



Michigan Technological University  
*Create the Future* Digital Commons @ Michigan Tech

---

Dissertations, Master's Theses and Master's  
Reports - Open

Dissertations, Master's Theses and Master's  
Reports

---

2010

## IMPROVING THE RELIABILITY OF AQUATIC BIOGEOCHEMICAL MODELS: INTEGRATING INFORMATION AND OPTIMIZING COMPLEXITY

Cory P. McDonald  
*Michigan Technological University*

Follow this and additional works at: <https://digitalcommons.mtu.edu/etds>

 Part of the [Environmental Engineering Commons](#)


Copyright 2010 Cory P. McDonald

---

### Recommended Citation

McDonald, Cory P., "IMPROVING THE RELIABILITY OF AQUATIC BIOGEOCHEMICAL MODELS: INTEGRATING INFORMATION AND OPTIMIZING COMPLEXITY", Dissertation, Michigan Technological University, 2010.  
<https://doi.org/10.37099/mtu.dc.etds/717>

Follow this and additional works at: <https://digitalcommons.mtu.edu/etds>

 Part of the [Environmental Engineering Commons](#)

IMPROVING THE RELIABILITY OF AQUATIC BIOGEOCHEMICAL MODELS:  
INTEGRATING INFORMATION AND OPTIMIZING COMPLEXITY

By

Cory P. McDonald

A DISSERTATION

Submitted in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

(Environmental Engineering)

MICHIGAN TECHNOLOGICAL UNIVERSITY

2010

© 2010 Cory P. McDonald



This dissertation, "Improving the reliability of aquatic biogeochemical models: Integrating information and optimizing complexity", is hereby approved in partial fulfillment of the requirements for the degree of DOCTOR OF PHILOSOPHY in the field of Environmental Engineering.

DEPARTMENT:  
Civil & Environmental Engineering

Signatures:

Dissertation Advisor \_\_\_\_\_  
Dr. Noel R. Urban

Program Chair \_\_\_\_\_  
Dr. David W. Hand

Date \_\_\_\_\_





# Contents

<b>List of Figures</b>	<b>xv</b>
<b>List of Tables</b>	<b>xviii</b>
<b>Preface</b>	<b>xix</b>
<b>Acknowledgments</b>	<b>xxi</b>
<b>Abstract</b>	<b>xxiii</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Aquatic biogeochemical modeling . . . . .	1
1.1.1 Current state of the Science . . . . .	3
1.2 Assessing the performance of biogeochemical models . . . . .	5
1.2.1 The role of data and <i>a priori</i> knowledge in model specification and estimation . . . . .	7
1.3 Outline . . . . .	8

<b>2</b>	<b>Using a model selection criterion to identify appropriate complexity in aquatic biogeochemical models</b>	<b>11</b>
2.1	Abstract . . . . .	11
2.2	Introduction . . . . .	12
2.3	Background and methods . . . . .	15
2.3.1	Model description . . . . .	17
2.3.2	Parameter estimation . . . . .	20
2.4	Results . . . . .	21
2.5	Discussion . . . . .	25
<b>3</b>	<b>Test-bed calibration of a Lake Superior biogeochemical model</b>	<b>29</b>
3.1	Abstract . . . . .	29
3.2	Introduction . . . . .	30
3.3	Methods . . . . .	35
3.3.1	Calibration data . . . . .	35
3.3.2	Ecosystem model description . . . . .	37
3.3.3	1-D hydrodynamic emulator . . . . .	43
3.3.4	Prior estimation . . . . .	45
3.3.5	Calibration . . . . .	46
3.4	Results . . . . .	48

3.4.1	Estimation of error . . . . .	51
3.4.2	3-D model implementation . . . . .	54
3.5	Discussion . . . . .	57
3.6	Conclusions . . . . .	60
<b>4</b>	<b>Modeling carbon fluxes through the Lake Superior lower food web: A parsimonious semi-mechanistic approach</b>	<b>63</b>
4.1	Abstract . . . . .	63
4.2	Introduction . . . . .	64
4.2.1	Parsimony in aquatic biogeochemical models . . . . .	66
4.2.2	Motivation . . . . .	67
4.3	Methods . . . . .	69
4.4	Results and Discussion . . . . .	75
4.4.1	Modeling rates vs. concentrations . . . . .	81
4.4.2	Extended simplified model . . . . .	82
4.4.3	Lake Superior C cycling . . . . .	86
4.4.4	Deep chlorophyll maximum . . . . .	90
4.4.5	Future data requirements . . . . .	93
4.5	Conclusions . . . . .	94
<b>5</b>	<b>Modeling historical trends in Lake Superior total nitrogen concentrations</b>	<b>97</b>

5.1	Abstract . . . . .	97
5.2	Introduction . . . . .	98
5.3	Model Development . . . . .	101
5.3.1	Atmospheric Inputs . . . . .	103
5.3.2	Tributary Inputs . . . . .	108
5.3.3	Model Scenarios . . . . .	111
5.4	Results . . . . .	111
5.5	Discussion . . . . .	115
5.6	Conclusions . . . . .	121
<b>6</b>	<b>Conclusions</b>	<b>123</b>
6.1	Scientific conclusions . . . . .	128
	<b>Bibliography</b>	<b>131</b>
<b>A</b>	<b>CAEDYM model descriptions and estimated parameter values</b>	<b>153</b>
<b>B</b>	<b>CAEDYM/Simulated Annealing MATLAB Code</b>	<b>159</b>
<b>C</b>	<b>Trout Lake dissolved oxygen and total dissolved phosphorus data</b>	<b>165</b>
<b>D</b>	<b>KITES calibration data</b>	<b>167</b>
<b>E</b>	<b>Lake Superior hydrodynamic emulator input data</b>	<b>173</b>

<b>F</b>	<b>Lake Superior biogeochemical model code</b>	<b>177</b>
<b>G</b>	<b>Lake Superior nitrogen Simulink model</b>	<b>195</b>
<b>H</b>	<b>Lake Superior nitrogen data</b>	<b>197</b>
<b>I</b>	<b>Elsevier copyright policy</b>	<b>205</b>



# List of Figures

1.1	Components of model uncertainty, and the idealized roles of data and <i>a priori</i> knowledge in the modeling process. . . . .	5
2.1	Simplified conceptual diagram of the candidate set of models. Each dashed line circumscribes a model; the entire set of models represents a nested structure. Models 1 and 2 consider only simple light (photosynthetically active radiation, PAR) or light/temperature limitation on phytoplankton (i.e., chl <i>a</i> ) growth. Model 3 adds a phosphorus cycle (particulate organic, POP; dissolved organic, DOP; phosphate). Model 4 includes sediment phosphorus (P) and dissolved oxygen (DO). . . . .	19
2.2	Modeled chlorophyll <i>a</i> (a) and 1992-1993 data (b) plotted using a common time axis and colormap. As model complexity (number of estimated parameters) increases from model 1 to 4, the patterns observed in the data are more accurately reproduced. . . . .	23
3.1	KITES Calibration data. Observations are from 47° 24.33'N, 88° 44.29'W, approximately 21 km offshore. Shading represents water temperature (lighter equals warmer). Annual trends in chlorophyll (chl <i>a</i> ) include elevated concentrations throughout the water column, followed by reduced concentrations in the epilimnion and a deep maximum at ~30 m, and elevated chlorophyll distributed evenly through the surface waters during fall overturn. DOC is relatively constant in space and time, other than being elevated ~20 $\mu$ M during the spring bloom. Spatiotemporal trends in phosphorus (TDP and POP) data are less apparent. . . . .	36



3.2	Conceptual diagram of ecosystem model. Phosphate ( $\text{PO}_4^{3-}$ ) is taken up by phytoplankton and lost to DOP and POP via a lumped “mortality” term. Algal carbon is derived from P concentrations using a fixed C:P stoichiometry. Particulate and dissolved P and C pools degrade directly to the mineral form (DIC is not explicitly accounted for in the model). Detrital organic matter (POP and POC) is also lost to settling. . . . .	38
3.3	Conceptual diagram of 1-D hydrodynamic emulator framework (not to scale). The cross-sectional area and cell heights correspond exactly to the GCM model grid. The emulator is forced by output from the GCM (neglecting horizontal transport processes), so that a parameterization developed using the emulator can be expected to produce similar results in the 3-D model. . . . .	43
3.4	Uncalibrated 1-D model output (contour maps) and observations (circles) plotted using identical color scales. When run using the mean prior parameter values, the model generally does not agree well with observations. This demonstrates the level of model performance that would be expected to be achieved using literature values with no additional parameter tuning. .	49
3.5	Calibrated (mean posterior parameter values) 1-D model output (contour maps) and observations (circles) plotted using identical color scales. A dramatic improvement in model fit is observed for chlorophyll, DOC, and phosphorus when compared to the uncalibrated model. The model does not accurately resolve the deep chlorophyll maximum (DCM), however. . . .	51
3.6	Calibrated (Marginal posterior distributions (solid lines) plotted with prior distributions (dashed lines). Posterior distributions for $Mort$ , $remin_P$ , and $remin_C$ indicate significantly lower parameter values than were estimated from the literature. Uninformative priors were used for $f_P$ , $f_C$ , $Chl : P_{max}$ , and $Chl : P_{min}$ (uniform distributions, see Table 3.1). . . . .	52
3.7	Lower (a) and Upper (b) 99% confidence limits for chlorophyll. . . . .	53
3.8	Temporal cross-sections of 1-D posterior model output for the model cells encompassing 15–20 (a) and 25–30 (b) meters water depth. The shaded region indicates the 99% confidence interval. Points indicate observations from various depths within the vertical limits of the model cell. While a number of points do fall outside of the predictive limits in both layers, the model produces a significantly better fit to observations in the mixed layer (a) than in the region of the DCM (b). . . . .	54

3.9	Vertical profiles of net primary production on three dates plotted with the data of Sterner (2010). Model runs were performed for 1999–2000; data are from 2007–2008. The shaded region represents the 99% confidence interval. While the model agrees quite well with measurements in the fall following turnover, it generally underpredicts productivity in the spring, and severely underpredicts productivity throughout the water column during the stratified period. . . . .	55
3.10	3-D GCM output at calibration site, using the parameterization identified in the test-bed calibration. Note the similarity to the 1-D emulator output (Figure 3.5). The discrepancies between the two may be attributed to horizontal transport phenomena and computational artifacts. . . . .	56
3.11	Discrepancy between model output in the 1-D and 3-D hydrodynamic frameworks using identical test-bed (1-D) calibrated parameters, measured as a fraction of the 1-D model value. Note the scale for DOC is reduced by 10%; the discrepancy between 1-D and 3-D model predictions for this variable are much smaller than for the others. . . . .	57
4.1	NPD (a) and NPZD (b) model structures. . . . .	68
4.2	Biogeochemical cycling of carbon in Lake Superior. The dashed ellipse demarcates the portion of the cycle represented by the simplified model structure. Solid lines represent real carbon transport pathways; dotted lines indicate “short-circuits” implemented in the extended simplified model while gray lines indicate the processes that are aggregated within these short-circuits. . . . .	72
4.3	NPD (a-d) and NPZD (e-h) model output for NPP, POP, TDP, and chl <i>a</i> . Symbols indicate calibration data; nutrients and chlorophyll are from the KITES project and NPP is from Sterner (2010). While the explicit representation of zooplankton in the model does improve model performance with respect to productivity, overall model goodness-of-fit is diminished. . . . .	77
4.4	Simplified model output for chl <i>a</i> and uncorrected NPP. Although nutrients and zooplankton are not explicitly modeled, the simplified model outperforms both the NPD and NPZD models with respect to both chlorophyll and production. . . . .	78

4.5	Extended simplified model output for POC and DOC. This output represents only the fraction of these pools derived from primary production. . . .	85
4.6	Modeled community respiration derived from photosynthesis, $CR_{auto}$ , derived by combining modeled phytoplankton respiration, zooplankton respiration, and remineralization rates. Respiration of allochthonous carbon is not included, but results suggest that $CR_{auto}$ only accounts for $\sim 22\%$ of total CR. . . . .	88
4.7	Sensitivity analysis (chl <i>a</i> ) results. Each panel represents one of the scenarios in the factorial design (see Table 4.4 for descriptions). . . . .	93
5.1	Conceptual diagram of model. Atmospheric and tributary loading include nitrate, ammonia/ammonium, and organic nitrogen inputs to the lake as a function of time. . . . .	102
5.2	Model Inputs. The entire shaded region represents modeled atmospheric N loading ( $W_{atm}$ ); the subregions represent the contribution of organic N, ammonium, and nitrate to the total. The solid line represents modeled tributary loading ( $W_{trib}$ ). . . . .	106
5.3	Log-log regression of instantaneous river flow rates and total N concentration (see text for equation). The data are pooled from the St. Louis, Ontonagon, Bad, Taquamenon, Nemadji, and Silver Rivers and Gomanche Creek. Data from individual rivers are more strongly correlated. . . . .	110
5.4	Results of model scenarios 1 (a) and 2 and 3 (b). $NO_3^-$ data shown in (a) are from Weiler 1978 ( $\times$ ), Environment Canada ( $\bigcirc$ ), US EPA ( $+$ ), and KITES ( $\square$ ); see text for complete citations. The solid points indicate the estimated total N concentration. The solid line in (a) is model scenario 1 with $k_{loss}=3.21 \times 10^9 \text{ mol yr}^{-1}$ , the dashed line is with $k_{loss}=1.33 \times 10^9 \text{ mol yr}^{-1}$ . Model output for scenario 2 is shown in (b) by the solid line and model output for scenario 2 is shown in (b) by the dashed line. . . . .	113

5.5	Modeled N loading for scenario 2 (a) and modeled N net loss for scenario 3 (b). The gray shaded area in (a) represents total N loading to the lake; the black area depicts the contribution of $\Delta(t)$ , the calibrated additional loading term. The solid line in (a) indicates (loading - $\Delta(t)$ ), the original estimate of total loading. The modeled loss (b) indicates that a significant decline in burial (or denitrification) rates must have occurred between 1900 and mid-century, followed by a return to higher levels. . . . .	114
5.6	Major nitrogen fluxes in Lake Superior. Inputs and output values are those used in the model, and represent present loading, though there is considerable interannual variability in inflows and outflows. The implied current imbalance between inputs and outputs ( $\sim 1.1 \text{ Gmol yr}^{-1}$ ) only represents 0.03% of the total nitrogen pool in the lake. Biotic uptake is as estimated by Kumar et al. (2008). . . . .	118
5.7	Forecast of scenario 1 with $k_{loss}=1.33 \times 10^9 \text{ mol yr}^{-1}$ (dashed line) and scenario 2 with constant future tributary loading and atmospheric deposition continuing to decrease at current rates (solid line). . . . .	119
C.1	Trout Lake DO and TDP profiles for 1992–1993. Note that elevated dissolved phosphorus in the hypolimnion is observed when DO is depleted (especially in 1992), suggesting sediment phosphorus release is occurring during the stratified period . . . . .	166
E.1	Forcing data for 1-D emulator. Temperature and diffusivity data are outputs from the 3-D hydrodynamic model. Solar radiation data are from NARR (Mesinger et al., 2006). . . . .	174



# List of Tables

2.1	Model squared errors and AIC scores (n=284). K is equal to the number of parameters present in the model plus one ( $\sigma^2$ ). While Model 4 (K=39) resulted in the smallest discrepancy between model output and calibration data, AIC indicates Model 3 (K=24) strikes the best balance between fit and estimation uncertainty. . . . .	22
3.1	Ecosystem model parameters and priors . . . . .	42
3.2	Posterior parameter values . . . . .	50
3.3	Measures of goodness-of-fit (Equation 3.11) for the uncalibrated (using the mean prior parameter values) and calibrated (mean posterior parameter values) models. . . . .	53
4.1	NP[Z]D model parameter descriptions and posterior distributions. . . . .	71
4.2	Simplified model parameter descriptions, values, and posterior distributions. The mean and standard deviation are given for parameters were included in the MCMC procedure. Values below the horizontal line are based on literature values and implemented in the extended simplified model (see text for sources). . . . .	73
4.3	Model normalized root mean square error (NRMSE) for individual variables and weighted cost function ( $\Sigma$ ). . . . .	79
4.4	Chlorophyll <i>a</i> sensitivity analysis results (values given are the normalized root mean square error (NRMSE) for chl <i>a</i> ). . . . .	92

A.1	CAEDYM model parameter descriptions. . . . .	154
A.2	Estimated values of model parameters and upper and lower constraints used in parameter estimation. The parameter ranges given in bold were based on literature values (see text for citations) . . . . .	157
D.1	Chlorophyll <i>a</i> data ( $\mu\text{g L}^{-1}$ ). . . . .	168
D.2	Total dissolved phosphorus data ( $\mu\text{M}$ ). . . . .	169
D.3	Particulate organic phosphorus data ( $\mu\text{M}$ ). . . . .	170
D.4	Dissolved organic carbon data ( $\mu\text{M}$ ). . . . .	171
E.1	1-D emulator grid characteristics (equivalent to the 3-D grid at KITES site HN 210). . . . .	175
H.1	Lake Superior total nitrogen model inputs: riverine total nitrogen loading and atmospheric nitrate deposition. Atmospheric deposition of $\text{NH}_4$ and organic N is held constant in the model (see Chapter 5). . . . .	198
H.2	Lake Superior nitrate concentrations used for model calibration (Chapter 5), listed by source. . . . .	201

## Preface

Chapters 2 and 5 of this dissertation represent previously published work, in (*Ecological Modelling* (doi:10.1016/j.ecolmodel.2009.10.021), and the *Journal of Great Lakes Research* (doi:10.1016/j.jglr.2010.07.008), respectively. The content of Chapter 2 is nearly (95%) all my own work; the contribution of my coauthor (N. R. Urban) was in an editorial and advising capacity. I am responsible for the modelling and analysis presented in Chapter 5 (90%); my coauthors (N. R. Urban and C. M. Casey) contributed background information and data on historical nitrogen loading to the lake.





## Acknowledgments

I would like to extend a heartfelt thank you to my advisor, Dr. Urban, for his patient guidance and generous support as I made the long journey from student to scholar. Thanks also to my committee members, Drs. Auer, Kerfoot, Mayer, and McKinley, for helping me along the way. Thanks to Dr. McKinley and Val Bennington for collaboration on Lake Superior carbon modelling. Thanks to Robert Sterner, his recent work on Lake Superior primary productivity made Chapter 4 of this work possible. Thanks to Marcel Dijkstra for the useful discussions on the DCM. Major funding was provided by the National Science Foundation (OCE 0628545; Sustainable Futures IGERT, DGE 0333401). Thanks to Michigan Tech's Sustainable Futures Institute, Dr. Judith Perlinger, the Center for Water and Society, the Earth, Planetary, and Space Sciences Institute, the DeVlieg Foundation, and the Graduate School for financial support throughout my studies. Thanks to The Center for Limnology, University of Wisconsin-Madison, Environment Canada, and the Centre for Water Research, University of Western Australia for providing data and models. Thanks to S. Gowtham for developing and sharing the LaTeX template used to typeset this work. Thanks to all the CEE staff for always helping me to get things done. Finally, thanks to my fellow students, friends, family, and especially my wife Rachel for all of their support.



## Abstract

The reliability of aquatic biogeochemical models is inversely proportional to the amount of uncertainty associated with model predictions. This uncertainty is a function of error arising from model specification, parameter estimation, and computational methods. Here, the role of observations (data) and *a priori* knowledge in mitigating the first two sources of error is explored, with an emphasis on the identification of the optimal level of model complexity for a given situation. Information-theoretic model selection techniques are applied to a set of simple one-dimensional biogeochemical models describing chlorophyll in Trout Lake using a large field data set; results demonstrate that even relatively modest levels of complexity are often statistically unsupportable by available data. This is one of the first applications of information theory to development of aquatic biogeochemical models. Since the scope of problems being addressed by environmental models often necessitates a greater level of complexity than can be mathematically justified, the application of formal parameter estimation techniques (rather than manual tuning based solely on *a priori* knowledge) is essential, yet is often prohibited by computational costs. A novel method is developed in which the vertical hydrodynamics of a three-dimensional model is emulated in one-dimensional space, allowing biogeochemical parameter values to be optimized using the available data at a reasonable computational expense. It is also demonstrated that a simple model formulation can, in some cases, provide a more useful tool than a more detailed representation of biogeochemical dynamics, due to the greater uncertainty in both

model structure and parameterization associated with the more complex model. A simple model of biogeochemical cycling is developed for Lake Superior that yields a better fit to existing data than do more mechanistically detailed formulations. Using this model, gross primary production (GPP) in Lake Superior is estimated to be  $\sim 10 \text{ Tg C yr}^{-1}$ , an amount equal to 12–77% of community respiration (CR) in the lake. The model also suggests that the deep chlorophyll maximum in Lake Superior may be explained primarily by algal shade adaptation. Finally, even in the case where both constraining data and adequate knowledge of system structure are insufficient, simple models can still be formulated to test hypotheses, though the amount of uncertainty associated with such models is high. A series of simple model structures are calibrated using historical Lake Superior nitrate data in order to identify the possible historical sequence of events that caused a precipitous rise in nitrate concentrations during the previous century. The model suggests that atmospheric deposition alone was likely insufficient to result in the observed increase, and that either loading was elevated or burial was depressed mid-century.

# **Chapter 1**

## **Introduction**

### **1.1 Aquatic biogeochemical modeling**

The practice of developing biogeochemical models of aquatic systems traces its roots to the Streeter-Phelps model (Streeter and Phelps, 1925), which was developed to predict the effect of waste loading on downstream dissolved oxygen levels in rivers. This and related models were mathematically simple and were solved analytically. The advent of digital computing (1960s–1970s) paved the way for the development of more complex (and more realistic) models requiring numerical solutions (Chapra, 1997). During the 1970s, mechanistic nutrient-phytoplankton-zooplankton-detritus (NPZD) models began to be widely employed, chiefly to deal with eutrophication of surface water (e.g., DiToro et al., 1971; Chen

and Orlob, 1975; Canale et al., 1976). Subsequent modeling efforts focused on toxics, and included additional media (sediment, air, suspended solids, etc.). Parallel developments took place in marine systems; building on the early work of Fleming, Riley, and Steele (Gentleman, 2002), NPZD models were applied in the oceans to study ecosystem structure and functioning (e.g., Jassby and Platt, 1976; Steele and Henderson, 1981). A growing recognition of the importance of the “microbial loop” in marine systems (Pomeroy, 1974) prompted the inclusion of microorganisms in food-chain models (Thingstad, 1987).

During the past several decades, rapid advances in computing have allowed for mechanistic aquatic biogeochemical models to grow in complexity, both physically and ecologically. Early spatially-explicit models were one-dimensional (vertical) representations of reservoir water quality (e.g., Imberger and Patterson, 1981; Ford and Johnson, 1986). Increased availability of computing power in the 1980s made possible the development of three-dimensional hydrodynamic models (Blumberg and Mellor, 1987); more recent efforts in estuaries and large lakes have routinely employed three-dimensional hydrodynamic frameworks (e.g., Cerco and Cole, 1993; Chen et al., 2002). Coupled 3-D models such as the MIT general circulation model (MITgcm), the Princeton Ocean Model (POM), and the Regional Ocean modeling System (ROMS) have become standard tools in oceanographic and climate change research (e.g., McKinley et al., 2004; Sarmiento et al., 1993; McKinley et al., 2006).

Advances in complexity in mechanistic aquatic biogeochemical modeling can be broadly classified into two categories: refinement of the formulations employed, and development of more comprehensive representations of ecosystems. In the first category, much effort has been focused on improving the mathematical representation of plankton physiology by simulating the variable nutrient, carbon, and chlorophyll content of cells and the interactive effect of these constituents (along with light and temperature) on cell growth (e.g., Auer and Canale, 1982; Droop, 1983; Flynn, 2008, 2001; Saito et al., 2008; Pahlow, 2005; Pahlow and Oschlies, 2009). Another area in which formulations have been improved is in the development of more realistic representations of zooplankton grazing (Buitenhuis et al., 2006; Mitra et al., 2007). In the second category, modelers have attempted to better approximate reality through the inclusion of additional detail (i.e., additional state variables), most notably by representing plankton using a number of functional types or groups based on phylogeny, trophic role, or size (e.g., Arhonditsis and Brett, 2005; Bissett et al., 1999; Bierman and Dolan, 1981; Le Quere et al., 2005; Bissett et al., 1999) but also by including components of the microbial loop, higher trophic levels, or increasingly detailed water chemistry.

### **1.1.1 Current state of the Science**

While mechanistic modeling has become a prominent tool in aquatic biogeochemistry, many remain skeptical of the reliability of modeling studies in general (Aber, 1997). A re-

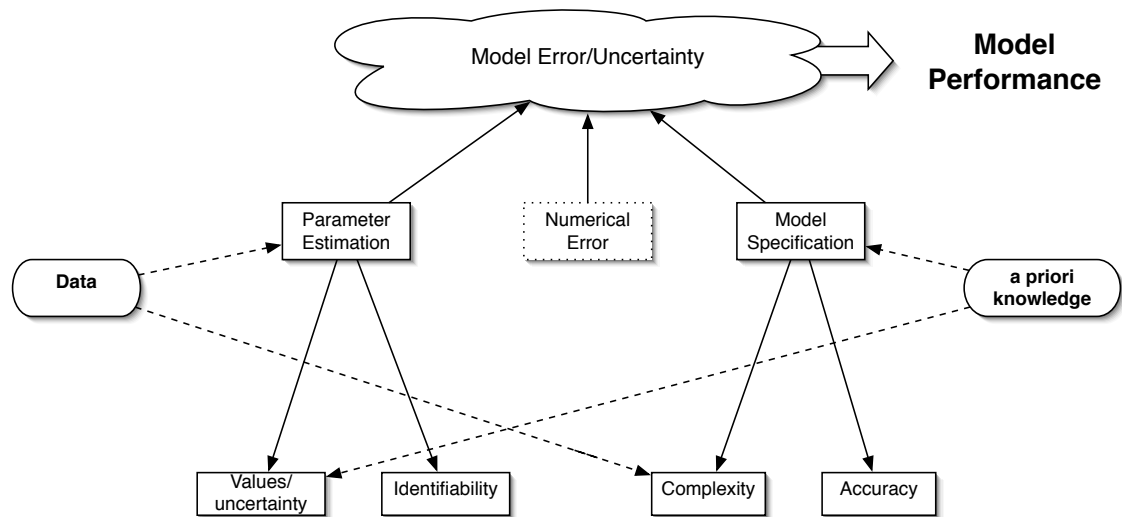


cent review of 153 studies by Arhonditsis and Brett (2004) reveals several probable causes of this skepticism, at least with respect to mechanistic aquatic biogeochemical models. They found that the models generally predicted nutrients and phytoplankton with only moderate success, while performing poorly with respect to zooplankton and bacteria. They also highlighted the failure of the modeling community to adopt and follow a universal methodology and the resistance of freshwater modelers to embrace methodological advances in data assimilation (relative to their oceanographic counterparts). Perhaps most significantly, they found no correlation between model complexity and performance.

The topic of model complexity has evoked a heated debate in the literature. One view is that if we know a process to be ecologically significant in a system, including that process in a biogeochemical model of that system will increase model accuracy (Le Quere, 2006). Another is that our understanding of many of these processes is incomplete at best (Anderson, 2005) and we generally lack sufficient data to validate their implementation in models (Anderson, 2010; Crout et al., 2009), so that the predictions of simpler models tend to be more reliable, although less detailed. In fact, Friedrichs et al. (2006, 2007) demonstrated that complex models do not always fit the data better than simpler ones. While some have championed model simplification using a variety of approaches (Cox et al., 2006; Crout et al., 2009; Denman, 2003), the overall trend continues to be towards increasing complexity.

## 1.2 Assessing the performance of biogeochemical models

Model performance or reliability is maximized when the amount of error or uncertainty associated with the model is minimized. Model error has three main components: specification error, estimation error, and numerical error (Figure 1.1). Model specification defines the structure chosen for the model. This includes state variables and the mathematical formulations used to represent biogeochemical processes or interactions among state variables, as well as the spatial and temporal structure and scale of the model. Specification error is a result of the mismatch between the model and reality, and is a function of both the mathematical complexity and the accuracy of the model. The term accuracy is used here to describe how closely the model structure approximates reality, relative to the level of aggregation of the model. All aquatic biogeochemical models represent a major simpli-



**Figure 1.1:** Components of model uncertainty, and the idealized roles of data and *a priori* knowledge in the modeling process.

fication of reality; all of the organisms, substances, and physiochemical interactions in a system are either partitioned into relatively few modeled state variables and processes or else neglected. Error arising from this aggregation or simplification of reality is unavoidable; while it can be mitigated to a certain extent through increases in model detail (i.e., complexity), doing so tends to increase other forms of uncertainty, as discussed below.

A second source of error and uncertainty arises from the choice of parameter values specified in the model (i.e., calibration). It is important to recognize that many parameters are model-specific, and as such their values are difficult (if not impossible) to determine solely based on field observations or laboratory experiments. Model parameters actually represent abstractions or aggregations of real, measurable rates and constants (O'Neill and Rust, 1979), and the more aggregated a model structure is, the less model parameters can be expected to have an observable analogue. Biological parameters, in particular, often tend to be less directly related to measurable quantities than physical or chemical parameters. For example, while one can model the growth of an algal culture under controlled conditions accurately by Michaelis-Menten kinetics (specifying a maximum growth rate and half-saturation coefficient), it is more difficult to parameterize the algal community in a natural system in this manner, since the factors that affect growth are constantly in flux (Jorgensen and Bendoricchio, 2001).

Poorly identified parameters often confound the issue of parameter estimation. Model structures often include a number of parameters that exert opposing effects on model out-

put, yet are not individually constrained by data, and thus no unique, optimal set of values exists for them. For example, algal maximum growth rates and loss (respiration/excretion/-grazing) rates are highly correlated, as an increase in one can be offset by a decrease in the other without affecting model output. Such a situation should, in general, be remedied by respecifying the model, but this is not always a practical solution when using mechanistic representations.

A third source of model error arises from inaccuracies in the numerical methods utilized to solve the large systems of differential equations typically embodied in a biogeochemical model. Advances in computing power have made it feasible to apply fairly sophisticated methods to solving these systems of equations with minimal error; this error can generally be assumed to be small relative to specification and estimation error for ecological models, and thus it will not be discussed further in this work.

