

Change in aquatic ecosystems: advancing resilience concepts towards practical applications

Cal Dunham Buelo
Middleton, Wisconsin

B.S. Biology and Mathematics, University of Wisconsin-Madison, 2014

A Dissertation presented to the Graduate Faculty
of the University of Virginia in Candidacy for the Degree of
Doctor of Philosophy

Department of Environmental Sciences

University of Virginia
December 2021

Abstract

Aquatic ecosystems are subject to many stressors including eutrophication, hydrologic alteration, invasive species, and climate change. These stressors alter ecosystems and their resilience – how an ecosystem responds to disturbances like storms, droughts, and other discrete events. Disturbance impacts can vary widely in magnitude, duration, and the changes they induce on ecosystem structure, function, and services. Understanding and predicting change has motivated development of theories and frameworks for several resilience concepts. However, applying resilience methods in practice is often challenged by data limitations and the inherent complexity of ecosystems.

The goal of this dissertation was to advance the application of resilience concepts to real world ecosystems using data intensive methods. I focused specifically on two ecosystem changes and resilience concepts: the prediction of algal blooms in lakes and understanding patterns and controls of disturbance in estuaries following tropical cyclones. I first evaluated if spatial early warning statistics (EWS), based on theory that generic changes in system dynamics are reflected in statistical properties, are expected to change prior to algal blooms. Using a spatial model incorporating physical forces that control transport in aquatic systems, I found that spatial standard deviation and autocorrelation distinguished between bloom states and changed predictably near thresholds. I then tested those findings and compared spatial EWS to previously studied EWS in time series data using a whole-lake nutrient addition experiment. Spatial EWS did not change consistently before the bloom, while temporal standard deviation did for 3 out of 4 variables. I then utilized high frequency time series from 18 lake-years of both experimental and non-experimental conditions to quantify temporal EWS performance at separating low from high resilience states, a necessary step for the method's potential use for bloom management.

Using high frequency data from the same lake fertilization experiments, I also explored the ability of near-term forecasting to accurately predict bloom initiation timing, a short but critical period for taking management action. Accurately forecasting bloom timing was difficult but possible and depended on both model initial conditions and flexibly adjusting parameters as new observations were collected. Finally, I used a new algorithm for detecting disturbance and recovery in high frequency data to quantify disturbance occurrence, timing, length, and severity in salinity and dissolved oxygen across 19 estuaries and 59 tropical cyclones in the eastern United States. Most estuaries recovered from hurricane-initiated disturbances within days, but some lasted weeks or months, and properties of both storms and the sites they impacted were related to disturbance characteristics.

This dissertation shows that resilience concepts can be operationalized to measurable properties, which can be used to understand and predict change with possibilities for application to ecosystem management. Operationalizing resilience is crucial to maintaining ecosystem services such as clean water, fisheries, and carbon sequestration into a future where stresses on aquatic systems are projected to intensify. My findings also demonstrate the power of ecosystem scale experiments as well as high-frequency and long-term data to test and advance understanding and management.

Acknowledgements

So many people have contributed to this dissertation and made it possible. First, I am extremely grateful to my advisor, Michael Pace, for his generous guidance and support at all stages of my graduate career and beforehand as a research technician in his lab. I'd also like to thank Cascade project PIs Steve Carpenter and Emily Stanley from the University of Wisconsin for sharing their expertise and guidance. Thank you to my committee members Patricia Wiberg and Xi Yang for their feedback and advice, and to Garrick Louis for serving as my Dean's Representative.

I'm very thankful for my colleagues in the Pace lab and Cascade group who I've worked with during this dissertation: Grace Wilkinson, Alice Besterman, Dat Ha, Spencer Tassone, Jonathan Walter, and David Ortiz. Thank you for your friendship and I am excited to continue collaborating with and learning from all of you. To the members of all the communities I've been lucky enough to be a part of, including UVA's Department of Environmental Sciences (and especially my cohort), Notre Dame's Environmental Research Center, and UW's Trout Lake Station, your friendship and camaraderie are greatly cherished. I have also enjoyed working with many amazing undergraduates through the years, including Anne Marie Saunders, Carson Lambert, Dat Ha, Kayla Wernsing, and Sara McCormack who helped with this dissertation - thank you.

Finally, I owe my perpetual gratitude to my family and loved ones. My parents, Bret and Patti, have been endlessly supportive in all aspects of my life, as have my siblings Audrey, Collin, and Aaron. To my partner Shayna – your constant love and support means so much to me, as has your unwavering encouragement through field seasons, successes, and setbacks. Thank you.

Table of Contents

Abstract	2
Acknowledgements	4
List of Figures	6
List of Tables	7
Chapter 1 Introduction.....	8
Chapter 2 A modeling analysis of spatial statistical indicators of thresholds for algal blooms.....	18
Chapter 3 Evaluating the performance of temporal and spatial early warning statistics for algal blooms.....	40
Chapter 4 Anticipating blooms: exploring the accuracy of algal bloom initiation forecasts.....	71
Chapter 5 Quantifying disturbance and recovery in estuaries: tropical cyclones and high frequency measures of oxygen and salinity.....	96
Chapter 6 Conclusion.....	125
Appendix 1 Supplemental Information for Chapter 2.....	132
Appendix 2 Supplemental Information for Chapter 3.....	137
Appendix 3 Supplemental Information for Chapter 4.....	145
Appendix 4 Supplemental Information for Chapter 5.....	153

List of Figures

Figure 2.1	Spatial bloom model output grids for three dynamic states.....	<u>36</u>
Figure 2.2	Temporal evolution of spatial indicators in each state.....	<u>37</u>
Figure 2.3	Spatial indicators for three physical forcing cases.....	<u>38</u>
Figure 2.4	Spatial indicator changes near critical transitions	<u>39</u>
Figure 3.1	Time series of bloom state variables	<u>66</u>
Figure 3.2	Spatial maps of bloom state variables.....	<u>67</u>
Figure 3.3	Temporal early warning statistics and alarms.....	<u>68</u>
Figure 3.4	Spatial early warning statistics.....	<u>69</u>
Figure 3.5	True and false positive alarm rates of temporal EWS.....	<u>70</u>
Figure 4.1	Chlorophyll-a concentrations up to bloom peak vs. P loading.....	<u>91</u>
Figure 4.2	Ensemble state and parameter fits for Tuesday Lake 2015.....	<u>92</u>
Figure 4.3	Seven-day forecasts and overall forecast error.....	<u>93</u>
Figure 4.4	Lh_t fits and forecasts for different initial parameter values.....	<u>94</u>
Figure 4.5	Bloom timing error of 7-day forecasts for different Lh_{t0} and IF values.....	<u>95</u>
Figure 5.1	Disturbance detection algorithm example.....	<u>119</u>
Figure 5.2	Time series and normalized difference statistics for several disturbances...	<u>120</u>
Figure 5.3	Disturbances detected during storm and pre-storm periods.....	<u>121</u>
Figure 5.4	Disturbance timing relative to storm passage.....	<u>122</u>
Figure 5.5	Disturbance length.....	<u>123</u>
Figure 5.6	Relationship between disturbance length and severity.....	<u>124</u>
Figure A2.1	Daily and cumulative phosphorus added to Peter Lake.....	<u>141</u>
Figure A2.2	Weekly spatial maps of phycocyanin in Peter and Paul Lakes.....	<u>142</u>
Figure A2.3	Weekly spatial maps of dissolved oxygen.....	<u>143</u>
Figure A2.4	Weekly spatial maps of pH.....	<u>144</u>
Figure A3.1	Bloom timing error of 3-day forecasts for different Lh_{t0} and IF values.....	<u>149</u>
Figure A3.2	Bloom timing error of 5-day forecasts for different Lh_{t0} and IF values.....	<u>150</u>
Figure A3.3	Relationship between chl-a and potential bloom driver variables.....	<u>151</u>
Figure A3.4	Time series of chl-a for all fertilization experiments.....	<u>152</u>

List of Tables

<i>Table 3.1</i>	Summary of temporal and spatial early warning statistics.....	65
<i>Table 4.1</i>	Forecasting model parameters.....	90
<i>Table 5.1</i>	Storm and station variables related to disturbance characteristics.....	118
<i>Table A1.1</i>	Nondimensional spatial bloom model parameters.....	135
<i>Table A1.2</i>	Dimensional and nondimensional parameter symbols and units.....	136
<i>Table A4.1</i>	Number of stations and tropical cyclones analyzed for each NERRS site...	153

Chapter 1

Introduction

Aquatic ecosystems are subject to numerous stressors and disturbances from local to global scales, with important consequences for these ecosystems and the services they provide. For example: land use change and agricultural intensification cause eutrophication and alter hydrological processes (Bennett et al. 2001, Gordon et al. 2008); humans facilitate the spread of non-native species (Lovell et al. 2006); climate change alters atmospheric temperature and precipitation, which in turn impacts water temperature, ice and stratification regimes, and water clarity (Adrian et al. 2009). These and other stressors affect the multitude of ecosystem services including water for drinking and irrigation, recreational and commercial fisheries, and carbon sequestration (Postal and Carpenter 1997, Mendonça et al. 2017).

An ecosystem's response to stress and disturbance depends in part on its *resilience* – a term that has many definitions, here I use it to broadly describe how a system responds to perturbations or shocks: the degree to which it changes from a given magnitude shock (also referred to as resistance; Pimm 1984), the size of shock it can absorb and remain in the same state (Carpenter et al. 2001) and how quickly it recovers (also referred to as return time or engineering resilience; Gunderson 2000). More comprehensive frameworks have included resilience as one component determining the dynamics of combined social-ecological systems (SES), along with adaptability (the capacity of actors to influence SES resilience) and transformability (the capacity to create a new system; Walker et al. 2004). Resilience has received considerable attention across the field of ecology, frequently with the goal of understanding what determines if an ecosystem is resilient or if resilience is changing. Doing so

offers insights into when and where changes are likely to occur, and potentially allows for actions to be taken to increase resilience and avoid unwanted changes (Folke et al. 2004).

Multiple approaches have been taken to understand resilience and predict change in ecosystems. Over approximately the past 15 years, the concept of early warning statistics (EWS; also referred to as early warning indicators or resilience indicators) of regime shifts has been developed, which predicts that ecosystem statistical properties change in specific ways before rapid transitions (Scheffer et al. 2015). Mechanistically, EWS often arise from critical slowing down – when a system’s return rate to equilibrium decreases as a system approaches a critical transition (Scheffer et al. 2009), resulting in increasing variance and autocorrelation (Carpenter and Brock 2006, Van Nes and Scheffer 2007) as well as changes in other statistics (Guttal and Jayaprakash 2008).

EWS have been studied in both spatial and temporal data across different ecosystems (Scheffer et al. 2015). The EWS concept has been extensively developed using data from model simulations, with a more limited number of empirical tests in laboratory experiments (e.g. Drake and Griffen 2010, Dai et al. 2012, Dai et al. 2013) and observational and experimental field studies (e.g. Litzow et al. 2008, Litzow et al. 2013, Eby et al. 2017, Ratajczak et al. 2017, Rindi et al. 2017, Rindi et al. 2018). At the ecosystem scale, some of the most powerful tests of EWS have been carried out using whole-lake manipulations to show that expected changes in EWS occur prior to known regime shifts (Carpenter et al. 2011, Seekell et al. 2012, Batt et al. 2013, Pace et al. 2013, Cline et al. 2014, Butitta et al. 2017, Pace et al. 2017, Wilkinson et al. 2018). However, even in these studies, questions important to potential application of EWS remain unanswered, including: 1) whether spatial or temporal EWS perform better and 2) if EWS can accurately differentiate between low and high resilience states.

Interest in predicting change and providing useful information for management has also stimulated research on direct, near-term forecasting of ecosystem properties including phenology (Taylor and White 2020), pest and disease outbreak (Jones et al. 2021), and biogeochemical cycling (Gao et al. 2011). In aquatic systems, forecasting of water temperatures (Thomas et al. 2020), dissolved oxygen (Carey et al. 2021), and algal blooms (Recknagel et al. 2017, Xiao et al. 2017, Page et al. 2018) is motivated by impacts on aquatic organisms and water quality. Algal bloom forecasting in particular has received considerable attention due to potential human health impacts of algal toxins in harvested shellfish (Trainer et al. 2007) and drinking water (Carmichael 2001). However, adoption of bloom forecasting by managers has been limited, likely due to the need for system-specific forecast model development (Rousso et al. 2020) arising from the wide diversity in aquatic ecosystem properties and mechanisms controlling bloom development (Paerl et al. 2001, Isles and Pomati 2021). Management-relevant model performance is also a concern and may not be captured by widely used evaluation metrics.

More generally, resilience to shocks or perturbations is a central component of disturbance ecology, which focuses on discrete events that disrupt “ecosystem, community, or population structure and changes resources, substrate, availability or the physical environment” (White and Pickett 1985). Quantifying resilience is easy when events initiating disturbances are discrete and recovery is obvious (e.g., regrowth of vegetation after fire; Goetz et al. 2006), but in many cases shocks can cause qualitatively different types of disturbances and recovery may be difficult to identify due to natural variability. The response of estuaries to tropical cyclones is one example: storm characteristics vary along several axes (e.g., precipitation, storm surge, winds and waves), causing different types of disturbances (increases or decreases in salinity and oxygen, changes in vegetation or other populations), and recovery determination is complicated

by the highly dynamic nature of these systems (daily, tidal, and seasonal cycles). These challenges have largely limited studies of tropical cyclone-caused disturbances to either one or a few locations or one or a few storms (Pruitt et al. 2019).

The goal of this dissertation was to advance understanding of resilience in aquatic ecosystems by bridging gaps between theory and application. Many of the concepts and frameworks of resilience have been developed from conceptual or relatively simple mathematical models, which are critical for formalizing assumptions and generating expectations and hypotheses that can be tested. However, doing so empirically is often not straightforward: it can be difficult or impossible to measure key variables precisely, forcing a reliance on more easily measured proxies; time and funding constraints may limit the frequency and extent at which data can be collected; and the inherent complexity of ecosystems may generate stochasticity that obscures expected relationships or patterns. Nonetheless, technological advances have made it possible to collect measurements on an increasing number of ecosystem variables at high frequencies and spatial extents, presenting the opportunity to advance our understanding of resilience and use it to improve ecosystem management. This dissertation addresses application of resilience concepts to real-world ecosystems using data-intensive methods, focused specifically on predicting algal blooms and understanding disturbances in estuaries caused by tropical cyclones.

Chapter 2 evaluates whether spatial EWS should be expected before algal bloom regime shifts. Temporal EWS have been well studied in aquatic ecosystems using both models and whole-lake experiments, including several experimentally induced algal blooms. In contrast, spatial EWS have been most well developed for terrestrial ecosystems, and studies in aquatic ecosystems have mostly been on species that have more control over their movement in the

environment than algae. Using a 2D algal bloom model with different dynamic states as well as realistic physical forcing (advection and diffusion), I test several statistics for spatial EWS using model simulations at a range of nutrient loading rates. Spatial standard deviation and spatial autocorrelation had distinct values in different bloom states and changed predictably near bloom thresholds, suggesting these statistics are good candidates for spatial EWS in field tests.

Chapter 3 tests the findings of Chapter 2 using a whole-lake experiment and directly compares temporal EWS to spatial EWS prior to algal blooms for the first time. Nutrients were added to Peter Lake while collecting high-frequency temporal and spatial measurements of key bloom variables; adjacent Paul Lake was a non-manipulated reference system. Analogous temporal and spatial EWS (standard deviation and autocorrelation) were computed and compared. Both spatial EWS performed poorly, with high sample-to-sample variability obscuring trends that would provide early warning of the bloom. Temporal standard deviation provided early warning for 3 out of 4 variables, while temporal autocorrelation did not provide warning in any variable. Analyzing temporal EWS across all past lake-years with high-frequency data (both with and without fertilizations) demonstrated that temporal EWS can provide performance potentially useful in management scenarios.

Chapter 4 approaches algal blooms from an ecological forecasting approach, using high frequency data from the above fertilization experiments and data assimilation to explore the dependence of bloom initiation forecast accuracy on model initial conditions and flexibility. Bloom initiation is a short time period but is critical for potential management actions. Model flexibility to learn parameters from observations can correct for inaccurate initial conditions but can also fit incorrect parameters from short-term trends, leading to inaccurate forecasts. Conclusions based on bloom timing forecast error contradicted those based on overall forecast

error, highlighting that commonly used model performance metrics underweight this crucial time period.

Chapter 5 uses a new algorithm for detecting disturbance and recovery with high-frequency and long-term dissolved oxygen and salinity data from estuaries across the eastern US and dozens of tropical cyclones. The algorithm detects diverse disturbances and characterizes their timing, length, and severity. Salinity disturbances generally started earlier and lasted longer than dissolved oxygen disturbances, suggesting different physical and biological mechanisms controlling disturbance in these variables. Most disturbances were less than 7 days long though some extremely long disturbances over 50 days long were observed. Disturbance severity was positively related to disturbance length (measures of ecosystem resistance and recovery time, respectively) for both variables. Site and storm properties associated with changes in disturbance characteristics demonstrate the ability of this approach to quantify disturbance and identify potential drivers across different variables, systems, and initiating events.

The final chapter summarizes the main findings of this dissertation and provides directions for future research. I conclude that resilience concepts can be operationalized to measurable properties. This quantification of features of resilience can be used to understand, predict, and in the future aid ecosystem management. Specifically, 1) early warning statistics and forecasting can be used to predict algal blooms in advance, and 2) robust and flexible methods for detecting disturbance and recovery can offer insights into patterns and controls of estuarine response to tropical cyclones, and likely other disturbances and ecosystems. Harnessing these approaches requires high frequency data coupled to real-time analysis and translation to useable outputs for managers. This dissertation also illustrates the importance of ecosystem-scale

experiments as well as the opportunity that long-term data provide to advance both understanding and management.

References

- Adrian, R, CM O'Reilly, H Zagarese, SB Baines, DO Hessen, W Keller, DM Livingstone, R Sommaruga, D Straile, E Van Donk, GA Weyhenmeyer, M Winder. 2009. Lakes as sentinels of climate change. *Limnology and Oceanography*. 54(6, pt 2): 2283-2297.
- Batt, RD, SR Carpenter, JJ Cole, ML Pace, RA Johnson. 2013. Changes in ecosystem resilience detected in automated measures of ecosystem metabolism during a whole-lake manipulation. *Proc. Nat. Acad. Sci. U.S.A.* 110(43): 17398-17403.
- Bennett, E.M., S.R. Carpenter, N.F. Caraco. 2001. Human impact of erodable phosphorus and eutrophication: a global perspective: increasing accumulation of phosphorus in soils threatens rivers, lakes, and coastal oceans with eutrophication. *BioScience*. 51(3): 227-234.
- Butitta, VL, SR Carpenter, LC Loken, ML Pace, and EH Stanley. 2017. Spatial early warning signals in a lake manipulation. *Ecosphere*, 8(10): e01941. <https://doi.org/10.1002/ecs2.1941>.
- Carey, CC, WM Woelmer, ME Lofton, RJ Figueiredo, BJ Bookout, RS Corrigan, W Daneshmand, AG Hounshell, DW Howard, ASL Lewis, RP McClure, HL Wander, NK Ward, RQ Thomas. 2021. Advancing lake and reservoir water quality management with near-term, iterative ecological forecasting. *Inland Waters*. doi: 10.1080/20442041.2020.1816421.
- Carmichael, WW. 2001. Health Effects of Toxin-Producing Cyanobacteria: "The CyanoHABs", Human and Ecological Risk Assessment: An International Journal, 7(5): 1393-1407.
- Carpenter, S, B Walker, J Marty Anderies, N Abel. 2001. From metaphor to measurement: resilience of what to what? *Ecosystems*. 4: 765-781.
- Carpenter, SR, and WA Brock. 2006. Rising variance: a leading indicator of ecological transition. *Ecol. Lett.* 9: 311–318.
- Carpenter, SR, JJ Cole, ML Pace, R Batt, WA Brock, T Cline, J Coloso, JR Hodgson, JF Kitchell, DA Seekell, L Smith, B Weidel. 2011. Early warnings of regime shifts: a whole-ecosystem experiment. *Science*, 332(6033): 1079–1082.
- Cline, TJ, DA Seekell, SR Carpenter, ML Pace, JR Hodgson, JF Kitchell, BC Weidel. 2014. Early warnings of regime shifts: evaluation of spatial indicators from a whole-ecosystem experiment. *Ecosphere*, 5(8): 102.

- Drake JM and BD Griffen. 2010. Early warning signals of extinction in deteriorating environments. *Nature*. 467(7314): 456-459.
- Dai L, D Vorselen, KS Korolev, J Gore. 2012. Generic indicators for loss of resilience before a tipping point leading to population collapse. *Science*. 336:1175–77
- Dai, L, KS Korolev, J Gore. 2013. Slower recovery in space before collapse of connected populations. *Nature*, 496: 355-358.
- Eby, S, A Agrawal, S Majumder, AP Dobson, and V Guttal. 2017. Alternative stable states and spatial indicators of critical slowing down along a spatial gradient in a savanna ecosystem. *Glob. Ecol. Biogeogr.* 26: 638–649.
- Folke, C, S Carpenter, B Walker, M Scheffer, T Elmqvist, L Gunderson, CS Holling. 2004. Regime shifts, resilience, and biodiversity in ecosystem management. *Annu. Rev. Ecol. Evol. Sys.* 35: 557-581.
- Gao, C, H Wang, E Weng, S Lakshmiarahan, Y Zheng, Y Luo. 2011. Assimilation of multiple data sets with the ensemble Kalman filter to improve forecasts of forest carbon dynamics. *Ecological Applications*. 21(5): 1461-1473.
- Goetz, SJ, GJ Fiske, AG Bunn. 2006. Using satellite time-series data to analyze fire disturbance and forest recovery across Canada. *Remote Sensing of Environment*. 101: 352-365.
- Gordon, LJ, GD Peterson, EM Bennett. 2008. Agricultural modifications of hydrological flows create ecological surprises. *Trends in Ecology and Evolution*. 23(4): 211-219.
- Gunderson, LH. 2000. Ecological resilience – in theory and application. *Annu. Rev. Ecol. Syst.* 31: 425-439.
- Guttal, V., and C. Jayaprakash. 2008. Changing skewness: an early warning signal of regime shifts in ecosystems. *Ecol. Lett.* 11: 450–460.
- Isles, PDF, F Pomati. 2021. An operational framework for defining and forecasting phytoplankton blooms. *Front. Ecol. Environ.* doi:10.1002/fee.2376.
- Jones, CM, S Jones, A Petrasova, V Petras, D Gaydos, MM Skrip, Y Takeuchi, K Bigsby, RK Meentemeyer. 2021. Iteratively forecasting biological invasions with PoPS and a little help from our friends. *Frontiers in Ecology and the Environment*. 19(7): 411-418.
- Litzow, MA, JD Urban, and BJ Laurel. 2008. Increased spatial variance accompanies reorganization of two continental shelf ecosystems. *Ecol. Appl.* 18(6): 1331–1337.
- Litzow, MA, FJ Mueter, JD Urban. 2013. Rising catch variability preceded historical fisheries collapses in Alaska. *Ecol. Appl.* 23(6): 1475-1487.
- Lovell, SJ, SF Stone, L Fernandez. 2006. The economic impacts of aquatic invasive species: a review of the literature. *Agricultural and Resource Economics Review*. 35(1): 195-208.

- Mendonça, R, RA Müller, D Clow, C Verpoorter, P Raymon, LJ Tranvik, S Sobek. 2017. Organic carbon burial in global lakes and reservoirs. *Nature Communications*. 8:1694.
- Pace, ML, SR Carpenter, RA Johnson, JT Kurtzweil. 2013. Zooplankton provide early warnings of a regime shift in a whole lake manipulation. *Limnology and Oceanography*, 58(2): 525-532.
- Pace, ML, RD Batt, CD Buelo, SR Carpenter, JJ Cole, JT Kurtzweil, and GM Wilkinson. 2017. Reversal of a cyanobacterial bloom in response to early warnings. *Proc. Nat. Acad. Sci. U.S.A.* 114(2): 352-357.
- Paerl, HW, RS Fulton, PH Moisander, and J Dyble. 2001. Harmful freshwater algal blooms, with an emphasis on cyanobacteria. *Sci. World*. 1:76–113.
- Page, T, PJ Smith, KJ Beven, ID Jones, JA Elliott, SC Maberly, EB Mackay, M De Ville, H. Feuchtmayr. 2018. Adaptive forecasting of phytoplankton communities. *Water Research*. 134: 74-85.
- Postal, S, and S Carpenter. 1997. Freshwater ecosystem services; in *Nature's Services: Societal Dependence on Natural Ecosystems*. Island Press.
- Pruitt, JN, AG Little, SJ Majumdar, TW Schoener, DN Fisher. 2019. Call-to-action: a global consortium for tropical cyclone ecology. *Trends in Ecology and Evolution*. 34(7): 588-590.
- Ratajczak Z, P D'Odorico, JB Nippert, SL Collins, NA Brunsell, S Ravi. 2017. Changes in spatial variance during a grassland to shrubland state transition. *Journal of Ecology*. 105: 750-760.
- Recknagel, F, PT Orr, M Bartkow, A Swanepoel, H Cao. 2017. Early warning of limit-exceeding concentrations of cyanobacteria and cyanotoxins in drinking water reservoirs by inferential modelling. *Harmful Algae*. 69: 18-27.
- Rindi, L, MD Bello, L Dai, J Gore, L Benedetti-Cecchi. 2017. Direct observation of increasing recovery length before collapse of a marine benthic ecosystem. *Nature Ecology & Evolution*. 1: 0153.
- Rindi, L, MD Bello, L Benedetti-Cecchi. 2018. Experimental evidence of spatial signatures of approaching regime shifts in macroalgal canopies. *Ecology*. 99(8): 1709-1715.
- Rousso, BZ, E Bertone, R Stewart, DP Hamilton. 2020. A systematic literature review of forecasting and predictive models for cyanobacteria blooms in freshwater lakes. *Water Research*. 182: 115959.
- Scheffer, M, J Bascompte, WA Brock, V Brovkin, SR Carpenter, V Dakos, H Held, EH van Ness, M Rietkerk, and G Sugihara. 2009. Early-warning signals for critical transitions. *Nature*. 461: 53–59.

- Scheffer, M, SR Carpenter, V Dakos, E van Ness. 2015. Generic Indicators of Ecological Resilience: Inferring the Chance of a Critical Transition. *Annu. Rev. Ecol. Evol. Syst.* 46:145–67.
- Seekell, DA, SR Carpenter, T Cline, and ML Pace. 2012. Conditional heteroskedasticity forecasts regime shift in a whole-ecosystem experiment. *Ecosystems*. 15: 741-747.
- Taylor, SD, EP White. 2020. Automated data-intensive forecasting of plant phenology throughout the United States. *Ecological Applications*. 30(1): e02025.
- Thomas, RQ, RJ Figueiredo, V Daneshmand, BJ Bookout, LK Puckett, CC Carey. 2020. A near-term iterative forecasting system successfully predicts reservoir hydrodynamics and partitions uncertainty in real time. *Water Resources Research*, 56, e2019WR026138.
- Trainer, VL, WP Cochlan, A Erickson, BD Bill, FH Cox, JA Borchert, KA Lefebvre. 2007. Recent domoic acid closures of shellfish harvest areas in Washington State inland waterways. *Harmful Algae*. 6(3): 449-459.
- Van Nes, EH, M Scheffer. 2007. Slow recovery from perturbations as a generic indicator of a nearby catastrophic shift. *The American Naturalist*. 169(6): 738-747.
- Walker, B, CS Holling, SR Carpenter, A Kinzig. 2004. Resilience, adaptability and transformability in social-ecological systems. *Ecology and Society*. 9(2):5.
- White, PS, STA Pickett. 1985. Natural Disturbance and Patch Dynamics: An Introduction. *The Ecology of Natural Disturbance and Patch Dynamics*, 3–13. doi:10.1016/b978-0-12-554520-4.50006-x
- Wilkinson, GM, SR Carpenter, JJ Cole, ML Pace, RD Batt, CD Buelo, and JT Kurtzweil. 2018. Early warning signals precede cyanobacterial blooms in multiple whole-lake experiments. *Ecol. Monogr.* 88(2): 188–203.
- Xiao, X, J He, H Huang, TR Miller, G Christakos, ES Reichwaldt, A Ghadouani, S Lin, X Xu, J Shi. 2017. A novel single-parameter approach for forecasting algal blooms. *Water Research*. 108: 222-231.

Chapter 2

A modeling analysis of spatial statistical indicators of thresholds for algal blooms

Published in *Limnology and Oceanography Letters* <https://doi.org/10.1002/lol2.10091>

Abstract

Predicting algal blooms both within and among aquatic ecosystems is important yet difficult because multiple factors promote and suppress blooms. Statistical indicators (e.g. variance and autocorrelation) based on time series can provide warning of transitions in diverse complex systems, including shifts from clear water to algal blooms. Analogous spatial indicators have been demonstrated with models and empirical data from vegetated terrestrial ecosystems. Here, we test the applicability of spatial indicators to algal blooms using a nutrient-phytoplankton spatial model. We found that standard deviation and autocorrelation successfully distinguished bloom state and proximity to transitions, while skewness and kurtosis were more ambiguous. Our findings suggest certain spatial indicators are applicable to aquatic ecosystems despite dynamic physical-biological interactions that could reduce detectable signals. The growing capacity to collect spatial data on algal biomass presents an exciting opportunity for application and testing of spatial indicators to the study and management of blooms.

Introduction

Algal blooms have large impacts on aquatic ecosystems. In oceans, the spring bloom supports growth of zooplankton and eventual fish production (Mann 1993). In more nutrient-rich lakes and coastal oceans, blooms leading to very high concentrations of algae have adverse effects including accelerated nutrient cycling, depletion of hypolimnetic dissolved oxygen that causes fish kills (Swingle 1968), and toxin release that harms grazers and higher consumers

(Christoffersen 1996, Ibelings & Chorus 2007). Blooms may also disrupt aquaculture, recreation, and drinking water supplies (Dodds et al. 2009) and blooms with negative impacts are designated harmful algal blooms, or HABs. High-profile blooms, e.g. in Lake Erie, underscore the scientific consensus that HAB occurrence is increasing worldwide (Ho & Michalek 2017; Heisler et al. 2008).

Understanding algal bloom drivers and dynamics is critical to mitigating their negative impacts. Excess nutrient loading, especially phosphorus in inland waters (Schindler et al. 2016), has been widely studied as a main driver of algal blooms. Grazing, temperature, physical mixing, and other factors promote or suppress phytoplankton and thereby affect blooms (Paerl et al. 2001). The varied drivers of algal blooms complicate predictions of the timing and location of blooms. Reliably predicting blooms can improve prevention strategies (e.g. decreasing nutrient loading) and provide time to minimize bloom impacts (e.g. using algicides).

Algae blooms in eutrophic waters are examples of critical transitions (i.e. abrupt shifts in response to small changes in a forcing) in ecosystems (Batt et al. 2013, Carpenter et al. 1999, Cottingham et al. 2015). An emerging body of literature from fields as diverse as financial markets and human physiology has found that generic “early warning indicators” can provide information on system state and proximity to thresholds (reviewed in Scheffer et al. 2015). Early warning indicators are statistics that change in predictable ways as transitions are approached due to critical slowing down – slowed recovery from perturbations as a system approaches a transition (Scheffer et al. 2009). Variance and standard deviation of ecosystem state variables have been widely used as early warning indicators (e.g. Carpenter and Brock 2006, Pace et al. 2013) and are expected to increase as a critical transition is approached (Biggs et al. 2009). Decreasing return rates near critical transitions due to critical slowing down also lead to

increasing autocorrelation (Dakos et al. 2012). Changes in skewness, another early warning indicator, can capture the dependence of return rate on perturbation direction as well as response to large external fluctuations as thresholds are approached (Guttal & Jayaprakash 2008).

In ecology, early warning indicators have been studied in a variety of ecosystems using modeling, laboratory, and field experiments as well as historical observations and a few whole-ecosystem manipulations (Scheffer et al. 2015). Most studies, especially in aquatic ecosystems, have focused on temporal statistics as early warning indicators (Carpenter et al. 2011, Pace et al. 2017, Wilkinson et al. 2018). Analogous changes in spatial statistics are also expected from critical slowing down near transitions and can potentially provide information on ecosystem state and proximity to thresholds without extensive prior data (Dai et al. 2013). In terrestrial ecosystems including grasslands and shrubland, spatial statistics of vegetation coverage change near critical transitions driven by precipitation and fire frequency (Kéfi et al. 2014, Ratajczak et al. 2016). Spatial early warning indicators have not been studied as widely in aquatic ecosystems, but there are a few examples where changes in spatial statistics are associated with transitions. Donangelo et al. (2010) extended a simple eutrophication model to a spatial grid and found pattern formation near the transition to a degraded state, as well as an increase in spatial variance before temporal variance. Litzow et al. (2008) and Cline et al. (2014) observed changes in fish distribution and spatial variance during regime shifts in the fish community. Rindi et al. (2017) found that the spatial recovery length from perturbations, another indicator of critical slowing down, increased in the rocky intertidal near a transition from canopy to turf dominated benthic algae. A recent study used spatial mapping of phytoplankton pigments to test for indicators of transitions during a whole-lake nutrient manipulation (Butitta et al. 2017). The utility of spatial early warning indicators in these studies is promising, but additional modeling

tests with precisely known system dynamics and transitions can provide understanding of when spatial indicators are expected to give reliable information on ecosystem state and threshold proximity and guide future research.

The spatially patchy distributions observed in algal blooms and captured by spatial models (e.g. Franks 1997) suggest that spatial early warnings may also be useful for analyzing and managing aquatic ecosystems. It is unknown if the highly dynamic (time-varying) spatial distribution of algal blooms could limit the possibility that spatial patterns and statistics provide early warning of critical transitions. Here, we ask: 1) can spatial indicators be used to distinguish between states (e.g. bloom vs. non-bloom) in aquatic ecosystems that undergo critical transitions? and 2) do spatial indicators provide reliable warning of approaching algal bloom critical transitions?

To answer these questions, we applied a published spatial model of algal blooms with known critical transitions (Serizawa et al. 2008). We evaluated the ability of spatial early warning indicators to distinguish algal bloom states and transition proximity in aquatic ecosystems by simulating the model through time for a range of nutrient input levels and then calculating spatial early warning statistics.

Methods

We use a model created by Serizawa et al. (2008) (see Supplemental Information Appendix A for full description). The model is defined on a two dimensional spatial grid (180x180 cells) and represents dynamics of phosphorus and phytoplankton in pelagic systems, which interact nonlinearly via Holling type-II responses. Neighboring grid cells exchange components physically via diffusion and advection, the latter generated by randomly-seeded eddies. Like Serizawa et al. (2008), we use the nondimensional version of the model. This

transformation reduces the number of parameters and does not affect the relative magnitude of the resilience indicators (see equations 3 and 4 in Appendix 1 for transformation between nondimensional and dimensional values). Because this model is dimensionless, the spatial components (extent, resolution, diffusion, and advection) of the grid are defined relative to each other. We've chosen this framework as a basis for exploring spatial indicators conceptually with a model that is reasonably complex but also tractable. The model framework creates patterns that are highly dynamic both spatially and temporally and includes advection and diffusion.

Advection is an important spatial flux in pelagic ecosystems and has not been considered in previous studies of spatial resilience indicators in terrestrial ecosystems. The original model was modified to include stochasticity in the phytoplankton dynamics to represent processes not explicitly included in the model and environmental noise, e.g. local variations in nutrient inputs or weather (equation 1, Appendix 1). Critical slowing down and resilience indicators are based on a system's response to such perturbations changing as a transition is approached. Simulations and calculations were carried out in R 3.4.1 (R Core Team 2017); code is available on GitHub at <https://github.com/cbuelo/SpatialBloomIndicators>.

The model exhibits critical transitions (Hopf bifurcations, where stable fixed points transition to limit cycles or vice versa) in phosphorus-phytoplankton dynamics as the phosphorus input rate, i , is varied. These critical transitions are functions of the model parameter values and define the bloom states, which Serizawa et al. (2008) identify using stability analysis. All analyses use the parameter values from Serizawa et al. (2008, Table 2, set I), and included in Table A1.1. At low i ($0.3 \leq i < 0.5$), the system is in a constant, low-phytoplankton, non-bloom stable state. As i increases, the system enters a stable limit cycle at the low-input transition ($i \approx 0.5$); this cycling bloom state is characterized by repeating cycles in phosphorus and

phytoplankton concentrations. At i above the high-input transition ($i \approx 1.25$) the system returns to a constant-bloom stable state with high phytoplankton concentrations (see Serizawa et al. 2008, Figure 4b).

Integration was done using the Euler-Maruyama method with a small time step (0.025) to approximate continuous development of the model (Higham 2001). The phosphorus input rate was constant for each simulation. Separate simulations, each with a fixed and distinct phosphorus input rate, were compared to assess the ability of spatial indicators to distinguish different states or indicate proximity to thresholds. We define “reliable warning” as an indicator for which neighboring bloom states have distinct indicator values and, within a given state, the indicator changes unambiguously as a transition is approached. For each phosphorus input rate simulation, the model was first run deterministically as in Serizawa et al. (2008; and see Supplemental Information). The long-run ($t = 1000$) deterministic state was used as the starting state for stochastic simulations. Stochastic simulations were run for 500 time units, with system state and spatial statistics retained every 1 time unit after a spin-up period of 100 time units. For each of the 400 “snapshots” at a given phosphorus input rate, the following spatial statistics were calculated from each grid’s phytoplankton concentrations: mean, standard deviation (SD), skewness, kurtosis, Moran’s I, and autocorrelation range (AC range). We tested kurtosis as a potential spatial indicator, expecting that there may be a changing proportion of extreme values as the transitions approached. Moran’s I measures the degree of correlation between neighboring grid cells (analogous to lag-1 temporal autocorrelation) while AC range is the maximal distance over which grid concentrations are correlated; these autocorrelation measures explicitly depend on the spatial distribution of phytoplankton concentrations. The other measures are sample statistics of the 180x180 values in each “snapshot” and do not depend on the specific spatial

distribution of the values on the grid. Skewness and kurtosis were calculated using the *moments* package in R. AC range was computed from the semivariogram using the *gstat* package function `fit.variogram()` and exponential fit.

We evaluated the robustness of our findings by varying the physical forcing. In extreme cases, high diffusion and low advection would eliminate spatial patterns and simulate 180x180 synchronous cells. Relative to the base case from Serizawa et al. (2008), we repeated the simulations for two additional cases: a high diffusion case (double the base case diffusivity) and low advection case (half of the base case advection velocity). Increased diffusivity decreased local concentration gradients and decreased advection slowed patch formation.

Results

Model simulations generated the expected spatial distributions: spatially uniform low and high phytoplankton concentrations in the low and high phosphorus input stable states, respectively, and patchy patterns of high and low concentrations in the intermediate-input cycling state (Figure 2.1). The spatial patterns in the cycling state changed and repeated through time, as in Serizawa et al. (2008).

