
1366 the time scale on which the pressure(s) operate.

1367 Although it is suggested that statistical and pragmatic methods are advanced more
1368 rapidly by bottom-up approaches, i.e. case studies (see DeAngelis & Yurek, 2017),
1369 few studies test the rigor of SRDMs using spatially-explicit, *empirical* data. The
1370 objective of this chapter is to determine the utility of Fisher Information [Eq. (4.4)]
1371 as a spatial regime detection measure. This chapter is also supported by original
1372 software developed for implementation in Program R, which is publicly available [see
1373 Appendix 8.4].

1374 Despite its controversial applicability to temporal data (Bestelmeyer *et al.*, 2011;
1375 Perretti & Munch, 2012; Burthe *et al.*, 2016; Sommer *et al.*, 2017; Dutta *et al.*, 2018),
1376 variance is proposed as a spatial regime detection measure (hereafter, SRDM; Brock

₁₃₇₇ & Carpenter, 2006). Here, variance is assumed to increase across space as an SRDM
₁₃₇₈ is approached, capturing the variability in the landscape relative to its surroundings.

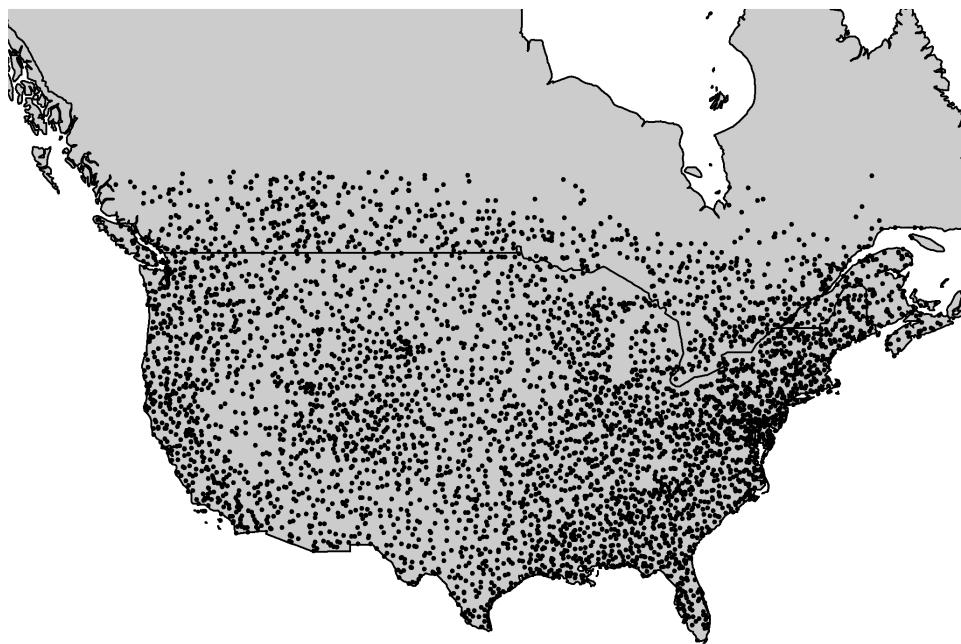
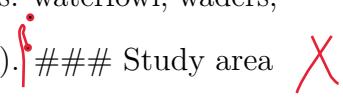


Figure 4.1: Locations of Breeding Bird Survey routes sampled between 1966 and 2017.

1380 4.2 Data and methods**1381 4.2.1 Data: North American breeding bird communities**

1382 I use community abundance data (Sauer *et al.*, 2014) from long-term monitoring
1383 programs to identify spatial and temporal regimes using the Fisher Information
1384 (FI) derivatives method (see Eq. (4.4)). The North American Breeding Bird Survey
1385 (NABBS) trains citizen scientist volunteers to annually collect data using a standardized
1386 roadside, single observer 3-minute point count protocol and has organized data
1387 collection annually across North America (Figure 4.1) since 1966. The roadside
1388 surveys consist of 50 point counts (by sight and sound) along ~ 24.5 mile stretch
1389 of road. Due to strict reliance on volunteers, some routes are not covered every
1390 year. Additionally, some routes are moved or discontinued due to changing landscape
1391 conditions and change in observer safety. Route-year combinations that were missing
1392 years but were not discontinued were treated as missing data. Although NABBS
1393 volunteers attempt identify all species as possible, persistent biases exist in this
1394 protocol. Despite a standardized survey protocol, some species are difficult to identify
1395 using these methods. For example, crepuscular species are less likely to be detected
1396 beyond the first few points of the BBS route, given they are most active at sunrise
1397 and the survey begins within 30 minutes of sunrise. Further, species which congregate
1398 in large groups and are highly mobile (e.g., waterfowl) tend to have less reliable
1399 inter-annual abundance estimates given their ability to move long-distances in a short
1400 period of time. To remove any potential influence of sampling bias on the Fisher
1401 Information result, I removed birds of these types from all analyses: waterfowl, waders,
1402 and shore species (BBS AOU numeric codes 0000 through 2880).
1403 Although the NABBS conducts surveys throughout much of North America (most of
1404 the United States, Canada, and Mexico), coverage of the boreal forests of Canada are

¹⁴⁰⁵ sparse in space, and many routes in Mexico have fewer than 25 years of observations.
¹⁴⁰⁶ For these reasons I limited analyses largely to the continental United States and parts
of Southern Canada (see Figure 4.1).



Figure 4.2: A single East-West transect of Breeding Bird Survey
routes used to calculate the Fisher Information.

¹⁴⁰⁷

¹⁴⁰⁸ Focal military base

¹⁴⁰⁹ The Mission of the U.S. Department of Defense is to provide military forces to deter
¹⁴¹⁰ war and protect the security of the country, and a primary objective of individual
¹⁴¹¹ military bases is to maintain military readiness. To maintain readiness, military
¹⁴¹² bases strictly monitor and manage their natural resources. Military bases vary in

1413 size and nature, and are heterogeneously distributed across the continental United
1414 States (See Figure 4.2). The spread of these bases (Figure 4.3), coupled with the top-
1415 down management of base-level natural resources presumably influences the inherent
1416 difficulties associated with collaborative management within and across military bases
1417 and other natural resource management groups (e.g., state management agencies,
1418 non-profit environmental groups.

J. S. Fenn
1419 Much like other actively managed landscapes, military bases are typically surrounded
1420 by non- or improperly-managed lands. Natural resource managers of military bases
1421 face environmental pressures within and surrounding their properties, yet their primary
1422 objectives are very different. Natural resource managers of military bases, whose
1423 primary objective is to maintain military readiness, are especially concerned with
1424 if and how broad-scale external forcings might influence their lands. Prominent
1425 concerns include invasive species, wildlife disease, and federally protected species
1426 (personal communication with Department of Defense natural resource managers at
1427 Eglin Air Force and Fort Riley military bases). For these reasons, natural resource
1428 managers attempt to create buffers along their perimeters (e.g., live fire/ammunition
1429 suppression, wide fire breaks). Identifying the proximity of military bases to historic
1430 and modern ecological shifts may provide insight into the effectiveness of their natural
1431 resource management efforts. The NABBS routes chosen for analyses in this Chapter
1432 lie within or near Fort Riley military base (located at approximately 39.110474° ,
1433 -96.809677° ; Kansas, USA). Fort Riley (Figure 4.4) is a useful reference site for this
1434 study. Woody encroachment of the Central Great Plains over the last century has
1435 triggered shifts in dominant vegetative cover and diversity (Ratajczak *et al.*, 2018) in
1436 the area surrounding Fort Riley military base (Van Auken, 2009). This phenomena
1437 should present itself as a regime boundary if Fisher Information is a reliable SRDM.

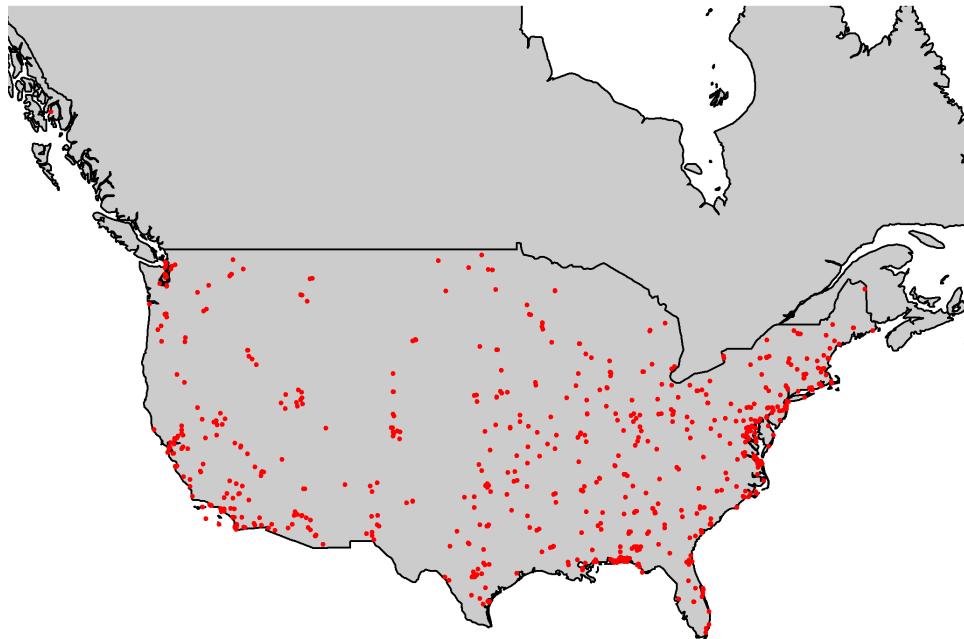


Figure 4.3: Locations of Fort Riley military base in Kansas, USA.

1439 **Spatial sampling grid**

1440 Fisher Information has been applied to empirical data as a SRDM in recent years
1441 (Sundstrom *et al.*, 2017; Eason *et al.*, 2019). The authors of Sundstrom *et al.* (2017)
1442 used the Fisher Information binning method to prove the concept of this method as
1443 an SRDM, suggesting that the metric should detect ‘regime changes’ when adjacent
1444 sampling points represented different ecoregions (which is a nationally-recognized,
1445 broad-scale vegetation classification scheme). Suggesting that each ecoregion in their
1446 analyses should be similarly represented ~~wit~~ respected to the number of NABBS
1447 routes in each ecoregion, the authors handpicked NABBS survey points (routes)

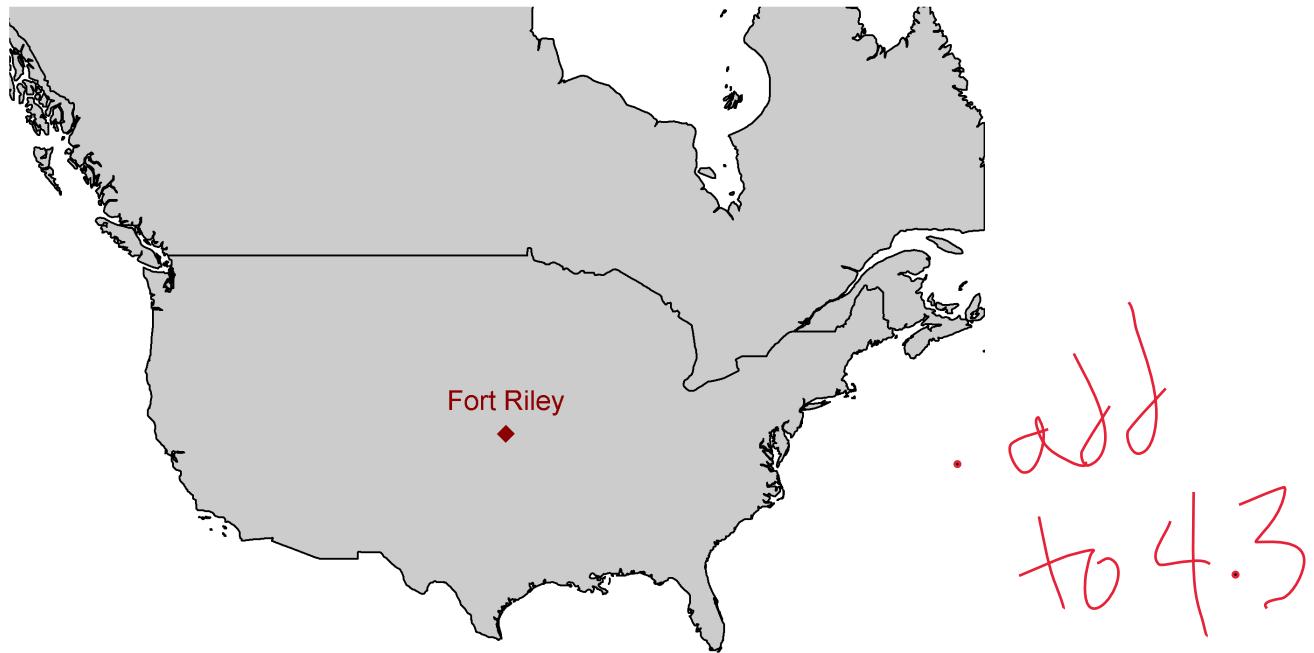


Figure 4.4: Locations of focal U.S. military bases, Eglin Air Force Base (AFB) and Fort Riley Military Base.

1448 which resulted in a transect which was neither North-South nor East-West running.

1449 Rather, the handpicked routes zigzagged across a Midwestern region through multiple
1450 ecoregions (Sundstrom *et al.*, 2017).

1451 To ameliorate any potential bias in handpicking NABBS routes, I first constructed
1452 a gridded system across the continental United States and parts of Canada. The
1453 gridded system comprises East-West running transects transects, ameliorating potential
1454 sampling bias as the transect location and widths were designed to capture large-scale
1455 shifts in bird communities at regular intervals. This spatial sampling grid approach also
1456 allows for raster stacking, or layering data layers (e.g., vegetation, LIDAR, weather),

1457 providing an opportunity to identify potential relationships with abiotic drivers, should
1458 regime shifts be observed in the avifauna data. This spatial sampling method also
1459 provides a simple vector for visualizing changes in the Fisher Information over space-
1460 time. For brevity, I present visual results of only three, spatially-adjacent, East-West
running transects (Figure 4.5) at multiple time periods.

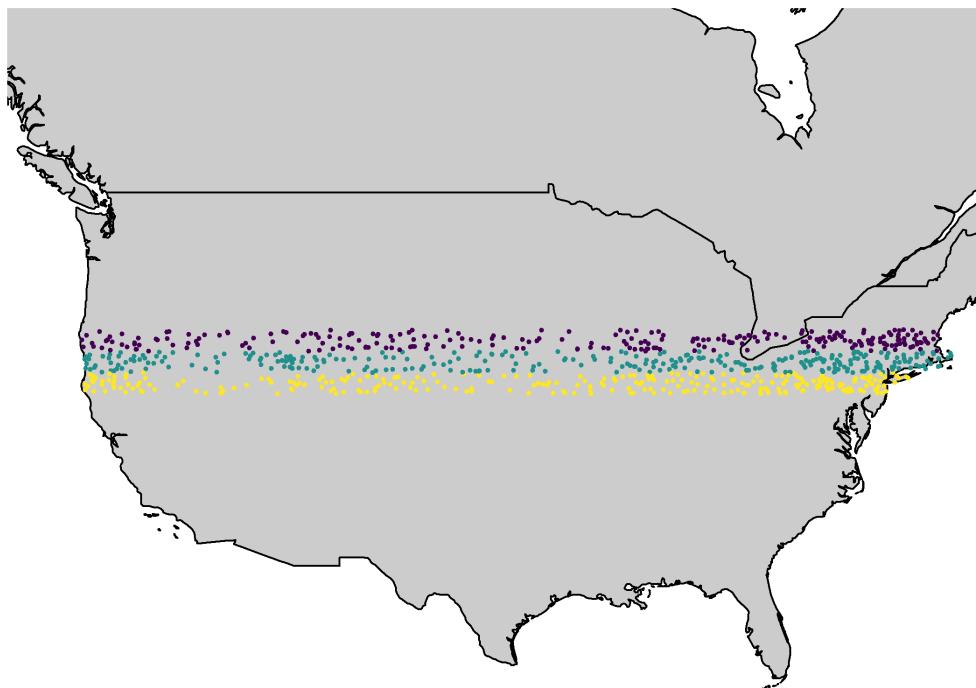


Figure 4.5: The three East-West running transects used to visualize results in this chapter.

redundant with 4.2

1462 4.2.2 Calculating Fisher Information (FI)

1463 Fisher Information, $I(\theta)$, was developed in 1922 by Ronald Fisher as a measure of
 1464 the amount of information that an observable variable, X , reveals about an unknown
 1465 parameter, θ . Fisher Information is a measure of indeterminacy (Fisher 1922) and is
 1466 defined as,

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

1467 where $p(y|\theta)$ is the probability density of obtaining the data in presence of θ . The Fisher
 1468 Information measure (FIM) is used to calculate the covariance matrix associated with
 1469 the likelihood, $p(y|\theta)$. Fisher Information is described as Extreme Physical Information
 } 1470 (EPI; @frieden1995lagrangians; Frieden *et al.*, 2002), a measure that has been used
 1471 to track the complexity of systems in many scientific disciplines including, physics,
 1472 cancer research, electrical engineering, and, recently, complex systems theory and
 1473 ecology

1474 Fisher Information as gathered from observational data provides insight as to the
 1475 dynamic order of a system, where an orderly system is one with constant (i.e.,
 1476 unchanging) observation points, and one whose nature is highly predictable. A
 1477 disorderly system is just the opposite, where each next data point is statistically
 1478 unpredictable. In ecological systems, patterns are assumed to be a realization of
 1479 ecosystem order; therefore, one should expect orderliness in a system with relatively
 1480 stable processes and feedbacks. Orderliness, however, does not necessarily infer long-
 1481 term predictability. Equation (4.1) is next adapted to estimate the dynamic order of
 1482 an entire system, s , as

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

1483 where $p(s)$ is the probability density for s . Here, a relatively high Fisher Information
 1484 value (I) infers higher dynamic order, whereas a lower value (approaching zero) infers

1485 less orderliness. To limit the potential values of I in real data, we can calculate the
1486 amount of Fisher Information by re-expressing it in terms of a probability amplitude
1487 function $q(s)$ (Fath *et al.*, 2003; Mayer *et al.*, 2007):

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

1488 A form specific to the pdf of distance traveled by the entire system, which I call the
1489 ‘derivatives’ method, is defined as (Mayer *et al.*, 2007, eq. 7.12):

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

1490 where T is the number of equally spaced time points over which the data are integrated.
1491 Numerical calculation of I using the binning method (Eq. (4.3) and (4.4)) each
1492 incorporate a moving-window procedure for calculating the probability of the system,
1493 $p(s)$, as being in one of an unidentified number of states (s). Although previously
1494 applied to spatially-explicit terrestrial community data, the binning method requires
1495 multiple parameters to be defined *a priori*, which have been shown to influence
1496 inference based on the metric. I therefore calculated FI using the derivatives equation
1497 [see Chapter 3].

1498 The binning procedure allows for a single point in time or space to be categorized into
1499 more than one state, which violating the properties of alternative stable states theory.
1500 The size of states (see Eason and Cabezas 2012) measure is required to construct
1501 $p(s)$. In the case of high dimensional data, a univariate binning procedure of $p(s)$ is
1502 not intuitive (i.e., reducing a multivariable system to a single probability distribution
1503 rather than constructing a multivariate probability distribution). Importantly, when
1504 using community or abundance data, rare or highly abundant species can influence
1505 the size of states criterion, thus influencing the assignment of each point into states.

1506 Finally, Eq. (4.3) assumes equal spacing (in space or time) between sampling points.
1507 Each of these violations can be avoided by using Eq. (4.4) (Cabezas & Fath, 2002;
1508 Fath *et al.*, 2003) to calculate the Fisher Information measure (see Chapters 3, 5 for
1509 detailed discussions on this topic). Briefly, derivatives method (Eq. (4.4)) estimates
1510 the trajectory of the system's state by calculating the integral of the ratio of the
1511 system's acceleration and speed in state space (Fath *et al.*, 2003). Here, I use the
1512 derivatives method (Eq. (4.4)) to calculate Fisher Information for all East-West
1513 transects (see Figure 4.5) at decadal intervals (years 1980, 1990, 2000, and 2010).
1514 Justification for using this method is provided in detail in Chapter 3.

1515 **4.2.3 Interpreting and comparing Fisher Information across
1516 spatial transects**

1517 **Interpreting Fisher Information values**

1518 Interpretation of FI is still a qualitative effort. Fisher Information is proposed as an
1519 indicator of system orderliness, where periods of relatively high values of FI indicate
1520 the system is in an “orderly” state, possibly fluctuating around a single attractor. A
1521 rapid change in FI is supposed to indicated the system is no longer orderly and may
1522 be undergoing a reorganization phase. Whether Fisher Information can identify a
1523 switch among basins of attraction within a single, stable state (or around a single
1524 attractor) remains unknown, as does the number of states which a system can occupy.
1525 When a system occurs within any number of states equally, i.e., $p(s)$ is equal for each
1526 state, both the derivative, $(\frac{dq(s)}{ds})$, and I are zero. As $(\frac{dq(s)}{ds} \rightarrow \infty)$, we infer the system
1527 is approaching a stable state, and as $\frac{dq(s)}{ds} \rightarrow 0$ the system is showing no preference
1528 for a single stable state and is on an unpredictable trajectory. Eq. (4.3) bounds the
1529 potential values of Fisher Information at [0, 8], whereas Eq. (4.1), Eq. (4.2), and

Contradictory

1530 Eq. (4.4) are positively unbounded $[0, \infty)$. If the Fisher Information is assumed to
1531 represent the probability of the system being observed in some state, s , then the
1532 absolute value of the Fisher Information index is relative within a single datum (here
1533 a single datum is a spatial transect). It follows that Fisher Information should be
1534 interpreted relatively, but not absolutely.

1535 Here I define a potential regime change as a point(s) having a non-zero derivative, and
1536 at which relatively large changes (manifested as either a sharp increase or decrease) in
1537 FI occurs. Regime shifts are identified as data changing from one state to another,
1538 thus, rapid shifts in the value of FI should indicate the locations of these shifts in the
1539 time *and* space, at which the system undergoes reorganization. Spatial and temporal
1540 Fisher Information calculation does not vary, but interpretation of either differ in that
1541 a spatial analysis will identify a spatial regime boundary (Sundstrom *et al.*, 2017)
1542 within a single time period, whereas temporal analysis identifies the point in time at
1543 which the system undergoes a regime shift. I follow published recommendations for
1544 interpreting the Fisher Information results in the context of identifying regime shifts
1545 (e.g., Karunани thi *et al.*, 2008; Fath *et al.*, 2003; Eason & Cabezas, 2012).

1546 Interpolating results across spatial transects

1547 NABBS are not regularly spaced, and pairwise correlations of adjacent transects
1548 (see Figure 4.5) is not possible without either (1) binning the Fisher Information
1549 calculations using a moving-window analysis, or (2) interpolating the results to
1550 regularly-spaced positions in space. To avoid potential biases associated with the
1551 former option (i.e. choosing window size, location of data aggregation), I linearly
1552 interpolated the calculated Fisher Information within each spatial transect to 50,
1553 evenly-spaced points along the longitudinal dimension. The 50 longitudinal points to
1554 which I interpolated were the same across each spatial transect, while latitude varied

1555 across transects. I used the function `stats::approx()` (with argument `rule=1`) to
1556 linearly approximate the Fisher Information. I did not interpolate values beyond the
longitudinal range of the original data (i.e., no extrapolation).

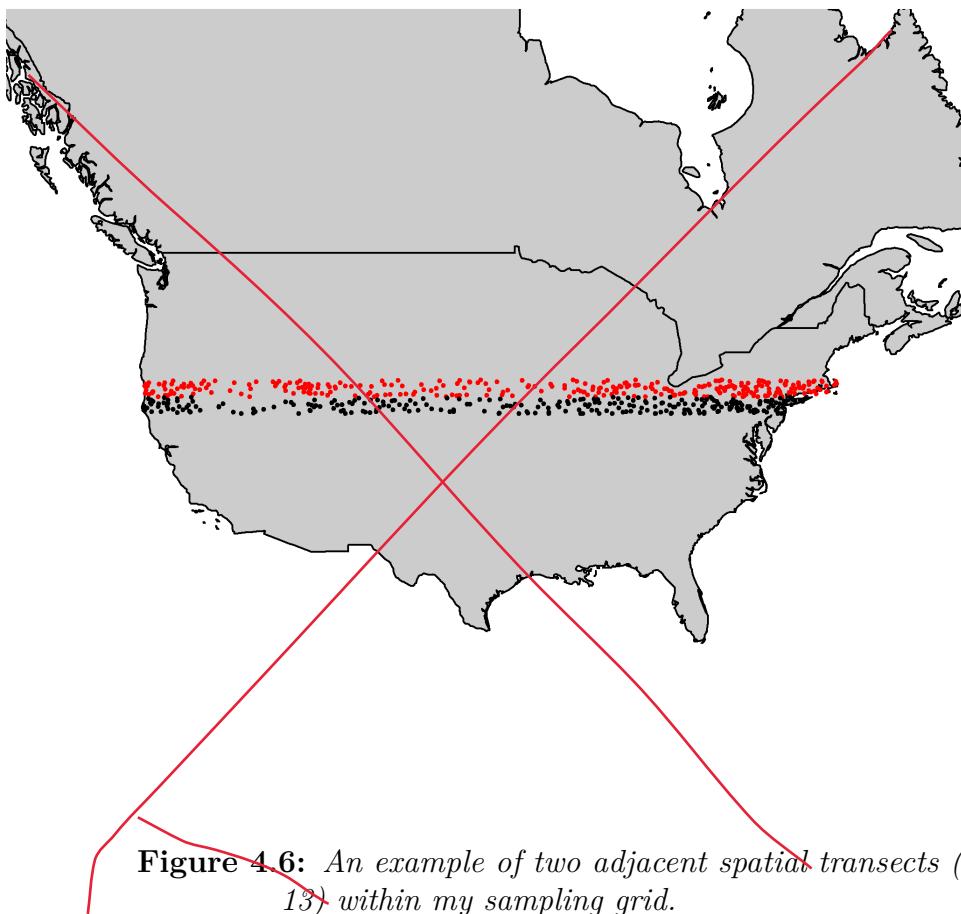


Figure 4.6: An example of two adjacent spatial transects (12, 13) within my sampling grid.

1557

1558 **Spatial correlation of Fisher Information**

1559 If Fisher Information captures and reduces information regarding abrupt changes in
1560 community structure across the landscape, then it follows that the values of Fisher
1561 Information should be spatially auto correlated. That is, the correlation of FI values
1562 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 ∞). In other words, FI values calculated are **not** relatively comparable outside of a single spatial transect (Figure 4.5). Fisher Information **is**, however, directly comparable within each spatial transect (e.g., 4.5). For these reasons, we can identify spatial regime shifts both within and among spatial transects by using pairwise correlations among two transects (e.g., 4.5) 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 4.6). 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

Interpreting the Fisher Information is currently a qualitative effort. As suggested earlier, rapid increases or decreases in FI are posited indicate a change in system orderliness, potentially suggesting the location of a regime shift. Using this method yields inconclusive results regarding the location of ‘spatial regimes’ (Figure 4.7). Of the three spatial transects analyzed in this chapter (see Figure 4.5), Figure 4.7 is representative of the lack of pattern observed in the Fisher Information values across all analyzed transects. I did not identify clear patterns within nor among spatial transects with respect to Fisher Information. Further, the log-transformed FI suppresses some of the extreme values which can visually dampen other sharp changes, but did not exhibit clear regime shifts in these data.

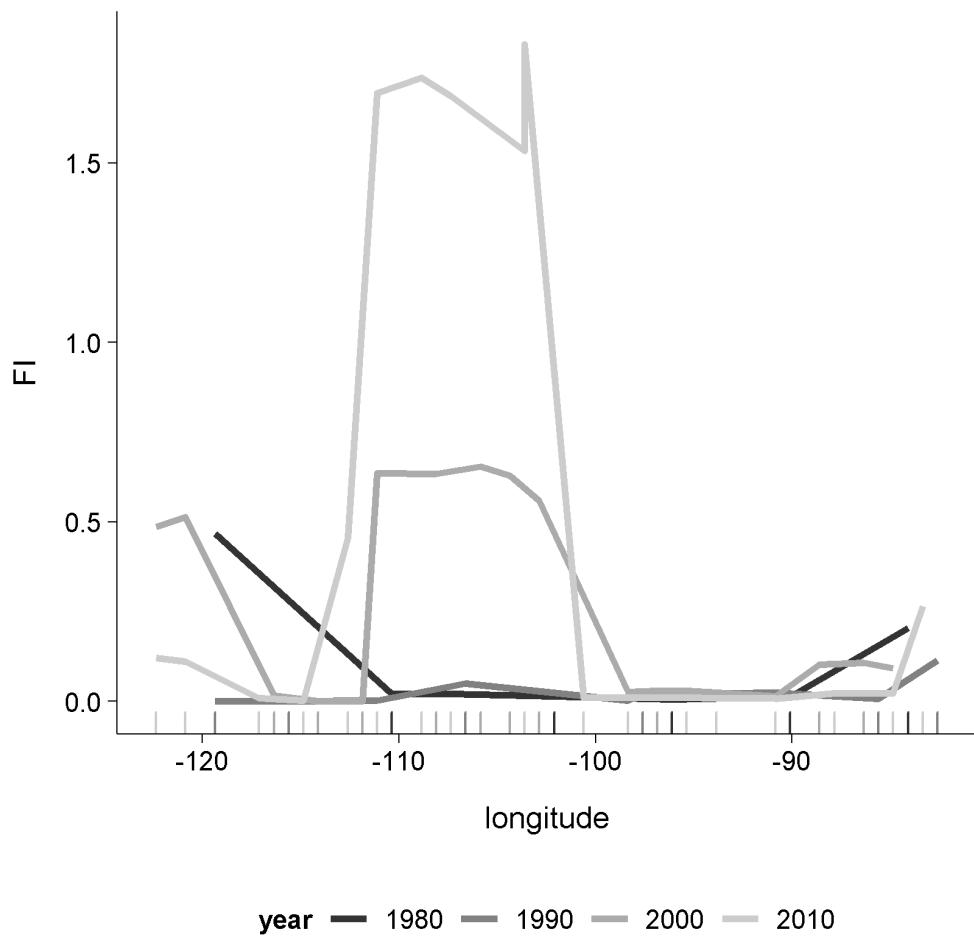


Figure 4.7: Fisher Information calculated for a single transect over time.



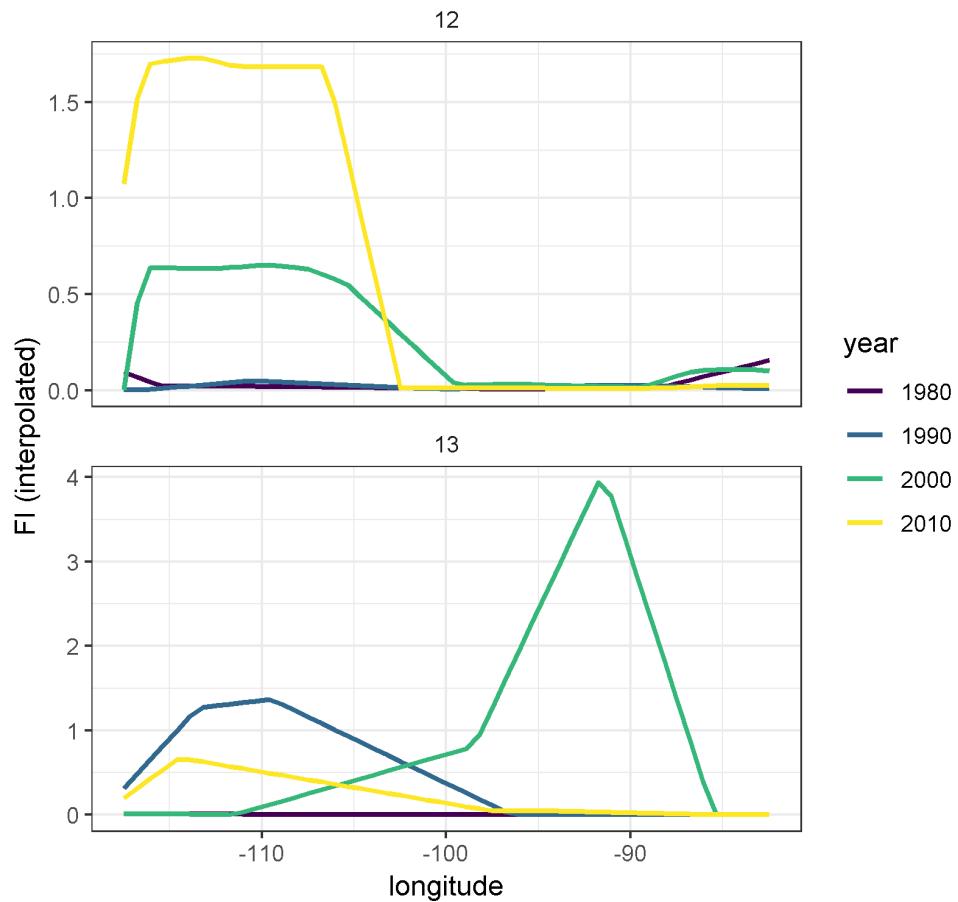


Figure 4.8: Fisher Information of two spatially adjacent transect pairs (transects 12, 13) over time. Interdecadal trends in FI are very different within each transect and are not highly correlated among transects over time.

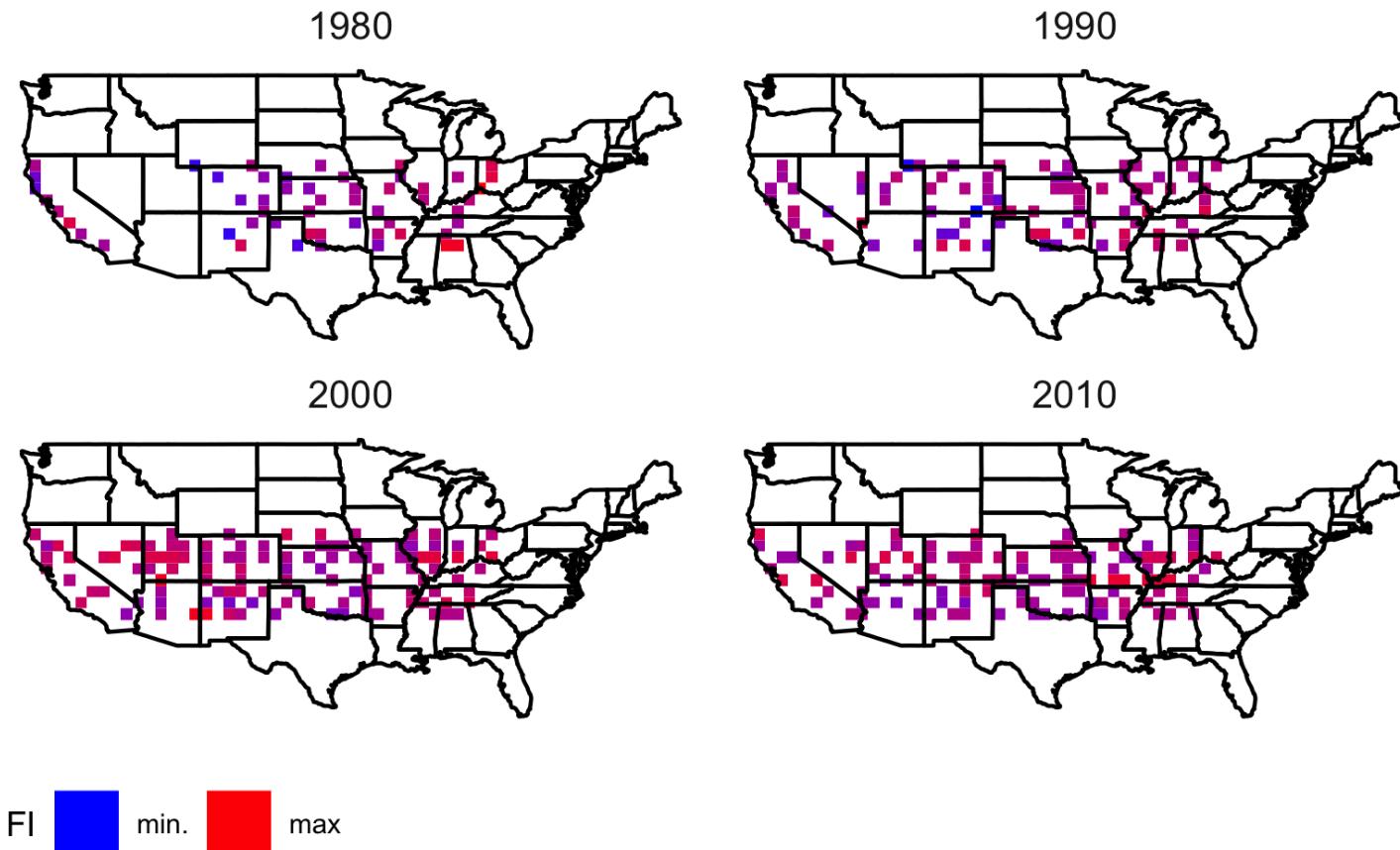


Figure 4.9: Fisher Information of 5 East-West spatial transects over time.