### **1.2.1 The role of data and *a priori* knowledge in model specification and estimation**

Figure 1.1 depicts the ideal roles of data and *a priori* information in model development. The linkage between *a priori* knowledge (expert opinion) and model specification is reasonably evident; the accuracy of the model structure is wholly dependent upon the modeler's knowledge and understanding of the system of interest. Prior knowledge must also

be drawn upon when estimating parameter values, especially when these values are poorly identified. Given the preceding discussion on the model-specific nature of parameters, however, it is clear that values must ultimately be chosen that maximize a model's ability to reproduce calibration data. Perhaps less obvious is the importance of considering the quality and quantity of calibration data available when specifying model structure. A reliable model should, ideally, not contain more degrees of freedom than can be reasonably constrained by observations. Allowing model specification to be guided by data availability can reduce error and uncertainty stemming from overparameterization.

## 1.3 Outline

Each of the following chapters examines a subset of the components of model uncertainty and demonstrates novel methods of incorporating data and *a priori* knowledge into the modeling process to optimize model performance:

- In Chapter 2, it is demonstrated that a maximum level of model complexity can be statistically defined, based on the amount of calibration data available. Using four one-dimensional biogeochemical models of varying complexity and a rich calibration data set from Trout Lake, WI, it is shown that increasing model complexity beyond this optimum results in little improvement in model fit while increasing model uncertainty.

- While the theoretical advantages of data assimilation are clear, in practice the computational resources required can be prohibitive, especially when considering large, three-dimensional models. In Chapter 3, a methodology for data assimilation is presented in which the spatial complexity of a model is artificially reduced. A one-dimensional “test-bed” is developed in which the parameters of a three-dimensional biogeochemical model for Lake Superior can be rigorously calibrated.
- In Chapter 4, it is shown that increasing model complexity does not necessarily lead to better agreement with calibration data, particularly when the added model components are poorly constrained by the data. A parsimonious model of carbon cycling in Lake Superior is developed that demonstrates that a highly aggregated structure, when based on *a priori* knowledge of the system and strongly supported by data, can be preferable to a higher-resolution (i.e., more complex) model.
- Chapter 5 illustrates a “worst-case” modeling scenario: one in which there are few data available and little is known about the system being modeled. By employing data assimilation techniques to estimate parameter values and varying model structure, useful inferences can still be made about the system, and areas in which future research efforts are best focused can be identified.



# Chapter 2

## Using a model selection criterion to identify appropriate complexity in aquatic biogeochemical models

It is vain to do with more what can be done with fewer.

---

*William of Occam (1285-1349)*

### 2.1 Abstract

Aquatic biogeochemical models are widely used as tools for understanding aquatic ecosystems and predicting their response to various stimuli (e.g., nutrient loading, toxic substances, climate change). Due to the complexity of these systems, such models are often elaborate and include a large number of estimated parameters. However, correspondingly

---

The material contained in this chapter was previously published in the journal *Ecological Modelling*: McDonald, C.P. and Urban, N.R. 2010. Using a model selection criterion to identify appropriate complexity in aquatic biogeochemical models. *Ecol Model* 221:428–432.



large data sets are rarely available for calibration purposes, leading to models that may be overfit and possess reduced predictive capabilities. We apply, for the first time, information-theoretic model-selection techniques to a set of spatially explicit (1-D) algal dynamics models of varying parameter dimension. We demonstrate that increases in complexity tend to produce a better model fit to calibration data, but beyond a certain degree of complexity the benefits of adding parameters are diminished (the risk of overfitting becomes greater). The particular approach taken here is computationally expensive, but several suggestions are made as to how multimodel methods may practically be extended to more sophisticated models.

## **2.2 Introduction**

Models are ideally designed to minimize the error resulting from both bias and variance (Ljung, 1987). Bias is reduced as model complexity increases— as the model structure becomes more flexible, the likelihood that it is able to describe the true mechanisms increases. However, variance also generally increases as model complexity increases (i.e., as more estimated parameters are introduced). The level of complexity that simultaneously minimizes bias and variance therefore represents an optimal balance for a given modeling scenario; this is the basis for the the well-known principle of parsimony. Model development should be guided not only by the goodness-of-fit of models, but also by model complexity (Myung, 2000).

Aquatic ecosystems are inherently complex, and data for these systems are often relatively scarce (Pomeroy, 2001). Models of such systems are necessarily simplified representations of reality. If not constrained by adequate calibration data, models may be built that contain unjustifiable degrees of freedom (i.e., are overfit to calibration data) (Aber, 1997; Denman, 2003). Nonetheless, a prevailing trend in biogeochemical modeling is towards increasing model complexity in an effort to develop mechanistically “correct” models (Arhonditsis et al., 2006). For example, the rapid expansion of plankton functional type (PFT) modeling has been called into question, as it may be outpacing data availability and ecological understanding (Anderson, 2005).

There are a variety of information-theoretic tools available for model selection that take into account model complexity (Ward, 2008; Myung, 2000). In this study, Akaike’s Information Criterion (AIC) (Akaike, 1974) is utilized for mechanistic aquatic biogeochemical model selection. AIC evaluates the likelihood of each of a “candidate set” of models, taking into account not only how well the model fits the calibration data (likelihood), but also the number of estimated parameters in the model (Burnham and Anderson, 2002). An advantage in the application of AIC and related criteria is that they attempt to identify the most likely representation of the underlying mechanism that created the data (generalization), rather than considering only the current data set (explanation) (Myung, 2000). AIC is therefore well-suited for use in situations where model predictive capability is important, as is often the case in aquatic biogeochemical modeling.

This work extends the application of AIC to spatially explicit models of aquatic biogeochemistry. While AIC is commonly used to evaluate the fit of multivariate linear regression models in the ecological sciences (Burnham and Anderson, 2001), it is rarely applied to mechanistic models (notable exceptions include Cox et al. (2006) and Poeter and Anderson (2005)), and only recently has been applied to process-based aquatic models (Boulêtreau et al., 2008). In this study we apply AIC to an ecosystem model coupled with a one-dimensional hydrodynamic model.

AIC is used here to evaluate a candidate set of models and to select the model most appropriate for simulating algal dynamics in Trout Lake, WI, USA, part of the North Temperate Lakes Long-Term Ecological Research (NTL-LTER) station (Magnuson and Bowser, 1990). The Trout Lake site has been intensively sampled from 1981-present for a comprehensive suite of physical, chemical, and biological parameters; it is therefore an ideal “test site” for the application of model selection techniques. Modeling is performed using the Dynamic Reservoir Simulation Model/Computational Aquatic Ecosystem Dynamics Model (DYRESM-CAEDYM) package (Hipsey et al., 2006; Imerito, 2007).

## 2.3 Background and methods

A mechanistic biogeochemical model may be viewed as a probabilistic model of a dynamical system (Ljung, 1987):

$$y(t) = g(t, Z^{t-1}; \theta) + \varepsilon(t, \theta), \quad (2.1)$$

where  $y(t)$  is the model output at time  $t$ , consisting of both a predictor term ( $g$ ) and prediction error ( $\varepsilon$ ).  $Z$  is past model data, and  $\theta$  is the parameter vector. The likelihood function is

$$\mathcal{L}(\theta) = \prod_{t=1}^n f_e(\varepsilon(t, \theta), t; \theta), \quad (2.2)$$

where  $f_e$  is the probability density function (PDF) of  $\varepsilon$ . If we assume the prediction errors are normally distributed with zero means and constant variance,  $\sigma^2$ , the likelihood function for the univariate case becomes

$$\mathcal{L}(\theta) = \left( \frac{1}{\sqrt{2\pi}\sigma} \right)^n e^{-\frac{1}{2} \sum_{i=1}^n \left[ \frac{\varepsilon_i}{\sigma} \right]^2}. \quad (2.3)$$

Equation 2.3 is maximized when  $\varepsilon_i/\sigma = 1$  for all  $i$ :

$$\mathcal{L}(\hat{\theta}) = \left( \frac{1}{\sqrt{2\pi}\sigma} \right)^n e^{-\frac{1}{2}n}, \quad (2.4)$$

and the maximized log-likelihood is therefore

$$\ln \mathcal{L}(\hat{\theta}) = -\frac{n}{2} \ln(\sigma^2) - \frac{n}{2} \ln(2\pi) - \frac{n}{2}. \quad (2.5)$$

For the Gaussian case, the maximum likelihood estimator (MLE) of  $\sigma^2$  is simply

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n \varepsilon^2. \quad (2.6)$$

While maximum likelihood (ML), least squares (LS), and other methods provide means to estimate the model parameters,  $\theta$ , based on the goodness-of-fit, another metric is required to evaluate model performance with respect to complexity. Akaike (1973, 1974) derived one such metric, based on Kullback-Leibler (K-L) information, which is now known as Akaike's Information Criterion (AIC). When the ratio of data to estimated parameters is small, as is often the case in aquatic biogeochemical modeling, an additional bias correction term is added to AIC (Sugiura, 1978; Hurvich and Tsai, 1989). The result is the “corrected” AIC:

$$AIC_c = -2\ln(\mathcal{L}(\hat{\theta})) + 2K + \frac{2K(K+1)}{n-K-1} \quad (2.7)$$

where  $K$  is the number of estimated parameters (the number of parameters present in the model plus one;  $\hat{\sigma}^2$  is an additional estimated parameter) and  $n$  is the number of data. Because the second and third terms in Equation 2.5 are constant and equivalent for all models, they may be neglected in the computation of  $-2\ln(\mathcal{L}(\hat{\theta}))$  in Equation 2.7. To

employ AIC in model selection, a set of candidate models is evaluated simultaneously. The model most likely emulating only the underlying mechanisms present in reality, based on goodness of fit and degrees of freedom with respect to the data, is indicated by the lowest AIC value.

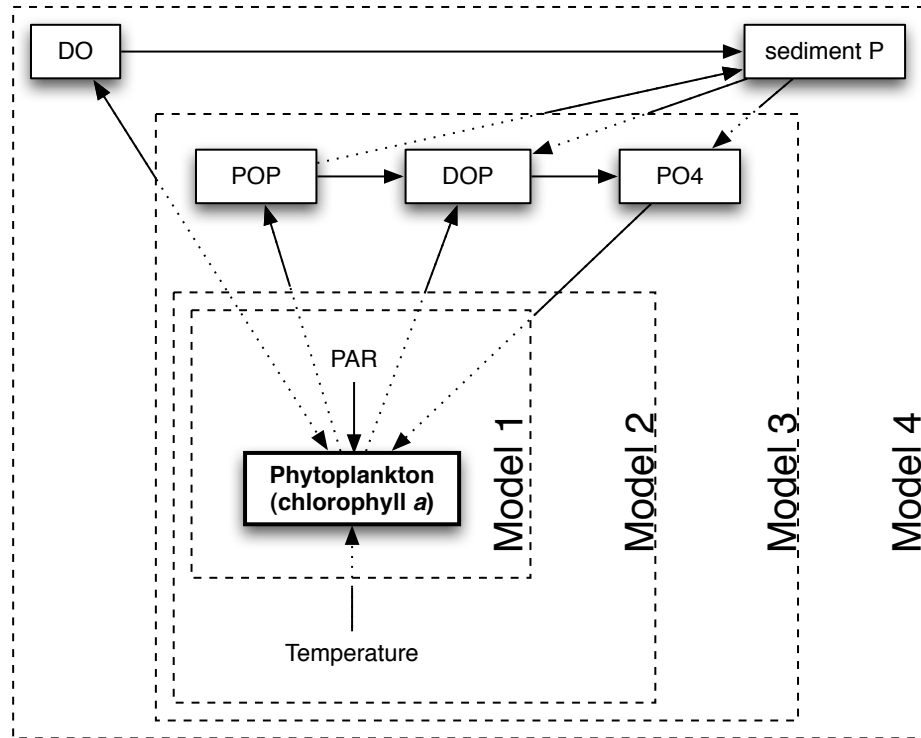
### **2.3.1 Model description**

The 1-D hydrodynamic model DYRESM was used as a physical framework in which to run the candidate set of biogeochemical models. The Trout Lake Basin is dominated by groundwater flow and groundwater flow-through may be an important component of the lake's hydrologic budget (Hunt et al., 2006). Baseflow in the inflowing streams (Allequash, Stevenson, North, Mann) is equal to approximately half of the baseflow in the outflow (Trout River) (Hunt et al., 1998); tributary streams were therefore assumed to equal half of the measured outflow, with the remainder consisting of direct groundwater inputs to the lake. Daily outflow data were obtained from the USGS stream gauges on the outflow (<http://waterdata.usgs.gov>). The model was forced using hourly meteorological data from nearby Woodruff Airport (North Temperate Lakes Long Term Ecological Research program, <http://lter.limnology.wisc.edu>).

DYRESM is not equipped to explicitly model winter ice cover. While an ice cover component has been added to the model in the past (Patterson and Hamblin, 1988), no generally

applicable version of this feature is publicly available. Previously, DYRESM has been applied to Trout Lake during the ice-free season only (Stasio et al., 1996). However, we found DYRESM is capable of running multi-year simulations using modified winter meteorological input data despite the fact that Trout Lake averages 135 days of ice cover annually. To achieve model stability during ice cover, a lower threshold of 6 degrees Celsius was applied to air temperatures and wind speed was set to a constant  $3 \text{ m s}^{-1}$ . This combination effectively prevented modeled surface water temperatures from becoming negative. The hydrodynamic model was calibrated to ten years of water temperature measurements (1992-2002, North Temperate Lakes LTER program, <http://lter.limnology.wisc.edu>) by optimizing the value of the extinction coefficient. The normalized root mean square error (RMSE) of the temperature output vs. observations was 2.5%. In evaluating ecological models, the DYRESM output was treated as “truth”; that is, estimated parameters present in the hydrodynamic model were not considered.

A set of four candidate models was constructed using CAEDYM to model algal dynamics in Trout Lake (Figure 2.1). There are 12 required state variables in CAEDYM. However, manipulation of various model parameters and settings makes it possible to deactivate many of these variables, or at least to sever their interaction with the variables of interest (e.g., chlorophyll *a* (chl *a*), phosphorus pools, dissolved oxygen (DO), etc.). Models were specified in a nested structure. Model 1 considered only the effect of light on phytoplankton growth while all losses were due to settling and respiration/excretion. Model 2 added the effect of temperature on metabolic processes. Model 3 included the effect of nutrient (phos-



**Figure 2.1:** Simplified conceptual diagram of the candidate set of models. Each dashed line circumscribes a model; the entire set of models represents a nested structure. Models 1 and 2 consider only simple light (photosynthetically active radiation, PAR) or light/temperature limitation on phytoplankton (i.e., chl *a*) growth. Model 3 adds a phosphorus cycle (particulate organic, POP; dissolved organic, DOP; phosphate). Model 4 includes sediment phosphorus (P) and dissolved oxygen (DO).

phorus) limitation on growth, and introduced variables for particulate organic phosphorus (POP), dissolved organic phosphorus (DOP), and phosphate in the water column. In an attempt to capture summer sediment P release with Model 4, a DO variable was introduced, as well as sediment pools of phosphorus. The state equations for the candidate models are presented in the Appendix.



### 2.3.2 Parameter estimation

Parameter estimation was performed via Simulated Annealing (Kirkpatrick et al., 1983). For models 1 and 2, the objective function was defined simply as the sum of squared error (SSE) between chl *a* measurements (North Temperate Lakes LTER program, <http://lter.limnology.wisc.edu>) and interpolated model output. For models 3 and 4, the following initial objective function was used:

$$f_{obj} = \sum_{i=1}^n [(\text{chl } a)_{model,i} - (\text{chl } a)_{data,i}]^2 + \sum_j \frac{\overline{\text{chl } a}}{\overline{x_j}} \left( \sum_{i=1}^n [(x_j)_{model,i} - (x_j)_{data,i}]^2 \right) \quad (2.8)$$

where  $n$  represents the number of data points, and  $x$  represents  $j$  additional model variables (phosphorus, DO). Once the algorithm began to stabilize (i.e., the rate of improvement in the initial objective function became low), the chl *a*-only SSE was used as the objective function to seek final optimized values of all fitted parameters. The models were each run under identical forcing for 785 days, beginning on 1 January 1992. This timeframe was chosen (rather than two years) because the observed top-to-bottom chlorophyll *a* bloom following fall turnover in 1993 persisted well into December of that year; choosing data that extended into 1994 ensured the models would be fit to two complete annual cycles.

The phytoplankton community of Trout Lake is dominated by diatoms and chrysophytes (Descy et al., 2000); parameter starting points and bounds were selected accordingly. Growth and respiration rates, half-saturation coefficients, and settling velocities were con-

strained based on reported literature values (Bowie et al., 1985; Carpenter et al., 1998); CAEDYM defaults were used to loosely constrain the other parameters. Final estimated values as well as upper and lower bounds for each parameter are listed in the Appendix. Field measurements were used to initialize the state variables.

## 2.4 Results

Models 1 (SSE=2932) and 2 (SSE=2003) fail to accurately reproduce the measured measured profiles chlorophyll *a* data (Figure 2.2, Table 2.1). This is the expected result because they lack adequate mechanisms to reproduce the dominant features of the chlorophyll *a* profile in the lake over the entire year (or the mathematical flexibility to do so). These models instead employ fairly unrealistic parameters to achieve a minimal discrepancy between model output and measured data (see Appendix). Model 1 accomplishes this by invoking a relatively large settling rate, so that spring production in the euphotic zone is quickly dispersed throughout the water column, which results in maximum chlorophyll *a* concentrations forming concurrently in the water column with the observed maximum caused by spring turnover. Model 2, on the other hand, invokes a large light saturation coefficient, which lessens the effect of light on algal growth. As a result, chlorophyll *a* patterns in Model 2 are strongly correlated with temperature patterns. While this model fails to capture the spring algal bloom, it does effectively reproduce the distribution of chlorophyll *a*

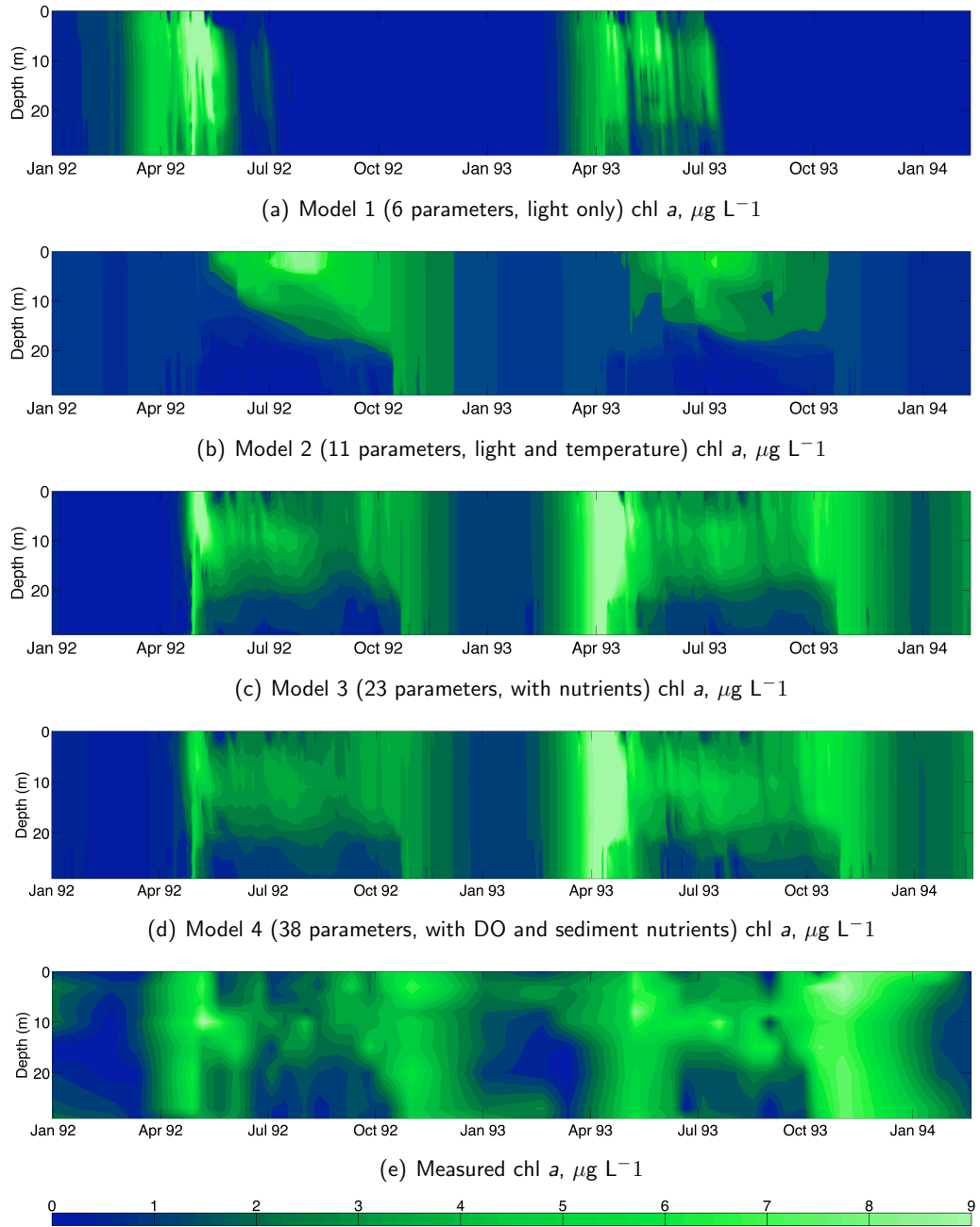
**Table 2.1:** Model squared errors and AIC scores (n=284). K is equal to the number of parameters present in the model plus one ( $\sigma^2$ ). While Model 4 (K=39) resulted in the smallest discrepancy between model output and calibration data, AIC indicates Model 3 (K=24) strikes the best balance between fit and estimation uncertainty.

	K	SSE	AIC	AICc
Model 1	7	2932	677	677
Model 2	12	2003	579	580
Model 3	24	689	<b>300</b>	<b>304</b>
Model 4	39	<b>638</b>	308	321

throughout the epilimnion during stratification, resulting in a better overall quantitative fit to the data.

The introduction of phosphorus in Model 3 (SSE=689) causes a marked improvement in model fit (Figure 2.2, Table 2.1). The timing and location of the spring bloom, deep chlorophyll *a* maximum, and fall bloom all agree well with the data, and the modeled concentrations are similar to measured values. The least agreeable fit is obtained during the winter months; this is to be expected given the manner in which DYRESM is configured to circumvent ice cover (see Section 2.3.1). There are two clear maxima in the data (Figure 2.2), presumably formed by under-ice algal production and settling. Due to the simulated continuous mixing during winter, no vertical patterns in chlorophyll *a* concentrations emerge in the model during this time period.

Graphical output from Model 4 (SSE=638) appears very similar to output from Model 3. It does, however, fit the calibration data slightly better (Figure 2.2, Table 2.1). Sub-thermocline chlorophyll *a* concentrations are reproduced more accurately in Model 4, most



**Figure 2.2:** Modeled chlorophyll  $a$  (a) and 1992-1993 data (b) plotted using a common time axis and colormap. As model complexity (number of estimated parameters) increases from model 1 to 4, the patterns observed in the data are more accurately reproduced.

notably during the summer of 1993. The increase in model performance gained by adding sediment phosphorus pools and DO, however, is clearly much less pronounced than the increase gained by adding the water column phosphorus cycle.

As expected, the SSE between model-generated chlorophyll *a* and data decreases as model complexity increases (Table 2.1). A large degree of improvement is seen between Model 1 and Model 2 ( $\Delta\text{SSE}=929$ ) and an even larger improvement between Model 2 and Model 3 ( $\Delta\text{SSE}=1314$ ). The increase in fit from Model 3 to Model 4, however, is much smaller ( $\Delta\text{SSE}=51$ ), indicating diminishing returns (model accuracy) on parameter investment.

AIC<sub>c</sub> (and AIC) values, on the other hand, show a clear minimum with Model 3. AIC values decrease from Model 1 through Model 3, but increase for Model 4. Between Models 3 and 4, the model error due to variance outweighs the model error due to bias. In other words, although the formulation of Model 4 more completely represents the true physical system, Model 3 is more reliable *given the data we have available*. It is important to note here that by “data”, we are only referring to chlorophyll *a* data. It is impossible to determine how well the multivariate models perform relative to the univariate models using all of the available data because the corresponding variables are not present in the smaller models; AIC must be calculated for all models using a common data set.

## 2.5 Discussion

Increasing model size leads to better model goodness-of-fit, but Model 4 contains a large number of estimated parameters relative to the data set used for calibration, and therefore produces a higher AIC value than Model 3. This does not indicate, however, that sediment phosphorus release is an unimportant process in Trout Lake. Hypolimnetic phosphorus concentration data are clearly elevated contemporaneous with late stratification, particularly in 1992. Diminished improvement in model fit with increasing parameter dimension is generally the case; the lower AIC value for Model 3 indicates it contains an amount of complexity that can be supported by the calibration data, but does not suggest it is more mechanistically correct than Model 4. Considering only this candidate set of model structures and calibration data set, Model 3 embodies the best trade-off between bias and variance error.

Were the goal simply to reproduce the calibration data set as accurately as possible, the highest-dimensioned model would be most appropriate, because it would result in the best fit to the data. Such a model, however, would have limited utility as a scientific or managerial tool. Models are typically constructed to forecast a system's behavior. We therefore seek models with a high degree of generality, rather than models configured to reproduce a particular data set. Failure to assess whether a model is truly being fit to major system events and patterns (underlying mechanisms) or being excessively tuned to noise in the calibration data increases the risk of model overfitting. Overfit models' performance tends

to be diminished when run using forcing data outside of the calibration timeframe (Pitt and Myung, 2002).

In the foregoing analysis, we have implicitly made the assumption that variation in model performance due to specification error (i.e., the correctness of model structure) is negligible. While AIC considers the number of estimated parameters in the model, it does not consider the suitability of the model configuration. Assuming the formulation of all model components is reasonable (reasonably reflects reality), however, average model performance will increase with complexity. This caveat illustrates the importance of *a priori* expertise when models are constructed. Further, the decision to include or not to include model variables is typically based on the desired functionality of the model; it is often necessary or desirable to include more variables than field data can actually support. In these cases, model selection criteria such as AIC may still be applied to identify the most parsimonious mathematical representations of the processes affecting the variables of interest, and to assist in identifying unnecessary (insensitive) parameters.

In order for the various candidate models to be evaluated, it is important that their parameters are set at optimal or near-optimal values, which is computationally expensive (or even practically impossible) to achieve for large models. An attractive alternative exists in stochastic parameter estimation methods. Bayesian model calibration (Kennedy and O'Hagan, 2001) has been demonstrated to be capable of efficiently estimating parameter values in aquatic biogeochemical models (Arhonditsis et al., 2008; Law et al.,

2009). While such stochastic approaches do not reduce parameter estimation uncertainty to the point where it may be neglected (as required by AIC), the estimation uncertainty is quantified, since probability density functions are generated for each parameter. Parameter interdependence may also be quantified. Several selection criteria do take variances and covariances into account (Minimum Descriptive Length (MDL) (Rissanen, 1987), Information Complexity Criterion (ICOMP) (Bozdogan, 2000), and Deviance Information Criterion (DIC) (Spiegelhalter et al., 2002)). A set of models calibrated using Bayesian techniques could, therefore, be evaluated for appropriate complexity, given the amount of information available, with these. Owing to the great computational expense associated with the application of deterministic parameter estimation methods, stochastic parameter estimation methods coupled with more sophisticated model selection techniques may represent a more practical framework for the widespread application of multimodel selection techniques to mechanistic aquatic biogeochemical models.





## Chapter 3

# Test-bed calibration of a Lake Superior biogeochemical model

On two occasions I have been asked, ‘Pray, Mr. Babbage, if you put into the machine wrong figures, will the right answers come out?’ I am not able rightly to apprehend the kind of confusion of ideas that could provoke such a question.

---

*Charles Babbage (1792-1871)*

### 3.1 Abstract

While the reliability of aquatic mechanistic biogeochemical models is highly dependent upon accurate parameterization, the high computational expense associated with 3-D models generally prohibits the application of formal parameter estimation techniques *in situ*. Here, a one-dimensional hydrodynamic emulator, driven by three-dimensional model output, is developed to provide an efficient test-bed environment in which model parameters

---

The material contained in this chapter will be expanded and submitted for publication in the peer-reviewed literature, with N. R. Urban, G. A. McKinley, and V. Bennington as co-authors. V. Bennington provided the 3-D model output presented here.

are estimated using a Markov Chain Monte Carlo (MCMC) approach. The spatial and temporal uncertainty of model predictions due to estimation error is quantified. A simple ecosystem model is calibrated for Lake Superior that is capable of reproducing most of the major features of observed concentration profiles of nutrients, dissolved organic carbon, and chlorophyll at a location in the western basin of the lake. The model structure appears to be insufficiently detailed (or inappropriately formulated) to reconcile observations of these variables with measured primary productivity during the stratified period. Implementation of the test-bed calibration approach increases the discrepancy between the model and data by less than 20% on average.

## **3.2 Introduction**

Parameter estimation (calibration) is a critical step in the development of environmental models (Chapra, 1997). The utilization of data and prior knowledge to inform parameter values (in addition to appropriate mechanistic formulations) leads to useful and credible tools (Jakeman et al., 2006). However, proper identification of parameter values remains a difficult task due to the inherent complexity of environmental systems, scarcity of data, and mathematical and computational limitations.

While models in general represent simplifications of real systems, mechanistic aquatic biogeochemical models in particular attempt to distill exceedingly intricate and heterogeneous

systems into relatively elementary systems of equations. While a number of arguments for and against increasing model complexity have been made from an ecological standpoint (e.g., Flynn, 2006, 2003; Anderson, 2005, 2010, 2006; Le Quere, 2006), even at their most complex, aquatic biogeochemical models represent a very simplified view of reality. In a sense, all these models are misspecified, and their optimal parameterization is less related to observable values of real-world analogs of model parameters (which often simply do not exist) and more related to the particular model specification.

There are also mathematical limitations on the identifiability of parameters. Because the observations available to calibrate aquatic biogeochemical models are generally scarce, the number of parameters that can be accurately estimated from the data is limited. The number of estimated parameters in a model increases approximately as the square of the number of state variables (Denman, 2003); the cost, in terms of model uncertainty stemming from a high-dimensional parameter space, of adding additional model compartments (e.g., plankton functional types) is high. Even when a relatively large data set is available, only a modest level of complexity can be statistically justified (McDonald and Urban, 2010), and typically only 10–15 parameters can be estimated from the available data (Oschlies, 2006). Calibration of an overly complex model to insufficient data will likely result in overfitting and a decreased predictive capability. Furthermore, because many parameters in aquatic biogeochemical models exhibit a high degree of correlation, a unique solution to the parameter vector often does not exist.