For the stable states at both low and high phosphorus input rates, spatially uniform phytoplankton concentrations were maintained through time and had relatively constant mean grid phytoplankton concentration and spatial indicators at a given input rate (Figure 2.2). In the intermediate-input rate cycling bloom state, temporally cycling spatial patterns in phytoplankton concentration resulted in cycling spatial statistics. Both the grid mean and spatial indicators exhibited repeating cycles in time for a given phosphorus input level.

All spatial indicators, except for Moran's I, were highly variable in the intermediate-input cycling bloom state and more constrained in both the non-bloom and constant-bloom stable

states (Figure 2.3). Mean phytoplankton concentration increased across the range of phosphorus input rates, however within the cycling bloom state there was significant overlap in the distributions of grid mean phytoplankton concentration between adjacent input rates (Figure 2.3A). Both standard deviation (SD) and skewness had “humped” patterns within the cycling state where indicator values peaked in the middle of the phosphorus input range and declined near the transitions (Figures 2.3B and 2.3E). In the cycling state, SD values had less overlap between adjacent input rates near the transitions and no overlap with the stable states, whereas skewness had a large degree of overlap especially at the high input transition. Autocorrelation range (AC range) had the opposite pattern within the cycling state; values were highest and variable near the transitions and decreased at phosphorus input rates in the middle of the cycling state (Figure 2.3C). Moran’s I was the only indicator that was highly constrained (near 1) in the cycling bloom state and was also the only indicator to decrease steadily within increasing input rates in the constant-bloom stable state (Figure 2.3D). Kurtosis, while constrained in the stable states, was highly variable in the cycling bloom state with a high degree of overlap and no trends near the transitions (Figure 2.3F).

Increasing diffusion or decreasing advection did not have a strong effect on the statistical moment spatial indicators (mean, standard deviation, skewness, and kurtosis) in any of the bloom states (Figure 2.3A, 2.3B, 2.3E, and 2.3F, respectively). Increasing diffusion increased both autocorrelation indicators relative to the base case. AC range was most strongly effected in the cycling bloom state (Figure 2.3C) while the increase in Moran’s I was stronger in the non-bloom and constant-bloom stable states (Figure 2.3D). Decreased advection had minimal impact on all statistics.

To determine if spatial indicators can reliably distinguish proximity to thresholds, we selected the spatial indicators from Figure 2.3 that were best at distinguishing between bloom states and studied them at a higher resolution of phosphorus input rate near the transitions. Bloom state could be unambiguously inferred from SD, AC range, and Moran's I at phosphorus input rates near the critical transitions (Figure 2.4). Standard deviation increased steadily with phosphorus input rate through the transition from the low-input stable state to intermediate-input cycling state, with little or no overlap between distributions of SD at adjacent input rates (Figure 2.4A). At the transition from the cycling bloom state to constant-bloom stable state, standard deviation decreased and was relatively constant at phosphorus input rates above the transition (Figure 2.4B). Autocorrelation range was low and overlapped at input rates from 0.3 – 0.4, increased sharply at $i = 0.45$ before the low-input transition occurred, then gradually declined from $i = 0.5$ to $i = 0.7$ (Figure 2.4C). AC range increased gradually with overlapping distributions as the high-input transition was approached at $i = 1.25$, then fell sharply and overlapped from $i = 1.3$ – 1.45 (Figure 2.4D). Moran's I also increased sharply at $i = 0.45$ below the low-input transition and was high (near 1) in the cycling bloom state (Figure 2.4E). Moran's I remained near 1 as the high-input transition was approached in the cycling bloom state and declined slightly at $i = 1.25$ before falling sharply at $i = 1.3$ (Figure 2.4F). From $i = 1.3$ to $i = 1.45$ in the high-input stable bloom state, Moran's I declined steadily with relatively little overlap among adjacent input values.

Discussion

Spatial indicators can discern ecosystem state and proximity to thresholds, even in a model of a highly dynamic pelagic system with interacting physical and biological components. SD, AC range, and Moran's I differed between the cycling and stable states, allowing for

classification of bloom state (non-bloom stable, cycling bloom, or constant-bloom stable) from a single spatial snapshot. At phosphorus input rates near both transitions in bloom state, trends in these indicators provided a reliable warning of approaching transition, indicating that changes in spatial indicators calculated from repeated sampling in time would occur prior to crossing thresholds. Skewness and kurtosis had somewhat distinct distributions between states but significant overlap within states, suggesting that these indicators would be unreliable for determining threshold proximity.

Spatial standard deviation was one of the most robust indicators of both ecosystem state and threshold proximity. An approximately order of magnitude difference in SD on either side of the transitions clearly indicated the state of the system. Additionally, the monotonically increasing and non-overlapping distributions prior to the low-input transition provide a relative measure of proximity to the approaching threshold. At the high-input transition from the cycling bloom state to the constant-bloom stable state there was a slightly less than order of magnitude difference in SD. The decline in SD as this transition was approached from lower phosphorus input levels may provide an indication of the approaching threshold, although there was some overlap between adjacent phosphorus input levels. The observed increases in SD as transitions were approached from the stable states match with findings of temporal early warning indicator studies (Carpenter and Brock 2006, Carpenter et al. 2012, Pace et al. 2013). The maximal values of SD occurred at intermediate input levels within the cycling state, corresponding to the maximum cycle amplitude in phytoplankton concentration (Figure 2.3 this study, Figure 4b from Serizawa et al. 2008).

Both measures of autocorrelation were indicative of bloom state and threshold proximity. AC range and Moran's I provided warning of the low input transition from the stable non-bloom

state to the cycling bloom state, increasing sharply from $i = 0.4$ to $i = 0.45$ prior to the transition at $i = 0.5$. The steady decrease in Moran's I at input rates above the high-input transition combined with a fairly small degree of overlap in this statistic between adjacent input rates could provide early warning of transition from the stable, constant-bloom state to the cycling bloom state (i.e. if input rates were high but decreasing, Figure 2.4F). Moran's I provides clearer warning of proximity to transition to the cycling state (from either low or high input rates) than AC range. In contrast, AC range provides more information than Moran's I on proximity to transitions from within the cycling state but is more ambiguous in the high-input stable state. The differences between AC range and Moran's I are perhaps not surprising as the interactions between physical processes, biological processes, and return rate from perturbations have different effects on correlation between neighboring cells (Moran's I) and those further apart (AC range). The minimum in AC range within the cycling state likely corresponds to a maximum in return rate at the phytoplankton cycle amplitude maximum. The increase in AC range in the high diffusion case may be the result of decreased local gradients in phytoplankton concentration near patch edges.

While SD and Moran's I changed unambiguously as at least one transition was approached from at least one direction, they also displayed statistic-dependent differences in sensitivity to the low- and high-input transitions, depending on the direction from which the transition is approached (SD, Figure 2.4B; Moran's I, Figures 2.4E and 2.4F). While critical transitions are most frequently studied in the context of ecosystem degradation (e.g., increased nutrient inputs), we also observed early warnings in the opposite scenario which may be useful in systems undergoing remediation (e.g. Lake Washington, Hampton et al. 2006). Asymmetric warnings prior to transitions (i.e. whether approached from lower or higher input rates) have also

been observed in studies of temporal resilience indicators using models with Hopf bifurcations (Batt et al. 2013). In our study, warning asymmetry may result from an interaction between the deterministic dynamics of the model and the structure of the stochasticity (Horsthemke and Lefever 1984, Benincà et al. 2011). These results reinforce previous suggestions that a resilience indicator approach should only be used when there is reason to suspect a critical transition and data on appropriate state variables are collected (Wilkinson et al. 2018, Gsell et al. 2016).

In this study, both skewness and kurtosis were highly constrained in the stable states and much more variable in the cycling states. Despite these differences between bloom states, the lack of trends near transitions and overlap of the indicator distributions in different states limit the applicability of skewness and kurtosis as spatial indicators for this model. Contrasting with our results, Guttal and Jayaprakash (2009) found spatial skewness an unambiguous indicator of an impending regime shift using a two dimensional model of terrestrial vegetation collapse. We suspect these differences arise from the different model types used in each study; their model includes a single fold-bifurcation separating two spatially homogenous states while the model used in this study has a spatially-patterned intermediate state separated from stable states by Hopf bifurcations. Overall, studies of spatial resilience are less numerous than studies of temporal resilience and more work is needed to understand differences among types of ecosystems.

This work demonstrates the potential applicability of spatial resilience indicators to pelagic aquatic systems. Spatial statistics have been successfully tested as resilience indicators in arid terrestrial ecosystem models containing critical transitions in vegetation state (Reitkerk et al. 2004, Kéfi et al. 2007, Guttal & Jayaprakash 2009, Dakos et al. 2011). Recent empirical tests in terrestrial ecosystems including experimental grassland manipulations (Ratajczak et al. 2017)

and applications of remote sensing (Eby et al. 2017) have supported model findings. Many of these terrestrial vegetation models include diffusive exchange between neighboring grid cells. Compared to rooted and relatively stationary vegetation on land, the physical forcing of pelagic environments (e.g. both diffusion and advective currents) plays a more immediate role in determining the movement of free-floating phytoplankton and nutrients. These forces could “wash out” signals of aquatic ecosystem state and transitions. However, a few empirical studies have shown promising results in the application of spatial resilience indicators to aquatic ecosystems with organisms that have more control over their spatial pattern than phytoplankton, including fish (Litzow et al. 2008, Cline et al. 2014) and intertidal benthic algae (Rindi et al. 2017).

It is also possible that the rapid time scales on which algal blooms develop and spatial patterns form and change could limit the use of spatial resilience indicators. However, a recent empirical study applied several of the indicators used here to spatial data from an experimentally fertilized lake (Butitta et al. 2017). Butitta et al. (2017) found that SD was highest prior to and during an induced algal bloom and declined after fertilization ended and the bloom abated. AC range was highest prior to and after the bloom peak, but declined for a short period coincident with peak bloom conditions. Both the SD and AC range observations of Butitta et al. (2017) are consistent with the findings in this study of a shift from the non-bloom stable state past the low-input transition into the cycling bloom state, and then a return to the stable non-bloom state when fertilization ceased. Contrasting with our findings, Butitta et al. found that skewness was elevated prior to and during the bloom and maximal just after the bloom peak. It is not possible to compare spatial statistics in the constant-bloom stable state as Butitta et al. (2017) stopped nutrient additions after temporal early warning signs were observed.

While our findings suggest that spatial indicators can differentiate bloom state and threshold proximity using a model that creates dynamic patches, evaluation and application of spatial indicators requires further research. The model could be expanded to include the vertical dimension of the water column and more ecological and physical detail (e.g. additional nutrients, light limitation, spatially and temporally varying zooplankton density). Most importantly, the model could be used in tandem with field experiments to improve understanding of spatial dynamics of blooms under diverse conditions. Empirical studies could also address questions related to the spatial and temporal scales at which spatial indicators occur, e.g. do spatial indicators change prior to blooms and early enough to be useful for management? What spatial resolution and temporal sampling frequency is required? How do temporal and spatial indicators compare?

The increasing ease and cost-efficiency of collecting spatial data on phytoplankton biomass provides the opportunity to test and improve the predictions in this study. Technological advances have made it possible to spatially map photosynthetic pigments both in-situ (Crawford et al. 2014) and via remote sensing (Tyler et al. 2016). We advocate additional work using modeling, empirical, and combined studies to further develop and test the applicability of these methods. Successful use of spatial indicators could allow the classification of the current state of individual aquatic systems as well as their proximity to thresholds. These goals are particularly important as continued nutrient inputs to inland waters and coastal marine systems increase the potential for eutrophication and harmful algal blooms (Sinha et al. 2017).

Acknowledgements

This material is based upon work supported by the National Science Foundation Graduate Research Fellowship Program under grant number DGE 1315231 and DEB grant numbers

1144624, 1144683, 1456151, and 148425. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation. We thank the reviewers and editor whose comments and suggestions improved an initial version of this paper.

References

- Batt, RD, WA Brock, SR Carpenter, JJ Cole, ML Pace, DA Seekell. 2013. Asymmetric response of early warning indicators of phytoplankton transition to and from cycles. *Theoretical Ecology*, 6(3): 285-293.
- Benincà, E, V Dakos, EH van Nes, J Huisman, M Scheffer. 2011. Resonance of plankton communities with temperature fluctuations. *The American Naturalist*, 178: E85-E95
- Biggs, R, SR Carpenter, WA Brock. 2009. Turning back from the brink: Detecting an impending regime shift in time to avert it. *Proceedings of the National Academy of Sciences*, 106(3): 826-831.
- Butitta, VL, SR Carpenter, LC Loken, ML Pace, EH Stanley. 2017. Spatial early warning signals in a lake manipulation. *Ecosphere* (in press).
- Carpenter, SR, D Ludwig, WA Brock. 1999. Management of eutrophication for lakes subject to potentially irreversible change. *Ecological Applications*, 9(3): 751-771.
- Carpenter, SR, WA Brock. 2006. Rising variance: a leading indicator of ecological transition. *Ecology Letters*, 9: 311-318.
- Carpenter, SR, JJ Cole, ML Pace, R Batt, WA Brock, T Cline, J Coloso, JR Hodgson, JF Kitchell, DA Seekell, L Smith, B Weidel. 2011. Early warnings of regime shifts: a whole-ecosystem experiment. *Science*, 332(6033): 1079-1082.
- Christoffersen, K. 1996. Ecological impacts of cyanobacterial toxins in aquatic food webs. *Phycologia*, 36(6): 42-50.
- Cline, TJ, DA Seekell, SR Carpenter, ML Pace, JR Hodgson, JF Kitchell, BC Weidel. 2014. Early warnings of regime shifts: evaluation of spatial indicators from a whole-ecosystem experiment. *Ecosphere*, 5(8): 102.
- Cottingham, KL, HA Ewing, ML Greer, CC Carey, KC Weathers. 2015. Cyanobacteria as biological drivers of lake nitrogen and phosphorus cycling. *Ecosphere*, 6(1): 1-19.
- Crawford, JT, LC Loken, NJ Casson, C Smith, AG Stone, LA Winslow. 2015. High-speed limnology: using advanced sensors to investigate variability in Biogeochemistry and Hydrology. *Environmental Science & Technology*, 49: 442-450.

- Dai, L, KS Korolev, J Gore. 2013. Slower recovery in space before collapse of connected populations. *Nature*, 496: 355-358.
- Dakos, V, S Kefi, M Rietkerk, EH van Nes, M Scheffer. 2011. Slowing down in spatially patterned ecosystems at the brink of collapse. *The American Naturalist*, 177(6): E153-E166.
- Dakos, V, EH van Nes, P D'Odorico, M Scheffer. 2012. Robustness of variance and autocorrelation as indicators of critical slowing down. *Ecology*, 93(2): 264-271.
- Dodds, WK, WW Bouska, JL Eitzmann, TJ Pilger, AJ Riley, JT Schloesser, DJ Thornbrugh. 2009. Eutrophication of U.S. freshwaters: analysis of potential economic damages. *Environmental Science & Technology*, 43(1): 12-19.
- Donangelo, R, H Fort, V Dakos, M Sheffer, EH Van Ness. 2010. Early warnings for catastrophic shifts in ecosystems: comparison between spatial and temporal indicators. *International Journal of Bifurcation and Chaos*, 20(2): 315-321.
- Eby, S, A Agrawal, S Majumder, AP Dobson, V Guttal. 2017. Alternative stable states and spatial indicators of critical slowing down along a spatial gradient in a savanna ecosystem. *Global Ecology and Biogeography*. 26:638-649.
- Franks, PJS. 1997. Spatial patterns in dense algal blooms. *Limnology and Oceanography*, 45(5, part 2): 1297-1305.
- Gsell, AS, U Scharfenberger, D Ozkundacki, A Walters, LA Hansson, ABG Janssen, P Noges, PC Reid, DE Schindler, E Van Donk, V Dakos, R Adrian. 2016. Evaluating early-warning indicators of critical transitions in natural aquatic ecosystems. *Proceedings of the National Academy of Sciences*, 113(50): E8089-E8095.
- Guttal, V, C Jayaprakash. 2008. Changing skewness: an early warning signal of regime shifts in ecosystems. *Ecology Letters*, 11: 450-460.
- Guttal, V, C Jayaprakash. 2009. Spatial variance and spatial skewness: leading indicators of regime shifts in spatial ecological systems. *Theoretical Ecology*, 2(1): 3-12.
- Hampton, SE, MD Scheuerell, DE Schindler. 2006. Coalescence in the Lake Washington story: interaction strengths in a pelagic food web. *Limnology and Oceanography*, 51(5): 2042-2051.
- Heisler, J, PM Gliber, JM Burkholder, DM Anderson, W Cochlan, WC Dennison, Q Dortch, CJ Gobler, CA Heil, E Humphries, A Lewitus, R Magnien, HG Marshall, K Sellner, DA Stockwell, DK Stoecker, M Suddleson. 2008. Eutrophication and harmful algal blooms: a scientific consensus. *Harmful Algae*, 8: 3-13.
- Higham, DJ. 2001. An algorithmic introduction to numerical simulation of stochastic differential equations. *SIAM Review*, 43(3): 525-546.

- Ho, JC, and AM Michalak. 2017. Phytoplankton blooms in Lake Erie impacted by both long-term and springtime phosphorus loading. *Journal of Great Lakes Research* 43: 221-228.
- Horsthemke, W, R Lefever. 1984. *Noise-induced transitions: theory and applications in physics, chemistry and biology*. Springer-Verlag.
- Kéfi, S, M Rietkerk, CL Alados, Y Pueyo, VP Papanastasis, A ElAich, PC de Ruiter. 2007. Spatial vegetation patterns and imminent desertification in Mediterranean arid ecosystems. *Nature*, 449: 213-218.
- Kéfi S, V Guttal, WA Brock, SR Carpenter, AM Ellison, VN Livina, DA Seekell, M Scheffer, EH van Nes, and V Dakos. 2014. Early warning signals of ecological transitions: methods for spatial patterns. *PLoS ONE* 9: e92097.
- Ibelings, BW, I Chorus. 2007. Accumulation of cyanobacterial toxins in freshwater “seafood” and its consequences for public health: a review. *Environmental Pollution*, 150(1): 177-192.
- Litzow, MA, JD Urban, BJ Laurel. 2008. Increased spatial variance accompanies reorganization of two continental shelf ecosystems. *Ecology Applications*, 18(6): 1331- 1337.
- Mann, KH. 1993. Physical oceanography, food chains, and fish stocks: a review. *ICES Journal of Marine Science* 50 (2): 105-119.
- Pace, ML, SR Carpenter, RA Johnson, JT Kurtzweil. 2013. Zooplankton provide early warnings of a regime shift in a whole lake manipulation. *Limnology and Oceanography*, 58(2): 525-532.
- Pace, ML, RD Batt, CD Buelo, SR Carpenter, JJ Cole, JT Kurtzweil, GM Wilkinson. 2017. Reversal of a cyanobacterial bloom in response to early warnings. *Proceedings of the National Academy of Sciences*, 114(2): 352-357.
- Paerl, HW, RS Fulton, PH Moisander, J Dyble. 2001. Harmful freshwater algal blooms, with an emphasis on cyanobacteria. *The Scientific World*, 1: 76-113.
- Ratajczak, Z, P D’Odorico, JB Nippert, SL Collins, NA Brunsell, S Ravi. 2017. Changes in spatial variance during a grassland to shrubland state transition. *Journal of Ecology*, 105: 750-760.
- Rietkerk, M, SC Dekker, PC de Ruiter, and J van de Koppel. 2004. Self-organized patchiness and catastrophic shifts in ecosystems. *Science*, 305:1926-1929.
- R Core Team. 2017. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria, URL <https://www.R-project.org>.
- Rindi, L, MD Bello, L Dai, J Gore, L Benedetti-Cecchi. 2017. Direct observation of increasing recovery length before collapse of a marine benthic ecosystem. *Nature Ecology & Evolution*. 1: 0153.

- Scheffer, M, J Bascompte, WA Brock, V Brovkin, SR Carpenter, V Dakos, H Held, EH van Ness, M Rietkerk, G Sugihara. 2009. Early-warning signals for critical transitions. *Nature*, 461: 53-59.
- Scheffer, M, SR Carpenter, V Dakos, EH van Ness. 2015. Generic indicators of ecological resilience: inferring the chance of a critical transition. *Annual Review of Ecology, Evolution, and Systematics*, 46: 145-167.
- Schindler, DW, SR Carpenter, SC Chapra, RE Hecky, DM Orihel. 2016. Reducing phosphorus to curb lake eutrophication is a success. *Environmental Science & Technology*, 50(17): 8923-8929.
- Serizawa, H, T Amemiya, K Itoh. 2008. Patchiness in a minimal nutrient-phytoplankton model. *Journal of Biosciences*, 33(3): 391-403.
- Sinha, E, AM Michalak, V Balaji. 2017. Eutrophication will increase during the 21st century as a result of precipitation changes. *Science*, 357, 405-408.
- Swingle, HS. 1968. Fish kills caused by phytoplankton blooms and their prevention. *FAO (F.A.O.U.N.) Fish. Rep.* 44(5):407-411.
- Tyler, A, P Hunter, E Spyrakos, S Groom, A Constantinescu, J Kitchen. 2016. Developments in Earth observation for the assessment and monitoring of inland, transitional, coastal and shelf-sea waters. *Science of the Total Environment*, 572, 1307-1321.
- Wilkinson, GM, SR Carpenter, JJ Cole, ML Pace, RD Batt, CD Buelo, JT Kurtweil. 2018. Early warning signals precede cyanobacterial blooms in multiple whole-lake experiments. *Ecological Monographs*, 88(2): 188-203.

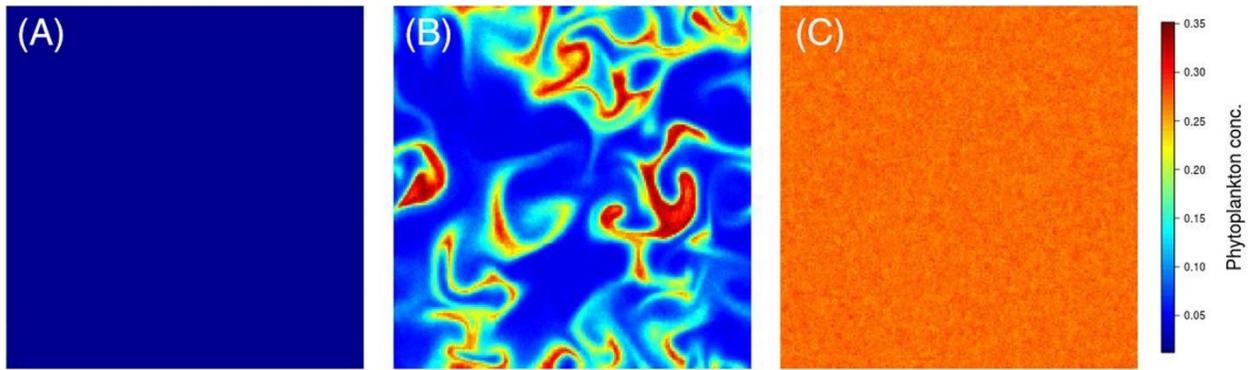


Figure 2.1. Example grids showing the spatial distribution of phytoplankton concentrations for the (A) non-bloom stable state ($i = 0.35$), (B) cycling bloom state ($i = 0.9$), and (C) constant-bloom stable state ($i = 1.6$) at $t = 100$ of the stochastic simulations.

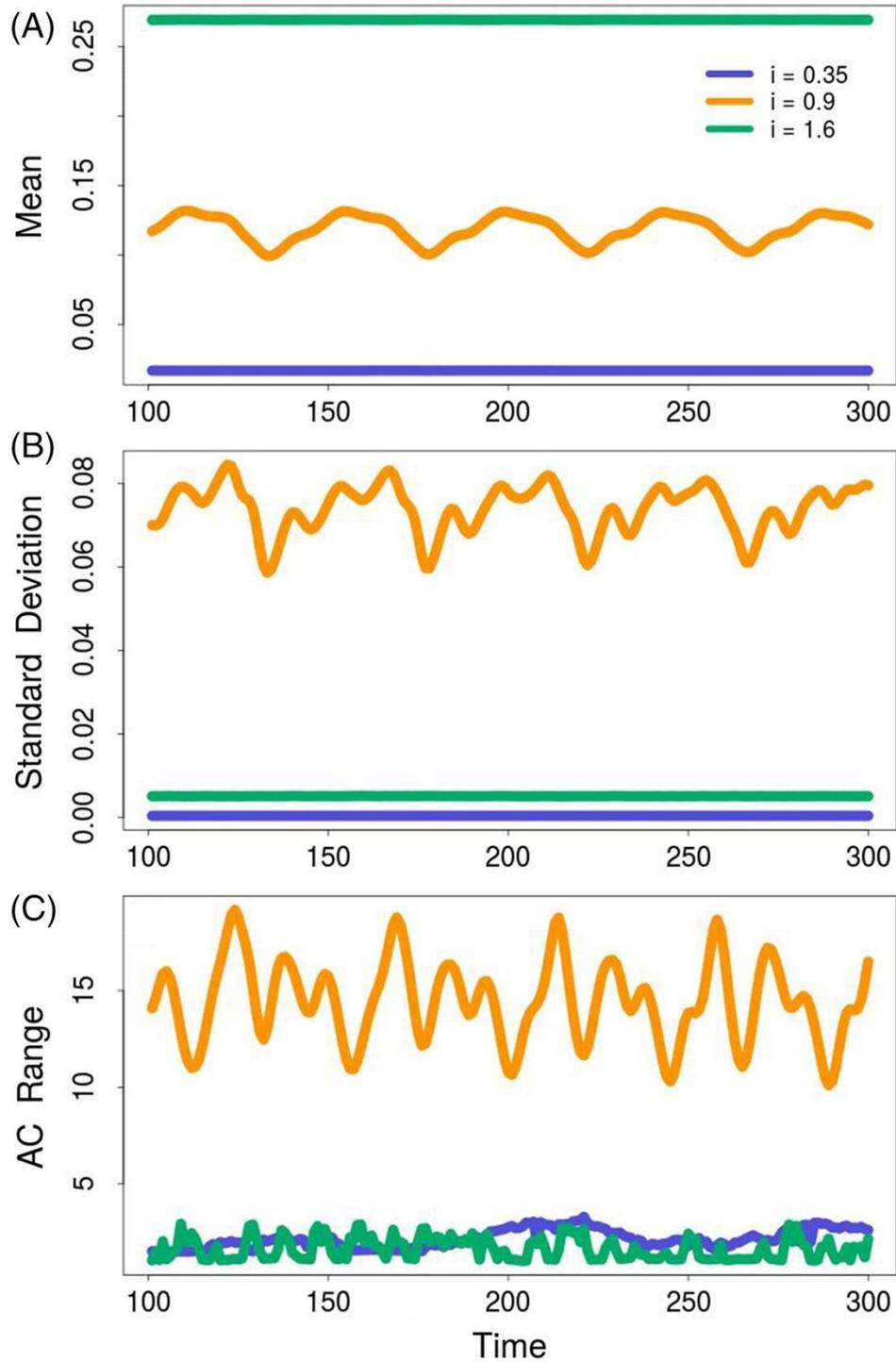


Figure 2.2. Spatial indicators (A) mean, (B) spatial standard deviation, (C) autocorrelation range vs. time (dimensionless) for the nonbloom stable state (blue, $i = 0.35$), cycling bloom state (gold, $i = 0.9$), and constant-bloom stable state (green, $i = 1.6$).

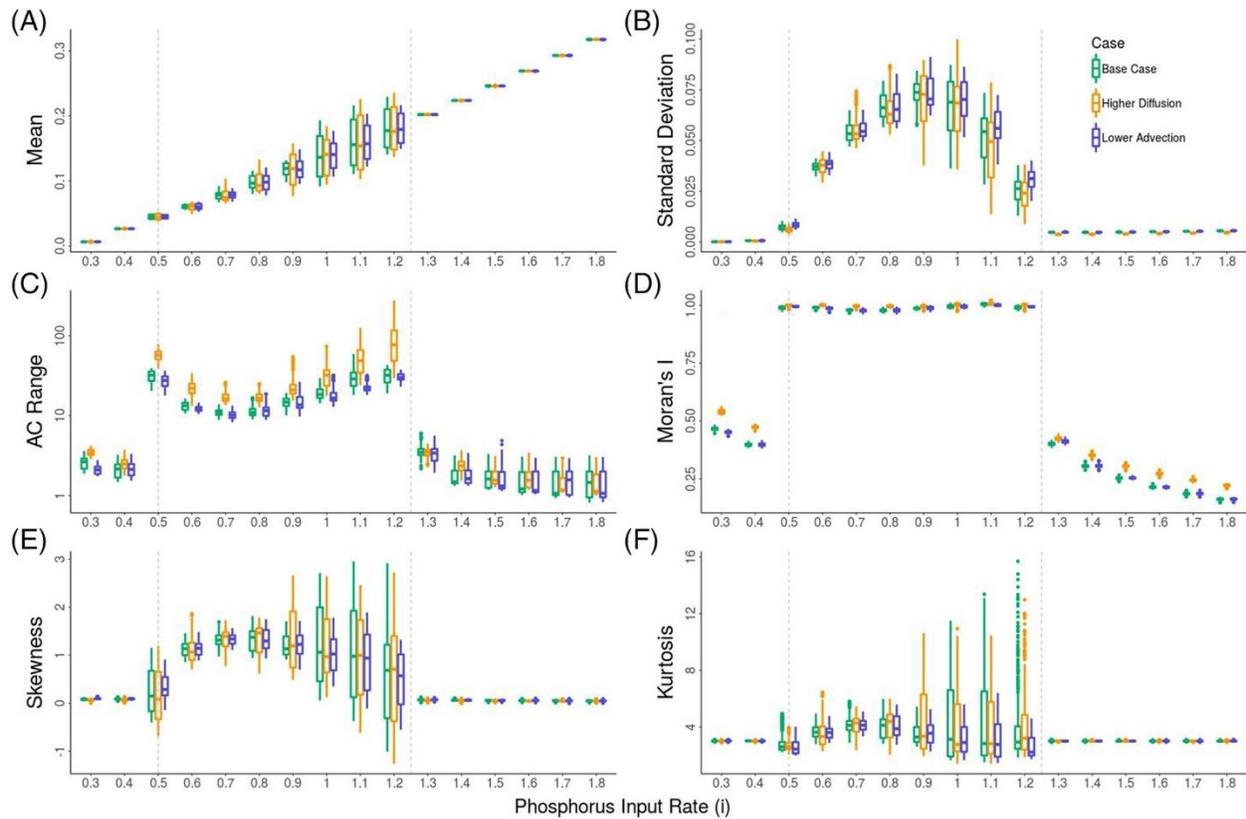


Figure 2.3. Spatial indicators calculated from 400 snapshots of grid phytoplankton concentration for a range of phosphorus concentrations including spatial mean (A), SD (B), autocorrelation range (C), Moran's I (D), skewness (E), and kurtosis (F). Physical forcing was varied from the base case (green) by increasing diffusion (gold; $2\times$ base case) and decreasing advection (blue; $0.5\times$ base case). Vertical dashed lines represent the phosphorus concentrations at which transitions in bloom state occur. Note different vertical axis scales for each indicator.

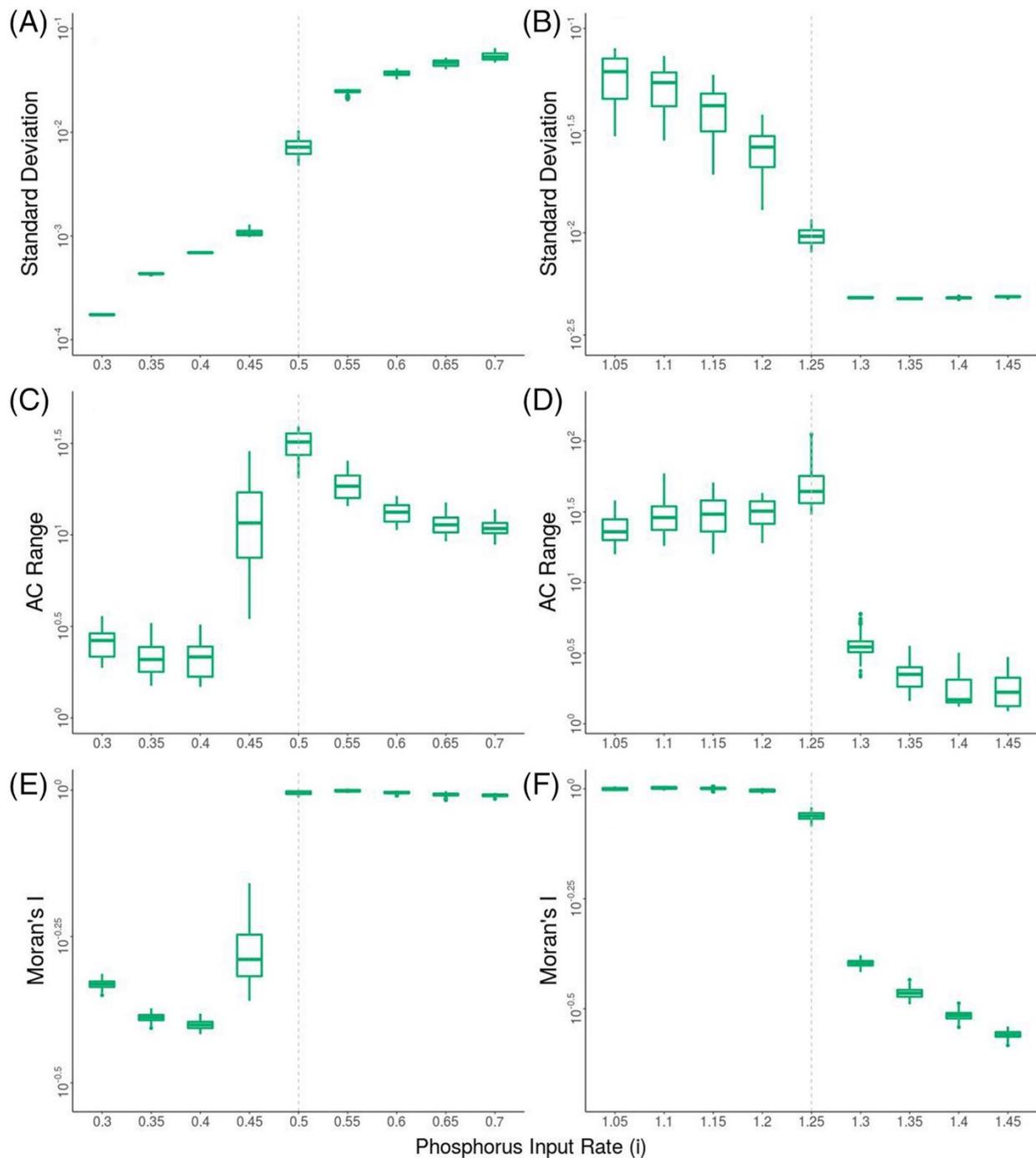


Figure 2.4. Spatial indicators (A and B: standard deviation, C and D: autocorrelation range, E and F: Moran's I) near critical transitions from the nonbloom stable state to the cycling bloom state (A, C, E) and from the cycling bloom state to constant-bloom stable state (B, D, F). Vertical dashed lines represent the phosphorus concentrations at which transitions in bloom state occur. Note independently scaled and log-transformed vertical axes used to highlight trends across transitions.

Chapter 3

Evaluating the performance of temporal and spatial early warning statistics of algal blooms

In Review at *Ecological Applications*

Abstract

Regime shifts have large consequences for ecosystems and the services they provide. However, understanding the potential for, causes of, proximity to, and thresholds for regime shifts in nearly all settings is difficult. Generic statistical indicators of resilience have been proposed and studied in a wide range of ecosystems as a method to detect when regime shifts are becoming more likely without direct knowledge of underlying system dynamics or thresholds. These early warning statistics (EWS) have been studied separately but there have been few examples that directly compare temporal and spatial EWS in ecosystem-scale empirical data. To test these methods, we collected high-frequency time series and high-resolution spatial data during a whole-lake fertilization experiment while also monitoring an adjacent reference lake. We calculated two common EWS, standard deviation and autocorrelation, in both time series and spatial data to evaluate their performance prior to the resulting algal bloom. We also applied the quickest detection method to generate binary alarms of resilience change from temporal EWS. One temporal EWS, rolling window standard deviation, provided advanced warning in most variables prior to the bloom, showing trends and between-lake patterns consistent with theory. In contrast, temporal autocorrelation and both measures of spatial EWS (spatial SD, Moran's I) provided little or no warning. By compiling time series data from this and past experiments with and without nutrient additions, we were able to evaluate temporal EWS performance for both constant and changing resilience conditions. True positive alarm rates were 2.5 – 8.3 times higher for rolling window standard deviation when a lake was being

pushed towards a bloom than the rate of false positives when it was not. For rolling window autocorrelation, alarm rates were much lower and no variable had a higher true positive than false positive alarm rate. Our findings suggest temporal EWS provide advanced warning of algal blooms and that this approach could help managers prepare for and/or minimize negative bloom impacts.

Introduction

Large changes in ecosystems often reveal important aspects of dynamics and have significant consequences. Changes can be abrupt and have obvious exogenous causes, like wildfires or introductions of exotic species, or have more subtle causes, such as gradual changes of internal regulation, that push drivers past thresholds (Scheffer et al. 2001, Ratajczak et al. 2018). Research on ecosystems, financial markets, and the human body have identified generic statistics that indicate relative resilience and proximity to regime shifts without knowing precise thresholds at which changes will occur (Scheffer et al. 2009). Statistics such as variance and autocorrelation of measurements of ecosystem state have been proposed as a potential management tool for predicting regime shifts (Biggs et al. 2009). However, early warning statistic (EWS) methods are not yet used in management due to performance questions arising from data requirements and ambiguity in the precise mechanisms driving ecosystem dynamics (Boettiger et al. 2013).

Despite these challenges, research using models, experiments, and historical data has used EWS successfully to detect coming transitions in complex systems. Ecosystem models have been used to explore how several time series statistics (e.g. variance, autocorrelation, skewness) change near thresholds (Carpenter and Brock 2006, Dakos et al. 2012, Guttal and

Jayaprakash 2008). These predictions have also been tested using lab and field experiments (Drake and Griffin 2010, Dai et al. 2012, Carpenter et al. 2011). Analogous spatial statistics have been proposed and tested (Dakos et al. 2011, Ratajczak et al. 2017). While these examples and others identified potential transitions in advance, other studies have failed to identify the hypothesized changes in EWS before regime shifts and question the generality of these methods (Hastings and Wysham 2010, Bestelmeyer et al. 2011, Sommer et al. 2017, Spears et al. 2017). Changes in resilience of real-world ecosystems may not be readily differentiated from other sources of variability (Perretti and Munch 2012). EWS across different statistics and ecosystem state variables can have variable reliability and low agreement, and are dependent on ecosystem-specific knowledge (Gsell et al. 2016). Others have suggested that studying EWS using historical observations only in systems that have undergone regime shifts inflates the rate of false positives (Boettiger and Hastings 2012).