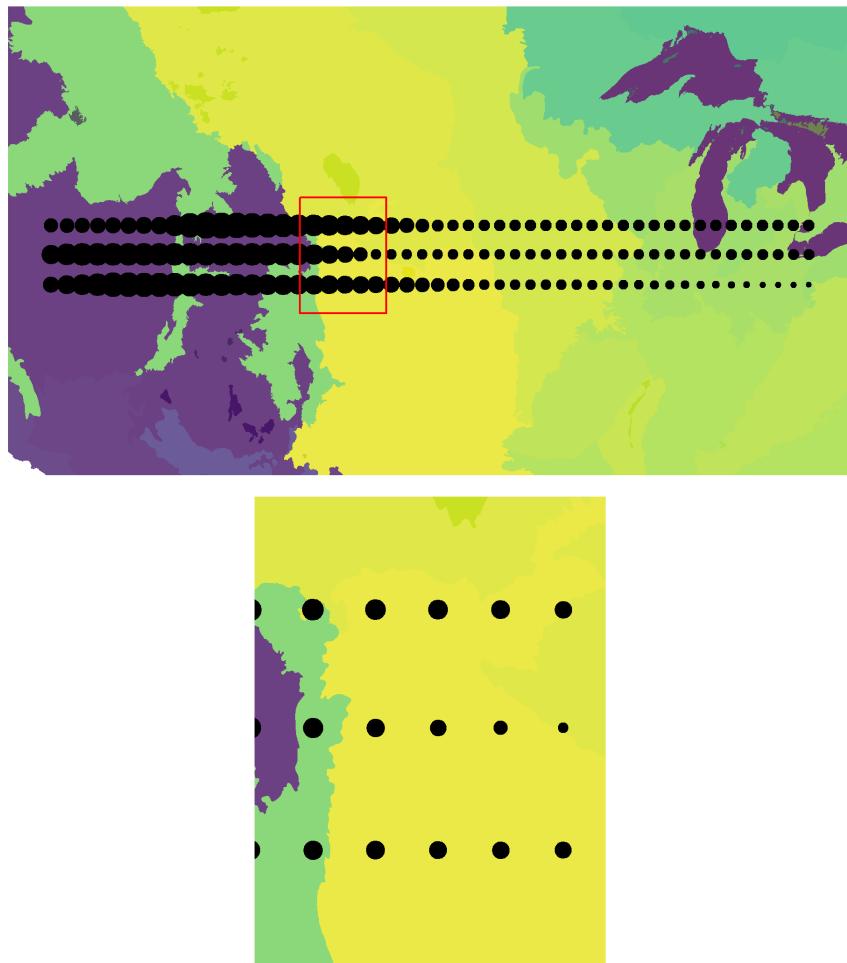


Figure 4.10: No patterns of abrupt change detected in Fisher Information along three transects in year 2010

1587 4.3.2 Spatial correlation of Fisher Information

1588 In addition to failing to identify clear geological boundaries across large swaths of our
 1589 study area, (Fig 4.9), this method did not identify spatial correlation of Fisher Informa-
 1590 tion among adjacent spatial transects (Figure 4.11)¹. For spatially-adjacent transects
 1591 (e.g, transects 11 and 12, or 12 and 13 in Figure 4.11), we should expect high and pos-
 1592 itive correlation values, and these values should stay consistent across time *unless* the

¹Pairs 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.

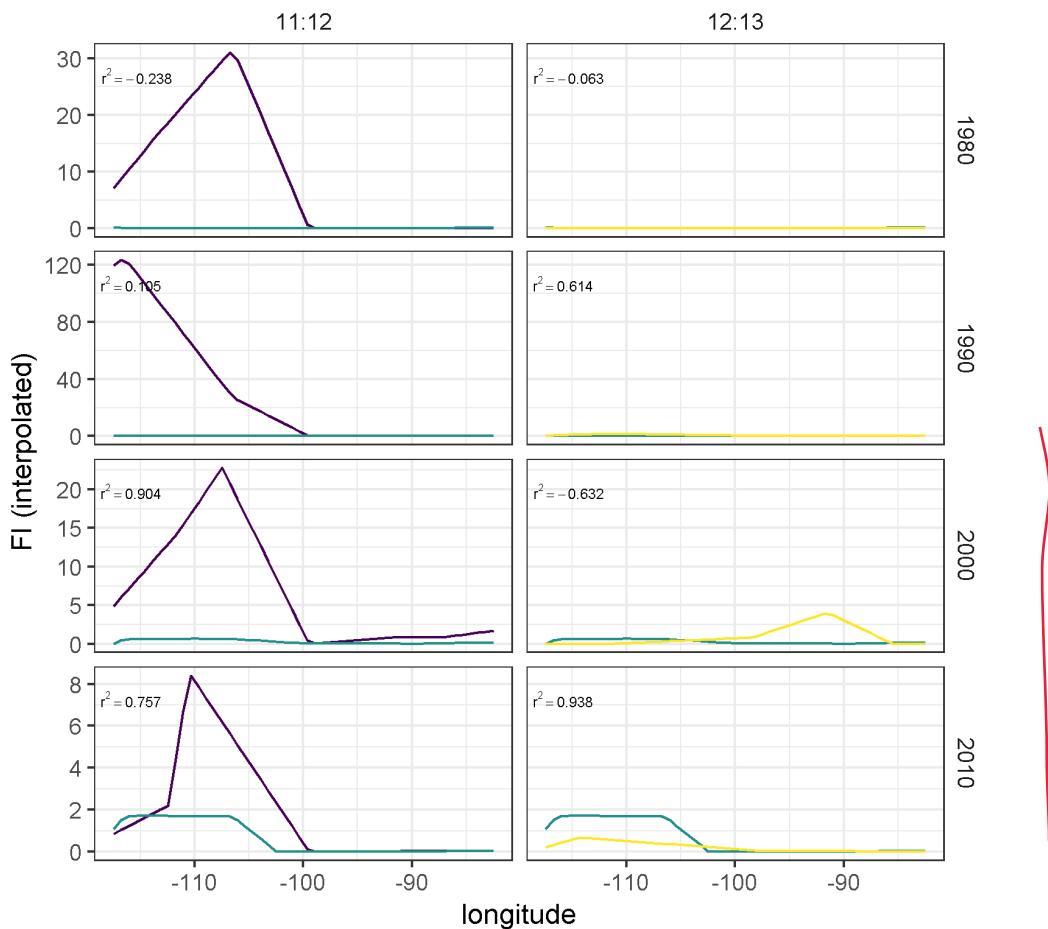


Figure 4.11: Pairwise relationships of Fisher Information (interpolated values) of spatially adjacent transects over time do not exhibit expected patterns of high positive correlation.

1593 spatial transects were separated by an East-West running physical or functional bound-
 1594 ary. This is not, however, what I expect in our East-West running transects (Figure
 1595 4.5), as the spatial soft-boundaries limiting the distribution and functional potential of
 1596 avian communities are largely North-South (Figure @ref(fig:ewRoutes_ecoRegions)).
 1597 Note spatial transects in Figure @ref(fig:ewRoutes_ecoRegions) overlap multiple, large
 1598 spatial ecoregion boundaries, such that we should expect our data to identify these
 1599 points (boundaries).

1600 Upon initial investigation, there are no obvious signs of broad-scale patterns in FI

1601 across space (Figure 4.10)². If Fisher Information is an indicator of spatial regime
1602 boundaries, we should expect to see large changes in its value (in either direction)
1603 near the edges of functional spatial boundaries (e.g., at the boundaries of ecoregions).

1604 No clear regime changes appeared in areas where we might expect rapid changes (e.g.,
1605 along the 105th meridian West, where a sharp change in altitude occurs).

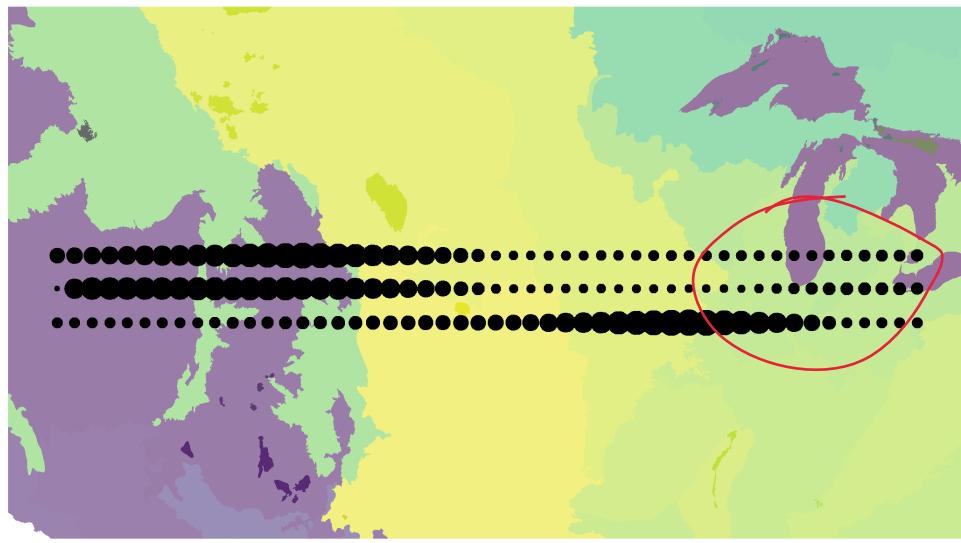
1606 Numerical investigation of the spatial correlation among adjacent transects also yielded
1607 no clear patterns. I did not identify any obvious correlation with changes in FI values
1608 and functional potential (using Omernick Ecoregion Level 2; see Figure 4.10). Rather
1609 than abrupt changes in Fisher Information I found gradual changes (e.g., see ~~some~~
1610 for years 2000 and 2010 in Figs. 4.10 and 4.12.

1611 4.4 Discussion

1612 The Fisher Information measure was introduced as a method to avoid analytical issues
1613 related to complex and noisy ecological data (Fath *et al.*, 2003; Karunanithi *et al.*,
1614 2008) and was recently suggested as an indicator of *spatial* regimes (Sundstrom *et*
1615 *al.*, 2017; Eason *et al.*, 2019). Using this method (Eq. (4.4)), I found no evidence
1616 of spatial regime shifts in the avifauna in my study area. Further, the absence of
1617 autocorrelation among spatially adjacent transects suggests Fisher Information may
1618 not be a reliable indicator of changes in bird community structure.

1619 Although the Fisher Information equation (Eq. (4.4)) used in this study is a relatively
1620 straightforward and fairly inexpensive computational calculation, extreme care should
1621 be taken when applying this index to empirical data. Fisher Information is capable of
1622 handling an infinite number of inputs (variables) and, given sufficiently low window
1623 size parameters, can technically calculate an index value for only two observations. It

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



BBS
rovers
in lake?

Figure 4.12: Fisher Information (scaled and centered; point size positively correlated with value) against ecoregion boundaries (EPA Level 2).

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):

1. the order or state(s) (s) of the system is observable,

- 1632 2. any observable change in the information observed in the data represents reality
1633 and the variables used in the analyses will not produce false negatives, and
1634 3. changes in I presumed to be regime shifts do not represent the peaks of cyclic
1635 (periodic) patterns.

1636 The first assumption is one of philosophical debate and is thus not controllable. To
1637 attempt to control for false negatives, the user should take caution in her choice of input
1638 variables. In the the case of a high dimensional data, relativization and/or variable
1639 reduction measures may be useful (Rodionov 2005). However, Fisher Information
1640 does not convey information on how specific variables relate to the calculated index.
1641 Finally, we can take measures to account for cyclic behavior in the data by ensuring
1642 integration periods capture at one full cycle of the system and, given sufficiently high
1643 number of observations, increasing the integration period may also alleviate some
1644 issues related to irreducible error, or white noise.

1645 The lack of patterns identified using Fisher Information may be influenced by one or
1646 more of the following: (1) the Breeding Bird Survey data collection scheme was designed
1647 to estimate and track **species** trends and not changes in entire communities; (2) these
1648 data consist of < 50 time points, and for some BBS routes much fewer. Ecological
1649 processes affecting large regions in this study area (e.g., the Central Great Plains)
1650 operate on larger time scales (i.e., » 50 points). A mismatch among the ecologically
1651 relevant scales and the temporal resolution and extent of our data may influence the
1652 ability of this index to capture large-scale changes in whole bird communities.

1653 Aside from the typical biases associated with the BBS data (e.g., species detection
1654 probability, observer bias), there are additional considerations to be made when using
1655 these data to identify ‘spatial regime shifts’. Breeding Bird Survey routes are spaced
1656 apart so as to reduce the probability of observing the same individuals, but birds
1657 which fly (especially in large flocks) overhead to foraging or roosting sites have a

*Not relevant
for SRC*

1658 higher probability of being detected on multiple routes. We have, however, removed
1659 these species (waders, shorebirds, waterfowl, herons) from analysis. Regardless, this
1660 study assumes there is potential for each unique BBS route to represent its own state.
1661 If routes were closer together, it is more probable that the same type and number of
1662 species would be identified on adjacent routes. Therefore, if this method does not
1663 detect slight changes in nearby routes which occupy the same ‘regime’, then it follows
1664 that the method is sensitive to loss or inclusion of new species, which are spatially
1665 bounded by geological and vegetative characteristics. What new information does this
1666 give us about the system? Fisher Information reduces and removes the dimensionality
1667 of these middle-numbered systems, which omits critical information.

1668 Effective regime detection measures should provide sufficient evidence of the drivers
1669 and/or pressures associated with the identified regime shifts (Mac Nally *et al.*, 2014).
1670 The Fisher Information index collapses a wealth of data into a single metric, thereby
1671 foregoing the ability to relate state variables to the observed changes in Fisher
1672 Information, unlike other dimension reduction techniques. For example, loadings, or
1673 the relative influence of variables on the ordinate axes, can be derived from a Principal
1674 Components Analysis—this cannot be achieved using Fisher Information. If Fisher
1675 Information clearly suggested a spatial regime boundary or shift, a before-and-after
1676 post-hoc analysis of the regional community dynamics might confirm the regime shift
1677 occurrence.

1678 This study found no evidence suggesting Fisher Information accurately and consistently
1679 detects spatial boundaries of avian communities. Rapid changes in either direction
1680 of Fisher Information is suggested to indicate of a regime shift (Mayer *et al.*, 2006;
1681 Eason & Cabezas, 2012). Although this interpretation has been applied to multiple
1682 case studies of Fisher Information, there is yet a statistical indicator to objectively
1683 identify these abrupt changes. After calculating the Fisher Information for each

1684 spatial transect (Figure 4.5) during each sampling year, I used pairwise correlation to
1685 determine whether spatial autocorrelation existed among pairs of spatial transects.
1686 If some set of points are close in space and are *not* separated by some physical or
1687 functional boundary (e.g., an ecotone, high altitude rock formations), then the Fisher
1688 Information calculate should exhibit a relatively high degree of spatial autocorrelation
1689 that is consistent over time. It follows that the correlation coefficient of spatially
1690 adjacent transects should be similar, diverging only as the distance between the
1691 transects differs and/or a functional or physical boundary separates them.

1692 Several questions remain regarding the efficacy of Fisher Information as a robust
1693 regime detection measure in both spatial and temporal empirical ecological data. If
1694 signals of regime shifts do exist, this results of this study suggest it is not possible to
1695 identify them using visual interpretation of Fisher Information. This study also did
1696 not identify evidence spatial autocorrelation of the FI values, further supporting this
1697 claim. I suggest future studies of Fisher Information focuses on temporal, rather than
1698 spatial data. Potential areas of research and questions include:

- 1699 1. Sensitivity of Fisher Information to data quality and quantity (this is explored
1700 in Chapter 6).
- 1701 2. What, if any, advantages does FI have over other density estimation techniques?
- 1702 3. Does FI provide signals in addition to or different than geophysical and vegetative
1703 (e.g. LIDAR) observations (data)?
- 1704 4. Relationship of Fisher Information to likelihood ratio-based unsupervised change-
1705 point detection algorithms (e.g., ChangeFinder; Liu *et al.*, 2013).
- 1706 5. How does Fisher Information perform relative to other regime detection measures
1707 (see Chapter 6)?

1708 This study provided an objective evaluation of the Fisher Information metric as a
1709 spatial regime detection measure. Future studies exploring the aforementioned areas

₁₇₁₀ can provide further insight into the efficacy of this method as a spatial and/or temporal

₁₇₁₁ method.

1712 **Chapter 5**

1713 **Velocity (v): using rate-of-change
1714 of system trajectory to identify
1715 abrupt changes**

1716 **5.1 Introduction**

1717 When, how and why ecological systems exhibit abrupt changes is a hallmark of
1718 modern ecological research, and changes which are unexpected and undesirable can
1719 have undesirable downstream consequences on, e.g., ecosystem services, biodiversity, and
1720 human well-being. Quantitatively detecting and forecasting these changes, however,
1721 has yet to be accomplished for most ecological systems (Chapter 2; Ratajczak *et al.*,
1722 2018). Moving from abrupt change methods requiring highly descriptive models and *a*
1723 *priori* assumptions of the state variable responses to drivers to methods requiring few,
1724 if any, *a priori* assumptions or knowledge is increasingly necessary for forecasting and
1725 managing complex ecosystems under an era of intensifying anthropogenic pressures. A
1726 few broad classes of quantitative approaches exist for quantitatively identifying abrupt

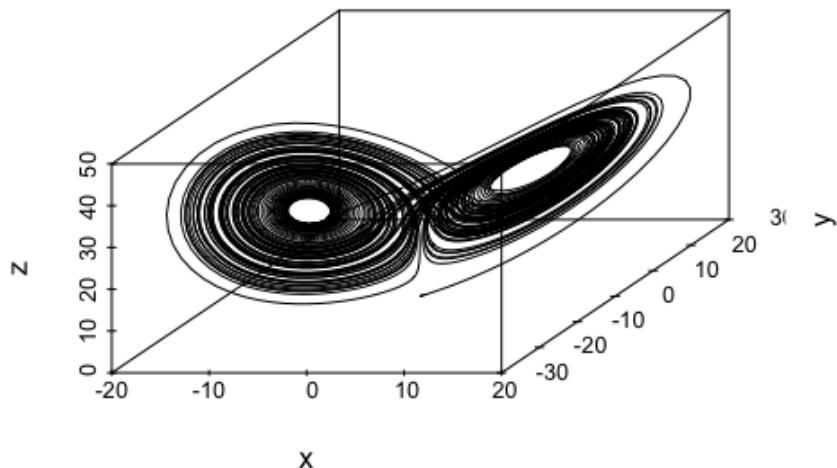


Figure 5.1: An example solution of the Lorenz ('butterfly') represented in 3-dimensional phase-space. Phase plots are typically used to visualize stable areas within a system's trajectory but reconstruction requires the difference models to be known and parameterized.

1727 changes in complex ecosystems. First, one can use simple mathematical models to
1728 describe the system and statistically test for discontinuities in the observed variables
1729 (e.g., in coral reefs, Mumby *et al.*, 2013). Although mathematical representations are
1730 ideal, very rarely are ecological systems easily and well-described by them and often
1731 fail to meet the assumptions of the model. Second, we can track changes in the mean
1732 or variance of state variables to identify departures from the norm (e.g., early-warning
1733 indicators such as variance and variance index, Brock & Carpenter, 2006). Much like
1734 the mathematical modelling approach, these early-warning indicators have shown to be
1735 useful in some simple driver-response systems (e.g., lake eutrophication Carpenter *et*
1736 *al.*, 2008), but are unreliable in other empirical systems (e.g., Perretti & Munch, 2012;
1737 Dakos *et al.*, 2012b; Dutta *et al.*, 2018). The last type of approach is the model-free
1738 approach [Dakos *et al.* (2012a); Chapter 2]. This group of abrupt change indicators
1739 can incorporate multiple state variables, and ideally requires no *a priori* assumptions
1740 about the expected driver-response relationships, or even about the drivers at all. It
1741 is this class of abrupt change indicators to which this chapter contributes.

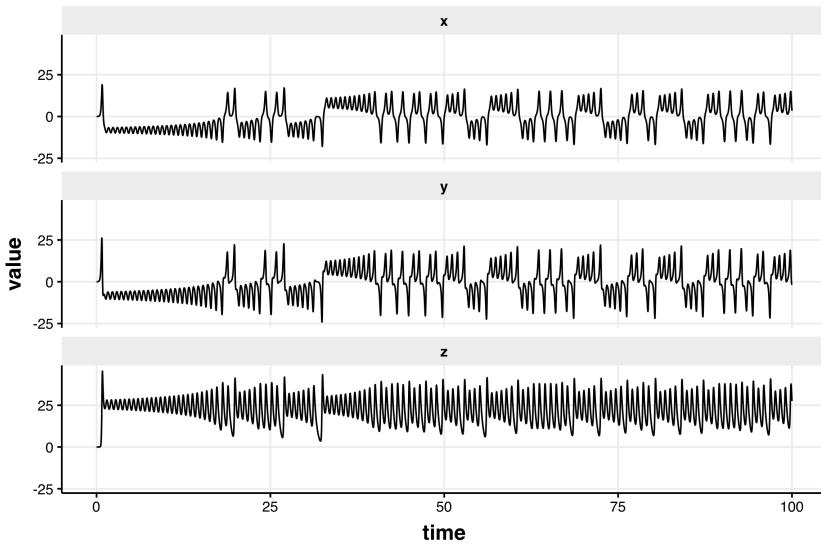


Figure 5.2: An example solution of the Lorenz ('butterfly') represented in individual system components.

1742 **5.1.1 Tracking ecosystem trajectory through time to explore
1743 system dynamics**

1744 A classic example of state-switching by a system is demonstrated in the Lorenz
1745 ('butterfly') attractor (Figure 5.1; Takens, 1981). This phase plot (Figure 5.1) provides
1746 an informative visual of the behavior of a chaotic system manifesting two attractors.
1747 Although the periodic, attractor behaviors are made clear when examining the time
1748 series of each dimension (Fig 5.2), identifying such behaviors in additional dimensions
1749 becomes increasingly difficult.

1750 System behavior/trajectory in phase space are used often in dynamical systems theory
1751 and systems ecology to make inference regarding system behavior and dynamics,
1752 but phase space (trajectory) dynamics are not commonly applied outside theoretical
1753 studies as a tool for ecological data analysis (c.f. Sugihara *et al.*, 2012 for an example
1754 of phase-space reconstruction using Taken's theorem of ecological time series). Some
1755 methods of attractor reconstruction have been applied to environmental data (e.g.,
1756 individual time series of fisheries stocks, climate, stock market; Sugihara *et al.*,

1757 2012; Ye *et al.*, 2015), yet they **do not incorporate the dynamics of whole-**
1758 **systems.** Model-free methods for exploring and describing the dynamics of whole
1759 (i.e. > 1 variable) ecological systems are restricted to the commonly-applied **dimnesion**
1760 reduction techniques and clustering algorithms (e.g., Principal Components Analysis,
1761 K-means clustering). In fact, this is true of many abrupt change and regime shift
1762 indicators.

1763 **5.1.2 Rate of change as an indicator of abrupt change in the**
1764 **system trajectory**

1765 How quickly a system switches states [e.g., moving from attractor to another; 5.1] may
1766 yield insights into the responses of ecological systems to perturbations (e.g., anthro-
1767 pogenically induced pressures such as climate change, urbanization) and community
1768 shifts (e.g., species introductions or extinctions, shifts in dominance). For example,
1769 Beck *et al.* (2018) tracked rate of change using chord distances—a data transformation
1770 for positive values and which is suitable prior to ordination analysis—to capture
1771 abrupt changes in community composition of a temperate, paleodiatom community.
1772 Chord distance, however, is greatest when the observations among data rows (e.g.,
1773 time, location) have no species in common. In other words, this measurement may be
1774 most useful in high community turnover conditions. Identifying alternative numerical
1775 methods for estimating system rates of change may be when the system does not
1776 exhibit, for example, high degrees of turnover.

1777 Rate of change (ROC, often represented as Δ) is a term used for various measures
1778 which describe the relationship among to variables, measuring the change in one
1779 variable relative to another. As a refresher ROC is represented as **speed (S)** or
1780 **velocity (V)**, where (**S**) is the adirectional magnitude (i.e. it is a scalar) of the
1781 displacement of an object over unit time and **V** describes both the direction and

1782 magnitude (i.e. it is a vector) of the object's movement in spacetime. \mathbf{S} is a scalar
1783 taking values of ≥ 0 and \mathbf{V} can take any value between $-\infty$ and ∞ . For example,
1784 consider a car travelling at a constant speed of $50 \frac{\text{km}}{\text{h}}$ around along a hilly landscape,
1785 where it is ascending and descending hills. Although \mathbf{S} is constant, \mathbf{V} changes in a
1786 sunusoidal fashion, where \mathbf{V} is $\mathbf{V} > 0$ when ascending, $\mathbf{V} < 0$ when descending, and
1787 $\mathbf{V} \approx 0$ at in the valleys and at the peaks of the hills. Although \mathbf{S} is useful when
1788 estimating other scalar quantities (e.g., $\frac{\text{miles}}{\text{gallon}}$), given a starting and/or final position
1789 in space, \mathbf{S} is not informative of its the path traveled.

1790 5.1.3 Aims

1791 Here, I propose a method which simply describes the rate of change behavior of
1792 system dynamics in phase space: **velocity**, V . An alternative to other ~~complicated~~,
1793 model-free approaches (e.g., Fisher Information; Cabezas & Fath, 2002), the velocity
1794 metric allows one to examine the behavior of an entire system along its trajectory
1795 (through space or time) without having to reconstruct the pahse space. The ability
1796 to handle noisy and high-dimensional data and the lack of subjective parameters in
1797 calculating the metric makes this method ~~an ideal~~ alternative to existing early warning
1798 indicators and phase-space reconstruction methods. *objektive*

1799 I first describe the steps for calculating this new metric (v), as both a dimension
1800 reduction technique and abrupt change indicator. Although this is the first instance
1801 of this calculation to, alone, be suggested as a regime detection metric, it has been
1802 used as part of a larger series of calculations of the Fisher Information metric [see
1803 Chapter 3], first introduced in Fath *et al.* (2003). I use this theoretical system to
1804 present baseline estimates of the expected behavior of v under various scenarios of
1805 changing mean and variability in a theoretical, discussing the contexts under which
1806 this metric may signal abrupt changes. Finally, I explore the utility of this metric in

1807 identifying known regime shifts in an empirical paleoecological time series data.

1808 **5.1.4 Analytical approach**

1809 I first describe the steps for calculating velocity by constructing a simple, two-variable
1810 system which exhibits only a rapid, discontinuous change in the means of the state
1811 variables. I next vary the mean and variance of the state variables of this system
1812 to demonstrate baseline expectations for the behavior of velocity under a simple
1813 rapid shift scenario. Next, I construct a second model system similar to the first,
1814 but one which exhibits a non-discontinuous rapid change in the state variables. The
1815 purpose of this section is three-fold. First, I demonstrate how velocity behaves when
1816 the system undergoes varying degrees of change (e.g., slow change versus nearly
1817 discontinuous, rapid). Second, I concurrently identify baseline expectations of velocity
1818 under varying conditions of mean and variability of the state variables before and after
1819 a shift. Third, by introducing a smoothing function to the rapid shift, we gain an
1820 understanding of how process variability (noise) impacts the shift detectability by the
1821 velocity metric. Finally, I calculate the velocity of an empirical, paleolithic freshwater
1822 diatom community time series to demonstrate the utility of the velocity metric in
1823 highly noisy, high dimensional, and irregularly-sampled data.

1824 **5.2 Steps for Calculating velocity, v**

1825 In this section, I first demonstrate the calculations of velocity using a very simple,
1826 two-variable toy system. The first system exhibits a rapid shift at a single point
1827 in time, where mean and variance are constant before and after the shift point. I
1828 demonstrate the signals achieved with and the variability within the v calculation
1829 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 μ is the mean and σ is the standard deviation. Consider a system (Figure

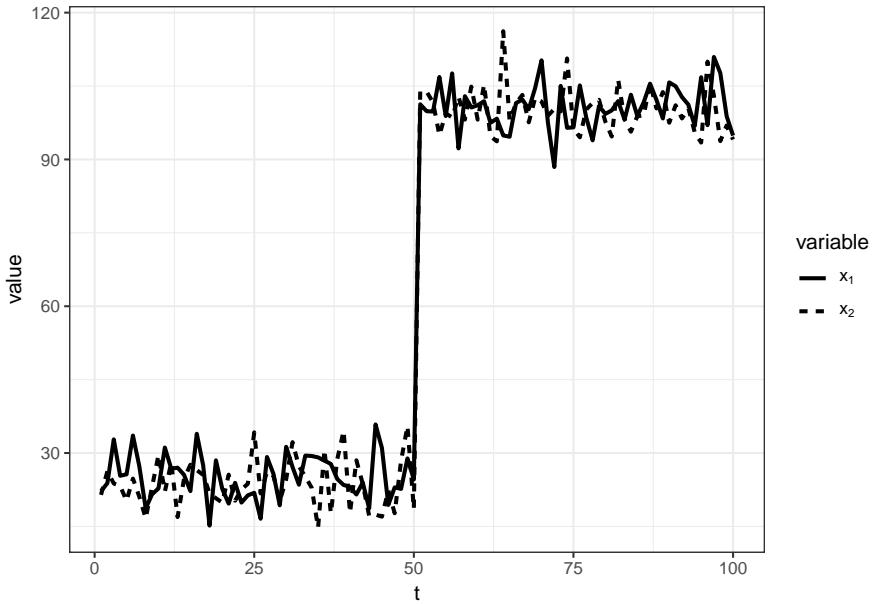


Figure 5.3: The 2-variable discrete time toy system used to demonstrate steps for calculating system velocity. Each variable, x , is drawn from a normal distribution with means that change at $t = 50$. State variables have constant standard deviation, $\sigma = 5$.

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 σ_i (sd) for 100 time steps, t .

¹⁸⁴³ The regime shift in this system occurs at $t = 50$, where a shift in either or both \bar{x}_i or σ_i .

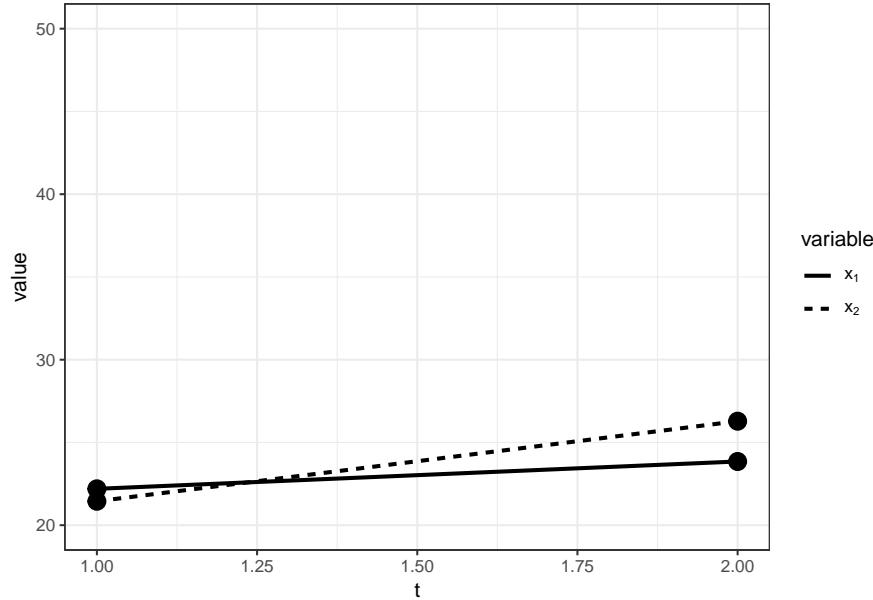


Figure 5.4: Data used to calculate velocity at the first two time points, t_1 and t_2 .

¹⁸⁴⁴

¹⁸⁴⁵ 5.2.1 Steps for calculating v

¹⁸⁴⁶ Step 1: Calculate Δx_i

¹⁸⁴⁷ The first step is to calculate the change in values for each state variables, x_i , between
¹⁸⁴⁸ two consecutive time points [e.g., from time t to $t + 1$ for the discrete-time system;
¹⁸⁴⁹ Figure 5.4; Equation (5.1)]:

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

¹⁸⁵⁰ Note that Δx_i can take any value between $-\infty$ and ∞ .

1851 **Step 2: Calculate distance traveled, s**

1852 Next, we calculate the total change in the multivariable system as a function of the
1853 change in all state variables x_i . First, we calculate Δs as the square root of the sum
1854 of squares of the changes in all state variables per Pythagora's theorem [Equation
1855 (5.2)]:

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

1856 Although Δs represents the absolute change in the system between consecutive points
1857 in time, this measure is not yet relative along the system's trajectory. To create a
1858 relative value we next calculate the total distance traveled along the system trajectory,
1859 s , as the cumulative sum of Δs [Equation (5.2)] since the first observation, such that a
1860 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

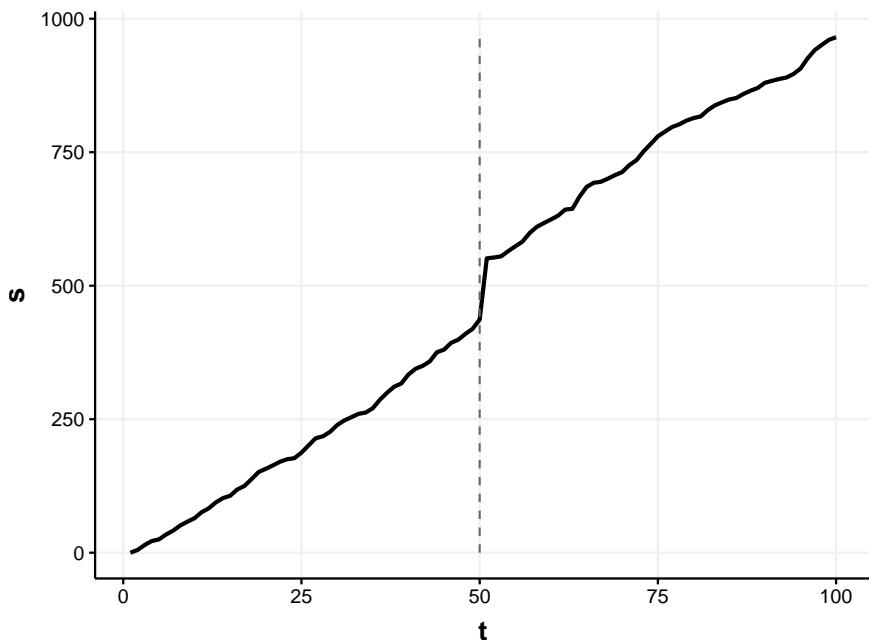


Figure 5.5: Distance traveled, s , for the 2-species toy system.

1861 discrete point in time in our N -dimensional system (Figure 5.5). It should be noted
1862 that s (Figure 5.5) is monotonically increasing since the value of Δs [Equation (5.2)]
1863 is a sum of squares. Although discussed in a later section, it is important to note that
1864 s is not unitless—that is, s has units of the state variables, x_i . For example, if our
1865 2-variable toy system represents biomass, then the units of s represents the cumulative
1866 absolute change in biomass of the entire system.

1867 **Step 3: Calculate velocity, v (or $\frac{\Delta s}{\Delta t}$)**

1868 Finally, we calculate the **system velocity**, v (or $\frac{\Delta s}{\Delta t}$), by first calculating the change
1869 in s [Equation (5.3)], and then divide by the total time elapsed between consecutive
1870 sampling points:

$$v = \frac{s_{t+1} - s_t}{\Delta t} \quad (5.4)$$

The numerical results for each step in the calculation of velocity [Equation (5.4)] is

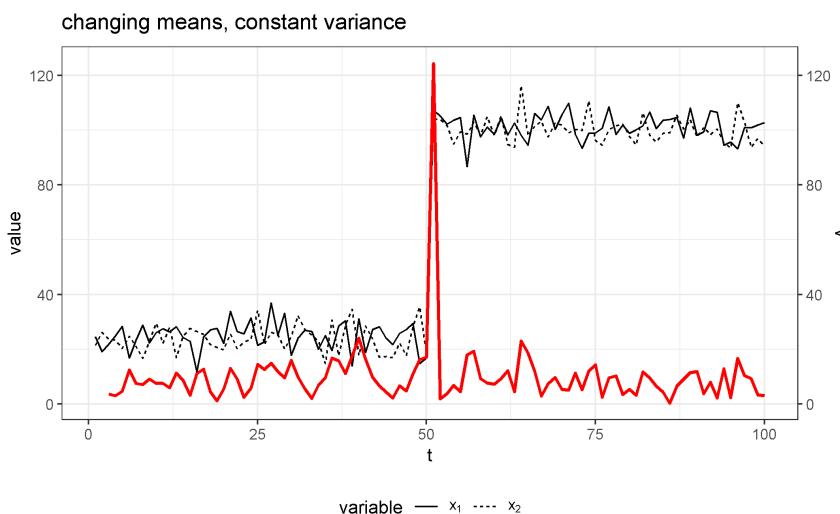


Figure 5.6: System change (s) and velocity (v) of the model system over the time period. Constant means ($\bar{x}_{pre} = 25$, $\bar{x}_{post} = 10$) and sharp change in variance for both state variables, $\sigma = 5$.

1871

1872 demonstrated using the first five time points of our toy system (Figure 5.3) in Table

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

t	x_1	x_2	Δx_1	Δx_2	Δt	$\sqrt{(\sum_{i=1}^N \Delta x_i^2)}$	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

1873 5.1.

1874 5.3 Velocity v performance under a discontinuous 1875 transition

1876 I used simulation techniques to determine the baseline expectations of the performance
1877 of velocity v under varying degrees of rapid shifts in the mean and variance of the toy
1878 system. The toy system in this section undergoes a discontinuous shift at $t = 50$ (see
1879 5.3). If the system undergoes a rapid and discontinuous change in one or more state
1880 variables, the velocity, because it is a rate of change, may $\rightarrow \infty$ as $\Delta t \rightarrow 0$. Therefore,
1881 it is important to understand the degree to which velocity can detect very sudden
1882 changes in mean values, despite effect sizes. Here, I varied each of the following system
1883 parameters at the regime shift location ($t = 50$): \bar{x}_1 , increase in the mean value of x_1
1884 and σ_1 , the change in variance of x_1 .

1885 Simulations consisted of 10,000 random samples drawn from the normal distribution for
1886 each parameter, I randomly drew the toy system samples 10,000 times under increasing
1887 values of \bar{x}_1 and σ_1 . To identify patterns in the influence of parameter values on velocity,
1888 I present the mean values of v across all simulations, with confidence intervals of ± 2
1889 standard deviations. As mentioned above, the state variables x_1 and x_2 were drawn

1890 from a normal distribution (using function *rnorm*), with parameters \bar{x}_i (mean) and σ_i
1891 (sd) for 50 time steps, t .

1892 **Varying post-shift mean**

1893 I examined the influence of the magnitude of change in x_1 in the period before (pre;
1894 $t < 50$) and after (post; $t \geq 50$) by varying the mean parameter, \bar{x}_1 in the set
1895 $W = \{25, 30, 35, \dots, 100\}$ (Figures 5.7, ??). As expected, the magnitude of v increases
1896 linearly as the total difference between $\bar{x}_{1_{pre}}$ and $\bar{x}_{1_{post}}$ increases (Figure 5.8). This is
1897 not surprising because s increases as the total change in abundance across the entire
1898 system increases [Equation (5.3)]. Consequently the potential of v also increases with
1899 total state variable values (e.g. abundance, biomass). The linear relationship among v
1900 and total state variable values indicates that while v is capable of identifying large
1901 shifts in data structure, it may fail to identify subtle changes (i.e. lower effect sizes).

1902

1903 **Varying post-shift variance**

1904 In the previous example, variance was constant before and after the abrupt shift at
1905 $t = 50$. To determine whether the signal emitted by v at the regime shift is lost or
1906 dampened when increasing variance I varied the variance parameter, σ_1 along the set
1907 $W = \{1, 2, 3, \dots, 25\}$. The variance for both state variables (x_1, x_2) prior to the regime
1908 shift, σ_{x_1} and σ_{x_2} , was 5, with the change occurring in σ_{x1post} .

1909 **Smoothing the data prior to calculating v**

1910 To determine whether process or observational noise influences the signal in v , I used
1911 linear approximation techniques to smooth the data prior to calculating the derivatives.