In spite of the obvious pitfalls associated with complex models, aquatic biogeochemical models are often *necessarily* overly complex; the level of detail required to provide useful answers to scientific and managerial questions is often greater than the level of detail that the available data can support. Sophisticated approaches to parameter estimation must therefore be employed in order to extract as much information from the calibration data as possible, and model uncertainty should be quantified. In a review of 153 mechanistic aquatic biogeochemical modeling studies, Arhonditsis and Brett (2004) found that only 8.5% utilized formal parameter estimation techniques, and only 30.1% quantified model goodness-of-fit. Quantification of uncertainty is even less common. Given the difficulties associated with parameter identification, it seems unlikely that a trial-and-error approach can successfully identify optimal parameterization of even the simplest models. Smith and Yamanaka (2007) significantly improved the performance of two previously hand-tuned photoacclimation models (Geider et al., 1998; Pahlow and Oschlies, 2009) by applying mathematically rigorous fitting techniques. While expert knowledge can arguably be applied judiciously to guide the manual parameter tuning process, the need to explore extensive areas of the parameter space while monitoring the performance of each state variable simultaneously (i.e., avoiding calibration bias (Arhonditsis and Brett, 2004)) constitutes a Sisyphean task.

A number of techniques have been applied to parameter identification in aquatic mechanistic biogeochemical modeling. All such approaches essentially seek to minimize the discrepancy between the model and data. Gradient descent methods, such as the conju-

gate gradient or adjoint methods, have been successfully applied to marine biogeochemical models (e.g., Friedrichs, 2002). Formulation of the adjoint model can present a challenge, however, especially for complex models. Gradient-free optimization methods may also be employed; while the computational cost of exhaustively searching the parameter space is generally prohibitive, stochastic methods such as genetic algorithms, simulated annealing, or Markov Chain Monte Carlo (MCMC) can explore the parameter space much more efficiently (e.g., Harmon and Challenor, 1997; Arhonditsis et al., 2007). For a detailed review of data assimilation techniques in biogeochemical modeling, see Oschlies (2006). The benefits of the Bayesian paradigm and MCMC approach for parameter estimation in aquatic biogeochemical models have been established (Law et al., 2009; Arhonditsis et al., 2008, 2007); prior knowledge of parameter values is combined with the information contained in calibration data to formulate posterior parameter distributions. This approach represents a compromise between the utilization of expert judgment and mathematical rigor. The stochastic nature of MCMC also facilitates estimates of parameter and model uncertainty.

Three-dimensional hydrodynamics are often necessary to address large-scale problems in the oceans, as well as in large estuarine and freshwater aquatic ecosystems (e.g., Cerco and Cole, 1993; Six and Maier-Reimer, 1996; Chen et al., 2002; McKinley et al., 2003, 2004; Dutkiewicz et al., 2005; Buitenhuis et al., 2006; Follows et al., 2007). The high computational demands of 3-D models, however, present an additional obstacle to the application of rigorous parameter estimation techniques. Because a single multi-year model run requires several days of computational time even when parallelized and run on computer clusters,

these models cannot practically be run the tens of thousands of times often required to sufficiently explore the parameter space. Instead, parameter values in 3-D biogeochemical models are typically chosen based on a combination of literature values and manual tuning.

Formal parameter estimation techniques have, on the other hand, been successfully applied to one-dimensional marine biogeochemical models. Friedrichs et al. (2007, 2006) constructed a 1-D test bed framework in which various model structures and physical forcing regimes could be implemented and calibrated using an adjoint method. Their 1-D framework was forced using a combination of observations and 3-D hydrodynamic model output. Schartau and Oschlies (2003) took a similar approach to calibrating a simple biogeochemical model simultaneously at three sites using a genetic algorithm; physical forcing was acquired from a 3-D hydrodynamic model run. Such a test bed approach represents a practical approach to identifying the biogeochemical parameters needed for 3-D models. Computational time demands are dramatically reduced because the spatial complexity of the model is diminished, and also because much of the hydrodynamic computation is performed externally.

In this paper we present the results of a test bed calibration of a basic biogeochemical model for Lake Superior. We construct an efficient 1-D (vertical) hydrodynamic emulator that corresponds to a specific location in a Lake Superior 3-D general circulation model (Desai et al., 2009; Bennington et al., 2010; Bennington, 2010) for which we have time-resolved 1-D biogeochemical data. We use an MCMC procedure to identify the optimal

model parameterization and estimate uncertainty *ex situ*. The resulting parameterization is then implemented in the 3-D framework.

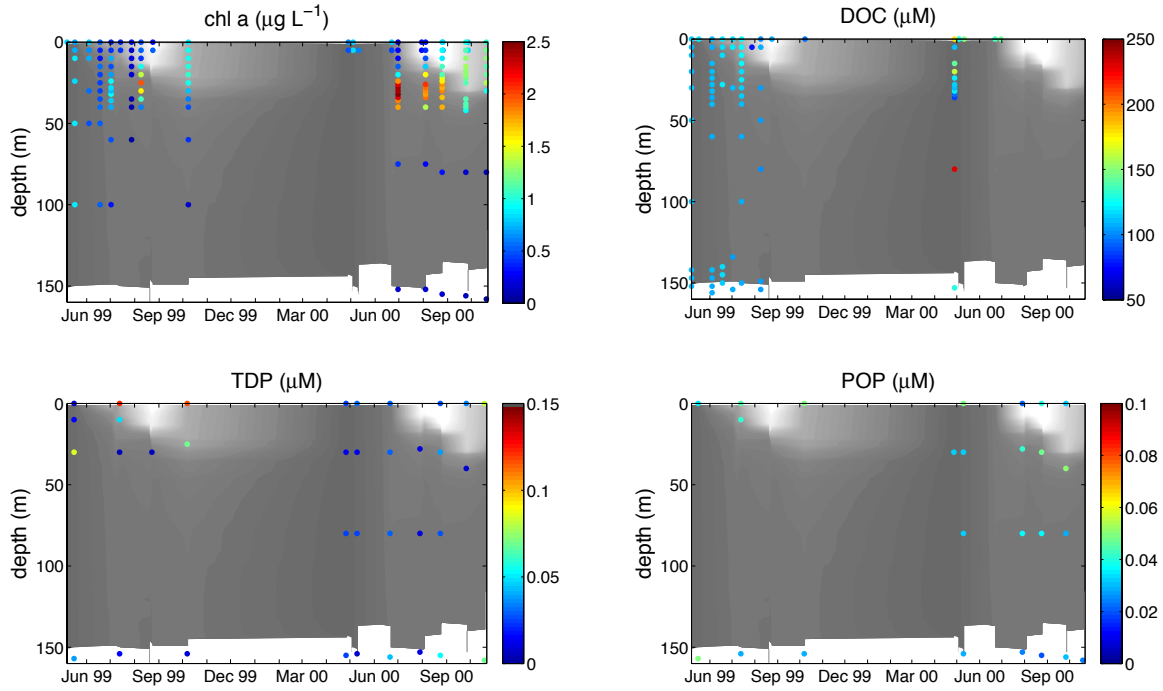
## **3.3 Methods**

### **3.3.1 Calibration data**

Intensive sampling was performed along several transects perpendicular to the Keweenaw Peninsula (located on the southern shore of Lake Superior), between 1998 and 2000 as part of the Keweenaw Interdisciplinary Transport Experiment in Superior (KITES), an NSF-sponsored program designed to examining cross-margin transport processes (Auer and Johnson, 2004; Green and Eadie, 2004) . For the purpose of calibrating our ecosystem model, we focus here on one of the sites located farthest offshore, 21 km northwest of the north entry of the Portage waterway ( $47^{\circ} 24.33'N$ ,  $88^{\circ} 44.29'W$ ) at a water depth of approximately 180 m. This site is representative of the deep pelagic regions that comprise the majority of Lake Superior, and thus is a good candidate for calibrating a generalized ecosystem model.

Despite representing a relatively rich data set for Lake Superior, the observations used to calibrate the model are sparse and irregularly distributed in both space and time (Figure 3.1). The four parameters for which sufficient data exist to be utilized for model calibra-





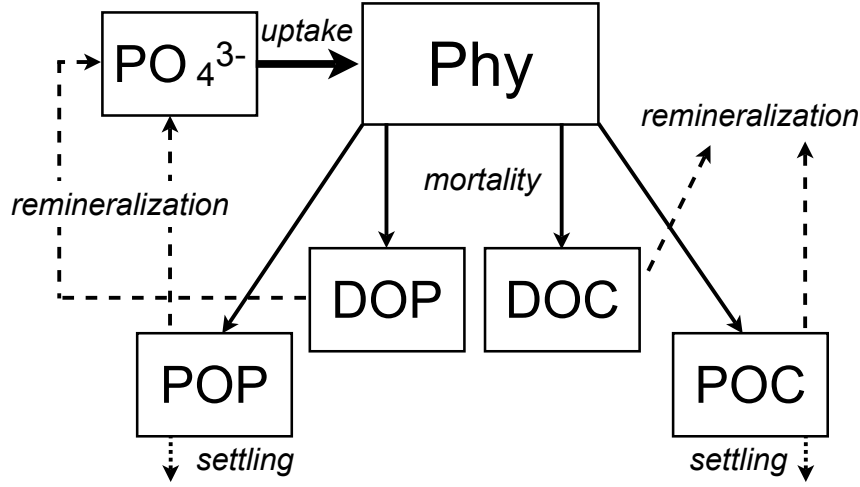
**Figure 3.1:** KITES Calibration data. Observations are from  $47^{\circ} 24.33'\text{N}$ ,  $88^{\circ} 44.29'\text{W}$ , approximately 21 km offshore. Shading represents water temperature (lighter equals warmer). Annual trends in chlorophyll (chl  $a$ ) include elevated concentrations throughout the water column, followed by reduced concentrations in the epilimnion and a deep maximum at  $\sim 30$  m, and elevated chlorophyll distributed evenly through the surface waters during fall overturn. DOC is relatively constant in space and time, other than being elevated  $\sim 20 \mu\text{M}$  during the spring bloom. Spatiotemporal trends in phosphorus (TDP and POP) data are less apparent.

tion are chlorophyll  $a$  (chl  $a$ ), dissolved organic carbon (DOC), total dissolved phosphorus (TDP), and particulate organic phosphorus (POP). The chl  $a$  observations show uniformly elevated chlorophyll concentrations throughout the water column prior to stratification, a deep chlorophyll maximum (DCM) occurring at a water depth of approximately 30 m during stratification, and increased epilimnetic chlorophyll toward the end of the stratified period in both years, but DCM concentrations are higher in 2000 than in 1999. These data have been explored in more detail by Auer and Bub (2004) and Auer et al. (2010). The

majority of DOC data are from 1999, and show a  $\sim 20 \mu\text{M}$  increase in the surface waters prior to stratification. The majority of the P data, on the other hand, were obtained in 2000. Measured TDP concentrations are fairly uniform throughout the water column. POP concentrations, however, are elevated in the surface layer and in the region of the deep chlorophyll maximum.

### 3.3.2 Ecosystem model description

The ecosystem model is a simplified adaptation of the marine biogeochemical model described by Dutkiewicz et al. (2005), shown conceptually in Figure 3.2. Modeled variables include phytoplankton, phosphate ( $\text{PO}_4^{3-}$ ), dissolved organic phosphorus (*DOP*), dissolved organic carbon (*DOC*), and particulate (detrital) organic carbon and phosphorus (*POC/POP*). Phytoplankton “mortality” (a lumped loss term essentially representing predation and excretion) is partitioned into the dissolved and particulate fractions. Both particulate and dissolved phosphorus/carbon pools remineralize directly into the inorganic forms ( $\text{PO}_4^{3-}$  and *DIC*, the latter of which is not explicitly included in the model). Only the detrital particulate pools (*POP* and *POC*) experience settling; phytoplankton biomass does not. Phytoplankton biomass is measured in units of phosphorus ( $\mu\text{M P}$ ) and is governed by the following equation:



**Figure 3.2:** Conceptual diagram of ecosystem model. Phosphate ( $PO_4^{3-}$ ) is taken up by phytoplankton and lost to DOP and POP via a lumped “mortality” term. Algal carbon is derived from P concentrations using a fixed C:P stoichiometry. Particulate and dissolved P and C pools degrade directly to the mineral form (DIC is not explicitly accounted for in the model). Detrital organic matter (POP and POC) is also lost to settling.

$$\frac{\partial Phy}{\partial t} = \mu \cdot Phy - mort \cdot Phy + D' \frac{\partial^2 Phy}{\partial z^2}. \quad (3.1)$$

Growth,  $\mu$  is a function of light, nutrients, and temperature:

$$\mu = \hat{\mu} \left( \frac{PAR}{K_I + PAR} \right) \left( \frac{PO_4^{3-}}{K_P + PO_4^{3-}} \right) \frac{1}{\theta_{norm}} \theta_{base}^T \quad (3.2)$$

light and nutrient limitation are modeled using the Monod formulation (with half-saturation coefficients,  $K$ ), and temperature effects are modeled according to a power function (with

coefficients  $\theta_{base}$  and  $\theta_{norm}$ ). Mortality losses of phytoplankton cease when biomass falls below a specified minimum,  $phymin$ .

Phosphate is removed from the water column by phytoplankton and replenished via remineralization of both  $DOP$  and  $POP$ :

$$\frac{\partial PO_4^{3-}}{\partial t} = -\mu \cdot Phy + remin_P(DOP + POP) \frac{1}{\theta_{norm}} \theta_{base}^T + D' \frac{\partial^2 PO_4^{3-}}{\partial z^2}, \quad (3.3)$$

with the rate of remineralization adjusted for temperature effects using the same temperature adjustment as for phytoplankton growth.

Dissolved organic phosphorus is produced from the fraction of phytoplankton mortality designated as dissolved:

$$\frac{\partial DOP}{\partial t} = f_P \cdot mort \cdot Phy - remin_P \cdot DOP \cdot \frac{1}{\theta_{norm}} \theta_{base}^T + D' \frac{\partial^2 DOP}{\partial z^2}. \quad (3.4)$$

The remaining fraction of phytoplankton P is partitioned to the particulate organic phosphorus pool:

$$\frac{\partial POP}{\partial t} = (1 - f_P) \cdot mort \cdot Phy - remin_P \cdot POP \cdot \frac{1}{\theta_{norm}} \theta_{base}^T + D' \frac{\partial^2 POP}{\partial z^2} + v_s \frac{\partial POP}{\partial z}. \quad (3.5)$$

Carbon is treated similarly to phosphorus in the model. A constant carbon to phosphorus ratio,  $x_{C:P}$ , is applied to algal biomass. Mortality is partitioned into dissolved organic carbon,

$$\frac{\partial DOC}{\partial t} = x_{C:P} \cdot f_C \cdot mort \cdot Phy - remin_C \cdot DOC \cdot \frac{1}{\theta_{norm}} \theta_{base}^T + D' \frac{\partial^2 DOC}{\partial z^2}, \quad (3.6)$$

and particulate organic carbon,

$$\frac{\partial POC}{\partial t} = x_{C:P} \cdot (1 - f_C) \cdot mort \cdot Phy - remin_C \cdot POC \cdot \frac{1}{\theta_{norm}} \theta_{base}^T + D' \frac{\partial^2 POC}{\partial z^2} + v_s \frac{\partial POC}{\partial z}, \quad (3.7)$$

Chlorophyll is calculated as a function of phytoplankton biomass and light availability (Doney et al., 1996):

$$chl : P = chl : P_{max} - (chl : P_{max} - chl : P_{min}) \cdot \min \left( \frac{PAR}{I^*}, 1 \right). \quad (3.8)$$

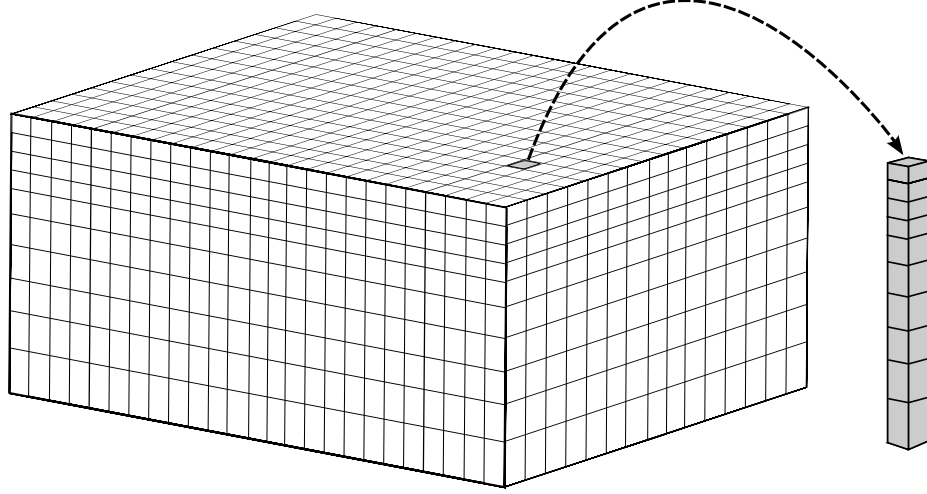
Complete parameter descriptions are provided in Table 3.1.

Several transformations were performed in order to make the model output compatible with observations. Only a small fraction of the total dissolved organic carbon pool in Lake Superior is readily available to the biota; the majority appears to be only moderately reactive (Urban et al., 2005). To simulate this, a “refractory” DOC constant was added to the parameter vector; this value was added to model *DOC* output prior to comparing to data. Model output for  $PO_4^{3-}$  and *DOP* were combined to compare with measurements of total dissolved phosphorus (TDP). Model output for *Phy* and *POP* were combined to compare with POP data.

Fitting model output to static observations of state variables does not guarantee that the model is kinetically accurate. The spatiotemporal distribution of POP or chl *a*, for example, does not provide any information on the flux of carbon or nutrients through the biota. In order to constrain modeled rates, we also included the *in situ* net primary productivity data of Sterner (2010) in our calibration. These data were collected at several offshore sites in the period 2006–2008. Since we run the model for the two years for which we have KITES data, we arbitrarily combined 2006 and 2007 data to compare to model output from 1999, and compared 2008 data directly to model output from 2000. Measured profiles of carbon uptake do not show large interannual differences, so the error introduced into the model calibration from the integration of these temporally incongruous data is assumed to be negligible when compared to specification and estimation error.

**Table 3.1:** Ecosystem model parameters and priors

Parameter	Description	Units	Prior Mean S.D.	Bounds Min Max	
$k_0$	Light extinction coefficient	$m^{-1}$	0.16 -	- -	This study
$k_c$	Light extinction due to algae	$m^{-1} (\mu g \text{ chl } L^{-1})^{-1}$	0.015 -	- -	(Krause-Jensen and Sand-Jensen, 1998)
$\hat{\mu}$	Maximum growth rate	$d^{-1}$	2.20 0.47	0 3.63	(Bowie et al., 1985)
$Mort$	Mortality rate	$d^{-1}$	0.05 0.03	0 0.14	(Bowie et al., 1985)
$K_I$	PAR half-saturation coefficient	$W m^{-2}$	12.0 8.2	2.6 28.3	(Bowie et al., 1985)
$K_P$	$PO_4^{3-}$ half-saturation coefficient	$\mu M$	0.91 0.71	0.03 2.60	(Bowie et al., 1985)
$remin_P$	Phosphorus remineralization rate	$d^{-1}$	0.14 0.17	0.001 0.650	(Bowie et al., 1985)
$remin_C$	Carbon remineralization rate	$d^{-1}$	0.04 0.04	0.001 0.100	(Bowie et al., 1985)
$f_P$	Fraction of mortality to DOP, Phy1	-	- -	0 1	Prior $U(0,1)$
$f_C$	Fraction of mortality to DOC, Phy1	-	- -	0 1	Prior $U(0,1)$
$x_{C:P}$	Molar C:P	-	219 28	147 291	(Barbiero and Tuchman, 2004)
$phym_{in}$	Minimum phytoplankton biomass	$\mu M$	- -	- -	Model default = $1.0 \times 10^{-5}$
$v_P$	Sinking velocity of POP	$m d^{-1}$	0.6 0.1	0.24 2.32	(Burns and Rosa, 1980)
$v_C$	Sinking velocity of POC	$m d^{-1}$	0.6 0.1	0.24 2.32	(Burns and Rosa, 1980)
$Chl : P_{max}$	Maximum ratio of chlorophyll to P	$\mu g \mu mol^{-1}$	28.3 -	21.7 34.9	(Barbiero and Tuchman, 2004), Prior $U(21.7, 34.9)$
$Chl : P_{min}$	Minimum ratio of chlorophyll to P	$\mu g \mu mol^{-1}$	7.7 -	7.1 8.3	(Barbiero and Tuchman, 2004), Prior $U(7.1, 8.3)$
$I^*$	Light coefficient for chlorophyll	$W m^{-2}$	24.0 16.4	5.2 56.6	Estimated as $2 \times K_I$
$\theta_{base}$	Temperature base coefficient	-	1.066 0.045	1.0 1.2	(Bowie et al., 1985)
$\theta_{norm}$	Temperature normalization coeff.	-	1.40 -	- -	This study



**Figure 3.3:** Conceptual diagram of 1-D hydrodynamic emulator framework (not to scale). The cross-sectional area and cell heights correspond exactly to the GCM model grid. The emulator is forced by output from the GCM (neglecting horizontal transport processes), so that a parameterization developed using the emulator can be expected to produce similar results in the 3-D model.

### 3.3.3 1-D hydrodynamic emulator

A 1-D hydrodynamic emulator was constructed to reproduce the vertical GCM behavior for the location at which the calibration data were measured. Grid cell dimensions are identical to those used in the GCM (surface area of 4 km<sup>2</sup> and depth-dependent variable height), so that the resulting 1-D domain represents a vertical “stack” of model cells (Figure 3.3). The model contains 22 vertical cells at the calibration location for a total modeled depth of 177 m. Vertical transport (turbulent diffusive mixing) was implemented as a simple finite-difference approximation:

$$\frac{\partial C_i}{\partial t} = D'_i \frac{C_i - C_{i-1}}{z_i - z_{i-1}} + D'_{i+1} \frac{C_i - C_{i+1}}{z_{i+1} - z_i} \quad (3.9)$$



where  $D'$  is the eddy diffusivity ( $\text{m}^2 \text{d}^{-1}$ ),  $i$  is the layer number,  $C$  is the tracer concentration, and  $z$  is the water depth at the midpoint of the layer (m). Vertical advective transport, which is small relative to diffusive mixing, was neglected in the emulator, as were horizontal transport processes. Particle settling (*POC* and *POP*) is modeled as:

$$\frac{\partial C_i}{\partial t} = v_s \left( \frac{C_{i-1}}{z_i - z_{i-1}} - \frac{C_i}{z_{i+1} - z_i} \right), \quad (3.10)$$

where  $v_s$  is the vertical settling velocity ( $\text{m d}^{-1}$ ).

The GCM was forced using climatological data from the North American Regional Reanalysis (NARR) (Mesinger et al., 2006). The emulator was in turn forced using water temperature and diffusivity output from the GCM and incident shortwave radiation, 45% of which was considered to be photosynthetically active radiation (PAR) (Papaioannou et al., 1993). Light intensity is attenuated with depth according to the Beer-Lambert equation, taking into account light absorption due to the chlorophyll concentration in the water column (i.e., algal self-shading). The model was run for two years beginning on January 1, 1999 using a multistep, variable-order numerical differentiation formula (Shampine and Reichelt, 1997).

### 3.3.4 Prior estimation

The base light extinction coefficient,  $k_e$ , was determined by fitting the Beer-Lampert equation ( $I = I_0 \exp[-k_e z]$ ) to PAR profiles measured in 1999–2000 at the three farthest offshore KITES locations. At 47°3.53'N, 89°26.19'W (160 m) the mean light extinction was calculated as  $0.16 \text{ m}^{-1}$  ( $n = 14$ ,  $sd = 0.08$ ,  $range = 0.08\text{--}0.34$ ); at the calibration site the mean was  $0.15 \text{ m}^{-1}$  ( $n = 23$ ,  $sd = 0.04$ ,  $range = 0.06\text{--}0.22$ ); and at 47°33.06'N, 88°9.22'W (260 m) the mean was  $0.16 \text{ m}^{-1}$  ( $n = 13$ ,  $sd = 0.04$ ,  $range = 0.10\text{--}0.26$ ). An average value of  $0.16 \text{ m}^{-1}$  was used. Additional light extinction due to chlorophyll (self-shading) was assumed to be equal to  $0.015 \text{ m}^{-1} (\mu\text{g chl L}^{-1})^{-1}$  (Krause-Jensen and Sand-Jensen, 1998).

Phytoplankton parameter (maximum growth rate, mortality, half-saturation coefficients, temperature coefficient) and remineralization rate priors were formulated based on the literature review of Bowie et al. (1985). Organic particle settling velocity priors were estimated using a weighted average of the size-classified data presented by Burns and Rosa (1980). Stoichiometric parameter ( $C : P$ ,  $chl : P$ ) priors were based on the data of Barbiero and Tuchman (2004). All priors were assumed to be normally distributed. Prior distribution parameter values ( $\mu, \sigma$ ) are provided in Table 3.1.

The initial condition of the state vector was estimated from KITES data. The average total phosphorus concentration in the water column at the calibration site, computed for dates and depths where simultaneous measurements of TDP and POP are available ( $n = 24$ ) was

0.072  $\mu\text{M}$ , with TDP and POP each comprising 50% of the total. Each of the four P pools was therefore initialized at 0.018  $\mu\text{M}$ . The exact initial distribution among the pools ( $PO_4^{3-}$ ,  $DOP$ ,  $POP$ , and  $Phy$ ) is not critical; the phosphorus component of the model spins up rapidly, before any observations are available.  $DOC$  was initialized at 116  $\mu\text{M}$  (average of observations at the calibration site), with 17  $\mu\text{M}$  in the biologically active DOC pool (Urban et al., 2005), and the initial POC concentration was estimated as 3  $\mu\text{M}$  based on the prior estimate of  $x_{C:P}$  and the initial estimated value of  $Phy$ .

### 3.3.5 Calibration

The goodness-of-fit of the model was defined as the sum of the normalized root mean squared error (NRMSE) of each data type:

$$\sum_{i=1}^5 \sqrt{\frac{\sum_{n=1}^{N_i} (f_{i,n}(\theta, x) - y_{i,n})^2}{N_i}} \div IQR_i \quad (3.11)$$

where  $N$  is the number of data of type  $i$ ,  $f$  is model output and  $y$  is the data. The interquartile range ( $IQR$ ) was used for normalization to mitigate the impact of outliers on the cost function. We are assuming that the mathematical model formulation is adequate (i.e., provides the flexibility necessary to model the system) and that measurement error is negligible.

Light extinction coefficients, minimum phytoplankton biomass (*phym<sub>in</sub>*), and initial conditions were excluded from the parameter estimation process, either because they were sufficiently identified from the data ( $k_0$  and  $k_e$ ) or the model was insensitive to their value. While the model is highly sensitive to the value of the temperature normalization coefficient ( $\theta_{norm}$ ), it is simply a multiplicative term distributed throughout the model; adjusting its value in addition to maximum growth and remineralization rates would be redundant.

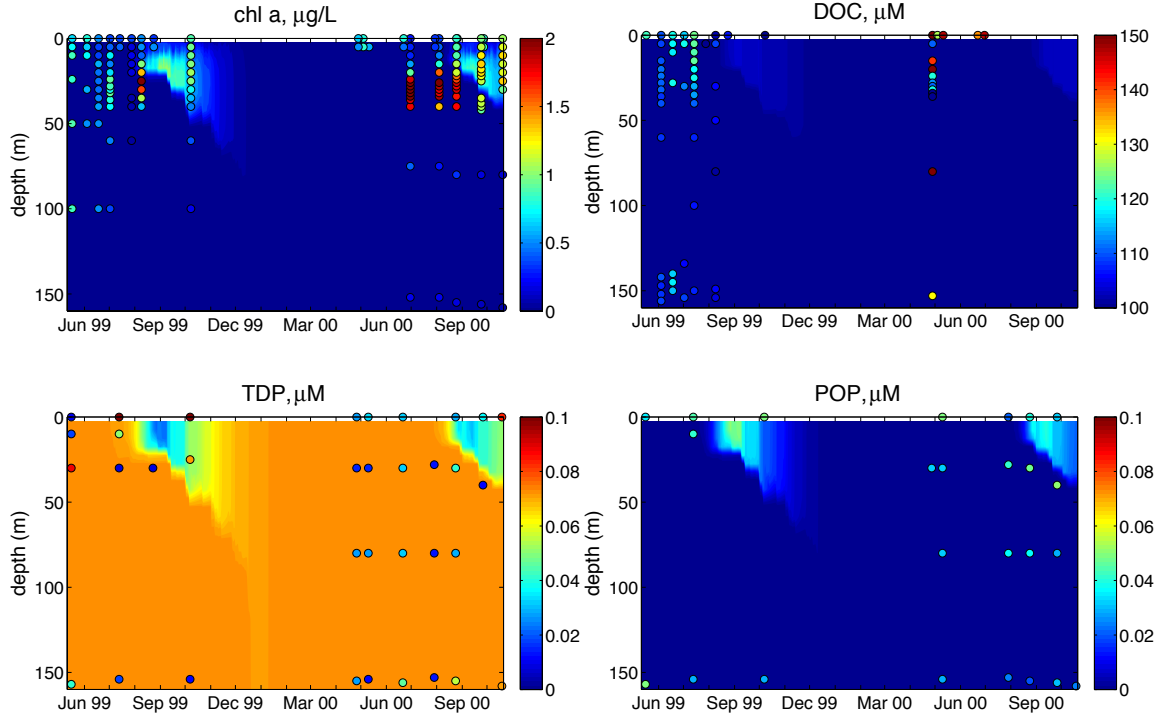
The joint posterior parameter distribution for the remaining parameters was estimated via Markov Chain Monte Carlo (MCMC). Initially, the parameter space was explored using a standard Metropolis-Hastings algorithm, in which parameter samples are randomly drawn from the current state, and either accepted or rejected based on the current and proposed states (Tierney, 1994). The strong correlation between many of the model parameters resulted in an unacceptably low rate of mixing when exploring the parameter space using this algorithm. Only minor improvement was achieved by applying an Adaptive Metropolis (AM) algorithm, in which the proposal distribution is periodically updated using the full information available in the current chain (Haario et al., 2001). However, the chains converged rapidly (<5000 iterations) when using a combined Delayed Rejection (DR) and Adaptive Metropolis sampler, DRAM (Haario et al., 2006). In a DR algorithm, upon rejection a second proposal is made based on the rejected proposal (Green and Mira, 2001), which allows for a more efficient exploration of the parameter space of highly nonlinear models. Chains were run to 10000 iterations and the posterior sample was obtained from

the final 5000 iterations. Model predictive limits were calculated using 500 random samples from the chain.