Given limitations and uncertainties, applying EWS requires understanding key mechanisms underlying ecosystem dynamics well enough to know the potential for a critical transition, the capacity to collect high resolution data on relevant ecosystem state variables to detect changes in resilience, and the ability to distinguish changes in EWS due to resilience loss from other causes (e.g. by using concomitant measurements in an unmanipulated reference ecosystem). Additionally, evaluating EWS performance requires observations at different levels of resilience that are either known from direct manipulation of drivers or measurements of them. However, examples from many ecosystems have demonstrated that regime shifts occur and data appropriate for EWS analysis are becoming available (Scheffer et al. 2015, Dakos et al. 2015). Thus, there is increasing potential to test EWS by experimentally manipulating ecosystem

drivers (Carpenter et al. 2011, Rindi et al. 2017) or by evaluating observations over space or time where drivers controlling resilience vary naturally (Litzow et al 2008, Eby et al. 2017).

Algal blooms are a phenomenon where EWS could benefit managers and the public. Algal blooms are a shift from a clear-water to an algae-dominated state involving a critical transition to runaway algal growth driven mainly by changing controls on nutrient availability (Carpenter et al. 1999, Serizawa et al. 2008, Cottingham et al. 2015). Blooms can comprise toxin-producing algae species, requiring treatment of surface water supplies at the source or in water treatment facilities. Bloom die-offs and subsequent decomposition can also harm species by depleting oxygen (Swingle 1968). In many aquatic ecosystems, blooms occur frequently but not predictably, with intermittent periods of relatively clear water. In these cases of recurring blooms, advanced warning could allow managers to take steps to avoid or minimize impacts. For example, managers could decrease point or nonpoint nutrient sources or apply alum or algicides (Welch and Cooke 1999). In other systems where blooms can't be avoided, advanced warning provided by EWS could still give managers time to alter public use (e.g. beach closings, fish/shellfish harvest advisory) or to switch water supply sources.

Prior studies of algal blooms provide evidence for early warnings. Models of algal dynamics approaching critical transition find both temporal (Batt et al. 2019) and spatial EWS (Donangelo et al. 2010; Buelo et al. 2018). These findings have also been tested in field studies. Pace et al. (2017) used temporal EWS to reverse a bloom by halting nutrient additions when EWS alarms were detected; in the same experiment spatial EWS were significantly different before, during, and after the bloom (Butitta et al. 2017). Wilkinson et al. (2018) found that temporal EWS were often consistent with predictions from theory, but that performance of different state variable-statistic combinations varied among experiments. Finally, Ortiz et al.

2020 demonstrated that temporal EWS increase prior to non-experimental algal blooms in 3 out of 4 lake-years.

While the studies above demonstrate the promise of EWS for algal blooms, several questions remain to be answered prior to applying EWS for bloom management:

- Are temporal or spatial EWS better predictors of algal blooms?
- How do temporal EWS perform under an approaching bloom (near threshold of change) and when a bloom is not imminent?
- Do spatial EWS change prior to bloom onset, and if so how reliable are those changes?

To answer these questions, we did a whole-lake fertilization experiment designed to promote an algal bloom while conducting high-resolution spatial and temporal data collection. We monitored both an experimental and unmanipulated lake during pre-manipulation and manipulation years to assess EWS performance for the spectrum of bloom and non-bloom conditions identified above. We also analyzed temporal data from other years (both fertilized and not) and an additional experimental lake to evaluate consistency of temporal EWS performance.

Methods

Experimental Design

Our experiment was carried out at the University of Notre Dame Environmental Research Center in the Upper Peninsula of Michigan, United States. Peter and Paul Lakes were the experimental and reference lakes, respectively. These lakes are immediately adjacent to one another and were a single lake before being divided by an earthen dam in 1951 (Leavitt et al. 1989). They have previously been used in whole-lake manipulation experiments (Carpenter and Pace 2018). Both are relatively small (2.6 ha for Peter Lake and 1.7 ha for Paul Lake) and deep

(mean depth 5.7 and 3.7 meters, max depth approximately 18 and 12 meters) with bowl-shaped basins and food webs dominated by largemouth bass (*Micropterus salmoides*). The lakes are oligo- to mesotrophic and algal blooms do not occur naturally without experimental fertilization. There are no major inflows or outflows with the exception of a small culvert that drains from Paul Lake to Peter Lake during high water conditions. This combined with their undisturbed, forested watersheds, make the lakes ideal systems in which to conduct experimental nutrient addition (hereafter fertilization) experiments. More detailed description of these lakes is available in Carpenter and Kitchell (1993).

During the first year of the experiment, 2018, data were collected without experimental fertilization to establish baseline conditions and typical values of early warning statistics when the lake is not being pushed towards a critical transition. In the second year, 2019, nutrients were added to Peter Lake daily starting on day of year 161 and ending on day of year 237. Solutions of phosphoric acid and ammonium nitrate were prepared and distributed by pumping them into the prop-wash of a boat propelled by an electric motor. Nutrients were added at a fixed 15:1 molar ratio of N:P, with the loading rate starting at $0.5 \text{ mg P} / \text{m}^2 / \text{day}$ for the first week. Every 7 days, the nutrient loading rate was increased by $0.5 \text{ mg P} / \text{m}^2 / \text{day}$ until reaching a rate of $5 \text{ mg P} / \text{m}^2 / \text{day}$ in week 10, which was then maintained for 7 additional days before stopping nutrient additions. Figure A2.1 in the Appendix 2 shows the daily and cumulative P added to Peter Lake in 2019.

The date of bloom onset was determined using a lake-specific threshold of 14 ug/L chlorophyll for Peter Lake derived from historical data (Wilkinson et al. 2018). For evaluation of early warning statistic performance, the experiment was divided into pre-manipulation (all of

2018 and 2019 before nutrient additions began) and pre-bloom fertilization (from the start of nutrient additions in 2019 to bloom onset) periods.

In addition to data from this experiment, data from previous nutrient addition studies were also used to assess the consistency of temporal EWS across multiple experiments. These include three additional years (2013 - 2015) of data in Peter and Paul lakes as well as nearby Tuesday Lake; both Peter and Tuesday lakes were experimentally fertilized in these years and Paul Lake serving as a reference system (see Wilkinson et al. 2018 for full description).

Additional lake-years with daily data and no fertilization experiments were also analyzed: 2011 in Peter and Paul lakes (variables include chlorophyll-a, dissolved oxygen saturation, and pH; but not phycocyanin fluorescence), and 2016 in Peter, Paul, and Tuesday Lakes (phycocyanin fluorescence, dissolved oxygen, and pH; but not chlorophyll-a).

Data Collection

Time series data were collected from a raft at the center of each lake. Two automated, multi-parameter water quality sensors (hereafter sondes) were deployed at each raft at a depth of 0.75 meters, one EXO 3 manufactured by YSI, Inc. and one Hydrolab DS5X manufactured by OTT Hydromet. The sondes measured phycocyanin fluorescence (phyco; a cyanobacteria pigment), dissolved oxygen saturation (D.O. sat.), and pH every 5 minutes. Sondes were calibrated monthly. Phycocyanin sensors used the manufacturer-provided calibration curves to convert fluorescence to physical units (ug/L of phycocyanin for the YSI EXO 3, cells/mL for the Hydrolab DS5X); Hydrolab DS5X phycocyanin fluorescence time series were used to allow direct comparison to prior experiments. These measurements should be interpreted as relative measures of cyanobacteria abundance as we did not calibrate them to extracted phycocyanin concentrations or microscopy cell counts. In previous studies, we found a strong relationship

between sonde fluorescence measurements and extracted phycocyanin concentrations (Pace et al. 2017). Outlier and instrument error values were removed from sonde time series data, as were data from periods when the sonde was out of the water for cleaning or recalibration. These gaps in the high frequency data were filled by fitting a multivariate autoregressive state space model (R package MARSS version 3.10.10, Holmes et al. 2010, 2018) to the two simultaneously collected time series (each variable measured by each sonde), and then averaged to daily values.

Chlorophyll-a (chl-a) was also measured daily by manually collecting water samples which were then filtered, frozen, methanol extracted, and measured in the laboratory using a Turner Trilogy benchtop fluorometer (Holm-Hansen 1978). A single outlier chl-a value of 13.1 ug/L in the reference lake on day of year 180 of 2019 was removed and filled by linear interpolation.

Spatial data were collected using the FLAMe system (**F**ast **L**imnological **A**utomated **M**easurements, Crawford et al. 2015). The FLAMe uses a flow-through design to collect spatial measures of water quality parameters, pumping water from an intake through a sensor array while the boat is underway, and simultaneously recording GPS position. Spatial variables included phycocyanin fluorescence (ug/L), D.O. saturation (%), and pH measured with a YSI, Inc. EXO 2 sonde. The FLAMe system was mounted to a flat bottom boat powered by an electric motor and driven at a speed of approximately 1.5 m/s in a grid pattern. Lakes were surveyed with adjacent tracks in the grid approximately 15 meters apart. Data and GPS coordinates were collected at a frequency of 1 Hz. Spatial sampling was conducted weekly during the non-fertilization year (2018) and 3 times per week (Monday, Wednesday, Friday) during the fertilization year (2019). Spatial data were processed and corrected for the hydrologic residence of the system as well as sensor-specific response times (Crawford et al. 2015).

Analysis

Temporal early warning statistics - Rolling window standard deviations (SD) and lag-1 autocorrelation (AR(1)) were calculated for each lake for all variables (chl-a, phyco, DO sat, and pH). We use a window width of 21 days as has been used in prior studies (Wilkinson et al. 2018, Ortiz et al. 2020) and shown to balance accuracy and speed in detecting changes in EWS. To test if changes in rolling window EWS were consistent with predictions from theory, that SD and AR(1) increase as a critical transition is approached, we calculated Kendall's tau on the difference between rolling window EWS (experimental lake - reference lake) during the pre-bloom fertilization period.

We also applied the quickest detection method to the calculated rolling window EWS to evaluate for evidence that the experimental lake had switched from a baseline, or high resilience, state to an alarm, or low resilience, state (Carpenter et al. 2014). The quickest detection method is an online method, updating with each data point to give a binary "alarm" or "no alarm" status (Polunchenko and Tartakovsky 2012). It is based on the Shiryaev-Roberts statistic, which accumulates evidence from the likelihood ratio that the experimental lake is in the alarm vs. baseline states. When the S-R statistic gets large enough, it generates an alarm, suggesting the experimental lake is in a low resilience state, and then resets and can be triggered again; see Appendix A2 for full description and equations for the method.

The baseline and alarm states are characterized by distributions of the rolling window statistics: $f(x)$ for the baseline state and $g(x)$ for the alarm state. Consistent with theory, the alarm state for both SD and AR(1) is higher relative to the baseline state. For rolling window SD, $f(x) \sim N(\mu_{\text{baseline}}, \sigma_{\text{pool}})$ and $g(x) \sim N(\mu_{\text{alarm}}, \sigma_{\text{pool}})$ where μ_{baseline} is the observed rolling window standard deviation in the reference lake, $\mu_{\text{alarm}} = \mu_{\text{baseline}} + 2 * \sigma_{\text{pool}}$, and σ_{pool} is the pooled standard

deviation of the reference and experimental lake (Wilkinson et al. 2018). For rolling window autocorrelation, $f(x)$ and $g(x)$ are given by the exact distribution for Pearson correlation coefficient; with ρ equal to the observed lag-1 correlation in the reference lake for the baseline distribution, and ρ equal to 0.95 for the alarm distribution. This formulation for QD of rolling window AR(1) is an update to that used in previous studies (e.g. Pace et al. 2017, Wilkinson et al. 2018) which was based on a first-order error propagation approximation; the new “exact” method avoids erroneous alarms that occur when experimental lake AR(1) is less than reference lake AR(1) (see SI of Wilkinson et al. 2018).

Quickest detection (QD) alarms were classified as either “true alarms” if they occurred during the pre-bloom fertilization period, or “false alarms” if they occurred when the experimental lake was not being fertilized, either in 2018 or 2019 prior to the start of nutrient additions. Alarms occurring after bloom onset were designated “late alarms” and were not included in any analyses. True and False Positive Rates (TPR and FPR) for each variable were calculated by dividing the total number of alarms observed during a period (pre-manipulation or pre-bloom fertilization) by the total number of days in that period. These analyses were also carried out for all lake-years and variables with daily data available as described in *Experimental Design* above.

Spatial early warning statistics - Analogous spatial early warning statistics were calculated for data collected by the FLAME system. Standard deviation of all spatial points collected from each lake on a sampling date was calculated for all variables. Spatial autocorrelation was quantified using Moran’s I with inverse distance weighting using the *ape* and *geosphere* R packages (Paradis and Schlieff 2019, Hijmans 2019). For each variable and statistic, a paired t test was performed to test if the spatial statistics were significantly different

between the experimental and reference lakes during the pre-bloom fertilization period, based on expectations that spatial EWS would be higher in the experimental than the reference lake.

Changes in spatial statistics through time during the pre-bloom fertilization period were also tested for as described for temporal EWS above, by calculating Kendall's tau on the difference in the statistics between lakes (experimental - reference).

Results

Nutrient additions in 2019 created a large algal bloom in the experimental lake, with chlorophyll-a crossing the 14 ug/L bloom threshold on day of year (DOY) 201 after 41 days of fertilization (Figure 3.1). Chlorophyll-a (chl-a) and phycocyanin fluorescence (phyco) peaked 8 days after the bloom threshold was crossed, at a chl-a concentration of 49 ug/L, and then declined and rebounded over the next 13 days. A secondary peak in pigment concentrations was followed by a 12-day crash, and then by an increase that crossed the bloom threshold for a second time on DOY 240. These bloom dynamics in the experimental lake were also reflected in D.O. saturation and pH, daily averages of which peaked at greater than 130% and 9.5, respectively (Figure 3.1). Reference lake dynamics closely tracked the experimental lake prior to nutrient additions, except for higher dissolved oxygen and less acidic pH in the experimental lake. These offsets are likely largely due to past liming experiments in the experimental lake (Leavitt et al. 1989).

Bloom dynamics were also reflected in spatial observations from the FLAME system. Representative spatial data from before the bloom, near its peak, and after the bloom decline for phycocyanin, D.O. saturation, and pH indicate the large bloom in the experimental lake and little change in the reference lake (Figure 3.2). Overall, spatial variability at any sampling event was modest; 94% of lake-date-variable combinations had a coefficient of variation < 0.03 and a

maximum value of 0.058 in experimental lake phycocyanin fluorescence during the bloom. In the non-fertilization year, spatial DO and pH were similar to the reference and experimental lakes (excepting the offset mentioned above) and no large increases in primary producers were observed. Weekly maps of spatial data are included in Appendix 2 (Figures A2.2 – A2.4), and a web app (R Shiny application) for visualizing any sampling event(s) is available in the *tvsews* package (<https://github.com/cbuelo/tvsews>) using the `plot_FLAMe_maps()` function.

Temporal early warning statistics largely matched expectations from theory for rolling window standard deviation (SD). For all variables, rolling window SD was much higher during the bloom relative to non-bloom periods in the experimental lake and relative to the reference lake (Figure 3.3). For three of the four temporal data variables (chl-a, phyco, and pH) rolling window SD was higher in the experimental lake than the reference lake during the pre-bloom fertilization period, generating “true” quickest detection (QD) alarms. There was little difference in rolling window SD of D.O. saturation during the entire pre-bloom period (both before and during nutrient additions), and no QD alarms were generated. Both chl-a and phyco during the pre-manipulation period had times when rolling window SD was higher in the experimental lake than the reference lake, resulting in “false” QD alarms. In total for rolling window SD, there were four false QD alarms and three true QD for chl-a, one false alarm and seven true alarms for phyco, zero false and zero true alarms for D.O. saturation., and zero false and two true alarms for pH (Figure 3.3, Table 3.1). The difference between experimental lake and reference lake rolling window SD increased significantly over the pre-manipulation period (Kendall’s tau positive, $p < 0.05$) for both phyco and pH consistent with theory, while there was no significant change for chl-a or D.O. sat. ($p > 0.05$).

Rolling window lag-1 autocorrelation (AR(1)) largely did not show expected differences between lakes and changes through time during the pre-bloom fertilization period. The only exceptions were two periods for chl-a (approximately DOY 161 - 188 and after 187 of 2019) and one period for phyco (after DOY 188) when AR(1) was higher in the experimental lake than in the reference lake (Figure 3.3). However, the differences between lakes were not large enough, and AR(1) in the experimental lake not high enough, during these periods to generate QD alarms. The only time that AR(1) generated QD alarms in the manipulation year occurred after the start of the bloom: five late Chl-a alarms occurred between DOY 204 and 215 of 2019, when rolling window AR(1) was approximately 0.9 in the experimental lake and less than 0.25 in the reference lake. While there were no true alarms for rolling window AR(1) of phyco during the pre-bloom fertilization period, there was one false alarm on DOY 192 of 2018 and one late alarm on DOY 236 of 2019. For D.O. sat. and pH, rolling window AR(1) did steadily increase during much of the pre-bloom fertilization period in the experimental lake, but those increases were mirrored by increases in reference lake AR(1) (Figure 3.3). While rolling window AR(1) was not high enough in the experimental lake relative to the reference lake to generate true QD alarms, the difference between experimental lake and reference lake AR(1) did increase for phyco and D.O. sat (Kendall's tau positive, $p < 0.05$) over the pre-bloom fertilization period (Table 3.1). There also was a significant decrease in the difference for pH (Kendall's tau negative, $p < 0.05$).

There was significant sample-to-sample variability in both spatial statistics of all variables (Figure 3.4). For all three primary producer indicator variables measured by the FLAMe, spatial standard deviation (SD) in the experimental lake peaked and was markedly elevated at times during the bloom relative to the non-bloom period. During the pre-bloom fertilization period, only D.O. sat. had higher mean spatial variability in the experimental lake

than the reference lake (paired t test, $p < 0.05$). Also during this period, temporal trends in the difference between experimental lake and reference lake SD were not significant for any variable (Kendall's tau, all $p > 0.05$), though spatial SD of pH increased steadily in the experimental lake over the four spatial sampling dates just prior to bloom onset (Figure 3.4).

There was significant overlap in spatial autocorrelation (Moran's I) across all lakes, variables, and time periods (Figure 3.4). Unlike for spatial SD, Moran's I was largely similar between the bloom and non-bloom periods, with the exception of Moran's I in phyco being consistently elevated in the experimental lake relative to the reference lake from DOY 205 to 217 after the start of the bloom. During the pre-bloom fertilization period, no variable had significantly different Moran's I between the lakes (paired t test, $p > 0.05$), and only pH had a significantly increasing difference through time (Kendall's tau positive, $p < 0.05$; Table 3.1).

For the 2018 – 2019 experiment, all variables with QD alarms in rolling window SD (i.e. excluding D.O. sat.) had a higher true positive rate (TPR) during the pre-bloom fertilization period than false positive rate (FPR) during the pre-manipulation period. Phyco and chl-a had the highest true positive rates, at 0.175 and 0.075 alarms per day respectively, pH had a slightly lower TPR (0.05), and D.O. sat. had a TPR of 0 as there were no alarms during pre-bloom fertilization period. D.O. sat. and pH had the lowest false positive rates (FPR = 0), followed by phyco (0.011), and then chl-a (0.056). Based on the difference between TPR and FPR, phyco was the best performing variable in 2018 – 2019 followed by pH, chl-a, and lastly D.O. sat.

Including all lakes and years of data from prior experiments, the relative performance of the pigment or biomass variables (chl-a and phyco) was better than the process (i.e., affected by primary production) variables (D.O. sat and pH). For rolling window SD, chl-a and phyco had TPR near 0.15 and FPR < 0.05 alarms per day (Figure 3.5). pH had the third highest TPR (0.089)

followed by D.O. sat. (0.055). The FPR of SD for pH (0.035) was slightly lower than for Chl-a and phyco, while D.O. sat had the lowest FPR at 0.007 (Figure 3.5).

With no rolling window autocorrelation (AR(1)) alarms in any variable during the pre-bloom fertilization period of the 2019 manipulation year, TPR was 0 for AR(1) across all variables for this 2018 – 2019 experiment (Figure 3.3). The single observed false alarm in the 2018 – 2019 experiment in phyco AR(1) yields a FPR rate of 0.011 alarms per day for phyco and 0 for all other variables. Including data from all lake-years, the AR(1)alarm rate was positive but low for both TPR and FPR of chl-a and phyco as well as FPR of D.O. sat, and zero for TRP of D.O. sat and TPR and FPR of pH (Figure 3.5). No variable had $TPR > FPR$ for AR(1)alarms whether looking at data from all lake-years or just the 2018 – 2019 experiment.

Discussion

We found that temporal EWS performed better than spatial EWS for predicting algal blooms and quantified the performance of temporal EWS in a realistic management scenario. Temporal EWS agreed with predictions from theory for more variable-statistic combinations than for analogous spatial EWS prior to an experimental bloom, with spatial EWS having significant variability between sampling events. Further, the rate of true positive alarms of resilience loss was higher than the rate of false positives in all variables across multiple lakes and experiments for rolling window standard deviation. In contrast, rolling window autocorrelation had much worse performance across all variables.

For temporal EWS, 50% of variable-statistic combinations showed the expected increase in the difference between the manipulated reference lake over the pre-bloom fertilization period (Table 3.1). As in previous studies of temporal EWS of experimental algal blooms (Pace et al. 2017, Wilkinson et al. 2018), we found that the quickest detection method generated “true”

alarms during the pre-bloom fertilization period. Also in agreement with those studies, rolling window standard deviation performed better (true alarms in 3 out of 4 variables and 12 total true alarms in our study) than rolling window autocorrelation (0 out of 4 variables and 0 total true alarms). The low number of pre-bloom alarms in autocorrelation in this experiment was surprising, though AR(1) alarms were observed when analyzing prior data from experiments (Figure 3.5). Notably, the updated “exact” quickest detection method for rolling window autocorrelation used here (Appendix 2) did not produce any alarms when AR(1) in the experimental lake was less than in the reference lake (see Appendix S1 of Wilkinson et al. 2018). We recommend the current implementation be used in future work to minimize misleading alarms.

The relatively poor performance of spatial EWS ran counter to expectations from previous studies. Donangelo et al. 2010 and Buelo et al. 2018 found that spatial variability and autocorrelation increased near critical transitions in models of eutrophication and algal blooms, respectively. But in our experiment, only pH had an increase in the difference between lakes for Moran’s I, and that difference was driven by low Moran’s I values in the experimental and reference lakes at the beginning and end of the pre-bloom fertilization period, respectively. Besides spatial SD of DO sat. being consistently higher in the experimental lake, the most conspicuous patterns in spatial EWS during the pre-bloom fertilization period was high sample-to-sample variability and a large degree of overlap between lakes. The lack of change in spatial EWS was likely caused by limited spatial heterogeneity throughout the pre-bloom period. However, as was found by prior study of spatial EWS (Butitta et al. 2017), we did observe large changes in spatial EWS when comparing bloom and non-bloom states; all variables had large spikes in spatial SD at the peak of the bloom, as did Moran’s I for phyco.

Our study design, a whole-lake manipulation including a reference system and non-manipulation year, avoids many of the limitations of EWS studies based on post-hoc analysis of historical data. Studying EWS in historical data can help develop intuition on EWS methodology, e.g. which variables, statistics, data resolutions, and rolling window widths are likely to be successful (Dakos et al. 2008, Lenton et al. 2012, Bell et al. 2017). However, studies based on existing data can also bias findings towards positive performance of EWS due to the selection of cases observed to have undergone large changes, whether or not the underlying driver was a loss of resilience (Boettiger et al. 2012). One way to minimize bias is by using historical data from an ideally random and representative set of systems with and without regime shifts, as was done in a recent study of temporal EWS in tree growth using dendrochronology (Cailleret et al. 2019). The growing proliferation of ecological data available from research and monitoring programs as well as earth observing systems, including for aquatic ecosystems and algal blooms, opens the possibility of applying these types of analyses more broadly (Hampton et al. 2013, Meinson et al. 2016, Eby et al. 2017). Properly designed experimental studies, while difficult to perform at the ecosystem scale, can provide powerful tests of findings from historical EWS studies and establish expectations for performance in applied management scenarios.

Control of nutrient inputs in this experiment and the availability of an immediately adjacent reference system allowed us to connect changes in EWS to changes in resilience with a high degree of confidence. This along with extensive historical data when these lakes both were and were not fertilized provided *a priori* knowledge of the regime shift and its underlying causes to robustly test EWS. Prior data and studies provided a bloom threshold that allowed us to classify QD alarms as true, false, or late. However, there is no widely accepted universal definition of a bloom (Isles and Pomati 2021) and a lower threshold could have resulted in some

true alarms being classified as late alarms, decreasing the true positive rate (TPR) of alarms, though in the current experiment only one or two SD phyco alarms would likely be affected. Conversely, a slightly higher bloom threshold could have reclassified late alarms in SD of chl-a, phyco, and pH as true alarms, raising the TPRs. In most systems, the drivers of regime shifts and their values are not likely to be known, and indeed a primary motivation for EWS methods is to derive insight on relative system resilience without complete knowledge of underlying dynamics (Scheffer et al. 2015). The success of EWS in this study (i.e. temporal EWS changes agreeing with theoretical predictions, and higher TPR than FPR) suggests that these methods may have utility to managers in predicting algal blooms when they have adequate reason to suspect a regime shift (Spears et al. 2017).

However, caution is also warranted. In the current experiment and across all lake-years, rolling window SD performed much better than rolling window autocorrelation, contradicting results from modeling studies that both standard deviation and autocorrelation should increase as critical transitions are approached (Batt et al. 2013a) or that increases in autocorrelation are more consistent than increases in variance before regime shifts (Carpenter et al. 2009; Dakos et al. 2012). The low number of quickest detection alarms in AR(1) is not surprising given the similarity in AR(1) between the experimental and reference lakes (Figure 3.3). Non-co-occurrence of EWS in different statistics has been noted in other EWS studies using data from long-term studies (Gsell et al. 2016, Burthe et al. 2016). The dependence of autocorrelation on time scale and sampling frequency is one possible explanation, though a recent study found that changes in temporal autocorrelation before a bloom were largely similar from time scales of minutes to days (Batt et al. 2019). The constrained range (-1 to 1) and expected value at the

critical point of a transition (1) of AR(1) likely also plays a role in its decreased sensitivity relative to SD, which can vary from 0 to infinity.

In addition to differences in temporal EWS performance between statistics, we also found significant performance differences across variables and between the current experiment and all lake-years. Across all lake-years, the two algal pigment variables (chl-a and phyco) had the highest TPR and difference between TPR and FPR for rolling window SD (Figure 3.5), while pH and DO sat. had lower TPR and FPR. This is consistent with previous studies that documented contrasting EWS performance depending on which ecosystem state variable is considered (Carpenter et al. 2008, Batt et al. 2013b, Cailleret et al. 2019). In comparison, rolling window SD performance for Chl-a was much worse for the current experiment (TPR more than 50% lower and FPR slightly higher), while phyco performed much better (higher TPR, FPR more than 50% lower). DO sat. and pH also had significant differences in performance comparing all lake-years vs. just the current experiment. These differences demonstrate that while intuition or best practices maybe emerge from prior studies (e.g. “algal pigments are the most reliable variables” or “rolling window SD alarms perform better than AR(1)”), such generalizations should not be taken as hard-and-fast rules. In practice, managers should consider the number of alarms, the variables, and statistics in which they occur in to determine the likelihood of a regime shift; but more and likely system-specific examples are needed to explicitly quantify that likelihood.

In conclusion, this study offers the first direct comparison we are aware of between temporal and spatial EWS at the ecosystem scale for an experimentally-induced regime shift. Spatial EWS did not perform well due to high sample-to-sample variability, suggesting that relatively frequent spatial sampling may be required to discern changes in spatial EWS caused by resilience loss. Sampling at a lower frequency than done in this study could have increased the

chances of observing a misleading trend due to limited sample size. However, these methods should be tested further in different systems where spatial patterns may be more likely due to wind, currents, and heterogeneity in drivers of algal growth. Increasing ability to collect spatial data on lake characteristics using autonomous gliders and remote sensing (Austin 2013, Kislik et al. 2018) will continue to lower the difficulty of conducting such studies. We also demonstrate and quantify for the first time that temporal EWS, and the quickest detection method specifically, provide higher true positive alarm rates when an ecosystem is being pushed towards a critical transition than false positive rates when it is not. While we have applied this method as was done in past studies, further work should evaluate which parameters (rolling window size, alarm threshold, offset between the baseline and alarm states, etc.) optimize alarm performance including cases without contemporaneous data from a reference lake. Quickest detection performance should also be compared to other methods for predicting algal blooms, both EWS based (Ortiz et al. 2020) and near-term forecasting methods (Carey et al. 2021) for bloom state variables. Ultimately, prediction performance and uncertainty (EWS based or otherwise) will have to be incorporated into a framework that includes the costs and benefits of different interventions. For example, prediction reliability may be less critical if prevention or mitigation strategies are relatively inexpensive, easy, or quick to implement (e.g. issuing public warnings at beaches, stopping point sources of nutrients) vs. expensive, difficult, or slow (e.g. treating an entire water body with alum or algicides, shutting down or switching water supplies, reducing non-point or internal nutrient loading). Managers will inevitably have to balance the reliability of any prediction with the costs of taking action.

Acknowledgements

We thank Kevin Gauthier, Carson Lambert, Sara McCormack, Anne Marie Saunders, Mike Smale, and Kayla Wernsing for their assistance carrying out the experiment and collecting data, and to Paul Schramm for help with the FLAMe system. Feedback from Jonathan Walter and Spencer Tassone improved an earlier version of this manuscript. This work was supported by the National Science Foundation under grants DGE 1315231, DEB 1754712 and 1753854, and a Long-Term Ecological Research (LTER) award to North Temperate Lakes (NTL) DEB 1440297. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

References

- Austin, J. 2013. The potential for Autonomous Underwater Gliders in large lake research. *Journal of Great Lakes Research*. 39: 8 - 13.
- Batt, RD, WA Brock, SR Carpenter, JJ Cole, ML Pace, and DA Seekell. 2013a. Asymmetric response of early warning indicators of phytoplankton transition to and from cycles. *Theor. Ecol.* 6(3): 285–293.
- Batt, RD, SR Carpenter, JJ Cole, ML Pace, RA Johnson. 2013b. Changes in ecosystem resilience detected in automated measures of ecosystem metabolism during a whole-lake manipulation. *Proc. Nat. Acad. Sci. U.S.A.* 110(43): 17398-17403.
- Batt, RD, T Eason, A Garmestani. 2019. Time scale of resilience loss: Implications for managing critical transitions in water quality. *PLOS ONE*. 14(10): e0223366. <https://doi.org/10.1371/journal.pone.0223366>
- Bell, S, V Baudrot, A Lami, S Musazzi, V Dakos. 2017. Rising variance and abrupt shifts of subfossil chironomids due to eutrophication in a deep sub-alpine lake. *Aquatic Ecology*. 51: 307-319.
- Bestelmeyer BT, AM Ellison, WR Fraser, KB Gorman, SJ Holbrook, CM Laney, MD Ohman, DPC Peters, FC Pillsbury, A Rassweiler, RJ Schmitt, S Sharma. 2011. Analysis of abrupt transitions in ecological systems. *Ecosphere*. 2(12): 1-26.
- Biggs, R, SR Carpenter, WA Brock. 2009. Turning back from the brink: Detecting an impending regime shift in time to avert it. *Proc. Nat. Acad. Sci. U.S.A.* 106(3): 826–831.
- Boettiger, C, and A Hastings. 2012. Early warning signals and the prosecutor’s fallacy. *Proc. R. Soc. B*, 279: 4734–4739.

- Boettiger, C, N Ross, A Hastings. 2013. Early warning signals: the charted and uncharted territories. *Theoretical Ecology*, 6(3): 255-264.
- Buelo, CD, SR Carpenter, ML Pace. 2018. A modeling analysis of spatial statistical indicators of thresholds for algal blooms. *Limnology and Oceanography Letters*, 3(5):384–392.
- Butitta, VL, SR Carpenter, LC Loken, ML Pace, and EH Stanley. 2017. Spatial early warning signals in a lake manipulation. *Ecosphere*, 8(10): e01941.
<https://doi.org/10.1002/ecs2.1941>.
- Burthe, SJ, PA Henrys, EB Mackay, BM Spears, R Campbell, L Carvalho, B Dudley, IDM Gunn, DG Johns, SC Maberly, L May, MA Newell, S Wanless, IJ Winfield, SJ Thackeray, F Daunt. 2016. Do early warning indicators consistently predict nonlinear change in long-term ecological data? *Journal of Applied Ecology*. 53: 666-676.
- Cailleret, M, V Dakos, S Jansen, EMR Robert, T Aakala, and 44 others. 2019. Early-warning signals of individual tree mortality based on annual radial growth. *Front. Plant Sci.* 9: 1964.
- Carey, CC, WM Woelmer, ME Lofton, RJ Figueiredo, BJ Bookout, RS Corrigan, W Daneshmand, AG Hounshell, DW Howard, ASL Lewis, RP McClure, HL Wander, NK Ward, RQ Thomas. 2021. Advancing lake and reservoir water quality management with near-term, iterative ecological forecasting. *Inland Waters*. doi: 10.1080/20442041.2020.1816421
- Carpenter, SR, and WA Brock. 2006. Rising variance: a leading indicator of ecological transition. *Ecol. Lett.* 9: 311–318.
- Carpenter, SR, WA Brock, JJ Cole, ML Pace. 2008. Leading indicators of trophic cascades. *Ecol. Lett.* 11: 128 - 138.
- Carpenter, SR, WA Brock, JJ Cole, ML Pace. 2009. Leading indicators of phytoplankton transitions caused by resource competition. *Theo. Ecol.* 2: 139 - 148.
- Carpenter, SR, WA Brock, JJ Cole, ML Pace. 2014. A new approach for rapid detection of nearby thresholds in ecosystem time series. *Oikos*. 123(3): 290-297.
- Carpenter, SR, JJ Cole, ML Pace, R Batt, WA Brock, T Cline, J Coloso, JR Hodgson, JF Kitchell, DA Seekell, L Smith, B Weidel. 2011. Early warnings of regime shifts: a whole-ecosystem experiment. *Science*, 332(6033): 1079–1082.
- Carpenter, SR and JF Kitchell (eds.). 1993. *The trophic cascade in lakes*. Cambridge.
- Carpenter, SR, D Ludwig, WA Brock. 1999. Management of eutrophication for lakes subject to potentially irreversible change. *Ecol. Appl.* 9(3): 751–771.

- Carpenter, SR, and ML Pace. 2018. Synthesis of a 33-year series of whole-lake experiments: effects of nutrients, grazers, and precipitation-driven water color on chlorophyll. *Limnol. Oceanogr. Lett.* 3: 419-427
- Crawford, JT, LC Loken, NJ Casson, C Smith, AG Stone, and LA Winslow. 2015. High-speed limnology: using advanced sensors to investigate variability in biogeochemistry and hydrology. *Environ. Sci. Technol.* 49: 442–450.
- Cottingham, KL, HA Ewing, ML Greer, CC Carey, and KC Weathers. 2015. Cyanobacteria as biological drivers of lake nitrogen and phosphorus cycling. *Ecosphere* 6(1): 1–19.
- Dai, L, D Vorselen, KS Korelev, J Gore. 2012. Generic indicators for loss of resilience before a tipping point leading to population collapse. *Science.* 336(6085): 1175-1177.
- Dakos V, SR Carpenter, EH van Nes, M Scheffer. 2015. Resilience indicators: prospects and limitations for early warnings of regime shifts. *Phil. Trans. R. Soc. B.* 370: 20130263.
- Dakos, V, S Kefi, M Rietkerk, EH van Nes, and M Scheffer. 2011. Slowing down in spatially patterned ecosystems at the brink of collapse. *Am. Nat.* 177(6): E153–E166.
- Dakos, V, M Scheffer, EH van Nes, V Brovkin, V Petoukhov, H Held. 2008. Slowing down as an early warning signal for abrupt climate change. *Proc. Nat. Acad. Sci. U.S.A.* 105(38): 14308 - 14312.
- Dakos, V, EH van Nes, P D'Odorico, and M Scheffer. 2012. Robustness of variance and autocorrelation as indicators of critical slowing down. *Ecology* 93(2): 264–271.
- Donangelo, R, H Fort, V Dakos, M Sheffer, and EH Van Nes. 2010. Early warnings for catastrophic shifts in ecosystems: comparison between spatial and temporal indicators. *Int. J. Bifurc. Chaos* 20(2): 315–321.
- Drake JM and BD Griffen. 2010. Early warning signals of extinction in deteriorating environments. *Nature.* 467(7314): 456-459.
- Eby, S, A Agrawal, S Majumder, AP Dobson, and V Guttal. 2017. Alternative stable states and spatial indicators of critical slowing down along a spatial gradient in a savanna ecosystem. *Glob. Ecol. Biogeogr.* 26: 638–649.
- Gsell, AS, U Scharfenberger, D Özkundakci, A Walters, LA Hansson, ABG Janssen, P Nöges, PC Reid, DE Schindler, E Van Don, V Dakos, R Adrian. 2016. Evaluating early-warning indicators of critical transitions in natural aquatic ecosystems. *Proc. Natl. Acad. Sci.*, 113 (50): E8089-E8095.
- Guttal, V, and C Jayaprakash. 2008. Changing skewness: an early warning signal of regime shifts in ecosystems. *Ecol. Lett.* 11: 450–460.
- Hampton, SE, CA Strasser, JJ Tewksbury, WK Gram, AE Budden, AL Batcheller, CS Duke, JH Porter. 2013. Big data and the future of ecology. *Front. Ecol. Environ.* 11(3): 156 -162.