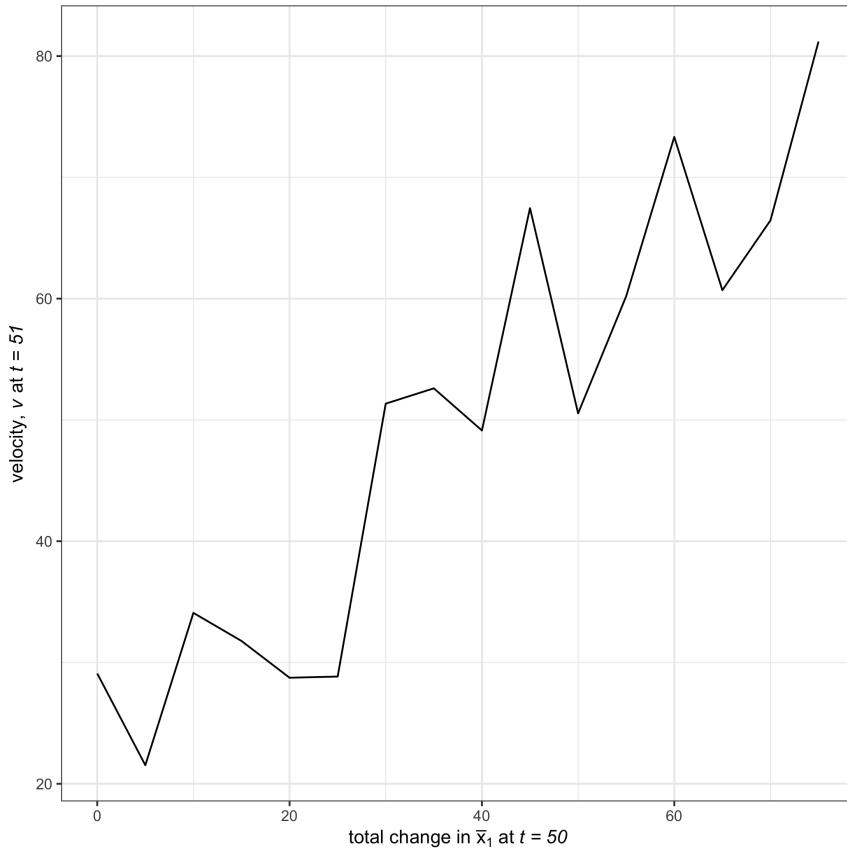


Figure 5.7: Velocity (v) generally increases as the total change in the mean value of $\bar{x}_{1_{t=50}}$ increases in a single iteration of our toy system ($N_{iter} = 1$, seed = 123). This 2-variable system exhibits a regime shift at $t = 50$, where variance is constant $\sigma = 5$, $\bar{x}_1 =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

1912 I used the function `stats::approx` which linearly interpolates the original data, x_1
 1913 and x_2 , to regularly-spaced time points along the set $t = \{1 : 100\}$. I then calculated
 1914 v as described in (Eqs. (5.1) through (5.4)). Increasing the number of points (t) at
 1915 which the original state variables were smoothed (i.e., t) did not influence the amount
 1916 of noise surrounding the signal of the regime shift (at $t = 50$) in system velocity, v
 1917 (Figure 5.10).

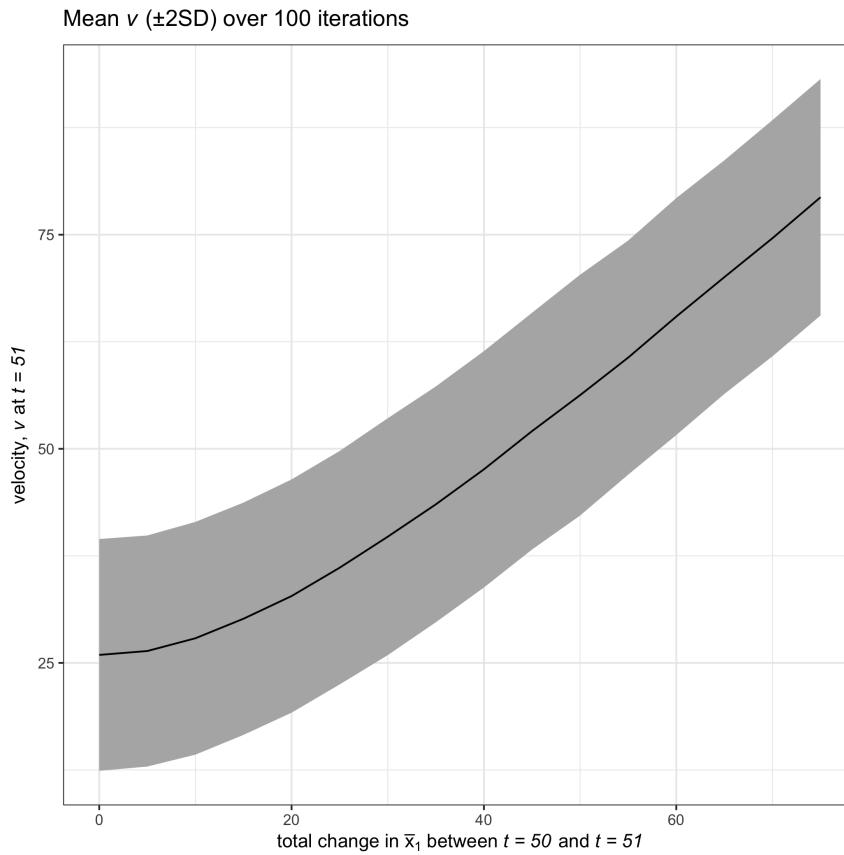


Figure 5.8: Change in velocity (v) as the total change in the mean value of $\bar{x}_{2_{t=50}}$ over 10,000 simulations. A regime shift was induced at $t = 50$ with constant variance $\sigma = 5$, $\bar{x}_2 = 25$ when $t < 50$, and changes in variable mean values, $\bar{x}_2 = 50$ when $t \geq 50$, $\bar{x}_1 = 25$ when $t < 50$.

1918 **5.4 Velocity performance under a smooth transi-**
 1919 **tion**

1920 In the previous section I presented expectations for velocity signals under a discontin-
 1921 uous transition in a discrete-time system. Given velocity is a measure of the rate of
 1922 a change in a system and the range of transition speeds ecological systems exhibit
 1923 (e.g., slow driver-response or threshold dynamics), it is important to understand if
 1924 and when the velocity signal is damped under varying degrees of transition speeds.

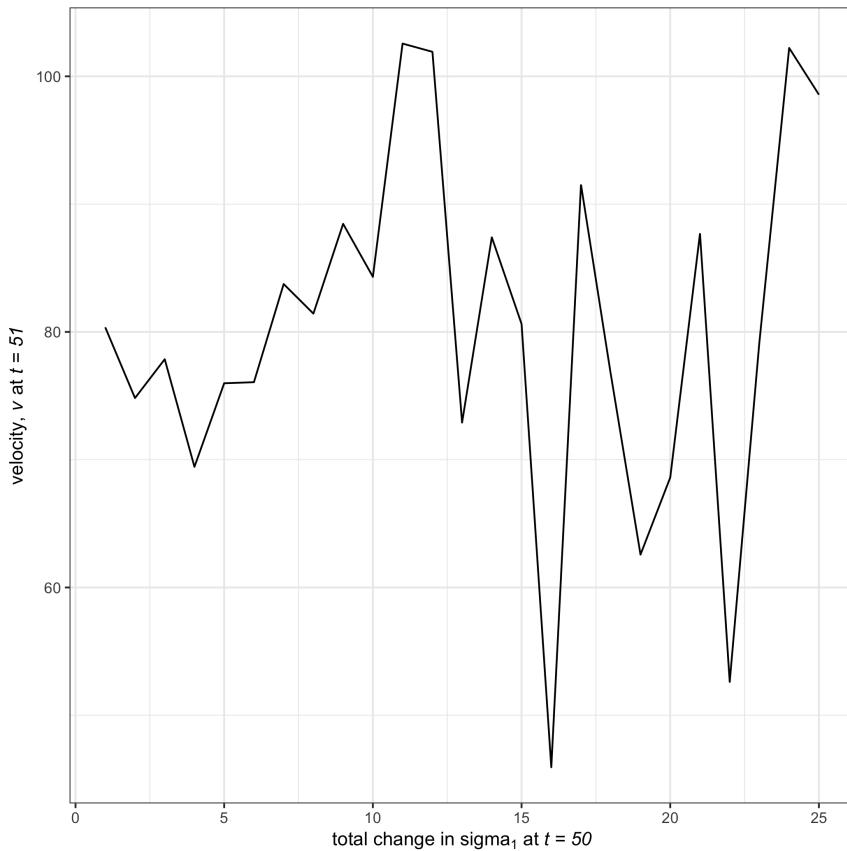


Figure 5.9: High variance of velocity (v) in a single iteration ($N_{iter} = 1$, seed = 123) of simulations as we increase σ_1 at $t = 50$.

1925 In this section I use a similar toy system, to demonstrate the expectations of velocity
 1926 under a smooth shift and under varying degrees of rapidity.

1927 Although the data constructed in this section are similar to that used in the previous
 1928 section in that we are manipulating the mean and variance of two state variables before
 1929 and/or after an abrupt shift, this section introduces a component of process noise into
 1930 the shift itself. This is important because the derivative of a nearly discontinuous
 1931 function is infinite. Although we are interested in identifying rapid shifts in systems,
 1932 velocity will approach infinity as the rate of change in the shift increases and the
 1933 sampling intervals decrease. In other words, if the system exhibits turnover in e.g. 25%
 1934 of the state variables, we expect the value of velocity to be similar to that of a turnover

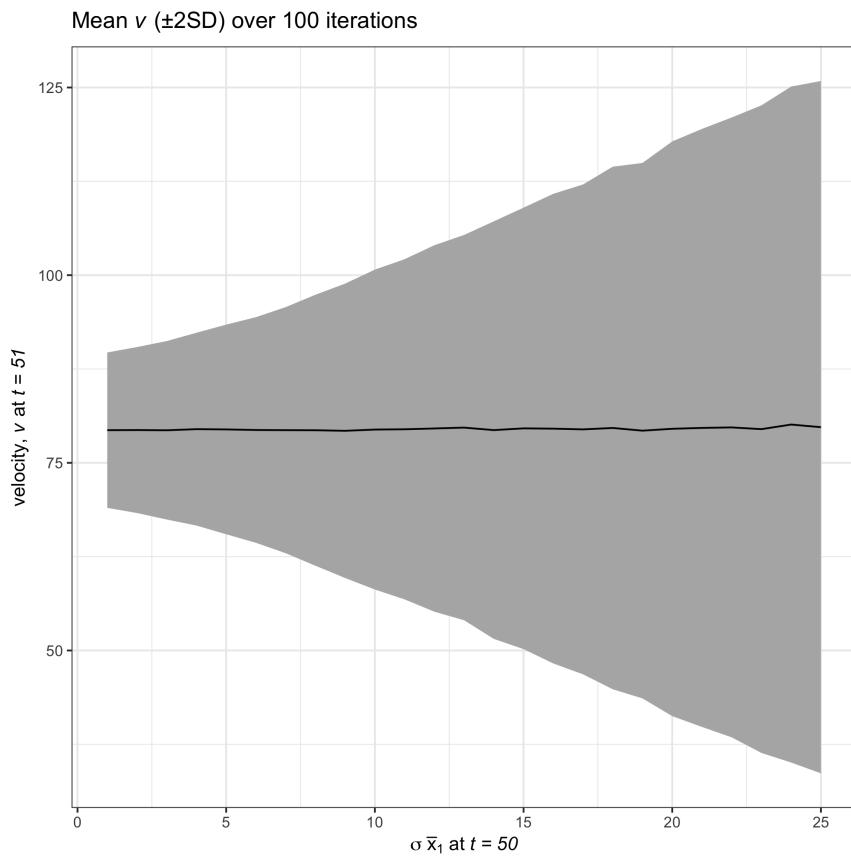


Figure 5.10: Average (± 2 SD) velocity (v) worsens as the variance of $\bar{x}_{2_{t=50(post)}}$ (post shift) increases. $\bar{x}_{1_{pre}} = 25$, $\bar{x}_{1_{post}} = 100$, $\bar{x}_{2_{pre}} = 25$, $\bar{x}_{2_{post}} = 50$, $\sigma_{1_{pre}} = 5$, $\sigma_{2_{pre,post}} = 5$

1935 in e.g. 75% of the variables. Removing the possibility of infinite values provides more
1936 relative measures within the community time series.

1937 5.4.1 Generating the data

1938 Here we consider a two-variable system over the time interval [1, 100] with state
1939 variables x_1 and x_2 which exhibits abrupt shifts in mean and/or variance of one or
1940 both variables at time $t = 50$. I generated species observations for the true process
1941 and the true process with process variability. The true process data were created
1942 using the parameters for μ and σ for each of the conditions in described in Table 5.2

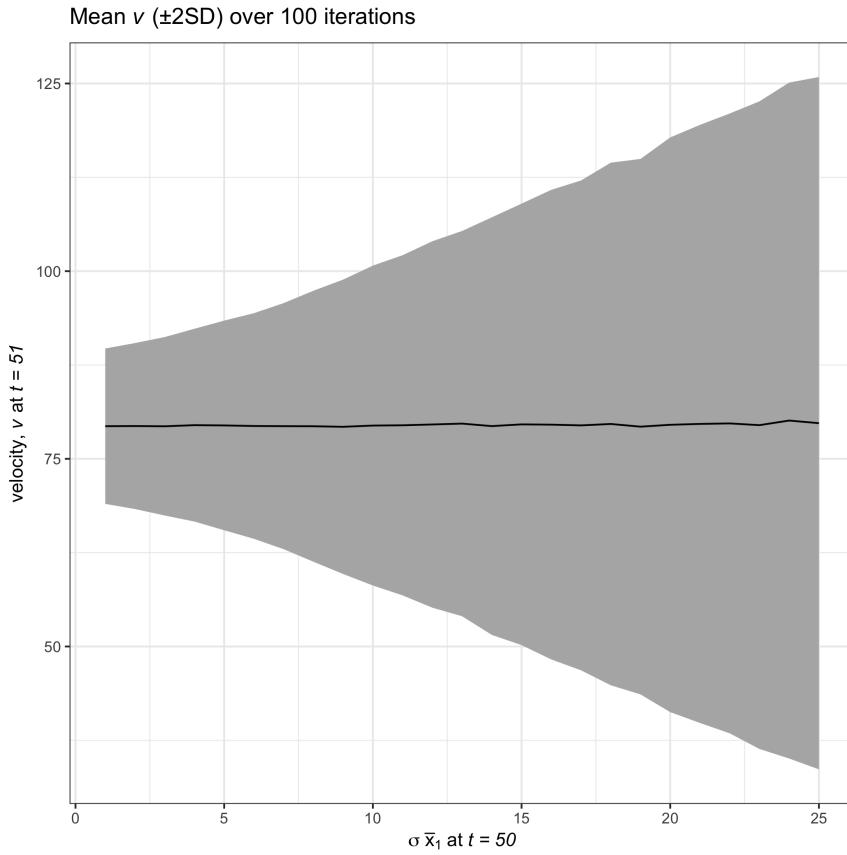


Figure 5.11: The noise in system velocity (v) is not obviously reduced in this system as the original data (x_1, x_2) is increasingly smoothed.

¹⁹⁴³ (random seed in Program R was 12345).

¹⁹⁴⁴ True process model

¹⁹⁴⁵ Data were generated from a normal distribution and an abrupt shift in the mean was
¹⁹⁴⁶ incorporated using a hyperbolic tangent function. The true process for each state
¹⁹⁴⁷ variable, x_i , was generated from [Equation (5.5); see Figure 5.13]:

$$\begin{aligned} \mu_{xipre} &\sim \text{Normal}(\mu_{xipre}, \sigma_{xipre}) \\ \mu_{xipost} &\sim \text{Normal}(\mu_{xipre}, \sigma_{xipost}) \end{aligned} \quad (5.5)$$

$$\mu_{x_i}(t) = \mu_{xipre} - 0.5(\mu_{xipre} - \mu_{xipost})(\tanh(\alpha(t - t_{shift})) + 1)$$

Table 5.2: Conditions for generating various scenarios of the hyperbolic tangent-induced abrupt change. σ_i represents the standard deviation of μ_{x_i} as the percent of μ_{x_i} , μ_{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
μ_{x_1}, σ_{x_1}	0.05	0.10	0.05	0.05	10	55	15	15
μ_{x_1}, μ_{x_2}	0.05	0.05	0.05	0.05	10	55	15	44
μ_{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
σ_{x_1}	0.05	0.10	0.05	0.05	10	10	15	15

1948 where $\mu_{x_i}(t)$ is the mean value of x_i at time t and *pre* and *post* are the periods before
1949 and after the abrupt shift (t_{shift}), respectively. The parameter α in Equation (5.5)
1950 controls for the rate of change at the point of the abrupt change, t_{shift} , where higher
1951 values of α correspond with a higher slope at t_{shift} . I simulated a single iteration
1952 (dataset) for various conditions of changing μ_{xi} and σ_{xi} (see Table 5.2), for two state
1953 variables x_1 & x_2 at intervals of $t = 1$ along the temporal interval $t = [1, 100]$.

1954 Observed process data

1955 I generated observations by imputing noise into the true process model [Equation
1956 (5.5)] through random sampling of σ_{xi} from a normal distribution [Equation (5.6);
1957 Figure 5.13]:

$$\begin{aligned}
 \mu_{xipre} &\sim Normal(\mu_{xipre}, \sigma_{xipre}) \\
 \sigma_{xipre} &\sim Normal(0, \sigma_{Xipre}\mu_{Xipre}) \\
 \mu_{xipost} &\sim Normal(\mu_{xipost}, \sigma_{xipost}) \\
 \sigma_{xipost} &\sim Normal(0, \sigma_{Xipost}\mu_{xipost})
 \end{aligned} \tag{5.6}$$

$$\mu_{xi}(t) = \mu_{xi}pre - 0.5(\mu_{xipre} - \mu_{xipost})(\tanh(\alpha(t - t_{shift})) + 1)$$

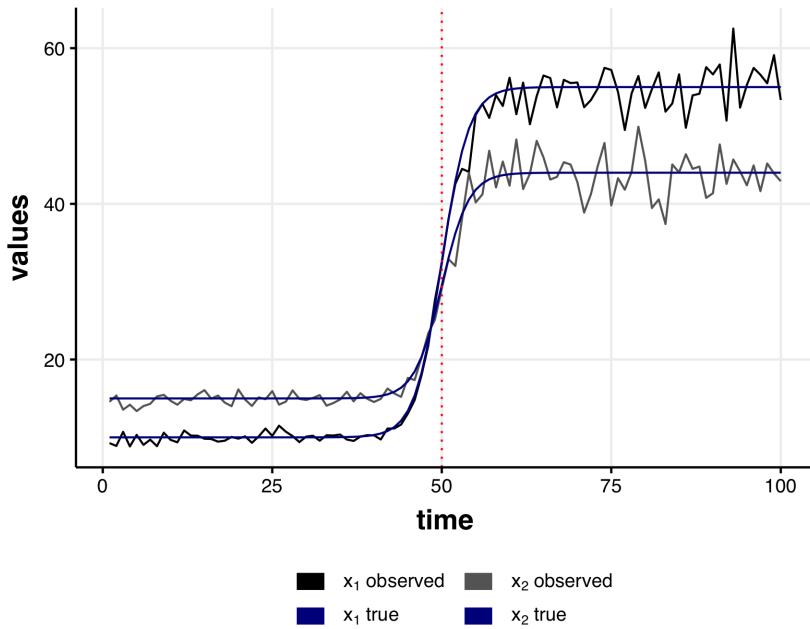


Figure 5.12: An example of the data generated by the true process model. In this example the mean values (μ_{xi}), but not the percent standard deviation (σ_{xi}), are varied before and after the transition point. The observed data are plotted against the true-process model for each state variable, x_i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

1958 where σ_{xi} is the observed error around μ_{xi} , and σ_{Xi} is X% of μ_{xi} under various
 1959 sampling conditions (as described in Table 5.2). I generated the error as a percent of
 1960 the mean as this scaling relationship is commonly observed in ecological data (Taylor,
 1961 1961).

1962 **5.4.2 Evaluating velocity performance under conditions of
 1963 changing means and/or variance**

1964 I simulated a single dataset (using in Program R) by randomly drawing a single
 1965 realisation (observed data) of the hyperbolic tangent process model with additive
 1966 noise process [Equation (5.6)]. I then calculated the distance traveled, s , and the

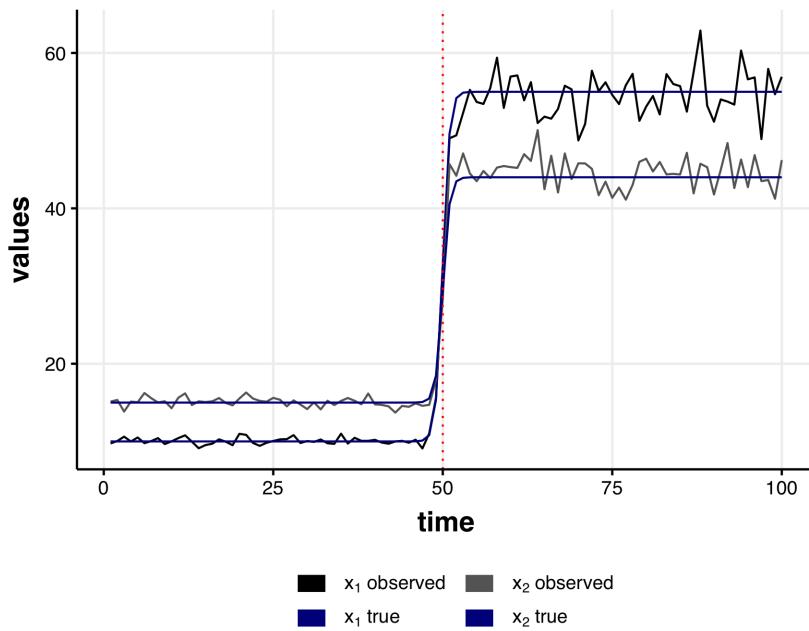


Figure 5.13: An example of the data generated by the true process model. In this example the mean values (μ_{xi}), but not the percent standard deviation (σ_{xi}), are varied before and after the transition point. The observed data are plotted against the true-process model for each state variable, x_i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

1967 velocity of the distance traveled, v (also referred to as $\frac{ds}{dt}$) using first differences. The
 1968 first differences approach is a simple alternative to numerical integration tecnicas,
 1969 requiring only simple algebraic techniques. This method is ideal for discrete time data,
 1970 or where computational power woudl not suffice for numerical integration. When using
 1971 the first differences method, however, v will demonstrate high variability, depending
 1972 on the amount of time between samples (i.e. as the intervals of $t - t + 1$ increase).

1973 I also calculated v using a numerical integration method for non-smooth, noisy
 1974 data, called total variation regularized differentiation (Chartrand, 2011). I used the
 1975 R package `tvdiff` (Price & Burnett, 2019) to perform numerically integrate the
 1976 distance traveled, s . The regularized differentiation method in this package (function
 1977 `tbdiff::tvRegDiff`; described fully in Chartrand, 2011) provides a numerical solution

1978 for calculating non-noisy derivatives of noisy, non-smooth data. Using this smooth-
1979 derivative estimation technique may be an ideal supplement to the velocity method
1980 in cases where process and observational error generate noisy observational data.
1981 Although not possible in most ecological systems data, here we can compare the
1982 fit of the smooth-derivative to the derivative of the true process, allowing us to
1983 determine the usefulness of calculating a smooth-derivative. There are two tuning
1984 ~~paramters~~ required to be chosen by the analyst when implementing the total-variation
1985 regularized differentiation, each of which influence the amount of noise smoothed
1986 out in the resulting derivative: α and the number of iterations. I implemented
1987 this numerical differentiation over 1,000 iterations, and selected α by comparing the
1988 antiderivatived distance traveled, s , to the true process values of s (e.g., see Figure
1989 5.14). For most conditions and smoothness I found the tuning parameter for `tvdiff`
1990 $\alpha = 0.50$ provided a good fit of s (Figure 5.14), however, when the hyperbolic tangent
1991 smoothing paramter, α was low (i.e. $\alpha_{tanh} = 0.25$) higher values of α_{tvdiff} yielded
1992 more abrupt changes in the derivative.

1993 **Smooth changes in the mean**

1994 As discussed earlier, the velocity of the distance traveled, v , is a measure of how quickly
1995 the sum of the squared system variables change between observations (i.e. time).
1996 Consequently, as the total change in state variables grows, so will the maximum
1997 potential of the velocity, v . Following this logic, we should expect to see a spike in
1998 the derivative of the distance traveled when the system changes quickly. I tested this
1999 hypothesis under two conditions of changing means, where either one or both variables
2000 underwent mean shifts (see Table 5.2), and under varying degrees of transition
2001 smoothness (i.e. $\alpha_{tanh} = 0.25, 0.50, 0.75, 1.00$). When the hyperbolic tangent
2002 smooth transition function is less steep (Figure 5.15) the observed velocity signal
2003 is dampened. This signal, however, quickly recovers when the transition function

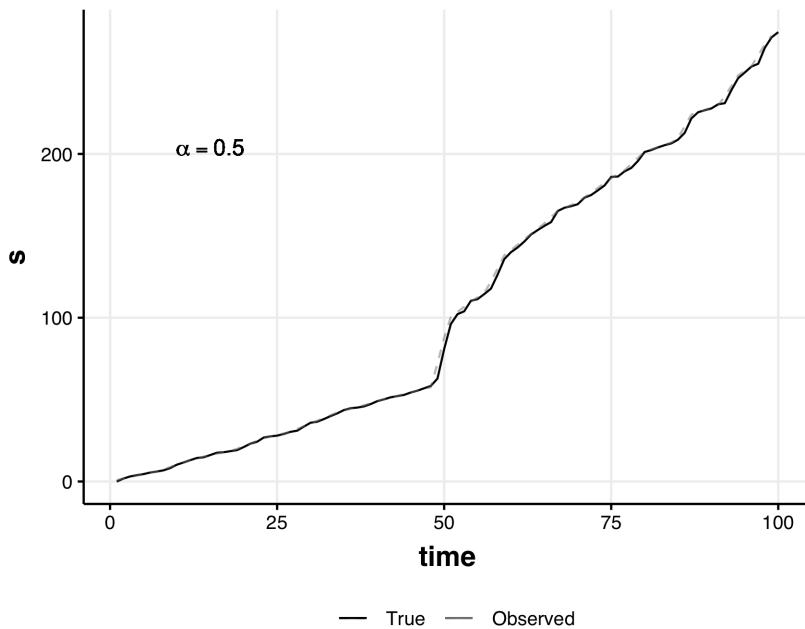


Figure 5.14: Antidifferentiated values ('observed') of the distance traveled, s , to the true process values of s ('true) provides a method for identifying the best values of the smoothing parameter, α . Under most conditions $\alpha \ll \text{tanh}$ sufficed. Here, we compare the true and antidifferentiated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

- 2004 becomes more abrupt (Figures 5.16, 5.17 ,5.18; $\alpha_{\text{tanh}} = 0.5, 0.75, \text{and} 1.00$, respectively).
- 2005 The velocity signal changes more abruptly when the means of both state variables
- 2006 while holding the relative variance constant (Figure 5.19) than when only a single
- 2007 variable shifts mean value (compare with Figure 5.15). Figure 5.20 is representative
- 2008 of the increasing signal in velocity as α_{tanh} increases.

2009 Smooth changes in variance

- 2010 Abrupt changes sometimes manifest first as a change in the variability, rather than
- 2011 the mean value, of the state variables. This condition manifests in the velocity signal

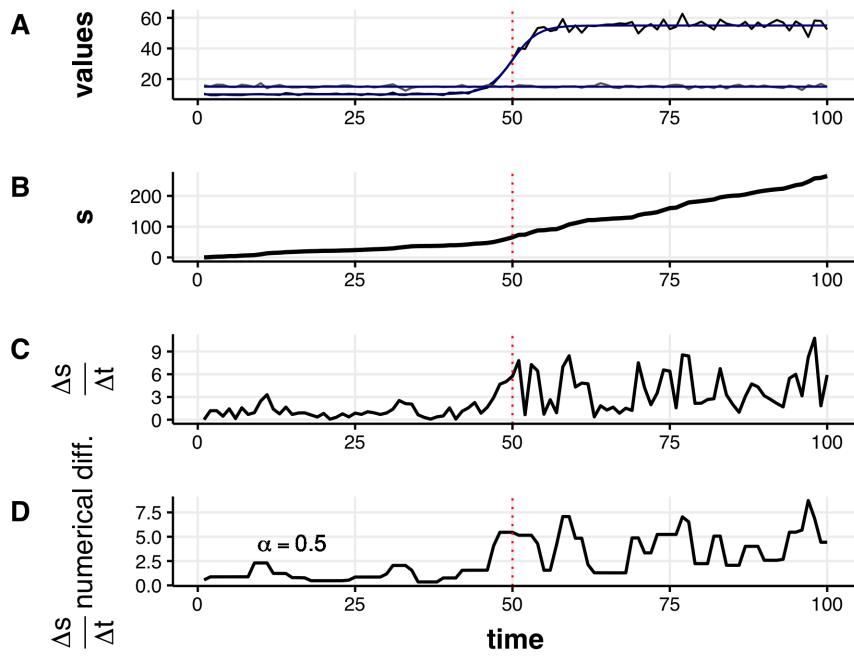


Figure 5.15: The velocity signal is muted when the hyperbolic smoothing parameter, α , is low (0.25). True and observed values of x_i (panel A), observed distance traveled (s , panel B), observed velocity (C), and the smoothed velocity (D).

when both variables experience a shift in relative variance (Figure 5.21), however, does not signal change when only one variable exhibits a shift in variance (Figure 5.22). 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 datasets. As such, I expect the velocity signal to be more pronounced when empirical systems undergo shifts in variance in at least one state variable.

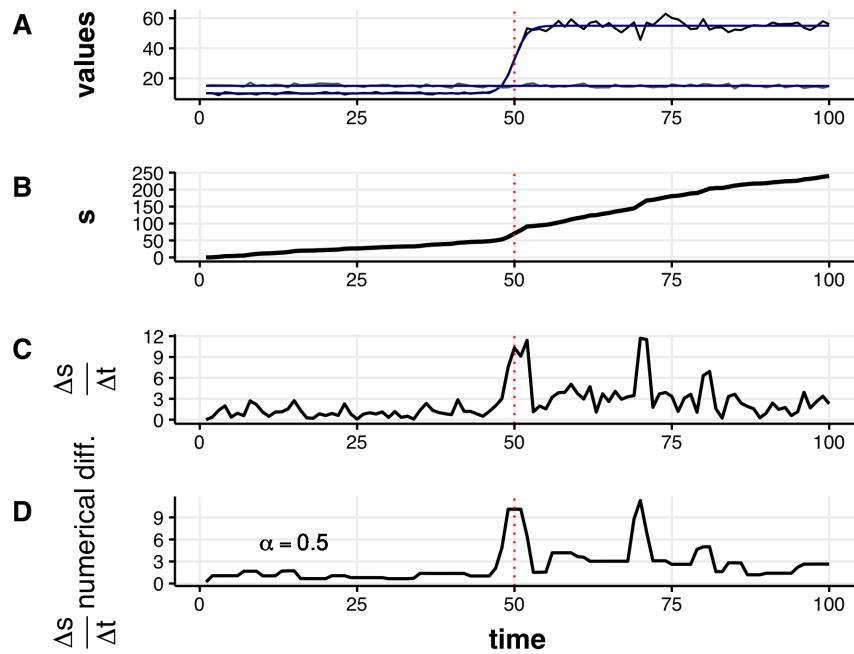


Figure 5.16: The velocity signal is muted when the hyperbolic smoothing parameter, α , is moderate (0.50). True and observed values of x_i (panel A), observed distance traveled (s , panel B), observed velocity (C), and the smoothed velocity (D).

2022 **Smooth changes in the mean and variance**

2023 Given the signals identified in the velocity when one or both state variables exhibits a
2024 shift in mean and/or variance, it is unsurprising that even under smooth transitions
2025 (when $\alpha_{tanh} = 0.25$), velocity manifests as a signal of change (Figure 5.23). This signal
2026 is most pronounced when the shift is abrupt (Figure 5.24).

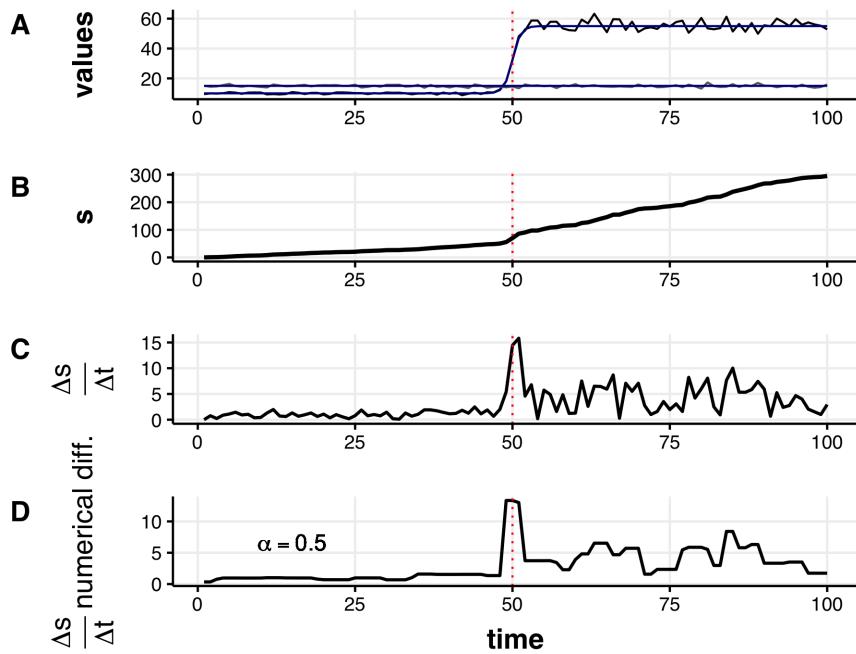


Figure 5.17: The velocity signal is muted when the hyperbolic smoothing parameter, α , is moderate (0.50). True and observed values of x_i (panel A), observed distance traveled (s , panel B), observed velocity (C), and the smoothed velocity (D).

2027 **5.5 Velocity performance under empirical transi-**
 2028 **tions: paleolithic freshwater diatom commu-**
 2029 **nity**

2030 To gather baseline information on the use of velocity in empirical systems data,
 2031 I calculated velocity for the paleodiatom system described in Chapter 6 (see also
 2032 Appendix ???. Briefly, the paleodiatom community comprises 109 time series over
 2033 a period of approximately 6936 years (Figure 5.25). As elaborated in Spanbauer
 2034 *et al.* (2014), the paleodiatom community is suggested to have undergone regime
 2035 shifts at multiple points. These abrupt changes are apparent when exploring the
 2036 relative abundances over time, as there are extreme levels of species turnover at multiple
 2037 points in the data (Figure 5.25). Using Fisher Information and climatological records,

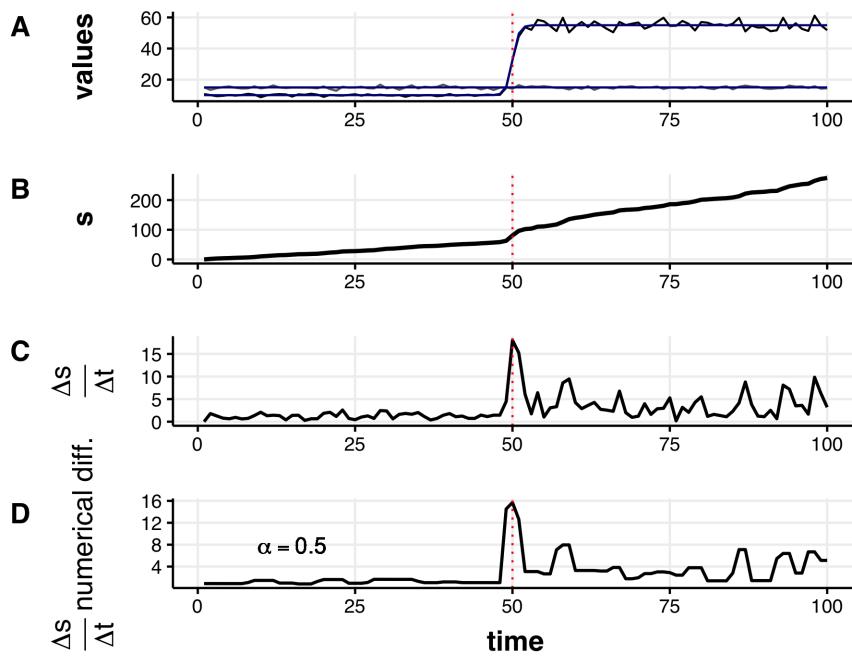


Figure 5.18: The velocity signal is muted when the hyperbolic smoothing parameter, α , is moderate (0.50). True and observed values of x_i (panel A), observed distance traveled (s , panel B), observed velocity (C), and the smoothed velocity (D).

2038 Spanbauer *et al.* (2014) suggest that regime shifts in this system at approximately
2039 1,300 years before present (where present is equal to year 1950). Spanbauer *et al.*
2040 (2014) used different regime detection metrics coupled with regional climatological
2041 events to identify regime shifts in the system, suggest that a regime shift occurred
2042 at \sim 1,300 years before present. Using the methods outlined above, I calculated the
2043 distance traveled (s) and velocity (v ; Figure 5.29). The results of v and s (Figure
2044 5.26) on the relative abundance data correspond with both the large shifts in species
2045 dynamics (see Fig 5.25, and also with the regime shift identified by Spanbauer *et al.*
2046 (2014)). However, two primary results can be made from the metrics v and s that are
2047 not obvious nor identified numerically in the results of Spanbauer *et al.* (2014):

- 2048 1. Two additional large shifts occurred at approximately 2,500, 4,800 and years
2049 before 1950

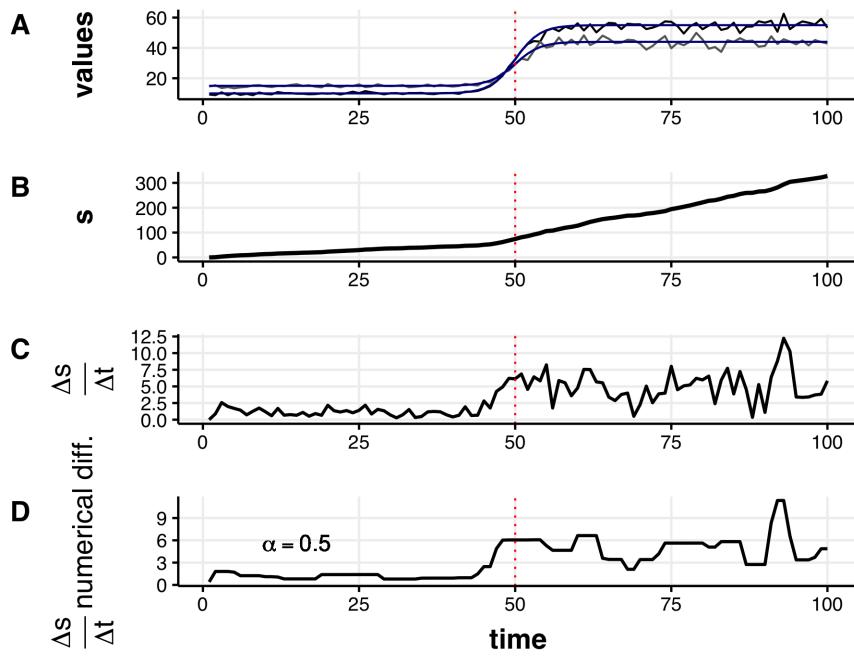


Figure 5.19: The velocity signal is regained under smooth transition ($\alpha_{\text{tanh}} = 0.25$) when both state variables undergo a shift in the mean. True and observed values of x_i (panel A), observed distance traveled (s , panel B), observed velocity (C), and the smoothed velocity (D).