Once the test-bed calibration was completed, the resultant parameter vector and initial conditions were implemented in the GCM. The model was spun up for one year and then run for 1999–2000.

### **3.4 Results**

The model does not agree well with the data when parameterized using the mean prior values (Figure 3.4). Significant nutrient uptake and algal growth do not occur in the model until the lake becomes stratified, but the data clearly show elevated chlorophyll throughout the completely mixed water column in the spring. Modeled chlorophyll concentrations do, however, agree well with observations during fall turnover. While the variable chl:P ratio in the model does produce higher chlorophyll concentrations at depth, the pronounced DCM at  $\sim 30$  m forming prior to stratification and continuing through the stratified period is absent from the model output. Modeled DOC does not show any of the variation observed in the data. Modeled phosphorus is disproportionately present in the dissolved pool (correspondingly, modeled POP is generally low). The performance of this “uncalibrated” model output is an example of a model parameterized using values from the literature with no additional tuning, and highlights the necessity of model-specific calibration.



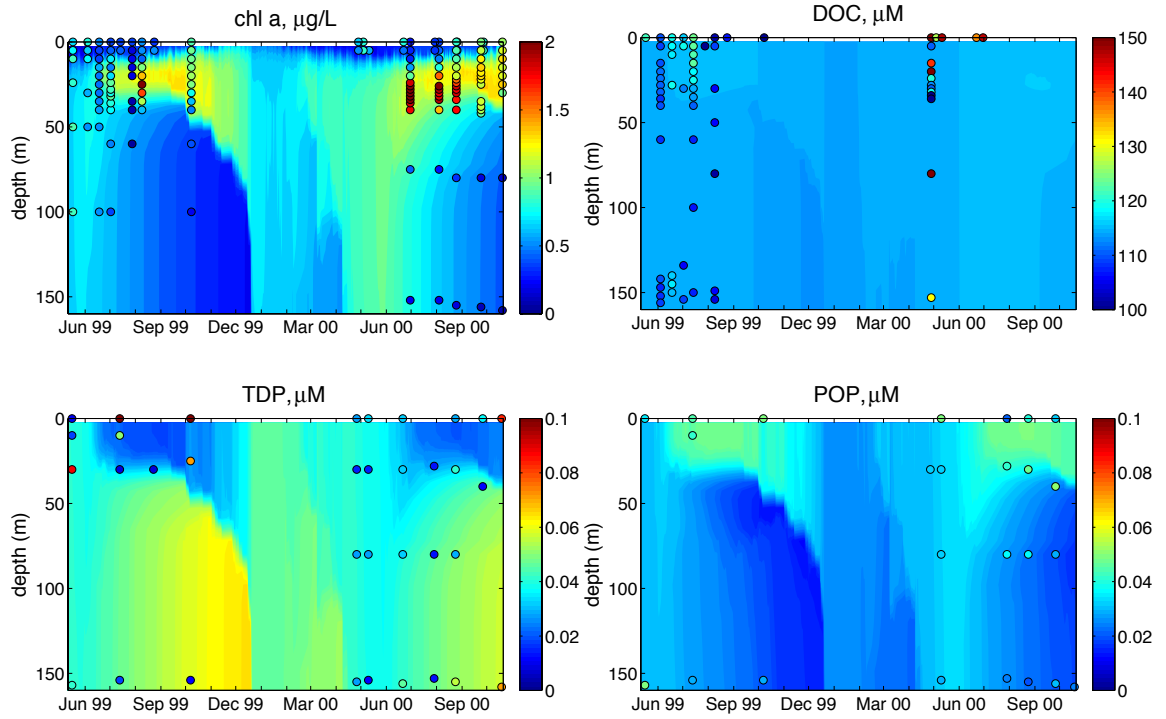
**Figure 3.4:** Uncalibrated 1-D model output (contour maps) and observations (circles) plotted using identical color scales. When run using the mean prior parameter values, the model generally does not agree well with observations. This demonstrates the level of model performance that would be expected to be achieved using literature values with no additional parameter tuning.

When run using the mean posterior parameter vector, the model fit is improved dramatically (Figure 3.5). The posterior parameter distributions identified using MCMC are shown in Figure 3.6, and posterior parameters ( $\mu, \sigma$ ) are listed in Table 3.2. Generally, the posterior distributions are similar to the prior distributions, highlighting the sensitivity of the model to slight variations in parameterization. Three parameters, however, are significantly lower than initial estimates: phytoplankton mortality ( $M_{ort}$ ), phosphorus remineralization ( $remin_P$ ), and carbon remineralization ( $remin_C$ ). While the model still does not accurately reproduce the DCM, the spring bloom is present, and predicted concentrations throughout

**Table 3.2:** Posterior parameter values

Parameter	Prior		Posterior	
	Mean	S.D.	Mean	S.D.
$\hat{\mu}$	2.20	0.47	1.35	0.21
$Mort$	0.05	0.03	0.004	0.002
$K_I$	12.0	8.2	7.04	3.87
$K_P$	0.91	0.71	0.45	0.19
$remin_P$	0.14	0.17	0.006	0.002
$remin_C$	0.04	0.04	0.002	0.001
$f_P$	-	-	0.92	0.02
$f_C$	-	-	0.97	0.03
$x_{C:P}$	219	28	214	21
$v_P$	0.6	0.1	0.58	0.06
$v_C$	0.6	0.1	0.65	0.06
$Chl : P_{max}$	28.3	-	28.9	1.5
$Chl : P_{min}$	7.7	-	7.9	0.3
$I^*$	24.0	16.4	34.1	8.8
$\theta_{base}$	1.066	0.045	1.037	0.045
$DOC_{refractory}$	-	-	99.9	2.1

the year agree reasonably well with the data. A stronger spatiotemporal trend is observed in DOC output, and the predicted patterns of TDP and POP agree well with observations. Variable-specific NRMSE values for both the calibrated and uncalibrated model are summarized in Table 3.3. The posterior estimate of the parameter vector generally reduces the NRMSE, although the accuracy of the model prediction of primary productivity is actually reduced relative to the prior condition.

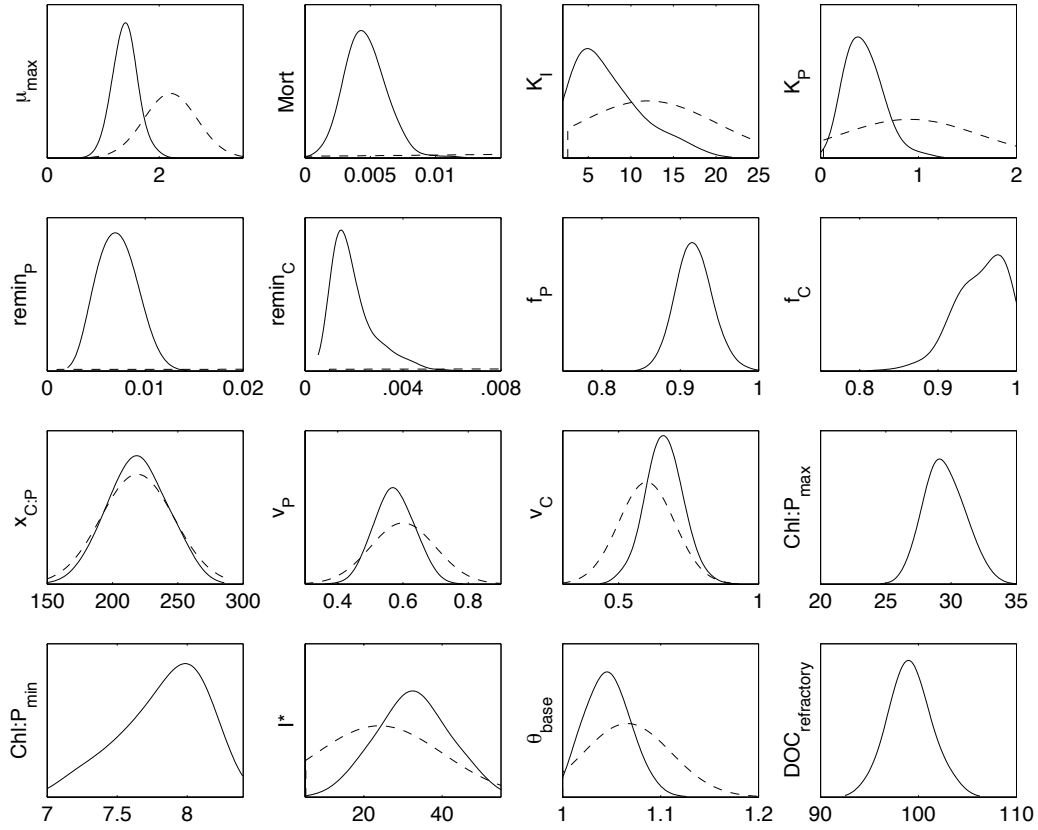


**Figure 3.5:** Calibrated (mean posterior parameter values) 1-D model output (contour maps) and observations (circles) plotted using identical color scales. A dramatic improvement in model fit is observed for chlorophyll, DOC, and phosphorus when compared to the uncalibrated model. The model does not accurately resolve the deep chlorophyll maximum (DCM), however.

### 3.4.1 Estimation of error

Our approach here does not explicitly account for model error due to specification (i.e., we assume that the mathematical formulation of the model is capable of simulating the major biogeochemical phenomena in the lake). Measurement error is also assumed to be negligible. The MCMC approach does, however, provide information about the uncertainty due to parameter estimation error. Figure 3.7 shows the upper and lower 99% Confidence Limits for model predictions of chlorophyll. In horizontal (time series) cross sections of the output, model performance is correlated with depth in the water column. Figure 3.8 depicts





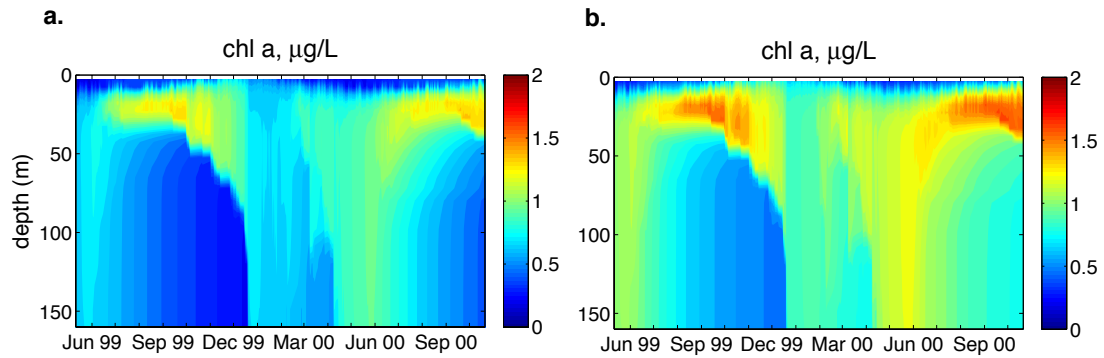
**Figure 3.6:** Calibrated (Marginal posterior distributions (solid lines) plotted with prior distributions (dashed lines). Posterior distributions for *Mort*, *remin<sub>P</sub>*, and *remin<sub>C</sub>* indicate significantly lower parameter values than were estimated from the literature. Uninformative priors were used for *f<sub>P</sub>*, *f<sub>C</sub>*, *Chl : P<sub>max</sub>*, and *Chl : P<sub>min</sub>* (uniform distributions, see Table 3.1).

the mean estimate of chlorophyll in two model cells (15–20 meters depth, representative of the lower epilimnion, and 25–30 meters depth, representative of the region the DCM occurs) with the 99% confidence interval and observations. While a number of points do fall outside the predictive limits in the 15–20-m cell, agreement between the model and data is reasonable. Deeper in the water column, however, the observations of chlorophyll corresponding to the DCM fall well above the upper 99% confidence limit (note the different scales of the y-axes in Figure 3.8).

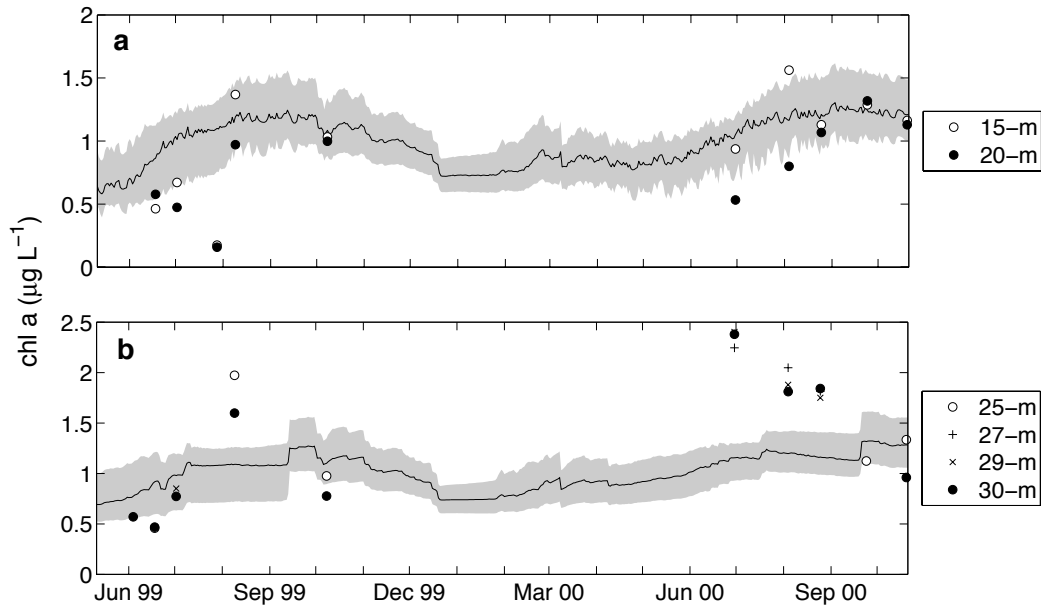
**Table 3.3:** Measures of goodness-of-fit (Equation 3.11) for the uncalibrated (using the mean prior parameter values) and calibrated (mean posterior parameter values) models.

	NRMSE	
	Uncalibrated	Calibrated
chl <i>a</i>	1.36	0.67
TDP	1.36	0.91
POP	2.60	0.72
DOC	2.42	1.87
NPP	0.71	0.79
$\Sigma$	8.45	4.98

Model performance with respect to primary productivity varies temporally (Figure 3.9). While excellent agreement between model output and data is achieved in the fall (November), for most of the spring (April) observations fall outside the 99% confidence interval. During the summer stratified period, however, the model severely underpredicts carbon uptake via photosynthesis; every summertime observation exceeds the upper 99% confidence limit of the model, and predicted summer productivity is actually lower than is productivity predicted for the unstratified season. The data show the reverse to be true.



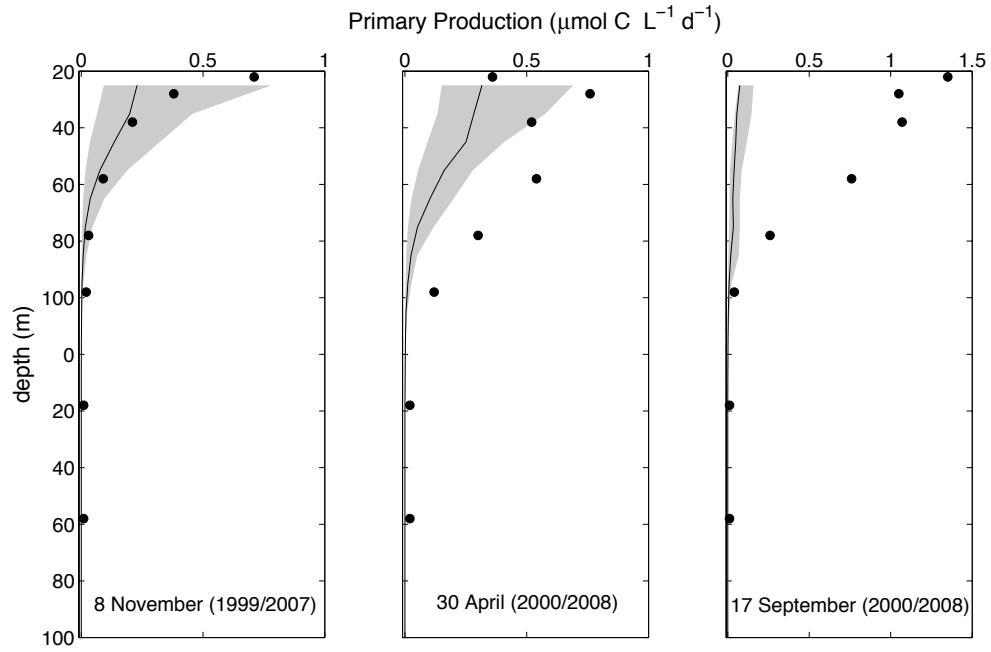
**Figure 3.7:** Lower (a) and Upper (b) 99% confidence limits for chlorophyll.



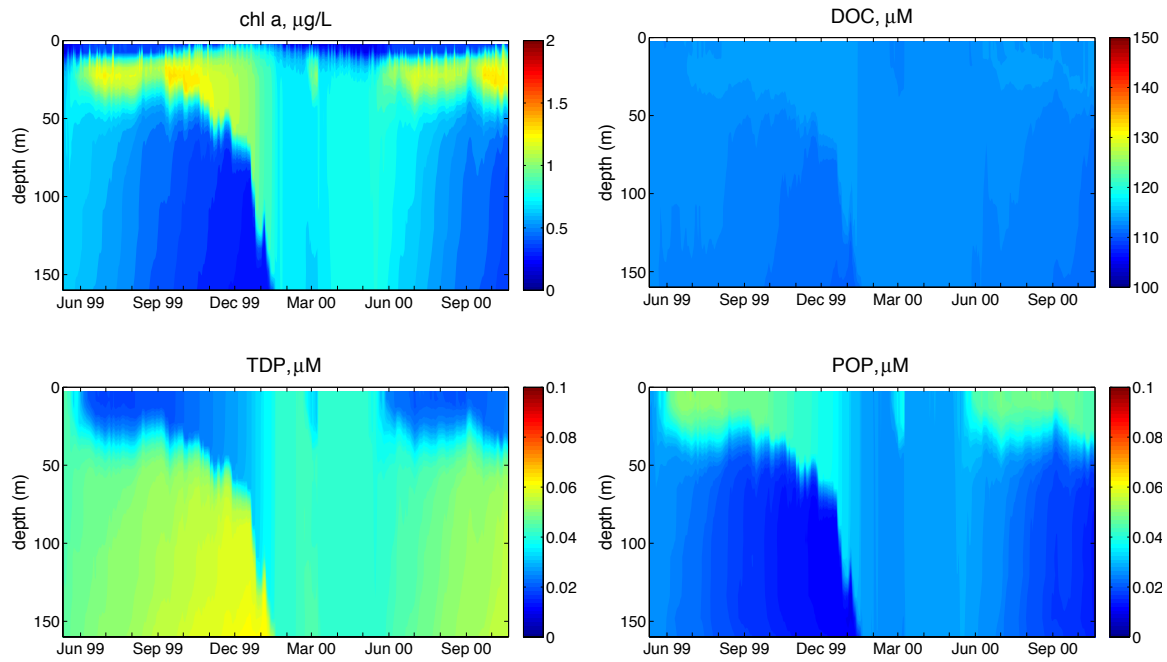
**Figure 3.8:** Temporal cross-sections of 1-D posterior model output for the model cells encompassing 15–20 (a) and 25–30 (b) meters water depth. The shaded region indicates the 99% confidence interval. Points indicate observations from various depths within the vertical limits of the model cell. While a number of points do fall outside of the predictive limits in both layers, the model produces a significantly better fit to observations in the mixed layer (a) than in the region of the DCM (b).

### 3.4.2 3-D model implementation

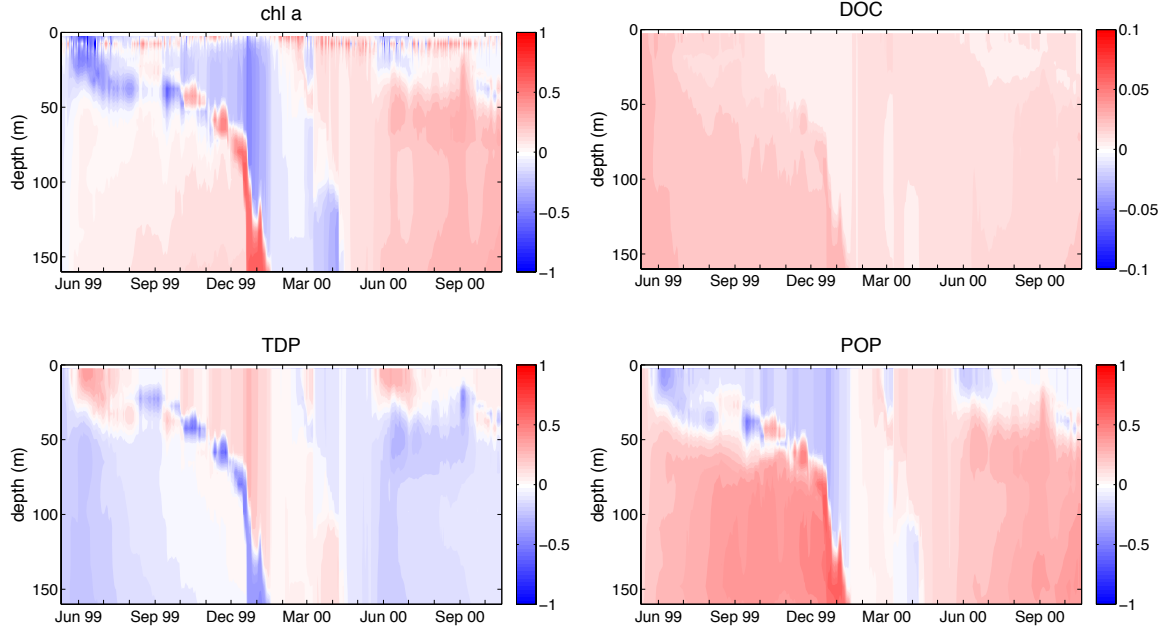
GCM output using the test-bed calibrated parameter vector appears nearly identical to the 1-D emulator output at the calibration site for all measured variables (Figure 3.10). The difference between the 1-D and 3-D model output, as a fraction of the 1-D model output, is plotted in Figure 3.11. The difference between the two is greatest for chl *a* and POP, and least for DOC.



**Figure 3.9:** Vertical profiles of net primary production on three dates plotted with the data of Sterner (2010). Model runs were performed for 1999–2000; data are from 2007–2008. The shaded region represents the 99% confidence interval. While the model agrees quite well with measurements in the fall following turnover, it generally underpredicts productivity in the spring, and severely underpredicts productivity throughout the water column during the stratified period.



**Figure 3.10:** 3-D GCM output at calibration site, using the parameterization identified in the test-bed calibration. Note the similarity to the 1-D emulator output (Figure 3.5). The discrepancies between the two may be attributed to horizontal transport phenomena and computational artifacts.



**Figure 3.11:** Discrepancy between model output in the 1-D and 3-D hydrodynamic frameworks using identical test-bed (1-D) calibrated parameters, measured as a fraction of the 1-D model value. Note the scale for DOC is reduced by 10%; the discrepancy between 1-D and 3-D model predictions for this variable are much smaller than for the others.

### 3.5 Discussion

The 1-D emulator makes accurate estimation of ecosystem parameters for the 3-D biogeochemical model computationally feasible. *A priori* knowledge, expert judgment, and information derived from the literature were incorporated in the formulation of the prior parameter distributions, but the model did not agree well with the data when run using the mean prior values (Figure 3.4). The improvement in goodness-of-fit after model calibration is striking (Figure 3.5; Table 3.3). While a great deal of improvement could be expected to

result from manually (trial and error) tuning the parameters (starting from the mean prior values), it is unlikely that such an approach could rival the goodness-of-fit resulting from the application of the MCMC method (or any other formal parameter estimation technique, for that matter). The large differences between Figures 3.4 and 3.5 clearly demonstrate the importance of model-specific parameter estimation.

The additional model uncertainty incurred by using the test-bed approach is reasonable and easily quantified (Figure 3.11). The mean difference between the emulator and GCM output, using the test-bed optimized parameters, is 16.8% for chl *a*, 1.7% for DOC, 11.4% for TDP, and 20.0% for POP. These errors do not appear to be systematic (i.e., the distribution of the discrepancies in Figure 3.11 is not identical from year to year). The source of this error is twofold: (1) computational artifacts arising from numerical solution techniques, time step size, etc., and (2) the influence of vertical advective and latitudinal transport processes on the biogeochemistry in the 3-D model, since these are absent in the 1-D emulator.

An advantage of the MCMC approach is the relative ease with which formal estimates of uncertainty may be made regarding predictions. Such estimates of uncertainty increase the utility of models as explanatory and predictive tools. In this study, we have quantified uncertainty in the estimated parameter values (Figure 3.6) and the resulting uncertainty in model output (Figures 3.7–3.9). Additional major sources of error include measurement uncertainty and model specification error. In relation to model specification error, measurement uncertainty is assumed to be small. Thus, we interpret any failure on the part of

the model to replicate specific signals in the data as structural inadequacies (i.e., the model does not contain the level of complexity required to adequately describe the system being modeled, or the mathematical formulations employed are erroneous).

While our calibrated model accurately reproduces the spatiotemporal distribution of chlorophyll (with the exception of the DCM) and particulate organic phosphorus in the lake (Figure 3.5), examination of the predicted primary production (Figure 3.9) indicates model kinetics are incorrect during the stratified period. This is despite the fact that primary production data were explicitly included in the model calibration scheme, and provides further evidence that a different model formulation is required to adequately describe the Lake Superior ecosystem. It appears as though the rate at which nutrients are made available to phytoplankton increases during stratification – presumably due to zooplankton grazing and/or the influence of the microbial loop. Since the model is constrained to adhere to a single nutrient cycling paradigm for the entire year, the calibration process tends to fit the behavior of the system during the time period for which the most data are available (i.e., the unstratified period). This result emphasizes the importance of utilizing kinetic data when calibrating biogeochemical models; concentrations of observed tracers can be accurately replicated by a model even if the underlying kinetics are poorly replicated by the model. The lack of a strong DCM signal in the model output also suggests that the true mechanisms responsible for this phenomenon are not present, or at least not described in adequate detail, in our simple model formulation.



Our cost function (Equation 3.11) weights model fit to each measurement of a given type equally. The sparse and heterogeneous nature of the observations therefore impacts the resulting parameterization. The ideal calibration data set is evenly spaced in both space and time, though in practice that is often not practical with environmental data collection. While beyond the scope of this study, it is possible to develop a cost function that attempts to correct for irregularly spaced data and space-time covariance (Stein, 2005; Fuentes, 2007). Additionally, the NRMSE associated with each data type is weighted equally in the overall cost function, so that model fit to chlorophyll ( $n = 146$ ) is given equal consideration as model fit to POP ( $n = 27$ ). Further, model performance with respect to certain variables is correlated (e.g., TDP, POP, and chl  $a$ ), such that parameterizations that favor improvement in fit to one data type likely improves the fit to the others as well, diminishing the influence of complementary constraining data (e.g., NPP) on the calibration.

## 3.6 Conclusions

We have demonstrated the improvement in model performance that may be attained by implementing a formal parameter estimation technique, and have outlined a method for accurately tuning the parameters of 3-D aquatic biogeochemical models. Owing to the significantly reduced computational demands of the 1-D hydrodynamic emulator, ecosystem models may feasibly be tested and parameterized *ex situ*, and the resulting parameter sets ported back to a 3-D hydrodynamic framework with an acceptably low increase in model

error (2–20%). Formal estimates of model uncertainty due to parameter error may also be developed in the emulator that may then also be translated to the 3-D model output. The simple model structure here appears inadequate to accurately describe ecosystem dynamics in Lake Superior, most notably failing to reproduce the high levels of primary productivity observed in the lake during the summer stratified period.



## Chapter 4

# Modeling carbon fluxes through the Lake Superior lower food web: A parsimonious semi-mechanistic approach

Perfection is achieved, not when there is nothing more to add, but when there is nothing left to take away.

---

*Antoine de Saint-Exupery (1900-1944)*

### 4.1 Abstract

Increasing model complexity is often assumed to lead to more realistic (and therefore more useful) models. This assumes, however, that the mathematics employed accurately simulate the modeled processes, and that appropriate parameter values can be determined using the available information. Sacrificing detail in favor of simple models that are well con-

strained by available observations can provide a useful foundation on which to build the necessary model complexity. Here we demonstrate the failure of NPZD-type models to accurately reproduce key signals in Lake Superior biogeochemical data. A simplified model, on the other hand, performs significantly better. Our model refines previous estimates of the carbon budget for the lake, and corroborates the importance of allocthonous inputs in subsidizing community respiration. Results indicate that approximately  $10 \text{ Tg C yr}^{-1}$  of respiration in the lake is supported by autotrophy ( $\sim 22\%$  of total community respiration), which agrees well with recent estimates of primary productivity. Our model also suggests the deep chlorophyll maximum in Lake Superior is formed primarily by shade adaptation, but is also influenced by particle settling. Data collection efforts should take into account the types of data with the greatest potential to constrain simulation models; such an approach would mitigate the development of unconstrained and overparameterized models and increase model reliability.

## 4.2 Introduction

Lake Superior is one of the earth's largest fresh water resources, yet biogeochemical cycling of carbon in this system is not yet fully understood. The lake is an ultraoligotrophic system characterized by low nutrient concentrations (total phosphorus  $2\text{--}4 \mu\text{M}$ ) and primary productivity, a dominant microbial loop, and a carbon budget dominated by dissolved organic matter (Cotner et al., 2004; Urban et al., 2005). Community respiration rates in the

lake exceed simultaneously measured rates of production (Urban et al., 2004), suggesting the lake is net heterotrophic; allochthonous inputs appear to provide an additional source of carbon for heterotrophy (Cotner et al., 2004). The magnitude of CO<sub>2</sub> emissions from the lake appears to be regionally significant, though the interaction between the lake and the surrounding terrestrial ecosystems remains unclear (Desai et al., 2008).

Evaluation of the carbon budget for the lake has resulted in an imbalance between estimated carbon sources and sinks, indicating that our understanding of the carbon cycle in the lake remains incomplete. Whole-lake budgets suggest that losses (dominated by respiration) exceed inputs (apparently dominated by photoautotrophy) by a factor of 1.5–6.6 (Cotner et al., 2004) or 3.4–20 (Urban et al., 2005). Recent in-situ measurements of productivity indicate that rates of carbon fixation in the lake are 1.2–3.2 times higher than previously estimated (Sturner, 2010), but a significant gap between productivity and respiration remains.