- Hastings, A and DB Wysham. 2010. Regime shifts in ecological systems can occur with no warning. *Ecology Letters*. 13: 464-472.
- Hijmans, RJ 2019. geosphere: Spherical Trigonometry. R package version 1.5-10.
<https://CRAN.R-project.org/package=geosphere>
- Holm-Hansen, O. 1978. Chlorophyll a determination: improvements in methodology. *Oikos* 30: 438-447.
- Holmes, EE, EJ Ward, and K Wills. 2012. MARSS: Multivariate Autoregressive State-space Models for Analyzing Time-series Data. *The R Journal*. 4(1):11-19
- Holmes, E, E Ward, and K Wills. 2018. MARSS: Multivariate Autoregressive State-Space Modeling. R package version 3.10.10.
- Isles, PDF, F Pomati. 2021. An operational framework for defining and forecasting phytoplankton blooms. *Front. Ecol. Environ.* doi:10.1002/fee.2376.
- Kislik, C, I Dronova, M Kelly. 2018. UAVs in support of algal bloom research: a review of current applications and future opportunities. *Drones*. 2(4): 35.
- Leavitt, PR, SR Carpenter, JF Kitchell. 1989. Whole-lake experiments: the annual record of fossil pigments and zooplankton. *Limnology and Oceanography*. 34(4): 700-717.
- Lenton, TM, VN Livina, V Dakos, EH van Nes, M Scheffer. 2012. Early warning of tipping points from critical slowing down: comparing methods to improve robustness. *Phil. Trans. R. Soc. A*. 370: 1185 - 1204.
- Litzow, MA, JD Urban, and BJ Laurel. 2008. Increased spatial variance accompanies reorganization of two continental shelf ecosystems. *Ecol. Appl.* 18(6): 1331–1337.
- Meinson, P, A Idrizaj, P Noges, T Notes, A Laas. 2016. Continuous and high-frequency measurements in limnology: history, applications, and future challenges. *Environ. Rev.* 24: 1 - 11.
- Ortiz, D, J Palmer, G Wilkinson. 2020. Detecting changes in statistical indicators of resilience prior to algal blooms in shallow eutrophic lakes. *Ecosphere*. 11(10): e03200.
- Pace, ML, RD Batt, CD Buelo, SR Carpenter, JJ Cole, JT Kurtzweil, and GM Wilkinson. 2017. Reversal of a cyanobacterial bloom in response to early warnings. *Proc. Nat. Acad. Sci. U.S.A.* 114(2): 352-357.
- Paradis, E and K Schliep. 2019. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* 35: 526-528.
- Perretti, CT, SB Munch. 2012. Regime shift indicators fail under noise levels commonly observed in ecological systems. *Ecological Applications*. 22(6): 1772-1779.

- Polunchenko, AS and AG Tartakovsky. 2012. State-of-the-art in sequential change-point detection. *Methodol. Comput. Appl. Probab.* 14: 649-684.
- Ratajczak, Z, S Carpenter, A Ives, K Kucharik, T Ramiadantsoa, AM Stegner, J Williams, J Zhang, and MG Turner. (2018) Abrupt change in ecological systems: diagnosis and inference. *Trends in Ecology and Evolution* 33: 513-526.
- Ratajczak Z, P D'Odorico, JB Nippert, SL Collins, NA Brunsell, S Ravi. 2017. Changes in spatial variance during a grassland to shrubland state transition. *Journal of Ecology.* 105: 750-760.
- Rindi, L, MD Bello, L Dai, J Gore, L Benedetti-Cecchi. 2017. Direct observation of increasing recovery length before collapse of a marine benthic ecosystem. *Nature Ecology & Evolution.* 1: 0153.
- Scheffer, M, J Bascompte, WA Brock, V Brovkin, SR Carpenter, V Dakos, H Held, EH van Nes, M Rietkerk, and G Sugihara. 2009. Early-warning signals for critical transitions. *Nature.* 461: 53–59.
- Scheffer, M, SR Carpenter, V Dakos, E van Nes. 2015. Generic Indicators of Ecological Resilience: Inferring the Chance of a Critical Transition. *Annu. Rev. Ecol. Evol. Cyst.* 46:145–67.
- Scheffer, M, S Carpenter, JA Foley, C Folke, B Walker. 2001. Catastrophic shifts in ecosystems. *Nature* 413, 591–596.
- Serizawa, H, T Amemiya, and K Itoh. 2008. Patchiness in a minimal nutrient-phytoplankton model. *J. Biosci.* 33(3): 391–403.
- Sommer S, KJ van Benthem, D Fontaneto, A Ozgul. 2017. Are generic early-warning signals reliable indicators of population collapse in rotifers? *Hydrobiologia.* 796: 111-120.
- Spears, BM, MN Futter, E Jeppesen, BJ Huser, S Ives, TA Davidson, R Adrian, DG Angeler, SJ Burthe, L Carvalho, F Daunt, AS Gsell, DO Hessen, ABG Janssen, EB Mackay, L May, H Moorhouse, S Olsen, M Søndergaard, H Woods, SJ Thackeray. 2017. Ecological resilience in lakes and the conjunction fallacy. *Nature Ecology and Evolution.* 1(11): 1616–1624.
- Swingle, HS. 1968. Fish kills caused by phytoplankton blooms and their prevention. *FAO (F.A.O.U.N.) Fish. LA Rep.* 44(5): 407–411.
- Welch, EB and GD Cooke. 1999. Effectiveness and longevity of phosphorus inactivation with alum. *Lake and Reservoir Management.* 15: 5 - 27.
- Wilkinson, GM, SR Carpenter, JJ Cole, ML Pace, RD Batt, CD Buelo, and JT Kurtzweil. 2018. Early warning signals precede cyanobacterial blooms in multiple whole-lake experiments. *Ecol. Monogr.* 88(2): 188–203.

Table 3.1. Summary of temporal and spatial EWS during the pre-bloom fertilization period.

	Temporal						Spatial			
	SD			AR(1)			SD		Moran's I	
Variable	True alarms	False alarms	Slope	True alarms	False alarms	Slope	Exp. vs. Ref.	Slope	Exp. vs. Ref.	Slope
Chl-a	3	4	0	0	0	0				
Phyco	7	1	+	0	1	+	≈	0	≈	0
DO sat.	0	0	0	0	0	+	>	0	≈	0
pH	2	0	+	0	0	-	≈	0	≈	+

- Slope denotes if the difference between lakes (experimental lake – reference lake) increased significantly through time (+, $p < 0.05$), decreased (–, $p < 0.05$), or was not significant (0, $p > 0.05$).
- “Exp. vs. Ref.” denotes if the mean spatial statistic was significantly higher (>, $p < 0.05$), lower (<, $p < 0.05$), or not significantly different (≈, $p > 0.05$) in the experimental lake relative to the reference lake.

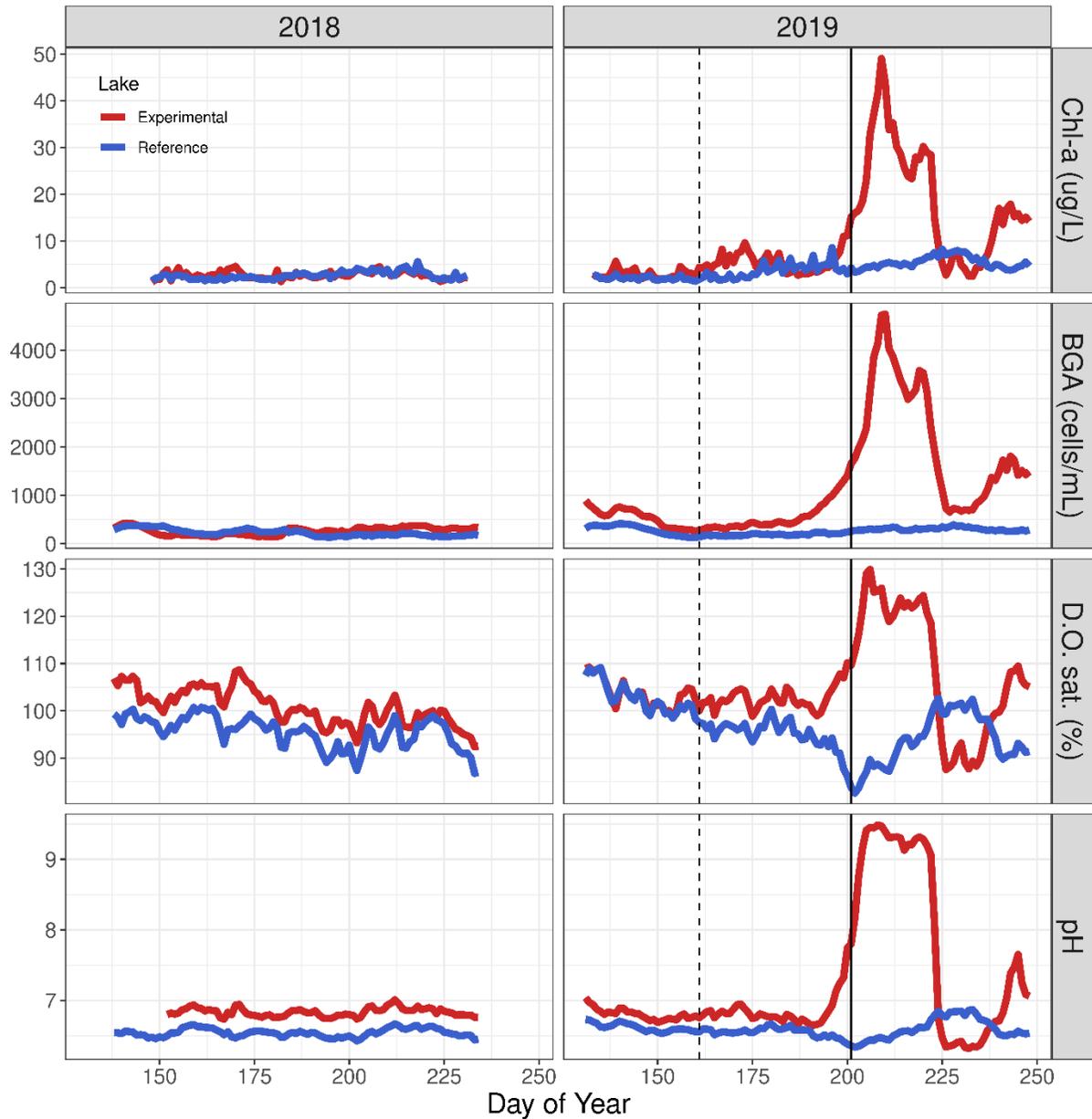


Figure 3.1. Daily time series of ecosystems state variables (chlorophyll, blue-green algae, dissolved oxygen saturation, and pH) measured from a single point at the center of each lake in the non-manipulation (2018) and manipulation (2019) years. Vertical lines represent the start of nutrient additions (dashed line) and bloom onset (solid line) in the experimental lake.

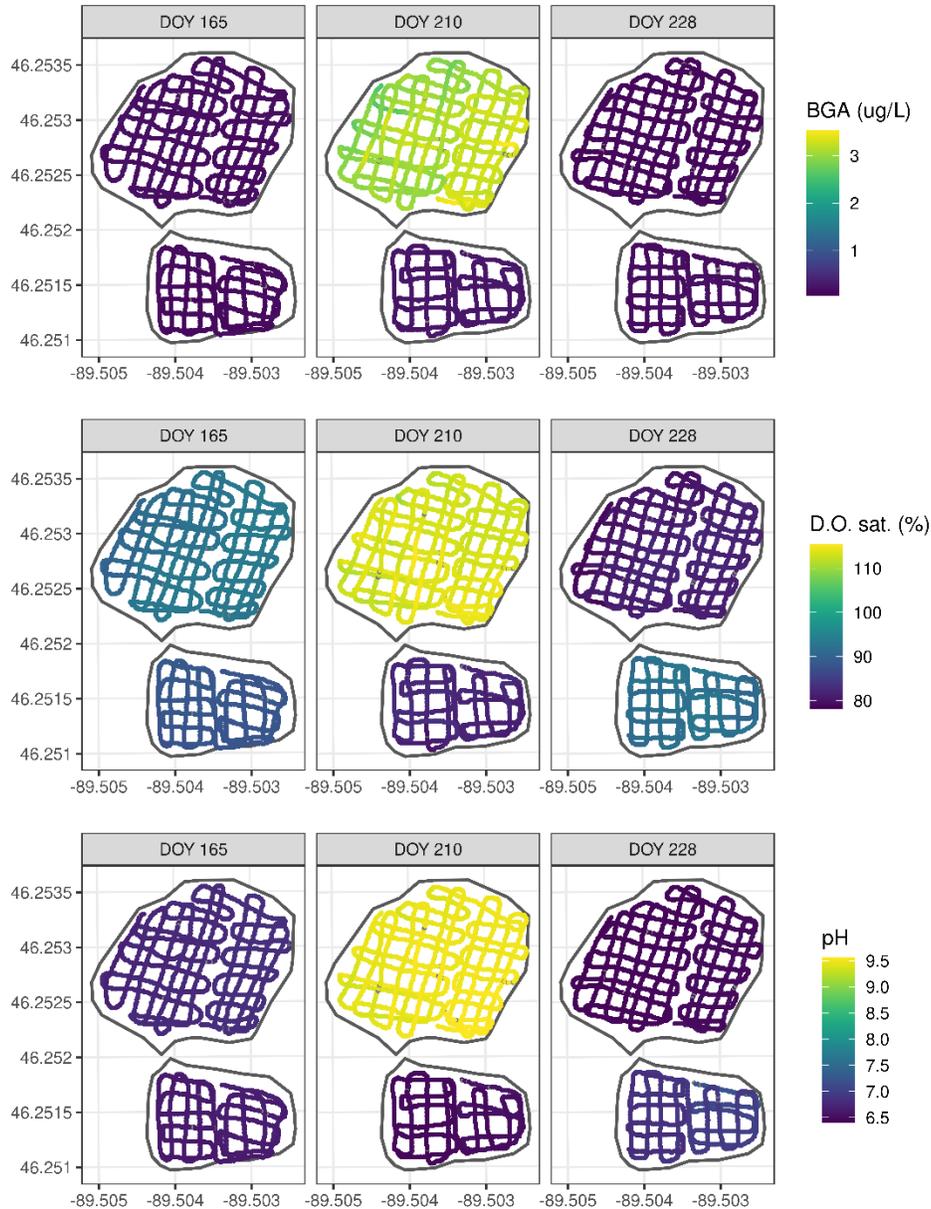


Figure 3.2. Maps showing the spatial distribution of phycocyanin, dissolved oxygen saturation, and pH on three representative dates: before the bloom in the experimental lake (DOY 165), near bloom peak (DOY 210), and after bloom crash (DOY 228). The top (northern) basin is the experimental lake, Peter Lake, and the bottom (southern) basin is the reference lake, Paul Lake.

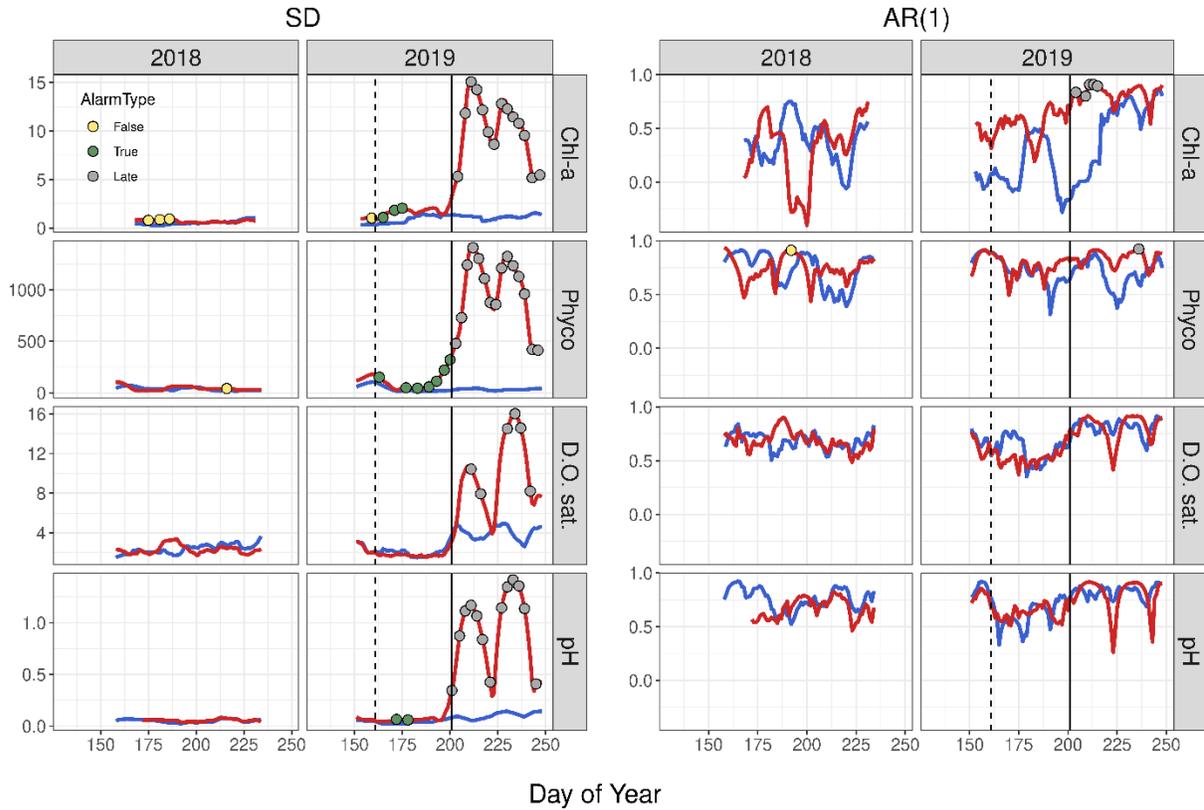


Figure 3.3. Temporal early warning statistics: rolling window standard deviation (SD) and lag-1 autocorrelation (AR(1)) with a 21-day window width. Points represent quickest detection alarms in the experimental lake; false alarms (yellow) before the start of nutrient additions (dashed vertical line), true alarms (green) after the start of nutrient additions and before bloom onset (solid vertical line), and late alarms (grey) occurred after bloom onset.

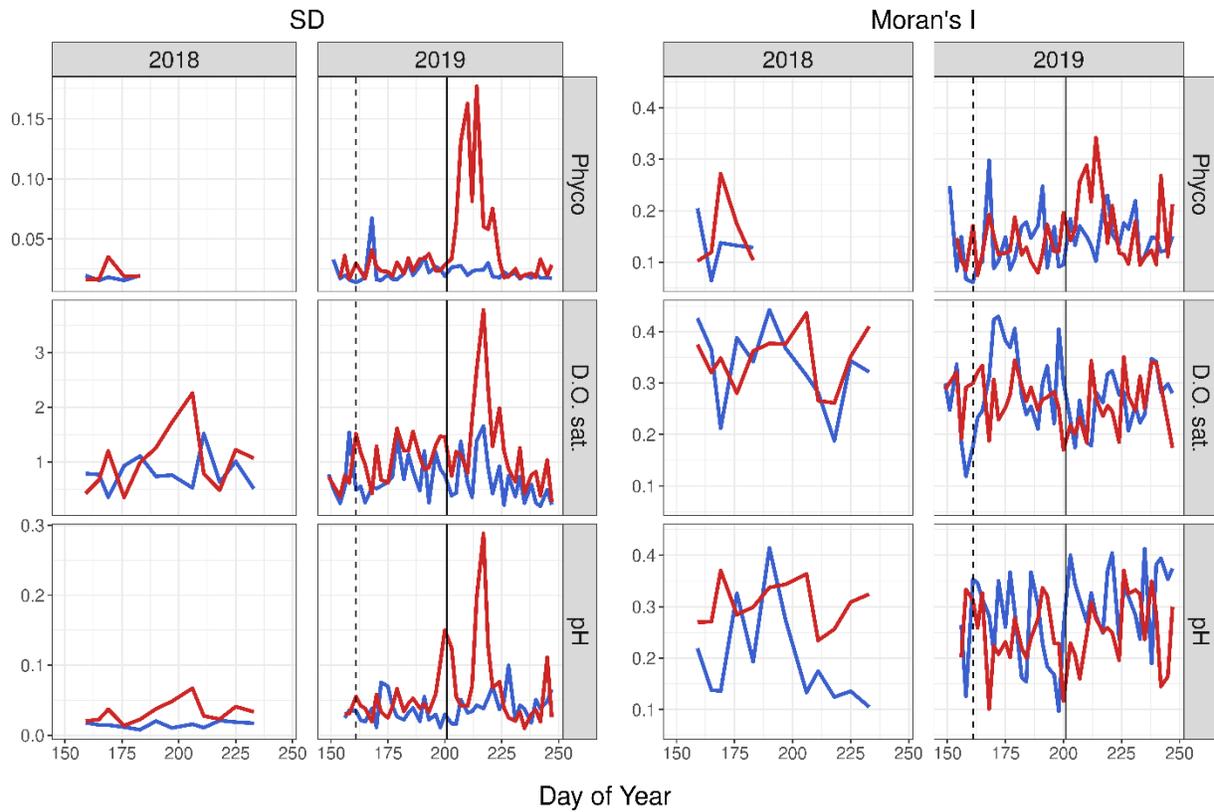


Figure 3.4. Spatial early warning statistics: spatial standard deviation (SD) and autocorrelation (Moran's I) for phycocyanin, dissolved oxygen saturation, and pH for all spatial sampling dates in the experimental (red) and reference (blue) lakes. Vertical lines represent the start of nutrient additions (dashed) and bloom onset (solid). Missing blue-green algae data from 2018 are due to a calibration error.

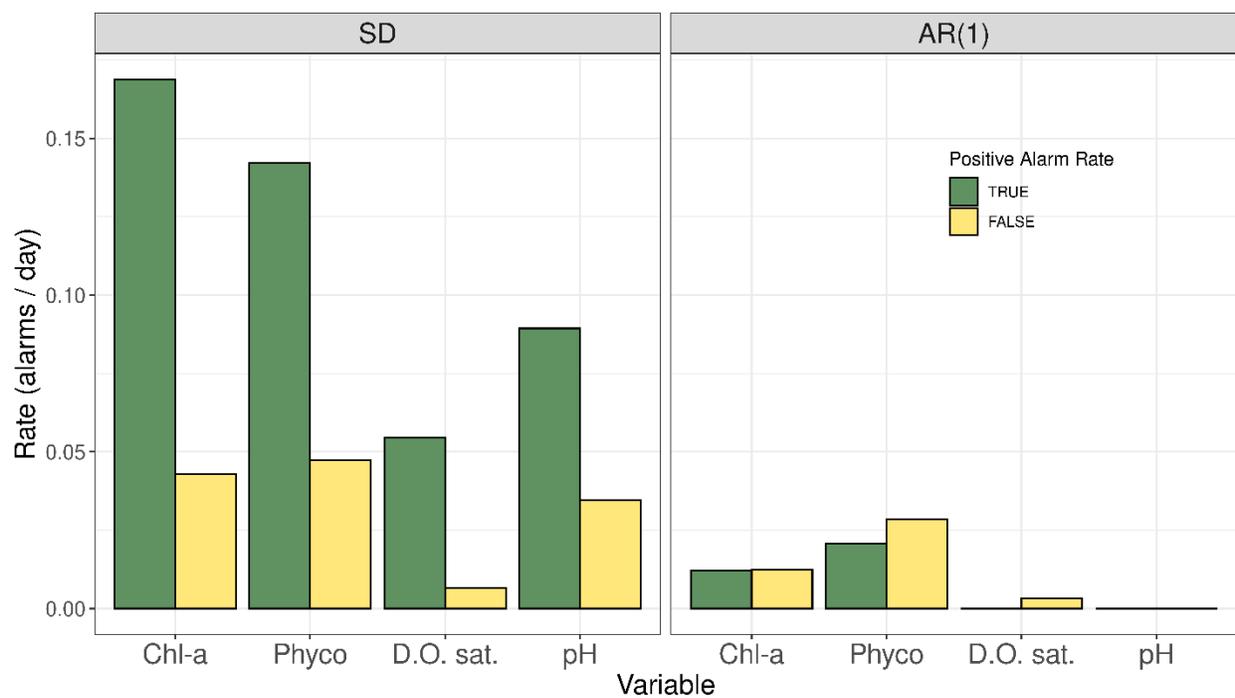


Figure 3.5. Rates of true (green) and false (yellow) quickest detection alarms for rolling window standard deviation (SD) and autocorrelation (AR(1)) across all lake and years for chlorophyll, phycocyanin, dissolved oxygen saturation, and pH.

Chapter 4

Anticipating blooms: exploring the accuracy of algal bloom initiation forecasts

Will be submitted to *Water Resources Research*

Abstract

In many aquatic ecosystems, eutrophication causes algal blooms that have negative effects on ecosystems and the services they provide. The desire to maintain water quality has driven increasing interest in near-term forecasting of algal blooms to enable proactive management. However, adoption of bloom forecasting systems has been limited due largely to the need for system-specific models to accurately forecast blooms. Another important but understudied hurdle to successful adoption is the performance of forecast models to predict bloom initiation timing, which is critical for informing management actions. This period is short and tends to be underweighted in traditional forecast evaluation metrics. In this study, we use high-frequency data from whole-lake fertilization experiments to explore bloom initiation forecast accuracy. A data assimilation model based on the relationship between cumulative nutrient loading and algal biomass was used to assess forecast accuracy, focused specifically on bloom timing. Both initial conditions for model parameters and model flexibility to iteratively adjust parameters based on observations were important to the accurate prediction of bloom initiation. Inaccurate initial conditions, which could arise from unreliable prior estimates or changes in the bloom-driver response relationship, can be compensated for by increased model flexibility. However, too much flexibility can also result in fitted parameters from short-term trends that lead to inaccurate forecasts. The importance of bloom initiation to proactive management and potentially conflicting conclusions when model evaluation is based on

forecasting overall accuracy vs. bloom timing points to the need to explicitly consider this metric in model development and application.

Introduction

Algal blooms are both a natural phenomenon and a consequence of eutrophication (Heisler et al. 2008). Some algal blooms, particularly those related to eutrophication, have negative consequences and are designated Harmful Algal Blooms, or HABs. Because of needs to manage water supplies, protect aquatic resources like fish stocks (Swingle 1968), and assure human health (Carmichael 2001), there is increasing interest in forecasting HABs. Accurate forecasts allow management actions such as beach and shellfish bed closures (Trainer et al 2007), application of algicides (McKnight et al. 1983), alternate uses of water bodies, and changes in water treatment.

In a systematic literature review of bloom forecasting models that focused on cyanobacteria, Rouso et al. (2020) found a wide range of model types, inputs, and outputs. Models tend to be either “process based” (explicitly quantifying known relationships in mathematical equations) or “data-driven” (algorithms and statistics that use observed patterns to create anticipating rules). The heterogeneity and site-specificity of bloom forecasting models has likely limited adoption by water resource managers. Additionally, the multitude of factors that play a role and interact to either promote or suppress individual blooms (Paerl et al. 2001) complicates model application.

Performance is another component of bloom forecasting models that is important to their potential adoption and application for management. Models are most often evaluated based on comparisons of predicted vs. observed values, for example with statistics such as correlation r ,

coefficient of determination R^2 , or root mean square error, RMSE (Rousso et al. 2020). The timing of blooms relative to forecasts, especially during the early stages of bloom growth, is a critical component of forecasting model performance, because management interventions are most effective prior to or early in blooms. While some studies qualitatively examine the timing of blooms in model forecasts relative to observations (e.g. Coad et al. 2014, Recknagel et al. 2017), and metrics like r , R^2 or RMSE also capture errors in bloom timing, quantitative assessment of bloom initiation is often underweighted in model performance evaluation.

Accurately forecasting bloom initiation is an inherently difficult task. It requires models that incorporate the response of algae to drivers, adapt to changing conditions that alter those responses, and have capability to know or predict drivers over relevant time scales. High frequency measurements of both algal abundance and driver variables are helpful to accurately capture the early stages of blooms and aid in evaluating this aspect of bloom forecasting. In this study, we use high-frequency data from a set of whole-lake fertilization experiments to identify key driver(s) of bloom initiation, which are then used in a forecasting model to explore the dependence of forecast accuracy of bloom initiation on parameter knowledge and model flexibility.

Methods

Field Experiments – Data from a series of whole-lake fertilization experiments were used to develop a model for forecasting phytoplankton biomass and test its performance. The manipulations were carried out at the University of Notre Dame Environmental Research Center in the upper peninsula of Michigan, USA. Two lakes were fertilized: Peter Lake and Tuesday Lake from 2013 to 2015, and Peter Lake in 2019. Ammonium nitrate and phosphoric acid were added to each lake daily during the summer while the lakes were stratified by pumping stock

solution from a carboy into the propwash of a small boat propelled by an electric motor. Full description of the experiments including nutrient loading regimes are available in Pace et al. 2017, Wilkinson et al. 2018, and Buelo et al. (In Review). Briefly, N:P ratios of nutrient additions were relatively low (5:1 – 15:1 molar ratio) to encourage cyanobacterial blooms; in 2013 and 2019 nutrient addition loading rates were started at low levels and increased weekly while in 2014 and 2015 constant moderate loading rates were used.

The lakes are small (Peter Lake 2.4 ha, Tuesday Lake 0.9 ha), oligo- to meso- trophic, and located in relatively undisturbed watersheds. They have been extensively studied and do not undergo large algal blooms in the absence of nutrient additions (Pace et al. 2019). Tuesday Lake is a bog lake with darker water (mean color based on light absorbance at 440 nm = 5.2 m⁻¹) due to high dissolved organic matter inputs from surrounding peatlands, while Peter Lake's watershed is more forested and its water more clear (mean color = 2.1 m⁻¹). Neither lake has significant inflow or outflows, though Peter Lake receives water from immediately adjacent Paul Lake via a culvert during high water conditions.

Measured Variables – Water samples for chlorophyll-a were collected daily at the center of the lake at a depth of 0.5m. Samples were filtered onto Whatman 47mm GF/F filters, extracted in methanol, and analyzed on a Turner Trilogy benchtop fluorometer (Holm-Hansen 1978). Additional water quality parameters (phycocyanin, dissolved oxygen, pH) and temperature were measured *in-situ* at 0.75m every 5 minutes using automated sensors. Water temperature was also measured every 5 minutes with a string of thermistors every 0.5m from 0.5 to 5 meters below the surface; these profiles were used to calculate mixed layer depth (Pace et al. 2021). Meteorological conditions (air temperature, wind speed, photosynthetically active radiation [PAR]) were measured every 5 minutes from a raft on Peter Lake. Additional potential

bloom drivers and other limnological variables including light extinction profiles, water color (g440), nutrient concentrations (TN and TP), dissolved oxygen profiles, zooplankton abundance were measured weekly as described in Wilkinson et al. (2018).

Forecast Model – To develop the forecasting model, exploratory data analysis was done graphically to identify driver(s) related to bloom timing. Cumulative phosphorus loading normalized by mean annual water color was closely related to chlorophyll during large blooms (see *Results* below) and this relationship was implemented in an Ensemble Kalman Filter (EnKF) model to predict chlorophyll concentrations and the bloom timing. EnKF is a data assimilation technique that uses a predictive model and observations to iteratively estimate system state, which can be used with forecasted (or in our case, known) future model driver values to predict future states. We use a variant of EnKF that employs parameter augmentation to also fit unobserved parameters of the predictive model to best match the observations (Zhang et al. 2017). Full model description is provided in Supporting Information Appendix 3, we focus here on the most relevant EnKF details. The predictive model $F()$ is a logistic growth type model where the carrying capacity is dynamic and changes through time based on the color-normalized cumulative phosphorus (CNCP) loading and three parameters that are also fit iteratively as observations are assimilated:

$$Chl_{i,t+1} = F(Chl_{i,t}, L_{t+1}) = Chl_{i,t} * [1 + \left(1 - \frac{Chl_{i,t}}{K_{i,t+1}}\right)] \quad (1)$$

$$K_{i,t+1} = B_{i,t} + A_{i,t} * \frac{L_{t+1}^k}{Lh_{i,t}^k + L_{t+1}^k} \quad (2)$$

Where i denotes a single ensemble member, Chl_t is chlorophyll at time t , K_t is the carrying capacity, A_t determines the asymptotic chlorophyll concentration at high loading values, L_t is the CNCP loading, Lh_t is the half-loading constant, B_t is the baseline chlorophyll concentration at low loading values, and k is a fixed, lake-specific constant determining how steep the increase in

Chl is at intermediate L (Table 4.1). $Chl_{i,t}$ is the EnKF state variable corresponding to the observed chlorophyll measurements, while A_t , Lh_t , and B_t are parameters augmented to the EnKF state matrix. State and parameters were initialized using draws from uniform distributions, see Appendix S1 for details.

Kalman filters assimilate data in two steps: first the forecast step, which uses the previous state of the system and the predictive model to forecast the state of the system at the current time step. In the update step, the difference between the predicted and observed state of the system is used to adjust the system state (including parameters) to better match the observations. To prevent filter divergence, when small parameter variance causes the ensemble to be overconfident in its forecasts and ignore observations, ensemble parameter inflation is applied using an inflation factor (IF) during the forecast step:

$$\alpha_{j,t}^{i-} = \overline{a_{j,t-1}} + IF(\alpha_{j,t-1}^i - \overline{a_{j,t-1}}) \quad (3)$$

where $a_{j,t}$ is the value of parameter j at time t , $i-$ denotes ensemble member i at the forecast step, i denotes ensemble member i at the update step, and $\overline{a_{j,t-1}}$ is the mean of parameter j across ensemble members after the update step at time $t - 1$. There are several methods for setting IF (Anderson 2007, Whitaker & Hamill 2012), we use a common one of setting IF to a constant value slightly larger than 1 (Evensen 2009).

Forecast Experiments – The ability of the model to accurately predict bloom timing was tested by varying both the initial values of the predictive model parameters and also the EnKF inflation factor IF . While all three parameters (A , Lh , B) of the predictive model $F()$ that determine the carrying capacity K contribute to the shape of the chlorophyll vs loading curve (Figure 4.1), we focus on Lh due to its importance in determining the cumulative loading values at which a rapid increase in chlorophyll occurs. Forecasting experiments with initial Lh values

(Lh_{t0}) starting at a range of values centered at the “correct” (fit) value were conducted to assess how the model might perform under more realistic scenarios, where Lh estimates from prior blooms are either inaccurate or not constant from bloom to bloom.

Forecast experiments were also carried out across a range of inflation factors IF from 1.001 to 1.025. Low values of IF lead to slow changes to parameter values as model predictions are compared to observations. Larger values of IF create more spread in parameter values across ensemble members, which in practice allows parameter values to change more quickly as observations are assimilated.

One hundred replicates for each combination of Lh_{t0} and IF of the EnKF model were carried out at a daily time step: for day t , first predictions were made for each ensemble member using the updated states and parameters from day $t-1$ and equations (1) and (2) with the cumulative loading on day t . Then the update step adjusted the state and parameters based on the difference between the predictions and the observed chlorophyll on day t . Finally, the updated states and parameters were used to iteratively forecast future states for days $t+1$ to $t+7$ using the known future CNCP loading; in non-experimental settings future loading would also need to be forecast. This process was repeated for each day to the end of the time series for each lake-year.

Seven-day forecasts were analyzed for errors in bloom timing; three and five day forecast horizon results are presented in Appendix 3 (Figures A3.1, A3.2). Bloom forecast timing error for each replicate was determined by calculating the difference between when the ensemble mean 7-day forecast and when the observed chlorophyll concentration first crossed a lake-specific bloom threshold determined from historical data (Wilkinson et al. 2018). A timing error of 0 means the 7-day forecast correctly predicted the day of bloom initiation; negative errors indicate the bloom prediction was too early while positive errors indicate the bloom prediction

was too late (though errors up to +6 days could still provide some warning depending on how quickly management action is taken). The 100 replicates for each Lh_{i0} , IF combination were averaged to determine the mean bloom timing error.

Results

Large blooms (> 40 ug/L chl-a) occurred in 3 out of 7 lake-years: Peter Lake in 2015 and 2019 and Tuesday Lake in 2015. For these years, chlorophyll concentrations up until the bloom peak displayed a threshold relationship with cumulative phosphorus loading, with the increase in chlorophyll in Tuesday Lake occurring at much higher cumulative P loading values (Figure 4.1a). Normalizing cumulative P loading by water color (g440) brought the responses of the lakes much closer together (Figure 4.1b), though there were still lake-specific differences in the sigmoid constant k (approximately 14 for Tuesday Lake and 7 for Peter Lake) as well as the loading half constant (approximately 60 for Tuesday Lake and 50 for Peter Lake) and asymptotic and baseline chlorophyll concentrations. Analysis of additional bloom drivers suggested other mechanisms that may have played a role limiting bloom development (Appendix 3, Figure A3.3), though the limited number of lake-years available precluded development of more complex predictive models with mechanisms explaining bloom occurrence and bloom die offs across all years. As our interest is in exploring limits to forecasting bloom timing, we limit our model analyses to the years with the largest bloom in each lake. These years were Peter Lake 2019 and Tuesday Lake 2015 when in both cases nutrients were applied continuously over the summer and large sustained blooms developed (Wilkinson et al 2018; Buelo et al. In Review).

The EnKF model successfully captured observed system dynamics. Figure 4.2 shows the evolution of model state and parameters for the 2015 Tuesday Lake bloom (Lh started at the “best estimate” value from Figure 4.1b and IF set to 1.02). The mean ensemble state estimate of

chlorophyll closely follows observations; the predictive model's instantaneous estimate of the carrying capacity (K_t) from the fit parameters A_t , Lh_t , and B_t generally does too although was not as variable from day-to-day (Figure 4.2a). The asymptotic chlorophyll parameter A_t remains steady for most of the time series before oscillating as the bloom approaches its peak (Figure 4.2b). The loading half constant Lh_t increases from day 55 to 70 before quickly declining around day 75 when chlorophyll starts to climb consistently (Figure 4.2c). The baseline chlorophyll-a constant B_t increases slightly and remains elevated from approximately day 10 to day 50 (Figure 4.2d).

Forecasts were generally accurate up until bloom peak, as expected based on the design of the model. Figure 4.3 shows 7-day forecasts for both Tuesday Lake 2015 and Peter Lake 2019 as well as model error (root mean square error, RMSE) for all forecasts for horizons from 1 to 7 days, with $IF = 1.02$ and parameters started at the lake-specific fit values from Figure 4.1. In Tuesday Lake, the initial ramp of the bloom before day 80 is missed by the 7-day forecast due to the higher value of Lh_t up until approximately day 75, but the forecasts generally match observations for the remainder of the bloom (Figure 4.3a). The 7-day forecasts in Peter Lake closely track observations up until the bloom peak, but then severely overestimate chlorophyll concentrations after the bloom collapses (Figure 4.3b). Overall, EnKF forecast error is lower in Tuesday than a persistence null model (future forecasts are the same as the last observation) for all forecast horizons, especially for 5- to 7-day forecasts (Figure 4.3c). The EnKF forecast error is higher than the null model for all horizons in Peter Lake (Figure 4.3d) due to poor forecasts after the bloom peaked and subsequently declined. This is due the strictly increasing relationship between CNCP loading and chlorophyll concentrations (Figure 4.1) used for forecasting, which

did not match the bloom collapse dynamics observed in the latter part the summer in Peter Lake but did agree with the late-occurring bloom in Tuesday Lake.

Varying the parameter inflation factor IF had an impact on parameter evolution and resulting forecasts, even for identical values of Lh_{t0} . Figure 4.4 illustrates Lh_t initialized at the “best” value ($\sim 58 \text{ mg/m}^2/\text{g440}$) and also for a low value ($47.5 \text{ mg/m}^2/\text{g440}$), for inflation factors of both 1.001 and 1.02. Prior to bloom initiation around day 70, observed chlorophyll declines slightly for most of the prior 25 days despite cumulative loading increasing. Ensemble members with higher Lh_t values better match these observations, so Lh_t increases over this period. However, the larger IF value (1.02) generates more variation in the ensemble, and Lh_t increases faster than for $IF = 1.001$ (Figure 4.4a,b) up to maximum values around day 70. The increases for both starting IF values overshoot the “best” Lh_t value (Figure 4.4a), as does the higher IF value for the low Lh_{t0} case (Figure 4.4b). As a result, forecasts for these cases miss the early stage of bloom initiation from day 75 – 80 (Figure 4.4c,d). The slower increase in Lh_t for the low Lh_{t0} case with $IF = 1.001$ overestimates chlorophyll for forecasts from day 60 – 75 but better captures the bloom ramp up from day 75 – 80 (Figure 4.4d). After day 80, forecasts for all cases largely converge as Lh ensemble variability decreases sharply.