2050

2051 2. The periods before the first and after the second large shifts appear oscillatory
2052 (Figure 5.27).

2053 To determine whether removing the noise in the data, I interpolated the each time
2054 series using function `stats::approx` to 700 time points. Next, I calculated the
2055 distance traveled of the entire system, s . Finally, I obtained the derivative of s by
2056 using a regularized differentiation (using function `tvdiff::TVRegDiffR`; parameters
2057 were $iter = 2000$, $scale = \text{small}$, $ep = 1 \times 10^{-6}$, and $\alpha = 100$). This method of
2058 regularized differentiation is an ideal approach to smoothing s because it assumes
2059 the data are non-smooth and incorporates finite differencing. The total variation
2060 regularized differentiation is described in Chartrand (2011), Price & Burnett (2019),
2061 and in the previous first-level section. The smoothed velocity (Figure 5.29) provides

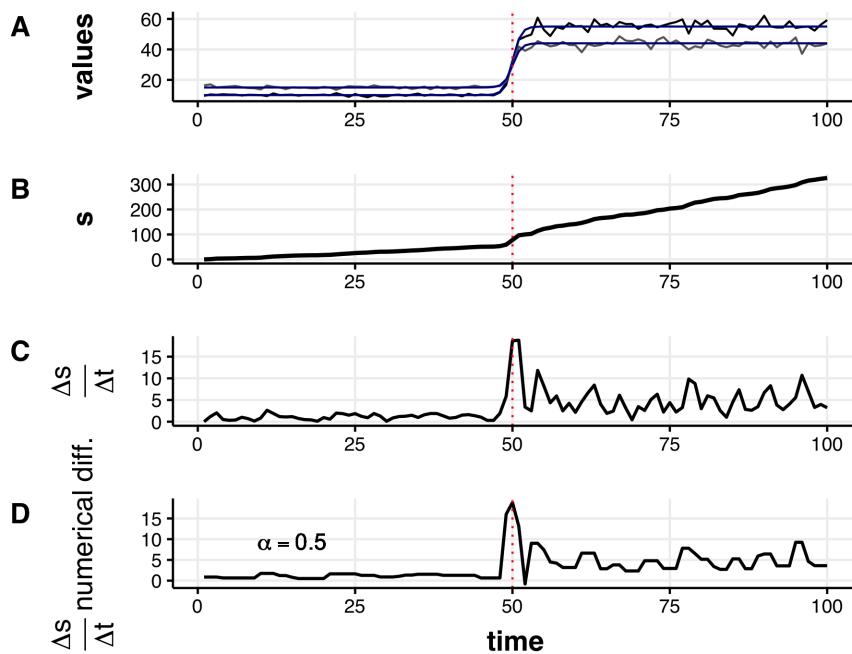


Figure 5.20: The velocity signal is regained under smooth transition ($\alpha_{\text{tanh}} = 0.75$) when both state variables undergo a shift in the mean. True and observed values of x_i (panel A), observed distance traveled (s , panel B), observed velocity (C), and the smoothed velocity (D).

2062 a similar but smoother picture of the velocity of the system trajectory. Comparing
2063 the smoothed (Figure 5.29) to the non-smoothed velocity (Figure 5.26) yields similar
2064 inference regarding the location of the regime shifts at 2,200 and 1,300 years before
2065 present, however, it more clearly demonstrates potential inter-regime dynamics (e.g.,
2066 between 7,000 and 4,800 years before present), which were not identified in previous
2067 study of this system (Spanbauer *et al.*, 2014).

2068 5.6 Discussion

2069 Here, I described the steps for calculating a novel regime detection metric, system
2070 velocity (v). First described in Fath *et al.* (2003), v is used as a single step for
2071 calculating a more complicated regime detection metric, Fisher Information (see also

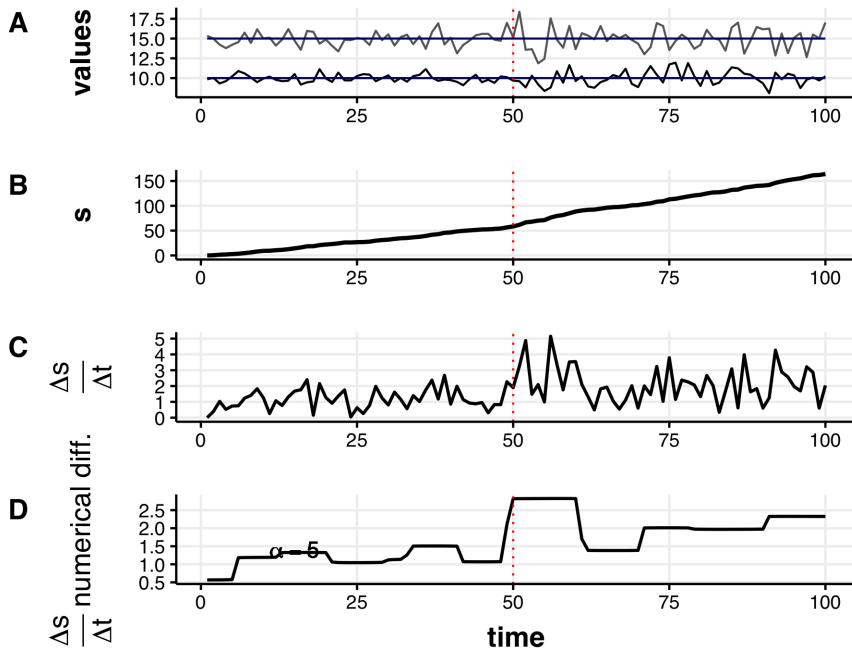


Figure 5.21: The velocity signals a rapid shift in the variance of both state variables under a moderately abrupt transition ($\alpha_{\text{tanh}} = 0.75$). True and observed values of x_i (panel A), observed distance traveled (s , panel B), observed velocity (C), and the smoothed velocity (D).

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 \gg a few. System velocity, v , may be useful in situations where the number of state variables is \gg few, and appears especially useful when the magnitude of change in one or more state variables is high (Figures 5.8, 5.20). 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 ,

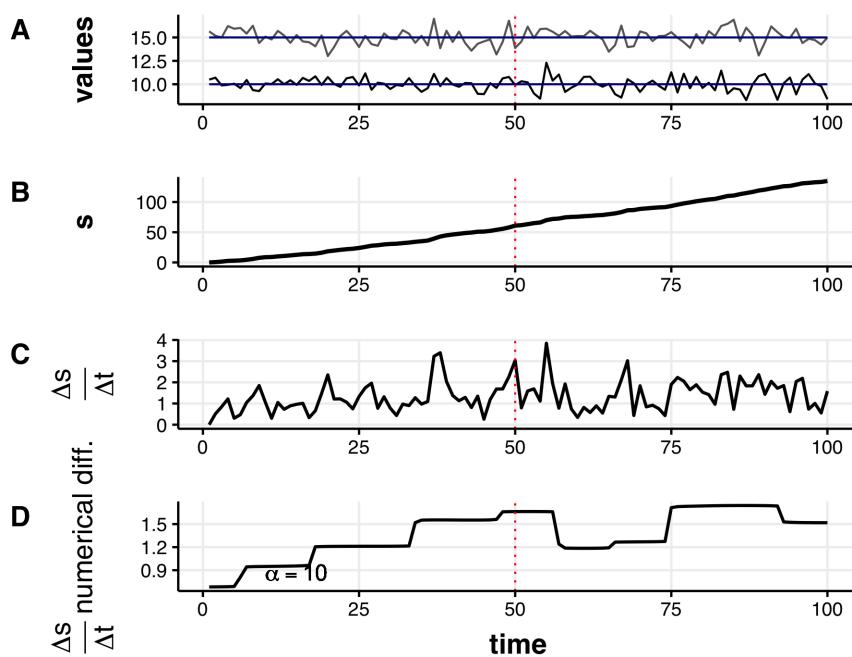


Figure 5.22: The velocity does not signal shifts in the variance of a single variable (x_1) under a moderately abrupt transition ($\alpha_{\text{tanh}} = 0.75$). True and observed values of x_i (panel A), observed distance traveled (s , panel B), observed velocity (C), and the smoothed velocity (D).

as an indicator of abrupt change in a multivariable 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 ??). 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 metric’s sensitivity to variance (Figure 5.8).

Further work is required to determine the utility of system velocity as a regime

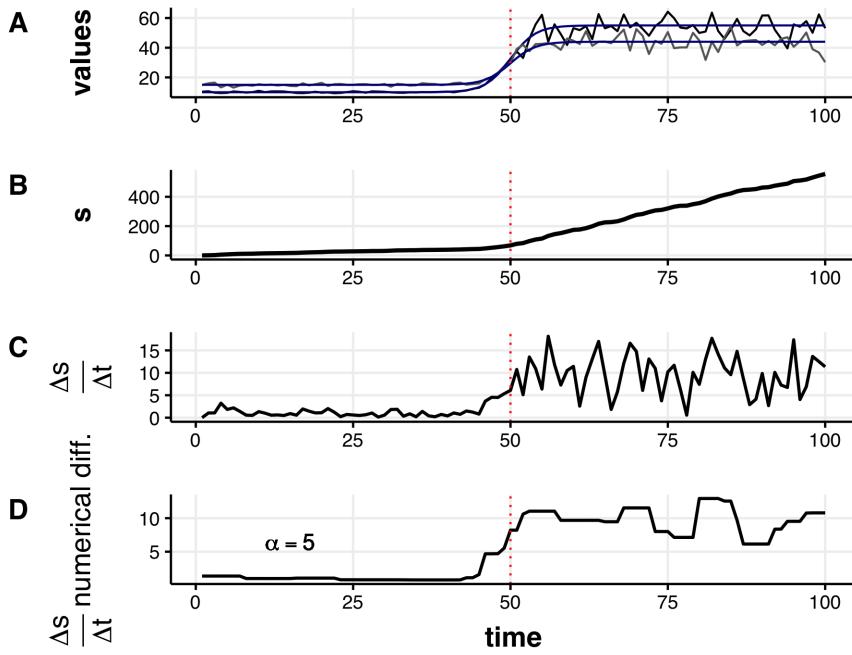


Figure 5.23: The velocity signals a shift when both variables undergo shifts in the mean and variance under a slightly abrupt transition ($\alpha_{\text{tanh}} = 0.25$). True and observed values of x_i (panel A), observed distance traveled (s , panel B), observed velocity (C), and the smoothed velocity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

2096 detection metric, however, this chapter demonstrates that the metric may indicate
 2097 clear shifts in variable means and variability about the means. In addition to examining
 2098 high-dimensional and noisy data, a study of the performance of v under conditions
 2099 where few variables exhibit large changes while many variables are relatively constant
 2100 may also prove useful. Additionally, this metric may be a useful tool for reducing the
 2101 dimensionality of high dimensional data. Although the metric loses much information,
 2102 as opposed to some dimension reduction techniques, e.g. Principal Components
 2103 Analysis PCA, the metric is simple to calculate (even by hand), is computationally
 2104 inexpensive, and is intuitive, unlike many clustering algorithms (e.g., Non-metric
 2105 Multidimensional Scaling NMDS). Like system velocity, methods of the latter variety
 2106 (e.g. NMDS) require post-hoc statistical analyses to confirm the location of clusters
 2107 (or abrupt change, regime shifts), while methods of the former variety (e.g. PCA)

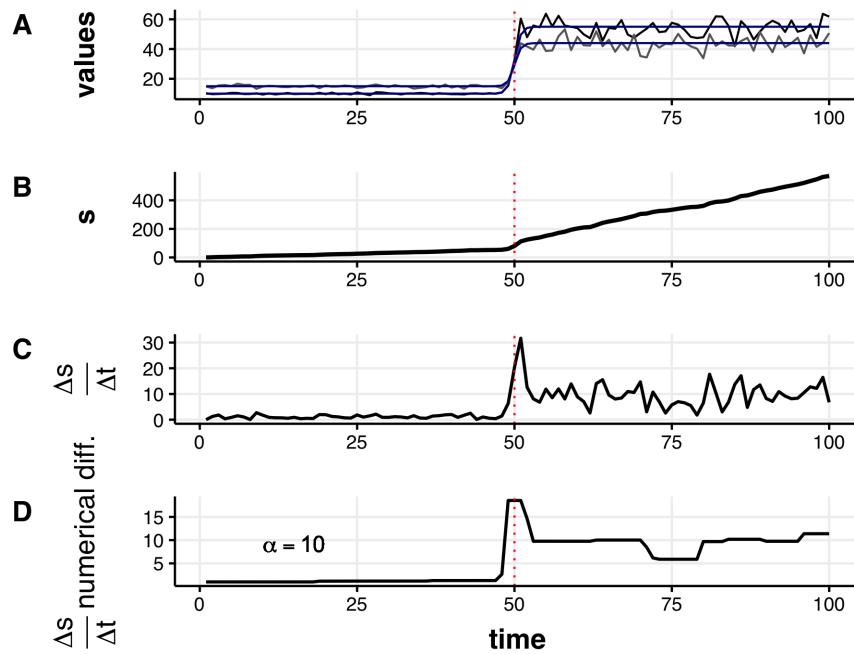


Figure 5.24: The velocity signals a shift when both variables undergo shifts in the mean and variance under a slightly abrupt transition ($\alpha_{\text{tanh}} = 1.00$). True and observed values of x_i (panel A), observed distance traveled (s , panel B), observed velocity (C), and the smoothed velocity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

2108 retain loadings but do not necessarily identify the locations of abrupt shifts.

2109 5.7 Supplementary Figures

2110 Figures 5.30, 5.31, and 5.32 provide additional examples of the behavior of velocity,
2111 v when varying the mean and/or variance prior to and/or after the induced abrupt
2112 shift in the toy system with a discontinuous transition at $t = 50$.

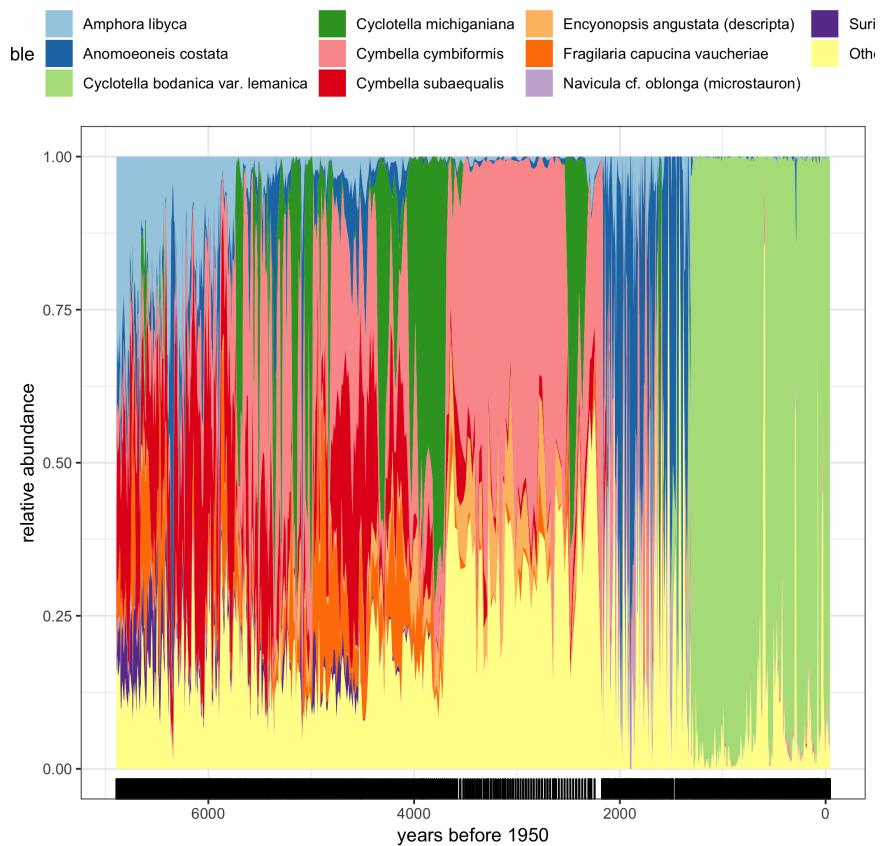


Figure 5.25: Relative abundances of the most common diatom species in the time series. Few species dominate the data over the entire time series, and turnover is apparent at multiple observations.

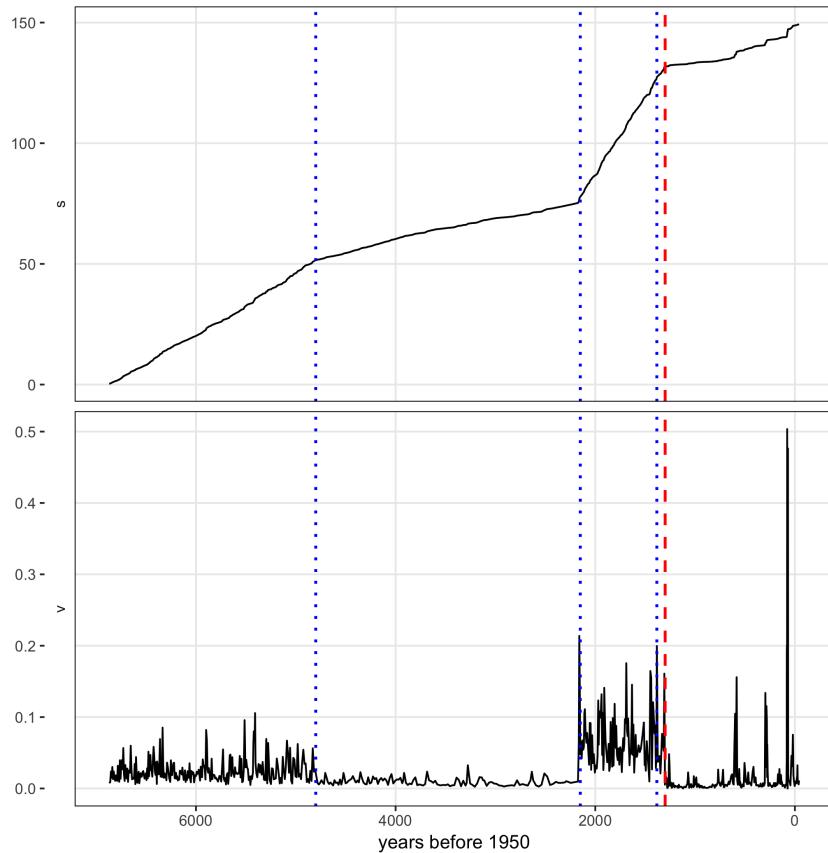


Figure 5.26: Velocity v and distance traveled s of the paleodiatom time series. Dashed line at 1,300 years before 1950 indicates the regime shift identified in Spanbauer et al. (2014). Dotted lines indicate regime shifts as visually identified on metrics s and v .

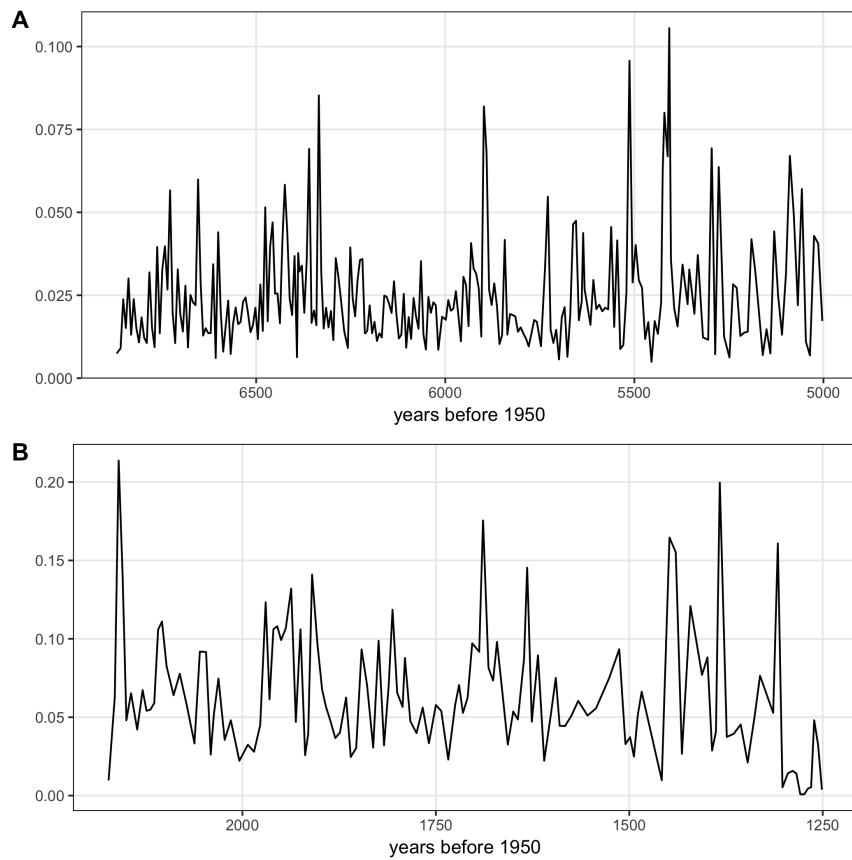


Figure 5.27: Velocity (v) indicates periodic-like conditions in the first (A) and second (B) regimes.

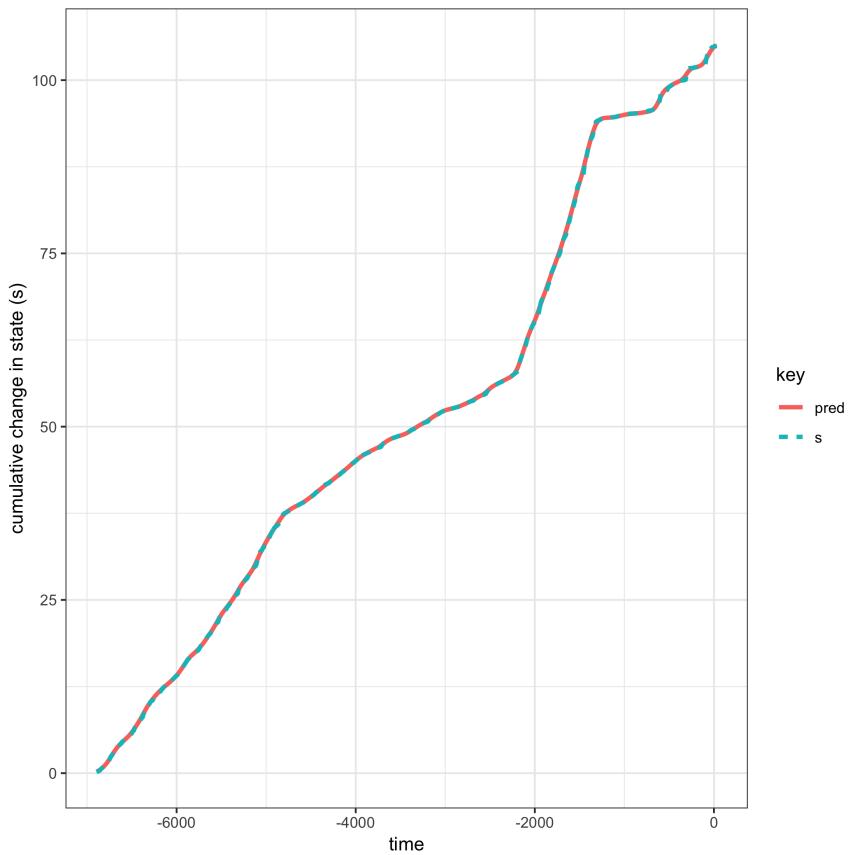


Figure 5.28: The regularized differentiation of s was best fit using $\alpha = 100$. Higher overlap of s and pred indicates a good fit of the regularized differentiated metric to the non-smoothed metric, s .

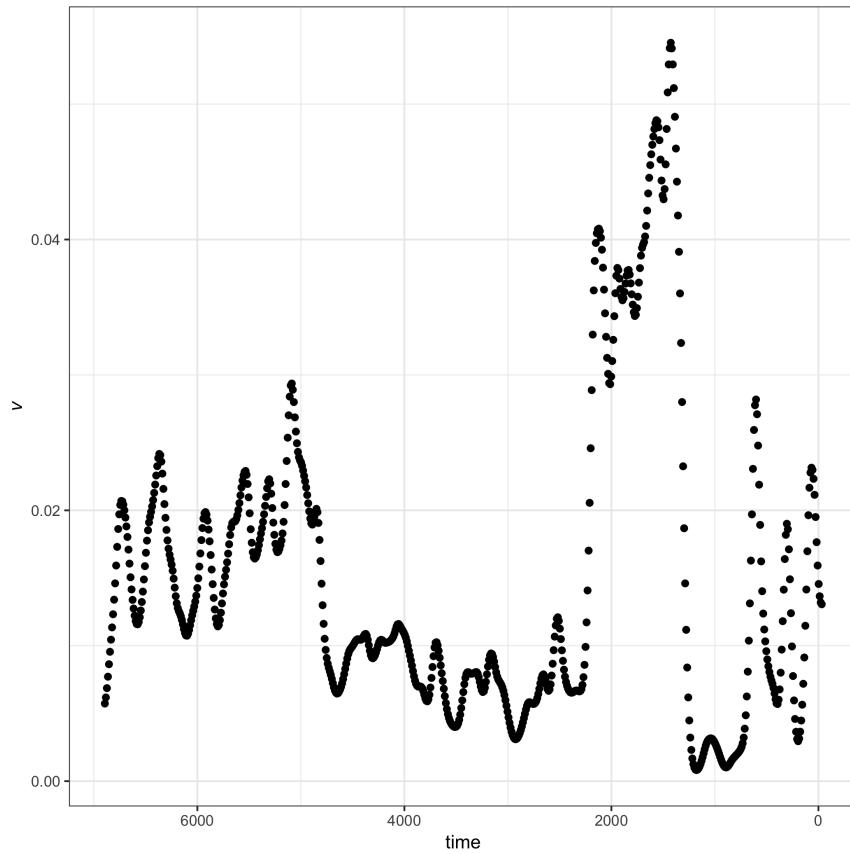


Figure 5.29: The velocity metric (v) signals potential periodicities in the paleo diatom time series data when the distance traveled metric, s , is smoothed using regularized differentiation methods [price2019tvdiff].

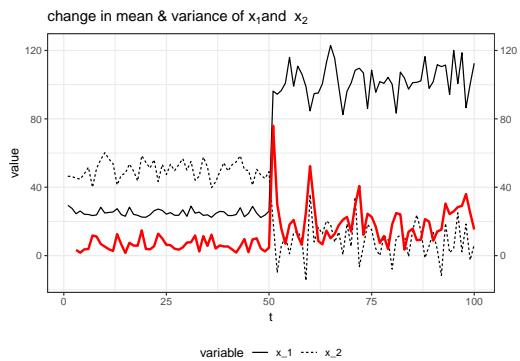


Figure 5.30: System change (s) and velocity (v) of the model system over the time period. Change in means ($\bar{x}_{1,pre} = 25$, $\bar{x}_{1,post} = 100$, $\bar{x}_{2,pre} = 50$, $\bar{x}_{2,post} = 10$) and an increase in variance ($\sigma_{1,pre} = 2$, $\sigma_{1,post} = 10$, $\sigma_{2,pre} = 5$,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

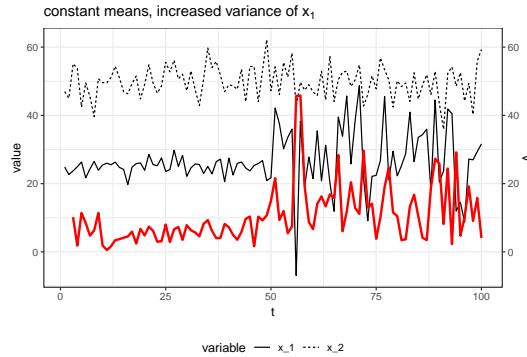


Figure 5.31: System change (s) and velocity (v) of the model system over the time period. Constant means ($\bar{x}_1 = 25$, $\bar{x}_2 = 50$) and sharp change in variance for one state variable $\sigma_{1_{pre}} = 2$, $\sigma_{1_{post}} = 12$, $\sigma_{2_{pre,post}} = 5$

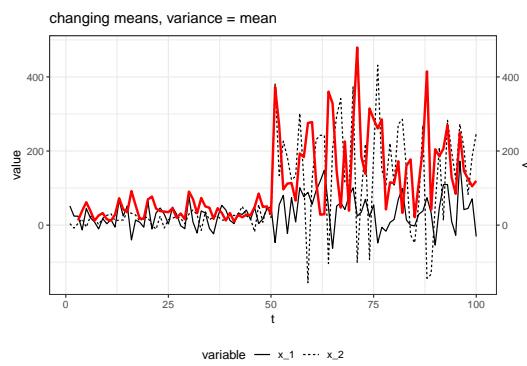


Figure 5.32: System change (s) and velocity (v) of the model system over the time period. Variance equal to mean ($/bar{x}_i = /sigma_i$), where means ($/bar{x}_{1_{pre}} = 25$, $/bar{x}_{1_{post}} = 50$, $/bar{x}_{2_{pre}} = 15$, $/bar{x}_{2_{post}} = 150$).

2113 **Chapter 6**

2114 **Using Resampling Methods to
2115 Evaluate the Relative Performance
2116 of Regime Detection Measures
2117 Under Varying Conditions of Data
2118 Quality and Quantity**

2119 **6.1 Introduction**

2120 Ecological systems have many unpredictable and variably interacting components.
2121 Methods for analyzing these complex systems, e.g. Dynamic Bayesian Networks,
2122 network models, and food webs are designed to handle these complexities, yet require
2123 data- and knowledge-intensive models. Although ecological data collection and data
2124 management techniques are improving (La Sorte *et al.*, 2018), the aforementioned
2125 approaches to modeling and understanding complex system are often infeasible in

2126 ecosystem research and management (Clements & Ozgul, 2016).

2127 A growing concern with anthropogenic impacts on the environment has increased
2128 the demand for mathematical and statistical techniques that capture these dynamics.
2129 These often undesirable changes in the structure or functioning of ecological systems
2130 are often referred to as *regime shifts*, *regime changes*, *state change*, *abrupt change*, etc.
2131 (Andersen *et al.*, 2009) . A yet-unattained goal of ecological research and management
2132 is to reach a point where these methods can predict impending regime shifts in real-
2133 time and with high confidence. Ideally, ecological regime shift detection methods
2134 (hereafter, regime detection measures) would require little knowledge of the intrinsic
2135 drivers of the system, and the users of the method would not be required to know if
2136 and where a regime shift occurred in the data.

2137 Despite the suite of regime detection measures in the environmental and ecological
2138 research literatures, they are not used in ecological management. We can describe the
2139 current state of regime detection measures as being either system specific (i.e., the
2140 method is not system agnostic) or not. Methods of the latter type are convenient in
2141 that they can be applied across various system and data types, but the results of these
2142 analyses require some degree of subjective interpretation (Clements & Ozgul, 2018; *c.f.*
2143 Batt *et al.*, 2013). Efforts to develop and/or improve regime detection measures that
2144 do not require such subjectivity will aid the advance of regime detection measures
2145 research and application.

2146 Current efforts to improve regime detection measures may be stunted by the lack of
2147 application beyond simple and/or theoretical (toy) systems data. Like most statistical
2148 and mathematical approaches, the evolution of many regime detection measures
2149 begins with application to theoretical data, followed by application to empirical data.
2150 Current applications of regime detection measures to empirical, ecological data are
2151 largely limited to data describing populations (Anderson & Piatt, 1999; Alheit *et al.*,

2152 2005; deYoung *et al.*, 2008), climatic, marine, and Paleolithic regime shifts (Yang
2153 & Wu, 2006; Spanbauer *et al.*, 2014; Kong *et al.*, 2017), with few applications to
2154 terrestrial data (*c.f.* Bahlai *et al.*, 2015; Sundstrom *et al.*, 2017). Although testing the
2155 performance and inference boundaries of theoretical and simple systems is important,
2156 they are of little use to ecosystem managers if they are not proven to be easily and
2157 reliably applicable to their system. Additionally, regime detection measures should
2158 be capable of handling empirical ecological data, which are often sparse, noisy, and
2159 irregularly sampled..

2160 Ecological systems data is expensive to capture, and has large process variation and
2161 observation errors. This variability reduces data quality and quantity, limiting the
2162 numerical tools for identifying trends and changes in the system (Thrush *et al.*, 2009).
2163 Some methods, new and old, proposed as regime detection measures are purported
2164 to handle the data limitation and quality issues inherent in ecological data, and
2165 minimize subjective decisions for choosing state variables and interpreting results. For
2166 example, variable reduction techniques, e.g. principal components analysis (Rodionov
2167 & Overland, 2005; Andersen *et al.*, 2009; Reid *et al.*, 2016), clustering algorithms
2168 (Weijerman *et al.*, 2005; Weissmann & Shnerb, 2016), an index of variance (Brock &
2169 Carpenter, 2006), and Fisher Information (Cabezas & Fath, 2002; Fath & Cabezas,
2170 2004; Karunanithi *et al.*, 2008) were introduced as methods which collapse the system
2171 into a single indicator of ecological regime shifts. Although these methods have been
2172 used on empirical ecological systems data, their robustness to empirical data quality
2173 and quantity have yet to be examined.

2174 In this Chapter I examine the influence of observation and process errors on the
2175 inference obtained from select multivariable regime detection measures. There are
2176 three major objectives:

- 2177 1. Identify the effects of data quality on regime detection measure inference.

- 2178 2. Identify the effects of data quantity on regime detection measure inference.
2179 3. Explore the relative performance of velocity (described in Chapter 5) to the
2180 above mentioned methods under multiple scenarios.

2181 This Chapter provides baseline relative performance estimates of select, multivariable
2182 regime detection measures under various scenarios of data quality and quantity. The
2183 results from this Chapter inform the practical ecologist of the potential limitations to
2184 consider when applying these regime detection measures to their data, and has potential
2185 to inform the data collection process. Additionally, the software accompanying this
2186 Chapter allows the end user to implement these methods on this diatom system, a
2187 toy system, or their own data.

2188 **6.2 Data and Methodology**

2189 **6.2.1 Study system and data**

2190 I used paleodiatom time series from a freshwater system in North America (Foy Lake,
2191 present day Montana) that apparently underwent rapid shifts in algal community
2192 dynamics at multiple points in time. This data comes from a single soil core sample,
2193 from which the relative abundances of 109 diatom species were identified at 768
2194 observations (time points) over \approx 7,000 years (Figure 6.1. Although the soil core was
2195 sampled at regular distances, the soil accumulation process is not necessarily linear
2196 over time, resulting in irregularly-sampled observations (i.e., time elapsed between
2197 sampling points varies; see Figure 6.2). The data were published in Spanbauer
2198 *et al.* (2014) and can be downloaded at the publisher's website.

²¹⁹⁹ **6.2.2 Regime detection measures**

²²⁰⁰ Fewer model-free regime detection metrics exist than do model-based metrics (Chapter
²²⁰¹ 2) and of these, only a few are suggested for multivariable data. Here, I compare
²²⁰² the results for three regime detection metrics that are model-free and can handle
²²⁰³ multivariable data: velocity (Chapter 5), the Variance Index (Brock & Carpenter,
²²⁰⁴ 2006) and Fisher Information (Fath *et al.*, 2003). I chose the Variance Index, as this
²²⁰⁵ is one of the more widely applied multivariate, model-free regime detection measures,
²²⁰⁶ and has been shown to, in some empirical data, identify regime shifts *post hoc*. I
²²⁰⁷ introduced the velocity in Chapter 5 as a new, potential regime detection metric. As
²²⁰⁸ this is the first time it has been used for such a purpose, including it in this approach
²²⁰⁹ allows us to further identify potential flaws with the method, but also to gain some
²²¹⁰ baseline estimates of its relative performance. In Chapter 3, I presented the Fisher
²²¹¹ Information metric as it is used in detecting ecological regime shfits, and discuss the
²²¹² situations under which it may or may not be a good metric.

²²¹³ **Velocity (v) calculation**

²²¹⁴ In Chapter 5, I describe a new method, **velocity**, v , as a potential dimension reduction
²²¹⁵ and regime detection method. First introduced by Fath *et al.* (2003) as one of
²²¹⁶ multiple steps in calculating their variant of Fisher Information, velocity calculates
²²¹⁷ the cumulative sum of the square root of the sum of the squared change in all state
²²¹⁸ variables over a period of time [Eq. (6.1)]. Steps for calculating this metric are
²²¹⁹ described in detail in Chapters 3 and 5.

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

2221 Variance Index (VI) calculation

2222 The Variance Index was first introduced by Brock & Carpenter (2006), and can be
2223 simple defined as the maximum eigenvalue of the covariance matrix of the system
2224 within some period, or window, of time. The Variance Index (also called Variance
2225 Indicator) was originally applied to a modelled system (Brock & Carpenter, 2006) and
2226 has since been applied to empirical systems data (Spanbauer *et al.*, 2014; Sundstrom *et*
2227 *al.*, 2017). Although rising variance has been shown to manifest prior to abrupt shifts
2228 in some empirical systems data (Nes & Scheffer, 2005; Brock & Carpenter, 2006), the
2229 Variance Index, which is intended for multivariate data, appears most useful when the
2230 system exhibits a discontinuous (non-linear) shift (Brock & Carpenter, 2006).

2231 Fisher Information (FI) calculation

2232 Fisher Information (I) is essentially the area under the curve of the acceleration to
2233 the fourth degree (s''^4) divided by the squared velocity (s'^2 ; also referred to as v in
2234 Chapter 5) of the distance travelled by the system, s over some period of time (T),
2235 and is given in Eq. (6.2):

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

2236 I refer the reader to Chapter 3 for a complete description and to Cabezas & Fath
2237 (2002) for a complete derivation of Fisher Information.