A deep chlorophyll maximum (DCM) forms during the stratified period (Auer and Bub, 2004; Barbiero and Tuchman, 2004), typically occurring below the thermocline. Primary production appears to peak during the period of DCM formation, but most production occurs at the surface rather than in the region of the DCM (Sturner, 2010). The mechanism(s) responsible for the formation of the DCM remains uncertain. Elucidation of the drivers of this strong vertical signal could help resolve biogeochemical carbon cycling in the lake on sub-annual time scales.

### **4.2.1 Parsimony in aquatic biogeochemical models**

The principle of parsimony dictates that a model should be sufficient to explain the phenomena modeled while invoking the fewest assumptions. This principle is not always followed in the development of aquatic biogeochemical models, in which a large number of assumptions regarding the mathematical representation of the relationships between organisms and their environment as well as the values of the associated parameters must be made (this is the case even for relatively elementary models). Attempts have been made, however, to simplify models systematically (Denman, 2003; Cox et al., 2006; Crout et al., 2009), as well as to apply information-theoretic approaches to identify the level of model complexity supported by available data (Boulêtreau et al., 2008; McDonald and Urban, 2010).

In addition to the danger of increasing mathematical complexity beyond a level that can be supported by the available data there is an additional danger of added detail actually diminishing the performance (i.e., fit to data and predictive capability) of models. The possibility of overfitting (i.e., overtuning parameters to calibration data at the expense of generality) complex models is well recognized; data are often scarce, and moreover there is often a fundamental mismatch between the measured parameters of a real system and model variables (e.g., chlorophyll *a* vs. phytoplankton biomass). Nevertheless, it is generally accepted that adding detail to models makes them more “realistic”. However, model dysfunctionality may arise from a model specification that provides an incorrect or inadequate description of the ecosystem (Flynn, 2005), and the probability of inaccurately describing

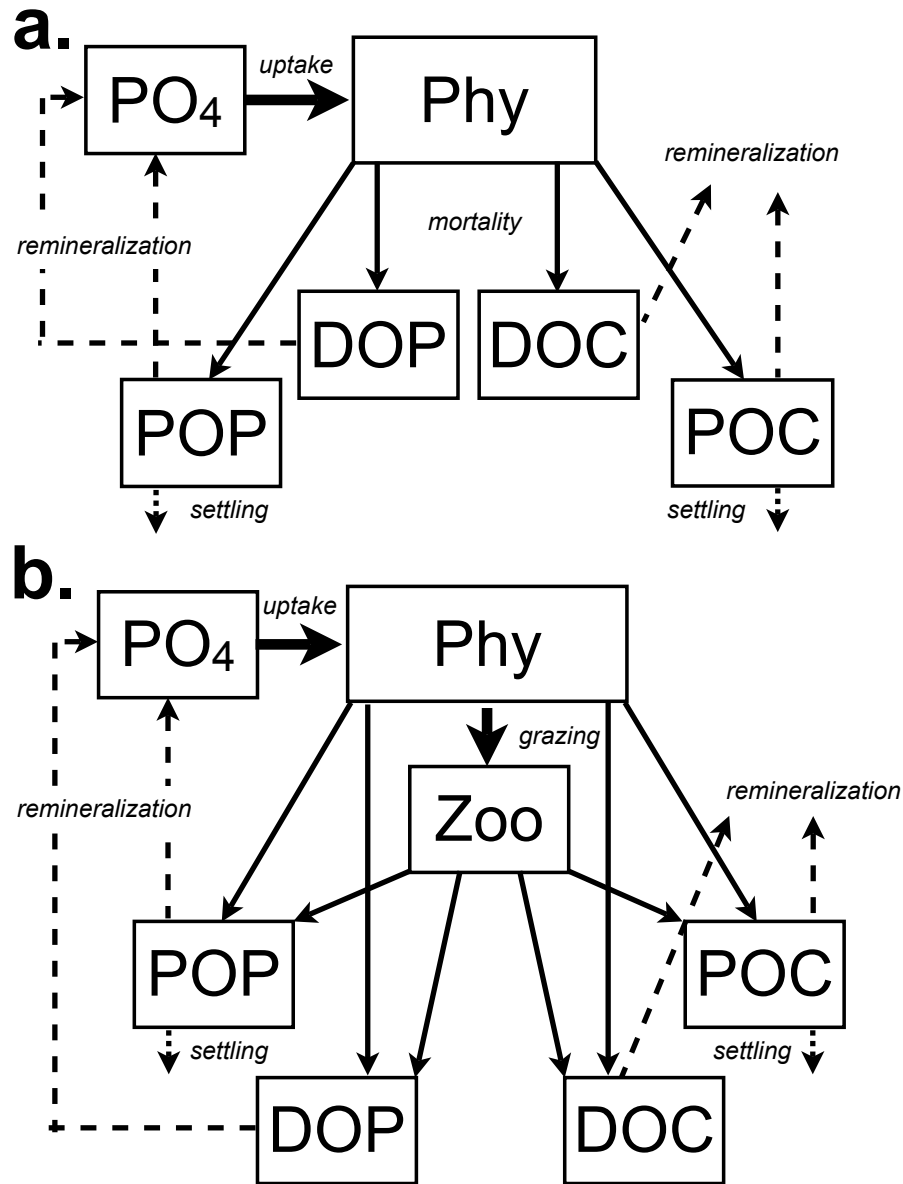
the system increases with increasing detail. Even when the model does accurately describe reality, there is generally too few data types (particularly data that describe model kinetics) available to calibrate the model. Indeed, in several cases, reduced or simpler models have been demonstrated to provide a better fit to calibration and validation data than their more complex counterparts (Cox et al., 2006; Crout et al., 2009).

### **4.2.2 Motivation**

We have previously developed and formally calibrated a simple 1-D biogeochemical model (nutrients/phytoplankton/detritus) for Lake Superior (Chapter 3). Despite rigorous calibration, the model poorly reproduced the spatiotemporal patterns of primary productivity, chlorophyll, and nutrients. Zooplankton grazing on phytoplankton in the surface layer could, in theory, reconcile observed profiles of primary productivity and chlorophyll; an explicit representation of this process may be required in a model formulation for the lake. Additionally, since the concentration data available for calibration (chl *a*, nutrients, etc.) outweigh the kinetic data available ( $^{14}\text{C}$  uptake), a weighted cost function may be required to identify appropriate parameter values. Sterner (2010) showed that primary productivity could be accurately predicted using only photosynthetically active radiation (PAR) and water temperature, suggesting that explicit representation of the phosphorus cycle may not be required in a Lake Superior ecosystem model to simulate carbon dynamics. Considering these factors, we examine the influence of model formulation on performance, with an em-



phasis on our ability to constrain model parameters using available data. We also identify the facets of the carbon cycle in Lake Superior that are least supported by data (i.e., the areas of our model containing the greatest uncertainty).



**Figure 4.1:** NPD (a) and NPZD (b) model structures.

### 4.3 Methods

Nutrient-phytoplankton-detritus (NPD) and nutrient-phytoplankton-zooplankton-detritus (NPZD) models were developed, based on the model of Dutkiewicz et al. (2005) (Figure 4.1). These models explicitly represent  $\text{PO}_4^{3-}$ , dissolved organic phosphorus and carbon (DOP and DOC), and particulate organic phosphorus and carbon (POP and POC). The degradation pathways are simple, and both particulate and dissolved pools are remineralized at the same rate. The particulate pools settle, with the exception of algal biomass. Variable chlorophyll content of algal cells is formulated based on Doney et al. (1996). Complete model equations for the NPD model are provided in Section 3.3.2. The nutrient-phytoplankton-zooplankton-detritus (NPZD) model is defined as follows (phytoplankton growth, phosphate and  $chl : P$  are formulated identically to the NPD model:

$$\frac{\partial Phy}{\partial t} = \mu \cdot Phy - mort \cdot Phy - graz \cdot Zoo + D' \frac{\partial^2 Phy}{\partial z^2} \quad (4.1)$$

$$\frac{\partial Zoo}{\partial t} = \phi \cdot graz \cdot Zoo - mort_Z \cdot Zoo + D' \frac{\partial^2 Zoo}{\partial z^2} \quad (4.2)$$

$$\begin{aligned} \frac{\partial DOP}{\partial t} = & f_P \cdot mort \cdot Phy + f_{P,mort,Z} \cdot mort_Z \cdot Zoo + f_{P,graz}((1 - \phi) \cdot graz \cdot Zoo) \\ & - remin_P \cdot DOP \cdot \frac{1}{\theta_{norm}} \theta_{base}^T + D' \frac{\partial^2 DOP}{\partial z^2}. \end{aligned} \quad (4.3)$$

$$\begin{aligned}
\frac{\partial POP}{\partial t} = & (1 - f_P) \cdot mort \cdot Phy + (1 - f_{P,mort,Z}) \cdot mort_Z \cdot Zoo \\
& + (1 - f_{P,graz})((1 - \phi) \cdot graz \cdot Zoo) \\
& - remin_P \cdot POP \cdot \frac{1}{\theta_{norm}} \theta_{base}^T + D' \frac{\partial^2 POP}{\partial z^2} + v_s \frac{\partial POP}{\partial z}.
\end{aligned} \tag{4.4}$$

$$\begin{aligned}
\frac{\partial DOC}{\partial t} = & x_{C:P} \cdot f_C \cdot mort \cdot Phy + f_{C,mort,Z} \cdot mort_Z \cdot Zoo \\
& + f_{C,graz}((1 - \phi) \cdot graz \cdot Zoo) \\
& - remin_C \cdot DOC \cdot \frac{1}{\theta_{norm}} \theta_{base}^T + D' \frac{\partial^2 DOC}{\partial z^2},
\end{aligned} \tag{4.5}$$

$$\begin{aligned}
\frac{\partial POC}{\partial t} = & x_{C:P} \cdot (1 - f_C) \cdot mort \cdot Phy + (1 - f_{C,mort,Z}) \cdot mort_Z \cdot Zoo \\
& + (1 - f_{C,graz})((1 - \phi) \cdot graz \cdot Zoo) \\
& - remin_C \cdot POC \cdot \frac{1}{\theta_{norm}} \theta_{base}^T + D' \frac{\partial^2 POC}{\partial z^2} + v_s \frac{\partial POC}{\partial z},
\end{aligned} \tag{4.6}$$

$$graz_z = graz' \left( \frac{\max(Phy - grazlim, 0)}{\max(Phy - grazlim, 0) + K_{graz}} \right) \tag{4.7}$$

In both the NPD and NPZD models, NPP is defined simply as  $\mu \cdot Phy \cdot x_{C:P}$ . Parameter descriptions are given in Table 4.1.

A simplified model was also formulated that describes algal growth (net primary production, NPP\*) in terms of only temperature and light availability (Sterner, 2010), and algal losses by a first-order, temperature-adjusted rate constant (Figure 4.2). Biomass is specified in carbon units. A variable chlorophyll to carbon ratio was computed as a function

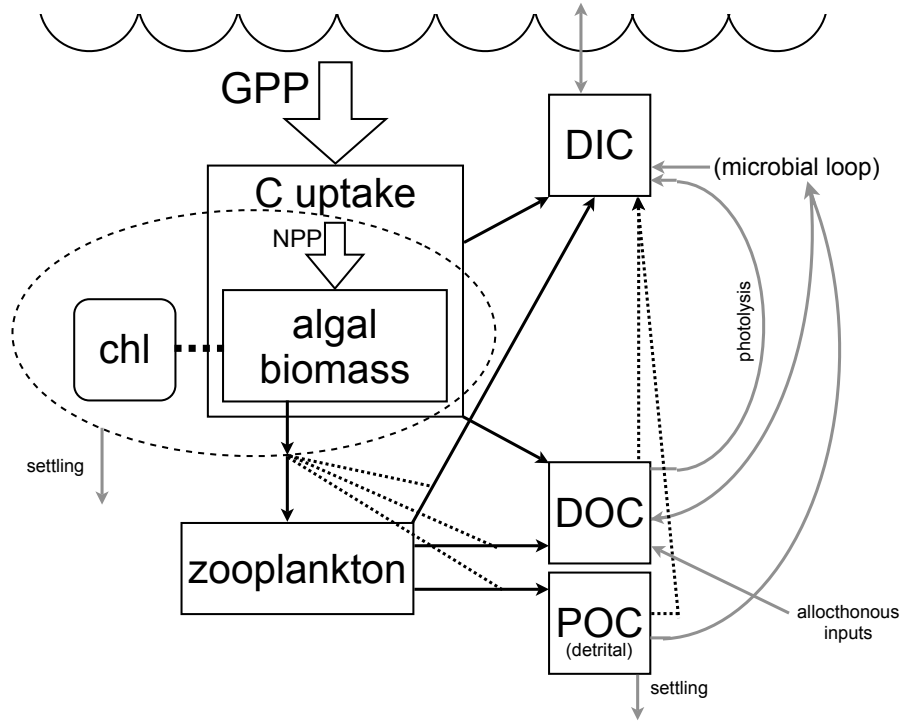
**Table 4.1:** NP[Z]D model parameter descriptions and posterior distributions.

Parameter	Description	Units	Posterior - NPD		Posterior - NPZD	
			Mean	S.D.	Mean	S.D.
$k_0$	Light extinction coefficient	$\text{m}^{-1}$	0.16	-	0.16	-
$k_c$	Light extinction due to algae	$\text{m}^{-1} (\mu\text{g chl L}^{-1})^{-1}$	0.015	-	0.015	-
$\hat{\mu}$	Maximum growth rate	$\text{d}^{-1}$	1.39	0.21	2.19	0.25
$Mort$	Mortality rate	$\text{d}^{-1}$	0.0047	0.0015	0.037	0.020
$K_I$	PAR half-saturation coefficient	$\text{W m}^{-2}$	7.30	3.87	12.6	5.07
$K_P$	$\text{PO}_4$ half-saturation coefficient	$\mu\text{M}$	0.45	0.19	0.76	0.36
$remin_P$	Phosphorus remineralization rate	$\text{d}^{-1}$	0.0072	0.0019	0.13	0.08
$remin_C$	Carbon remineralization rate	$\text{d}^{-1}$	0.0020	0.0009	0.049	0.020
$f_P$	Fraction of mortality to DOP	-	0.92	0.02	0.97	0.02
$f_C$	Fraction of mortality to DOC	-	0.96	0.03	0.91	0.05
$x_{C:P}$	Molar C:P	-	219	21	234	12
$phymin$	Minimum phytoplankton biomass	$\mu\text{M}$	1e-5	-	1e-5	-
$v_P$	Sinking velocity of POP	$\text{m d}^{-1}$	0.57	0.06	0.63	0.05
$v_C$	Sinking velocity of POC	$\text{m d}^{-1}$	0.66	0.06	0.60	0.06
$Chl : P_{max}$	Maximum ratio of chlorophyll to P	$\mu\text{g } \mu\text{mol}^{-1}$	29.5	1.5	27.6	1.6
$Chl : P_{min}$	Minimum ratio of chlorophyll to P	$\mu\text{g } \mu\text{mol}^{-1}$	7.8	0.3	7.6	0.3
$I^*$	Light coefficient for chlorophyll	$\text{W m}^{-2}$	32.6	8.8	17.2	7.2
$\theta_{base}$	Temperature base coefficient	-	1.046	0.021	1.079	0.022
$\theta_{norm}$	Temperature normalization coeff.	-	1.40	-	1.40	-
$DOC_{refrac}$	Refractory DOC	$\mu\text{M}$	99.1	2.1	99.0	3.3
$gfaz$	Zooplankton max grazing rate	$\text{d}^{-1}$	-	-	0.78	0.48
$mort_z$	Zooplankton mortality rate	$\text{d}^{-1}$	-	-	0.053	0.024
$K_{graz}$	Grazing half-saturation coefficient	$\mu\text{M}$	-	-	0.074	0.019
$\phi$	Assimilation efficiency	-	-	-	0.24	0.15
$f_{P,mort,Z}$	Fraction of mortality to DOP	-	-	-	0.66	0.20
$f_{P,graz,Z}$	Fraction of grazing to DOP	-	-	-	0.84	0.12
$f_{C,mort,Z}$	Fraction of mortality to DOC	-	-	-	0.61	0.15
$f_{C,graz,Z}$	Fraction of grazing to DOC	-	-	-	0.72	0.20
$zoomin$	Minimum zooplankton biomass	$\mu\text{M}$	-	-	1e-4	-
$grazlim$	Minimum Phy biomass for grazing	$\mu\text{M}$	-	-	5e-5	-

of light availability using a formulation identical to that of the NP[Z]D models. The core of the simplified model, representing the extent of the model that could be strongly constrained by observations, consists of only algal biomass in carbon units and chlorophyll a (parameter descriptions are provided in Table 4.2):

$$\frac{\partial Phy}{\partial t} = NPP - mort \cdot Phy \cdot \theta_m^T + D' \frac{\partial^2 Phy}{\partial z^2} + v_s \frac{\partial Phy}{\partial z}, \quad (4.8)$$

where net primary production,  $NPP$ , is calculated in  $\mu\text{mol L}^{-1} \text{d}^{-1}$  as:



**Figure 4.2:** Biogeochemical cycling of carbon in Lake Superior. The dashed ellipse demarcates the portion of the cycle represented by the simplified model structure. Solid lines represent real carbon transport pathways; dotted lines indicate “short-circuits” implemented in the extended simplified model while gray lines indicate the processes that are aggregated within these short-circuits.

$$NPP = C \exp \left[ \frac{-E_a}{8.62 \times 10^{-5} \cdot T} \right] P_{opt} \left( 1 - \exp \left[ \frac{-\alpha \cdot PAR}{P_{opt}} \right] \right) \div 12 \quad (4.9)$$

after Sterner (2010), with  $T$  in K and  $PAR$  in  $W \, m^{-2}$ . Chlorophyll concentrations are calculated according to Equation 3.8 and replacing  $P$  with  $C$ .

All models were implemented in a 1-D hydrodynamic emulator, described in detail in Section 3.3.3. Vertical diffusivity and water temperature output from a 3-D hydrodynamic

**Table 4.2:** Simplified model parameter descriptions, values, and posterior distributions. The mean and standard deviation are given for parameters were included in the MCMC procedure. Values below the horizontal line are based on literature values and implemented in the extended simplified model (see text for sources).

Parameter	Description	Units	Posterior Mean	S.D.
$k_0$	Light extinction coefficient	$\text{m}^{-1}$	0.15	-
$k_c$	Light extinction due to algae	$\text{m}^{-1} (\mu\text{g chl L}^{-1})^{-1}$	0.015	-
$C$	Fitting constant	-	1158	-
$E_a$	Activation energy	eV	0.282	-
$P_{opt}$	Max. rate of production	$\mu\text{g L}^{-1} \text{d}^{-1}$	838	-
$\alpha$	Light dependence of production	$\text{m}^2 \text{W}^{-1}$	499	-
$mort$	Phytoplankton mortality rate	$\text{d}^{-1}$	0.038	0.016
$\theta_m$	Mortality temperature coefficient	-	1.057	0.014
$Chl : C_{min}$	Maximum ratio of chl a to C	$\mu\text{g } \mu\text{mol}^{-1}$	0.076	0.028
$Chl : C_{max}$	Minimum ratio of chl a to C	$\mu\text{g } \mu\text{mol}^{-1}$	0.219	0.083
$I^*$	Light coefficient for chlorophyll	$\text{W m}^{-2}$	11.1	5.5
$\rho_{phy}$	Phytoplankton particle density	$\text{g cm}^{-3}$	1.027	-
$d$	Phy. particle effective diameter	$\mu\text{m}$	27.1	6.1
$f_{res}$	Ratio of respiration to NPP	-	1.35	-
$f_{exc}$	Ratio of excretion to NPP	-	0.35	-
$f_{DOM}$	Fraction of Phy mortality to DOM	-	0.14	-
$f_{POM}$	Fraction of Phy mortality to POM	-	0.57	-
$remin_{DOM}$	Remineralization rate of DOM	$\text{d}^{-1}$	0.01	-
$remin_{POM}$	Remineralization rate of POM	$\text{d}^{-1}$	0.01	-
$\theta_r$	Remineralization temp. coeff.	-	1.06	-
$v_p$	Settling rate of POM	$\text{m d}^{-1}$	0.2	-
$DOC_{refrac}$	Refractory DOC	$\mu\text{M}$	112.5	-

model of Lake Superior (Bennington et al., 2010; Bennington, 2010), along with incident radiation (Mesinger et al., 2006), were used to force the 1-D emulator, such that the resulting parameterization is transferrable to the 3-D model. Settling of POP and POC in the models was simply formulated using a constant settling velocity. Algal particles in the simplified model, however, settle according to Stoke's law, taking into account water density, particle density (assumed to be a constant  $1.027 \text{ g cm}^{-3}$ ), and effective particle diameter.

The models were calibrated for a site 21 km offshore of the Keweenaw Peninsula (47° 24.33'N, 88° 44.29'W) using data collected in 1999–2000 as part of the Keweenaw Interdisciplinary Transport Experiment in Superior (KITES) (Auer and Johnson, 2004; Green and Eadie, 2004). Observations included chlorophyll *a* (chl *a*), total dissolved phosphorus (TDP), POP, and POC. NP[Z]D model output was modified to be compatible with available measurements; detrital POP was combined with algal and zooplankton biomass to compare with measured POP, and modeled phosphate and DOP were combined to compare with measured TDP. The KITES data were augmented with the in-situ <sup>14</sup>C uptake measurements of Sterner (2010), referred to here as NPP\*.

Model calibration was performed via Markov Chain Monte Carlo (MCMC) using the Delayed Rejection Adaptive Metropolis algorithm (DRAM) (Haario et al., 2006). For more detail on this method, see Section 3.3.5. The objective function was defined as:

$$\sum_{i=1}^5 w_i \sqrt{\frac{\sum_{n=1}^{N_i} (f_{i,n}(\theta, x) - y_{i,n})^2}{N_i}} \div IQR_i \quad (4.10)$$

where  $N$  is the number of data of type  $i$ ,  $f$  is model output as a function of estimated parameters,  $\theta$ , and forcing data,  $x$ , and  $y$  is the observations. The interquartile range ( $IQR$ ) was used for normalization to mitigate the impact of outliers on the cost function. The weighting factor,  $w$ , was typically set to 1. For calibration of the NPZD model, however,  $w_{NPP}$  was set to 3 because TDP, POP, and chl *a* are highly covariant, and a parameterization that favors a better fit to these three variables at the expense of the fit to NPP\* makes a

disproportionate impact on the overall model goodness-of-fit. The NP[Z]D models were calibrated to all data types simultaneously. The simplified model was calibrated in two steps:

- Because the NPP\* equation was previously parameterized by Sterner (2010), an initial adjustment of the four associated parameters ( $C, E_a, P_{opt}, \alpha$ ) was performed using nonlinear least-squares regression. The original values of  $C, E_a, P_{opt}$  were confirmed to be near optimal, and  $\alpha$  was adjusted to accommodate the use of average rather than integral daily PAR.
- Holding NPP\* parameters constant, the model was calibrated via the MCMC method described above but using only chl *a* data (i.e, excluding nutrients).

The resulting posterior parameter values are provided in Tables 4.1 and 4.2.

## 4.4 Results and Discussion

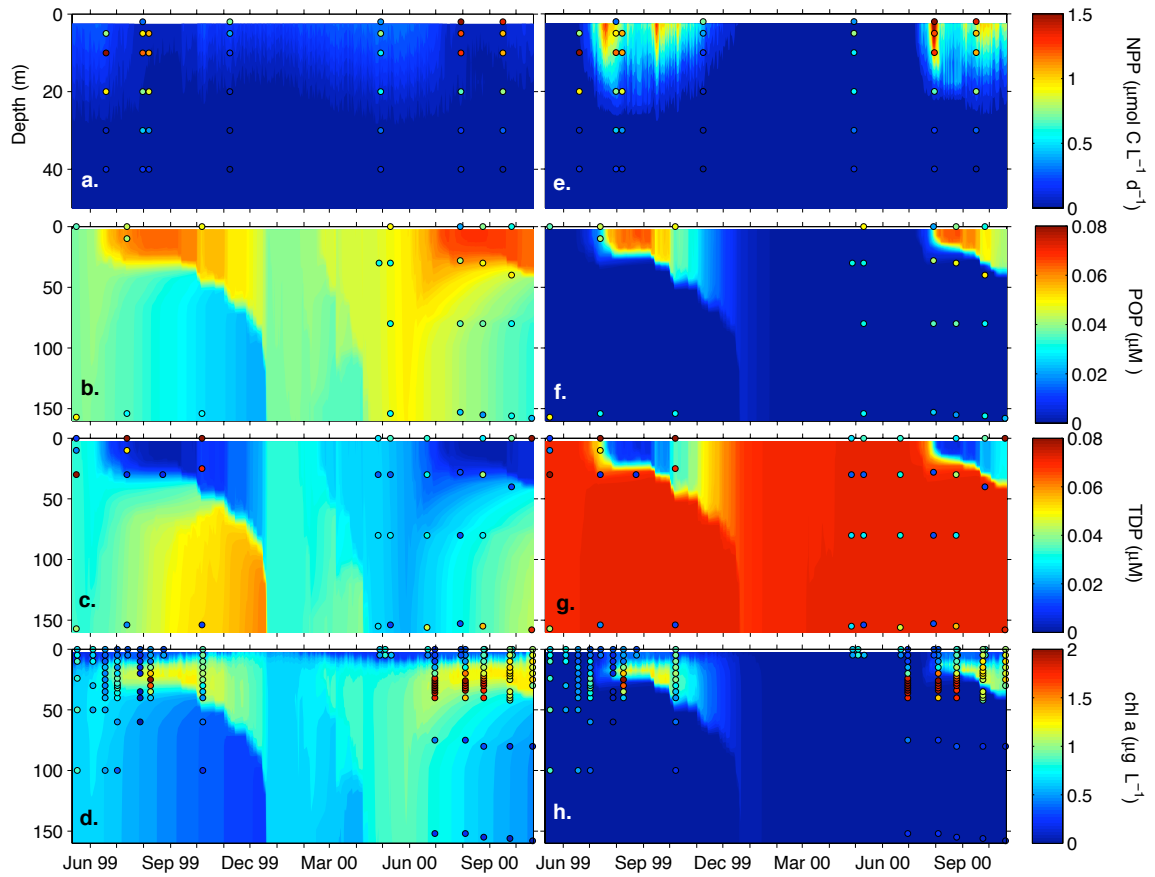
NPD and NPZD model output is shown together with calibration data in Figure 4.3. In the NPD model, chlorophyll profiles (Figure 4.3d) appear to be generally consistent with data, exhibiting increased concentrations throughout the water column in the spring, a DCM during the stratified period (though not located at the proper depth), and increased concentrations in the mixed layer during fall turnover. Modeled nutrient profiles (Figures



4.3b,c) show much more pronounced variability than is actually observed; it would appear as though modeled phosphate is depleted in the epilimnion during stratification, and maintained at low levels until fall. Examination of modeled NPP reveals a major discrepancy between modeled rates of carbon fixation and observations. Not only are predicted rates of NPP much lower than observed, the seasonal pattern is inverted relative to reality (i.e., higher productivity is predicted in the winter and spring, while production nearly ceases in the summer). When one views model output for NPP\*, POP, and DOP together (Figure 4.3), it becomes apparent that while the model formulation may be accurate for the unstratified period (to the extent that we can characterize the unstratified period with the available data), it does not accurately reproduce primary productivity during the warmer months. Calibrated and modeled rates (growth/loss/remineralization) are all relatively low (Table 4.1).

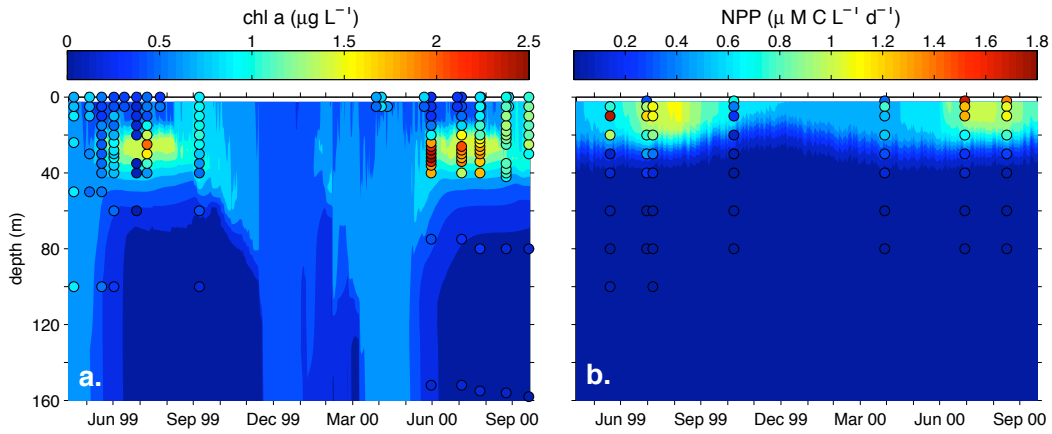
It is important to note that the distribution of the standing crop of chlorophyll in space and time agrees fairly well with observations despite the unrealistically modeled productivity. This illustrates the importance of considering kinetic data when evaluating biogeochemical models, as **concentration data can easily be reproduced without properly identifying the dynamic processes that produce the concentrations.**

Although the remineralization rate of nutrients in the NPD model is dependent upon water temperature, the rate of P turnover does not increase with temperature sufficiently in the summer to allow phytoplankton to continue to grow. Model results suggest that turnover



**Figure 4.3:** NPD (a-d) and NPZD (e-h) model output for NPP, POP, TDP, and chl *a*. Symbols indicate calibration data; nutrients and chlorophyll are from the KITES project and NPP is from Sterner (2010). While the explicit representation of zooplankton in the model does improve model performance with respect to productivity, overall model goodness-of-fit is diminished.

of phosphorus (and biomass) should occur more quickly in the epilimnion. When the cost function is altered to place more weight on NPP\* data, predicted NPP\* matches observations more closely (Figure 4.3). There remains, however, an inverse relationship between POP\* and NPP during the stratified period, suggesting the role of nutrient limitation on algal growth is still exaggerated. The model produces two peaks in NPP\* in the summer, at



**Figure 4.4:** Simplified model output for chl a and uncorrected NPP. Although nutrients and zooplankton are not explicitly modeled, the simplified model outperforms both the NPD and NPZD models with respect to both chlorophyll and production.

the beginning and end of the stratified period, bracketing a period of reduced productivity corresponding to nutrient depletion. The modeled chlorophyll profile does not appear to be as accurate as that produced by the NPD model; overall concentrations are lower than observed and the spring bloom is not resolved.