Analysis of 100 replicate simulations for each combination of initial loading half constant Lh_{t0} and parameter inflation factor IF reveal large variations in the error of 7-day forecast bloom timing predictions. The range of these forecasts is from 22 days too early to over 13 days too late (Figure 4.5). For Tuesday Lake, low inflation factors very close to 1 have a very small range of Lh_{t0} values that lead to bloom timing errors close to zero; for example for $IF = 1.002$, an increase in Lh_{t0} from 45 to 55 changes the mean forecasting error from approximately 9 days too early to approximately 4 days too late. More moderate IF values (~ 1.01 to 1.025) generally have positive

but small bloom timing error over a wide range of Lh_{t0} values. Peter Lake has larger ranges of Lh_{t0} values with bloom timing errors near zero for a given IF value, and smaller errors overall (Figure 4.5b).

Discussion

Our findings suggest that bloom timing can be forecast relatively accurately at horizons potentially useful for management, but that the consistency of bloom response to drivers and model flexibility is critical. Increasing the ensemble Kalman filter's parameter inflation factor increases the model's ability to correct from inaccurate starting parameters, but also can lead to the model fitting parameters from short-term trends that cause inaccurate predictions of bloom timing. Additionally, the more strongly nonlinear bloom initiation in Tuesday Lake was more difficult to forecast.

The steady increase in chlorophyll past thresholds in cumulative loading (Figure 4.1a) across the large blooms suggests total (as opposed to instantaneous) loading and is a crucial determinant of bloom timing for these lakes. The relationship could arise from several possible mechanisms of bloom initiation: saturating benthic and other non-phytoplankton nutrient sinks (Vadeboncoeur & Steinman 2002), overcoming grazing pressure (Carpenter et al. 2001), and shifting the phytoplankton community to bloom dominating taxa (Brauer et al 2012) prior to rapid bloom growth. In non-experimental systems where nutrient loading is more sporadic, cumulative loading over a set time period (as opposed to total annual loading) may perform better. That normalizing cumulative P load by water color increased response similarity between lakes (Figure 4.1b) suggests that nutrient interactions with light availability and possibly dissolved organic matter uptake/release are important, consistent with other studies (Olson et al.

2020), but additional experiments would be needed to definitively identify the mechanism(s) involved.

While our study does not represent a true forecasting scenario because it applied the forecasting model to the same data it was developed from, it illustrates both the challenge of accurately forecasting bloom timing and the potential of data assimilation methods like the ensemble Kalman filter. Identifying which drivers control bloom initiation is critical but difficult. Even in these relatively simple lakes where the key bloom driver was experimentally controlled, we were unable to identify other variable(s) that concretely separated years with large blooms from years with similar nutrient loads that had only moderate increases in phytoplankton (Wilkinson et al. 2018; Appendix 3 Figure A3.3). Doing so in larger and more complex systems is likely to be even more challenging, though machine learning methods have shown promise for identifying bloom drivers and thresholds in complex datasets (Isles et al 2017). A recently proposed framework by Isles and Pomati (2021) defines a bloom classification system based on bloom dynamics and causes and suggests modeling approaches for forecasting each type; such classifications will likely also be helpful for identifying which drivers are important and useful methods when focused specifically on bloom timing. Our model performs poorly at predicting chlorophyll concentrations after the bloom peak (Figure 4.3); similar findings have sometimes been observed in more complex forecasting models as well (Page et al. 2017). Models that include more drivers will be more flexible in the predicted driver(s)-bloom response relationship, but also will require more data to fit parameters and can lead to less accurate forecasts due to overfitting (Cawley and Talbot 2010).

In addition to identifying important drivers and how they relate to bloom development, forecasting models need to correctly parameterize the bloom-driver relationship either *a priori* or

dynamically. If controls of bloom initiation are constant from bloom-to-bloom, it may be possible to establish fixed model parameters for forecasting. More often, bloom responses are likely to vary due to stochasticity and interactions between the many drivers not included in forecasting models. Data assimilation methods like the ensemble Kalman filter with parameter augmentation can iteratively fit parameter values that best match the data, potentially leading to more accurate forecasts. We evaluated our model's ability to correctly fit parameters qualitatively by comparing fit carrying capacity to chlorophyll values. The calculated K_t values generally tracked chlorophyll observations up until bloom peak (Figure 4.2a), though with periods of consistent under- and over-estimation. Process error (reflecting error in the model's predictions of chlorophyll in the forecast EnKF step) limited more dynamic updates to parameter values that would have led to K_t more closely tracking observations. Changes in individual model parameter values occurred during the expected time periods based on cumulative loading. B_t is the main determinant of carrying capacity K_t at low nutrient loading; consequently, B_t increased early in the experiment when observed chlorophyll is consistently higher than the calculated carrying capacity (Figure 4.2a). Lh_t has a strong control on K_t at moderate cumulative loads, and differences between the observed chlorophyll and K_t just before and after the start of bloom initiation drive directional changes in Lh_t (Figure 4.2c). Parameter variability of the ensemble around mean values was generally maintained throughout the experiment except for a sharp decrease in ensemble variability of Lh_t during the bloom initiation period, underscoring the importance of this parameter to bloom initiation.

Another option for assessing model ability to correctly fit parameters is using model simulations where “true” parameter values are known and running the data assimilation model on synthetic time series generated from simulation outputs (Zwart et al. 2019). Where parameters

represent physical characteristics (e.g. sediment temperature in a water temperature forecasting model, Thomas et al. 2020), measurements could be used to confirm model fits, though this becomes more challenging when parameters represent more complex processes (e.g. hypo- and epi-limnetic oxygen demand, Carey et al 2021; DOC turnover rate or partitioning between different pools, Zwart et al. 2019). In the case of algal blooms, some parameters maybe difficult or impossible to measure in ways that realistically match field conditions (e.g., parameters that make up carrying capacity in this study), while checking other possible parameters may be feasible (e.g., the partitioning of total chlorophyll by different phytoplankton species or functional groups).

The degree of model flexibility to fit dynamic model parameters interacted with parameter initial condition values and observed chlorophyll dynamics to determine bloom timing forecasts. In Tuesday Lake 2015, IF values greater than ~ 1.01 led to similar forecast errors for a wide range of Lh_{t0} values, reflecting model flexibility to learn from assimilated data when started at non-optimal parameter values (Figure 4.5). Flexibility can also cause the model to fit “wrong” parameter values, for example elevated Lh_t values just prior to bloom initiation resulting from steady declines in chlorophyll (Figure 4.4) and led to late predictions of bloom timing for most Lh_{t0} , IF parameter combinations. Bloom timing errors near 0 for the Tuesday Lake 2015 bloom only occurred at low inflation factors for a narrow band of Lh_{t0} values less than the optimal value fit from all the data (Figure 4.1), which we interpret as coincidental matching of Lh_{t0} values and the IF -determined rate at which Lh_t changes.

Despite the relatively narrow ranges of parameters that led to forecast timing errors near 0, there was a much wider range of Lh_t and IF values that produced forecasts that would produce some warning for Tuesday in 2015. For example, a 7-day forecast with a bloom timing error of

+4 days would still provide three days for action to be taken prior to chlorophyll crossing the bloom threshold. Negative timing errors could also still be useful, though at some point raise the question of false positives as opposed to true prediction of an impending bloom. Peter Lake in 2019 had a wider range of Lh_{t0} values that lead to bloom timing errors near 0 for a given IF value, reflecting the less severe nonlinearity of the chl-CNCP loading response in this lake. The dependence of bloom timing error on IF suggests this EnKF model parameter can be an important determinant of model performance that should be evaluated. Some forecasting studies using EnKF have fit IF dynamically (Zwart et al. 2019) or consider a few different values (Page et al. 2018), but often only a single value is typically considered or presented.

The specification of a bloom threshold is necessary to determine bloom timing forecast error. There is no universal definition of bloom (Isles and Pomati 2021), and while extensive historical and prior studies allowed us to define a threshold for these lakes (Wilkinson et al. 2018), doing so in most systems will be challenging. Lower thresholds closer to baseline chl-a concentrations would provide less time for model parameters (especially Lh_t) to be adjusted as observations are assimilated, leading to more positive bloom timing forecast errors. Conversely, higher thresholds would provide more time for parameter adjustment and move forecast timing errors closer to zero.

While we found that parameter initial conditions and IF contributed strongly to bloom forecasts and forecast error, this may not be the case for all forecasting models and scenarios. Uncertainty partitioning can be used to quantify the relative importance of parameter values to forecast uncertainty (Dietze et al. 2017). For example, Thomas et al. (2020) found that parameter and initial condition uncertainty were much less important than process and driver uncertainty for forecasting reservoir water temperatures. Similar analyses could be carried out on bloom

forecasts, though several bloom instances would be needed to robustly partition uncertainty specifically during bloom initiation periods.

Focusing on bloom timing as opposed to overall forecast accuracy is an important distinction. Our findings demonstrate that which metric is being used to evaluate forecasts can lead to conflicting conclusions. Comparing RMSE of all forecasts would suggest our model performs worse for Peter Lake than for Tuesday Lake due to inaccurate forecasts after the bloom collapsed, whereas Peter Lake had smaller errors in bloom timing forecasts. The period when algal biomass is increasing presents the best opportunity to control blooms and minimize or avoid negative consequences. Bloom suppression methods like treatment with algicides are most effective when applied during the early stages of blooms (Burch et al. 2021), which can also minimize the release of toxins when algal cells are lysed (Tsai 2015). Overall, where bloom management is a primary goal for forecasting systems, bloom timing should be a central consideration in forecast model development and evaluation.

Acknowledgements

This work was supported by a Presidential Fellowship from the University of Virginia School of Data Science and the National Science Foundation under grants DGE 1315231, DEB 1754712, DEB 1753854 and DEB 1440297. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

References

- Anderson, JL. 2007. An adaptive covariance inflation error correction algorithm for ensemble filters. *Tellus A: Dynamic Meteorology and Oceanography*. 59(2): 210-224.
- Brauer, VS, M Stomp, J Huisman. 2012. The Nutrient-Load Hypothesis: Patterns of Resource Limitation and Community Structure Driven by Competition for Nutrients and Light. *American Naturalist*. 179(6): 721-740.

- Burch, M, J Brookes, I Chorus. 2021. Assessing and controlling the risk of cyanobacterial blooms. In: Toxic cyanobacteria in water - Second edition, World Health Organization, ISBN: 978-1-003-08144-9.
- Buelo, CD, ML Pace, SR Carpenter, EH Stanley, DA Ortiz, DT Ha. *In Review at Ecological Applications*. Evaluating the performance of temporal and spatial early warning statistics of algal blooms.
- Carey, CC, WM Woelmer, ME Lofton, RJ Figueiredo, BJ Bookout, RS Corrigan, W Daneshmand, AG Hounshell, DW Howard, ASL Lewis, RP McClure, HL Wander, NK Ward, RQ Thomas. 2021. Advancing lake and reservoir water quality management with near-term, iterative ecological forecasting. *Inland Waters*. doi: 10.1080/20442041.2020.1816421.
- Carmichael, WW. 2001. Health Effects of Toxin-Producing Cyanobacteria: “The CyanoHABs”, Human and Ecological Risk Assessment: An International Journal, 7(5): 1393-1407.
- Carpenter, SR, JJ Cole, JR Hodgson, JF Kitchell, ML Pace, D Bade, KL Cottingham, TE Essington, JN Houser, DE Schindler. 2001. Trophic cascades, nutrients, and lake productivity: whole-lake experiments. *Ecological Monographs*: 71(2): 163-186.
- Cawley, GC, NLC Talbot. 2010. On Over-fitting in Model Selection and Subsequent Selection Bias in Performance Evaluation. *Journal of Machine Learning Research*. 11: 2079-2107
- Coad, P, B Cathers, JE Ball, R Kadluczka. 2014. Proactive management of estuarine algal blooms using an automated monitoring buoy coupled with an artificial network. *Environmental Modeling and Software*. 61: 393-409.
- Dietze, MC. 2017. Prediction in ecology: a first-principles framework. *Ecological Applications*. 27(7): 2048-2060.
- Evensen, G. 2009. The ensemble Kalman filter for combined state and parameter estimation. *IEEE Control Systems Magazine*. 29,(3): 83-104.
- Heisler, J, P Glibert, J Burkholder, D Anderson, W Cochlan, W Dennison, C Gobler, Q Dortch, C Heil, E Humphries, A Lewitus, R Magnien, H Marshall, K Sellner, D Stockwell, D Stoecker, M Suddleson. 2008. Eutrophication and harmful algal blooms: a scientific consensus. *Harmful Algae*. 8:3–13.
- Holm-Hansen, O. 1978. Chlorophyll a determination: improvements in methodology. *Oikos* 30: 438-447.
- Isles, PDF, F Pomati. 2021. An operational framework for defining and forecasting phytoplankton blooms. *Front. Ecol. Environ*. doi:10.1002/fee.2376.
- Isles, PDF, DM Rizzo, Y Xu, AW Schroth. 2017. Modeling the drivers of interannual variability in cyanobacterial bloom severity using self-organizing maps and high-frequency data. *Inland Water*. DOI: 10.1080/20442041.2017.1318640.

- McKnight, DM, SW Chisholm, DRF Harleman. 1983. CuSO₄ treatment of nuisance algal blooms in drinking water reservoirs. *Environmental Management*. 7:311-320.
- Olson, CR, CT Solomon, SE Jones. 2020. Shifting limitation of primary production: experimental support for a new model in lake ecosystems. *Ecology Letters*. 23(12): 1800-1808.
- Pace, ML, RD Batt, CD Buelo, SR Carpenter, JJ Cole, JT Kurtzweil, and GM Wilkinson. 2017. Reversal of a cyanobacterial bloom in response to early warnings. *Proc. Nat. Acad. Sci. U.S.A.* 114(2): 352-357.
- Pace, ML, SR Carpenter, GM Wilkinson. 2019. Long-term studies and reproducibility: Lessons from whole-lake experiments. *Limnology and Oceanography*. 64(S1): S22-S33.
- Pace, ML, CD Buelo, SR Carpenter. 2021. Phytoplankton biomass, dissolved organic matter, and temperature drive respiration in whole lake nutrient additions. *Limnology and Oceanography*. 66(6): 2174-2186.
- Paerl, HW, RS Fulton, PH Moisaner, and J Dyble. 2001. Harmful freshwater algal blooms, with an emphasis on cyanobacteria. *Sci. World*. 1:76–113.
- Page, T, PJ Smith, KJ Beven, ID Jones, JA Elliott, SC Maberly, EB Mackay, M De Ville, H. Feuchtmayr. 2017. Constraining uncertainty and process-representation in an algal community lake model using high frequency in-lake observations. *Ecological Modelling*. 357: 1-13.
- Page, T, PJ Smith, KJ Beven, ID Jones, JA Elliott, SC Maberly, EB Mackay, M De Ville, H. Feuchtmayr. 2018. Adaptive forecasting of phytoplankton communities. *Water Research*. 134: 74-85.
- Recknagel, F, PT Orr, M Bartkow, A Swanepoel, H Cao. 2017. Early warning of limit-exceeding concentrations of cyanobacteria and cyanotoxins in drinking water reservoirs by inferential modelling. *Harmful Algae*. 69: 18-27.
- Rouso, BZ, E Bertone, R Stewart, DP Hamilton. 2020. A systematic literature review of forecasting and predictive models for cyanobacteria blooms in freshwater lakes. *Water Research*. 182: 115959.
- Swingle, HS. 1968. Fish kills caused by phytoplankton blooms and their prevention. *FAO (F.A.O.U.N.) Fish. LA Rep.* 44(5): 407–411.
- Thomas, RQ, RJ Figueiredo, V Daneshmand, BJ Bookout, LK Puckett, CC Carey. 2020. A near-term iterative forecasting system successfully predicts reservoir hydrodynamics and partitions uncertainty in real time. *Water Resources Research*, 56, e2019WR026138.
- Trainer, VL, WP Cochlan, A Erickson, BD Bill, FH Cox, JA Borchert, KA Lefebvre. 2007. Recent domoic acid closures of shellfish harvest areas in Washington State inland waterways. *Harmful Algae*. 6(3): 449-459.

- Tsai, KP. 2015. Effects of two copper compounds on *Microcystis aeruginosa* cell density, membrane integrity, and microcystin release. *Ecotoxicology and Environmental Safety*. 120: 428-435.
- Vadeboncoeur, Y, AD Steinman. 2002. Periphyton function in lake ecosystems. *TheScientificWorld*. 2: 1449-1468.
- Whitaker, JS, TM Hamill. 2012. Evaluating Methods to Account for System Errors in Ensemble Data Assimilation. *Monthly Weather Review*. 140: 3078-3089.
- Wilkinson, GM, SR Carpenter, JJ Cole, ML Pace, RD Batt, CD Buelo, and JT Kurtzweil. 2018. Early warning signals precede cyanobacterial blooms in multiple whole-lake experiments. *Ecol. Monogr.* 88(2): 188–203.
- Zhang, H, HJH Franssen, X Han, JA Vrugt, H Vereecken. 2017. State and parameter estimation of two land surface models using the ensemble Kalman filter and the particle filter. *Hydrol. Earth Syst. Sci.* 21: 4927–4958.
- Zwart, JA, O Hararuk, YT Prairie, SE Jones, CT Solomon. 2019. Improving estimates and forecasts of lake carbon dynamics using data assimilation. *Limnology and Oceanography: Methods*. 17: 97-111.

Table 4.1. Model parameters.

Variable / Parameter	Description
Chl _{i,t}	Chlorophyll concentration, state variable
K _{i,t}	Carrying capacity, parameter (eqn 2)
L _t	Color-normalized cumulative P loading, driver variable
A _{i,t}	Asympote chlorophyll parameter, fit by EnKF
L _{hi,t}	Loading half-constant parameter, fit by EnKF
B _{i,t}	Baseline chlorophyll parameter, fit by EnKF

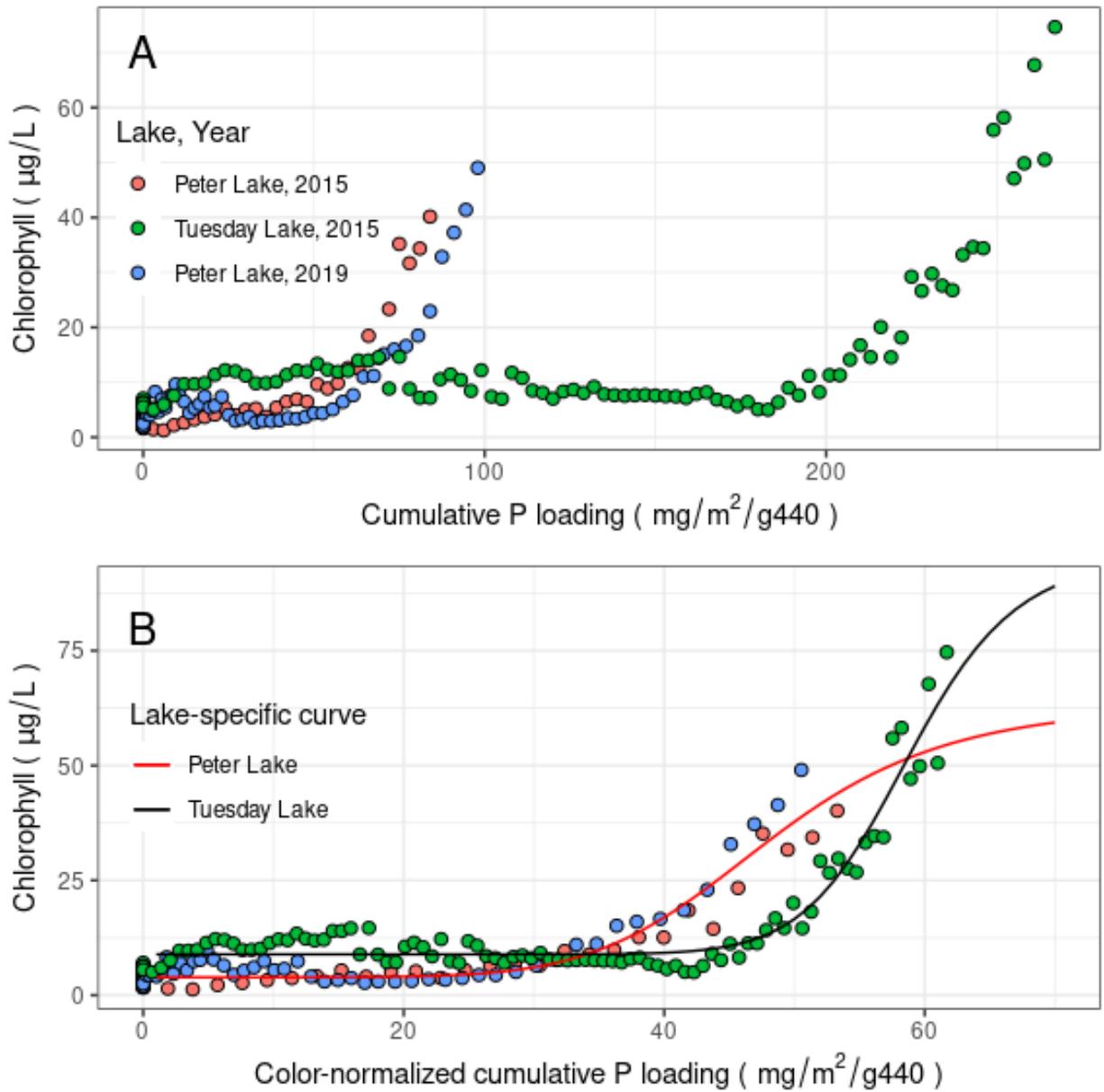


Figure 4.1. Chlorophyll-a concentration up to bloom peak vs. a) cumulative P loading and b) cumulative P-loading normalized by color. Points are daily measured values. Lines are fits to threshold response for each lake. Time series of chlorophyll for all years are in Appendix 3 Figure A3.4.

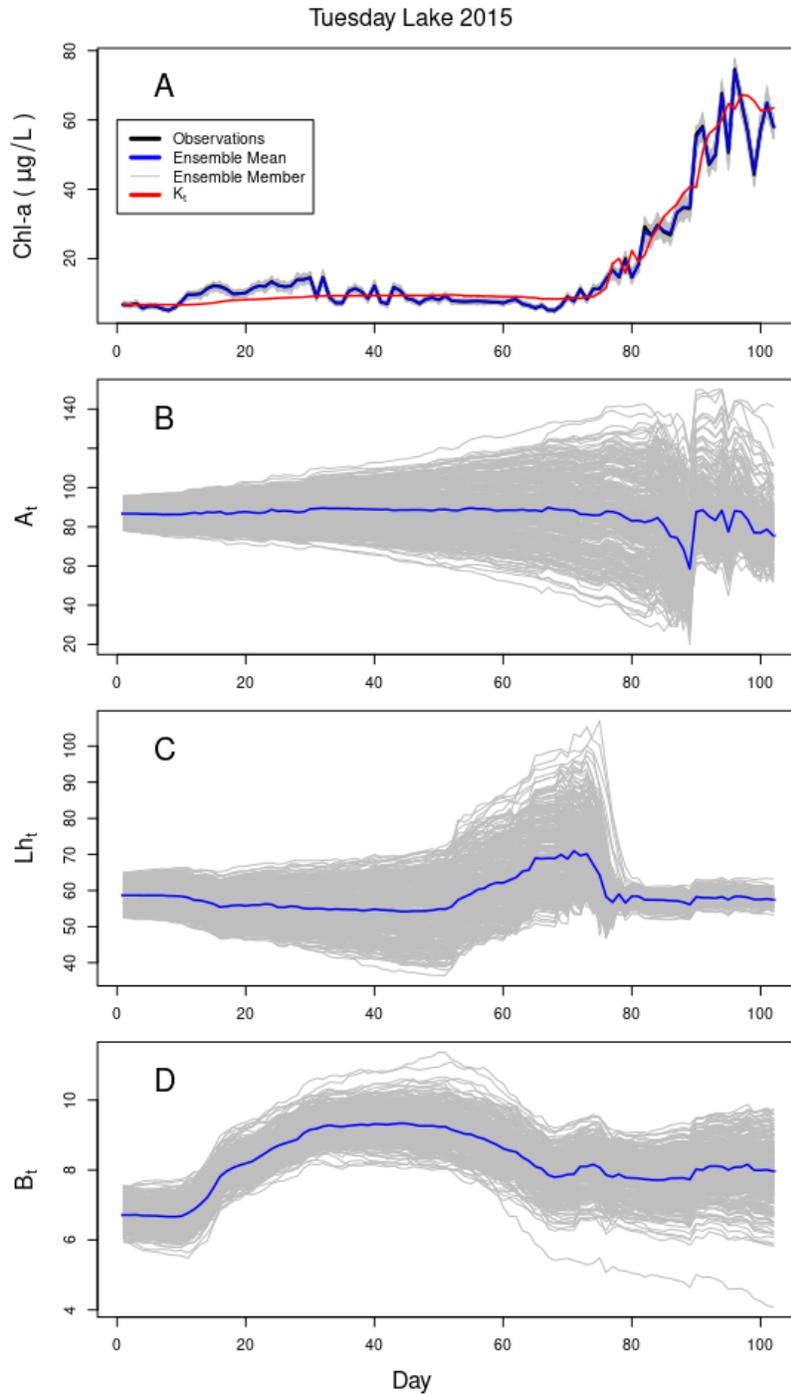


Figure 4.2. Evolution of ensemble state and parameters. A) Observed chlorophyll-a concentrations (black line), individual ensemble state fits of chlorophyll (grey lines), ensemble mean (blue line), and estimate of carrying capacity (red line). B) Individual ensemble parameter fits (grey lines) and ensemble mean for asymptotic chlorophyll concentration A_t . C) Individual ensemble parameter fits (grey lines) and ensemble mean for loading half constant Lh_t . D) Individual ensemble parameter fits (grey lines) and ensemble mean for baseline chlorophyll concentration B_t .

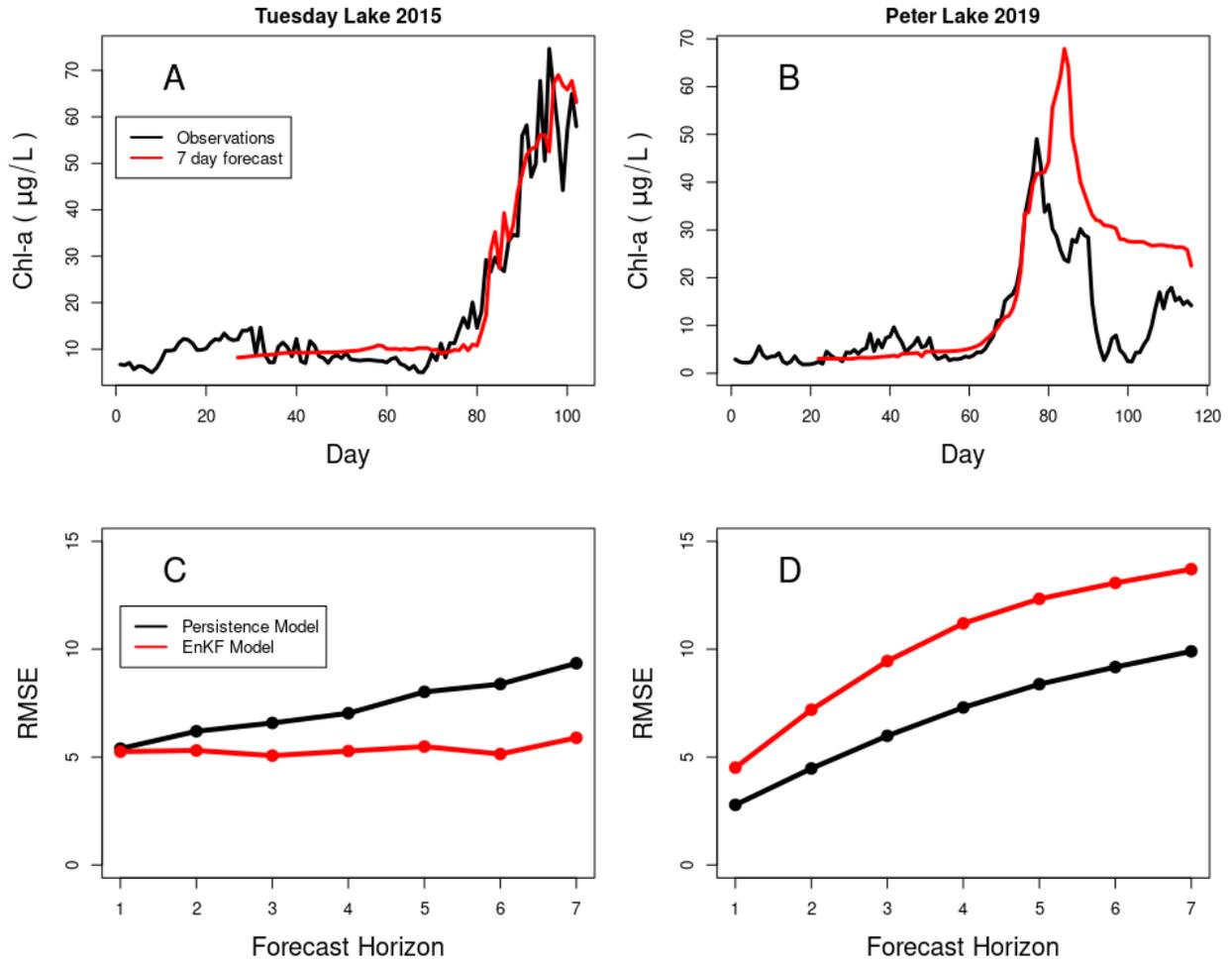


Figure 4.3. Seven day forecasted and observed chlorophyll concentrations (A and B) and overall model error for 1 to 7 day forecast horizons (C and D) in Tuesday Lake 2015 (A, C) and Peter Lake 2019 (B, D).

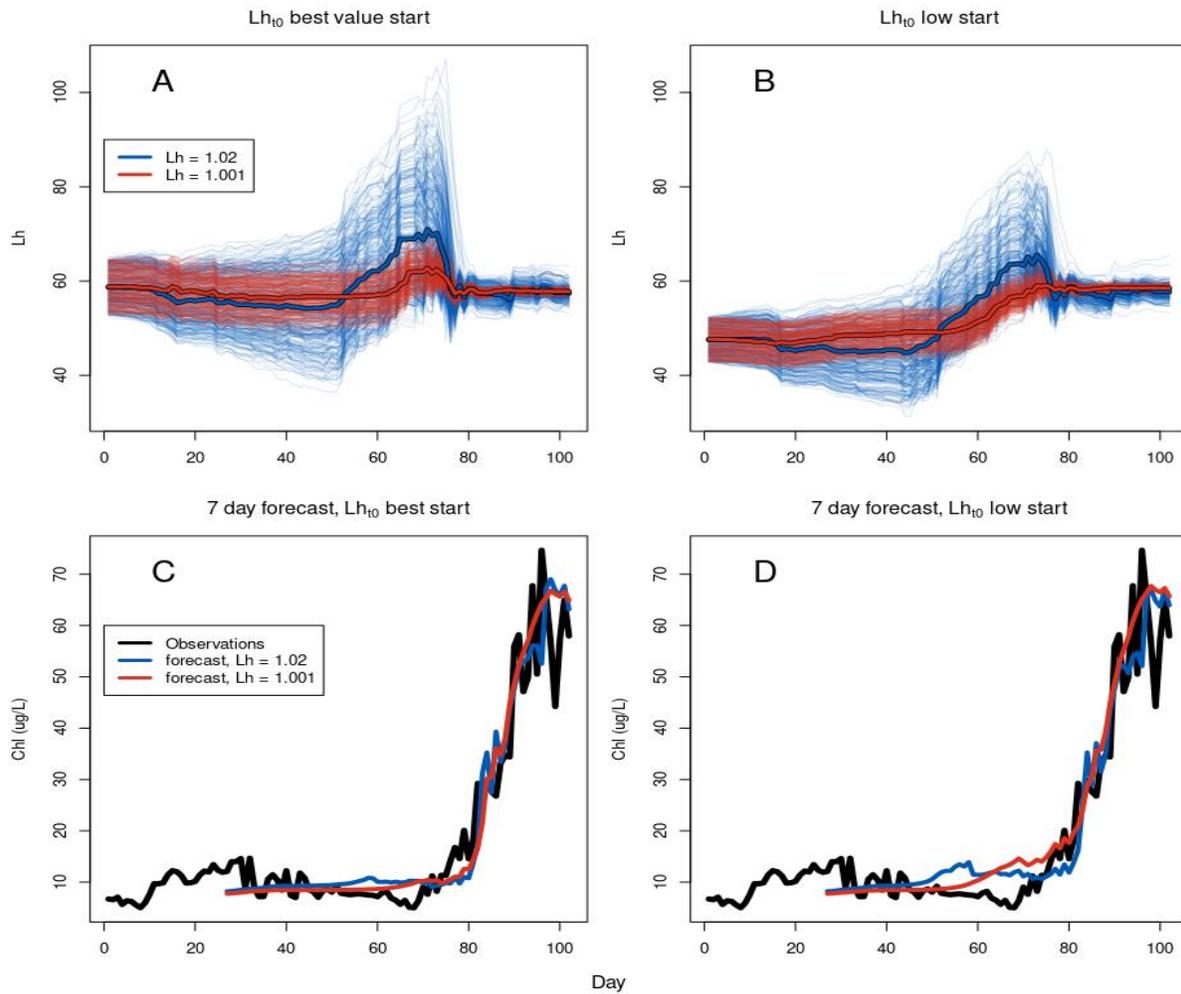


Figure 4.4. Lht parameter fits (A, B) and 7-day forecasts (C,D) for different parameter inflation factors (1.001 = red, 1.02 = blue). A and C start L_h at the “best” value fit from all data (Figure 4.1); B and D start L_h at a lower initial value. Thin, transparent lines in A and B represent individual ensemble member, thick lines are the ensemble mean.

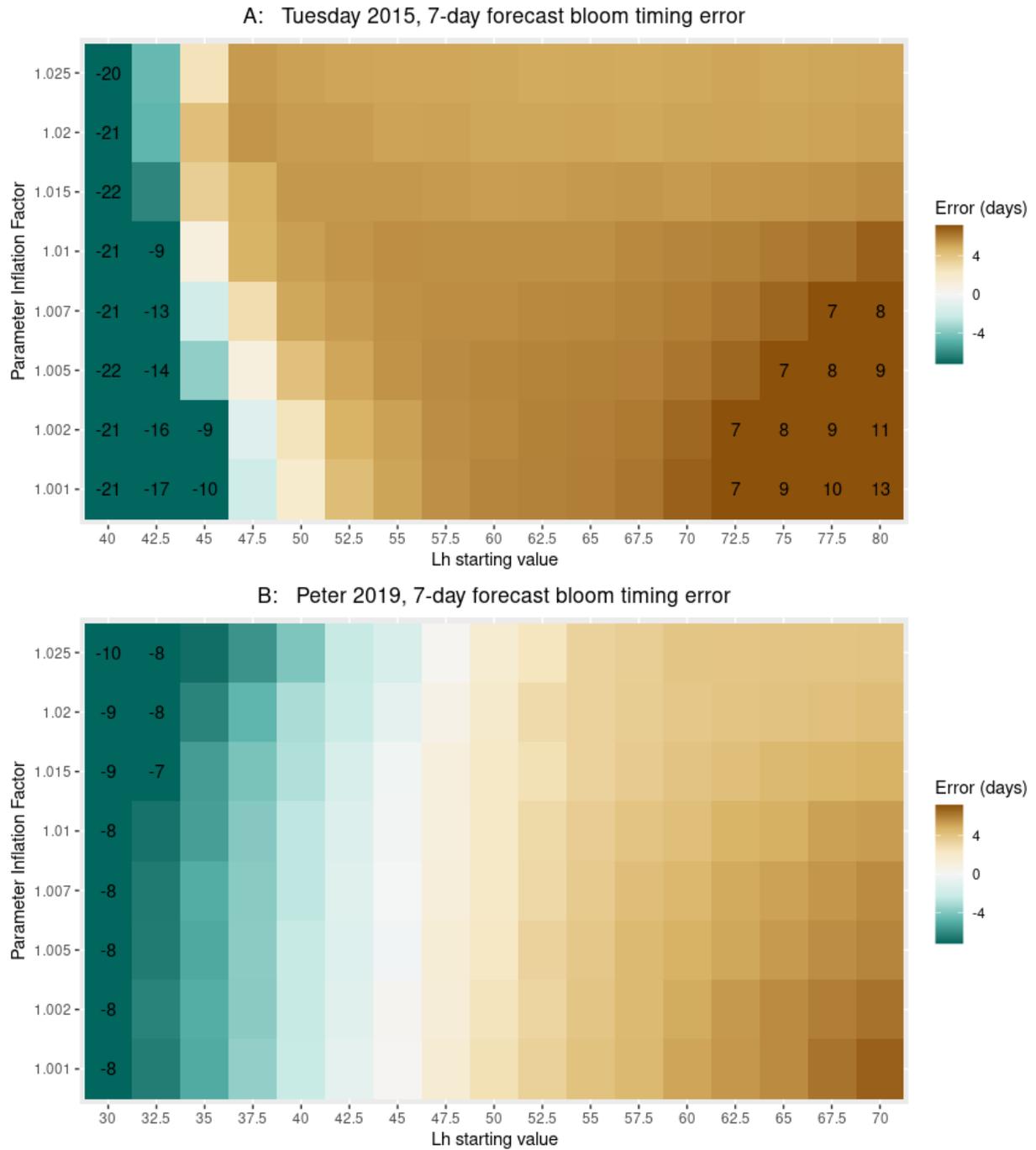


Figure 4.1. Average error in bloom timing for 7-day forecast across 100 replicate simulations for each combination of starting loading half-constant parameter (Lh_t) and parameter inflation factor (IF) for A) Tuesday Lake 2015 and B) Peter Lake 2019. The color scalebar was truncated at ± 8 days to allow for visualization of smaller errors; Lh , IF combinations with < -7 or $> +7$ day errors are printed on the grid, rounded to the nearest whole number.