**2238 Using moving window analysis to calculate Fisher Information and Vari-
2239 ance Index**

2240 Unlike *velocity*, the Variance Index and Fisher Information are calculated using moving
2241 window analysis. That is, over the entire time series, T^* , these metrics are calculated
2242 within multiple windows of time, T . In this approach, all state variables, x_i , are used

2243 to inform the calculations (of Variance Index and Fisher Information) over a time
2244 interval, T , where T is the length in [time] units of the time interval and satisfies the
2245 following condition: $2 \leq T < (T^* - 1)$. If $T = T^* - 1$, then only a single value of the
2246 metric will be calculated for entire time series, which does not allow for any estimate
2247 of change.

2248 When using these metrics in the context of identifying abrupt changes in ecological
2249 systems data across T^* , it is ideal the value of T meets the following conditions:
2250 $3 < T \ll T^* - 1$. The length of a time window dictates the number of calculations
2251 one can obtain over T^* , such that the number of potential metric calulations increases
2252 as $\frac{T}{T^*}$ decreases. Previous applications of moving window analyses to calculate
2253 Fisher Information found that at least eight observations (time points) should be used
2254 [citation].

2255 An additional parameter is required when conducting moving window analyses: the
2256 number of time points by which the window advances. In order to maximize the
2257 data, I advance the window at a rate of one time unit. However, it is important to
2258 note that because these data are not sampled annually and the because the window
2259 always advances by a single time unit, the number of observations included in each
2260 calculation will not be the same. If fewer than 5 observations are in a window, I did
2261 not calculate metrics, advancing the window forward. I assigned the calcuated values
2262 of Fisher Information and Variance Index within each moving window to the **end** (the
2263 last time unit) of the moving window. In temporal analyses, assigning the value[which
2264 value] to any other point in time (e.g., the beginning or the middle) muddles the
2265 interpretation of the metric over T^* . Also note that this method has the potential to
2266 result in calculating a metric for all integers between $0.20T^*$ and T^* .

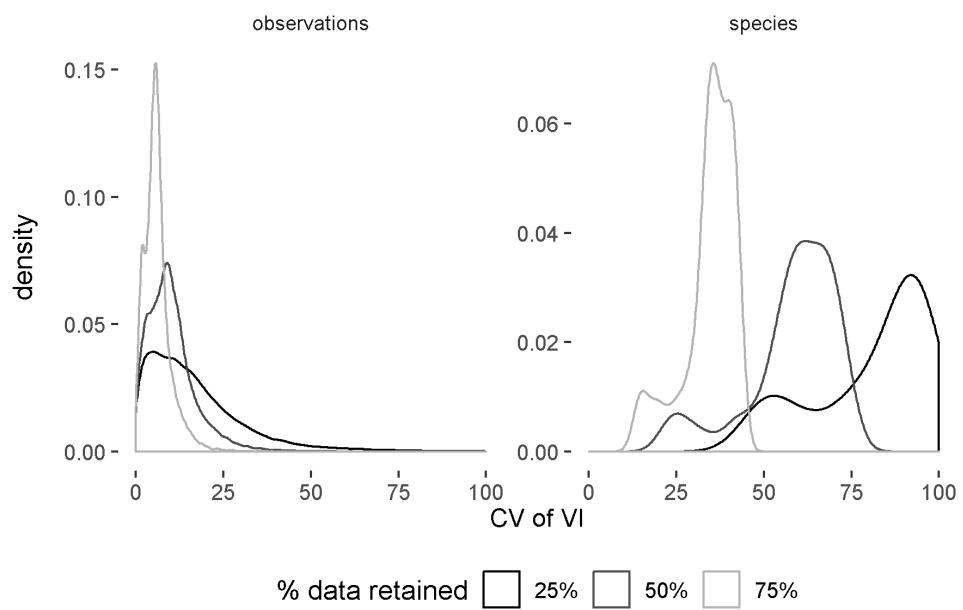
2267 **6.2.3 Simulating data quality and quantity issues using re-**
2268 **sampling techniques**

2269 Using a resampling approach I calculated the regime detection measures over different
2270 scenarios simulating data quality and data quantity issues common to ecological data
2271 analysis. The scenarios are categorized as *observations* and *species*. The observations
2272 scenario simulates a loss of temporal observations (decreasing the number of times the
2273 system was observed), and the species scenario simulates a loss of information about the
2274 system by removing some proportion of the species. The loss of temporal observations
2275 and the loss of species were examined at three proportions: $\mathbf{P} = [0.25, 0.50, 0.75, 1.00]$,
2276 where \mathbf{P} is the proportion of species and time points **retained** for analysis. For
2277 example, when $\mathbf{P} = 0.25$, a random selection of 25% of the species are retained for
2278 analysis in the species scenario. I resampled the data over 10,000 iterations (N_{samp})
2279 for each scenario and \mathbf{P} combination. Note that because when $\mathbf{P} = 1.00$, all data are
2280 retained. Therefore, no resampling was conducted at this level because only a single
2281 metric (e.g. Velocity) value is possible.

2282 **6.2.4 Comparing regime detection measures**

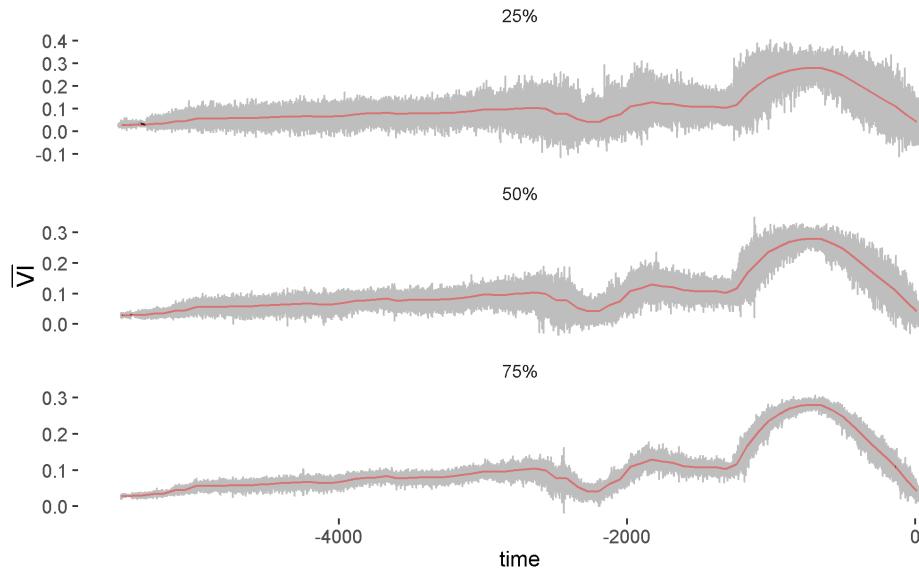
2283 Interpretation of the regime detection measures used in this analysis are currently
2284 limited to visual inspection. Therefore, I limit inference in this study largely to the
2285 impact of data loss on the variability with a regime detection measure (i.e. how robust
2286 is the measure to data loss). It is important to not only identify the influence of data
2287 quality and quantity on the performance of individual regime detection metrics, but
2288 also to somehow relate these qualities. I visually inspect the relative performance of
2289 these metrics by comparing the coefficient of variation of the resampled samples for the
2290 results of resampling method (\mathbf{M} ; species, observations) and sampling percentage (\mathbf{P} ;

2291 25%, 50%, 75%) combination for each metric (FI, VI, v). The coefficient of variation
 2292 measures provides a relative measure of the variability in the estimated metric across
 2293 resampled samples as $100\frac{\sigma}{\mu}$, where σ is the standard deviation and μ is the mean
 2294 value. I observed the distributions of the CV to identify potential flaws in the metrics
 2295 should data quality or quantity (**M**, **P**) decrease. First, within a value of **P** a low error
 2296 to mean ratio (CV) indicates that the metric value is similar across the resampled
 2297 samples ($N_{samp} = 10,000$). The efficacy of the metric should be questioned as $CV \rightarrow 1$,
 2298 and perhaps even abandoned as $CV \gg 1$. Next, we can examine how the distribution
 2299 of CV changes within **M** and across **P**. As we increase **P**, we are increasing the volume
 2300 of data we are feeding the metric. Intuitively, we can assume that as we add more data
 2301 (volume), we are supplying the metric with more *information*, theoretically increasing
 2302 the signal-to-noise ratio. Following this logic, we should expect the distribution of CV
 2303 to generally decrease in mean CV value and also become less variable (less dispersion
 2304 around the mean CV). A visual examination of the distribution of CV across **P** and
 2305 within **M** was sufficient to achieve inference regarding the quality of these metrics
 2306 upon data loss and lessened quality.

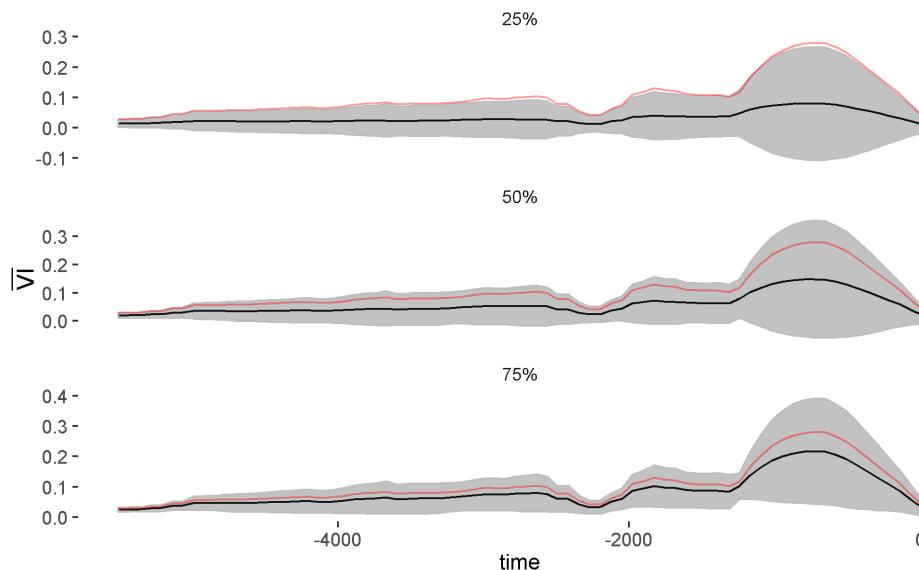


2307 \begin{figure}

2308 \caption{Density plot of the coefficient of variation (CV) as a percentage (%) of the
2309 Variance Index resampled values over 10,000 iterations. Densities are drawn based on
2310 all values of CV but values >100% are not printed.} \end{figure}

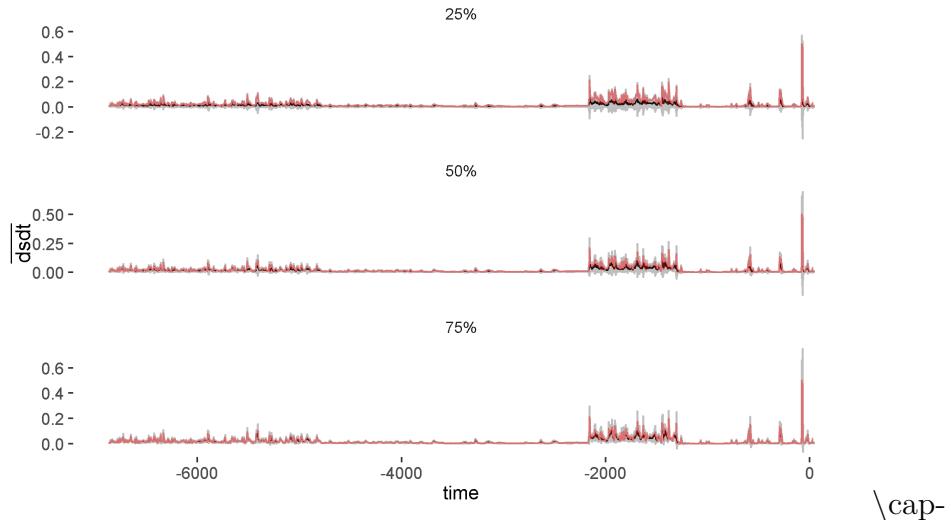


2311 \begin{figure}
2312 \caption{Mean Variance Index (VI) and associated 95% confidence intervals over
2313 10,000 iterations using the **observations** resampling method. Red line indicates the
2314 value of VI when **M** and **P** = 100%.} \end{figure}

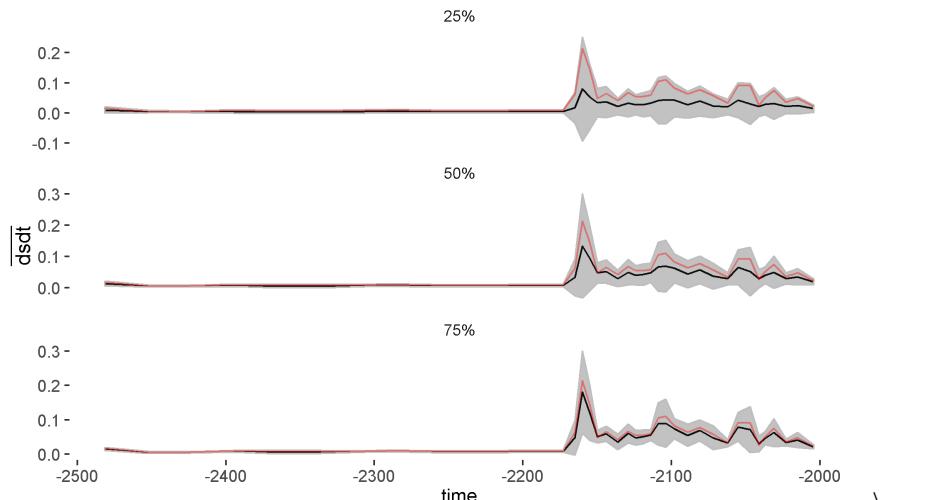


2315 \begin{figure}
2316 \caption{Mean Variance Index (VI) and associated 95% confidence intervals over

2317 10,000 iterations using the **species** resampling method. Red line indicates the value
 2318 of VI when **M** and **P** = 100%.} \end{figure}



2319 \begin{figure} \cap-
 2320 tion{Mean velocity and associated 95% confidence intervals over 10,000 iterations
 2321 using the observations resampling method. Red line indicates the value of velocity
 2322 when **M** and **P** = 100%.} \end{figure} \cap-



2323 \begin{figure} \cap-
 2324 tion{Mean velocity and associated 95% confidence intervals over 10,000 iterations
 2325 using the observations resampling method for a subset of the time series (the second
 2326 'regime' identified). Red line indicates the value of velocity when **M** and **P** = 100%.}
 2327 \end{figure}

2328 **6.3 Results**

2329 **6.3.1 Velocity of the distance travelled (v)**

2330 The velocity of the distance travelled, $\frac{ds}{dt}$ or v , exhibited dispersion across the values
2331 of \mathbf{P} , however, yielded consistent results (i.e., high overlap in the densities of the
2332 CV across values of \mathbf{P} and across methodologies; see Figure ??). Further, it should
2333 be noted that because v is calculated using first differences, it will be sensitive to
2334 large changes in the state variables. By examining the density plot of the CV of the
2335 dsitance travelled, s , we notice that this measure is highly *insensitive* to data loss
2336 (Figure ??), suggesting that a finite differencing appraoch (e.g., using total variation
2337 regularized differentiation; see Chapter) which can yield a much smoother derivative
2338 than the approach used here, may decrease the sensitivity of v to data loss. This
2339 hypothesis is further supported when examining the effect of species (Figure 6.2.4)
2340 and temporal observation loss (Figure 6.2.4) on the velocity metric. These conditions
2341 are representative of the other $\mathbf{P} - \mathbf{M}$ combinations.

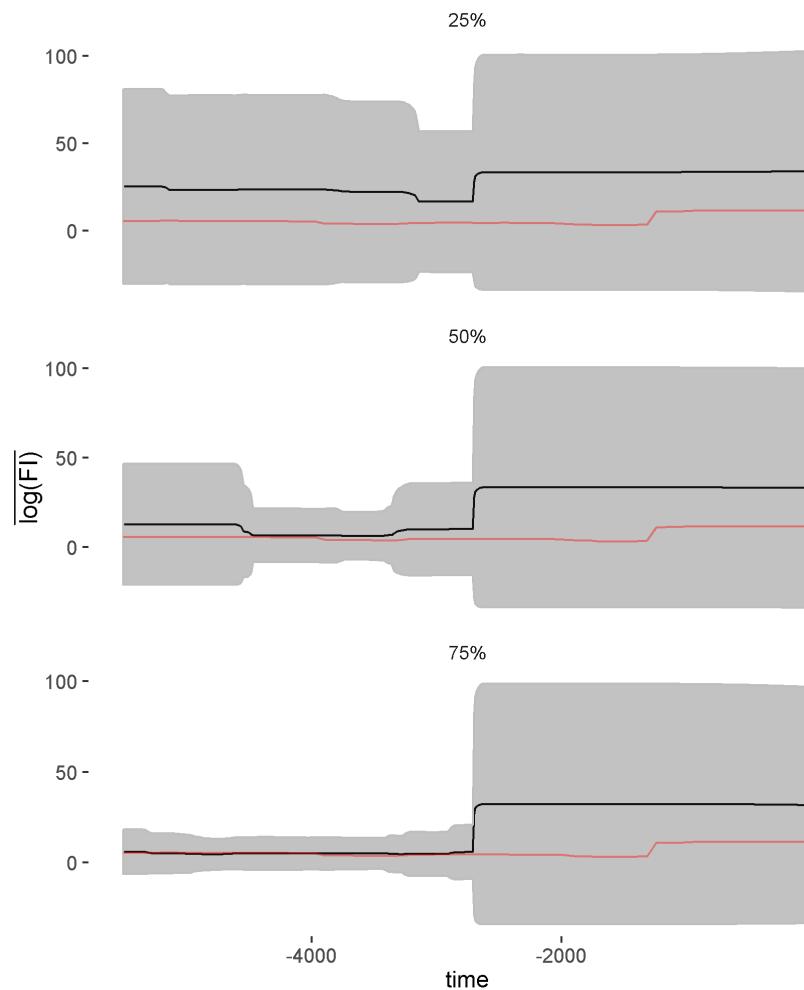
2342 **6.3.2 Variance Index**

2343 The Variance Index (VI) performed best under the the observations resampling method,
2344 exhibiting low values for and low dispersion in the CV density (Figure 6.2.4) across
2345 iterations. However, the VI appears sensitive to high losses of species information,
2346 where the density of the CV still exhibits low dispersion but with higher overall mean
2347 values (Figure 6.2.4). Surprisingly, the Variance Index was insensitive to temporal
2348 observation loss (Figure 6.2.4), exhibiting a similar amount of noise across various
2349 degrees of data loss (\mathbf{P}). Although the signal was damped under the species method,
2350 the signals for the shifts in community composition were not lost across levels of \mathbf{P}
2351 (Figure 6.2.4). This is likey due to the high probably that the dominant species were

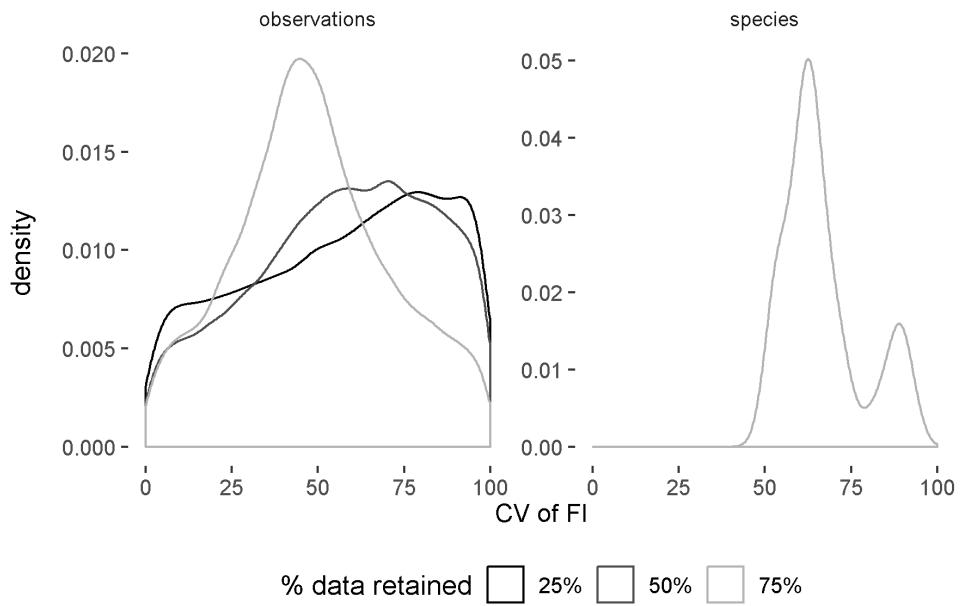
2352 rarely always excluded from the resampled observations.

2353 **6.3.3 Fisher Information is highly sensitive to information
2354 loss**

2355 The Fisher Information method did not yield conclusive results regarding the abrupt
2356 shifts in the paleodiatom community composition. Further, this method appears highly
2357 sensitive to varying quality and quantitatis of data (Figures 6.3.3, 6.3.3). Although
2358 the Fisher Information identifies the shift in community composition at $\sim 1,300$ years
2359 before present, it fails to identify shifts outside this period. Further, it is difficult
2360 to visually analyze any value of the Fisher Information on the original scale as the
2361 values range from ≈ 0 to 10^{15} (Figure 6.3.3). In addition to failing to identify the
2362 shifts in community composition, the standard deviation of Fisher Information far
2363 exceeded the mean value of Fisher Information under all $\mathbf{M} - \mathbf{P}$ scenarios (Figure
2364 6.3.3). When I resampled the data using 25% and 50% of the species the ratio of mean
2365 Fisher Information to standard deviation (CV) of Fisher Information is always $\gg 1$
2366 (i.e, not pictured in Figure 6.3.3). The high variation in FI values across resampled
2367 iterations coupled with the high dispersion within each $\mathbf{M} - \mathbf{P}$ combination (Figure
2368 6.3.3) suggests Fisher Information will not produce similar trends when we lose or
2369 distort the data collected. This is also suggested by the high confidence intervals
2370 surrounding each $\mathbf{M} - \mathbf{P}$ combination (Figure 6.3.3).



2371 \begin{figure} \cap-
2372 tion{Mean Fisher Information (FI; note the scale) and associated 95% confidence
2373 intervals over 10,000 iterations using the species resampling method. Red line indicates
2374 the value of FI when M and $P = 100\%$. A very small value was added to the mean
2375 FI prior to log transformation.} \end{figure}



2376 \begin{figure}

2377 \caption{Density plot of the coefficient of variation (CV) as a percentage (%) of the
 2378 Fisher Information resampled samples (10,000 iterations). Densities are drawn based
 2379 on all values of CV, but values >100% are not printed.} \end{figure}

2380 6.4 Detrending the Data Prior to Calculations

2381 If and how to manipulate the original data prior to calculating various regime detection
 2382 methods is an important consideration, and a line of research that has not yet been
 2383 fully explored. Although most of the multivariate methods identified in the literature
 2384 review do not require data conforms to a specific distribution, how th results of
 2385 these methods can vary as we change the quality and characteristics of the original
 2386 data (Michener & Jones, 2012). In fact, since many of the methods for regime shift
 2387 detection are specifically looking for changes in variance structure and autocorrelation,
 2388 standardizing variances is not counterintuitive.

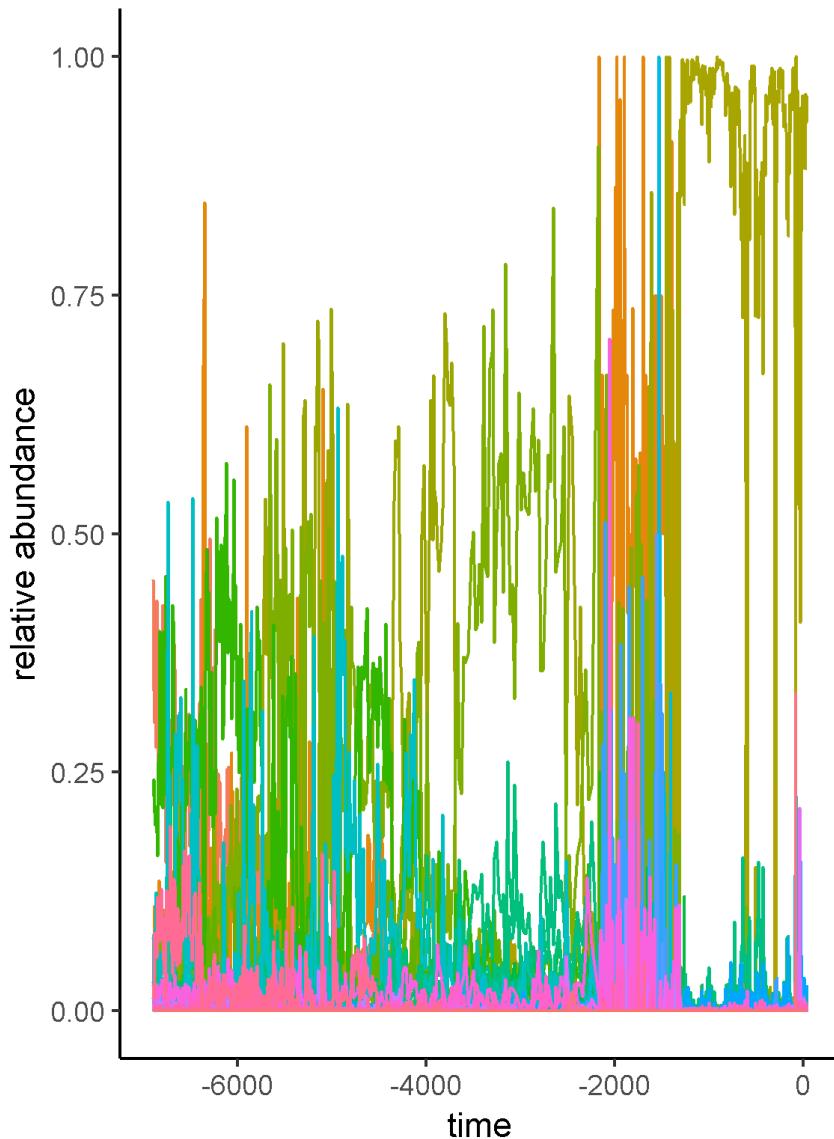


Figure 6.1: Relative abundances of the diatom species in Foy Lake over the time period.

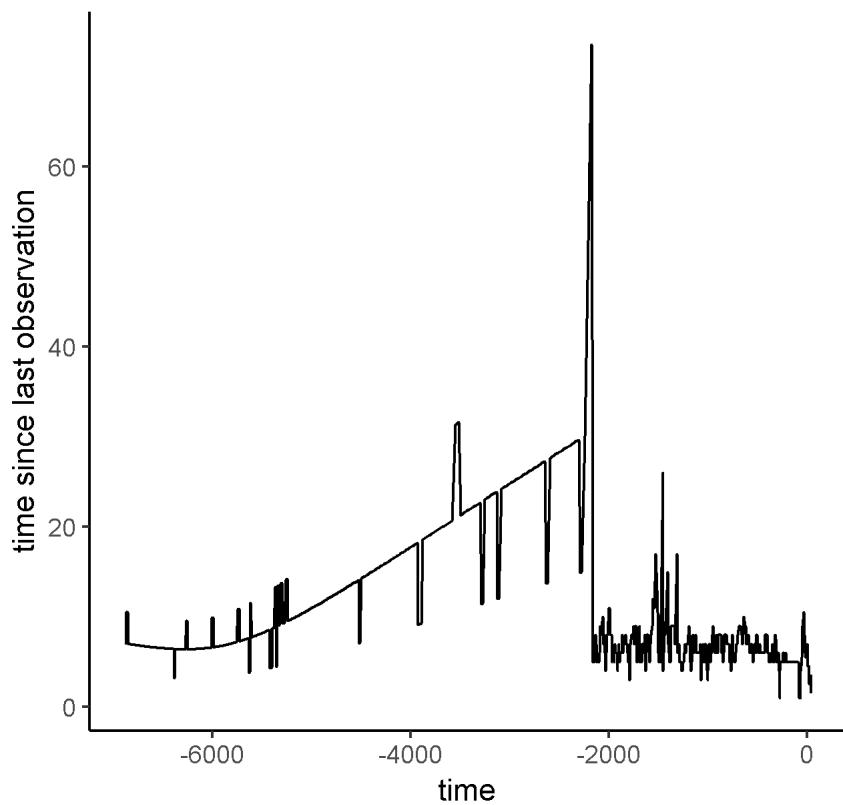


Figure 6.2: The amount of time elapsed between observations.

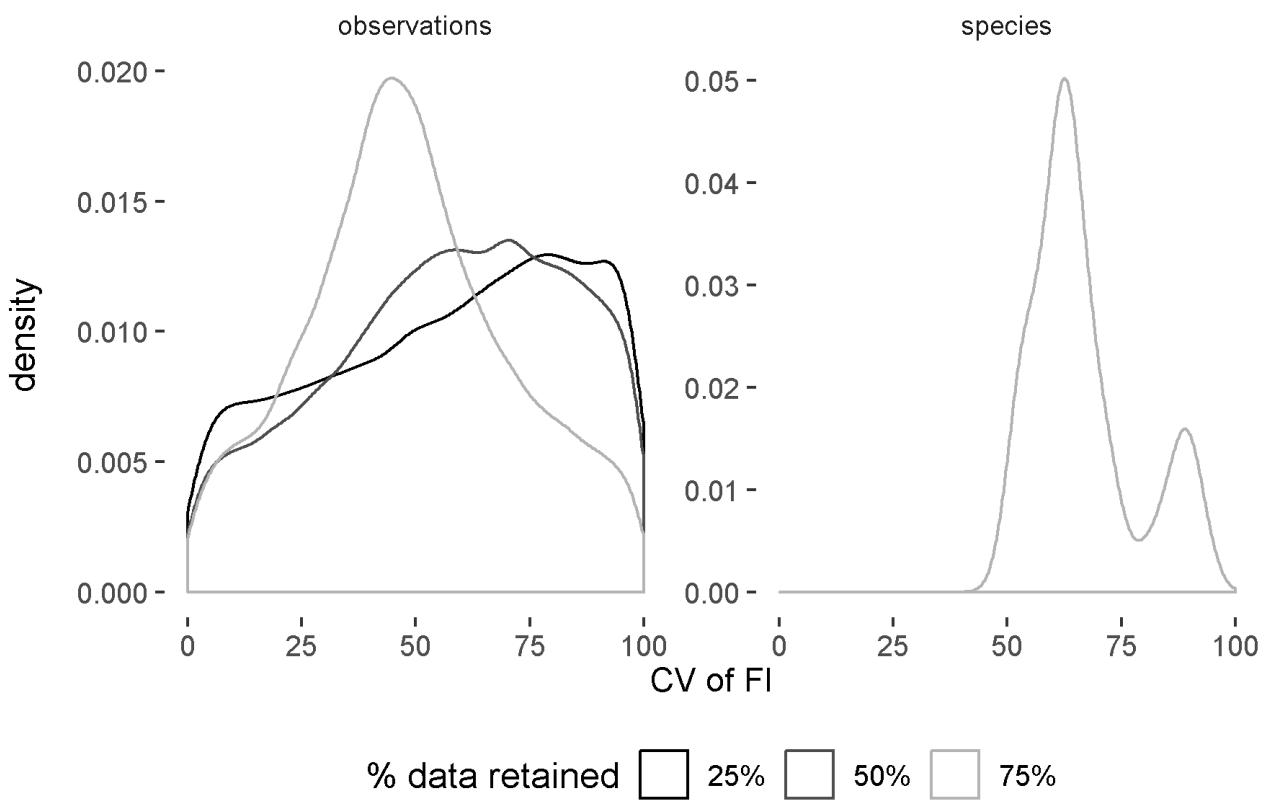


Figure 6.3: Local regression (loess) smoothing of a dominant species in the paleodiatom community, extit*Anomoeoneis costata* varies with the span parameter, making it difficult to justify smoothing the data prior to calculating various regime detection metrics.

2389 Some studies detrend the original time series prior to data aggregation and calculation
2390 of regime detection metrics. I did not detrend the original data for two reasons. First,
2391 the authors of the original paper analysing this dataset (Spanbauer *et al.*, 2014) did
2392 not detrend species time series. Like Spanbauer *et al.* (2014) I only scaled the original
2393 data, rather than detrending. Second, detrending a time series requires yet another
2394 subjective decision by the data analyst. For example, a “spanning” parameter must
2395 be chosen when detrending (smoothing) non-linear time series using local regression
2396 (Loess) regression (see Figure 6.3). Other smoothing methods are being explored for
2397 both detrending (e.g., PcR; Beck *et al.*, 2018) and regime shift identification (e.g.,
2398 generalized additive modelling; Beck *et al.*, 2018). Finally, this data exhibits rapid
2399 and drastic shifts in community composition *and* contains a disproportionate amount
2400 of dominant versus non-dominant species. Consequently, most species contain more
2401 zero than non-zero observations, which makes loess smoothing difficult. Although this
2402 chapter concerns impacts of data quality and quantity based on hypothetical data
2403 collection and analytical decisions, adding yet another parameter necessitates another
2404 layer of comparative analysis. Future work studying the impact of detrending, data
2405 scaling, outlier removal, and other related decisions would be of value in understanding
2406 the efficacy of these and other regime detection measures in real-world situations.

2407 6.5 Conclusion

2408 In this chapter I provide additional evidence for the sensitivity of select regime detection
2409 measures to information (data) quality and quantity loss. The loss of data quantity
2410 was simulated by randomly sampling subsets of both the species and the temporal
2411 observations, and the reduction in data quality manifests as a function of removing
2412 whole species from the community profile. Previous studies of the robustness of
2413 univariate regime detection metrics have found similar results, suggesting the measures

2414 fail in numerous real-world ecological conditions (Andersen *et al.*, 2009; Contamin
2415 & Ellison, 2009). This chapter also highlights the relative insensitivity of the new
2416 velocity metric (see Chapters 3, 5) to data and information quality and quantity (e.g.,
2417 Figure 6.2.4) loss.

2418 **6.6 Acknowledgements**

2419 This study was conceptualized at the International Institute for Applied Systems
2420 Analysis (IIASA) as part of the Young Scholars Summer Program in 2018. I thank
2421 my IIASA program supervisors, Drs. Brian Fath and Elena Rovenskaya, for their
2422 advisement of this project and IIASA scientists Drs. Matthias Jonas, Chai Molina,
2423 Piotr Zebrowski for feedback on study design.

2424 **Chapter 7**

2425 **Grassland Obligates and Declining**
2426 **Birds Operate Near Edges of Body**
2427 **Mass Distributions**

2428 **7.1 Introduction**

2429 Animal body mass distributions have been used to identify scaling structures of
2430 ecological communities (Holling, 1992; Allen & Holling, 2002; Allen, 2006). Using
2431 statistical methods to identify gaps, or discontinuities, in body mass distributions,
2432 some patterns are observed within and across taxonomic groups and biomes. Given
2433 the ubiquity of discontinuities identified in body mass distributions of fauna and
2434 social systems (Allen, 2006), the ecological significance of these patterns may prove
2435 useful in understanding ecosystem structure and functioning (Angeler *et al.*, 2016).
2436 Various hypotheses are postied as drivers of the observed discontinuities in animal
2437 body mass distributions, including those related to resource use (the Energetic and
2438 Textural Discontinuity hypotheses), community interactions, biogeography, and evolu-

tion/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). A recent study of the Central United States, including and beyond the Great Plains ecoregions, used discontinuity analysis of body mass distributions to identify the locations of what they refer to as ‘spatial regimes’ over an approximately 50 year period (Roberts *et al.*, 2019). The authors concluded that a spatial regime boundary exists in this region, and has been moving poleward as a consequence of large-scale drivers including woody planty encroachment, fire suppression, and climate change. Using the boundaries identified in this study I seek to determine whether this ‘shifting spatial regime’ manifests in the grasslands of the Central Great Plains. 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 first use discontinuity analysis of avian body mass distributions to identify the scaling

2466 structures of local avian communities in the Prairie Potholes, Central Mixed Grass,
2467 and Eastern Tall Grass regions of the central Great Plains of North America. I then
2468 use these distributions to determine whether the shifting spatial regime proposed by
2469 Roberts *et al.* (2019) manifests in the grassland bird community which is most suscep-
2470 tible to native grassland habitat loss or degradation (Murray *et al.*, 2008). Although
2471 I find no evidence to support the hypothesis of the spatial regime boundary suggested
2472 in Roberts *et al.* (2019), the results from this study support previous hypotheses that
2473 vulnerable species operate at the ‘edge’ of body mass aggregations.