The simplified model does an excellent job of reproducing observed carbon fixation rates using only temperature and light as independent variables (Figure 4.4b). Chlorophyll is also modeled accurately (Figure 4.4a); while predicted concentrations are slightly less than measurements in the DCM ( $0.5\text{--}1.0\ \mu\text{g L}^{-1}$  lower), the depth of the DCM agrees well with observations. Spring and fall blooms also agree well with data.

A quantitative comparison of goodness-of-fit of the three models is provided in Table 4.3. As is apparent from model output (Figure 4.3), the addition of zooplankton and weighting NPP in the NPZD model significantly improves the fit of modeled NPP to observa-

**Table 4.3:** Model normalized root mean square error (NRMSE) for individual variables and weighted cost function ( $\Sigma$ ).

	NPD	NPZD	Simplified
Chl <i>a</i>	0.67	1.26	0.60
TDP	0.93	1.38	-
POP	0.77	2.43	-
DOC	1.86	2.41	1.83 <sup>1</sup>
NPP*	0.78	0.54	0.32
$\Sigma$	6.58	9.10	-

<sup>1</sup> Derived from the extended version of the simplified model

tions (NRMSE of 0.54 as compared to 0.78, or a 69% reduction in error). However, the goodness-of-fit for all four of the other variables actually decreases when zooplankton are added (NRMSE increases), meaning the NPD model is able to fit the calibration data better overall. The simplified model fits NPP\* data much better than the NP[Z]D models (NRMSE = 0.32), and also achieves a better fit with the chlorophyll data (Table 4.3).

Although numeric metrics of fit indicate the NPD model is more accurate, an examination of model output suggests that the NPZD model may better reproduce temporal trends in productivity. In this case, adding complexity (zooplankton) to the model formulation does not improve quantitative model performance when all variables are considered. It improved the model fit to one measured variable (NPP), which was weighted more heavily, but reduced the fit to four others. Further, the inclusion of zooplankton increases the number of fitted parameters in the model by approximately 50%, yet no further data are available to identify parameter values. While zooplankton grazing in the summer would be expected to result in a more accurate simulation of algal biomass, without more information on grazing rates, or at least zooplankton biomass, it is impossible to accurately implement this process

in the model. The case could be made that since the NPZD model more accurately depicts seasonal trends in productivity (a critical component of the carbon cycle in the lake), it does “outperform” the NPD model in a sense even though it results in a greater overall discrepancy between model output and data.

The simplified model suggests it is not necessary to consider nutrients in order to model NPP\*. This does not suggest that nutrient limitation is unimportant on a lake-wide scale. On the contrary, it is known that phytoplankton and bacteria are phosphorus limited in Lake Superior (Sterner et al., 2004). Vertical gradients in phosphorus in the water column are weak (Baehr and McManus, 2003), but spatial gradients do exist across the lake, with higher P concentrations located nearshore, close to tributary inputs (Baehr and McManus, 2003; Weiler, 1978; Urban, 2009). The availability of phosphorus regulates algal growth in Lake Superior as a whole (Chapra, 1977), and perhaps laterally within the lake. The dominant vertical and seasonal trends in productivity, however, can be explained better as functions of just light and temperature than by also invoking phosphorus as was done in the NP[Z]D models. Even when considering data from disparate locations in the lake, the formulation utilized here explained 93% of the variability in carbon uptake at water depths >2 m (Sterner, 2010).

The improved performance of the simplified model also does not suggest that top-down regulation of phytoplankton biomass by zooplankton grazing is an insignificant component of the Lake Superior food web. This process is implicitly included in the simplified model,

albeit in an extremely aggregated form – as a temperature-dependent, first-order loss rate. If we had more information available on zooplankton (biomass, grazing rates), the addition of a more detailed representation of grazing (such as the one implemented in the NPZD model) could theoretically improve the performance of the simplified model. In the absence of zooplankton biomass measurements, such an approach would, in this case, represent an unwarranted increase in complexity (especially considering the satisfactory performance of the simpler model).

#### **4.4.1 Modeling rates vs. concentrations**

The simplified approach presented here represents a fundamentally different modeling approach than that represented by the NP[Z]D models. Typically, in a mechanistically formulated aquatic biogeochemical model, primary productivity is determined by not only independent environmental parameters (light, temperature), but also by feedbacks from within the system (algal biomass, nutrient availability). By making the assumption that these processes may be neglected in Lake Superior, we have isolated modeled NPP from the other model variable (chl *a*), which allows the associated parameters to be independently determined. Such an approach would clearly not be feasible in all ecosystems; when a strong relationship does exist between NPP and nutrient availability, for example, nutrient concentrations (and the associated rates) would need to be parameterized simultaneously with NPP.

While this approach results in a parsimonious model that is supported by the available calibration data (NPP\* and chl *a*), this approach is not without drawbacks; the utility of such a simple model is severely restricted because only a small portion of the carbon cycle is actually included (Figure 4.2).

#### **4.4.2 Extended simplified model**

The simplified model, on its own, does not provide a complete picture of carbon cycling in the lake, but does provide a basic foundation, supported by data, that can be expanded to encompass the complete carbon cycle. While the simplified model accurately predicts NPP and chlorophyll concentrations, in order to create a useful model of the carbon cycle in Lake Superior we must make adequate assumptions to relate gross primary production (GPP), respiration, excretion, grazing, settling, and remineralization to NPP. First, we aggregate these processes to the extent possible (see the “short-circuit” pathways depicted in Figure 4.2). Rather than explicitly modeling zooplankton biomass, we can simply define the fate of algal carbon subject to grazing losses (inorganic, dissolved, and particulate carbon fractions). Similarly, rather than explicitly modeling the ecosystem components contributing to remineralization processes (i.e., bacterial biomass and activity, photolysis), we define simple pathways from the dissolved and particulate carbon forms to dissolved

inorganic carbon (DIC):

$$\frac{\partial Phy}{\partial t} = GPP - mort \cdot Phy \cdot \theta_m^T - NPP \cdot f_{res} - NPP \cdot f_{exc} + D' \frac{\partial^2 Phy}{\partial z^2} + v_s \frac{\partial Phy}{\partial z} \quad (4.11)$$

where

$$GPP = (1 + f_{res} + f_{exc}) \cdot NPP \quad (4.12)$$

and the fate of phytoplankton “mortality” (mainly grazing by zooplankton) is tracked in the dissolved and particulate carbon pools, with the activity of the microbial loop aggregated into first-order “remineralization” terms:

$$\frac{\partial DOC}{\partial t} = NPP \cdot f_{exc} + mort_{phy} \cdot f_{DOM} \cdot Phy \cdot \theta_m^T - remin_{DOM} \cdot DOC \cdot \theta_r^T + D' \frac{\partial^2 DOC}{\partial z^2} \quad (4.13)$$

$$\frac{\partial POC}{\partial t} = mort_{phy} \cdot f_{POM} \cdot Phy \cdot \theta_m^T - remin_{POM} \cdot POC \cdot \theta_r^T + D' \frac{\partial^2 POC}{\partial z^2} + v_s \frac{\partial POC}{\partial z} \quad (4.14)$$

Parameter descriptions are given in Table 4.2.

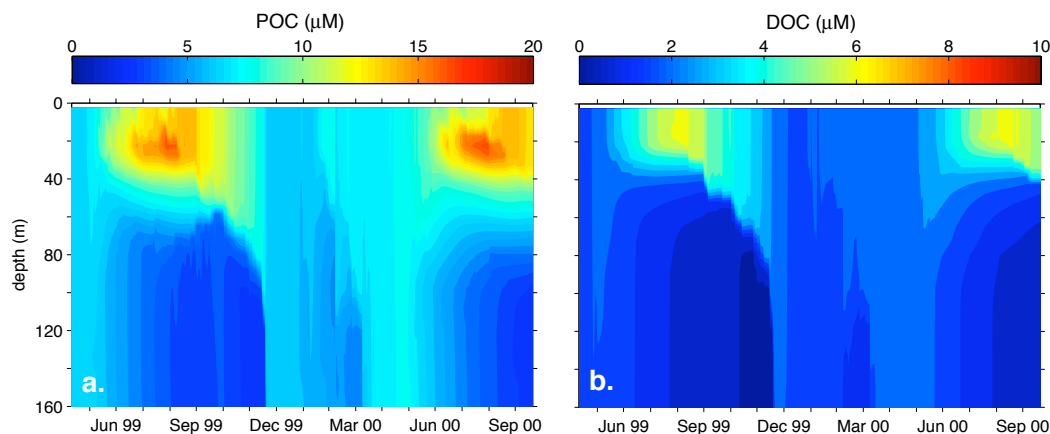
The additional parameters required for the extended simplified model could not be identified from data, thus their values were estimated from literature values (Table 4.2). Several studies (Bender et al., 1999; Laws et al., 2000) have shown  $^{14}\text{C}$ -derived primary production (net carbon fixation) to be equal to  $\sim 45\%$  of gross primary production (GPP) as measured via  $\text{O}_2$  production. Additionally, Bender et al. (1999) estimated that 15% of GPP measured via  $\text{O}_2$  production was due to the combined effects of photorespiration and the Mehler reaction. Excretion of DOC is approximated as 13% of NPP (Baines and Pace, 1991).



Combining these figures results in NPP being equal to 53% of GPP, excretion being equal to 7% of GPP, and the remainder (40%) of GPP being lost via mitochondrial respiration. Since measuring  $^{14}\text{C}$  uptake during daylight hours measures net carbon fixation during the day (i.e., photosynthesis less respiration and excretion), but does not account for the fraction of previously fixed carbon that is respired and excreted at night, we made an additional adjustment to Sterner's  $^{14}\text{C}$  uptake measurements/equation (Sterner, 2010) to account for the unmeasured fraction of carbon fixed via photosynthesis but lost at night to excretion and respiration. The average duration of *in situ*  $^{14}\text{C}$  uptake incubations was 16 h; taking into account the fraction of GPP lost to respiration and excretion estimated above and assuming these processes occur at a constant rate, NPP\* is higher than NPP by 16% of GPP (47% / 3). Therefore, actual NPP is equal to approximately 77% of  $^{14}\text{C}$  uptake:

$$NPP = 0.77NPP^* \quad (4.15)$$

The fractionation of phytoplankton mortality (assumed to be predominantly due to grazing) into DIC, DOC, and POC was specified based on the zooplankton model parameters described by Buitenhuis et al. (2006). Combining the effects of sloppy feeding, egestion, respiration, and mortality, we estimate that 14% of phytoplankton mortality is ultimately transformed to DOC, 50% to POC, and the remainder is respired. Both POC and DOC are assumed to be transformed to DIC according to a temperature-dependent, first-order rate



**Figure 4.5:** Extended simplified model output for POC and DOC. This output represents only the fraction of these pools derived from primary production.

formulation, with  $0.01 \text{ d}^{-1}$  as the base remineralization rate (Bowie et al., 1985). Particulate carbon is assumed to settle at a rate of  $0.2 \text{ m d}^{-1}$  (Chapra, 1997).

The POC and DOC output from the extended simplified model is shown in Figure 4.5. POC, as defined here, represents not only detrital POC, but also algal biomass. Zooplankton and bacterial biomass are not explicitly modeled. The predicted patterns of POC are similar to predicted patterns of chl a in time and space, though the variability is not as pronounced. Modeled DOC (of photosynthetic origin) ranges from  $1.9$  to  $6.2 \mu\text{M}$ , with elevated concentrations occurring in the surface layer during the stratified period. This seasonal increase is consistent with observations (Figure 4.5), but the magnitude is approximately  $100 \mu\text{M}$  lower than DOC concentrations commonly observed in Lake Superior. To simulate a refractory pool of DOC, a constant value was estimated using nonlinear least squares to obtain a best fit between model output and data; the refractory pool was esti-

mated in this manner to be  $112.5 \mu\text{M}$ . With this adjustment applied, the simplified model fits the DOC calibration data better than either NP[Z]D model (Table 4.3).

#### **4.4.3 Lake Superior C cycling**

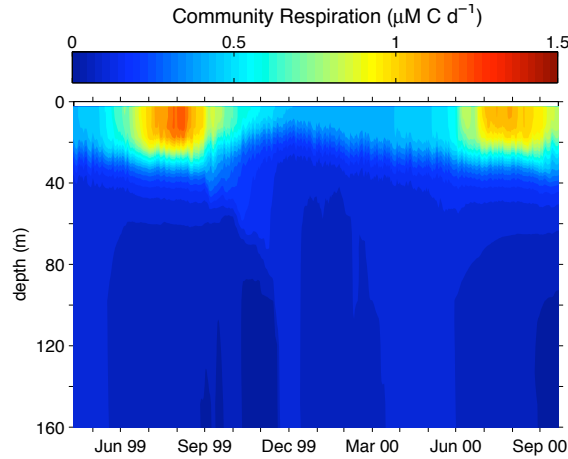
Model results support the existence of a large refractory DOC pool and a small, rapidly cycling labile DOC pool, as described by Urban et al. (2005). The modeled fraction of autochthonous DOC varies from  $\sim 5.5\%$  in the summer, which falls just within the 5–15% estimated for the summer epilimnion (Urban et al., 2005), to  $\sim 1.7\%$  in the winter. Results suggest that the seasonal increase in DOC observed in surface waters can only be partially explained by the combined effects of photoautotrophy and grazing (and offset by microbial heterotrophy). It must be noted, however, that the modeled ambient concentration of DOC is determined largely by the specified remineralization rate, our estimate of which is highly uncertain; thus model results can not be interpreted as proof that autochthonous carbon inputs alone could not be responsible for the seasonal variation in DOC. The constant calculated refractory DOC term ( $112.5 \mu\text{M}$ ) can be interpreted as a rough estimate of allochthonous DOC, although it is estimated that 13–19% of this pool is transformed and utilized for primary production (Cotner et al., 2004).

The model output can be used to estimate the amount of community respiration (CR) supported by primary production in the lake:

$$CR_{auto} = \text{algal respiration} + \text{zooplankton respiration} + \text{C remineralization}, \quad (4.16)$$

where C remineralization is equivalent to bacterial respiration. The resulting pattern of  $CR_{auto}$  is shown in Figure 4.6. Estimated CR in the surface layer ranges from  $\sim 0.4 \mu\text{M C d}^{-1}$  in the winter to  $\sim 1.3 \mu\text{M C d}^{-1}$  in the summer. This agrees well with the  $1.5 \mu\text{M C d}^{-1}$  (average value; range  $0.2\text{--}14 \mu\text{M C d}^{-1}$ ) reported for the epilimnion in the summer (Urban et al., 2004).

The spatial resolution of the model output allows for productivity and  $CR_{auto}$  estimates to be integrated through the water column to compare with previous estimates of annual fluxes. Doing so results in a value of between  $9.5$  and  $10.2 \text{ Tg C yr}^{-1}$  (lower and upper bounds based on numerical integration of GPP and  $CR_{auto}$  for 2000, respectively; while the  $CR_{auto}$  is constrained to be approximately equal to GPP in the model, error is introduced in the numerical integration of model output). This value ( $115\text{--}125 \text{ g C m}^{-2} \text{ yr}^{-1}$ ) is similar to the estimate of whole-lake annual net production made by Sterner (2010), though arrived at using a different methodology. It is lower, however, than previous estimates of community respiration of  $13\text{--}81 \text{ Tg C yr}^{-1}$  (Urban et al., 2005) and  $13\text{--}39 \text{ Tg C yr}^{-1}$  (Cotner et al., 2004). This is expected, since bacterial respiration has been shown to account for  $82\text{--}91\%$  of community respiration in the epilimnion during the stratified period (Biddanda



**Figure 4.6:** Modeled community respiration derived from photosynthesis,  $CR_{auto}$ , derived by combining modeled phytoplankton respiration, zooplankton respiration, and remineralization rates. Respiration of allocthonous carbon is not included, but results suggest that  $CR_{auto}$  only accounts for  $\sim 22\%$  of total CR.

et al., 2001), yet is constrained to approximately 40% of respiration by our model parameterization. To satisfy both our modeled rates of photosynthetically-supported community ( $\sim 10 \text{ Tg C yr}^{-1}$ ) and bacterial ( $\sim 4 \text{ Tg C yr}^{-1}$ ) respiration and the fact that  $\sim 87\%$  of CR in the lake is due to bacterial activity, an additional  $36 \text{ Tg C yr}^{-1}$  of allocthonous carbon must be remineralized.

While extrapolating our modeled rate of CR to the entire lake in this fashion provides a rough estimate of lakewide CR, it is likely that lateral variability in CR (and primary production) is significant in the nearshore region of the lake where riverine inputs are a factor. Modeling by Bennington (2010) suggests this is the case, and that taking such lateral variability into account may be the key to balancing the carbon budget for Lake Superior.

The water column is modeled in the 1-D emulator as a closed system; while settling takes place on a seasonal basis, permanent burial of carbon in the sediments is neglected. Therefore the value of  $CR_{auto}$  derived above is likely a slight overestimate. Sediment burial of carbon has been estimated at approximately  $0.5 \text{ Tg C yr}^{-1}$ , or  $\sim 5\%$  of our estimate of GPP.

While the simplified model presented here is applicable to the offshore regions of the lake, additional considerations may be required in the nearshore regions in order to improve the accuracy of lake-wide estimates. Although primary productivity can be accurately predicted without considering nutrient effects in the offshore waters where SRP concentrations are uniformly low, it is likely that increased nutrient availability about the periphery of the lake leads to increased productivity as evidenced by the higher chlorophyll concentrations found there (Auer and Bub, 2004), especially in the western arm of the lake (Munawar and Munawar, 1978). Labile terrestrial DOM inputs are presumably also transformed rapidly by heterotrophs in the nearshore region. If a complete picture of the carbon cycle of Lake Superior is to be developed, including explicit representation of air-water  $\text{CO}_2$  exchange, these factors may need to be considered in a 3-D implementation of the simplified model presented here.

#### 4.4.4 Deep chlorophyll maximum

The DCM in Lake Superior was first documented over four decades ago (Olson and Odlaug, 1966), yet the mechanisms responsible for its formation remain unclear (Auer and Bub, 2004; Barbiero and Tuchman, 2004). This region of maximum chlorophyll concentration typically occurs at approximately 30 m water depth, and contains chlorophyll concentrations approximately 2 times greater than epilimnetic concentrations. However, POC concentrations measured in August in the DCM are similar to those in the epilimnion (Barbiero and Tuchman, 2004). Some possible explanations for the DCM include increased nutrient availability at depth, reduced grazing at depth, settling of algal particles, and shade adaptation of phytoplankton (variable C:chl *a*).

The simple model includes three of these mechanisms (settling, grazing, and shade adaptation) and reproduces the DCM accurately while producing a less severe vertical gradient in particulate carbon (Figures 4.4 and 4.5), which is consistent with previous studies. While it is not immediately obvious which processes are driving DCM formation in the model, a sensitivity analysis can be employed to identify the relative importance of each. A binary factorial design was specified in which the effects of settling, grazing, and shade adaptation were explored. Each process was defined as activated or deactivated as follows:

- For settling, either algal particles settle according to Stoke's law, or they do not experience settling at all (i.e., vertical transport only occurs as a result of diffusive mixing).
- For grazing, we want to examine not the effect of predation on algae as a whole, but rather the effect of depth-dependent predation on algae. In the model, only the temperature dependence of phytoplankton mortality can simulate such a gradient. Therefore, phytoplankton mortality is either a function of temperature or it is not (i.e., it is held constant).
- For shade adaptation, either cell carbon to chlorophyll ratios are calculated according to Equation 3.8, modified to use carbon rather than phosphorus, or C:chl is held constant at  $C:chl_{max}$ .

In addition, the model was run with none of these processes activated. The scenarios are summarized in Table 4.4.

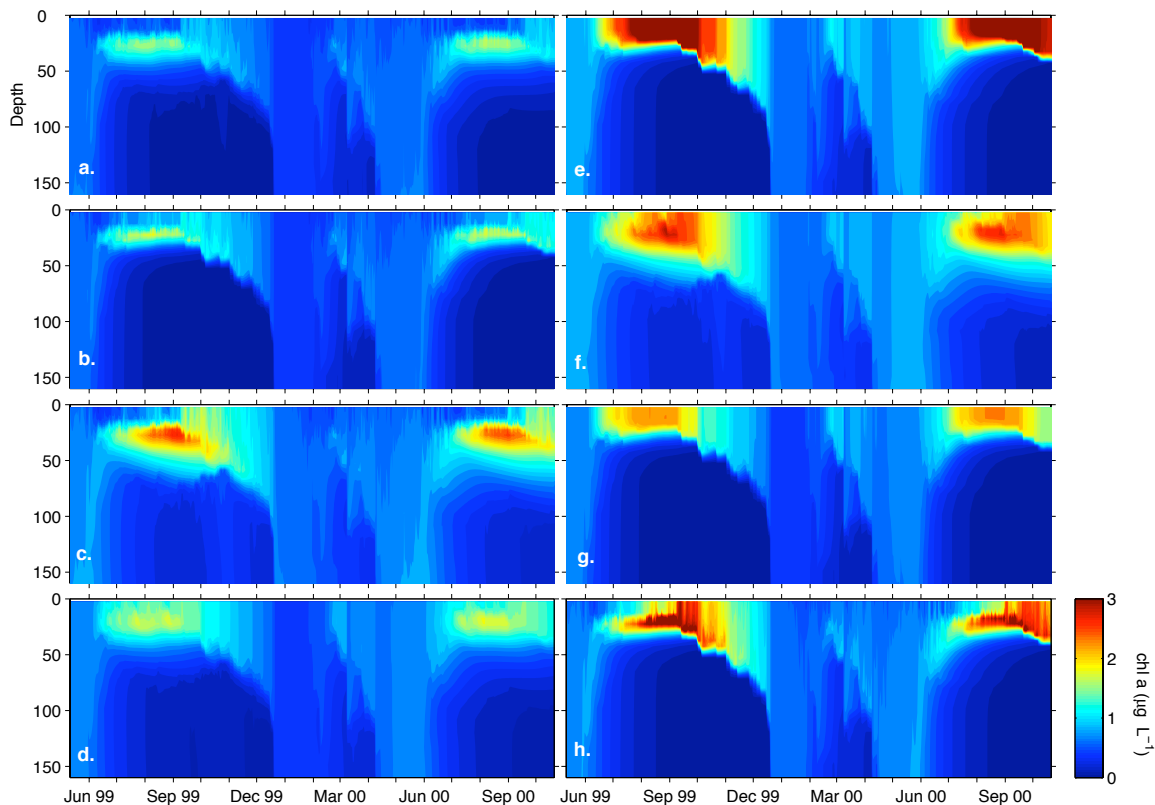
Omitting settling from the model (Figure 4.7b) had significantly less effect on the model's ability to resolve the DCM than did removal of temperature-dependent grazing or shade adaption (Figures 4.7c,d). Interestingly, however, the model run with only settling (i.e., grazing and shade adaptation deactivated) does form a DCM (Figure 4.7f), though not as well-defined as the data show (Figure 4.3). Removal of temperature-dependent grazing appears to decrease grazing pressure approximately an equal amount through the water



column, but does not appear to change the timing or location of the DCM (Figure 4.7c), whereas grazing alone (Figure 4.7g) creates no DCM at all, suggesting it is the least influential of the three processes on DCM formation. Shade adaptation alone (Figure 4.7h) also forms a DCM. In the absence of shade adaptation (Figure 4.7d), a less well-defined DCM forms, due to algal settling. While the goodness-of-fit metrics (Table 4.4) cannot be strictly interpreted as a measure of the importance of each process (the model was not re-calibrated for each scenario), their values do provide evidence for the relative influence each process has on modeled chlorophyll concentrations; they suggest that variable chlorophyll is the dominant driver. Examination of model output (Figure 4.7), however, reveals that phytoplankton settling also contributes to successfully reproducing the DCM in lake Superior, most notably by reducing epilimnetic chlorophyll concentrations. The simple representation of grazing activity in the model makes it impossible to rule out this process as an important driver of DCM formation, but the model does show that the DCM is able to develop in the absence of depth-dependent grazing pressure.

**Table 4.4:** Chlorophyll *a* sensitivity analysis results (values given are the normalized root mean square error (NRMSE) for chl *a*).

Algal particle settling	Temperature dependent grazing	Variable C:chl ratio	Panel in Figure 4.7	NRMSE
X	X	X	a	0.60
	X	X	b	0.74
X		X	c	0.92
X	X		d	0.93
			e	2.88
X			f	1.56
	X		g	1.53
		X	h	1.46



**Figure 4.7:** Sensitivity analysis (chl *a*) results. Each panel represents one of the scenarios in the factorial design (see Table 4.4 for descriptions).

#### 4.4.5 Future data requirements

Figure 4.2 reveals how ineffective existing data (mainly concentrations) are at constraining the parameterization of the extended simplified model. The ratio of net to gross primary production is implicitly defined in our model by  $f_{res}$  and  $f_{exc}$  as 0.37. This value is highly uncertain, however, and since more than 60% of GPP is modeled as contributing directly to DOC or DIC, the uncertainty associated with it is propagated throughout the model results. The most efficient and reliable way to improve confidence in model results would be to collect additional kinetic observations. Simultaneous measurements of gross and net primary

production would help to improve the reliability of a Lake Superior carbon cycle model. Secondly, direct measurements of remineralization/bacterial activity would constrain our coarse estimation of carbon remineralization and support a more detailed representation of microbial processes. Finally, measurements of zooplankton process rates are needed to better resolve the fate of fixed autochthonous carbon, and to improve understanding of the role grazing plays in the spatiotemporal distribution of carbon stocks. Simultaneous measurement of all of the carbon pools shown in Figure 4.2 would also increase our ability to accurately parameterize the model. These measurements are not ranked in order of ecological importance, *per se*, but rather by their potential to reduce uncertainty significantly and/or to warrant increases in complexity in the model presented here.

## 4.5 Conclusions

While conventional wisdom may suggest that the application of a more sophisticated model structure results in greater flexibility and thus a greater ability to mimic the system being modeled, this is not always true. We have demonstrated a case in which adding detail (zooplankton) to a typical NPD model, using widely employed formulations, actually decreases overall model performance. The role of nutrient limitation in vertical algal dynamics in the Lake Superior ecosystem appears to be much less significant than our NPZD formulation implies. On the other hand, a simple model in which primary productivity is only dependent upon light and temperature, and in which all processes that are unconstrained by observa-

tions are represented in the simplest manner possible, is able to better fit the calibration data.

Our parsimonious model supports previous studies examining the carbon cycle of Lake Superior (Cotner et al., 2004; Urban et al., 2005; Sterner, 2010). We suggest that primary production in the lake is higher than previous estimates, yet a significant gap remains between autotrophy and heterotrophy in the lake. Our model estimate of community respiration supported by autotrophy in Lake Superior is  $10 \text{ Tg C yr}^{-1}$ , which is between 12 and 77% of estimates of total community respiration; the remainder may be due to bacterial respiration of allocthonous carbon sources. Further exploration of spatial (lateral) variability in productivity and respiration may also help close the gap. Processing of autochthonous carbon may account for the seasonal variation in dissolved organic carbon observed in the offshore regions of the lake, though in the absence of additional constraining data there is a great deal of uncertainty surrounding the processes controlling DOC concentrations. Our model also suggests that the deep chlorophyll maximum (DCM) in Lake Superior can be explained by a combination of algal shade adaptation and settling. Future efforts to simulate biogeochemical cycling in Lake Superior should be preceded by additional data collection; of primary importance is clarifying the relationship between gross and primary productivity in the lake.



# Chapter 5

## Modeling historical trends in Lake Superior total nitrogen concentrations

History is the only science enjoying the ambiguous fortune of being required to be at the same time an art.

---

*Johann Gustav Droysen (1808-1884)*

### 5.1 Abstract

Nitrate concentrations in Lake Superior increased fivefold between 1900 and 1980, and have remained nearly constant since that time. Such rapid changes in concentration in a lake with a long hydraulic residence time ( $\sim 190$  years) are surprising. We developed a model to better understand the causes of the historical changes and to predict future changes in nitrate concentrations. Historical loadings were reconstructed based on average national

---

The material contained in this chapter was previously published in the *Journal of Great Lakes Research*: McDonald, C.P., et al. 2010. Modeling historical trends in Lake Superior total nitrogen concentrations. *J Great Lakes Res*, doi:10.1016/j.jglr.2010.07.008.

NO<sub>x</sub> emissions estimates, recent (past ~30 years) atmospheric N deposition data, recent tributary concentration data, and basin-wide runoff estimates. Increases in atmospheric N deposition alone were insufficient to have resulted in the observed trends. However, model runs combining increased atmospheric deposition with increased tributary N loading and/or decreased burial+denitrification mid-century reproduced the observed accumulation of N. Because internal N fluxes are an order of magnitude greater than external fluxes, relatively small changes in the lake's internal N cycle may produce relatively large changes in total N concentrations. Land-use changes in the watershed, particularly increases in logging activity, may have altered riverine N inputs. Regardless of the historical mechanisms leading to the rise in nitrate concentrations, it appears as though the system is currently at or is approaching peak N content.

## **5.2 Introduction**

Nitrate concentrations in Lake Superior increased approximately fivefold over the past century, resulting in a severe stoichiometric imbalance in the lake (Sterner et al., 2007). The upward trend in nitrate concentrations was first documented by Weiler (1978), and was examined in more detail by Bennett (1986) and more recently by (Sterner et al., 2007). Atmospheric emissions of NO<sub>x</sub> increased tenfold over the same period (U.S.EPA, 2000); atmospheric nitrate deposition was previously thought to fully explain the rise in lake nitrate concentrations (Bennett, 1986). A stable nitrogen isotope study revealed very low

$\delta^{15}\text{N-NO}_3$  values, consistent with the hypothesis that atmospheric N inputs are directly responsible for rising nitrate concentration in the lake (Ostrom et al., 1998). Recent studies, however, have questioned whether incomplete biological assimilation of atmospherically-deposited N can account for the observed rise in nitrate concentrations (Sturner et al., 2007; Finlay et al., 2007). Stable oxygen isotope ratios of in-lake nitrate ( $\delta^{18}\text{O-NO}_3^-$ ) indicate that the majority of nitrate in the lake has undergone in-situ oxidation (Finlay et al., 2007), suggesting that nitrogen inputs to the lake are assimilated quickly.

Studies of nutrient uptake and primary productivity confirm that nitrogen cycles through the biological pool in Lake Superior much more quickly than it is added to or removed from the system. In-situ measurements of ammonium and nitrate uptake suggest that uptake of both inorganic forms of nitrogen exceeds inputs to the lake on an annual scale (Kumar et al., 2008). Recent estimates of primary production (Urban et al., 2005) also indicate that annual biological uptake of nitrogen exceeds the annual supply to the lake, as do measurements of seasonal nitrate drawdown in the water column (Urban, 2009).