Chapter 5

Quantifying disturbance and recovery in estuaries: tropical cyclones and high frequency measures of oxygen and salinity

Will be submitted to a journal to be determined

Abstract

Tropical cyclones are important drivers of disturbances in coastal ecosystems. These storms can impact estuaries via a variety of mechanisms including storm surge, flooding from precipitation, and direct damage from high winds and wave action. Prior studies have documented and described disturbances caused by tropical cyclones, including prolonged periods of depressed salinity from high freshwater discharge as well as both increased and decreased dissolved oxygen concentrations from increased loading of organic matter and/or nutrients. However, most studies of disturbance and recovery in these highly diverse and dynamic ecosystems have been limited to one or a few locations or storm events. Thus, there are limited generalizations about storms impacts and whether there are general patterns associated with ecosystem response and recovery. In this study, we apply a new method for detecting disturbance and recovery to long-term and high-frequency measurements of salinity and dissolved oxygen from NOAA's National Estuarine Research Reserve System. By analyzing the impacts of 59 tropical cyclones at 19 estuaries throughout the eastern United States, we were able quantify disturbance occurrence, timing, recovery time, and severity and ask what storm and location properties are associated with changes in disturbance characteristics. Salinity disturbances generally started earlier than dissolved oxygen disturbances and lasted longer. In most cases recovery time was within days, though some disturbances lasted weeks or months. Recovery time was positively correlated with disturbance severity for both variables. Both storm (especially precipitation) and location properties were related to disturbance characteristics. Our

findings demonstrate the power of high-frequency, long-term, and cross-system data, when combined with appropriate statistical methods, to provide insights that improve understanding and potentially management of estuarine resilience to disturbances.

Introduction

Tropical cyclones are large and severe events that can have substantial impacts on coastal ecosystems including estuaries. Storm surge causes flooding and pushes saltwater inland, while high winds and wave action can damage terrestrial (Danielson et al. 2017) and aquatic vegetation (Congdon et al. 2019). Intense precipitation increases freshwater discharge to estuaries with high inputs of nutrients (Angels et al. 2015), dissolved organic carbon (Crosswell et al 2014), and suspended solids (Bukaveckas et al. 2020). These disturbances impact ecosystem function and the ecosystem services estuaries provide including, for example, declines in species and habitats that provide protection from storm surge (Armitage al. 2020), die-offs of recreationally and commercially important fish and shellfish populations (Parker et al. 2013), and increases in carbon emissions to the atmosphere (Crosswell et al. 2014).

Given the impacts of tropical cyclones and their global occurrence and projected increase in severity and geographic distribution due to climate change (Knutson et al. 2010, Sobel et al. 2016), understanding the drivers of estuary disturbance and recovery from storms is crucial to managing these critical ecosystems. Knowing what event and ecosystem characteristics determine if a disturbance occurs, the disturbance effects, and how systems respond can help build resilience (Grimm et al 2017). Additionally, disturbances may have differing effects on different ecosystems, variables, processes, and services that create tradeoffs for management. For example, release of water from upstream impoundments may be necessary to restore ecosystems

and protect infrastructure and property, while negatively impacting downstream estuaries via depressed salinity and increase nutrient loading (Steinman et al. 2002, Rogers and Allen 2008).

Identifying disturbances requires separating event-driven changes in ecosystem parameters from natural variability. However, doing so for tropical cyclone impacts on estuaries is difficult because they are inherently dynamic through time and spatially heterogeneous. At a given location within an estuary, what constitutes “normal” values of a parameter is determined by interacting processes (e.g., tidal, diel, and seasonal cycles, weather) and characteristics (e.g., hydrologic position, depth, bottom substrate). The influence that these and other factors play changes within and between estuaries and depending on what variable or parameter is being considered. As such, establishing baseline conditions requires either a strong understanding of what processes dominate at a given location or extensive prior data.

Despite the many difficulties that have limited cross-system and cross-storm examination of disturbances, the data to do so is increasingly available for a number of ecosystem parameters (Gaiser et al 2020). Developments in sensor technology and remote sensing platforms have made it possible to measure key physical, chemical, and biological variables at higher and higher frequency. In aquatic ecosystems, in-situ sensors can measure water quality parameters like temperature, salinity, turbidity, pH, dissolved oxygen, phytoplankton pigment fluorescence, and nutrient concentrations on the scale of seconds to minutes (Glasgow et al. 2004, Fries et al. 2008). Satellite remote sensing can be used to infer shellfish and submerged aquatic vegetation coverage (Niewhof et al. 2015, Wang et al. 2007), as well as the distribution of turbid waters (Doxaran et al 2006) and phytoplankton (Jiang et al 2020). For some locations, high-frequency measurements have been collected for decades by monitoring programs like NOAA’s National Estuarine Research Reserve System (NERRS; <https://coast.noaa.gov/nerrs/>).

While having high frequency and long-term data is a helpful first step to understanding patterns and controls of disturbances, statistical methods and algorithms are required for assessing impacts. However, the different mechanisms by which tropical cyclones can cause disturbances and their potentially contrasting effects make comparison across different variables, storms, and sites using simple metrics difficult. For example, storm surge may increase salinity of a given site during one storm, whereas river discharge may decrease it following another storm that lacks significant storm surge but has much higher precipitation (Davis et al. 2004). Dissolved oxygen concentrations may crash in response to increased bacterial respiration of terrestrial carbon, increase sharply if algal blooms develop from nutrient spikes, or stay about the same but with decreased amplitude of daily cycles if flushing limits primary producer growth and ecosystem respiration.

The combined effects of inherent spatial and temporal variability and heterogeneous potential responses have limited comprehensive study of the patterns and controls of post-hurricane disturbance and recovery. Most studies of estuarine response to tropical cyclones cover one or a few storms in one or a few locations, and often use experimental designs based on opportunistic sampling due to the unpredictability in when and where storms will occur (Pruitt et al. 2019). Such studies have provided important insights on the impacts that tropical cyclones can have on estuaries, including the importance of storm and site characteristics in determining storm impact (Wetz and Paerl 2008), cascading effects of multiple hurricanes hitting a single location over a short time period (Paerl et al. 2001), and relationships between system resistance and resilience to disturbance (Patrick et al 2020). Paerl et al. (2018) used long-term monitoring to distinguish how differing storm types cause different responses for several tropical cyclones that struck a large estuary-sound complex over more than 20 years. Increasingly available data

for numerous estuaries (Mills et al. 2008) provides the opportunity to assess hurricane impacts across many different storms, locations, and variables. Here, we present a synthesis of disturbance and recovery for estuaries monitored by the National Estuarine Reserve System (NERRS) of the U.S. National Oceanographic and Atmospheric Administration (NOAA). We apply a new disturbance detection method designed to robustly identify disturbance and recovery in high frequency data to water quality measurements from the NERRS monitoring program. Based on hundreds of storm-station time series, we ask: 1) Can the method quantitatively distinguish disturbances for a variety of estuaries and tropical cyclones? 2) What are the occurrence, timing, length, and severity characteristics of tropical cyclone disturbances in estuaries? And 3) What storm and site properties are associated with changes in disturbance characteristics?

Methods

Study Sites and Data – High frequency time series of water quality parameters from the National Oceanic and Atmospheric Administration’s National Estuarine Research Reserve System (NERRS) were analyzed for disturbance events associated with tropical storms. NERRS is composed of 29 U.S. estuaries, with each site containing several monitoring stations that collect water quality, meteorological, nutrient, and pigment data at various sampling frequencies. For this study, we focus on the water quality measurements collected by automated sensors at high frequency (15 minutes since 2007, 30 minutes prior) at 19 Atlantic NERRS sites from 2000 to 2018 (Appendix 4 Table A4.1). The stations measure several water quality parameters: temperature, specific conductivity, salinity, dissolved oxygen (in percent saturation and mg L^{-1}), depth, pH, turbidity, and optionally chlorophyll fluorescence. Here we focus on two variables, salinity and dissolved oxygen percent saturation (DO % sat). Salinity was chosen as it is critical

determinant of habitat suitability for aquatic organisms, varies across most NERSS sites from coastal to inland stations, and is controlled largely by physical processes. DO % saturation was chosen also as a critical determinant of habitat suitability. DO percent saturation (as opposed to concentration) controls for the effect of temperature and is driven by biological processes (i.e., primary production and respiration) and physical processes (e.g., atmospheric exchange). Data were obtained from the NERSS Central Data Management Office's Advanced Query System and all measurements with data quality flags were removed prior to analysis.

Storm identification – Tropical storms that potentially impacted water quality at each NERSS site were identified using a two-step process. First, storms that passed within 250 km of each NERSS site were ascertained using the *hurricaneexposure* and *hurricaneexposuredata* R packages (Anderson et al. 2020a, Anderson et al. 2020b). Second, for each identified storm, potential impacts were determined by plotting salinity and DO % saturation from 30 days prior to 60 days after the date the storm passed closest to the NERSS site. If any variable at any station within a NERSS site appeared to be affected by the storm (defined as an increase or decrease in the mean or variability), all stations and parameters for that site and storm were classified as “potentially impacted” and included in further analyses. While the second step is subjective, including some storms with minor impacts and possibly missing storms with subtle impacts, we aimed to be inclusive in classifying potentially impacted sites/storms to allow the disturbance detection algorithm to quantitatively distinguish events that fell outside the range of historic variability (see below). Alternative methods for identifying storms that potentially impacted water quality at sites were explored, such as thresholds in meteorological variables (high wind speed, heavy precipitation, drops in barometric pressure). However, the many potential mechanisms by which estuarine water quality can be impacted by storms (e.g., storm surge, wind

driven waves, local precipitation, increased discharge from the watershed) as well as frequent missing meteorological data during storms precluded the use of such a method. Ultimately, tropical cyclone-associated disturbances were not observed for over 60% of the station-storm-variable combinations analyzed (see Results). Non-detection was expected given the distance threshold we used and the minor impacts of weaker hurricanes. Our method identified storm impacts on salinity and dissolved oxygen in approximately 40% of the cases and these detections were not limited to only events that caused widespread and severe disturbances.

Disturbance detection – After potentially impacted NERRS sites for each storm were determined, individual stations within each site with sufficient data were identified for disturbance detection analysis. Station-storm-variable combinations with more than 25% missing data or a 5 day or longer gap in measurements during the period from 14 days prior to 60 days after the storm was closest were excluded from further analysis, as were combinations with fewer than 8 other years of data to use as reference data (see below) during the same date range meeting the same requirements.

We applied a recently developed disturbance detection method designed for use with high frequency data (Walter et al. In Review). The method is implemented in an R package available on GitHub (<https://github.com/jonathan-walter/disturbhf>). It compares the distribution of a variable for rolling windows of the time series within a test period to a reference period using the empirical cumulative distribution function, (ECDF). The analysis consists of three steps. First, the difference statistic time series $d_w(t)$ is calculated for each window within the test period:

$$d_w(t) = \sum_{i=1}^N \left| ECDF(x_{test,W}(t)) - ECDF(x_{ref,W}(t)) \right|_i * dx$$

where $x_{test,W}(t)$ are the variable values within a rolling window of width W centered at time t within the test period (Figure 5.1A), $x_{ref,W}(t)$ are the variable values in a reference period, N is the number of intervals at which to evaluate the EDCFs (here we use 1000) over the range of observed values in the test and reference windows, and dx is the width of those intervals (equal to $(x_{max} - x_{min}) / N$; Figure 5.1B,C). x_{ref} can be defined to be either fixed (all values within the reference period are used) or adaptive, where rolling windows of a specified width within the reference period are used to account for seasonable trends. As tropical cyclones occur from summer into late fall when seasonal changes in water quality might be expected, we use the adaptive reference period so rolling windows within the test period are compared to windows centered at the same day of year in the reference years.

In the second step, the d_w time series is rescaled based on the variability observed in the reference period. This is done by calculating $d_{w,ref}$ as above, but $x_{test,W}(t)$ is instead defined by rolling windows of observations within the reference period. The mean ($\mu_{d_w,ref}$) and standard deviation ($\sigma_{d_w,ref}$) of $d_{w,ref}$ are used to rescale $d_w(t)$ as a z-score (Figure 5.1D):

$$z(t) = (d_w(t) - \mu_{d_w,ref}) / (\sigma_{d_w,ref})$$

Finally, user-specified thresholds in $z(t)$ that define disturbance ($thresh_{dist}$) and recovery ($thresh_{recov}$) are applied to identify the timing of disturbance events (initiation and conclusion). Short disturbances and recoveries can optionally be combined or removed using minimum disturbance and recovery lengths. We used a test window width of 3 days and a reference window width of 6 days as a balance between the ability to detect short disturbances vs. power to robustly characterize variable distributions. We also required that disturbances and recoveries last for at least 24 hours and set $thresh_{recov} = 0.5 * thresh_{dist}$.

Disturbance threshold sensitivity – The disturbance detection algorithm is based on the idea that storms can cause changes in the distribution of water quality values that are distinguishable from background variability. As the method compares the entire distribution of data as opposed to a single metric like the mean or median, it has the potential to capture a wide range of storm impacts (e.g., increases or decreases, changes in variance or other moments like skewness). To evaluate the dependence of disturbance detection on $thresh_{dist}$, we conducted a sensitivity analysis by varying $thresh_{dist}$ from 1 to 3 and quantifying the number of detected disturbances in the storm and pre-storm periods across all station-storm-variable combinations identified above. The storm period was defined as 3 days prior to 7 days after when a storm was closest to each NERRS site for this analysis to limit disturbances to those that were highly likely to be caused by the storm. The pre-storm period was set to a period of the same length from 14 days to 4 days prior to when the storm was closest.

Disturbance event characteristics and drivers – After setting the algorithm parameters and identifying disturbance events, we quantified disturbance characteristics and explored potential relationships with station and storm variables. Disturbance event characteristics included disturbance occurrence, timing (relative to when the storm was closest), length of time between disturbance initiation and recovery, and severity (peak $z(t)$ during the disturbance). Disturbance events were limited to first occurring disturbances starting from 3 days before to 30 days after each storm was closest to a NERRS site. Potentially explanatory station variables included mean salinity and depth as proxies for relative location within the estuary (oceanic vs. inland), mean tidal range as a proxy for tidal influence, and standard deviation of salinity as a proxy for variability in the contribution of upstream vs. ocean water sources. Storm traits included closest storm distance to the NERRS site, maximum wind gust speed, duration of wind

gusts over 20 m/s, total storm precipitation, and storm surge height. Storm distance and winds were determined from the *hurricaneexposure* R package (Anderson et al. 2020a); winds were from the population-weighted center of the closest county. Total storm precipitation was obtained from the PRISM reanalysis product (PRISM Climate Group 2021); daily precipitation totals from 3 days before to 7 days after the storm was closest were summed. Storm surge height was calculated from depth observations at each station as the difference between the maximum depth observed from 3 days before to 7 days after storm passage and the maximum depth from the preceding two weeks. While imperfect due to not accounting for if storm surge occurred at high or low tide, or longer period tidal cycles, the storm surge metric provides an indicator of how high water got at each location within an estuary. Relationships between disturbance characteristics and potential driver variables were assessed using general multiple logistic and multiple linear regression; the best regression model was determined separately for each combination of disturbance characteristic and ecosystem variable (salinity or dissolved oxygen) using AIC and stepwise model selection (Venables and Ripley 2002).

Results

The disturbance detection method successfully identified anomalous salinity and oxygen conditions associated with tropical cyclones, and these anomalies included a wide range of responses and disturbance lengths. For example, Hurricane Florence in 2018 caused significant declines in salinity at North Inlet-Winyah Bay NERR's Debidue Creek station, and it took over a month to return to normal values (Figure 5.2A). At Jacques Cousteau NERR, there was a sharp but short (~ 2 days) increase in salinity in response to storm surge from Hurricane Sandy (2012), after which salinity was lower than before the storm but still within the range of natural variability (Figure 5.2B). The distribution difference statistic also increased in response to storms

that impacted variability more strongly than mean values; at Grand Bay NERR's Squamscott River station, Hurricane Hanna in 2008 significantly decreased the amplitude of dissolved oxygen oscillations but the mean remained near 85% saturation (Figure 5.2C). The normalized distribution difference statistic (z-score) increased over this period, but not enough to trigger a disturbance event for a disturbance threshold of 2. A more severe disturbance in dissolved oxygen occurred at Rookery Bay NERR's Lower Henderson Creek station, where Hurricane Irma caused a crash in dissolved oxygen to near 0% saturation (Figure 5.2D). The z-score for this disturbance peaked at almost 8, indicating a highly anomalous event for the location.

Disturbance threshold sensitivity analysis – The number of disturbances detected during 11-day pre-storm and storm periods across all 972 station-storm-variable cases decreases steadily with higher disturbance thresholds (Figure 5.3A,B). For dissolved oxygen, the number of disturbances detected for a given disturbance threshold was very similar in the pre-storm and storm periods. The number of salinity disturbances was also very similar to the number of oxygen disturbances during the storm period, but significantly lower during the pre-storm period at a given disturbance threshold. This translates to a higher proportion of cases where a disturbance was detected in the storm period but not the pre-storm period for salinity than for dissolved oxygen (Figure 5.3C), indicating there is generally more separation between d_w values before and after a storm for salinity and that dissolved oxygen disturbances had a similar probability of starting in the pre-storm and storm periods. Disturbance characteristic and driver variable analyses were carried using disturbance threshold of 2 and only for the first occurring disturbances starting from 3 days before to 30 days after storm passage, in order to base our conclusions on disturbances that were relatively severe and more likely to be driven by storm impacts as opposed to other causes.

Disturbance event characteristics – A wide range of disturbance timing, length, and severities were observed. Over 50% of detected disturbances began within 5 days of when the eye of a tropical storm passed closest to each NERRS site for both dissolved oxygen and salinity (Figure 5.4C). The peak in timing for disturbance initiation for salinity oxygen was from 0.5 to 1.5 days after storm passage, with the cluster of most frequently observed times from 3 days before to 4.5 days after (Figure 5.4A). For dissolved oxygen, disturbances most frequently began from 2.5 to 3.5 days after storm passage; the cluster of most frequent times peaked much lower and was wider relative to salinity. (Figure 5.4B).

Most detected disturbances were relatively short (median length of 6.4 days for salinity and 4.5 days for DO % sat.; Figure 5.5C), but there was a long tail in the distribution of disturbance lengths, with disturbances more than 50 days long observed for both DO % sat. and salinity (Figure 5.5A,B). Salinity disturbances between 1 and 8 days were most common, with a significant drop in the number of observed disturbances longer than 14 days (Figure 5.5A). The cluster of most common disturbance lengths for dissolved oxygen was shifted to slightly lower values; the most common disturbance lengths was 1 – 6 days and the drop in observed disturbance lengths occurred after 12 days (Figure 5.5B).

There was significant variation in the peak severity value of each disturbance, with peak severities between 2 and 5 being relatively common and several values between 5 and 15 for each variable (Figure 5.6). There was a significant, positive relationship between log-transformed peak severity and log-transformed disturbance length for both salinity and DO % saturation (Figure 5.6), with very similar correlation coefficients ($r = 0.52$ and 0.57 for salinity and oxygen, respectively).

Potential disturbance drivers – At least one potential driver was identified for all disturbance event characteristics for both dissolved oxygen and salinity based on regression analysis (Table 5.1). Higher precipitation and mean salinity were associated with increased occurrence of salinity disturbances, while increases in tidal range were associated with lower salinity disturbance occurrence. For dissolved oxygen, storms with higher precipitation and stations with larger depths had more disturbances, while stations with higher tidal ranges had fewer disturbances. Storms with longer durations of wind gusts over 20 m/s were associated with later occurring salinity disturbances; increased precipitation led to earlier occurring salinity disturbances. For dissolved oxygen disturbance timing, stations with higher salinity had later starting disturbances while storms with more precipitation that passed closer had earlier disturbances. Longer salinity disturbances were positively correlated with tidal range, wind gust duration, and precipitation. Longer dissolved oxygen disturbances were positively correlated with storm precipitation and station salinity variability, and negatively correlated with station mean salinity and storm wind gust duration. Several storm and station variables were positively and negatively related to peak severity for salinity disturbance, while the only driver variable that had a significant relationship with peak disturbance severity for DO % sat was maximum storm surge height.

Discussion

Combining the new disturbance detection method with long-term, high frequency data allowed us to robustly identify and characterize tropical cyclone-associated disturbances across 19 estuaries and 59 different storms in the eastern United States. The method was able to distinguish diverse disturbance types in two important ecosystem state variables (Figure 5.2, Introduction Question 1). Disturbances had a wide range of timing, length, and severity (Figures

5.3, 5.4, 5.5; Question 2), and both storm and site properties were important in determining disturbance characteristics (Table 5.1; Question 3).

Across all storms-estuary combinations analyzed, dissolved oxygen saturation and salinity had a similar number of disturbances detected during the time period around storms, though there were slightly more disturbances detected for salinity at disturbance thresholds > 2 (Figure 5.3). The lower number of disturbances detected prior to storms for salinity relative to DO % saturation suggests that disturbances in this variable are more closely tied to storm events, while in addition to tropical cyclones, other mechanisms are important generators of anomalous DO % saturation values. Biological processes that alter oxygen concentrations (i.e., respiration and primary production) are highly dynamic in time and related to several environmental drivers. The dynamic processes that drive respiration and production are a likely explanation for the observation of oxygen anomalies not associated with storms (Caffrey et al. 2014, Murrell et al. 2017). Biological mechanisms could also explain the observed lag in dissolved oxygen disturbance initiation relative to salinity. For example, physical processes that alter salinity like storm surge or precipitation-driven increases in freshwater discharge may occur simultaneously with or shortly after storm passage, whereas respiration and production impacts on oxygen concentration may take time to longer to develop. Dissolved oxygen disturbances were generally shorter than salinity disturbances. While these variables share many of the physical processes that promote recovery (e.g., stream and river discharge, tidal exchange), equilibration with atmospheric oxygen concentrations especially under vigorous mixing in estuaries could explain the faster recovery of DO % sat. relative to salinity (Kremer et al. 2003).

While short disturbances (< 7 days) were most frequent for both variables, longer disturbances were also common. Forty-six percent of salinity disturbances lasted longer than 7

days as did 28% of dissolved oxygen disturbances. For many organisms that inhabit estuaries, these disturbances likely represent prolonged periods of stress and require substantial movement for mobile species. Low oxygen can cause fish and shellfish die-offs as documented for some hurricanes (Paerl et al. 1998, Paerl et al. 2001, Parker et al. 2013). While many estuarine species are euryhaline, large salinity ranges associated with storm-disturbances may exceed tolerance (Du et al. 2021). Though rare, we also found several cases where disturbances persisted for more than 30 days (salinity: n = 17, maximum = 99 days; dissolved oxygen: n = 3, maximum = 59 days). The extremely slow recovery rates for these events could arise from different mechanisms. Long duration, low salinity disturbances result from high freshwater discharge (Paerl et al. 2001, Du et al. 2019). For dissolved oxygen disturbances, extreme loading of organic matter from the watershed into estuaries can stimulate high respiration that draws down oxygen concentrations (Paerl et al. 2018). Alternatively, storms can increase nutrient loading from internal or external sources, leading to algal blooms that increase oxygen concentrations, and though nutrient concentrations often decline quickly following hurricanes, recycling can maintain primary producer biomass (Peierls et al. 2003).

We observed a positive relationship between disturbance length (also referred to as recovery or return time) and disturbance severity, with slightly more variability in the relationship for salinity than dissolved oxygen. These two disturbance characteristics quantify metrics of ecosystem resilience and stability; terms that have taken varying definitions and received considerable attention in ecology and other fields. To facilitate comparison, here we adopt the same definitions used by a recent analysis of coastal ecosystem response to Hurricane Harvey in 2017 (Patrick et al. 2020). Disturbance length corresponds to resilience and quantifies the ability of a system to recover after a perturbation, while disturbance severity corresponds to

ecosystem resistance and measures a system's ability to oppose change (Pimm 1984). Patrick et al. (2020) found a negative relationship between resilience and resistance in the response of several categories of estuary variables (hydrology, hydrography, biogeochemistry, biota) in the response of several Texas estuaries to Hurricane Harvey. The variables we studied (salinity and dissolved oxygen) fall in the hydrography category of Patrick et al. (2020). Our disturbance lengths are similar to the return times they observed, though the resistance measures are not directly comparable (log response ratio vs. peak z-score). Taken together, the wide range in observed resilience and resistance values and differing strength and direction of their correlation suggests the relationship is scale-related: negative across different variable categories for a given storm (Patrick et al. 2020); and positive for individual variables across many estuaries and storms (this analysis).

By analyzing the impact of many storms on several estuaries, we examined the role that storm and site properties play in determining disturbance characteristics. Both storm and site properties were important (Question 3). Total storm precipitation was the most common storm variable related to disturbance responses, with higher precipitation associated with more, earlier starting, and longer disturbances for both salinity and dissolved oxygen. This finding is important given the projected increase in tropical cyclone severity and precipitation amounts due to climate change (Patricola and Wehner 2018). Mean salinity was the most common station property associated with disturbance responses, with saltier locations having more and more severe salinity disturbances, as well as later and shorter dissolved oxygen disturbances. Identifying storm and site characteristics associated with disturbances is important to identify current estuarine locations that are highly susceptible to disturbance and how climate change in the future may impact responses. Insights may offer ways to increase ecosystem resilience

through management of infrastructure (e.g., water retention/release), habitat restoration to promote refuges and portfolio effects (Schindler et al. 2015), and watershed land-management to limit run-off during extreme storms.

While the disturbance detection method was able to identify and characterize disturbances, it also has limitations. The method quantifies any difference between the distribution of values in the test and reference periods, but doesn't distinguish between different types/directions of disturbances (e.g., if a change is to higher or lower values, or increased/decreased variability). There also is not a direct correspondence between the disturbance statistics (d_w or z-score) and physically meaningful ecosystem state values or thresholds (e.g., oxygen or salinity concentrations at which organisms are harmed). These issues could be addressed by first using the method as described above to demarcate disturbances, then to compare differences in the mean, variance, etc. of the test and reference years within those periods. Finally, the method requires both high frequency and long-term measurements to identify disturbances, which limits the variables and locations to which it can be applied. However, these types of data are increasingly available from sensors that can measure important ecosystem state variables (Porter et al. 2012). The data from NOAA's National Estuarine Research Reserve System illustrates the immense value of long-term programs measuring the same variables at different sites, especially for events that are unpredictable but have large consequences like tropical cyclones.

Despite these limitations, our findings represent an advance in the study patterns and drivers of disturbance at broad spatial and temporal scales in estuaries. By detecting disturbances that cause diverse deviations from baseline variability and quantifying several important disturbance characteristics, our method overcomes many of the constraints that have previously

limited studies to one or a few storms and/or locations (Pruitt et al. 2019). The approach fits naturally within proposed frameworks for understanding disturbance in ecological and socio-economic systems (Gaiser et al. 2020), including specifically for responses to tropical cyclone (Hogan et al. 2020). Disturbances to salinity and dissolved oxygen in estuaries generally start soon after tropical cyclones pass and typically recover within days, though weeks and months long disturbance do occur. Properties of both storms and the locations they impact are related to disturbance response. Future work extending these findings including real time assessments with coupled data-analytical systems offer exciting possibilities that could improve understanding and management of hurricane impacts on estuarine ecosystems.

Acknowledgements

This work was supported by a Rapid Response Grant from the University of Virginia Environmental Resilience Institute as well as high performance computing resources provided by University of Virginia Research Computing's Rivanna system.

References

- Anderson, B, M Yan, J Ferreri, W Crosson, M Al-Hamdan, A Schumacher, D Eddelbuettel. 2020a. hurricaneexposure: Explore and Map County-Level Hurricane Exposure in the United States. R package version 0.1.1. <http://CRAN.R-project.org/package=hurricaneexposure>.
- Anderson B, A Schumacher, W Crosson, M Al-Hamdan, M Yan, J Ferreri, Z Chen, S Quiring, S Guikema. 2020b. hurricaneexposedata: Data Characterizing Exposure to Hurricanes in United States Counties. R package version 0.1.0. <https://github.com/geanders/hurricaneexposedata>.
- Anglès, S, A Jordi, L Campbell. 2015. Responses of the coastal phytoplankton community to tropical cyclones revealed by high-frequency imaging flow cytometry. *Limnology and Oceanography*. 60: 1562–1576.
- Armitage, AR, CA Weaver, JS Kominoski, SC Pennings. 2020. Resistance to Hurricane Effects Varies Among Wetland Vegetation Types in the Marsh–Mangrove Ecotone. *Estuaries and Coasts*. 43: 960–970.

- Bukaveckas, PA, S Tassone, W Lee, RB Franklin. 2020. The influence of storm events on metabolism and water quality of riverine and estuarine segments of the James, Mattaponi, and Pamunkey Rivers. *Estuaries and Coasts*. 43(7): 1585-1602.
- Caffrey, JM, MC Murrell, KS Amacker, JW Harper, S Phipps, MS Woodrey. 2014. Seasonal and Inter-annual Patterns in Primary Production, Respiration, and Net Ecosystem Metabolism in Three Estuaries in the Northeast Gulf of Mexico. *Estuaries and Coasts*. 37: S222-S241.
- Congdon, VM, C Bonsell, MR Cuddy, KH Dunton. 2019. In the wake of a major hurricane: differential effects on early vs. late successional seagrass species. *Limnology and Oceanography Letters* 4: 155–163.
- Crosswell, JR, MS Wetz, B Hales, and HW Paerl. 2014. Extensive CO₂ emissions from shallow coastal waters during passage of Hurricane Irene (August 2011) over the mid-Atlantic Coast of the U.S.A. *Limnology and Oceanography*. 59: 1651–1665.
- Danielson, TM, VH Rivera-Monroy, E Castañeda-Moya, H Briceño, R Travieso, BD Marx, E Gaiser, LM Farfán. 2017. Assessment of Everglades mangrove forest resilience: Implications for above-ground net primary productivity and carbon dynamics. *Forest Ecology and Management*. 404:115-125.
- Davis, SE, JE Cable, DL Childers, C Coronado-Molina, JW Day Jr, CD Hittle, CJ Madden, E Reyes, D Rudnick, F Sklar. 2004. Importance of Storm Events in Controlling Ecosystem Structure and Function in a Florida Gulf Coast Estuary. *Journal of Coastal Research*. 20(4):1198-1208.
- Doxaran, D, P Castaing, SJ Lavender. 2006. Monitoring the maximum turbidity zone and detecting fine-scale turbidity features in the Gironde estuary using high spatial resolution satellite sensor (SPOT HRV, Landsat ETM+) data. *International Journal of Remote Sensing*. 27(11): 2303-2321.
- Du, J, K Park, TM Dellapenna, JM Clay. 2019. Dramatic hydrodynamic and sedimentary responses in Galveston Bay and adjacent inner shelf to Hurricane Harvey. *Science of the Total Environment*. 653: 554-564.
- Du, J, K Park, C Jensen, TM Dellapenna, WG Zhang, Y Shi. 2021. Massive oyster kill in Galveston Bay caused by prolonged low-salinity exposure after Hurricane Harvey. *Science of the Total Environment*. 774: 145132.
- Fries, DP, SZ Ianov, PH Bhanushali, JA Wilson, HA Broadbent, AC Sanderson. 2007. Broadband, low-cost, coastal sensor nets. *Oceanography*. 20(4): 150-155.
- Gaiser, EE, DM Bell, MCN Castorani, DL Childers, PM Groffman, CR Jackson, JS Kominoski, DPC Peters, STA Pickett, J Ripplinger, JC Zinnert. 2020. Long-Term Ecological Research and Evolving Frameworks of Disturbance Ecology. *BioScience*. 70(2): 141-156.

- Glasgow, HB, JM Burkholder, RE Reed, AJ Lewitus, JE Kleinman. 2004. Real-time remote monitoring of water quality: a review of current applications, and advancements in sensor, telemetry, and computing technologies. *Journal of Experimental Marine Biology and Ecology*. 300: 409 – 448.
- Grimm NB, Pickett STA, Hale RL, Cadenasso ML. 2017. Does the ecological concept of disturbance have utility in urban social–ecological–techno-logical systems? *Ecosystem Health and Sustainability* 3: e01255.
- Hogan JA, RA Feagin, G Starr, M Ross, T Lin, C O’Connell, TP Huff, BA Stauffer, KL Robinson, and 34 others. 2020. A Research Framework to Integrate Cross-Ecosystem Responses to Tropical Cyclones. *BioScience*. 70(6): 477–489.
- Jiang, L, T Gerkema, JC Kromkamp, D van der Wal, PMC De La Cruz, K Soetaert. 2020. Drivers of the spatial phytoplankton gradient in estuarine–coastal systems: generic implications of a case study in a Dutch tidal bay. *Biogeosciences*. 17: 4135-4152.
- Knutson, TR, JL McBride, J Chan, K Emanuel, G Holland, C Landsea, I Held, JP Kossin, AK Srivastava, M Sugi. 2010. Tropical cyclones and climate change. *Nature Geoscience*. 3: 157-163.
- Kremer, JN, A Reischauer, C D’Avanzo. 2003. Estuary-specific variation in the air-water gas exchange coefficient for oxygen. *Estuaries*. 26(4A): 829-836.
- Mills, K, MJ Kennish, KA Moore. 2008. Research and Monitoring Components of the National Estuarine Research Reserve System. *Journal of Coastal Research*. 10055: 1-8.
- Murrell, MC, JM Caffrey, DT Marcovich, MW Beck, BM Jarvis, JD Hagy III. 2018. Seasonal Oxygen Dynamics in a Warm Temperate Estuary: Effects of Hydrologic Variability on Measurements of Primary Production, Respiration, and Net Metabolism. *Estuaries and Coasts*. 41: 690-707.
- Nieuwhof, S, PMJ Herman, N Dankers, K Troost, D van der Wal. 2015. Remote Sensing of Epibenthic Shellfish Using Synthetic Aperture Radar Satellite Imagery. *Remote Sensing*. 7: 3710 – 3734.
- Paerl, HW, JL Pickney, JM Fear, BL Peierls. 1998. Ecosystem responses to internal and watershed organic matter loading: consequences for hypoxia in the eutrophying Neuse River Estuary, North Carolina, USA. *Marine Ecology Progress Series*. 166: 17-25.
- Paerl, HW, JD Bales, LW Ausley, CP Buzzelli, LB Crowder, LA Eby, JM Fear, M Go, BJ Peierls, TL Richardson, JS Ramus. 2001. Ecosystem impacts of three sequential hurricanes (Dennis, Floyd, and Irene) on the United States’ largest lagoonal estuary, Pamlico Sound, NC. *Proceedings of the National Academy of Sciences, USA*. 98(10): 5655-5660.
- Paerl, HW, JR Crosswell, B Van Dam, NS Hall, KL Rossignol, CL Osburn, AG Hounshell, RS Sloup, LW Harding Jr. 2018. Two decades of tropical cyclone impacts on North

- Carolina's estuarine carbon, nutrient and phytoplankton dynamics: implications for biogeochemical cycling and water quality in a stormier world. *Biogeochemistry*. 141: 307-332.
- Parker, ML, WS Arnold, SP Geiger, P Gorman, EH Leone. 2013. Impacts of Freshwater Management Activities on Eastern Oyster (*Crassostrea virginica*) Density and Recruitment: Recovery and Long-Term Stability in Seven Florida Estuaries. *Journal of Shellfish Research*. 32(3): 695–708.
- Patrick CJ, L Yeager, AR Armitage, F Carvallo, VM Congdon, KH Dunton, M Fisher, AK Hardison, JD Hogan, J Hosen, X Hu, B Kiel Reese, S Kinard, JS Kominoski, X Lin, Z Liu, PA Montagna, SC Pennings, L Walker, CA Weaver, M Wetz. 2020. A System Level Analysis of Coastal Ecosystem Responses to Hurricane Impacts. *Estuaries and Coasts*. 43: 943-959.
- Patricola, CM, MF Wehner. 2018. Anthropogenic influences on major tropical cyclone events. *Nature*. 563: 339-346.
- Peierls, BJ, RJ Christian, HW Paerl. 2003. Water quality and phytoplankton as indicators of hurricane impacts on a large estuarine ecosystem. *Estuaries*. 26(5): 1329-1343.
- Pimm, SL. 1984. The complexity and stability of ecosystems. *Nature*. 307: 321-326.
- Porter, JH, PC Hanson, C Lin. 2012. Staying afloat in the sensor data deluge. *Trends in Ecology and Evolution*. 27(2): 121-129.
- PRISM Climate Group, Oregon State University, <http://prism.oregonstate.edu>, 2021.
- Pruitt, JN, AG Little, SJ Majumdar, TW Schoener, DN Fisher. 2019. Call-to-action: a global consortium for tropical cyclone ecology. *Trends in Ecology and Evolution*. 34(7): 588-590.
- Rogers, MW, MS Allen. 2008. Hurricane Impacts to Lake Okeechobee: Altered Hydrology Creates Difficult Management Trade Offs. *Fisheries*. 33(1): 11-17.
- Schindler, DE, JB Armstrong, TW Reed. 2015. The portfolio concept in ecology and evolution. *Frontiers in Ecology and the Environment*. 13(5): 257-263.
- Sobel, AH, SJ Camargo, TM Hall, C Lee, MK Tippet, AA Wing. 2016. Human influence on tropical cyclone intensity. *Science*. 353(6296): 242-246.
- Steinman, A, K Havens, L Hornung. 2002. The managed recession at Lake Okeechobee, Florida: integrating science and natural resource management. *Conservation Ecology* 6(2):17.
- Venables, WN, Ripley, BD. 2002. *Modern Applied Statistics with S*. New York: Springer (4th ed).

- Walter, JA, AF Besterman, CD Buelo, SJ Tassone, JW Atkins, ML Pace. *In Review at Limnology and Oceanography: Methods*. An algorithm for detecting and quantifying disturbance and recovery in high-frequency time series.
- Wang, Y, M Traber, M Milstead, S Stevens. 2007. Terrestrial and Submerged Aquatic Vegetation Mapping in Fire Island National Seashore Using High Spatial Resolution Remote Sensing Data, *Marine Geodesy*, 30: 77-95.
- Wetz, MS, HW Paerl. 2008. Estuarine Phytoplankton Responses to Hurricanes and Tropical Storms with Different Characteristics (Trajectory, Rainfall, Winds). *Estuaries and Coasts*. 31: 419-429.

Table 5.1. Storm and station variables related disturbance characteristics: occurrence/non-occurrence, timing (days after storm passage), length (days), and peak severity (z-score). Storm characteristics are highlighted in blue; site/station characteristics are highlighted in green. Drivers are separated into positive and negative significant relationships with each disturbance event characteristic.* denotes $0.05 \leq p \leq 0.1$; other p-values are < 0.05 .