2474 **7.2 Methods**

2475 **7.2.1 Study area**

2476 A recent study (Roberts *et al.*, 2019) identified what they refer to as spatial regimes
2477 across a large portion of the central United States [see Figure ??]. The authors
2478 hypothesize that a spatial regime boundary exists in the Central Great Plains and
2479 suggests it has exhibited a Northward shift at a rate of $\sim \frac{0.05^\circ \text{ latitude}}{\text{year}}$. The authors
2480 used discontinuity analysis to identify these ‘spatial regimes’, using the body massess
2481 of breeding bird communities. Their hypothesized spatial regime boundary occurs
2482 at 39° latitude in year 1970, 39.5° latitude in year 1985, 40° latitude in year 2000,
2483 and 40.5° latitude in year 2015 (see Figure 7.2). Sampling sites were classified each
2484 year as belonging to either the Southern or Northern regime according to whether the
2485 location was below (Southern) or above (Northern) the regimes identified by Roberts
2486 *et al.* (2019). The study area is designed such that there is minimal crossing of very
2487 different BCR, or habitat types (Figure 7.1). In other words, this study area largely
2488 falls within BCRs which can be generally classified as grassland habitat (BCR 11,
2489 Prairie Potholes; BCR 19, Central Mixed Grass ; BCR 22, Eastern Tall Grass). Using

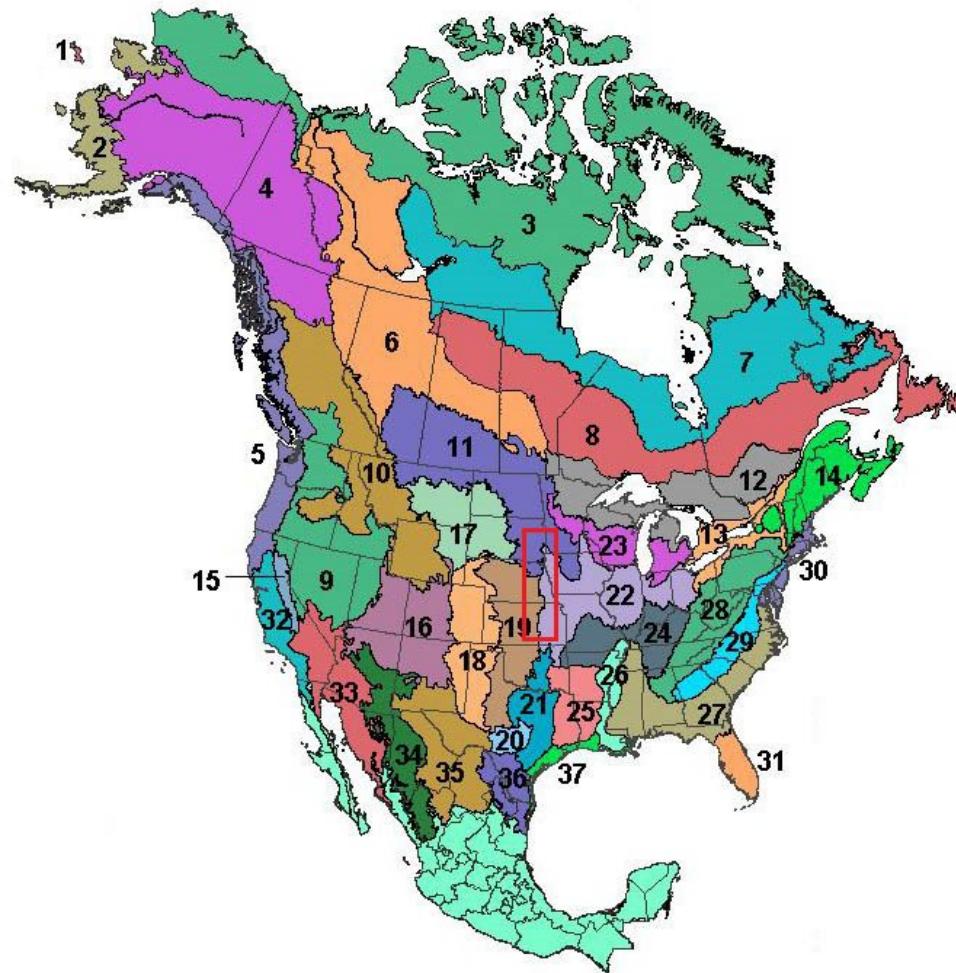


Figure 7.1: My study area (red box) overlaying the terrestrial Breeding Conservation Regions (BCR) in North America.

2490 this design we should expect that the functional groups within our avian communities
2491 should be similar across BCR boundaries, despite a potential turnover in species
2492 identity. Accounting for change in habitat across space allows us to assume that any
2493 observed change in the scaling structure of the avian community is due to changes in
2494 habitat and resource availability such that similar species are or are not included in
2495 the community.

2496 7.2.2 Data**2497 Avian census data - North American Breeding Bird Survey**

2498 I constructed body mass distributions using route-level data from the North American
2499 Breeding Bird Survey hereafter (NABBS; Sauer *et al.*, 2014). The NABBS uses
2500 citizen scientist volunteers to annually collect data using a standardized roadside,
2501 single observer, 3-minute point count protocol and has organized data collection
2502 annually across North America (Figure 4.1) since 1966. Each roadside survey consists
2503 of 50, 3-minute point counts (data collected using sight and sound) along ~ 24.5
2504 mile stretch of road. Although the point counts are designed to collect estimates of
2505 relative abundance, the method for building body mass distributions used in this
2506 chapter require only presence absence data. I therefore converted abundances to
2507 presence-absence data. I considered a species as ‘present’ if it was detected on the
2508 year in question or the ± 1 year to account for potential false negative observations
2509 (i.e., a species was not detected in the NABBS route despite its presence in the local
2510 community).

2511 **Identifying avian census locations** To determine whether the spatial regime
2512 shifts identified in Roberts *et al.* (2019) manifested in local avian community structure,
2513 I restricted analysis to the grassland habitat of the Central Great Plains. All routes
2514 falling within a rectangular area bounded by coordinates 37.8° and 44.5° latitude, and
2515 -101° and -95.5° longitude [see Figure 7.2]. I retained all NABBS routes which used
2516 the sampling protocol ‘101’, which is the standard method for conducting NABBS
2517 point count surveys.

2518 **Avian body mass data** Species operating at similar spatial and temporal scales
2519 are those which are close in body size as identified using statistical aggregation

2520 identification techniques [Allen *et al.* (1999); Section ??]. The interactions among
2521 species within a single body mass aggregation are presumed to experience a higher
2522 frequency and intensity of interspecific interactions with each other as opposed to
2523 those in different aggregations (Peterson *et al.*, 1998). Although some species of birds
2524 are sexually size dimorphic, I am unaware of any sexually size-dimorphic birds that
2525 would likely operate at different spatial and temporal. Therefore, I constructed body
2526 mass distributions of each avian census using the sex-averaged body masses published
2527 in Dunning Jr (2007) (available for download at CRC press).

2528 **Removing species from analysis** Due to strict reliance on volunteers, some routes
2529 are not covered every year. Although NABBS volunteers attempt identify all species in
2530 the point-count area, biases exist in data collection. Rather than retain observations of
2531 cryptic or species with low detection rates, I removed select species from the censuses
2532 (see Methods section in Chapter 4 for further discussion of this topic). I analyzed
2533 species of the following taxonomic families: Accipitriformes, Apodiformes, Cathartiformes,
2534 Charadriiformes, Columbiformes, Coraciiformes, Cuculiformes, Galliformes,
2535 Gruiformes, Passeriformes and Piciformes. Although removing cryptic, nocturnal,
2536 and some crepuscular species (e.g. Caprimulgiformes) from the analysis may yield
2537 a more conservative body mass distribution, including them may result in correctly
2538 identifying additional scaling structures (or body mass aggregations) in some routes
2539 but not in others. This method of exclusion also results in a loss of some medium- and
2540 larger-bodied Ciconiiformes (Podicipediformes, Phoenicopteriformes, Ciconiiformes;
2541 e.g. grebes, pelicans).

2542 **Taxonomic munging of the census data** Although the NABBS survey reports
2543 species-specific abundances, some birds are only classified to genera or order. Common
2544 examples of these species are those which are nearly indistinguishable from each

2545 other (e.g., Glossy Ibis and White-faced Ibis), birds which are difficult to see under
2546 certain conditions (e.g., hummingbirds, fast-moving hawks or accipiters), or species
2547 whose songs are similar. Numerous species were presented as identified to family or
2548 genus (e.g., Accipiter sp., Buteo sp., and Trochilids sp.) and others are categorized as
2549 hybrid.

2550 I made decisions regarding species-specific classification based on the North American
2551 breeding range maps provided by the Cornell Lab of Ornithology. Many unidentified
2552 species were easily categorized given the lack of overlap in species' ranges in our study
2553 area. For example, *Baeolophus bicolor* is nearly indistinguishable from *Baeolophus*
2554 *atricristatus*, however *B. atricristatus* is not known to occur in our study area (Figure
2555 7.1)—therefore all accounts classified as either *B. bicolor* or *B. atricristatus* were
2556 classified as the former. This example occurred for. Informed decisions like of this
2557 nature were made regarding the following unidentified species, where the second name
2558 in the binomial was assigned as the species preceding the “/”: *Passerina cyanea* /
2559 *amoena*, *Corvus brachyrhynchos* / *ossifragus*, *Petrochelidon pyrrhonota* / *fulva*, *Corvus*
2560 *brachyrhynchos*, *Quiscalus major* / *mexicanus*, *Pipilo maculatus* / *erythrophthalmus*,
2561 *Sturnella magna* / *neglect*, *Plegadis chihi* / *falcinellus*, *Coccyzus erythrophthalmus* /
2562 *americanus*, *Empidonax traillii* / *alnorum*, *Icterus galbula* / *bullockii*, *Nyctanassa*
2563 *nycticorax* / *violacea*, and *Poecile atricapillus* / *carolinensis* were all classified according
2564 to their known distributions. I classified unidentified hummingbirds (*Trochilid* sp.)
2565 as *Selasphorus rufus*, and unidentified Terns (Tern sp.) as *Chlidonias niger*. All
2566 unidentified Accipiters (Accipiter sp.), Buteos (Buteo sp.), and Gulls (Gull sp.) were
2567 removed from analysis entirely as there are no clear differences in the probability of
2568 occurrence in our study area.

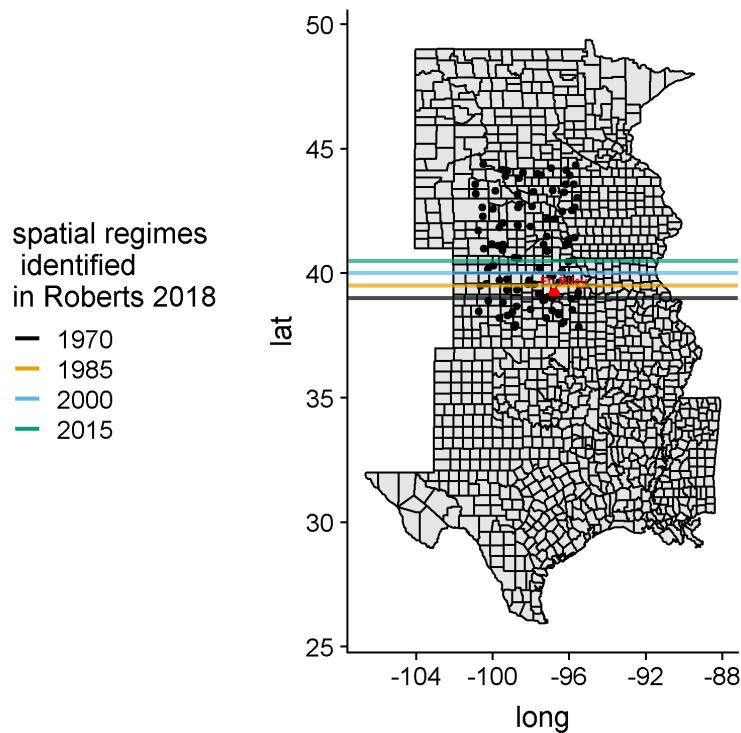


Figure 7.2: North American Breeding Bird Survey routes (points), latitudinal locations (horizontal bars) of the spatial regimes identified by *roberts2019shifting*.

2569 **Identifying species of interest**

2570 Allen *et al.* (2006) propose alternative hypotheses for the relative locations of species
2571 within the body mass aggregation distributions as a function of ‘distance-to-edge’,
2572 a measure indicating the distance (in log-mass units) of each species to the edge of
2573 a body mass aggregation (methods described in Section ??). This distance-to-edge
2574 measure is zero when the species falls at the edge of a statistically identified body
2575 mass aggregation. This species is often referred to as an ‘edge species’. To determine
2576 the effect of spatial regime shifts on edge species, I identified three types of species
2577 of interest: (1) grassland obligates species, (2) species with widespread population
2578 declines in the study area, and (3) a combination of these groups. All remaining
2579 species were classified as ‘other’.

2580 **Grassland obligate species** The spatial regimes identified in Roberts *et al.* (2019)
2581 are attributed to large-scale changes in the landscape, including woody plant invasion.
2582 The loss of native grassland in our study area due to land conversion is largely
2583 attributed to anthropogenic land use change (e.g., development) and fire suppression.
2584 Numerous species have been negatively impacted by this widespread habitat loss,
2585 but grassland obligates are particularly at risk. Grassland obligates should be strong
2586 indicators of the large-scale spatiel regime shifts identified in Roberts *et al.* (2019),
2587 given their high sensitivity to grassland habitat loss (Herkert, 1994). I identified North
2588 American grassland obligate species from the grey literature (Shriver *et al.*, 2005;
2589 Initiative *et al.*, 2009) and white literatures (Peterjohn & Sauer, 1999). Although some
2590 grassland obligates were positively impacted by the Conservation Reserve Program
2591 (CRP; Peterjohn & Sauer, 1999), this group of birds exhibited strong declines in North
2592 America until approximately 2003, the year the Farm Bill was adopted (Initiative *et*
2593 *al.*, 2009).

↑
which farm bill?

2594 **Declining species** I classified a species as 'declining' based on the results of the
2595 North American Breeding Bird Survey (Sauer *et al.*, 2014). The Patuxent Wildlife
2596 Research Center uses hierarchical modelling techniques to estimate the trends of
2597 species using the NABBS data at various spatial scales. Sauer *et al.* (2014) also
2598 provides estimates of data credibility according to data availability, number of routes
2599 used to build the population trend estimate, abundance, and probability of detecting
2600 a small change in population trend. These credibility scores are generated for multiple
2601 spatial extents: state-level, BCR-level, and across the three regions of the United
2602 States (Western, Central, and Eastern). Given the extent of this study, I considered
2603 the data credibility estimates using the Central Breeding Bird Survey Region, the
2604 Prairie Pothole BCR (BCR 11) and Eastern Tallgrass Prairie (BCR 22). A species was
2605 considered as declining only if the trend estimate was categorized as having moderate

2606 precision and abundance (blue) or having a deficiency (yellow). I considered the
2607 population trend estimates provided for the period of 1966 - 2015.

2608 **7.2.3 Statistical analysis**

2609 **Identifying scaling structure of avian communities using body mass distri-**
2610 **butions**

2611 Discontinuities in body mass distributions been quantified using various methods (e.g.,
2612 multivariate time series models, regression trees, and gap rarity index) which are
2613 collectively referred to as ‘discontinuity analyses’ (Allen, 2006; Stow *et al.*, 2007; Nash
2614 *et al.*, 2014a; Barichievy *et al.*, 2018). Using various methods, the discontinuous nature
2615 of body masses of ecological communities is well-documented, having been observed in
2616 various taxa of both terrestrial (Allen *et al.*, 2006) and aquatic (Spanbauer *et al.*, 2016)
2617 communities. Multiple methods are proposed for identifying discontinuities in body
2618 mass aggregations (Allen & Holling, 2001), including clustering algorithms (Stow *et*
2619 *al.*, 2007), body mass difference indices (Holling, 1992), gap rarity index (Restrepo &
2620 Arango, 2008), and more recently the discontinuity detector (Barichievy *et al.*, 2018),
2621 an extension of the gap rarity index (Restrepo & Arango, 2008).

2622 I used the discontinuity detector described in Barichievy *et al.* (2018), which uses likeli-
2623 hood to determine whether the observed data contains multiple modes as compared to
2624 that of a Gaussian (unimodal) distribution. This method requires multiple user-defined
2625 parameters, including an imputation resolution (1000) and a bootstrap sample size
2626 (1000) over which the null distribution is randomly sampled. I provided a slightly
2627 altered and annotated version of the functions `Neutral.Null` and `Bootstrap.Gaps`
2628 (first printed in Barichievy *et al.* (2018)) used to identify discontinuities in a contin-
2629 uous 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 ≥ 90 . I built route-level body mass distributions for each route-year combination by using presence absence data from the current, previous, and following year to account for observational and process errors impacting the detectability of a species within a single route. Using this method reduces the amount of species-specific, and consequently specific-body mass turnover within a route over time. This also assumes that an unobserved species is truly absent, an assumption which is difficult to avoid without a sophisticated occupancy modelling approach for each species in the community.

2646 Determining Effects of a Spatial Regime Boundary on Grassland 2647 Birds

2648 If the spatial regime shift occurred in the bird community, it should manifest in the
2649 local community scaling structure through one or both of species turnover and a
2650 shift in the number of body mass aggreagsations. I used linear mixed modelling to
2651 determine whether the local scaling structure and the location of grassland obligates
2652 and declining species within these scaling structures are impacted by the spatial regime
2653 boundaries proposed by Roberts *et al.* (2019).

2654 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 *aggreagtion*. 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 shift zones (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

2664 I therefore modelled the distance to edge of each species within each NABBS route as
 2665 a function of time, the regime location [South or North of the boundaries proposed
 2666 in Roberts *et al.* (2019); see Figure 7.2], and the species group (grassland obligate,
 2667 declining, declining grassland obligate, other). Although distance to edge was not
 2668 strongly correlated with body mass (Figure 7.7), I scaled and centered the response
 2669 variable (distance to edge; Y_i , where $i = \text{year}$), to avoid predictions beyond zero for the
 2670 unscaled response. Fixed effects included an interaction among year (β_{1i}) and regime
 2671 location (β_{2i}), and an interaction among regime (β_{2i}) and species group (β_3). Random
 2672 intercepts were estimated for each species (b_1) within each NABBS route (b_2). That
 2673 is, species was nested within route. An autoregressive lag-1 correlation structure was
 2674 assumed for the random intercept estimates. The model was fitted using restricted
 2675 maximum likelihood. The model was fitted using `nlme::lme` and was coded as >
 2676 `nlme::lme(distEdge.scaled ~ year.center * regime + regime * sppGroup, random = ~`
 2677 `1 | loc/aou, correlation = corAR1(form = ~ 1 | loc / aou), method = "REML")`

2678 7.3 Results

2679 7.3.1 Summary statsitics of censuses (NABBS data)

2680 Given the location of the study area (Figure 7.1) with respect to the location of the
 2681 contiguous Central Great Plains, fewer NABBS routes falling into the Southern regime
 2682 were anlaysed than those in the Northern (Table 7.3, Figure ??). Likely due to the
 2683 increase in the total number of routes surveyed over time across the entire North
 2684 American Breeding Bird Survey region, species richness increased over time within

2685 our study area (Figure 7.13). Annual turnover rates were relatively low but became
2686 more dispersed over time (Figure 7.14; Table 7.4).

2687 **Species of interest**

2688 A total of 163 species were considered for analysis across the entire study area. 18
2689 were classified as grassland obligate species, or species deemed highly sensitive to
2690 changes in amount and quality of grassland habitat. Of the grassland obligates,
2691 13 were considered as declining species. An additional 49 species were classified as
2692 non-grassland obligate and declining (Table 7.1).

2693 The total (Figure 7.3) and mean (Figure 7.4) number of birds counted within each
2694 species group was relatively constant across time in the Prairie Potholes and Eastern
2695 Tallgrass Prairie bird conservation regions, but fluctuation in stop totals appeared
2696 greater in the Badlands and Prairies BCR (Figures 7.4, 7.3)). The latter BCR
2697 comprises a much smaller portion of the study area (Figure 7.1) and accordingly
2698 the coefficient of variation (CV; ratio of deviation to the mean) around the stop
2699 totals (Figure 7.5) was highest in this region. It is worth noting the high CV (CV is
2700 considered low when $< \sim 40$) in all regions.

2701 **7.3.2 Statistical Analysis**

2702 **Identifying scaling structure in body mass distributions**

2703 Discontinuity analysis was conducted to identify the aggregations in the body mass
2704 distributions of 103 routes over a 50 period across the Central Great Plains (Figure
2705 7.1)). Discontinuity analysis suggested discontinuities existed in all routes analysed,
2706 and were relatively similar within NABBS routes over time. The number of body mass
2707 aggregations identified within each NABBS route using the discontinuity detector

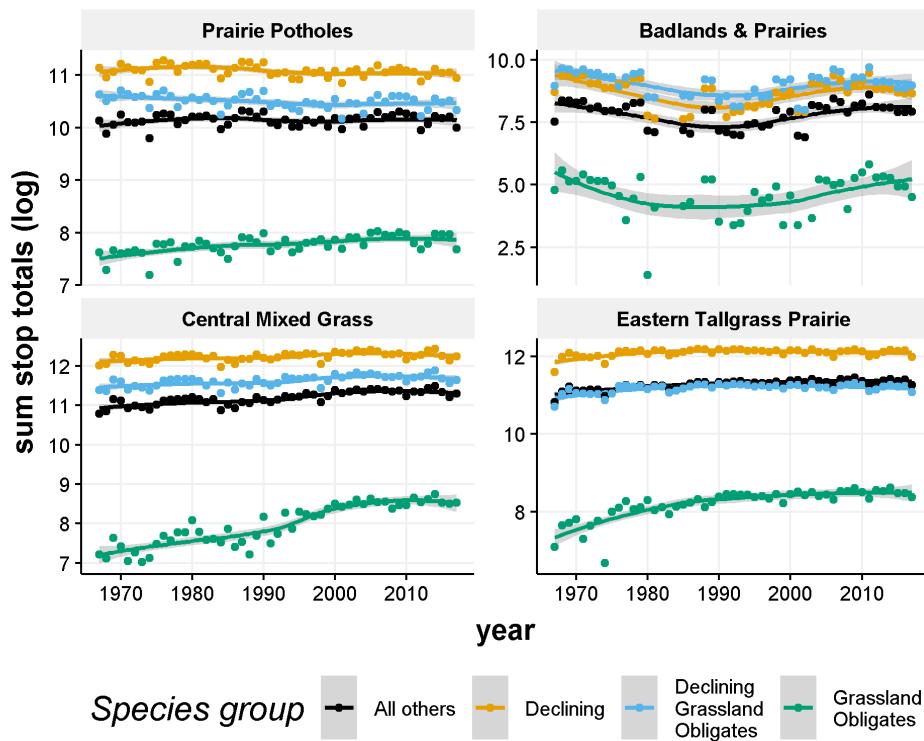


Figure 7.3: Total number of birds across the entire study area per species group per year.

(Barichievy *et al.*, 2018) was similar was similar across time (Figure 7.6a) and was approximately normally distributed across all survey-year combinations ($\bar{x} = 4.7, \sigma = 1.6$; Figure 7.6a). Species richness at the route level was strongly positively correlated with the number of aggregations (Figure 7.7). The distance to edge variable was statistically, but not strongly, correlated with body mass $r = -0.02, p = < 0.01$, and this relationship was similar across species groups except grassland obligates.

The discontinuities in the body mass distribution identified appeared relatively similar over time at most NABBS routes (Figure 7.8). 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 7.8 is representative of the ~5 NABBS locations falling in this area of expected changes shifting).

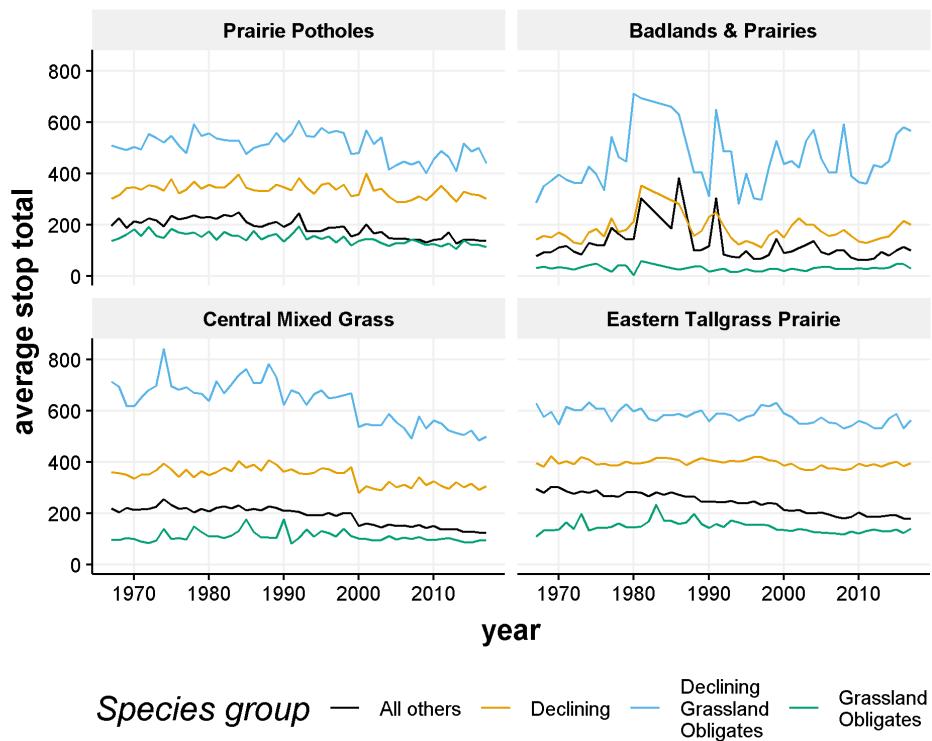


Figure 7.4: Average number of birds across the entire study area per species group per year.

2720 Linear Mixed Effects Analysis of Distance to Edge

2721 Declining species and declining grassland obligate species were located closer to the
 2722 edge than the ‘other’ species, while grassland obligate species were further from the edge
 2723 than ‘other’ species (Table 7.2). Similar trends held for declining grassland obligates
 2724 and grassland obligates in the Northern regime location. There was, however, no
 2725 evidence to suggest additive effects of the regime location or year, or their interactions
 2726 (Table 7.2). The lack of additive effects but the presence of multiplicative effects
 2727 of the regime location (Table 7.2) strengthens the support for the differences in
 2728 grassland obligates and declining grassland obligates with respect to all other species,
 2729 however, the confidence intervals around the estimates of declining grassland obligates,
 2730 grassland obligates suggests the evidence for such an effect is relatively weak (Figure
 2731 @ref(fig:intrxnPlot_regime1)).

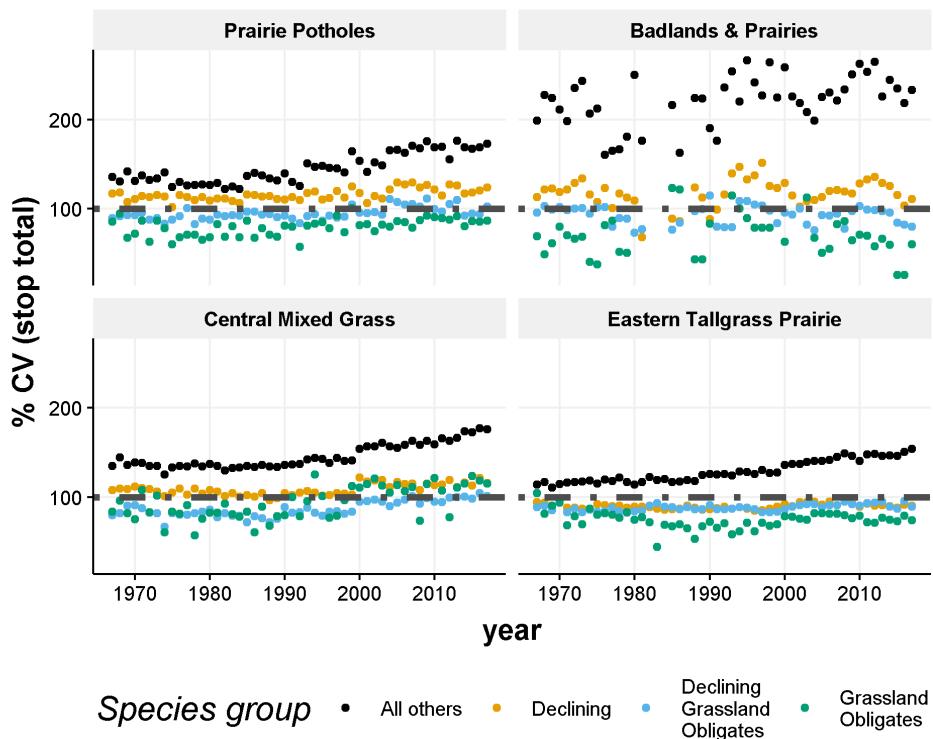


Figure 7.5: Average number of birds across the entire study area per species group per year.

2732 This is unsurprising given the distribution of grassland obligate body masses is highly
 2733 skewed right relative to the remaining species (Figures 7.9), ??)). Many grassland
 2734 obligate species have small body masses, reducing the probability that they will appear
 2735 in different body mass aggregations, however, depending on the local bird community
 2736 identities, local grassland obligates may all have similar body masses, removing this
 2737 effect. The larger, non-declining grassland obligates occur in a relatively small portion
 2738 of our study area, Shortgrass Prairie BCR (Figures 7.1,7.11).

2739 7.4 Discussion

2740 South-North shifts in the past 50+ years have been demonstrated in large scale pro-
 2741 cesses, including bird populations and ranges (Sorte & III, 2007) and plant hardening
 2742 zones (Mckenney *et al.*, 2014). The concept of spatial regimes was recently introduced

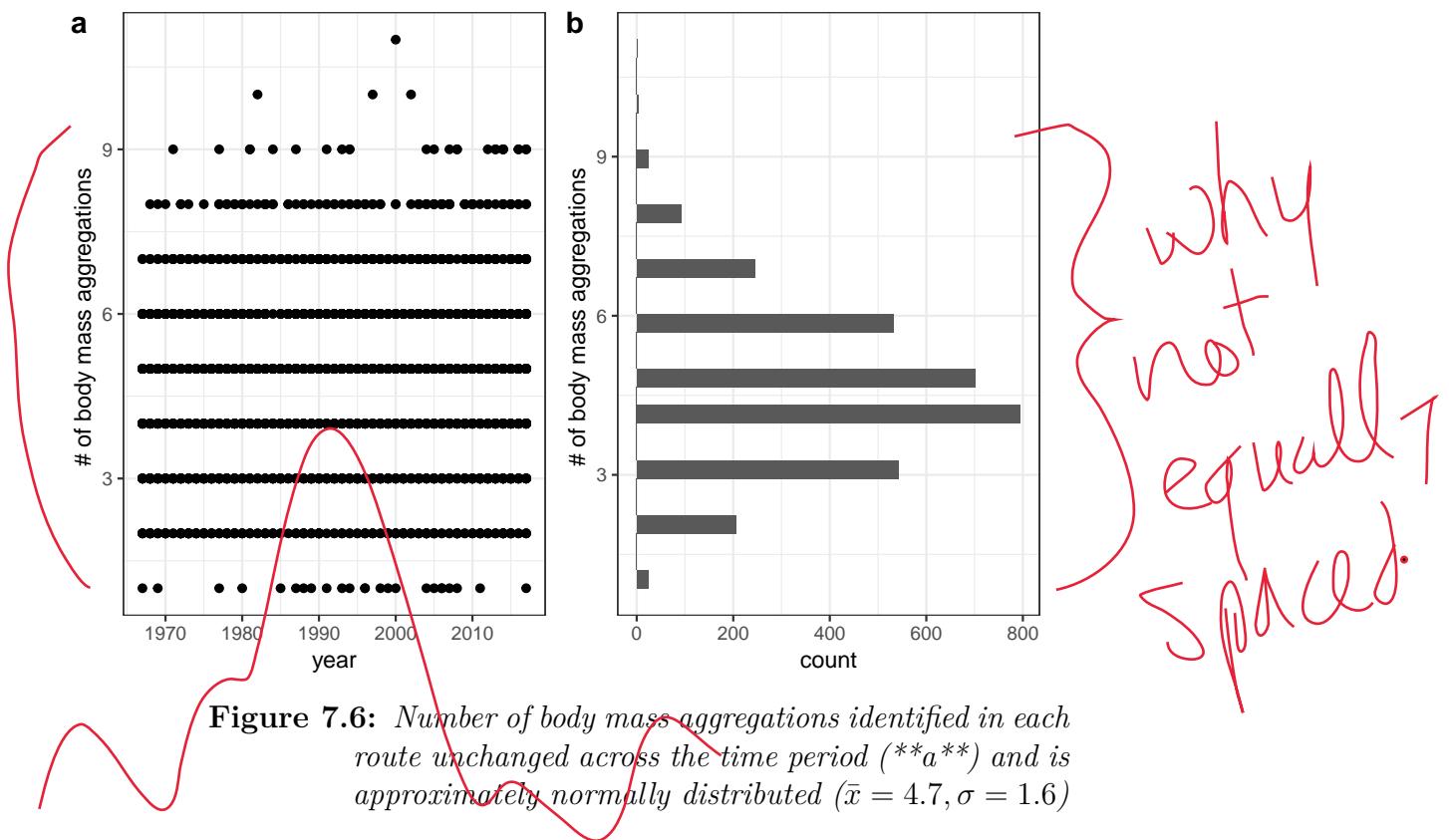


Figure 7.6: Number of body mass aggregations identified in each route unchanged across the time period (**a**) and is approximately normally distributed ($\bar{x} = 4.7, \sigma = 1.6$)

as a way of describing or identifying large-scale shifts of this nature (Sundstrom *et al.*, 2017; Roberts *et al.*, 2019), despite the lack of rigorous testing on the methods used to identify these shifts. In this Chapter, I used discontinuity analysis of bird communities, under the assumption of a zone of shifting ‘spatial regimes’ (Roberts *et al.*, 2019) to determine whether local-scale changes in the scaling structure were impacted by these shifts. Although I found moderate support for the hypothesis that declining species operate at the edges of body mass aggregations, I did not find sufficient support for the impact of these proposed spatial regime shifts.

Although the body mass distributions of terrestrial communities at small scales tends to differ from the distribution at larger spatial scales (Blackburn & Gaston, 1994), numerous studies confirm the evidence for ‘discontinuities’ in these distributions at across scales for a variety of phenomenon (Allen *et al.*, 1999; Wardwell & Allen, 2009; Nash *et al.*, 2014a; Spanbauer *et al.*, 2016). The jury is still out, however, as numerous

Table 7.2: 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)

2756 studies are finding a lack of evidence for discontinuities in their study sysytems (Manly,
 2757 1996; Siemann & Brown, 1999; Bibi *et al.*, 2019).

Table 7.3: 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

Table 7.4: Summary statistics for annual species richness and annual turnover in all NABBS routes in study area.

Year	Annual Richness			Annual Turnover		
	\bar{x}	σ	N	\bar{x}	σ	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

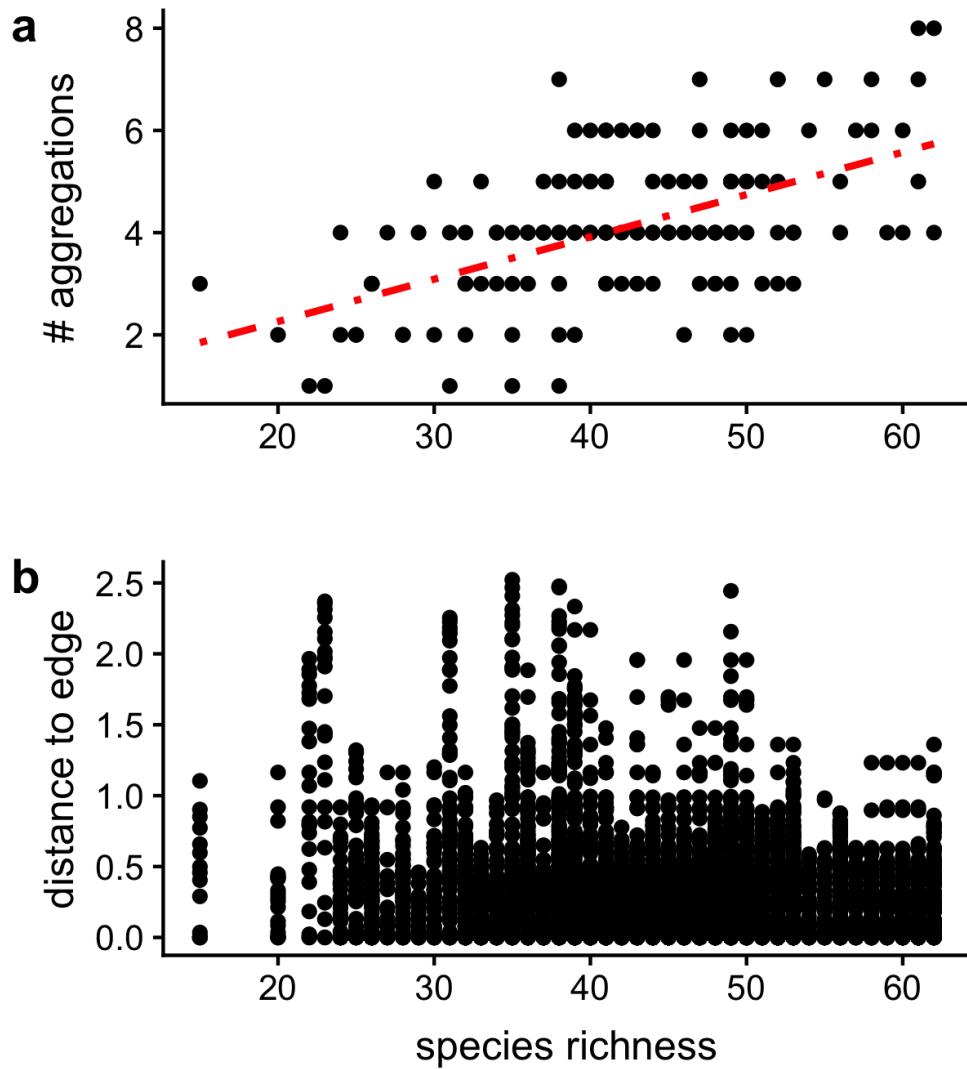


Figure 7.7: Relationship between species richness per route and (a) the number of aggregations identified in body mass distributions and (b) distance to the edge (units log body mass) of aggregations.

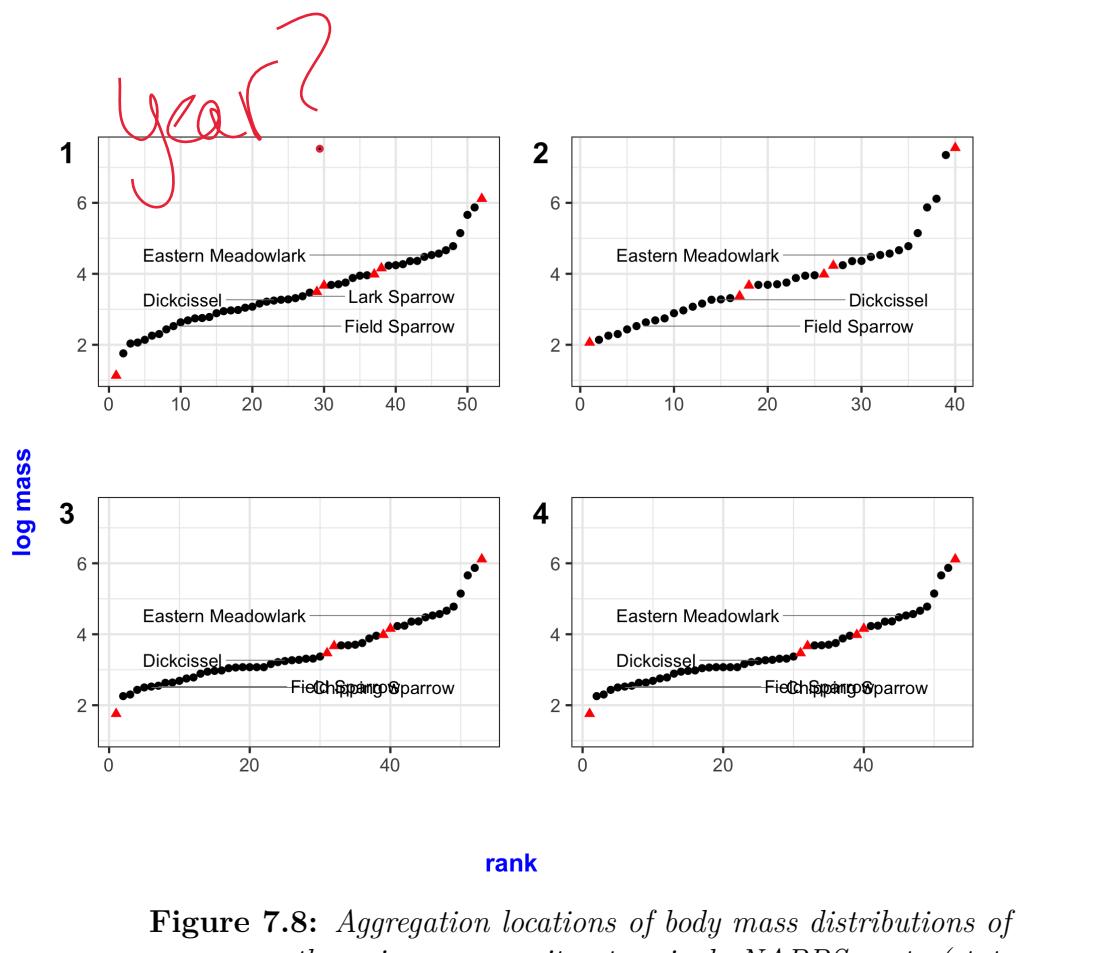


Figure 7.8: Aggregation locations of body mass distributions of the avian community at a single NABBS route (state 7 route 24) appear relatively similar across time.