Rates of nitrification and denitrification are less well known. Because Lake Superior is an ultraoligotrophic system, the flux of labile organic matter to the sediments is low, and the majority of sediment metabolism in the lake is oxic (Kumar et al., 2008; Carlton et al., 1989). Nitrification in and subsequent nitrate efflux from lake sediments has been observed (Carlton et al., 1989; Heinen and McManus, 2004), whereas the few published measure-



ments of denitrification indicate that it is a minor process (Carlton et al., 1989). Sediment burial of nitrogen has not been well quantified.

Many questions surrounding the historical increase in nitrate concentrations in Lake Superior remain unanswered. It is unclear whether the lake is already responding to reductions in  $\text{NO}_x$  emissions to the atmosphere or how quickly such a response will occur. Little is known about the effect historical changes in the watershed may have had on terrigenous N inputs to the lake, or if biological cycling of nitrogen in the lake has varied over time. Chapra et al. (2009) employed a mass-balance model to perform an inverse analysis of chloride loading to the Great Lakes; in this paper, we take a similar approach to reconstructing historical N loading and losses in Lake Superior. Dependence on scarce data produces a high level of uncertainty when performing this type of inverse calculation, particularly when dealing with nonconservative substances such as nutrients (Chapra et al., 2009). We model total nitrogen rather than individual species to minimize the number of assumptions made about transformation processes internal to the lake. The model is used to refine our understanding of historical nitrogen sources and sinks and current rates of in-lake nitrogen cycling processes. We then explore several input/output conditions that may have produced the observed historical record of nitrate in the lake. Finally, we forecast nitrogen concentrations in Lake Superior over the next century.

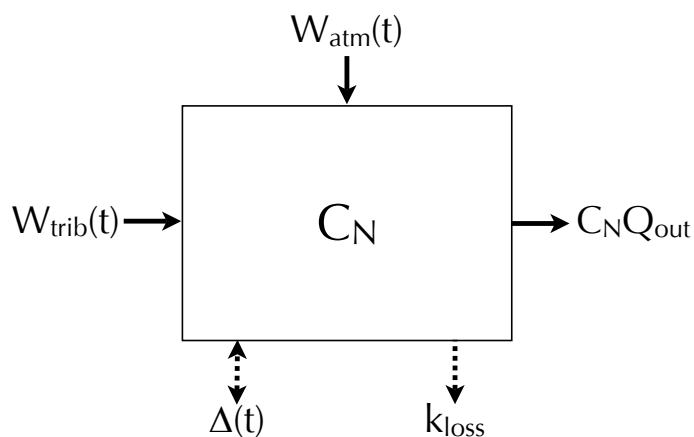
## 5.3 Model Development

We constructed a simple mass-balance model to explore total nitrogen dynamics in Lake Superior (Figure 5.1), governed by the following ordinary differential equation:

$$V \frac{dC_N}{dt} = W_{atm}(t) + W_{trib}(t) - Q_{out}C_N - k_{loss} \pm \Delta(t) \quad (5.1)$$

where  $C_N$  is the concentration of total nitrogen (M) in the lake,  $W_{atm}(t)$  is atmospheric loading in moles  $\text{yr}^{-1}$ ,  $W_{trib}(t)$  is tributary loading in moles  $\text{yr}^{-1}$ ,  $Q_{out}$  is the flow rate through the St. Mary's River ( $\text{L yr}^{-1}$ ),  $\Delta(t)$  is an optional estimated loading/burial term ( $\text{mol yr}^{-1}$ ),  $k_{loss}$  is the net loss (sediment burial and denitrification) of nitrogen ( $\text{mol yr}^{-1}$ ), and  $V$  is the volume of the lake ( $1.2 \times 10^4 \text{ km}^3$ ). The outflow through the St. Mary's River has varied throughout the past century due to fluctuations in lake level (Lenters, 2004); we used mean annual flows from 1900-1999 (<http://www.lre.usace.army.mil/>) and estimated the outflow post-1999 to be equal to the 1999 value of  $61.0 \text{ km}^3 \text{ yr}^{-1}$ . The average annual flowrate between 1900 and 1999 is  $67.4 \text{ km}^3 \text{ yr}^{-1}$ .

Little is known about burial of nitrogen in Lake Superior; few direct measurements exist. Bennett (1986) used a N:P ratio of 2.7 to extrapolate from the P burial rate in the International Joint Commission (IJC) budget (Upper Lakes Reference Group, 1977) to arrive at a N burial rate of  $7.1 \times 10^9 \text{ mol yr}^{-1}$ . However, a burial rate of  $5.2 \times 10^9 \text{ mol yr}^{-1}$  may be calculated by difference from the fluxes provided in the IJC nitrogen budget for Lake Superior.



**Figure 5.1:** Conceptual diagram of model. Atmospheric and tributary loading include nitrate, ammonia/ammonium, and organic nitrogen inputs to the lake as a function of time.

Heinen and McManus (2004) measured organic carbon burial in the relatively productive western arm of Lake Superior. Combining their reported organic C burial rate and molar C:N in settling material and extrapolating to the entire lake basin results in an estimated N burial rate of  $1.33 \times 10^9 \text{ mol yr}^{-1}$ . Using a whole-lake organic C burial estimate of  $0.45 \text{ Tg yr}^{-1}$  (Urban et al., 2005) and a mass C:N of 10 (Klump et al., 1989) results in an estimated N burial rate of  $3.21 \times 10^9 \text{ mol yr}^{-1}$ .

For model evaluation and calibration purposes, historical nitrate concentrations were compiled from several sources, (Great Lakes Environmental Database, [http://www.epa.gov/greatlakes/monitoring/data\\_proj/glenda](http://www.epa.gov/greatlakes/monitoring/data_proj/glenda); Environment Canada; Ke-

weenaw Interdisciplinary Transport Experiment in Superior (KITES); Weiler, 1978). For a more comprehensive compilation of historical  $\text{NO}_3^-$  data see Sterner et al. (2007). Nitrate accounts for about 75% of the current total nitrogen inventory in Lake Superior, with a current average concentration of  $\sim 25 \mu\text{M}$ . The majority of the remainder is dissolved organic nitrogen (DON), measured in western Lake Superior to be  $6.8 \pm 0.2 \mu\text{M}$  by Feuerstein et al. (1997) and  $5.0 \pm 1.5 \mu\text{M}$  for the period 1998–2001 (Lu, 2004). Current ammonium concentrations are two orders of magnitude lower than nitrate concentrations, with a lake-wide average of  $0.21 \mu\text{M}$  (Kumar et al., 2007). Particulate organic nitrogen (PON) concentrations are also relatively small, less than  $0.3 \mu\text{M}$  (Ostrom et al., 1998). To account for DON, PON, and  $\text{NH}_4^+$  contributions to total N concentrations in the lake, historical  $\text{NO}_3^-$  data were adjusted upward by  $7 \mu\text{M}$ . While we do not have direct evidence that the concentration of minor forms of N have remained constant throughout the past century, their contribution to total N, and therefore to trends in total N has been small for the majority of the past century.

### 5.3.1 Atmospheric Inputs

Atmospheric deposition of nitrogen includes multiple chemical forms ( $\text{NO}_3^-$ ,  $\text{NH}_4^+$ , organic N,  $\text{HNO}_3$ ,  $\text{NO}_x$ ,  $\text{NH}_3$ ) and modes of deposition (wet, dry particulate, dry gaseous). Seldom are all forms and modes of deposition measured in any monitoring program or reported in a single study. Spatial trends in deposition also contribute uncertainty to estimates for a sys-

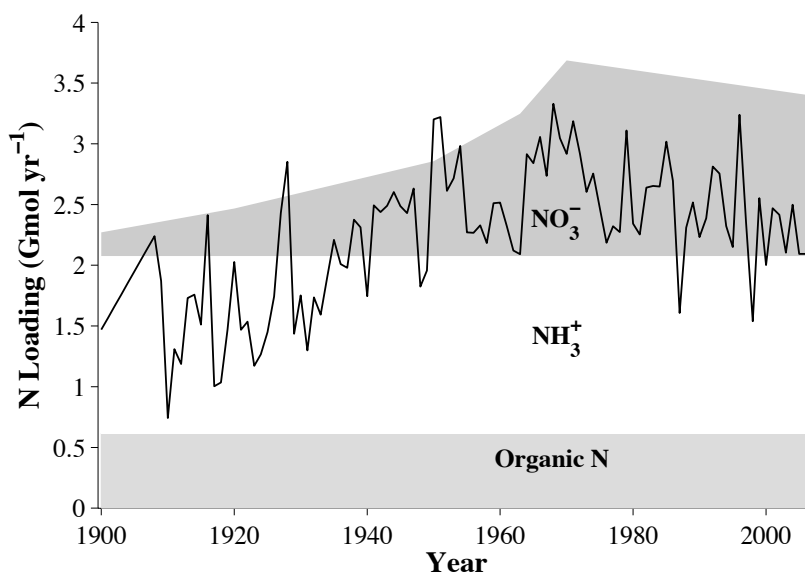
tem as large as Lake Superior. Below we assess current and historical rates of atmospheric deposition of oxidized N (including  $\text{NO}_3^-$ ,  $\text{HNO}_3$ ,  $\text{NO}_x$ ), inorganic reduced N ( $\text{NH}_3$ ,  $\text{NH}_4^+$ ), and organic N.

The National Atmospheric Deposition Program measures wet-only deposition at five sites in close proximity to Lake Superior (Hovland and Wolfridge, MN; Chassell, Seney Wildlife Refuge, and Racoon, MI). For nitrate, there is a spatial gradient in deposition (averaged for 2005–2007) at these sites increasing from 11.7 at the northern-most site (Hovland, MN) to 13.7  $\text{mmol NO}_3^- \text{ m}^{-2} \text{ yr}^{-1}$  at the southeastern-most site (Racoon, MI). When deposition rates for the five stations are weighted by their respective lake areas (Thiessen polygons), the average, lake-wide wet deposition flux of  $\text{NO}_3^-$  for 2005–2007 is calculated to be 13.1  $\text{mmol m}^{-2} \text{ yr}^{-1}$  or  $1.08 \times 10^9 \text{ mol yr}^{-1}$ .

Dry deposition of oxidized N was estimated by Shannon and Voldner (1992) to equal one-third of wet deposition. This estimate was based on atmospheric stability-dependent deposition velocities for  $\text{NO}_x$  and  $\text{HNO}_3$ . Although the estimate of Shannon and Voldner neglects aerosol deposition, it is very similar to the estimate of Baker (1991) that is based on both gas phase and aerosol deposition to nearby forests. A wet to dry ratio of 3:1 yields a total oxidized N deposition rate to Lake Superior (2005–2007) of  $1.44 \times 10^9 \text{ mol yr}^{-1}$ , a value nearly equal to the estimate of present-day  $\text{NO}_3^-$  inputs ( $1.42 \times 10^9 \text{ mol yr}^{-1}$ ) made by Sterner et al. (2007).

Recent estimates of  $\text{NO}_x$  emissions (National Emissions Inventory (NEI), <http://www.epa.gov/ttnchie1/trends>) show a reduction of  $\sim 7\%$  between 1970 and 1995, from 27 million to 25 million tons per year, followed by a much more dramatic reduction of  $32\%$  between 1995 and 2007, from 25 million to 17 million tons per year. Pooled and normalized NADP data from 6 sites around Lake Superior show a statistically significant but noisy ( $p < 0.05$ ,  $r = 0.38$ ) downward trend in wet  $\text{NO}_3^-$  deposition of  $\sim 25\%$  between 1979 and 2007. Individual sites generally show stronger downward trends in wet deposition. The NADP site in Racine, WI has recorded a reduction of  $\sim 41\%$  between 1984 and 2007 ( $p < 0.01$ ,  $r = 0.59$ ). The acceleration in national emissions reductions between 1995 and 2007, however, is not observable in regional deposition data. Our model assumes atmospheric deposition of  $\text{NO}_3^-$  to Lake Superior has declined steadily since 1970 at the approximate rate shown by the NADP data,  $1.3 \times 10^7 \text{ mol yr}^{-1}$ , with a mean deposition rate of  $1.4 \times 10^9 \text{ mol yr}^{-1}$  between 2005 and 2007 (Figure 5.2). This results in an estimate of  $1.9 \times 10^9 \text{ mol yr}^{-1}$  in 1970.

The NADP record of wet deposition is approximately 40 years long; nitrate deposition prior to this is poorly documented. National emissions of  $\text{NO}_x$  increased exponentially between 1900 and 1970 (U.S.EPA, 2000), from  $\sim 2.5$  million to  $\sim 25$  million tons per year. Our model assumes atmospheric deposition of oxidized N to Lake Superior increased proportionally over that same time period, from  $0.17 \times 10^9 \text{ mol yr}^{-1}$  to  $1.90 \times 10^9 \text{ mol yr}^{-1}$  (Figure 5.2).



**Figure 5.2:** Model Inputs. The entire shaded region represents modeled atmospheric N loading ( $W_{atm}$ ); the subregions represent the contribution of organic N, ammonium, and nitrate to the total. The solid line represents modeled tributary loading ( $W_{trib}$ ).

In contrast with oxidized N, wet deposition of reduced inorganic N shows no spatial or temporal trends over the 40-year NADP record. The average for all five stations for all available years yields a wet deposition rate of  $14.9 \text{ mmol m}^{-2} \text{ yr}^{-1}$ . Baker (1991) estimated the wet:dry ratio of  $\text{NH}_3:\text{NH}_4^+$  in the upper midwest to be 5.0. This ratio yields a total atmospheric deposition of reduced inorganic N to Lake Superior of  $1.46 \times 10^9 \text{ mol yr}^{-1}$ . Our estimate is 27% higher than the estimate of Kumar et al. (2007) based on data from the NADP and the EPA's Clean Air Status and Trends Network (CASTNET, <http://www.epa.gov/castnet>). Because there is little agriculture in the vicinity of Lake Superior, most atmospheric  $\text{NH}_3$  is likely derived from biogenic emissions from local forests. Our model assumes atmospheric deposition of  $\text{NH}_4^+$  has remained

constant throughout the past century. It is possible that historical emissions in the region were slightly higher due to greater anthropogenic influences (low-heat combustion of coal, biomass burning, etc.). Conversely, deforestation in the basin may have had an opposing effect. There is evidence that  $\text{NH}_4^+$  deposition declined from 1900–1970 in the northeast United States (Bowen and Valiela, 2001; Likens and Bormann, 1974), but no such evidence exists for the upper Great Lakes. If there were variations in atmospheric  $\text{NH}_4^+$  deposition to the lake in the early 20th century, they were likely less pronounced than trends in  $\text{NO}_3^-$  deposition.

Organic nitrogen often comprises around one-third of total atmospheric N loading (Neff et al., 2002), but the magnitude of this input to the Lake Superior basin is poorly quantified. A site in northern Minnesota received  $3.6 \text{ mmol organic N m}^{-2} \text{ yr}^{-1}$  in wet-only precipitation and  $16.4 \text{ mmol organic N m}^{-2} \text{ yr}^{-1}$  in bulk precipitation, suggesting a dry deposition flux of  $12.8 \text{ mmol m}^{-2} \text{ yr}^{-1}$  (Urban and Eisenreich, 1988; Verry and Timmons, 1975). The Acidic Precipitation in Ontario study reported wet-only deposition of organic N of  $8.6\text{--}10.7 \text{ mmol m}^{-2} \text{ yr}^{-1}$  at Dorion, Ontario in the early 1980s (Chan et al., 1983, 1984). Data from the Experimental Lakes Area (ELA) in northwestern Ontario show an average wet deposition of  $7.1 \text{ mmol N m}^{-2} \text{ yr}^{-1}$  (or an extrapolated deposition of  $6.16 \times 10^8 \text{ mol yr}^{-1}$  to Lake Superior) from 1970–2005 (Dr. Robert E. Hecky, personal communication). While the ELA data do show elevated deposition of organic nitrogen between 1980 and 1990, there is no clear trend over the entire period of record. All of these studies show remarkable agreement. Our model incorporates a constant rate of deposition of organic ni-



trogen to Lake Superior of  $0.62 \times 10^9 \text{ mol yr}^{-1}$  (average based on ELA observations). This rate may have varied prior to 1970. Atmospheric organic N can form from interactions of  $\text{NO}_x$  and non-methane hydrocarbons (NMHC) (Neff et al., 2002); the deposition of organic N could therefore have increased with increased atmospheric  $\text{NO}_x$  from 1900–1970. Biomass burning releases reduced organic N compounds to the atmosphere (Lobert et al., 1991), and vegetation produces NMHCs; deforestation and changes in anthropogenic activity may have affected atmospheric deposition of organic N over time as well.

The International Joint Commission (IJC) reported estimates of total annual atmospheric nitrogen ( $\text{NO}_3^- + \text{NH}_4^+$ ) deposition, as sampled during the period October 1973–June 1975, to be  $4 \times 10^9$  moles (Upper Lakes Reference Group, 1977); our estimate of total atmospheric N loading in 1974 is  $3.9 \times 10^9$  moles. The IJC value was estimated using atmospheric sampling data with an atmospheric transport model. The consistency of the two estimates lends support to our reconstruction of the historical record of atmospheric N deposition.

### **5.3.2 Tributary Inputs**

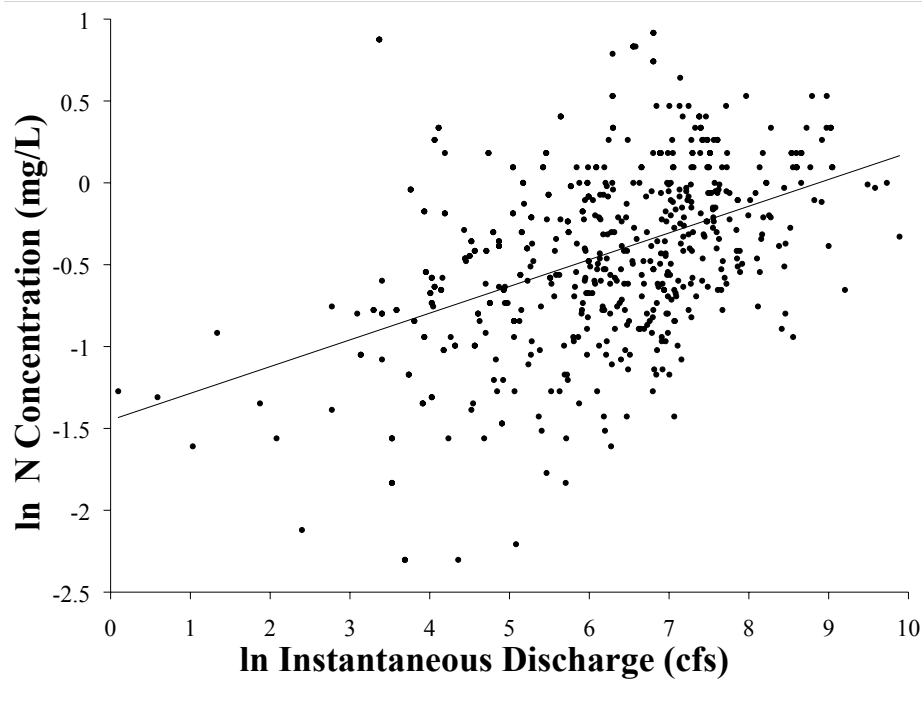
The United States Geological Survey maintains flow and water quality measurement records for major tributaries located in the U.S. portion of the watershed (<http://waterwatch.usgs.gov>). Applying logarithmic least-squares regression to a merged data set consisting of total ni-

trogen and corresponding flow rates from 7 tributaries of varying size (Tahquamenon, Ontonagon, Silver, Nemadji, Bad, and St. Louis Rivers and Gomanche Creek; Figure 5.3) produced the following weak but significant relationship ( $r=0.45$ ,  $p<0.01$ ,  $\sigma_{est}=0.38$ ):

$$\ln(C) = 0.16\ln(Q) + 1.45 \quad (5.2)$$

where  $C$  is total nitrogen concentration ( $\text{mg L}^{-1}$ ) and  $Q$  is instantaneous discharge (cfs). Similar regressions for data from individual rivers generally resulted in stronger correlations ( $\bar{r}=0.59$ ). Because of the logarithmic nature of the relationship between flow and nitrogen concentration, N concentrations are highly dependent on flow rates at relatively low flows, but become more nearly constant at high flows. The loading associated with major runoff events, therefore, is linearly correlated with runoff.

The IJC published the results of a lake-wide sampling campaign to quantify tributary inputs to Lake Superior during the period July 1973–June 1975 (Upper Lakes Reference Group, 1977). They estimated the annual riverine input of total nitrogen to be  $2.61 \times 10^9$  mol, which consisted of (as N) 71% organic nitrogen, 10% ammonia, and 19% nitrate plus nitrite. Monthly runoff estimates for the Lake Superior basin (Croley and Hunter, 1993) indicate a systematic increase in total average runoff to the lake from  $\sim 1000 \text{ m}^3\text{s}^{-1}$  at the turn of the 20th century to  $\sim 1800 \text{ m}^3\text{s}^{-1}$  in the mid-1970s, followed by a decrease to  $\sim 1500 \text{ m}^3\text{s}^{-1}$  currently. Assuming tributary inputs of total nitrogen are positively correlated with flow



**Figure 5.3:** Log-log regression of instantaneous river flow rates and total N concentration (see text for equation). The data are pooled from the St. Louis, Ontonagon, Bad, Taquamenon, Nemadji, and Silver Rivers and Gomanche Creek. Data from individual rivers are more strongly correlated.

rates, the annual riverine total N input can be estimated as:

$$W_t = W_{1974} \frac{Q_t}{Q_{1974}}, \quad (5.3)$$

where  $W_t$  and  $W_{1974}$  are the loadings at time  $t$  and in 1974 (IJC value) and  $Q_t$  and  $Q_{1974}$  are the annual runoff estimates to the lake at time  $t$  and in 1974. The resulting reconstruction of historical tributary total N loadings is shown in Figure 5.2.

### 5.3.3 Model Scenarios

The model was run using the atmospheric and riverine inputs described above as well as the following three conditions:

1. Constant loss rate,  $k_{loss}$ , with no additional loading ( $\Delta(t)=0$ ),
2. Constant loss rate,  $k_{loss}$ , with an additional calibrated loading function, positive  $\Delta(t)$ .
3. Zero loss rate,  $k_{loss}$ , with an additional calibrated loading function, negative  $\Delta(t)$ .

The model was run from 1900 to 2010 using the Dormand-Prince method (Dormand and Prince, 1980) with a fixed time step of 1 year. For scenarios 2 and 3,  $\Delta(t)$  was defined as a linear interpolation of  $\{(1900, \delta_1); (1910, \delta_2); \dots; (2010, \delta_{12})\}$ , where the values of  $\delta_i$  were determined by a least-squares fit to the available data using a bounded pattern search algorithm with  $\min_{\delta_i} = 0$ .  $\Delta(t)$  was defined as positive (additional loading) in scenario 2 and negative (loss) in scenario 3.

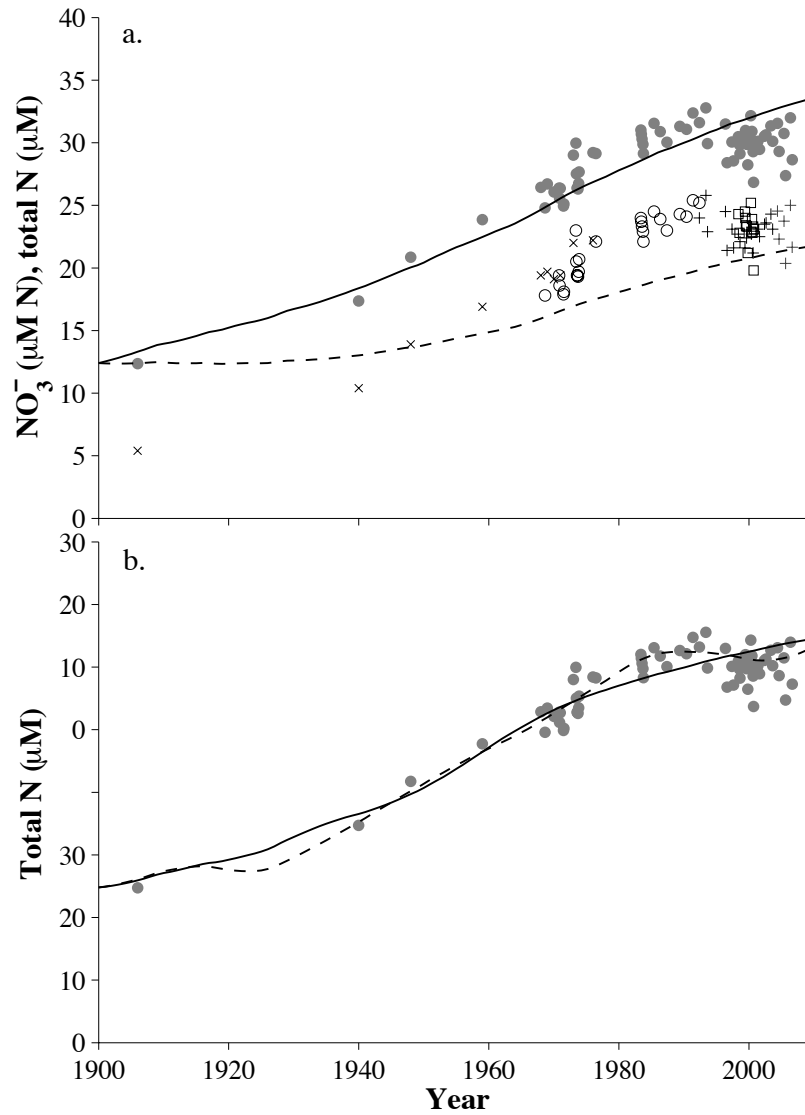
## 5.4 Results

Assuming a constant net loss rate,  $k_{loss}$ , of  $3.21 \times 10^9 \text{ mol yr}^{-1}$  (based on Urban et al. 2005) results in hindcasted total N concentrations substantially lower than the adjusted observations (Figure 5.4a). Using a rate of  $1.33 \times 10^9 \text{ mol yr}^{-1}$  (based on Heinen and McManus

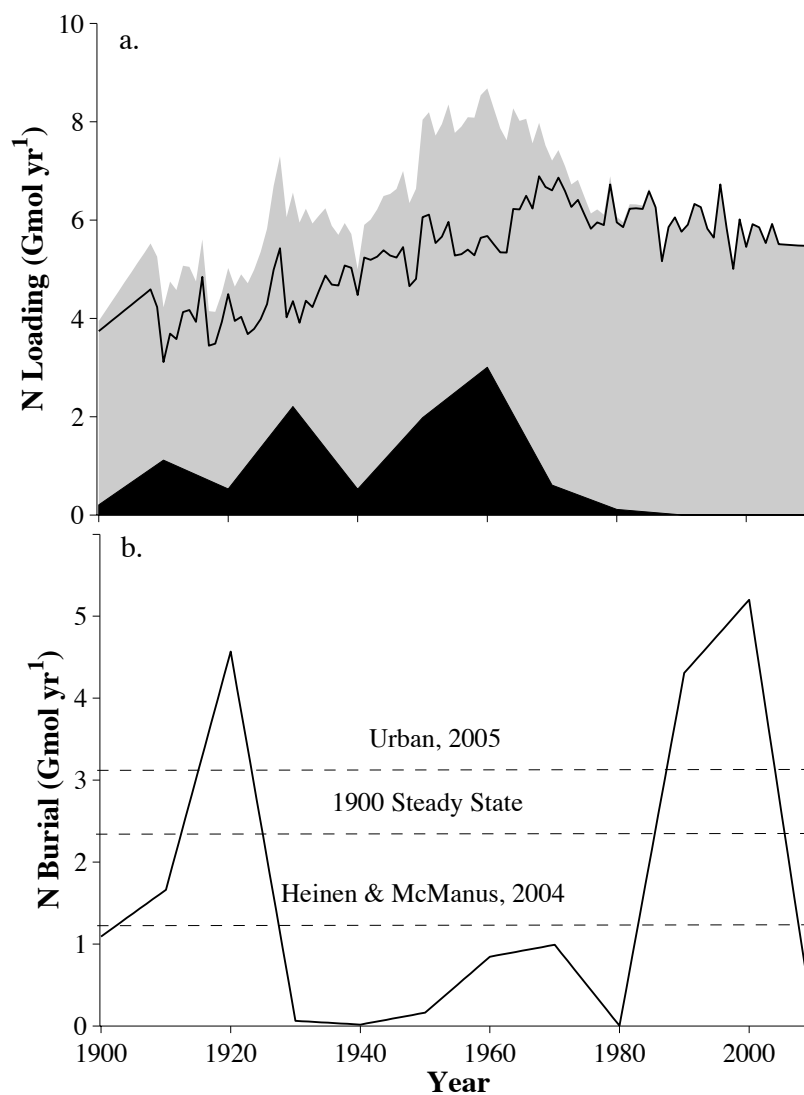
2004), however, results in a rough fit to adjusted observations, though the residuals are clearly not evenly distributed (Figure 5.4a). The modeled total N concentrations do not rise as high or as quickly as observations between 1940 and 2000. Given our estimates of historical loading (Figure 5.2), a constant loss rate of  $1.33 \times 10^9 \text{ mol yr}^{-1}$  would imply the system was already unbalanced in 1900, and N concentrations were increasing at a rate of approximately  $0.1 \mu\text{M yr}^{-1}$ . This implication is reasonable, given the uncertainty in loading estimates and concentrations in the lake circa 1900. To specify true steady-state conditions in 1900, our model requires a net loss rate of  $2.52 \times 10^9 \text{ mol yr}^{-1}$ .

Applying this 1900 steady-state predicted net loss rate of  $2.52 \times 10^9 \text{ mol yr}^{-1}$  and using an additional loading term,  $\Delta(t)$ , to calibrate the model to observations resulted in a better model fit (Figure 5.4b). Figure 5.5a illustrates the additional loading required to achieve a least-squares fit to the adjusted observations.  $70 \times 10^9$  moles N must be added to our original loading estimates over the entire modeled period, with a maximum instantaneous increase of approximately  $5 \times 10^9 \text{ mol yr}^{-1}$  in 1960. The estimated additional loading generally increases between 1920 and 1960, before dropping off to insignificant amounts in 1970.