Disturbance Characteristic	Salinity	DO % saturation
Occurrence	Positive: precipitation, mean(salinity) Negative: tidal range*	Positive: precipitation, depth Negative: tidal range
Timing relative to storm	Positive: wind gust duration Negative: precipitation	Positive: mean(salinity) Negative: precipitation, storm distance
Length	Positive: tidal range, wind gust duration, precipitation* Negative:	Positive: precipitation, sd(salinity)* Negative: mean(salinity), max wind gust speed*
Peak Severity	Positive: mean(salinity), storm distance, wind gust duration Negative: sd(salinity)	Positive: storm surge height Negative:

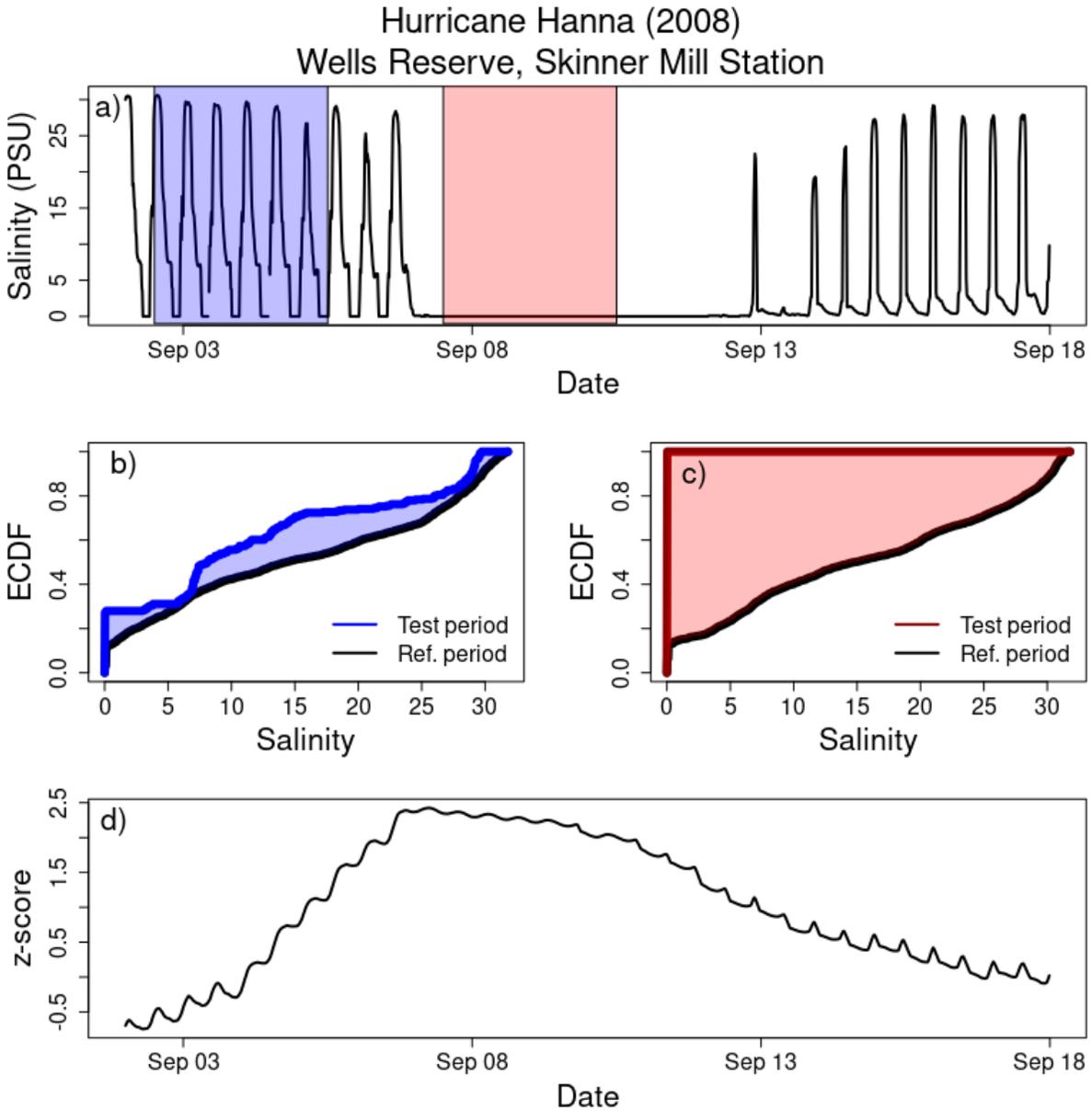


Figure 5.1. Illustration of the disturbance detection algorithm. A) Observations of salinity before and after Hurricane Hanna impacted the Wells Reserve (Maine, USA) in 2008. Shaded periods are example three-day test periods before (blue) and after (red) storm passage. B) Empirical cumulative distributions (ECDF) for the before-impact example test period (blue) and corresponding reference periods in other years (black). The area of the shaded region corresponds to d_w , the distribution difference statistic. C) ECDFs for the after-impact example test period (red) and corresponding reference periods in other years (black). D) Time series of the normalized (z-score) distribution difference test statistic.

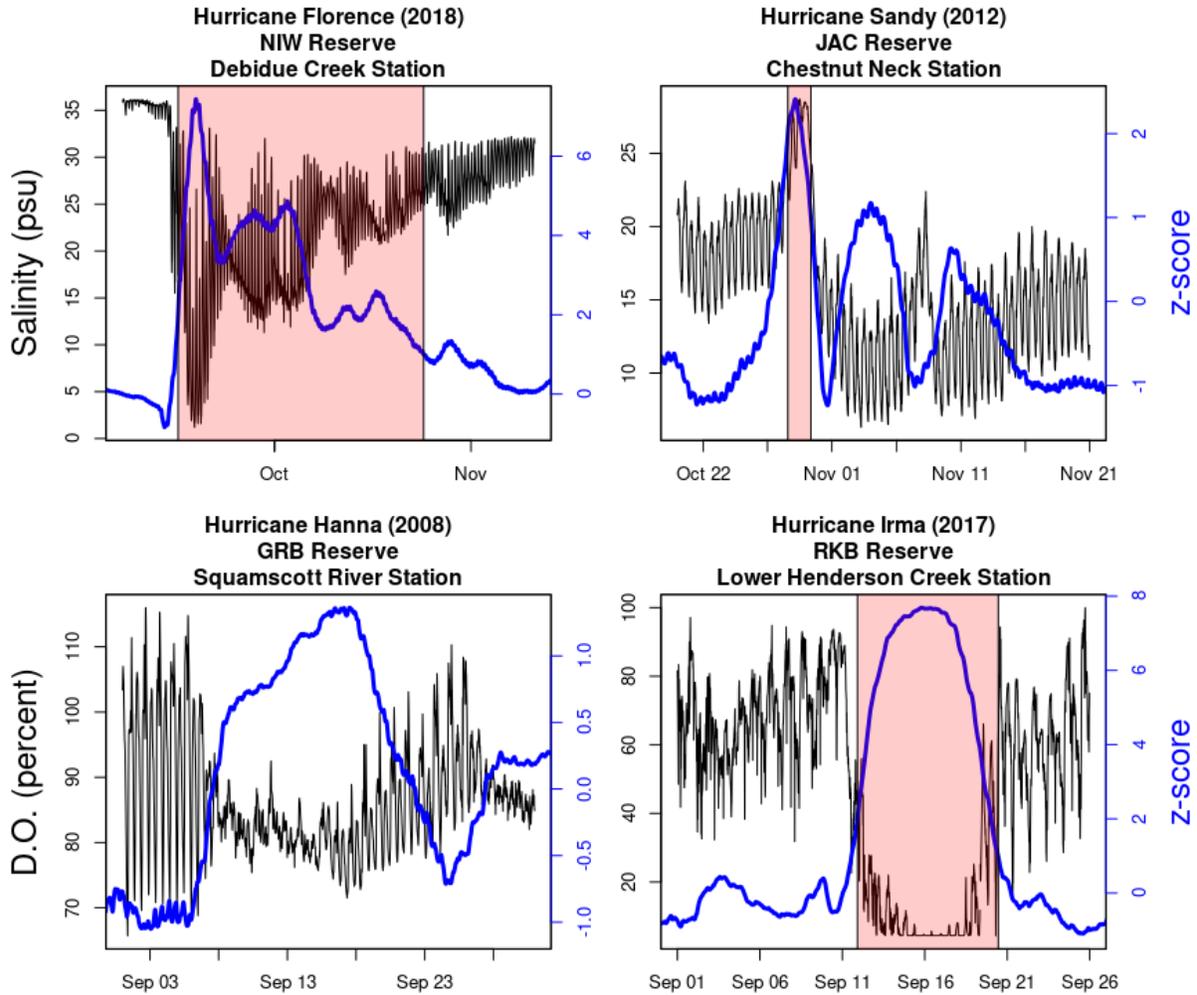


Figure 5.2. Example time series showing hurricane impacts and performance of the normalized distribution difference statistic. Black lines are observations of salinity (A, B) and dissolved oxygen percent saturation (C, D). Blue lines are the normalized distribution difference statistic for three-day wide rolling windows. Red shaded areas represent disturbances identified with z-score a disturbance threshold of 2 and recovery threshold of 1. Note different x and y axis scales.

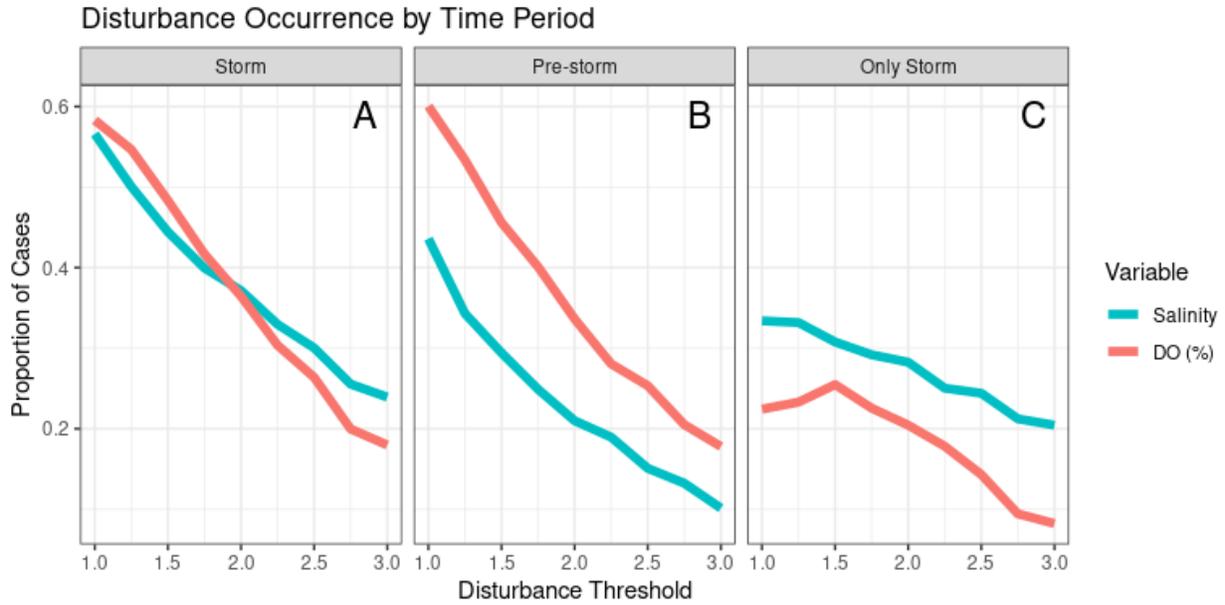


Figure 5.3. Proportion of analyzed cases where a disturbance was detected for different disturbance threshold values during the storm period (A; 3 days before to 10 days after the storm was closest to each site), pre-storm period (B; 14 days before to 4 before the storm was closest to each site), and during the storm period but not the pre-storm period (C). Red lines are dissolved oxygen percent saturation, blue lines are salinity.

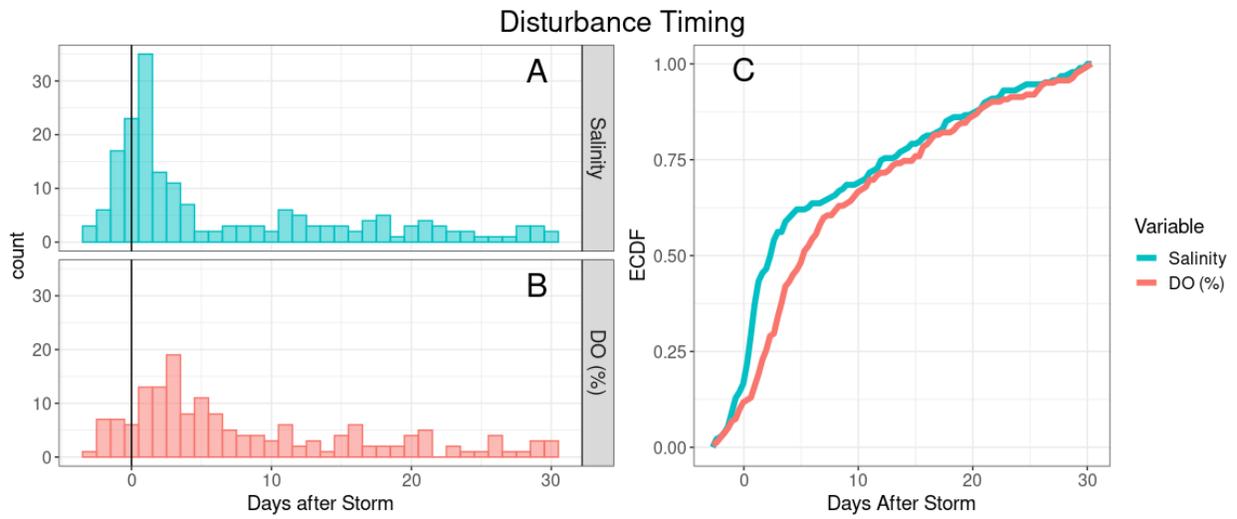


Figure 5.4. Histograms (A: dissolved oxygen, B: salinity) and empirical cumulative distribution curves (C) showing the timing of disturbance start relative to when the storm was closest to each NERRS site.

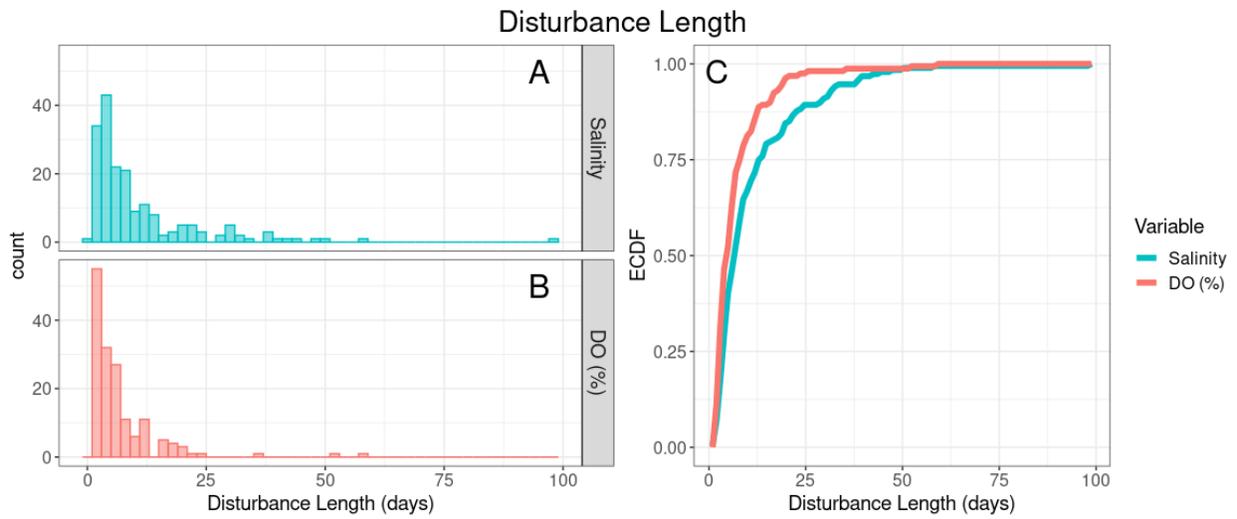


Figure 5.5. Histograms (A: dissolved oxygen, B: salinity) and empirical cumulative distribution curves (C) showing the length of time between disturbance initiation and recovery.

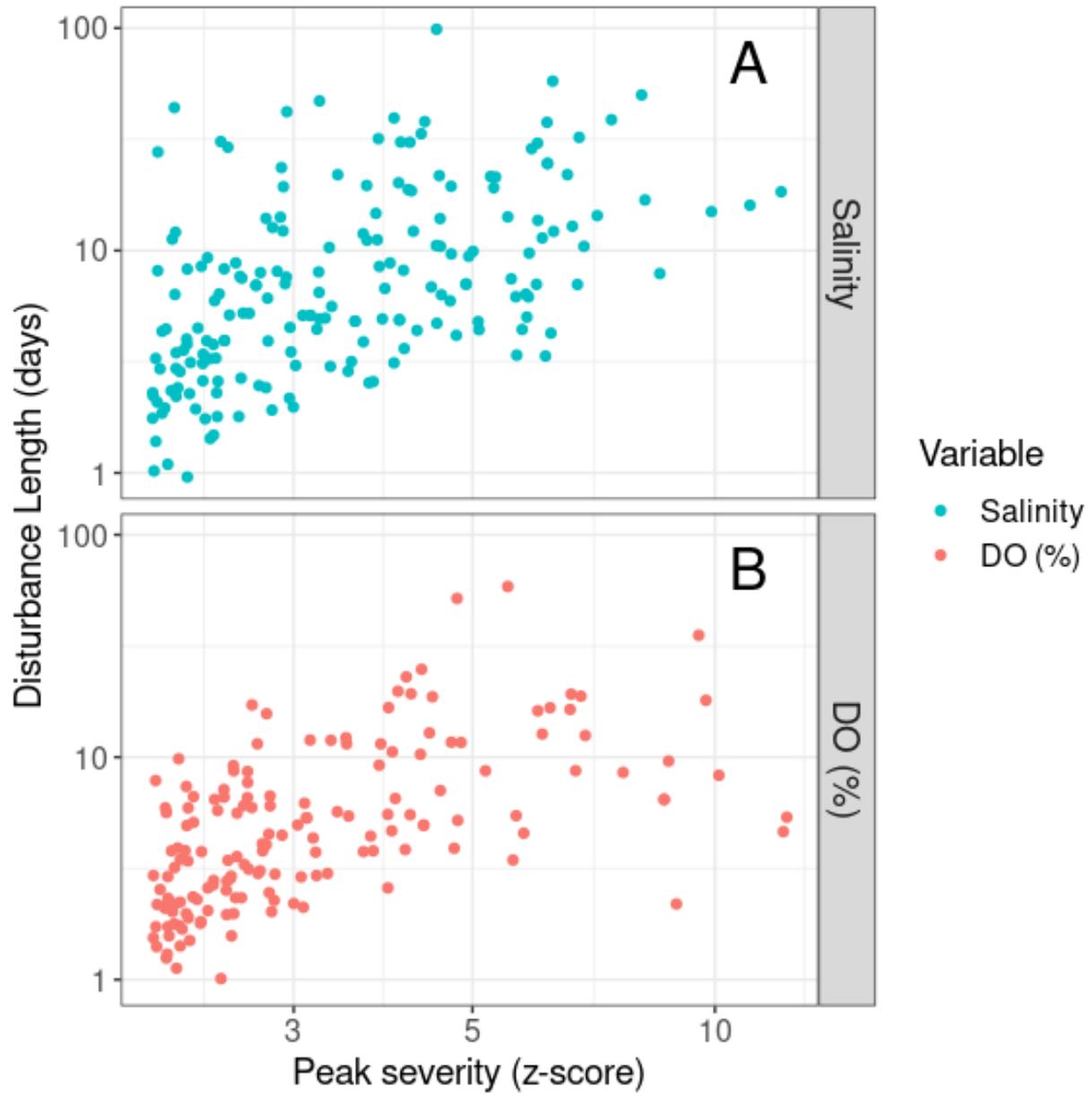


Figure 5.6. Relationship between disturbance length and peak disturbance severity for dissolved oxygen (A) and salinity (B).

Chapter 6

Conclusion

This dissertation advances understanding of ecosystem resilience and applies several approaches to assessing large changes in aquatic ecosystems. Chapter 2 established expectations for spatial early warning statistics (EWS) for algal blooms, finding that spatial standard deviation and autocorrelation provide early warning even in a model with highly dynamic spatial patterns controlled by physical transport. Chapter 3 tested these findings in a real aquatic ecosystem; spatial EWS did not perform as expected, likely due to the limited spatial heterogeneity observed in the small lakes studied. One temporal EWS, rolling window standard deviation, provided early warning prior to the bloom for most variables in this experiment and successfully differentiated between low and high resilience across multiple lakes and years. Chapter 4 used the same fertilization experiments to evaluate accurate forecasting of bloom timing, a crucial period for potential management actions. This analysis explored the possibility of using flexible models that assimilate information from both model predictions and high frequency observations to improve forecasts. Chapter 5 used a new disturbance detection method to quantify disturbance characteristics for many estuaries and tropical cyclones, revealing patterns and potential drivers of disturbance and recovery. Taken together, these findings illustrate how resilience concepts from theory and modeling studies can be applied to the real world.

Early warning statistics have generated significant excitement and attention due to their potential broad applicability, with the hope that “generic” statistics can be used to quantify resilience and predict changes across systems as diverse as the Earth’s climate, ecosystems, financial systems, and the human body (Scheffer et al. 2009). Numerous modeling studies as well as laboratory experiments (e.g., Drake and Griffen 2010) and analyses of historical data

(e.g., Dakos et al. 2008) have found support for that excitement. However, others have raised valid concerns about EWS, including whether signals of changing resilience can be separated from other sources of variability (Perretti and Munch 2012), if case study selection biases findings (Boettiger and Hastings 2012), and the prevalence of the mechanism underlying EWS theory (Spears et al. 2017). Ecosystem experiments, while often logistically difficult to perform, are strong tests as they are both carried out at the same scale that EWS would be applied at and can be designed to create changes in resilience. This dissertation builds upon previous whole-lake experiments that showed EWS do change in expected ways prior to algal bloom regime shifts (Pace et al. 2017, Wilkinson et al. 2018) by showing that temporal EWS (specifically, rolling window standard deviation) perform better than spatial EWS and can distinguish between high and low resilience states to provide management-relevant performance (Chapters 2 and 3).

The health and economic impacts of algal blooms have also stimulated extensive study of directly forecasting blooms over time periods from days to weeks. Doing so is difficult due to the diversity of aquatic ecosystems and varied drivers that can promote or suppress blooms (Isles and Pomati 2021, Paerl et al. 2001). However, data on algal biomass and driver variables is increasingly available for forecasting (Bertone et al. 2018), and a number of forecasting methods have been developed (Rousso et al. 2020). Bloom initiation is often a short, but important, period for taking management actions to minimize negative bloom impact. Accurately predicting bloom timing is difficult but possible with flexible modeling approaches like data assimilation and high frequency observations (Chapter 4). Due to the importance of bloom initiation and its potential underweighting in traditional model evaluation metrics, future forecasting studies should explicitly consider errors in bloom timing forecasts during model development. Further progress on the mechanisms for and best indicators of bloom initiation is also critical.

Disturbances are ubiquitous in ecosystems, and how a system responds to disturbance events depends on and can affect its resilience. Tropical cyclones cause disturbances in estuaries that can negatively impact these ecosystems and the services they provide. This dissertation builds on prior studies of tropical cyclone disturbances that have generally used one or a few locations and storms to elucidate disturbance mechanisms and their effects (e.g., Paerl et al. 2001, Wetz and Paerl 2008, Patrick et al. 2020). Utilizing a new method that is able to systematically identify disturbances across many tropical cyclones and estuaries allowed disturbance characteristics to be robustly quantified and tied to properties of storms and the locations they impact. Across many storm-estuary combinations salinity and dissolved oxygen recovered in estuaries usually within days to weeks (Chapter 5). However, periods of anomalous conditions are sufficiently long to stress both mobile and stationary organisms. These and future insights into tropical cyclone-caused disturbances will be crucial to maintaining and building estuarine resilience in light of projected increases in tropical cyclone severity (Knutson et al. 2010).

This dissertation takes important steps towards operationalizing resilience and applying it in the real world to improve management. The underlying motivation for studying resilience concepts is to maintain or improve the state of a system, but there are several hurdles between developing conceptual or mathematical models and putting them into practice. All four chapters of this dissertation were conducted at the ecosystem scale, providing a more realistic test than laboratory or smaller scale field experiments. The three chapters on algal blooms (Chapters 2 – 4) utilized whole-ecosystem experiments, allowing us to test methods for prediction and forecasting in cases where a change was expected with a high degree of confidence. However, fully testing and operationalizing these methods in non-experimental settings, where bloom

drivers are less likely to be known and accurately quantified, will require additional research before these approaches can be used effectively to inform management decisions.

The failure of spatial EWS to predict the experimental algal blooms (Chapter 3) despite model predictions to the contrary (Chapter 2) deserves further investigation. It's possible that spatial EWS may work better in larger or more complex systems that have more spatial heterogeneity; indeed a recent study (Ortiz and Wilkinson 2021) documented high and changing spatial variability across different bloom stages in a 40 ha lake in Iowa, USA (20-40 times the size of lakes considered in Chapter 3). In experimental lakes study of this dissertation, temporal EWS performed better than spatial statistics. However, temporal EWS methods require further development to operationalize them for application in scenarios in which reference systems are not available. If available, historic data from when a system did/did not subsequently experience a bloom could be used to define the distribution of temporal EWS in a low and high resilience states. Alternatively, trends or patterns in temporal EWS may be useful for distinguishing changes (Ortiz et al. 2020). For direct bloom forecasting, the accuracy of bloom timing forecasts should be determined quantitatively for both process-based and data-driven models (e.g., Recknagel et al. 2017, Page et al. 2018) and conclusions on the best-performing models compared those based on common evaluation metrics like RMSE and R^2 (Chapter 4). Both EWS and direct forecasting methods also need to be incorporated into a framework that includes management actions to determine if warnings come early enough for effective action (Biggs et al. 2009).

For the resilience of estuaries to tropical cyclones (Chapter 5), studying naturally occurring disturbances as opposed to experimentally caused ones is an important step toward operationalizing this method, as these are the same ecosystems, disturbances, and data that would

be used to understand and build resilience in practice. While the current method and results quantify disturbance characteristics at a much broader scale than has previously been done, additional work is needed to fully understand the drivers of estuarine resilience and use that understanding to improve management. The storm and location properties related to disturbance characteristics in Chapter 5 are a good starting point, but other, more detailed properties have the opportunity to provide deeper understanding and actionable insights, such as: which tide a storm impacts a location on; natural (e.g., marsh or SAV) or man-made (e.g., engineered structures like seawalls) characteristics; the degree of eutrophication; and watershed characteristics (e.g., total area, proportion covered by development, agriculture, and forested/natural landscapes). Finally, Chapter 5 can also be extended by analyzing additional variables that can be measured at high frequency (e.g., turbidity, chlorophyll-a) to build a more complete picture of estuarine disturbance and controls of resilience.

Resilience as a concept has been questioned as being too vague and multipurpose to be useful (Brand and Jax 2007). However, resilience can be defined for specific situations, given context, and operationalized (Quinlan et al. 2015). This dissertation evaluates several specific approaches to measuring aspects of ecosystem resilience. Expansion, elaboration, and application of resilience methodologies needs to continue to be a research priority and is especially critical given rapid local, regional, and global environmental changes.

References

Bertone, E, MA Buford, DP Hamilton. 2018. Fluorescence probes for real-time remote cyanobacteria monitoring: A review of challenges and opportunities. *Water Research*. 141: 152-162.

- Biggs, R, SR Carpenter, WA Brock. 2009. Turning back from the brink: Detecting an impending regime shift in time to avert it. *Proc. Nat. Acad. Sci. U.S.A.* 106(3): 826–831.
- Boettiger, C, and A Hastings. 2012. Early warning signals and the prosecutor’s fallacy. *Proc. R. Soc. B*, 279: 4734–4739.
- Brand, FS, K Jax. 2007. Focusing the meaning(s) of resilience: resilience as a descriptive concept and boundary object. *Ecology and Society*. 12(1): 23.
- Dakos, V, M Scheffer, EH van Nes, V Brovkin, V Petoukhov, H Held. 2008. Slowing down as an early warning signal for abrupt climate change. *Proc. Nat. Acad. Sci. U.S.A.* 105(38): 14308 - 14312.
- Drake JM and BD Griffen. 2010. Early warning signals of extinction in deteriorating environments. *Nature*. 467(7314): 456-459.
- Isles, PDF, F Pomati. 2021. An operational framework for defining and forecasting phytoplankton blooms. *Front. Ecol. Environ.* doi:10.1002/fee.2376.
- Knutson, TR, JL McBride, J Chan, K Emanuel, G Holland, C Landsea, I Held, JP Kossin, AK Srivastava, M Sugi. 2010. Tropical cyclones and climate change. *Nature Geoscience*. 3: 157-163.
- Ortiz, D, J Palmer, G Wilkinson. 2020. Detecting changes in statistical indicators of resilience prior to algal blooms in shallow eutrophic lakes. *Ecosphere*. 11(10):e03200.
- Ortiz, DA, GM Wilkinson. 2021. Capturing the spatial variability of algal bloom development in a shallow temperate lake. *Freshwater Biology*. <https://doi.org/10.1111/fwb.13814>
- Pace, ML, RD Batt, CD Buelo, SR Carpenter, JJ Cole, JT Kurtzweil, and GM Wilkinson. 2017. Reversal of a cyanobacterial bloom in response to early warnings. *Proc. Nat. Acad. Sci. U.S.A.* 114(2): 352-357.
- Paerl, HW, RS Fulton, PH Moisander, and J Dyble. 2001. Harmful freshwater algal blooms, with an emphasis on cyanobacteria. *Sci. World*. 1:76–113.
- Paerl, HW, JD Bales, LW Ausley, CP Buzzelli, LB Crowder, LA Eby, JM Fear, M Go, BJ Peierls, TL Richardson, JS Ramus. 2001. Ecosystem impacts of three sequential hurricanes (Dennis, Floyd, and Irene) on the United States’ largest lagoonal estuary, Pamlico Sound, NC. *Proceedings of the National Academy of Sciences, USA*. 98(10): 5655-5660.
- Page, T, PJ Smith, KJ Beven, ID Jones, JA Elliott, SC Maberly, EB Mackay, M De Ville, H. Feuchtmayr. 2018. Adaptive forecasting of phytoplankton communities. *Water Research*. 134: 74-85.
- Patrick CJ, L Yeager, AR Armitage, F Carvallo, VM Congdon, KH Dunton, M Fisher, AK Hardison, JD Hogan, J Hosen, X Hu, B Kiel Reese, S Kinard, JS Kominoski, X Lin, Z

- Liu, PA Montagna, SC Pennings, L Walker, CA Weaver, M Wetz. 2020. A System Level Analysis of Coastal Ecosystem Responses to Hurricane Impacts. *Estuaries and Coasts*. 43: 943-959.
- Perretti, CT, SB Munch. 2012. Regime shift indicators fail under noise levels commonly observed in ecological systems. *Ecological Applications*. 22(6): 1772-1779.
- Quinlan, AE, M Berbes-Blazquez, LJ Haider, GD Peterson. 2016. Measuring and assessing resilience: broadening understanding through multiple disciplinary perspectives. *Journal of Applied Ecology*. 53: 677-687.
- Recknagel, F, PT Orr, M Bartkow, A Swanepoel, H Cao. 2017. Early warning of limit-exceeding concentrations of cyanobacteria and cyanotoxins in drinking water reservoirs by inferential modelling. *Harmful Algae*. 69: 18-27.
- Rouso, BZ, E Bertone, R Stewart, DP Hamilton. 2020. A systematic literature review of forecasting and predictive models for cyanobacteria blooms in freshwater lakes. *Water Research*. 182: 115959.
- Scheffer, M., J. Bascompte, W. A. Brock, V. Brovkin, S. R. Carpenter, V. Dakos, H. Held, E. H. van Ness, M. Rietkerk, and G. Sugihara. 2009. Early-warning signals for critical transitions. *Nature*. 461: 53–59.
- Spears, BM, MN Futter, E Jeppesen, BJ Huser, S Ives, TA Davidson, R Adrian, DG Angeler, SJ Burthe, L Carvalho, F Daunt, AS Gsell, DO Hessen, ABG Janssen, EB Mackay, L May, H Moorhouse, S Olsen, M Søndergaard, H Woods, SJ Thackeray. 2017. Ecological resilience in lakes and the conjunction fallacy. *Nature Ecology and Evolution*. 1(11): 1616–1624.
- Wetz, MS, HW Paerl. 2008. Estuarine Phytoplankton Responses to Hurricanes and Tropical Storms with Different Characteristics (Trajectory, Rainfall, Winds). *Estuaries and Coasts*. 31: 419-429.
- Wilkinson, GM, SR Carpenter, JJ Cole, ML Pace, RD Batt, CD Buelo, and JT Kurtzweil. 2018. Early warning signals precede cyanobacterial blooms in multiple whole-lake experiments. *Ecol. Monogr*. 88(2): 188–203.

Appendix 1: Supplemental Information for Chapter 2

Model description

We describe the model of Serizawa et al. (2008) and modifications below, and refer readers to the original manuscript for additional details.

The full spatial model consists of a grid of 180x180 cells, each governed by a pair of partial differential equations representing two dynamic variables, the single nutrient (n , parameterized for phosphorus) and phytoplankton (p):

$$\partial p = \left[d\nabla^2 p - (\nabla \cdot vp) + \frac{n}{1+n}p - f_p \frac{p}{1+p} \right] \partial t + p\sigma dW \quad (1)$$

$$\partial n = \left[d\nabla^2 n - (\nabla \cdot vn) + i_n - a \frac{n}{1+n}p - m_n n \right] \partial t \quad (2)$$

Parameter values are from Table 2, Set I in Serizawa et al. (2008), and given below in Table A1.1.

In both (1) and (2), the first term inside the brackets represents diffusion: d is the diffusion coefficient and ∇^2 is the Laplacian operator (divergence of the gradient of p or n). In (1) and (2), the second terms represent advection, where v is the local velocity of the velocity field created by 100 eddies randomly seeded on the grid domain (see equation 7 in Serizawa et al. 2008).

The third term in (1) represents growth of phytoplankton from nutrient uptake, while the fourth represents grazing by zooplankton. Note that f_p is fixed (not dynamic), and so zooplankton density is constant. The final term in (1) was added in this study to represent small environmental fluctuations. It is a Weiner stochastic process with standard deviation $\rho = 0.01$. For the deterministic portions of the simulations, this term is set to zero. The long-run deterministic system state ($t=1000$) is used as the starting system state for the stochastic portions of simulations.

The third term in (2) is the input rate of phosphorus, which is constant for a given simulation. The fourth term represents the uptake rate of phosphorus by phytoplankton. The fifth term represents non-phytoplankton phosphorus losses from processes like sedimentation or outflow.

As in Serizawa et al. (2008), we start our simulations from initial distributions where both p and n vary around the fixed point for a given phosphorus input level, and gradients in the two variables are perpendicular to each other (see equations 10 and 11 and Figure 5a from Serizawa et al. 2008). The initial distribution scale parameter, A , determines the range of initial values around the fixed point at the start; we use the same value ($A = 1$) as Serizawa et al. (2008) do for their base case. Serizawa et al. (2008) show that changes in A for the same nutrient input rate create different specific patterns at a given time, but do not effect pattern formation (see Serizawa et al. 2008 Figure 6).

The model in (1) and (2) is nondimensional, meaning that the parameters above and in Tables A1.1 are unitless. Serizawa et al. (2008) gives a dimensional version of the model in equations (1) and (2) of that manuscript. Serizawa et al. (2008) provide equations to convert between the dimensional variables and parameters and their nondimensional equivalents in equations (3) and (4), which are also reproduced below:

$$t = \mu T \quad , \quad n = \frac{N}{H_N} \quad , \quad p = \frac{P}{H_P} \quad (3)$$

$$i_n = \frac{I_N}{\mu * H_N} \quad , \quad a = \frac{k H_P}{H_N} \quad , \quad m_n = \frac{m_N}{\mu} \quad , \quad f_p = \frac{f_P}{\mu H_P} \quad (4)$$

Variable and parameter meanings and units are presented in Tables A1.1 and A1.2 below.

Serizawa et al. (2008) also provide dimensional parameter values in Table 1 along with parameter value ranges from the literature. Like Serizawa et al. (2008), we have chosen to use the nondimensional version for our simulation to minimize the number of parameters and

because we are interested in differences in the spatial indicators as the model transitions between blooms states, and not the specific parameter values themselves. The use of non-dimensional models and parameters is standard for stability analysis of dynamic models in ecology and other disciplines (Hastings & Gross 2012).

Because this model is nondimensional, the spatial components (extent, resolution, diffusion and advection) of the grid are defined relative to each other, as is the temporal component of the model. For the parameter values used by Serizawa et al. (2008) and in this study, the time step of 0.025 corresponds to 1.2 hours and we calculate the spatial indicators every 1 time unit (unitless), which corresponds to 2 days. As both the temporal and spatial parameters can vary over several orders of magnitude (Table 1 and Section 4.1 in Serizawa et al. 2008), future work is needed to identify the specific time and spatial scales at which spatial patterns and derived spatial indicators occur using spatiotemporal data from individual aquatic ecosystems. Our conclusions do not depend on the specific time scale as simulations cover several complete cycles in spatial patterns and statistics (Chapter 2, Figure 2.2).

References

- Hastings, AM and L Gross (editors). 2012. *Encyclopedia of Theoretical Ecology*. Univ. California Press, Berkeley CA. 848 pages.
- Serizawa, H, T Amemiya, K Itoh. 2008. Patchiness in a minimal nutrient-phytoplankton model. *Journal of Biosciences*, 33(3): 391-403.

Table A1.1. Nondimensional model parameters and values used in this study.

Parameter	Symbol	Value(s)
Scaled nutrient input rate (phosphorus)	i_n	0.3 – 1.8
Scaled nutrient conc. in phytoplankton	A	8
Scaled nutrient loss rate	m_n	0.03
Scaled zooplankton feeding rate	f_p	0.9
Diffusion coefficient	D	0.04
Maximum velocity	v_{max}	0.3
Length of half grid domain	L	100
Initial distribution scale factor	A	1
SD of environmental noise	σ	0.01

Table A1.2. Variables and parameters from Equations 3 and 4 above.