What are red points?

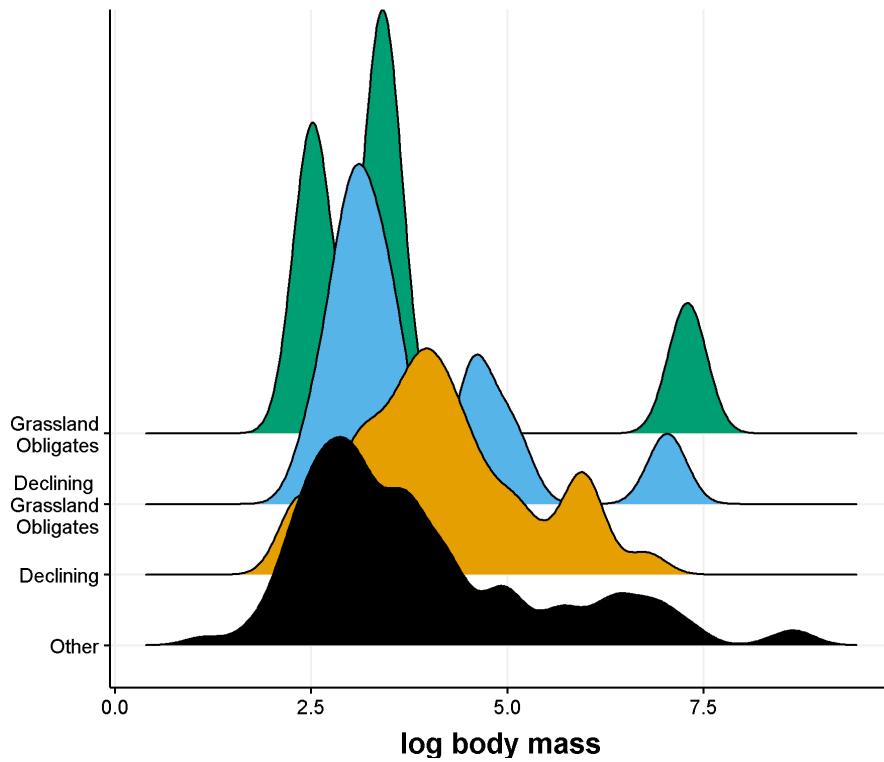


Figure 7.9: *Body mass distribution for species in the study area over the entire time period varies by species group. Distributions represent the species pool for each group over the entire study area and all years.*

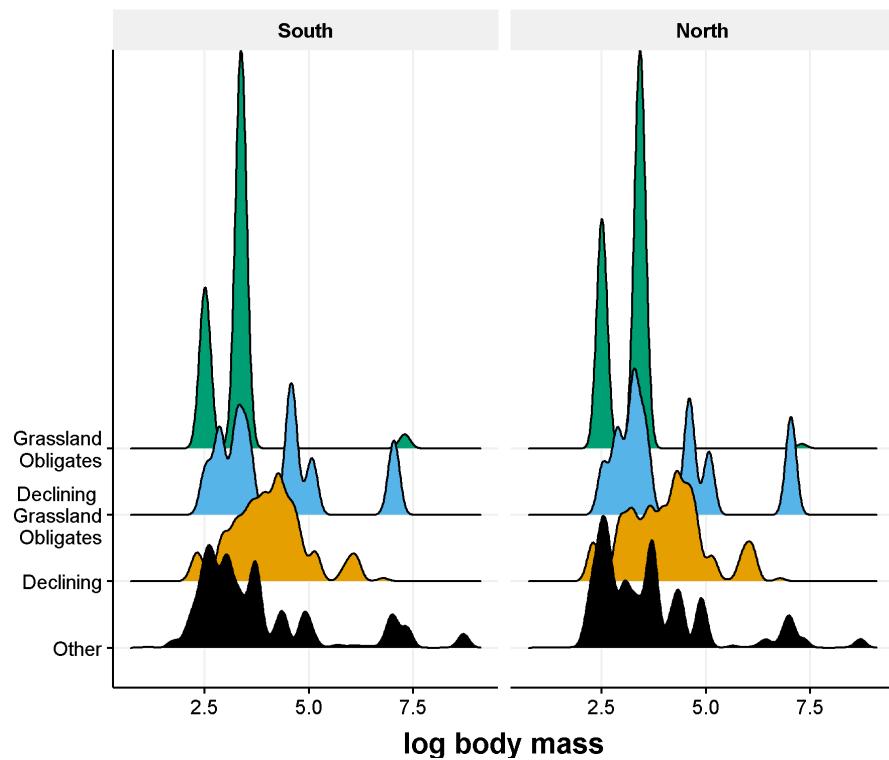


Figure 7.10: The body mass distribution of declining species appears to differ slightly between the Southern and Northern regimes.

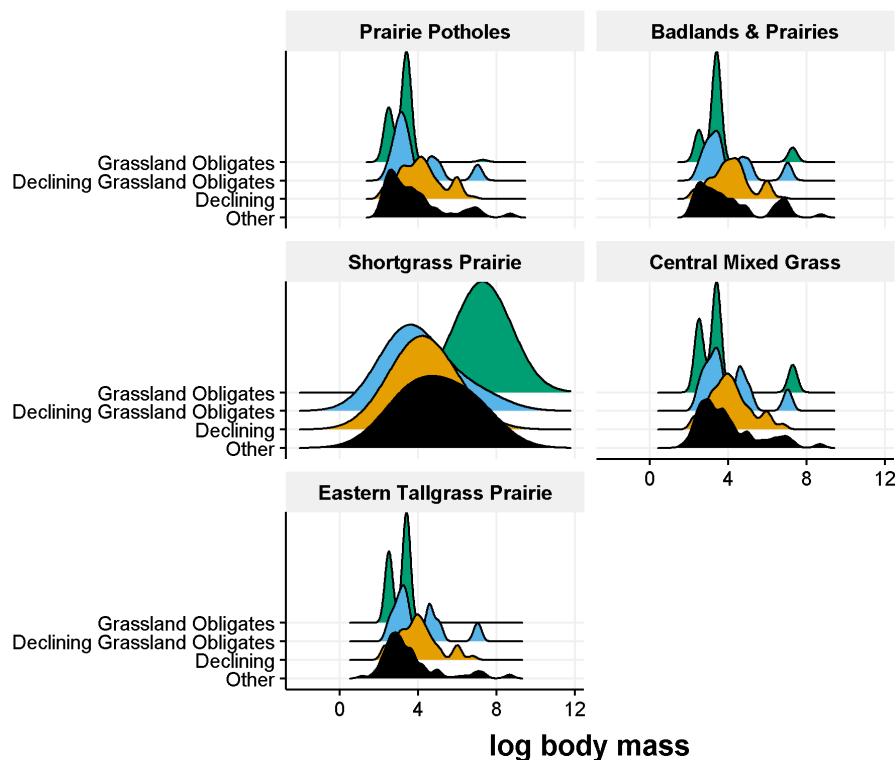


Figure 7.11: *Body mass distribution for species in the study area over the entire time period varies by species group. Distributions represent the species pool for each group and Bird Conservation Region over all years.*

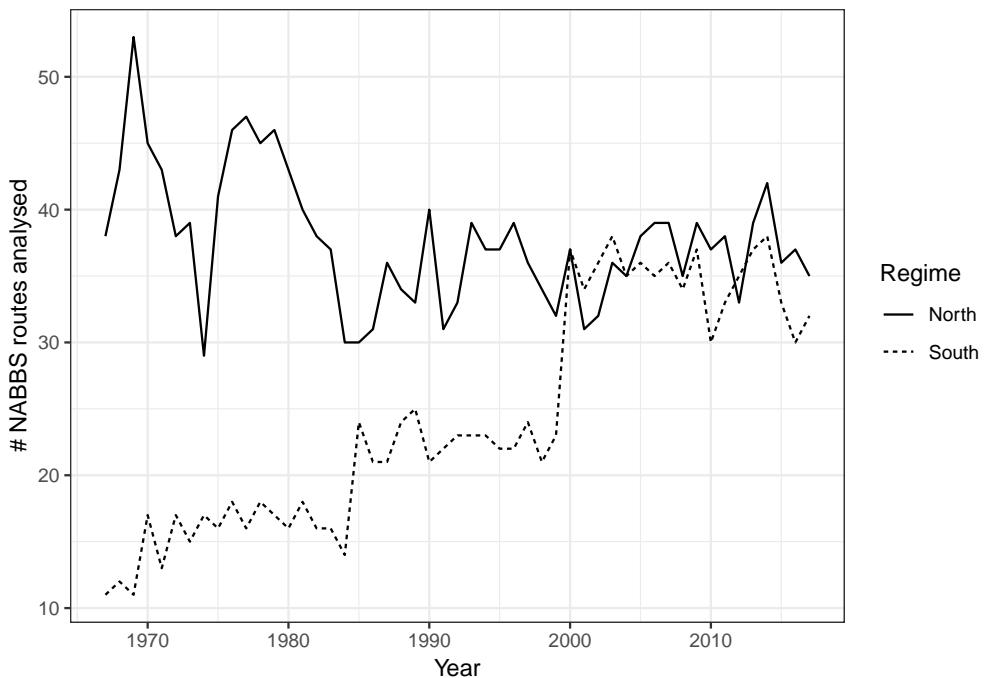
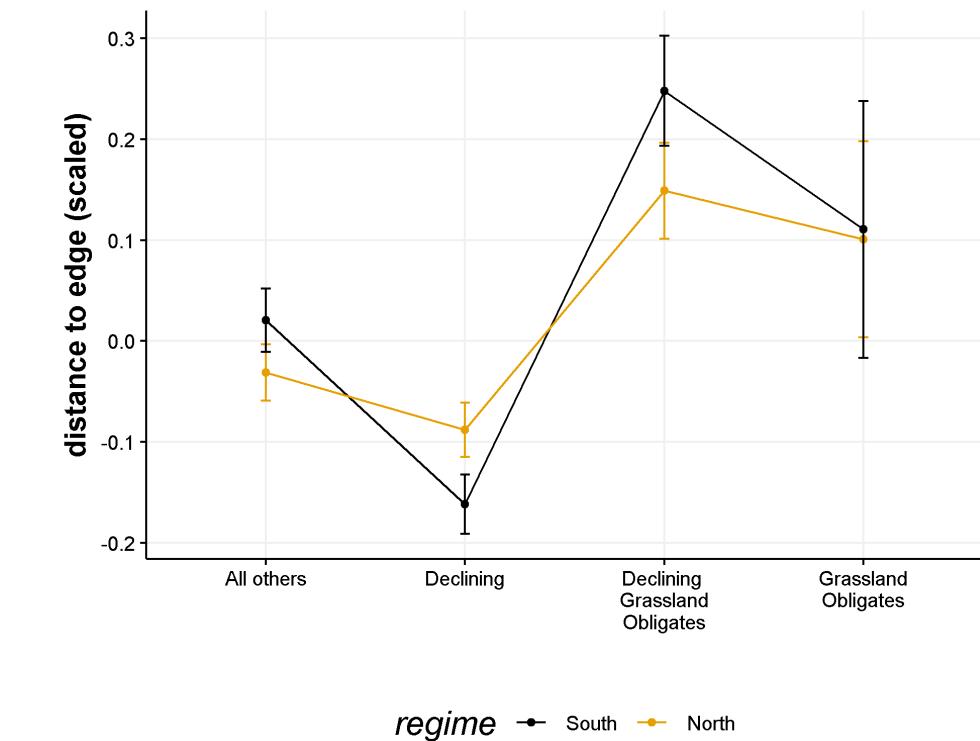


Figure 7.12: Number of NABBS routes analysed per year.
Some routes are not sampled annually due to volunteer availability, environmental conditions, or route discontinuation.

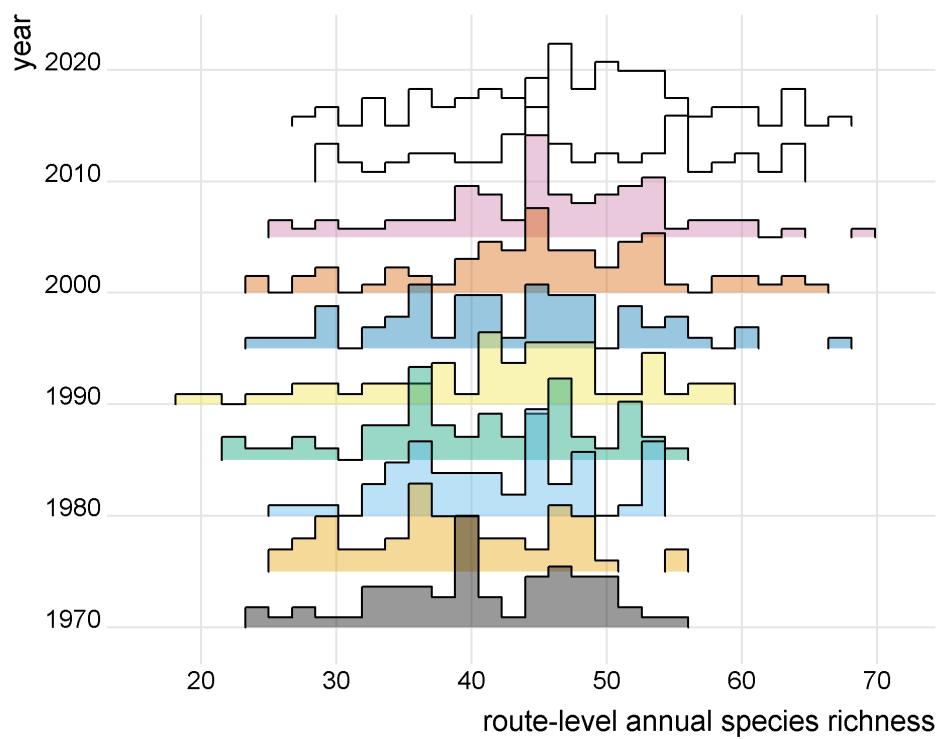


Figure 7.13: Species richness increases over time across the entire study area.

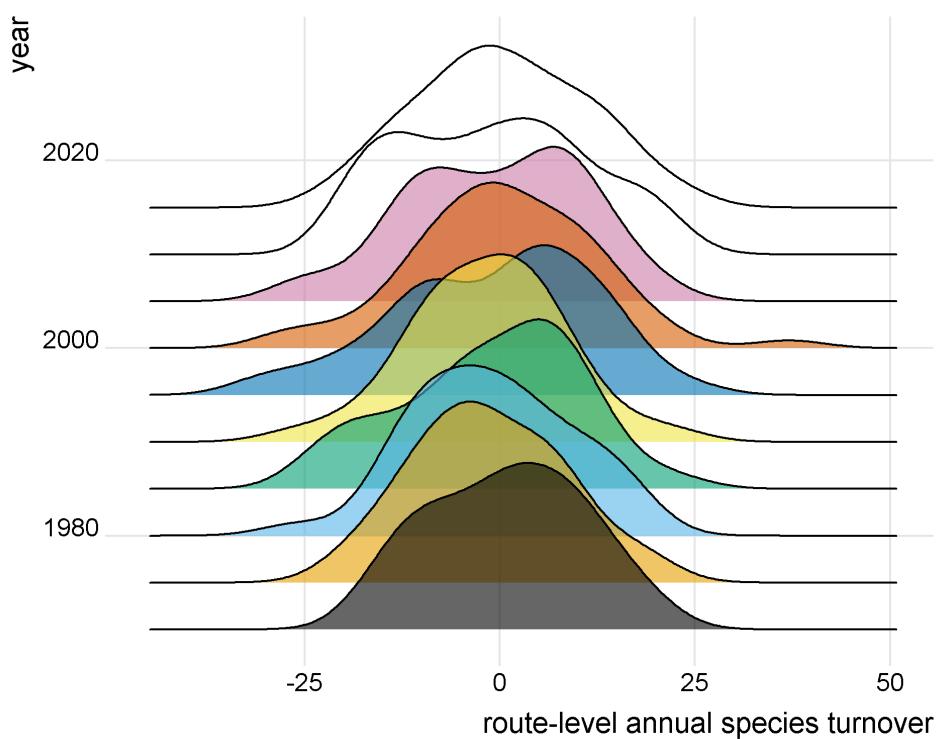


Figure 7.14: Variance in species turnover increases over time across the entire study area.

2758

Chapter 8

2759

Conclusions

2760 Climate change is expected to induce an increase in both the intensity and frequency
2761 of rapid ecological change or disturbance, impacting social systems, potentially to
2762 the detriment of human communities most vulnerable. Identifying and forecasting
2763 these changes is critical for community and ecological planning, management, and
2764 disaster mitigation. Because ecological and social systems are tightly coupled, it is
2765 commonplace to use ecological indicators to identify change and potential changes that
2766 may impact these systems. Many papers introducing or discussing regime detection
2767 measures suggest the ecologist uses multiple lines of evidence, ranging from historical
2768 observations to ecological modelling results, for identifying an ecological regime shift
2769 (Lindegren *et al.*, 2012). Although valid, comparing results of multiple methods or lines
2770 of evidence within a single system has yielded inconsistent results, and inconsistent
2771 results can result in either improper conclusions, or in what I am calling **method**
2772 **mining**. That is, a dataset is analyzed using until a sufficient number of methods
2773 yield affirmative results.

2774 8.1 Method mining regime detection methods

2775 Many regime detection measures have yet to be properly and statistically (or nu-
2776 merically) scrutinized. However, it should be noted that, in part due to both (i)
2777 the popularity and (ii) the sheer number of ‘new’ methods a handful of authors¹.
2778 Ecological indicators (a.k.a. indices, metrics) have been suggested as ‘early-warning
2779 indicators’ of ecological regime shifts or abrupt change (Chapters 1 and 2) and are
2780 methods of measurement designed to provide inference about one or more unobserved
2781 or latent processes, are inherently biased. Regardless of the state of the theory sup-
2782 porting *regime shifts* in ecology, ecological indicators and the methods for calculating
2783 them should be heavily scrutinized prior to being used in an ecological management
2784 or policy-making setting. Rather, new methods (indices, metrics, etc.) are being
2785 introduced into the literature at a rate exceeding that at which they are scrutinized
2786 (Chapter 2). This dissertation demonstrates that, while potentially useful, regime
2787 detection metrics are inconsistent, not generalizable, and are currently not validated
2788 using probabilities or other statistical measurements of certainty.

2789 8.2 Ecological data are noisy

2790 Regime detection metrics appear more reliable when the signal-to-noise ratio is high
2791 (Chapter 2, Chapter 5, Taranu *et al.*, 2018). Ecological systems are noisy, and the
2792 observational data we are collecting at large scales (e.g., the North American Breeding
2793 Bird survey), is noisy. Using methods incapable of identifying meaningful signals in
2794 noisy data appears futile, yet, methods for doing so are increasingly introduced in the
2795 scientific literature (Chapter 2).

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

2796 8.3 Data collection and munging biases and limits

2797 findings

2798 Regime detection measures and other ecological indicators can signal (see (??)) various
2799 changes in the data, however, understanding what processes are embedded in the
2800 signals (i.e., removing the noise) requires expert judgement. And because a consequence
2801 of data collection and data analysis limits the extent to which we can identify and
2802 infer processes and change within an ecological system, **I suggest the practical**
2803 **ecologist scrutinizes her data prior to identifying and conducting analyses,**
2804 including those that are purely exploratory. By collecting and analysing data, the
2805 ecologist has defined the boundaries of the system *a priori*^+ (+ Beisner *et al.*, 2003
2806 states this eloquently as, “The number and choice of variables selected to characterize
2807 the community will be determined by what we wish to learn from the model”). The
2808 influence of state variable selection is ignored by some metrics (e.g. Fisher Information,
2809 Eason *et al.*, 2014 and velocity, Chapter 5), in that the resulting measure is composite
2810 and carries no information regarding the influence of state variables on the metric
2811 result. The actual limitations to the system should be, theoretically, known as a result
2812 of bounding the system. Inference beyond this system is extrapolation, and should be
2813 treated as speculation, especially when not accompanied by a measure of uncertainty
2814 around one’s predictions.

2815 8.4 Common Limitations of Regime Detection

2816 Measures

2817 Limitations of the findings in this dissertation and of the regime detection methods
2818 used herein are largely influenced by the **data collection, data munging** processes.

2819 Although the below mentioned points may seem logical to many, these assumptions are
2820 overlooked by many composite indicators, including regime detection measures.

- 2821 1. Signals in the indicators are restricted to the ecological processes captured by
2822 the input data. Extrapolation occurs when processes manifest at scales different
2823 than the data collected [resolution; Chapter 4]
- 2824 2. Normalization and weighting techniques often impact results (whether ecological
2825 or numerical) (Appendices .1.3 and 8.4)
- 2826 3. Data aggregation techniques often impact results (Chapter 6)
- 2827 4. Some indices fail to generalize across systems or taxa (see Chapters 1 and 2)

2828

Appendix A: R package

2829

regimeDetectionMeasures

2830 This appendix contains example analysis associated with the R Package,
2831 `regimeDetectionMeasures`. Development source code for this package is available
2832 on GitHub as a compressed file at <https://github.com/TrashBirdEcology/>
2833 `regimeDetectionMeasures/archive/master.zip` or at <https://github.com/>
2834 `TrashBirdEcology/regimeDetectionMeasures`.

2835

.1 Example Analysis

2836

.1.1 Measures/metrics calculated

2837 This package will calculate a various regime detection metrics that have been used to
2838 ‘detect ecological regime shifts’. A ‘new’ metric, **distance travelled** is also calculated
2839 (Burnett and others, *in prep*).

2840 **Composite measures:**

- 2841 1. Distance travelled -see also package `distanceTravelled`.
- 2842
- 2843 2. Fisher Information

2844 3. Variance Index

2845 Single-variable measures:

2846 1. Skewness (mean and mode versions)

2847

2848 2. Kurtosis

2849

2850 3. Variance

2851

2852 4. Mean

2853

2854 5. Mode

2855

2856 6. Coefficient of variation, CV

2857

2858 7. Autocorrelation lag-1 (using stats::acf)

2859 **.1.2 Example analysis**

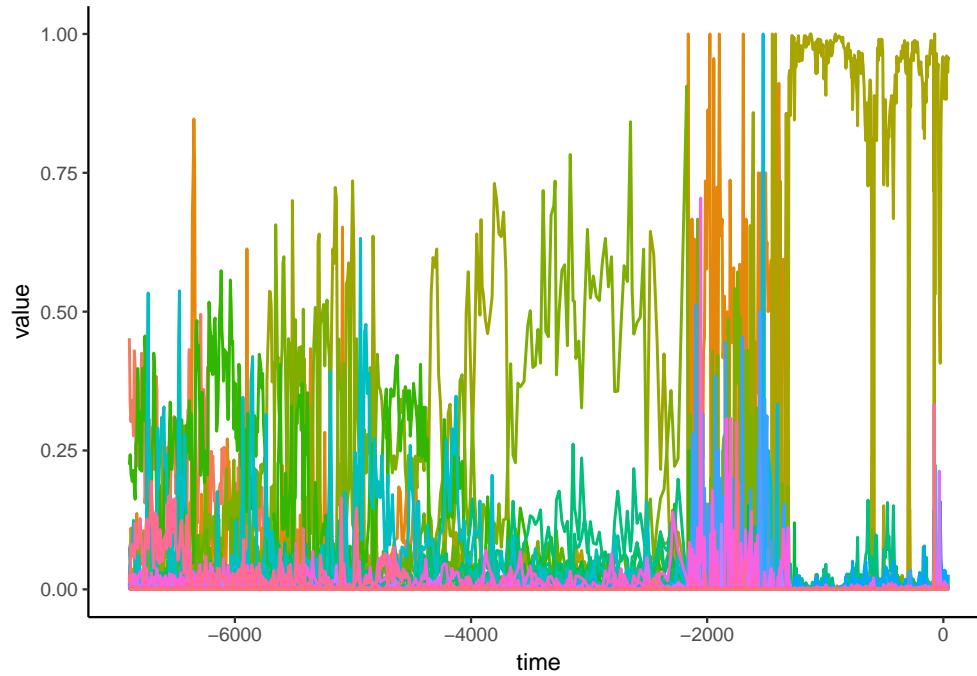
```
knitr::opts_chunk$set(echo=FALSE, eval=TRUE, message=FALSE, warning=FALSE, out.width=1000)

library(regimeDetectionMeasures)

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

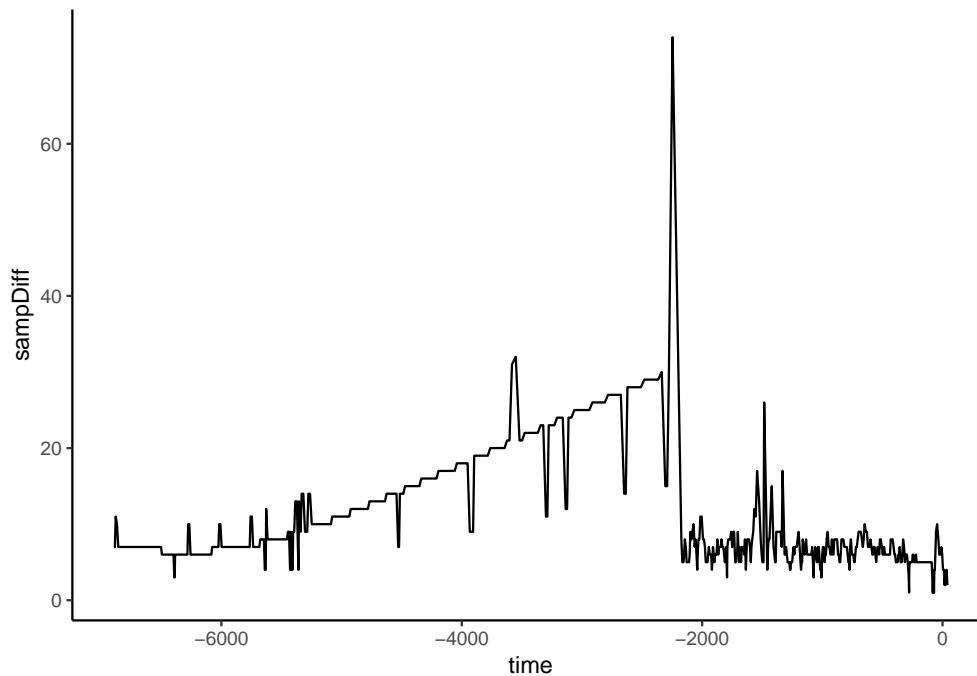
.1.3 Plots

2861 Plot the original time series as relative abundances.



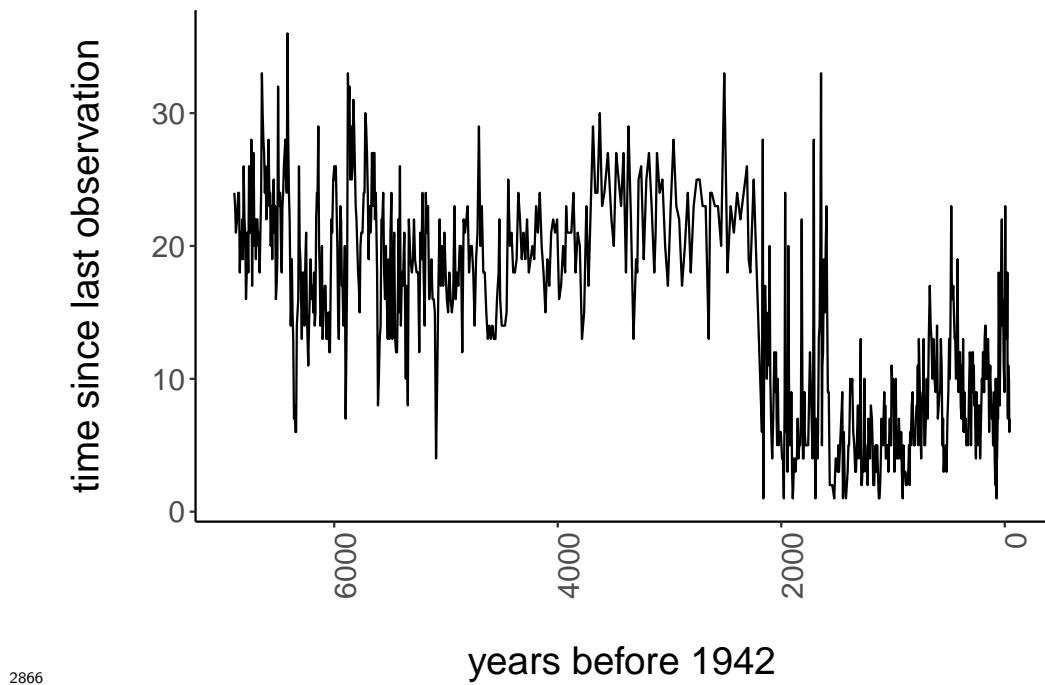
2862

2863 Plot the time elapsed among observations.



2864

2865 Plot species richness over time.



²⁸⁶⁷ Plot the distance travelled (s) and velocity of distance travelled (v)

²⁸⁶⁸ Plot the moving-window metrics (FI, VI, and the univariate EWSs)

	metricValue	cellID_min	cellID_max	winStart	winStop	metricType	
1	0.157575450026098			1	-1746.63	-12.6975	VI
2	0.186533784869325			1	-1820.16	-86.2275	VI
3	0.211304354488511			1	-1893.69	-159.7575	VI
4	0.235451240210219			1	-1967.22	-233.2875	VI
5	0.25306928580876			1	-2040.75	-306.8175	VI
6	0.260835104190261			1	-2114.28	-380.3475	VI

	variable	cellID
1	NA	NA
2	NA	NA
3	NA	NA
4	NA	NA
5	NA	NA

6 NA NA

2869 **Appendix B: R package**

2870 **bbsRDM**

2871 This appendix contains a vignette associated with the R Package, bbsRDM. De-
2872velopment source code for this package is available on GitHub as a compressed
2873file, <https://github.com/TrashBirdEcology/bbsRDM/archive/master.zip> or at
2874 <https://github.com/TrashBirdEcology/rRDM>.

2875 This vignette runs through the capabilities of the bbsRDM package, which relies on the
2876package `trashbirdecology::regimeDetectionMeasures`. Although this package
2877can be used to calculate and visualize BBS data using time series, the example at
2878hand runs presents an application to spatial transects.

2879 **.1 Example Analysis**

2880 **.1.1 Load packages & create local directories**

2881 There are a lot of dependencies to load.

```
2882 ## Re-install often as this package is under major development.  
2883 # devtools::install_github("trashbirdecology/regimedetectionmeasures", force = F  
2884 library(regimeDetectionMeasures)
```

```
library(sp)
library(raster)
library(feather)
library(bbsRDM)
library(here)
```

2882 .1.2 Download the BBS data and save to file locally

- 2883 If necessary, download all the state data. This takes 10-15 minutes, so only run if you
2884 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 or 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] "NOT DOWNLOADING BBS DATA. If you wish to download the BBS data, please remove",
  [1] "NOT DOWNLOADING BBS DATA. If you wish to download the BBS data, please remove"
}

```

2885 .1.3 Create a sampling grid

2886 Next, build a sampling grid to force route information onto a regular gridded area.

2887 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)

```

- 2888 See if the results are already saved to file. This will save us some computational
 2889 time when knitting the document. Otherwise, will take about a minute to calculate,
 2890 depending on the size of birdsData
- 2891 Now we load in the BBS data from the feathers we created and align with the sampling
 2892 grid. This requires a bit of memory, proceed with caution. If it's already saved in the
 2893 tempResultsDir (see Section @ref(import.ind)) then it will just upload 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")))

# A tibble: 5,884,135 x 12
# ... with 5,884,125 more rows, and 2 more variables: rowID <int>,
#       colID <int>

  year countrynum statenum route   bcr     lat    long    aou stoptotal cellID
  <int>      <int>    <int> <int> <int> <dbl> <dbl> <int>      <dbl> <int>
1  1967        840      2     1     27  34.9 -87.6  2890       28  4833
2  1967        840      2     1     27  34.9 -87.6  3131       5   4833
3  1967        840      2     1     27  34.9 -87.6  3160      23   4833
4  1967        840      2     1     27  34.9 -87.6  3870      14   4833
5  1967        840      2     1     27  34.9 -87.6  3940      12   4833
6  1967        840      2     1     27  34.9 -87.6  4050       4   4833
7  1967        840      2     1     27  34.9 -87.6  4090       8   4833
8  1967        840      2     1     27  34.9 -87.6  4120       4   4833
9  1967        840      2     1     27  34.9 -87.6  4160       4   4833
10 1967        840      2     1     27  34.9 -87.6  4171       1   4833
# ... with 5,884,125 more rows, and 2 more variables: rowID <int>,
#       colID <int>

```

2894 .1.4 Subset the BBS data by species and/or functional traits

2895 (OPTIONAL but highly recommended)

2896 Although subsetting the speices is optional, I recommend removing waterfowl, wading
2897 birds, and shorebirds from analyses, especially as the spatial extent of the analysis
2898 increases.

2899 .1.5 Subset species according to AOU species codes (i.e. by
2900 family, genera, etc..)

2901 For this example we will remove shorebirds, wading birds, and waterfowl (i.e., AOU
2902 species' codes 0000:2880). *See R/`subsetByAOU.R` source code or documentation for
2903 options (see: `subset.by`)

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

2904 .1.6 Subset species by trait, body mass, taxonomically,
2905 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)
```

2906 .1.7 Calculate regime detection metrics across space or
2907 time

2908 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 sites)  
fill = 0  
  
# Minimum number of sites (if spatial) or years (if temporal) required to be included  
min.samp.sites = 8
```

```

# Minimum number of sites (if spatial) or years (if temporal) required to be within a
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 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()
}

```

2909 Define the years we want to analyze. For this (spatial) example, we will analyze only
 2910 every fifth year in the available data.