Parameter	Symbol	Units
Nondimensional time	t	NA
Dimensional time	T	Days
Phytoplankton growth rate	μ	days ⁻¹
Nondimensional nutrient concentration (phosphorus)	n	NA
Dimensional nutrient concentration (phosphorus)	N	g/m ²
Nutrient half-saturation constant	H_N	g/m ²
Nondimensional phytoplankton concentration	p	NA
Dimensional phytoplankton concentration	P	g/m ²
Phytoplankton half-saturation constant	H_P	g/m ²
Dimensional nutrient input rate (phosphorus)	I_N	g/m ² /day
Nutrient content in phytoplankton	k	NA
Nondimensional loss rate of nutrient	m_n	NA
Dimensional loss rate of nutrient	m_N	day ⁻¹
Nondimensional zooplankton feeding rate	f_p	NA
Dimensional zooplankton feeding rate	f_P	g/m ² /day

Appendix 2: Supplemental Information for Chapter 3

Quickest Detection Method Details

The Quickest Detection method, also called the Shiryaev-Roberts procedure, is an ‘online’ method that updates with each newly collected data point. It is based on the Shiryaev-Roberts (S-R) statistic, which accumulates evidence as to whether an event has occurred based on the likelihood ratio of two models (Polunchenko and Tartakovsky 2012). For our purposes, an event corresponds to a shift from a baseline (high resilience) state to an alarm (low resilience) state, with each state defined by a distribution of rolling window early warning statistics (Carpenter et al. 2014). When the S-R statistic crosses a threshold, it triggers an alarm indicating it is likely the event has occurred and the system has changed state. After an alarm is triggered, the S-R statistic resets and is able to accumulate more evidence and trigger additional alarms. The alarm response is described in equations 1 - 3:

$$R_t = \begin{cases} (1 + R_{t-1}) * \Lambda_t, & R_{t-1} < A \\ \Lambda_t, & R_{t-1} \geq A \end{cases} \quad (1)$$

$$\Lambda_t = \frac{g(x_t)}{f(x_t)} \quad (2)$$

$$I_t = \begin{cases} 0, & R_t < A \\ 1, & R_t \geq A \end{cases} \quad (3)$$

where R_t is the Shiryaev-Roberts statistic at time t , A is the alarm threshold, Λ_t is the likelihood ratio of the alarm state model $g()$ and baseline state model $f()$ evaluated at the data point x_t , and I_t is an indicator for whether an alarm has been triggered at time t . The alarm threshold A can be set based on the user’s tolerance for false alarms assuming the input data meet certain statistical assumptions (Pollak and Tartakovsky 2009); in practice we have found alarm timing to be insensitive across broad ranges of A (Carpenter et al. 2014) and use $A = 10^7$ (Wilkinson et al. 2018).

Baseline and alarm models

The data points x_t are the EWS time series, rolling window standard deviation (SD) and autocorrelation (AR(1)), observed in the experimental lake. The models for the baseline $f()$ and alarm $g()$ states are determined by the reference lake EWS and expectations from theory, respectively, and are statistic specific. For both SD and AR(1), the baseline state model $f()$ is centered at the observed EWS value in the reference lake to control for variation in EWS not caused by the manipulation in the experimental lake (storms, seasonal trends, etc).

SD

For rolling window SD, both $f()$ and $g()$ are normal distributions:

$$f(x_t) \sim N(\mu_{baseline}, \sigma_{pool}) \quad (4)$$

$$g(x_t) \sim N(\mu_{alarm}, \sigma_{pool}) \quad (5)$$

where x_t is the observed rolling window SD in the experimental lake, $\mu_{baseline}$ is the observed rolling window SD in the reference lake and $\mu_{alarm} = \mu_{baseline} + 2 * \sigma_{pool}$. σ_{pool} is the pooled standard deviation of rolling window SD in both lakes as in Wilkinson et al. 2018 (Supplemental Information):

$$\sigma_{pool} = \sqrt{\sigma_{rw,exp}^2 + \sigma_{rw,ref}^2} \quad (6)$$

where $\sigma_{rw,exp}$ and $\sigma_{rw,ref}$ are the standard deviations of the observed rolling window standard deviations, sd_{rw} in the experimental and reference lakes, respectively:

$$\sigma_{rw,i} = sd(rolling\ window\ SD_i) = sd(x_{t,i}) = \sqrt{1 \frac{1}{4 * x_{t,i}^2} * x_{t,i}^4 * (\frac{2}{N-1} + \frac{\kappa}{N})} \quad (7)$$

where x_t is the observed rolling window SD, N is the sample size (rolling window width), and κ is sample kurtosis.

AR(1)

For rolling window AR(1), we use the exact probability distribution for Pearson sample correlation coefficient:

$$f(r, \rho, N) = \frac{(N-2)*\Gamma(N-1)*(1-\rho^2)^{\frac{(N-1)}{2}}*(1-r^2)^{\frac{(N-4)}{2}}}{\sqrt{2\pi}*\Gamma(N-\frac{1}{2})*(1-\rho r)^{N-\frac{3}{2}}} * {}_2F_1\left(\frac{1}{2}, \frac{1}{2}; \frac{1}{2}(2N-1); \frac{1}{2}(\rho r + 1)\right) \quad (8)$$

where r is the observed (sample) rolling window AR(1) coefficient, ρ is the population correlation coefficient, N is sample size (rolling window width), $\Gamma()$ is the gamma function, and ${}_2F_1()$ is the Gaussian hypergeometric function. Note $f(r, \rho, N)$ in equation 8 refers generically to the probability density function for correlation coefficient and not the baseline state; below the baseline state likelihood model is denoted $f_{baseline}()$ to differentiate:

$$f_{baseline}(x_t) = f(x_t, \rho_{baseline}, N) \quad (9)$$

$$g(x_t) = f(x_t, \rho_{alarm}, N) \quad (10)$$

where x_t is the observed rolling window AR(1) in the experimental lake at time t , $\rho_{baseline,t}$ is the observed rolling window AR(1) in the reference lake, and ρ_{alarm} is the “true” correlation coefficient of the alarm state. We use $\rho_{alarm} = 0.95$ based on theoretical expectations that AR(1) should increase towards 1 as a critical transition is approached (Dakos et al. 2012). The above formulation for AR(1) quickest detection alarms differs from previous applications, which used normal distributions for $f()$ and $g()$ with $\mu_{baseline} = x_{t,ref}$ and $\mu_{alarm} = 1$ and either fixed σ values (Carpenter et al. 2014) or time-varying σ derived from first-order error propagation (Wilkinson et al. 2018). While qualitatively very similar ($f(x_t)$ centered on the observed autocorrelation in the reference lake, $g(x_t)$ at/near the theoretical value of 1 at the critical point), the new “exact” formulation presented here represents an improvement as the probability density function of $f()$ and $g()$ lies entirely within the domain of the Pearson correlation coefficient (-1 to 1) and avoids

erroneous alarms that occur when experimental lake AR(1) is less than reference lake AR(1) (see Appendix S1 and Figure S3 of Wilkinson et al. 2018 Supporting Information).

References

- Polunchenko, A and AG Tartakovsky. 2012. State of the art in sequential change-point detection. *Meth. Comput. Appl. Probabil.* 14: 649-683.
- Carpenter, SR, WA Brock, JJ Cole, ML Pace. 2014. A new approach for rapid detection of nearby threshold in ecosystem time series. *Oikos* 123(3): 290-297.
- Pollak, M and AG Tartakovsky. 2009. Optimality of the Shiryaev-Roberts procedure. *Stat. Sin.* 19: 1729–1739.
- Wilkinson, GM, SR Carpenter, JJ Cole, ML Pace, RD Batt, CD Buelo, JT Kurtzweil. 2018. Early warning signals precede cyanobacterial blooms in multiple whole-lake experiments. *Ecol. Monogr.* 88(2): 188–203.
- Dakos, V, EH van Nes, P D’Odorico, M Scheffer. 2012. Robustness of variance and autocorrelation as indicators of critical slowing down. *Ecology* 93(2): 264–271

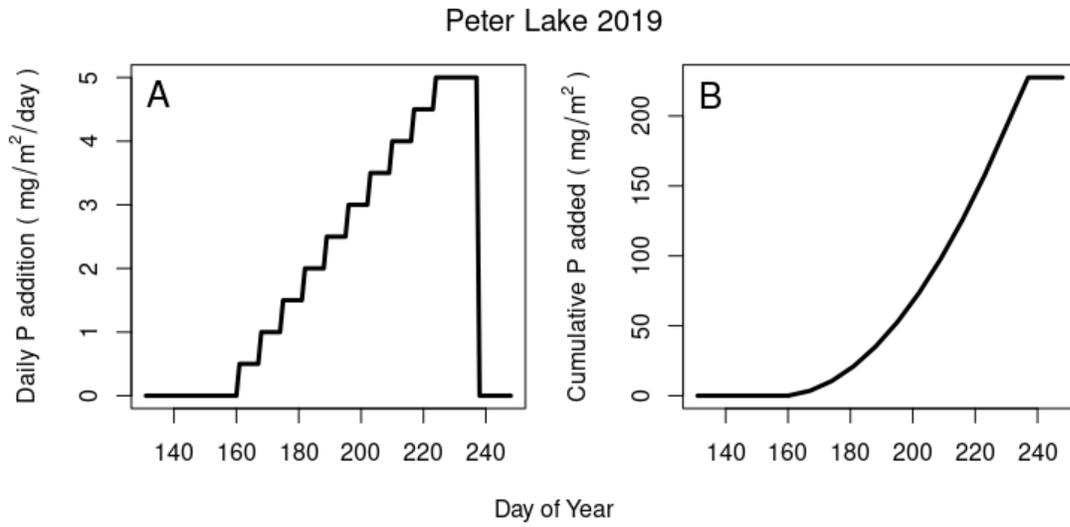


Figure A2.1. Daily (A) and cumulative (B) phosphorus added to Peter Lake in 2019. Nitrogen additions followed the same pattern at 15:1 N:P molar ratio.

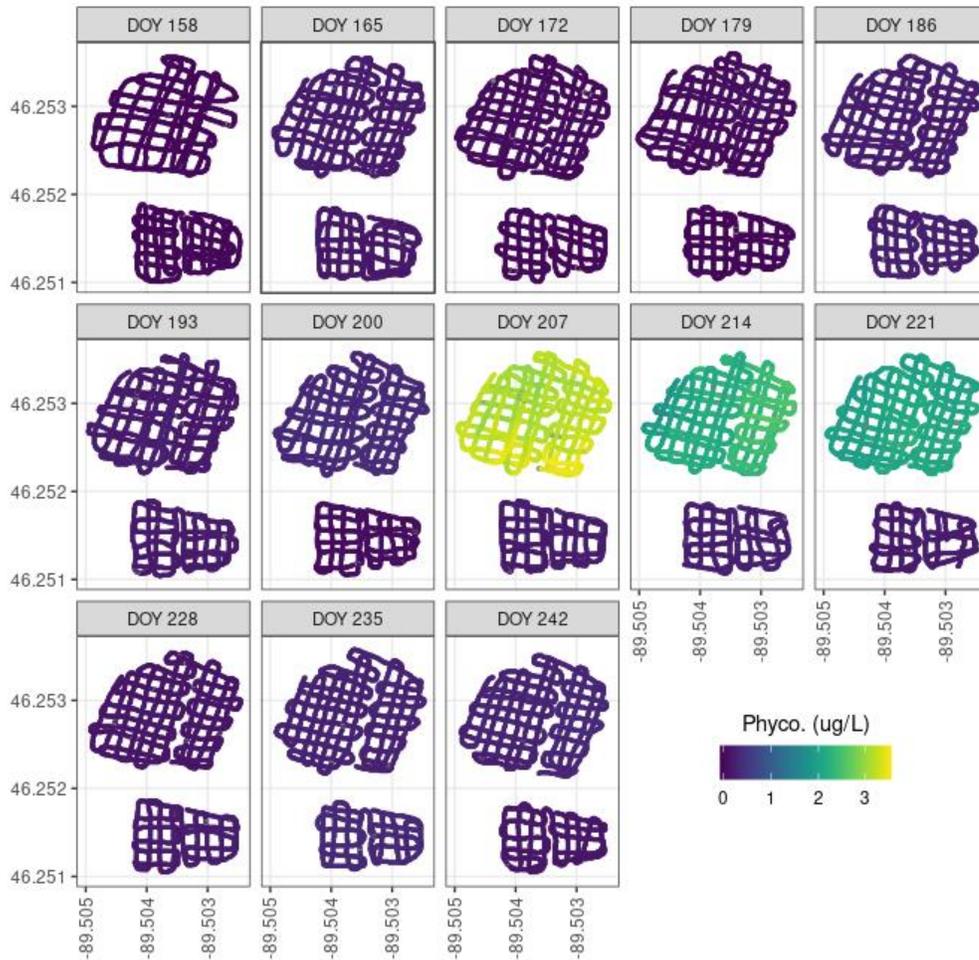


Figure A2.2. Spatial measurements of phycocyanin in Peter (top basin) and Paul (bottom basin) lakes every 7 days in 2019.

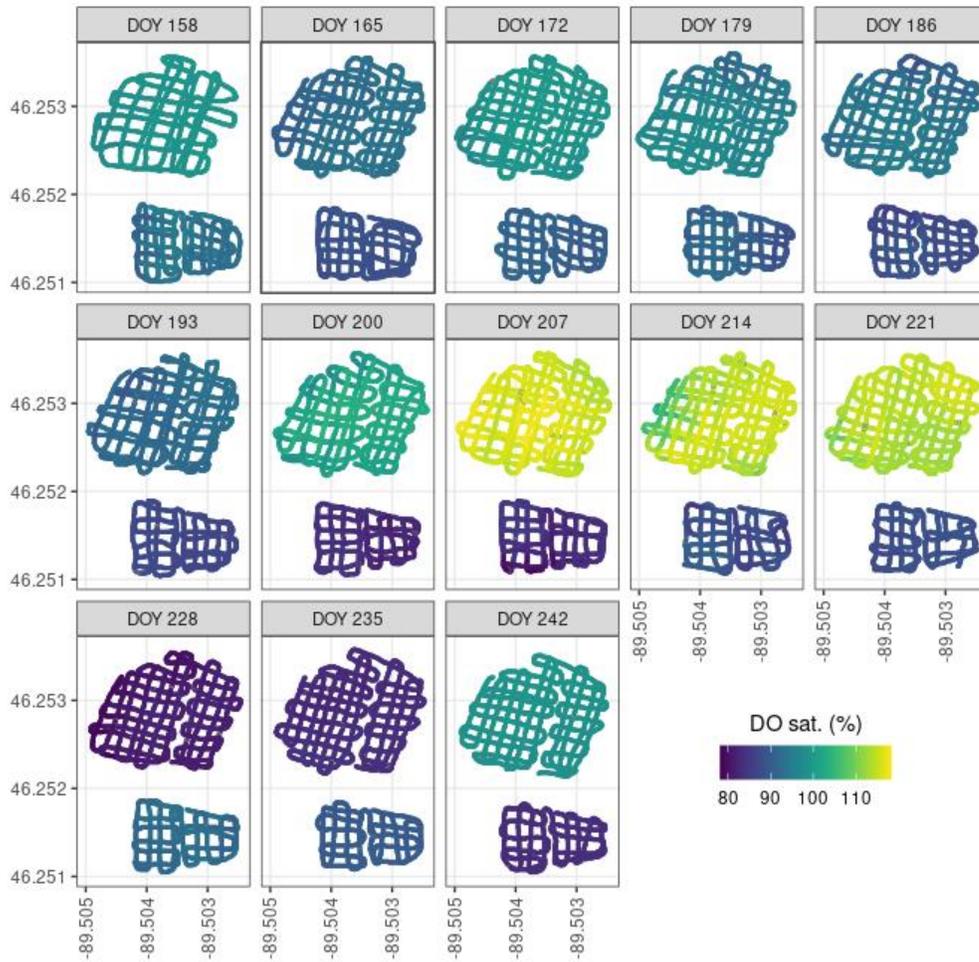


Figure A2.3. Spatial measurements of dissolved oxygen in Peter (top basin) and Paul (bottom basin) lakes every 7 days in 2019.

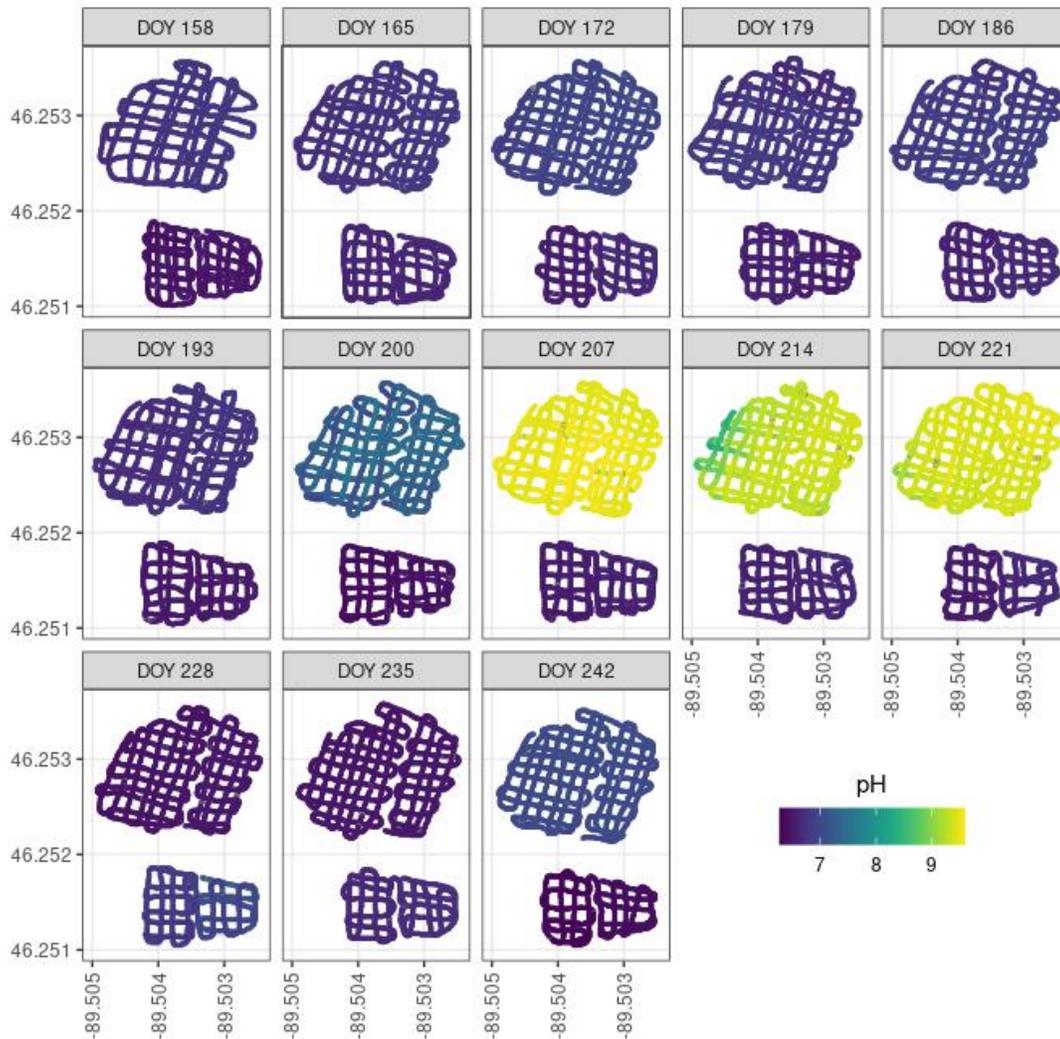


Figure A2.4. Spatial measurements of pH in Peter (top basin) and Paul (bottom basin) lakes every 7 days in 2019.

Appendix 3: Supplemental Information for Chapter 4

Full model description

The forecasting model is an ensemble Kalman filter (EnKF), a data assimilation technique that iteratively compares model predictions to observations to produce updated estimates of system state, which can then be used to forecast model state into the future (Evensen 2009). We use a version of EnKF, EnKF with parameter augmentation, to fit model parameters in addition to estimating model state (Zhang et al. 2017). Notation below generally follows that of Thomas et al. 2020. The model has one state variable, the concentration of chlorophyll-a, which is a proxy for algal biomass and corresponds directly to observed concentrations measured daily. Three parameters are fit that determine the relationship between color-normalized cumulative phosphorus loading (CNCP loading; units mg P / m² / g440, where g440 is the absorbance at 440 nm with units m⁻¹) on a given day and the instantaneous “carrying capacity” K_t of the system (units ug/L): B_t , Lh_t , and A_t . B_t is the baseline chlorophyll-a carrying capacity at zero CNCP loading, Lh_t is the loading half-constant at which there is a rapid increase in chlorophyll with increasing CNCP loading, and A_t is how much higher than B_t the carrying capacity asymptote is at high CNCP loading:

$$K_t = B_t + A_t * \frac{L_t^k}{Lh_t^k + L_t^k} \quad (1)$$

where L_t is the CNCP loading on a given day, and k is a lake-specific sigmoid constant determining the steepness of K_t at Lh_t (see Chapter 4 Figure 4.1).

The EnKF consists of N ensemble members (we use $N = 300$) with different values of state and parameter values. State and parameter values were initialized in the following manner: the state variable (chlorophyll-a concentration, Chl_t) and B_t were initialized at $t = 1$ with a uniform distribution ranging from 90% to 110% of the first observed chlorophyll concentration

and standard deviation equal to 10% of the mean; A_t was initialized with a uniform distribution ranging from 90% to 110% of the lake-specific asymptotic value fit from chlorophyll observations up to bloom peak (Chapter 4 Figure 4.1); and Lh_t was initialized with a uniform distribution from 90% to 110% of the specified Lh_{t0} for a given analysis.

The EnKF carries out data assimilation in two steps: a forecasting step and an update step. The forecasting step uses the previous state and the current driver values (for our model, CNCP loading) to generate a prediction of the system state at the next time point using a logistic-type response:

$$Chl_{i,t+1} = F(Chl_{i,t}, L_{t+1}) = Chl_{i,t} * \left[1 + \left(1 - \frac{Chl_{i,t}}{K_{i,t+1}} \right) \right] + \varepsilon \quad (2)$$

Where ε represents model process error. For our model there is only one state variable so ε is a random draw from normal distribution with mean 0 and variance determined by the difference between predicts and observation; in cases with more than one state variable ε is a draw from a multivariate normal distribution.

In the update step, predictions from the forecast step are compared to the next observation and used to adjust the system state and parameter values:

$$x_t^i = x_t^{i-} + K_x(\widehat{y}_t^i - Hx_t^{i-}) \quad (3)$$

$$\alpha_t^i = \alpha_t^{i-} + K_\alpha(\widehat{y}_t^i - Hx_t^{i-}) \quad (4)$$

Where x is the system state variable (chlorophyll-a concentrations) and α corresponds to each of the model parameters (B_t , A_t , Lh_t). The superscript i corresponds to each ensemble member after the update step, while i^- is the corresponding value before the update (after the forecast step).

Using $\widehat{y}_t^i = y_t + \varepsilon$ quantifies observation error, where y_t is the observed chlorophyll concentration at time t and ε is a random draw from a normal distribution for each ensemble member, with mean 0 and standard deviation $R = 0.05 * y_t$ based on the measured observation error determined

from 5 replicate chlorophyll samples collected in August 2018. H represents an observation matrix, with ones corresponding to observed state variables in x and zeros elsewhere; as our model has only one state variable and observations are available each day it is always [1] and could be dropped; we include it for consistency of notation. K_x and K_α represent the Kalman gain for the state variable and parameters, respectively:

$$K_x = C_{xx}H^T(HC_{xx}H^T + R)^{-1} \quad (5)$$

$$K_\alpha = C_{\alpha x}H^T(HC_{\alpha x}H^T + R)^{-1} \quad (6)$$

Where C_{xx} is the covariance of the state variables (for us variance, as there's only one state variable) across the ensemble after the forecast step, and $C_{\alpha x}$ is the covariance of the parameters with the state variable across the ensemble after the forecast step. Before calculating $C_{\alpha x}$ an inflation factor $IF > 1$ is applied to prevent filter divergence (Evensen 2009):

$$\alpha_t^{i-} = \overline{\alpha_{t-1}} + IF(\alpha_{t-1}^i - \overline{\alpha_{t-1}}) \quad (7)$$

Where $\overline{\alpha_{t-1}}$ is the mean value of a parameter after the update step at time $t - 1$.

After the update step was carried out using the observed chlorophyll concentration on a given day t , forecasts from 1 to 7 days into the future were created using the parameter values on day t and the known (because they were experimentally controlled) future loading values as in equations 2 and 3:

$$Chl_{i,t+h} = Chl_{i,t+h-1} * \left[1 + \left(1 - \frac{Chl_{i,t+h-1}}{K_{i,t+h}} \right) \right] \quad (8)$$

$$K_{i,t+h} = B_{i,t} + A_{i,t} * \frac{L_t^k}{Lh_{i,t}^k + L_{t+h}^k} \quad (9)$$

Where h is the forecast horizon (1 to 7 days).

References

- Evensen, G. 2009. The ensemble Kalman filter for combined state and parameter estimation. *IEEE Control Systems Magazine*. 29,(3): 83-104.
- Thomas, RQ, RJ Figueiredo, V Daneshmand, BJ Bookout, LK Puckett, CC Carey. 2020. A near-term iterative forecasting system successfully predicts reservoir hydrodynamics and partitions uncertainty in real time. *Water Resources Research*, 56, e2019WR026138.
- Zhang, H, HJH Franssen, X Han, JA Vrugt, H Vereecken. 2017. State and parameter estimation of two land surface models using the ensemble Kalman filter and the particle filter. *Hydrol. Earth Syst. Sci.* 21: 4927–4958.

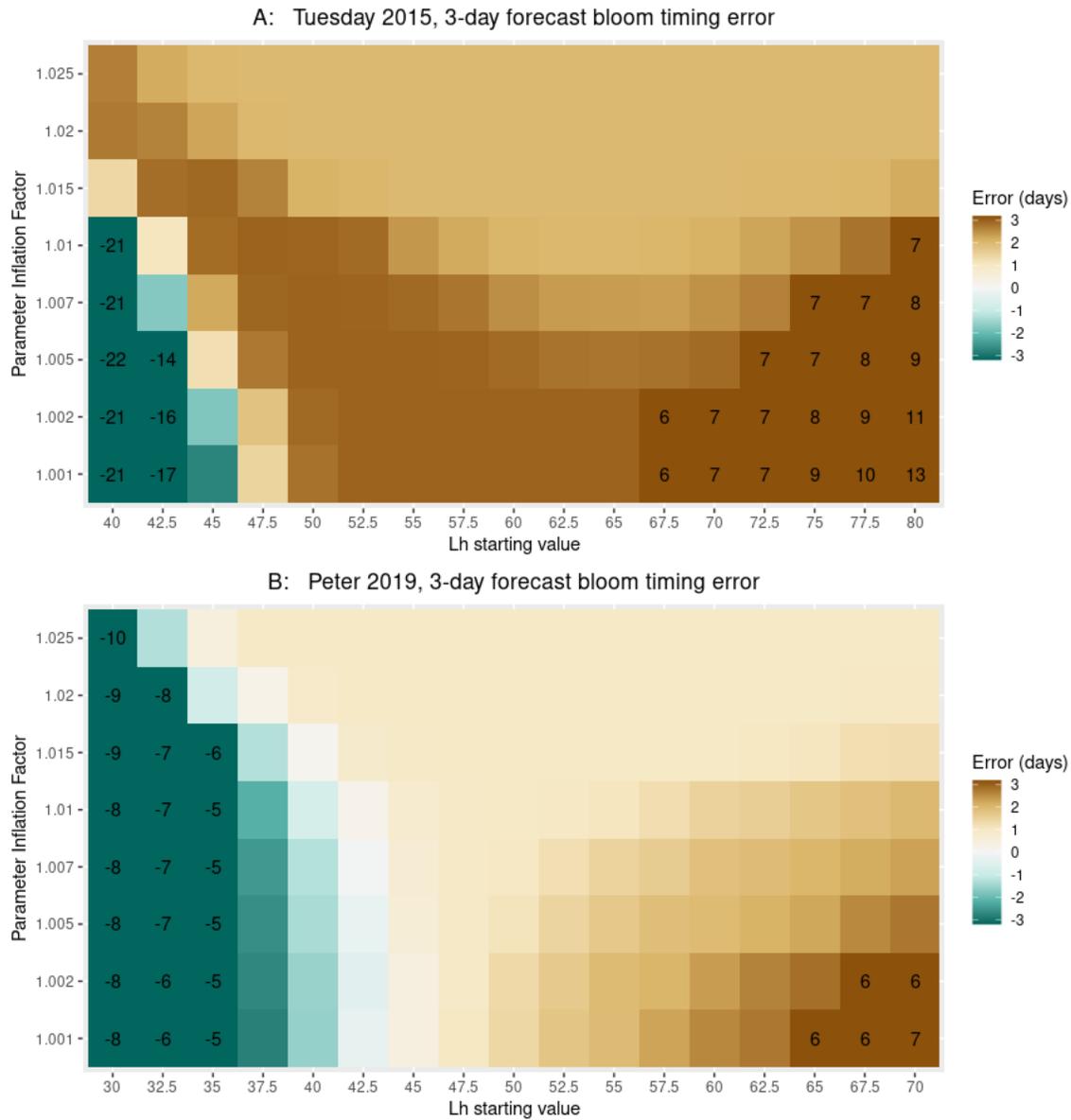


Figure A3.1. Average error in bloom timing for 3-day forecast across 100 replicate simulations for each combination of starting loading half-constant parameter (Lh_i) and parameter inflation factor (IF) for A) Tuesday Lake 2015 and B) Peter Lake 2019. The color scalebar was truncated at ± 3 days to allow for visualization of smaller errors; Lh , IF combinations with < -3 or $> +3$ day errors are printed on the grid, rounded to the nearest whole number.

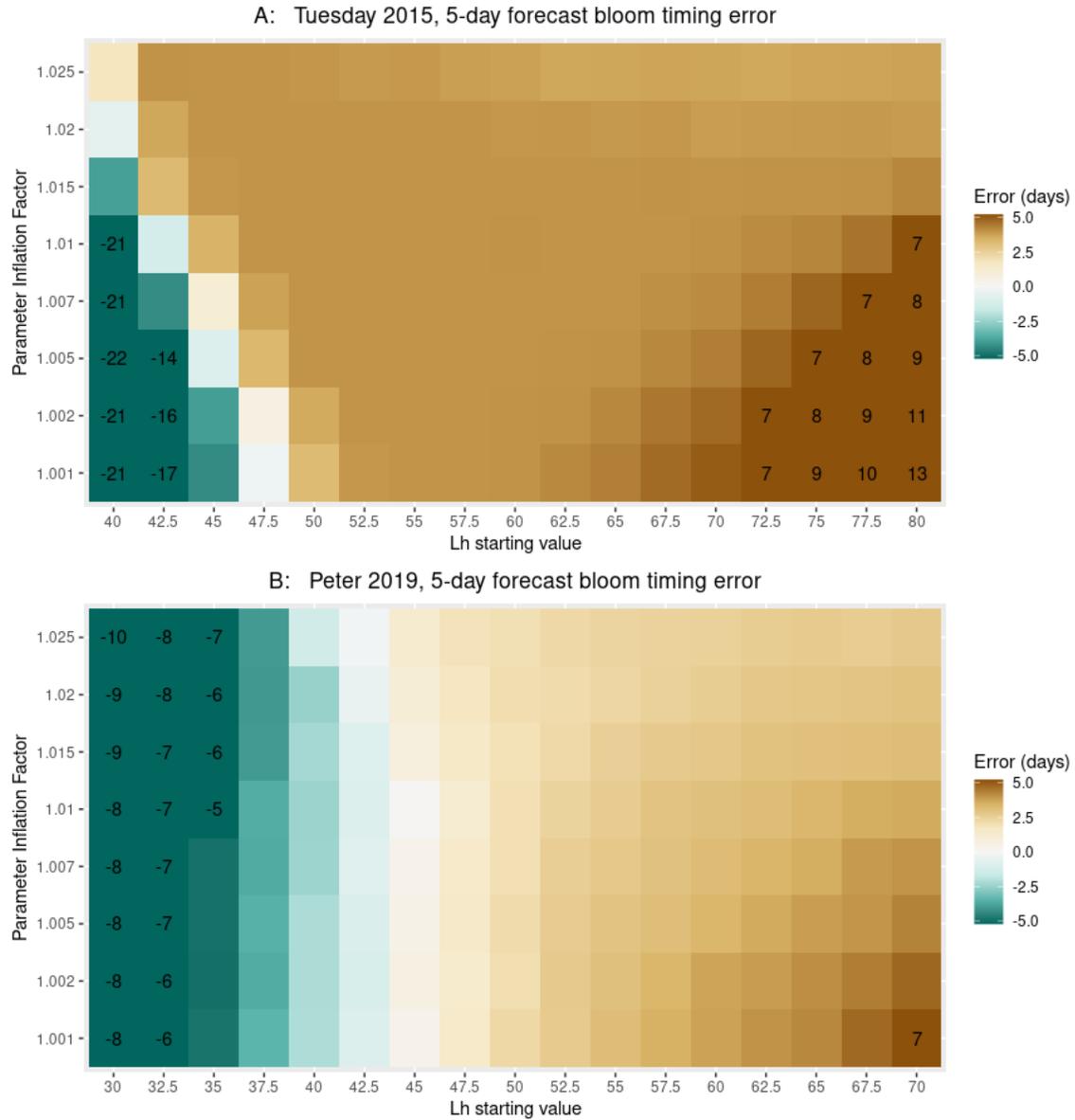


Figure A3.2. Average error in bloom timing for 5-day forecast across 100 replicate simulations for each combination of starting loading half-constant parameter (Lh_i) and parameter inflation factor (IF) for A) Tuesday Lake 2015 and B) Peter Lake 2019. The color scalebar was truncated at ± 5 days to allow for visualization of smaller errors; Lh , IF combinations with < -5 or $> +5$ day errors are printed on the grid, rounded to the nearest whole number.



Figure A3.3. Plots of chlorophyll-a concentration (ug/L) vs. potential bloom driver variables for Peter Lake (left column) and Tuesday Lak (right column). Point color represents different years. Descriptions and units for each variable (from top to bottom) are: color – absorbance at 440nm, m⁻¹; daily phosphorus loading mg P / m² / day; total phosphorus measured at the surface, ug / L; water temperature at a depth of 0.75 meters, degrees Celsius; and mixed layer depth, meters.

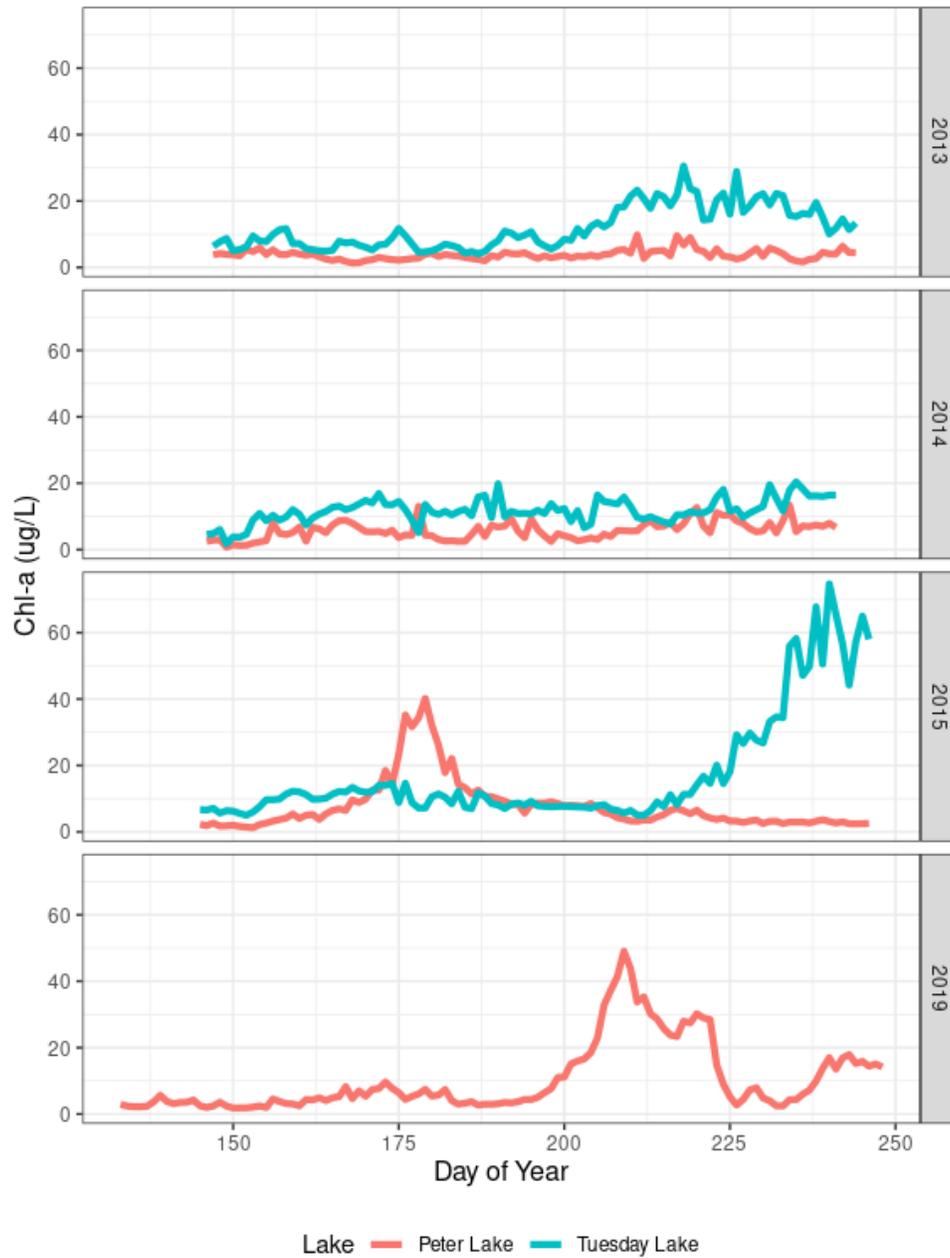


Figure A3.4. Daily time series of chlorophyll-a (ug/L) for all years in which fertilization experiments were done for Peter Lake (red) and Tuesday Lake (blue).

Appendix 4: Supplemental Information for Chapter 5.

Table A4.1. NERRS site ID codes, reserve names, location, and the number of stations and tropical cyclones analyzed for each NERRS site.

NERRS Site ID	Reserve Name	Location	Number of Stations	Number of Storms
ace	Ashepoo Combahee Edisto (ACE) Basin	Bennett's Point, South Carolina	6	7
apa	Apalachicola	Eastpoint, Florida	4	9
cbm	Chesapeake Bay (Maryland)	Washington DC and Baltimore, Maryland	4	7
cbv	Chesapeake Bay (Virginia)	Gloucester Point, Virginia	4	7
del	Delaware	Townsend and Dover, Delaware	4	10
gnd	Grand Bay	Mass Point, Mississippi	4	13
grb	Great Bay	Greenland, New Hampshire	4	5
gtm	Guana Tolomato Mantanzas	Ponte Vedra Beach, Florida	4	18
hud	Hudson River	Staatsburg, New York	3	1
jac	Jacques Cousteau	Tuckerton, New Jersey	4	12
mar	Mission-Aransas	Port Aransas, Texas	4	2
nar	Narragansett Bay	Prudence Island, Rhode Island	4	4
niw	North Inlet-Winyah Bay	Georgetown, South Carolina	6	12
noc	North Carolina	Corolla, Beaufort, and Wilmington, North Carolina	4	10
rkb	Rookery Bay	Naples, Florida	4	9
sap	Sapelo Island	Sapelo Island, Georgia	4	13
wel	Wells	Wells, Maine	4	5
wkb	Weeks Bay	Fairhope, Alabama	4	14
wqb	Waquoit Bay	Falmouth, Massachusetts	4	6