2911 .1.8 Conduct analysis

2912 This section will loop through `years.use` and `dir.use`, running each BBS route
 2913 (temporal analysis) or spatial transect by year (spatial analysis) at a time. Results are
 2914 saved in directories created in @ref(#createDirs)

```

# **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 subdirectory

# browser()

calculateMetrics(dataIn = birdData, metrics.to.calc, direction = direction, ye

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

```

2915 **.1.9 Import and munge the results to prepare for visualiza-**
2916 **tion**

2917 First, import and combine the results as created in @ref(#calcMetrics) This chunk
2918 will import the EWS results and the distance results separately, combining each into
2919 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


# Import distance results

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

  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)
}

```

2920 Next, get the results to align with our sampling grid for visualizing results across
 2921 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")
```

2922 .1.10 Visualize results: one regime detection metric at a
2923 time

2924 First, specify plotting parameters. We can visualize either the distance results
2925 (`distResults`) or the early-warning signal results (`ewsResults`). Define the results
2926 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")
```

2927 Plot the individual transects. Note: please specify dirID.ind as desired spatial
2928 transect number, and dirInd as direction (E-W or N-S)

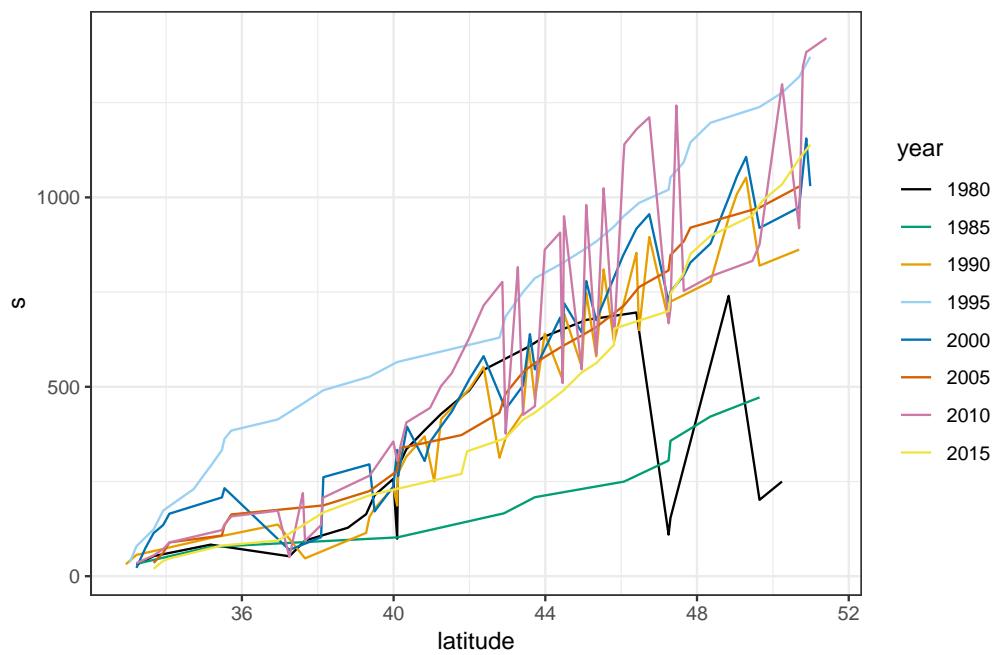
```
# Specify the transect # we want to see
dirID.ind <- 13

metric.ind <- "s"
```

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

```
pl1
```

South–North # 13



2929

2930 **Appendix C: Functions used to**

2931 **calculate discontinuities in avian**

2932 **body mass distributions.**

2933 **.1 About**

2934 This code was first published in Barichievy *et al.* (2018) and has been slightly modified
2935 and annotated for the purposes of this dissertation. This code was used to produce
2936 body mass discontinuities in Chapter #ref(discontinuity)

2937 **.2 Neutral.Null function**

```
2938 Neutral.Null <- function(log10.data, resolution = 4000) { Dmax = max(log10.data,  
2939 na.rm = FALSE) Dmin = min(log10.data, na.rm = FALSE) ds = (Dmax - Dmin) /  
2940 resolution MaxK = (Dmax - Dmin) / 2 MinK = ds * 2  
2941 #define h's to analyze ks = seq(MinK, MaxK, by = 1 / resolution)  
2942 # generate matrix bws = matrix(data = NA, nrow = length(ks), ncol = 1)  
2943 for (i in c(1:length(ks))) { # Calculate KS density estimate KSdens <- den-
```

```

2944 sity(log10.data, bw = ks[i], "gaussian", adjust = 1)

# Test if the ksdensity is unimodal

TF <- which(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)

2946 }
```

.3 Bootstrapping Function

```

2947 DD <- function(log10.data, hnull, Sample.N = 1000) { NNull <- density(log10.data,
2948   bw = hnull, "gaussian", adjust = 1) N <- length(log10.data)

2949   # generate matrix null.samples <- matrix(data = 0, ncol = Sample.N, nrow = N) for
2950   (i in 1:Sample.N) { #sample the null model rand.N <- sample(NNullx, N, replace =
2951     TRUE, prob = NNully) #calculate the gaps null.samples[, i] <- sort(rand.N, decreas-
2952     ing = FALSE) #put into the matrixi }

2953   # generate gaps gaps.log10.data <- diff(log10.data) gaps.null.samples <-
2954   diff(null.samples, decreasing = FALSE) # difference between random samples and 1st
2955   diff orig dat gap.percentile <- matrix(data = 0, nrow = length(gaps.log10.data), ncol
2956   = 1) for (i in 1:length(gaps.log10.data)) { # generate distribution of gaps per row
2957     (per gap rank) gap.percentile[i] <- ecdf(gaps.null.samples[i, ])(gaps.log10.data[i]) #
2958     returns the percentile at each observation

2959   } Bootstrap.gaps <- rbind(gap.percentile, 0) Bootstrap.gaps <- cbind(log10.data,
```

```
2961 Bootstrap.gaps) return(Bootstrap.gaps) }
```

2962 References

- 2963 Abadi, F., Gimenez, O., Arlettaz, R. & Schaub, M. (2010) An assessment of integrated
2964 population models: Bias, accuracy, and violation of the assumption of independence.
2965 *Ecology*, **91**, 7–14.
- 2966 Alheit, J., Möllmann, C., Dutz, J., Kornilovs, G., Loewe, P., Mohrholz, V. & Wasmund,
2967 N. (2005) Synchronous ecological regime shifts in the central Baltic and the North
2968 Sea in the late 1980s. *ICES Journal of Marine Science: Journal du Conseil*, **62**,
2969 1205–1215.
- 2970 Allen, C.R. (2006) Discontinuities in ecological data. *Proceedings of the National
2971 Academy of Sciences*, **103**, 6083–6084.
- 2972 Allen, C.R., Forys, E.A. & Holling, C. (1999) Body mass patterns predict invasions
2973 and extinctions in transforming landscapes. *Ecosystems*, **2**, 114–121.
- 2974 Allen, C.R., Garmestani, A., Havlicek, T., Marquet, P.A., Peterson, G., Restrepo, C.,
2975 Stow, C. & Weeks, B. (2006) Patterns in body mass distributions: Sifting among
2976 alternative hypotheses. *Ecology Letters*, **9**, 630–643.
- 2977 Allen, C.R. & Holling, C.S. (2001) Cross-scale morphology. *Encyclopedia of Environ-
2978 metrics*.
- 2979 Allen, C.R. & Holling, C.S. (2002) Cross-scale structure and scale breaks in ecosystems

- 2980 and other complex systems. *Ecosystems*, **5**, 315–318.
- 2981 Andersen, T., Carstensen, J., Hernandez-Garcia, E. & Duarte, C.M. (2009) Ecological
2982 thresholds and regime shifts: Approaches to identification. *Trends in Ecology &
2983 Evolution*, **24**, 49–57.
- 2984 Anderson, P.J. & Piatt, J.F. (1999) Community reorganization in the Gulf of Alaska
2985 following ocean climate regime shift. *Marine Ecology Progress Series*, 117–123.
- 2986 Angeler, D.G., Allen, C.R., Barichievsky, C., Eason, T., Garmestani, A.S., Graham,
2987 N.A., Granholm, D., Gunderson, L.H., Knutson, M., Nash, K.L. & others (2016)
2988 Management applications of discontinuity theory. *Journal of Applied Ecology*, **53**,
2989 688–698.
- 2990 Aria, M. & Cuccurullo, C. (2017) Bibliometrix: An r-tool for comprehensive science
2991 mapping analysis. *Journal of Informetrics*, **11**, 959–975.
- 2992 Bahlai, C.A., Werf, W. vander, O’Neal, M., Hemerik, L. & Landis, D.A. (2015)
2993 Shifts in dynamic regime of an invasive lady beetle are linked to the invasion and
2994 insecticidal management of its prey. *Ecological Applications*, **25**, 1807–1818.
- 2995 Barichievsky, C., Angeler, D.G., Eason, T., Garmestani, A.S., Nash, K.L., Stow, C.A.,
2996 Sundstrom, S. & Allen, C.R. (2018) A method to detect discontinuities in census
2997 data. *Ecology and Evolution*, **8**, 9614–9623.
- 2998 Batt, R.D., Carpenter, S.R., Cole, J.J., Pace, M.L. & Johnson, R.A. (2013) Changes
2999 in ecosystem resilience detected in automated measures of ecosystem metabolism
3000 during a whole-lake manipulation. *Proceedings of the National Academy of Sciences*,
3001 **110**, 17398–17403.
- 3002 Beaugrand, G. (2004) The north sea regime shift: Evidence, causes, mechanisms and
3003 consequences. *Progress in Oceanography*, **60**, 245–262.

- 3004 Beck, K.K., Fletcher, M.-S., Gadd, P.S., Heijnis, H., Saunders, K.M., Simpson, G.L.
3005 & Zawadzki, A. (2018) Variance and rate-of-change as early warning signals for
3006 a critical transition in an aquatic ecosystem state: A test case from Tasmania,
3007 Australia. *Journal of Geophysical Research: Biogeosciences*.
- 3008 Beisner, B.E., Haydon, D.T. & Cuddington, K. (2003) Alternative stable states in
3009 ecology. *Frontiers in Ecology and the Environment*, **1**, 376–382.
- 3010 Benedetti-Cecchi, L., Tamburello, L., Maggi, E. & Bulleri, F. (2015) Experimental
3011 perturbations modify the performance of early warning indicators of regime shift.
3012 *Current biology*, **25**, 1867–1872.
- 3013 Bennett, E.M., Peterson, G.D. & Gordon, L.J. (2009) Understanding relationships
3014 among multiple ecosystem services. *Ecology letters*, **12**, 1394–1404.
- 3015 Bestelmeyer, B.T., Ellison, A.M., Fraser, W.R., Gorman, K.B., Holbrook, S.J., Laney,
3016 C.M., Ohman, M.D., Peters, D.P., Pillsbury, F.C., Rassweiler, A. & others (2011)
3017 Analysis of abrupt transitions in ecological systems. *Ecosphere*, **2**, 1–26.
- 3018 Bibi, H., Raffaelli, D., Iqbal, M., Sharif, M. & Khattak, M.J.K. (2019) Body size pat-
3019 terns in stream communities: A test of holling?S textural discontinuity hypothesis.
3020 *Journal of the Entomological Research Society*, **21**, 17–35.
- 3021 Blackburn, T.M. & Gaston, K.J. (1994) Animal body size distributions: Patterns,
3022 mechanisms and implications. *Trends in Ecology & Evolution*, **9**, 471–474.
- 3023 Boettiger, C. & Hastings, A. (2012) Quantifying limits to detection of early warning
3024 for critical transitions. *Journal of the Royal Society Interface*, **9**, 2527–39.
- 3025 Boettiger, C., Ross, N. & Hastings, A. (2013) Early warning signals: The charted and
3026 uncharted territories. *Theoretical ecology*, **6**, 255–264.
- 3027 Brock, W.A. & Carpenter, S.R. (2010) Interacting regime shifts in ecosystems: Impli-

- 3028 cation for early warnings. *Ecological Monographs*, **80**, 353–367.
- 3029 3029 Brock, W. & Carpenter, S. (2006) Variance as a Leading Indicator of Regime Shift in
3030 Ecosystem Services. *Ecology and Society*, **11**.
- 3031 3031 Burthe, S.J., Henrys, P.A., Mackay, E.B., Spears, B.M., Campbell, R., Carvalho,
3032 L., Dudley, B., Gunn, I.D., Johns, D.G., Maberly, S.C. & others (2016) Do early
3033 warning indicators consistently predict nonlinear change in long-term ecological
3034 data? *Journal of Applied Ecology*, **53**, 666–676.
- 3035 3035 Butitta, V.L., Carpenter, S.R., Loken, L.C., Pace, M.L. & Stanley, E.H. (2017) Spatial
3036 early warning signals in a lake manipulation. *Ecosphere*, **8**, n/a–n/a.
- 3037 3037 Byrski, J. & Byrski, W. (2016) A double window state observer for detection and isolat-
3038 tion of abrupt changes in parameters. *International Journal of Applied Mathematics*
3039 and Computer Science
- 3040 3040 and Computer Science
- 3041 3041 Cabezas, H., Campbell, D., Eason, T., Garmestani, A., Heberling, M., Hopton,
3042 M., Templeton, J., White, D., Zanowick, M. & Sparks, R. (2010) San luis basin
3043 sustainability metrics project: A methodology for evaluating regional sustainability.
3044 USEPA. USA.
- 3045 3045 Cabezas, H. & Fath, B.D. (2002) Towards a theory of sustainable systems. *Fluid
Phase Equilibria*, **194**, 3–14.
- 3046 3046 Carpenter, S. & Brock, W. (2011) Early warnings of unknown nonlinear shifts: A
3047 nonparametric approach. *Ecology*, **92**, 2196–2201.
- 3048 3048 Carpenter, S.R. & Brock, W.A. (2006) Rising variance: A leading indicator of ecological
3049 transition. *Ecology letters*, **9**, 311–318.
- 3050 3050 Carpenter, S.R., Brock, W.A., Cole, J.J., Kitchell, J.F. & Pace, M.L. (2008) Leading
3051 indicators of trophic cascades. *Ecology Letters*, **11**, 128–138.

- 3052 Carpenter, S.R., Cole, J.J., Pace, M.L., Batt, R., Brock, W., Cline, T., Coloso, J.,
3053 Hodgson, J.R., Kitchell, J.F., Seekell, D.A. & others (2011) Early warnings of
3054 regime shifts: A whole-ecosystem experiment. *Science*, **332**, 1079–1082.
- 3055 Chartrand, R. (2011) Numerical differentiation of noisy, nonsmooth data. *ISRN
3056 Applied Mathematics*, **2011**.
- 3057 Clements, C.F. & Ozgul, A. (2016) Including trait-based early warning signals helps
3058 predict population collapse. *Nature Communications*, **7**.
- 3059 Clements, C.F. & Ozgul, A. (2018) Indicators of transitions in biological systems.
3060 *Ecology letters*, **21**, 905–919.
- 3061 Cobo, M.J., López-Herrera, A.G., Herrera-Viedma, E. & Herrera, F. (2011) An
3062 approach for detecting, quantifying, and visualizing the evolution of a research field:
3063 A practical application to the fuzzy sets theory field. *Journal of Informetrics*, **5**,
3064 146–166.
- 3065 Contamin, R. & Ellison, A.M. (2009) Indicators of regime shifts in ecological systems:
3066 What do we need to know and when do we need to know it. *Ecological Applications*,
3067 **19**, 799–816.
- 3068 Dakos, V., Carpenter, S.R., Brock, W.A., Ellison, A.M., Guttal, V., Ives, A.R., Kefi,
3069 S., Livina, V., Seekell, D.A. & Nes, E.H. van (2012a) Methods for detecting early
3070 warnings of critical transitions in time series illustrated using simulated ecological
3071 data. *PloS one*, **7**, e41010.
- 3072 Dakos, V., Carpenter, S.R., Nes, E.H. van & Scheffer, M. (2015) Resilience indica-
3073 tors: Prospects and limitations for early warnings of regime shifts. *Philosophical
3074 Transactions of the Royal Society B: Biological Sciences*, **370**, 20130263.
- 3075 Dakos, V., Van Nes, E.H., D'Odorico, P. & Scheffer, M. (2012b) Robustness of

- 3076 variance and autocorrelation as indicators of critical slowing down. *Ecology*, **93**,
3077 264–271.
- 3078 Davis, E.P. & Karim, D. (2008) Comparing early warning systems for banking crises.
3079 *Journal of Financial stability*, **4**, 89–120.
- 3080 DeAngelis, D.L. & Yurek, S. (2017) Spatially explicit modeling in ecology: A review.
3081 *Ecosystems*, **20**, 284–300.
- 3082 deYoung, B., Barange, M., Beaugrand, G., Harris, R., Perry, R.I., Scheffer, M. &
3083 Werner, F. (2008) Regime shifts in marine ecosystems: Detection, prediction and
3084 management. *Trends in Ecology & Evolution*, **23**, 402–409.
- 3085 Ditlevsen, P.D. & Johnsen, S.J. (2010) Tipping points: Early warning and wishful
3086 thinking. *Geophysical Research Letters*, **37**.
- 3087 Ducré-Robitaille, J.-F., Vincent, L.A. & Boulet, G. (2003) Comparison of techniques
3088 for detection of discontinuities in temperature series. *International Journal of
3089 Climatology: A Journal of the Royal Meteorological Society*, **23**, 1087–1101.
- 3090 Dunning Jr, J.B. (2007) *CRC handbook of avian body masses*, CRC press.
- 3091 Dutta, P.S., Sharma, Y. & Abbott, K.C. (2018) Robustness of early warning signals
3092 for catastrophic and non-catastrophic transitions. *Oikos*, **127**, 1251–1263.
- 3093 Eason, T. & Cabezas, H. (2012) Evaluating the sustainability of a regional system
3094 using Fisher information in the San Luis Basin, Colorado. *Journal of environmental
3095 management*, **94**, 41–49.
- 3096 Eason, T., Ching-Chuang, W., Sundstrom, S. & Cabezas, H. (2019) An information
3097 theory-based approach to assessing spatial patterns in complex systems. *Entropy*,
3098 **21**, 182.
- 3099 Eason, T., Garmestani, A.S. & Cabezas, H. (2014) Managing for resilience: Early de-

- 3100 tection of regime shifts in complex systems. *Clean Technologies and Environmental*
3101 *Policy*, **16**, 773–783.
- 3102 Fath, B.D. & Cabezas, H. (2004) Exergy and Fisher Information as ecological indices.
3103 *Ecological Modelling*, **174**, 25–35.
- 3104 Fath, B.D., Cabezas, H. & Pawłowski, C.W. (2003) Regime changes in ecological
3105 systems: An information theory approach. *Journal of theoretical biology*, **222**,
3106 517–530.
- 3107 Filatova, T., Polhill, J.G. & Ewijk, S. van (2016) Regime shifts in coupled socio-
3108 environmental systems: Review of modelling challenges and approaches. *Environmental*
3109 *modelling & software*, **75**, 333–347.
- 3110 Fisher, R.A. (1922) On the Mathematical Foundations of Theoretical Statistics.
3111 *Philosophical Transactions of the Royal Society of London. Series A, Containing*
3112 *Papers of a Mathematical or Physical Character*, **222**, 309–368.
- 3113 Folke, C., Carpenter, S., Walker, B., Scheffer, M., Elmqvist, T., Gunderson, L.
3114 & Holling, C.S. (2004) Regime shifts, resilience, and biodiversity in ecosystem
3115 management. *Annu. Rev. Ecol. Evol. Syst.*, **35**, 557–581.
- 3116 Frieden, B.R. (1990) Fisher information, disorder, and the equilibrium distributions of
3117 physics. *Physical Review A*, **41**, 4265–4276.
- 3118 Frieden, B.R., Plastino, A., Plastino, A.R. & Soffer, B.H. (2002) Non-equilibrium
3119 thermodynamics and Fisher information: An illustrative example. *Physics Letters*
3120 *A*, **304**, 73–78.
- 3121 Frieden, R. & Gatenby, R.A. (2010) *Exploratory data analysis using fisher information*,
3122 Springer Science & Business Media.
- 3123 Garmestani, A.S., Allen, C.R. & Bessey, K.M. (2005) Time-series analysis of clusters

- 3124 in city size distributions. *Urban Studies*, **42**, 1507–1515.
- 3125 Graham, R. & Tél, T. (1984) Existence of a potential for dissipative dynamical systems.
- 3126 *Physical review letters*, **52**, 9.
- 3127 Groffman, P.M., Baron, J.S., Blett, T., Gold, A.J., Goodman, I., Gunderson, L.H.,
3128 Levinson, B.M., Palmer, M.A., Paerl, H.W., Peterson, G.D., Poff, N.L., Rejeski,
3129 D.W., Reynolds, J.F., Turner, M.G., Weathers, K.C. & Wiens, J. (2006) Ecological
3130 Thresholds: The Key to Successful Environmental Management or an Important
3131 Concept with No Practical Application? *Ecosystems*, **9**, 7:e01614.
- 3132 Guttal, V. & Jayaprakash, C. (2009) Spatial variance and spatial skewness: Leading
3133 indicators of regime shifts in spatial ecological systems. *Theoretical Ecology*, **2**,
3134 3–12.
- 3135 Guttal, V., Jayaprakash, C. & Tabbaa, O.P. (2013) Robustness of early warning
3136 signals of regime shifts in time-delayed ecological models. *Theoretical ecology*, **6**,
3137 271–283.
- 3138 Hastings, A. & Wysham, D.B. (2010) Regime shifts in ecological systems can occur
3139 with no warning. *Ecology letters*, **13**, 464–472.
- 3140 Havlicek, T.D. & Carpenter, S.R. (2001) Pelagic species size distributions in lakes:
3141 Are they discontinuous? *Limnology and Oceanography*, **46**, 1021–1033.
- 3142 Hawkins, S.J., Bohn, K. & Doncaster, C.P. (2015) Ecosystems: The rocky road to
3143 regime-shift indicators. *Current Biology*, **25**, R666–R669.
- 3144 Hefley, T.J., Tyre, A.J. & Blankenship, E.E. (2013) Statistical indicators and state–
3145 space population models predict extinction in a population of bobwhite quail.
3146 *Theoretical Ecology*, **6**, 319–331.
- 3147 Herkert, J.R. (1994) The effects of habitat fragmentation on midwestern grassland

- 3148 bird communities. *Ecological applications*, **4**, 461–471.
- 3149 Holling, C.S. (1992) Cross-scale morphology, geometry, and dynamics of ecosystems.
3150 *Ecological monographs*, **62**, 447–502.
- 3151 Holling, C.S. (1973) Resilience and stability of ecological systems. *Annual review of*
3152 *ecology and systematics*, **4**, 1–23.
- 3153 Hughes, T.P. (1994) Catastrophes, phase shifts, and large-scale degradation of a
3154 caribbean coral reef. *Science*, **265**, 1547–1551.
- 3155 Hughes, T.P., Carpenter, S., Rockström, J., Scheffer, M. & Walker, B. (2013) Multiscale
3156 regime shifts and planetary boundaries. *Trends in ecology & evolution*, **28**, 389–
3157 395.
- 3158 Initiative, N.A.B.C., Committee, U. & others (2009) The state of the birds, united
3159 states of america, 2009. *US Department of Interior: Washington, DC*, **5**.
- 3160 Jorgensen, S.E. & Svirezhev, Y.M. (2004) *Towards a Thermodynamic Theory for*
3161 *Ecological Systems*, Elsevier.
- 3162 Karunanithi, A.T., Cabezas, H., Frieden, B.R. & Pawlowski, C.W. (2008) Detection
3163 and assessment of ecosystem regime shifts from fisher information. *Ecology and*
3164 *society*, **13**.
- 3165 Kaufmann, H., Hutter, R., Skopik, F. & Mantere, M. (2015) A structural design for a
3166 pan-european early warning system for critical infrastructures. *e & i Elektrotechnik*
3167 *und Informationstechnik*, **132**, 117–121.
- 3168 Kefi, S., Guttal, V., Brock, W.A., Carpenter, S.R., Ellison, A.M., Livina, V.N., Seekell,
3169 D.A., Scheffer, M., Nes, E.H. van & Dakos, V. (2014) Early warning signals of
3170 ecological transitions: Methods for spatial patterns. *PloS one*, **9**, e92097.
- 3171 Kong, X., He, Q., Yang, B., He, W., Xu, F., Janssen, A.B., Kuiper, J.J., Van Gerven,

- 3172 L.P., Qin, N., Jiang, Y. & others (2017) Hydrological regulation drives regime shifts:
3173 Evidence from paleolimnology and ecosystem modeling of a large shallow chinese
3174 lake. *Global change biology*, **23**, 737–754.
- 3175 La Sorte, F.A., Lepczyk, C.A., Burnett, J.L., Hurlbert, A.H., Tingley, M.W. &
3176 Zuckerberg, B. (2018) Opportunities and challenges for big data ornithology. *The
3177 Condor: Ornithological Applications*, **120**, 414–426.
- 3178 Lenton, T.M. (2011) Early warning of climate tipping points. *Nature climate change*,
3179 **1**, 201.
- 3180 Lindegren, M., Dakos, V., Gröger, J.P., Gårdmark, A., Kornilovs, G., Otto, S.A. &
3181 Möllmann, C. (2012) Early detection of ecosystem regime shifts: A multiple method
3182 evaluation for management application. *PLoS One*, **7**, e38410.
- 3183 Litzow, M.A. & Hunsicker, M.E. (2016) Early warning signals, nonlinearity, and signs
3184 of hysteresis in real ecosystems. *Ecosphere*, **7**, –.
- 3185 Liu, J., Dietz, T., Carpenter, S.R., Alberti, M., Folke, C., Moran, E., Pell, A.N.,
3186 Deadman, P., Kratz, T., Lubchenco, J. & others (2007) Complexity of coupled
3187 human and natural systems. *science*, **317**, 1513–1516.
- 3188 Liu, S., Yamada, M., Collier, N. & Sugiyama, M. (2013) Change-point detection
3189 in time-series data by relative density-ratio estimation. *Neural Networks*, **43**,
3190 72–83.
- 3191 Mac Nally, R., Albano, C. & Fleishman, E. (2014) A scrutiny of the evidence for
3192 pressure-induced state shifts in estuarine and nearshore ecosystems. *Austral Ecology*,
3193 **39**, 898–906.
- 3194 Manly, B.F. (1996) Are there clumps in body-size distributions? *Ecology*, **77**, 81–
3195 86.

- 3196 Mantua, N. (2004) Methods for detecting regime shifts in large marine ecosystems: A
3197 review with approaches applied to North Pacific data. *Progress in Oceanography*,
3198 **60**, 165–182.
- 3199 May, R.M. (1977) Thresholds and breakpoints in ecosystems with a multiplicity of
3200 stable states. *Nature*, **269**, 471.
- 3201 Mayer, A.L., Pawlowski, C., Fath, B.D. & Cabezas, H. (2007) *Applications of fisher
3202 information to the management of sustainable environmental systems. Exploratory
3203 data analysis using fisher information*, pp. 217–244. Springer.
- 3204 Mayer, A.L., Pawlowski, C.W. & Cabezas, H. (2006) Fisher information and dynamic
3205 regime changes in ecological systems. *Ecological modelling*, **195**, 72–82.
- 3206 Mckenney, D.W., Pedlar, J.H., Lawrence, K., Papadopol, P., Campbell, K. & Hutchin-
3207 son, M.F. (2014) Change and evolution in the plant hardiness zones of canada.
3208 *BioScience*, **64**, 341–350.
- 3209 Michener, W.K. & Jones, M.B. (2012) Ecoinformatics: Supporting ecology as a
3210 data-intensive science. *Trends in ecology & evolution*, **27**, 85–93.
- 3211 Möllmann, C., Folke, C., Edwards, M. & Conversi, A. (2015) Marine regime shifts
3212 around the globe: Theory, drivers and impacts.
- 3213 Mumby, P.J., Steneck, R.S. & Hastings, A. (2013) Evidence for and against the
3214 existence of alternate attractors on coral reefs. *Oikos*, **122**, 481–491.
- 3215 Murray, L.D., Ribic, C.A. & Thogmartin, W.E. (2008) Relationship of obligate
3216 grassland birds to landscape structure in wisconsin. *The Journal of Wildlife
3217 Management*, **72**, 463–467.
- 3218 Nash, K.L., Allen, C.R., Angeler, D.G., Barichievy, C., Eason, T., Garmestani,
3219 A.S., Graham, N.A., Granholm, D., Knutson, M., Nelson, R.J. & others (2014a)

- 3220 Discontinuities, cross-scale patterns, and the organization of ecosystems. *Ecology*,
3221 **95**, 654–667.
- 3222 Nash, K.L., Allen, C.R., Barichievy, C., Nyström, M., Sundstrom, S. & Graham, N.A.
3223 (2014b) Habitat structure and body size distributions: Cross-ecosystem comparison
3224 for taxa with determinate and indeterminate growth. *Oikos*, **123**, 971–983.
- 3225 Nes, E.H. van & Scheffer, M. (2005) Implications of spatial heterogeneity for catas-
3226 trophic regime shifts in ecosystems. *Ecology*, **86**, 1797–1807.
- 3227 Nicholls, K.H. (2011) Detection of regime shifts in multi-species communities: The
3228 Bay of Quinte phytoplankton example. *Methods in Ecology and Evolution*, **2**,
3229 416–426.
- 3230 Nicholls, K., Hoyle, J., Johannsson, O. & Dermott, R. (2011) A biological regime shift
3231 in the bay of quinte ecosystem (lake ontario) associated with the establishment of
3232 invasive dreissenid mussels. *Journal of Great Lakes Research*, **37**, 310–317.
- 3233 Parmesan, C. (2006) Ecological and evolutionary responses to recent climate change.
3234 *Annu. Rev. Ecol. Evol. Syst.*, **37**, 637–669.
- 3235 Perretti, C.T. & Munch, S.B. (2012) Regime shift indicators fail under noise levels com-
3236 monly observed in ecological systems. *Ecological Applications*, **22**, 1772–1779.
- 3237 Perretti, C.T., Munch, S.B. & Sugihara, G. (2013) Model-free forecasting outperforms
3238 the correct mechanistic model for simulated and experimental data. *Proceedings of
3239 the National Academy of Sciences*, **110**, 5253–5257.
- 3240 Peterjohn, B.G. & Sauer, J.R. (1999) Population status of north american grassland
3241 birds from the north american breeding bird survey. *Studies in Avian Biology*, **19**,
3242 27–44.
- 3243 Peters, R.H. & Wassenberg, K. (1983) The effect of body size on animal abundance.

- 3244 Petersen, J.K., Hansen, J.W., Laursen, M.B., Clausen, P., Carstensen, J. & Conley,
3245 D.J. (2008) Regime shift in a coastal marine ecosystem. *Ecological Applications*,
3246 **18**, 497–510.
- 3247 Peterson, G., Allen, C.R. & Holling, C.S. (1998) Ecological resilience, biodiversity,
3248 and scale. *Ecosystems*, **1**, 6–18.
- 3249 Price, Nathaniel B. & Burnett, J.L. (2019) Tvdif: An r package for numerical
3250 differentiation of noisy, nonsmooth data.
- 3251 Ratajczak, Z., Carpenter, S.R., Ives, A.R., Kucharik, C.J., Ramiadantsoa, T., Stegner,
3252 M.A., Williams, J.W., Zhang, J. & Turner, M.G. (2018) Abrupt change in ecological
3253 systems: Inference and diagnosis. *Trends in ecology & evolution*.
- 3254 Reid, P.C., Hari, R.E., Beaugrand, G., Livingstone, D.M., Marty, C., Straile, D.,
3255 Barichivich, J., Goberville, E., Adrian, R., Aono, Y., Brown, R., Foster, J., Grois-
3256 man, P., Hélaouët, P., Hsu, H.-H., Kirby, R., Knight, J., Kraberg, A., Li, J., Lo,
3257 T.-T., Myneni, R.B., North, R.P., Pounds, J.A., Sparks, T., Stübi, R., Tian, Y.,
3258 Wiltshire, K.H., Xiao, D. & Zhu, Z. (2016) Global impacts of the 1980s regime
3259 shift. *Global Change Biology*, **22**, 682–703.
- 3260 Restrepo, C. & Arango, N. (2008) Discontinuities in the geographical range size of north
3261 american birds and butterflies. *Discontinuities in ecosystems and other complex*
3262 *systems*. Columbia University Press, New York, New York, USA, 101–135.
- 3263 Roberts, C.P., Twidwell, D., Burnett, J.L., Donovan, V.M., Wonkka, C.L., Bielski,
3264 C.L., Garmestani, A.S., Angeler, D.G., Eason, T., Allred, B.W. & others (2018)
3265 Early warnings for state transitions. *Rangeland Ecology & Management*, **71**, 659–
3266 670.
- 3267

- 3268 Roberts, C.R., Allen, C.R., Angeler, D.G. & Twidwell, D. (2019) Shifting avian spatial
3269 regimes in a changing climate. *xxxx*, **xx**, xx–xx.
- 3270 Rockström, J., Steffen, W.L., Noone, K., Persson, Å., Chapin III, F.S., Lambin,
3271 E., Lenton, T.M., Scheffer, M., Folke, C., Schellnhuber, H.J. & others (2009)
3272 Planetary boundaries: Exploring the safe operating space for humanity. *Ecology*
3273 and society.
- 3274 Rodionov, S.N. (2005) A brief overview of the regime shift detection methods. *Large-*
3275 *Scale Disturbances (Regime Shifts) and Recovery in Aquatic Ecosystems: Challenges*
3276 *for Management Toward Sustainability*, 17–24.
- 3277 Rodionov, S. & Overland, J.E. (2005) Application of a sequential regime shift detection
3278 method to the Bering Sea ecosystem. *ICES Journal of Marine Science*, **62**, 328–
3279 332.
- 3280 Roy Frieden, B. (1998) *Physics from Fisher information*, Cambridge University Press,
3281 Cambridge.
- 3282 Sagarin, R. & Pauchard, A. (2012) *Observation and ecology: Broadening the scope of*
3283 *science to understand a complex world*, Island Press.
- 3284 Salehpour, S., Gustafsson, T. & Johansson, A. (2011) An on-line method for estimation
3285 of piecewise constant parameters in linear regression models. *IFAC Proceedings*
3286 *Volumes*, **44**, 3171–3176.
- 3287 Sauer, J.R., Niven, D.K., Hines, J.E., Ziolkowski, D J, Pardieck, K.L., Fallon, J.E. &
3288 Link, W.A. (2014) The north american breeding bird survey, results and analysis
3289 1966 - 2015. Version 2.07.2017.
- 3290 Scheffer, M. (2009) *Critical transitions in nature and society*, Princeton University
3291 Press.

- 3292 Scheffer, M., Carpenter, S., Foley, J.A., Folke, C. & Walker, B. (2001) Catastrophic
3293 shifts in ecosystems. *Nature*, **413**, 591.
- 3294 Scheffer, M. & Carpenter, S.R. (2003) Catastrophic regime shifts in ecosystems:
3295 Linking theory to observation. *Trends in ecology & evolution*, **18**, 648–656.
- 3296 Scheffer, M., Carpenter, S.R., Dakos, V. & Nes, E.H. van (2015) Generic indicators of
3297 ecological resilience: Inferring the chance of a critical transition. *Annual Review of
3298 Ecology, Evolution, and Systematics*, **46**, 145–167.
- 3299 Seddon, A.W., Froyd, C.A., Witkowski, A. & Willis, K.J. (2014) A quantitative
3300 framework for analysis of regime shifts in a galápagos coastal lagoon. *Ecology*, **95**,
3301 3046–3055.
- 3302 Shriver, W.G., Jones, A.L., Vickery, P.D., Weik, A. & Wells, J. (2005) *The distribution
3303 and abundance of obligate grassland birds breeding in new england and new york.
3304 Bird conservation implementation and integration in the americas: Proceedings of
3305 the third international partners in flight conference. Eds: Ralph, c. John; rich,
3306 terrell d., US Dept. of Agriculture, Forest Service, Pacific Southwest Research
3307 Station.*
- 3308 Siemann, E. & Brown, J.H. (1999) Gaps in mammalian body size distributions
3309 reexamined. *Ecology*, **80**, 2788–2792.
- 3310 Silva, M. & Downing, J.A. (1995) The allometric scaling of density and body mass:
3311 A nonlinear relationship for terrestrial mammals. *The American Naturalist*, **145**,
3312 704–727.
- 3313 Skillen, J.J. & Maurer, B.A. (2008) The ecological significance of discontinuities in
3314 body-mass distributions. *Discontinuities in ecosystems and other complex systems*,
3315 193–218.

- 3316 Smith, V.H. & Schindler, D.W. (2009) Eutrophication science: Where do we go from
3317 here? *Trends in ecology & evolution*, **24**, 201–207.
- 3318 Sommer, S., Benthem, K.J. van, Fontaneto, D. & Ozgul, A. (2017) Are generic early-
3319 warning signals reliable indicators of population collapse in rotifers? *Hydrobiologia*,
3320 **796**, 111–120.
- 3321 Sorte, F.A.L. & III, F.R.T. (2007) Poleward shifts in winter ranges of north american
3322 birds. *Ecology*, **88**, 1803–1812.
- 3323 Spanbauer, T.L., Allen, C.R., Angeler, D.G., Eason, T., Fritz, S.C., Garmestani, A.S.,
3324 Nash, K.L. & Stone, J.R. (2014) Prolonged instability prior to a regime shift. *PLoS
3325 One*, **9**, e108936.
- 3326 Spanbauer, T.L., Allen, C.R., Angeler, D.G., Eason, T., Fritz, S.C., Garmestani,
3327 A.S., Nash, K.L., Stone, J.R., Stow, C.A. & Sundstrom, S.M. (2016) Body size
3328 distributions signal a regime shift in a lake ecosystem. *Proceedings of the Royal
3329 Society B: Biological Sciences*, **283**, 20160249.
- 3330 Steel, E.A., Kennedy, M.C., Cunningham, P.G. & Stanovick, J.S. (2013) Applied
3331 statistics in ecology: Common pitfalls and simple solutions. *Ecosphere*, **4**, 1–
3332 13.
- 3333 Stow, C., Allen, C.R. & Garmestani, A.S. (2007) Evaluating discontinuities in complex
3334 systems: Toward quantitative measures of resilience.
- 3335 Sugihara, G., May, R., Ye, H., Hsieh, C.-h., Deyle, E., Fogarty, M. & Munch, S. (2012)
3336 Detecting causality in complex ecosystems. *science*, **338**, 496–500.
- 3337 Sundstrom, S.M., Eason, T., Nelson, R.J., Angeler, D.G., Barichievy, C., Garmestani,
3338 A.S., Graham, N.A., Granholm, D., Gunderson, L., Knutson, M. & others (2017)
3339 Detecting spatial regimes in ecosystems. *Ecology letters*, **20**, 19–32.

- 3340 Takens, F. (1981) *Detecting strange attractors in turbulence. Dynamical systems and*
3341 *turbulence, warwick 1980*, pp. 366–381. Springer.
- 3342 Taranu, Z.E., Carpenter, S.R., Frossard, V., Jenny, J.-P., Thomas, Z., Vermaire, J.C.
3343 & Perga, M.-E. (2018) Can we detect ecosystem critical transitions and signals of
3344 changing resilience from paleo-ecological records? *Ecosphere*, **9**.
- 3345 Taylor, L. (1961) Aggregation, variance and the mean. *Nature*, **189**, 732.
- 3346 Thrush, S.F., Hewitt, J.E., Dayton, P.K., Coco, G., Lohrer, A.M., Norkko, A., Norkko,
3347 J. & Chiantore, M. (2009) Forecasting the limits of resilience: Integrating empirical
3348 research with theory. *Proceedings of the Royal Society B: Biological Sciences*, **276**,
3349 3209–3217.
- 3350 Van Auken, O. (2009) Causes and consequences of woody plant encroachment into
3351 western north american grasslands. *Journal of environmental management*, **90**,
3352 2931–2942.
- 3353 Vasilakopoulos, P., Raitsos, D.E., Tzanatos, E. & Maravelias, C.D. (2017) Resilience
3354 and regime shifts in a marine biodiversity hotspot. *Scientific reports*, **7**, 13647.
- 3355 Walker, B., Holling, C.S., Carpenter, S. & Kinzig, A. (2004) Resilience, adaptability
3356 and transformability in social–ecological systems. *Ecology and society*, **9**.
- 3357 Walther, G.-R., Post, E., Convey, P., Menzel, A., Parmesan, C., Beebee, T.J., Fro-
3358 mentin, J.-M., Hoegh-Guldberg, O. & Bairlein, F. (2002) Ecological responses to
3359 recent climate change. *Nature*, **416**, 389.
- 3360 Wardwell, D. & Allen, C.R. (2009) Variability in population abundance is associated
3361 with thresholds between scaling regimes.
- 3362 Weijerman, M., Lindeboom, H. & Zuur, A.F. (2005) Regime shifts in marine ecosystems
3363 of the north sea and wadden sea. *Marine Ecology Progress Series*, **298**, 21–39.

- 3364 Weissmann, H. & Shnerb, N.M. (2016) Predicting catastrophic shifts. *Journal of*
3365 *theoretical biology*, **397**, 128–134.
- 3366 Wolkovich, E.M., Cook, B.I., McLauchlan, K.K. & Davies, T.J. (2014) Temporal
3367 ecology in the Anthropocene. *Ecology Letters*, **17**, 1365–1379.
- 3368 Yang, Q. & Wu, X. (2006) 10 challenging problems in data mining research. *Interna-*
3369 *tional Journal of Information Technology & Decision Making*, **05**, 597–604.
- 3370 Ye, H., Beamish, R.J., Glaser, S.M., Grant, S.C., Hsieh, C.-h., Richards, L.J., Schnute,
3371 J.T. & Sugihara, G. (2015) Equation-free mechanistic ecosystem forecasting using
3372 empirical dynamic modeling. *Proceedings of the National Academy of Sciences*,
3373 **112**, E1569–E1576.
- 3374 Yin, D., Leroux, S.J. & He, F. (2017) Methods and models for identifying thresholds
3375 of habitat loss. *Ecography*, **40**, 131–143.
- 3376 Zhou, T. & Shumway, R. (2008) One-step approximations for detecting regime changes
3377 in the state space model with application to the influenza data. *Computational*
3378 *Statistics & Data Analysis*, **52**, 2277–2291.