Applying the same loadings used in scenario 1 and using a time-variant net loss function ( $\Delta(t)$ ) resulted in the best fit to adjusted observations (Figure 5.4b). Figure 5.5b illustrates the historic net burial function that would have been required for our estimated inputs to have produced the observed N concentrations in the lake. Many of the characteristics of this



**Figure 5.4:** Results of model scenarios 1 (a) and 2 and 3 (b).  $\text{NO}_3^-$  data shown in (a) are from Weiler 1978 (x), Environment Canada (○), US EPA (+), and KITES (□); see text for complete citations. The solid points indicate the estimated total N concentration. The solid line in (a) is model scenario 1 with  $k_{\text{loss}} = 3.21 \times 10^9 \text{ mol yr}^{-1}$ , the dashed line is with  $k_{\text{loss}} = 1.33 \times 10^9 \text{ mol yr}^{-1}$ . Model output for scenario 2 is shown in (b) by the solid line and model output for scenario 3 is shown in (b) by the dashed line.



**Figure 5.5:** Modeled N loading for scenario 2 (a) and modeled N net loss for scenario 3 (b). The gray shaded area in (a) represents total N loading to the lake; the black area depicts the contribution of  $\Delta(t)$ , the calibrated additional loading term. The solid line in (a) indicates (loading -  $\Delta(t)$ ), the original estimate of total loading. The modeled loss (b) indicates that a significant decline in burial (or denitrification) rates must have occurred between 1900 and mid-century, followed by a return to higher levels.

function are likely the result of overfitting and observations that are unevenly distributed in time (e.g., the extremely low net loss rates predicted for the present or reconstructed for 1980). However, the general trend that emerges is a pattern of decreasing loss rates between 1920 and 1940, followed by a return to higher rates between  $\sim 1960$  and the present.

## 5.5 Discussion

Given a sufficiently low constant net N loss rate of around  $1.33 \times 10^9 \text{ mol yr}^{-1}$  and our estimated historical loading, the model reproduces the observed historical trend and present-day N concentrations in the lake (Figure 5.4a). While this loss rate is lower than the estimated 1900 steady-state loss rate based on our model inputs and constraints, a great deal of uncertainty surrounds model output for 1900. Calculation of a steady-state loss term requires the steady-state lake concentration and loading estimates. Atmospheric loading in 1900 is back-calculated from an assumed trend in nitrate deposition (see Model Development), and tributary inputs are extrapolated from the earliest estimation (1908) of basin-wide runoff. The concentration of total N in the lake in 1900 is estimated based on a single data point for nitrate ( $5.4 \mu\text{M}$ ) from 1906 (Weiler, 1978) and assuming the organic N concentration in the lake was equal to the current level. However, 11 measurements of nitrate taken at Sault Ste. Marie, MI between 1906–7 have a median value of  $7.3 \mu\text{M}$  (mean=8.5,  $\sigma$ =5.6) (Dole, 1909), suggesting total concentrations of N in Lake Superior at the turn of the century may have been slightly higher than our model assumes. Further-



more, it is entirely possible that the lake was not in steady-state in 1900, as the lake already may have begun responding to increased N inputs.

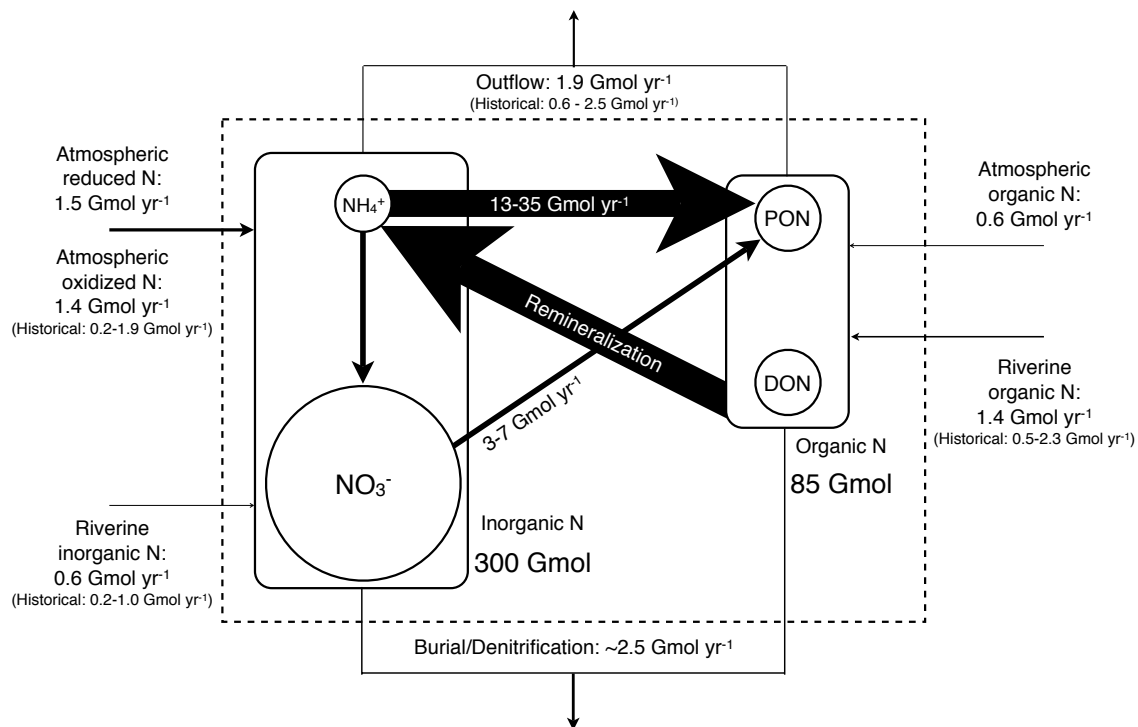
Though the constant-loss scenario reproduces the general trend and magnitude of Lake Superior nitrogen concentrations, there is a clear systematic discrepancy between the model and observations between 1940 and 2000 (Figure 5.4a), indicating actual inputs or outputs (or internal cycling) during this period differed from the assumed inputs and outputs of scenario 1. In order to reproduce the rate of increase in N concentrations observed in the lake during the latter half of the century, either the rate of N loading during this time must have been greater than assumed or the rate of N loss during this time must have been less than assumed.

In scenario 2, modeled “additional” N loading comprises ~50% of the total between 1930 and 1960, but contributes a negligible amount for the remainder of the simulation period. The assumption of a linear relationship between national  $\text{NO}_x$  emissions and  $\text{NO}_3^-$  deposition to Lake Superior prior to 1970 may be erroneous. The fact that measured deposition rates do not reflect the sharp decline seen in national loading estimates suggests this is the case. Further, while atmospheric deposition of reduced N has remained constant for the past 40 years, it is possible that it was altered by land transformation prior to that.

Another explanation for the discrepancy between reconstructed and modeled N inputs may lie in the assumption of a consistent relationship between tributary flow and N concentration throughout the modeled period. The average total N concentration in runoff may have

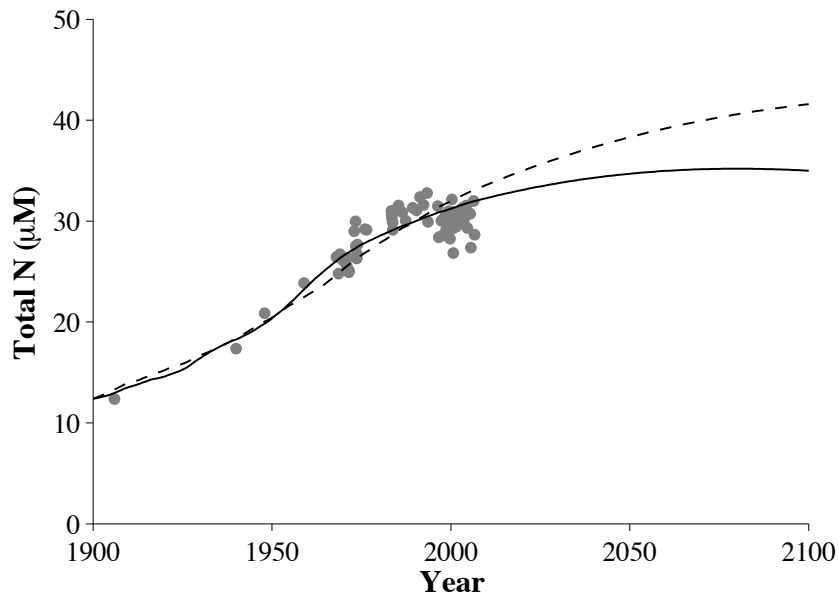
been higher during years with lower average flows (pre-1960), such that the total riverine loading has remained more constant over time than assumed. Land use changes may also have produced changes over time in riverine loading. The majority of the Lake Superior watershed lies in northern Ontario. This area is dominated by white and black spruce, jack pine, and balsam fir (OMNR Forests Division, 2006), typically harvested by clearcutting. Increased nitrogen export has been documented immediately following clearcutting in northern hardwoods ecosystems, though regrowth tends to induce increased nutrient retention (Holmes and Zak, 1999; Mroz et al., 1985); acceleration in the rate of clearcutting may therefore have led to increased N in tributary input. Increases in stream nitrate losses have been shown to occur following harvest in the northeastern U.S. (Likens et al., 1970; Aber et al., 2002). Nitrogen fertilization has also been shown to lead to increased N export from northern hardwoods ecosystems (Pregitzer et al., 2004), although similar experiments on red pine and mixed hardwoods have led to increased N retention (Magill et al., 1997). It is not clear how any of these findings might translate to the nutrient-poor, low atmospheric-N-deposition Boreal ecosystems of northern Ontario.

The observed rise in N concentrations in the lake could, alternatively, have been caused by a mid-century decline in net loss rates, as in scenario 3. Burial of organic N and denitrification are likely correlated with the rate of delivery of organic matter to the sediments. There exists some evidence that phosphorus levels in Lake Superior were elevated mid-century (Weiler, 1978; Urban, 2009), presumably resulting in greater productivity; this would seem to suggest that burial rates should have increased rather than decreased at this point in time.



**Figure 5.6:** Major nitrogen fluxes in Lake Superior. Inputs and output values are those used in the model, and represent present loading, though there is considerable interannual variability in inflows and outflows. The implied current imbalance between inputs and outputs ( $\sim 1.1 \text{ Gmol yr}^{-1}$ ) only represents 0.03% of the total nitrogen pool in the lake. Biotic uptake is as estimated by Kumar et al. (2008).

However, both algae and heterotrophic bacteria in Lake Superior are phosphorus-limited, and algae appear to be on the verge of iron limitation as well (Stern et al., 2004). It is possible, then, that increased phosphorus availability could have induced a relatively greater increase in remineralization processes than in primary productivity, leading to a net decrease in burial of organic N (Figure 5.6). It is also possible that the low sedimentation rates in Lake Superior, as low as  $0.1 \text{ mm yr}^{-1}$  (Kemp et al., 1978), produced a considerable time lag between increased productivity and increased burial of organic matter.



**Figure 5.7:** Forecast of scenario 1 with  $k_{loss}=1.33 \times 10^9 \text{ mol yr}^{-1}$  (dashed line) and scenario 2 with constant future tributary loading and atmospheric deposition continuing to decrease at current rates (solid line).

Comparing the rates of internal N cycling processes with the rates of N input and output for the system shows that a relatively small perturbation in the internal cycle does have the potential to lead to significant changes in the N burial rate (Figure 5.6). Total N uptake in the lake is estimated as being between  $16 \times 10^9 \text{ mol yr}^{-1}$  and  $42 \times 10^9 \text{ mol yr}^{-1}$  (Kumar et al., 2008), an order of magnitude greater than inputs and outputs. The standing pools of inorganic nitrogen ( $300 \times 10^9 \text{ mol}$ ) and organic nitrogen ( $85 \times 10^9 \text{ mol}$ ) are relatively large as well. Remineralization remains a poorly quantified process in Lake Superior, but can be assumed to be of comparable magnitude to uptake; a mass balance around either inorganic or organic nitrogen shows that a significant imbalance between uptake and remineralization would lead to much more rapid changes in pool sizes than have been observed. It is essen-

tially the imbalance between these two components of the N cycle that defines the sediment burial rate; an increase in productivity or a reduction in remineralization would result in a greater N burial rate and vice versa. Because the magnitude of these internal processes is much greater than the magnitude of inputs and outputs, slight changes in these components may have a greater effect on total N in the system than relatively larger changes in external inputs. A  $\sim 1\%$  increase in remineralization relative to primary production could theoretically lead to a  $\sim 10\%$  decrease in burial; a  $\sim 7\%$  relative increase could account for the entire  $\sim 80\%$  mid-century reduction in burial predicted by the model. It therefore seems quite possible that small perturbations in the biological processes affecting N cycling in the lake contributed to the rapid build-up of nitrogen in the lake.

Figure 5.7 illustrates model forecasts for scenarios 1 and 3. Both forecasts assume tributary inputs and outflow rates remain constant at current levels and atmospheric N deposition continues to decline at current rates. Applying a constant burial rate of  $1.33 \times 10^9 \text{ mol yr}^{-1}$  (scenario 1) yields a prediction of a continuing rise in total N concentrations until circa 2100, at which time the average concentration will be  $\sim 38 \mu\text{M}$ . This corresponds to an average  $\text{NO}_3^-$  concentration of  $\sim 31 \mu\text{M}$ . Applying a higher constant net loss rate of  $2.52 \times 10^9 \text{ mol yr}^{-1}$  and our adjusted historical loading function (scenario 3), however, suggests that the lake is currently near peak total N concentration, and will commence a gradual decline in total N content circa 2080.

## 5.6 Conclusions

Our results suggests several factors may have played a role in producing the observed rapid rise in nitrate (and total nitrogen) in Lake Superior. Increased atmospheric deposition in response to  $\text{NO}_x$  emissions and a historical increase in river inflows were the major factors causing increased  $\text{NO}_3^-$  concentrations in the lake. Mid-century loading may have been higher than assumed in our model, due either to (1) a nonlinear relationship between national atmospheric  $\text{NO}_x$  emissions and atmospheric inputs of nitrate to the lake, or (2) land-use changes in the watershed resulting in additional nutrient inputs to the lake. However, because the magnitude of internal N fluxes in the lake (biological uptake and remineralization) are much greater than the magnitude of external fluxes (atmospheric and riverine inputs, burial and denitrification, and outflow), relatively small perturbations in the ecology of the lake may also have had implications for nitrogen cycling- specifically reducing the net loss of total nitrogen mid-century. Current evidence is insufficient to further constrain the historical record, but paleolimnological investigations have the potential to better define historical N burial rates and to improve significantly the accuracy of the model.

Regardless of the conditions that produced a sharp rise in nitrogen concentrations in the lake during the 20th century, our model suggests that the lake is at or is approaching peak N concentration, in contrast with previously published model results (Bennett, 1986). Unless major perturbations in the watershed, airshed, or in-lake biogeochemical cycles occur, it

appears very unlikely that nitrate concentrations in Lake Superior will continue to rise more than 10–15  $\mu\text{M}$ ; our best model prediction places the figure closer to 3  $\mu\text{M}$ .

# Chapter 6

## Conclusions

The performance (i.e., reliability) of mechanistic aquatic biogeochemical models is optimized when the sources of model error and uncertainty are minimized. The most significant sources are generally those due to model specification and parameter estimation. Error and uncertainty can be minimized when the entire model development process is constrained by the available information (i.e., *a priori* knowledge and data).

An important step in optimizing model performance is the identification of the appropriate level of model complexity. Biogeochemical models can span a broad range of detail, from simple one-box models (e.g., the total nitrogen mass balance model presented in Chapter 5) to extensive systems of differential equations (e.g., the NPZD model presented in Chapter 4. More complex models can, in theory, more accurately simulate reality, but in practice



this is not necessarily the case. Each additional layer of complexity added to an ecological model represents an opportunity for the model to misrepresent reality (as demonstrated by the dysfunctionality of the NPZD model in Chapter 4) or for the uncertainty associated with the additional parameters to greatly reduce the certainty of the model predictions.

The biggest obstacle to properly implementing complexity in biogeochemical models is the lack of sufficient calibration data. Through the application of statistical model selection criteria, it can be demonstrated that an optimal level of model complexity exists that balances a model's ability to fit calibration data with the inherent uncertainty associated with the model structure (Chapter 2). In general, the more calibration data that are available, the higher is the appropriate level of complexity. However, the scientific and managerial questions at hand cannot always be answered by models that are parsimonious enough to be fully constrained by available data. The simple Lake Superior carbon cycle model presented in Chapter 4, while well-constrained, required the addition of much more (unconstrained) detail before it could represent the system adequately (i.e., before it included all of the processes of interest). In such cases it is important to recognize (and to quantify, when possible) the uncertainty associated with model predictions.

The importance of using formal data assimilation techniques to estimate model parameters is well established (e.g., Arhonditsis and Brett, 2004; Oschlies, 2006). Data assimilation is particularly important when considering complex models; even if such models contain greater degrees of freedom than can be justified by the available data, the amount of uncer-

tainty associated with model predictions can still be minimized. While the computational demands of complex models sometimes prohibit the application of formal parameter estimation *in situ*, simplified emulators of the model may be developed that are less computationally expensive, in which parameters may be efficiently optimized (Chapter 3).

Bayesian Markov Chain Monte-Carlo (see Chapter 3) is a particularly useful approach to data assimilation. *a priori* knowledge of parameter values is used to guide the estimation (via the specification of prior distributions). This is especially important when parameters are not uniquely identifiable, since without this information the Markov Chains for highly correlated parameters will not converge. The stochastic nature of the Monte-Carlo approach also allows for the amount of uncertainty due to parameter estimation to be quantified (e.g., Figures 3.8, 3.9). Ideally, the modeling process would always include a multi-model selection step, in which models of varying degrees of complexity are evaluated with model selection criteria and the most appropriate model is chosen. In practice, however, such an approach is often impractical. In addition to the common disconnect between data availability and modeling goals described above, there are practical barriers to applying existing model selection tools to mechanistic aquatic biogeochemical models. The computational cost of optimizing complex models is one such limitation. In a multi-model comparison, it must be ensured that all competing models are equally “tuned”. This can only be accomplished by (generally computationally expensive) data assimilation techniques. In addition, models of varying degrees of complexity generally include different numbers of state variables, so that direct comparison of fit between such models is not possible. In Chapter 2,

while the candidate models considered did include different numbers of state variables, they were only compared on the basis of chlorophyll prediction; for most modeling applications the relative performance of the additional state variables would also be important.

Just as expert judgement alone is generally inadequate to parameterize biogeochemical models, quantitative techniques alone are not always sufficient to evaluate relative model performance. In Chapter 4, it was demonstrated that while a NPD model fit the calibration data better than a NPZD model (as measured by the overall NRMSE of the model that measured fit to nutrients, chlorophyll, and net primary productivity), examination of the model output for individual variables revealed that an important component of the carbon cycle in Lake Superior (i.e., net primary productivity) is better represented in the NPZD model. Since primary productivity represents one of the dominant carbon fluxes in the system whereas the spatiotemporal distribution of phosphorus and chlorophyll have a smaller overall impact on the carbon cycle, the NPZD model could be considered to be more useful than the NPD model, at least for the task at hand. Such an assessment could only be made based on *a priori* knowledge of the system and the intended function of the model.

Another important role of expert judgement is in evaluating the quality of calibration data available. The most abundant field observations tend to be concentration data (e.g., chlorophyll, nutrients), while measurements of *in situ* rates are less common. As shown in Chapter 3, concentration data alone are inadequate to fully constrain a biogeochemical model; fitting the spatiotemporal distributions of state variables to data does not constrain process

rates in any way, yet it is often exactly these rates that are desired from model output. It may be necessary to define a weighting scheme in the goodness-of-fit function (Chapters 3 and 4) either to add or to remove bias related to specific data types in the assessment of model fit. Finally, the model fit to data in one spatial and/or temporal region can reasonably be considered more important than others. The calibration data themselves are also often irregular in space and time, further complicating the interpretation of goodness-of-fit metrics. For example, if a goal is to simulate algal dynamics, the relative fit of modeled chlorophyll concentrations to summertime epilimnetic observations may be deemed more important than the relative fit to wintertime observations. It is difficult to incorporate such a judgement into a purely quantitative assessment of model fit.

Optimal model performance cannot always be achieved using a purely mathematical approach. The work presented here highlights points in the modeling process in which the available information can be efficiently incorporated to minimize model error and uncertainty. Good modeling practice should include an assessment, based upon both data availability and model output requirements, of appropriate model complexity. Models should be as constrained by observations as possible, and include no more detail than necessary to adequately describe the system of interest. While *a priori* knowledge is essential in guiding model formulation and constraining parameter values, the optimal parameterization for each model structure should be determined via data assimilation. When proper consideration is given to model limitations and uncertainty, simple models can be as effective, or more so, than their complex counterparts when applied to biogeochemical problems.

## 6.1 Scientific conclusions

The ultimate goal of improving model performance is to provide reliable insights into the systems being modeled. Each of the models presented in the preceding chapters provided such insights.

In Trout Lake (Chapter 2), it appears as though sediment sequestration and release of nutrients (phosphorus) does not exert a significant influence on the seasonal progression of phytoplankton (as indicated by chlorophyll concentrations), as does the phosphorus cycle in the water column. Peaks in observed TDP concentrations in the hypolimnion do correspond with periods of hypolimnetic oxygen depletion (Appendix B), suggesting sediment P release does occur. Nonetheless, chlorophyll concentrations can be accurately simulated without the explicit inclusion of this process.

In Lake Superior, the failure of traditional NPZD-type models to reproduce algal dynamics on an annual scale (Chapters 3 and 4) suggests that nutrient limitation is not a key driver of the seasonal patterns in the vertical distribution of algal biomass. This is not to say that the phytoplankton of Lake Superior do not exist in a nutrient-limited state, or that nutrient availability does not limit the overall distribution of biomass in the lake, however.

The carbon cycle model developed in Chapter 4 provides the first estimate of lake-wide gross primary productivity (GPP). GPP is estimated at approximately 10 Tg C yr<sup>-1</sup>, mean-

ing autotrophy supports between 12 and 77% of community respiration in the lake (based on the estimates of community respiration made by (Urban et al., 2005; Cotner et al., 2004). The remainder of community respiration can presumably may be partially attributed to the processing of allocthonous carbon inputs by heterotrophs. The disproportionately large contribution of nearshore regions to biological processing of carbon, however, may be responsible for this apparent imbalance (Bennington, 2010). The simple model also examined the influence of several factors (settling, shade adaptation, and grazing) on DCM formation and found shade adaptation to be the dominant driver.

Finally, a simple model of total nitrogen in Lake Superior coupled with a detailed accounting of historical inputs and outputs indicates that increases in atmospheric nitrogen deposition alone were insufficient to have resulted in the precipitous rise in nitrate concentrations observed in the lake during the past century. Various model scenarios suggest that either increased loading from other sources (i.e., riverine inputs) or possibly reduced sediment burial mid-century also contributed to the rise in nitrate concentrations.



# References

Aber, J. D. (1997). Why don't we believe the models? *Bull Ecol Soc Amer*, 78:232–233.

Aber, J. D., Ollinger, S. V., Driscoll, C. T., Likens, G. E., Holmes, R. T., Freuder, R. J., and Goodale, C. L. (2002). Inorganic nitrogen losses from a forested ecosystem in response to physical, chemical, biotic, and climatic perturbations. *Ecosystems*, 5(7):648–658.

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In *2nd International Symposium on Information Theory, Tsahkadsor, Armenian SSR*, pages 267–281.

Akaike, H. (1974). A new look at the statistical model identification. *IEEE T Automat Contr*, 19(6):716–723.

Anderson, T. R. (2005). Plankton functional type modelling: running before we can walk? *J Plankton Res*, 27(11):1073–1081.

Anderson, T. R. (2006). Confronting complexity: reply to Le Quere and Flynn. *J Plankton Res*, 28(9):877–878.



- Anderson, T. R. (2010). Progress in marine ecosystem modelling and the “unreasonable effectiveness of mathematics”. *J Marine Syst*, 81:4–11.
- Arhonditsis, G. B., Adams-Vanharn, B. A., Nielsen, L., Stow, C. A., and Reckhow, K. H. (2006). Evaluation of the current state of mechanistic aquatic biogeochemical modeling: citation analysis and future perspectives. *Environ Sci Technol*, 40:6547–6554.
- Arhonditsis, G. B. and Brett, M. T. (2004). Evaluation of the current state of mechanistic aquatic biogeochemical modeling. *Mar Ecol-Prog Ser*, 271:13–26.
- Arhonditsis, G. B. and Brett, M. T. (2005). Eutrophication model for lake washington (usa):: Part i. model description and sensitivity analysis. *Ecol Model*, 187:140–178.
- Arhonditsis, G. B., Papantou, D., Zhang, W., Perhar, G., Massos, E., and Shi, M. (2008). Bayesian calibration of mechanistic aquatic biogeochemical models and benefits for environmental management. *J Marine Syst*, 73(1-2):8–30.
- Arhonditsis, G. B., Qian, S. S., Stow, C. A., Lamon, E. C., and Reckhow, K. H. (2007). Eutrophication risk assessment using bayesian calibration of process-based models: Application to a mesotrophic lake. *Ecol Model*, 208(2-4):215–229.
- Auer, M. T. and Bub, L. A. (2004). Selected features of the distribution of chlorophyll along the southern shore of Lake Superior. *J Great Lakes Res*, 30(Supplement 1):269–284.

- Auer, M. T., Bub, L. A., Auer, N. A., and Urban, N. R. (2010). Primary production, carbon flux, and the distribution of *Diporeia* in Lake Superior. *Verh Internat Verein Theor Angew Limnol*, 30(10):1499–1505.
- Auer, M. T. and Canale, R. P. (1982). Ecological studies and mathematical modeling of cladophera in lake huron: 3. the dependence of growth rates on internal phosphorus pool size. *J Great Lakes Res*, 8(1):93–99.
- Auer, M. T. and Johnson, T. C. (2004). Foreward, special issue on Lake Superior. *J Great Lakes Res*, 30(Supplement 1):i–ii.
- Baehr, M. and McManus, J. (2003). The measurement of phosphorus and its spatial and temporal variability in the western arm of Lake Superior. *J Great Lakes Res*, 29(3):479–487.
- Baines, S. and Pace, M. (1991). The production of dissolved organic matter by phytoplankton and its importance to bacteria: Patterns across marine and freshwater systems. *Limnol Oceanogr*, 36(6):1078–1090.
- Baker, L. A. (1991). Regional estimates of atmospheric dry deposition. In Charles, D. F., editor, *Acidic Deposition and Aquatic Ecosystems, Regional Case Studies*, pages 645–652. Springer-Verlag.
- Barbiero, R. T. and Tuchman, M. L. (2004). The deep chlorophyll maximum in Lake Superior. *J Great Lakes Res*, 30:256–268.

- Bender, M., Orchardo, J., Dickson, M., Barber, R., and Lindley, S. (1999). In vitro O<sub>2</sub> fluxes compared with <sup>14</sup>C production and other rate terms during the JGOFS equatorial Pacific experiment. *Deep-Sea Res Pt I*, 46(4):637–654.
- Bennett, E. B. (1986). The nitrifying of Lake Superior. *Ambio*, 15(5):272–275.
- Bennington, V. (2010). *Carbon Cycle Variability of the North Atlantic Ocean and Lake Superior*. PhD thesis, University of Wisconsin - Madison.
- Bennington, V., McKinley, G. A., Kimura, N., and Wu, C. H. (2010). The general circulation of Lake Superior: mean, variability, and trends from 1979-2006. *J Geophys Res*, doi:10.1029/2010JC006261.
- Biddanda, B., Ogdahl, M., and Cotner, J. (2001). Dominance of bacterial metabolism in oligotrophic relative to eutrophic waters. *Limnol Oceanogr*, 46(3):730–739.
- Bierman, V. J. and Dolan, D. M. (1981). Modeling of phytoplankton-nutrient dynamics in saginaw bay, lake huron. *J Great Lakes Res*, 7(4):409–439.
- Bissett, W. P., Walsh, J. J., Dieterle, D., and Carder, K. L. (1999). Carbon cycling in the upper waters of the Sargasso Sea: I. numerical simulation of differential carbon and nitrogen fluxes. *Deep-Sea Res Pt I*, 46:205–269.
- Blumberg, A. F. and Mellor, G. L. (1987). A description of a three-dimensional coastal ocean circulation model. In Heaps, N. S., editor, *Three Dimensional Coastal Ocean Models*, pages 1–16. American Geophysical Union.

- Boulêtreau, S., Izagirre, O., Garabétian, F., Sauvage, S., Elozegi, A., and Sánchez-Pérez, J.-M. (2008). Identification of a minimal adequate model to describe the biomass dynamics of river epilithon. *Riv Res App*, 24(1):36–53.
- Bowen, J. L. and Valiela, I. (2001). Historical changes in atmospheric nitrogen deposition to cape cod, massachusetts, usa. *Atmos Environ*, 35(6):1039–1051.
- Bowie, G. L., Mills, W. B., Porcella, D. B., Campbell, C. L., Pagenkopf, J. R., Rupp, G. L., Johnson, K. M., Chan, P. W., Gherini, S. A., and Chamberlain, C. E. (1985). *Rates, constants, and kinetics formulations in surface water quality modeling (second edition)*. United States Environmental Protection Agency, Athens, GA. EPA/600/3-85/040.
- Bozdogan, H. (2000). Akaike’s information criterion and recent developments in information complexity. *J. Math. Psych.*, 44(1):62–91.
- Buitenhuis, E., Le Quere, C., Aumont, O., Beaugrand, G., Bunker, A., Hirst, A., Ikeda, T., O’Brien, T., Piontkovski, S., and Straile, D. (2006). Biogeochemical fluxes through mesozooplankton. *Global Biogeochem Cy*, 20(2):GB2003.
- Burnham, K. P. and Anderson, D. R. (2001). Kullback-leibler information as a basis for strong inference in ecological studies. *Wildlife Research*, 28:111–119.
- Burnham, K. P. and Anderson, D. R. (2002). *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*. Springer, New York, NY. 488 pp.

- Burns, N. M. and Rosa, F. (1980). In situ measurement of the settling velocity of organic carbon particles and 10 species of phytoplankton. *Limnol Oceanogr*, 25(5):855–864.
- Canale, R., DePalma, L., and Vogel, A. (1976). A plankton-based food web model for Lake Michigan. In Canale, R., editor, *Modeling Biochemical Processes in Aquatic Ecosystems*, page 33. Ann Arbor Science.
- Carlton, R. G., Walker, G. S., Klug, M. J., and Wetzel, R. G. (1989). Relative values of oxygen, nitrate, and sulfate to terminal microbial processes in the sediments of Lake Superior. *J Great Lakes Res*, 15(1):133–140.
- Carpenter, S. R., Cole, J. J., Kitchell, J. F., and Pace, M. L. (1998). Impact of dissolved organic carbon, phosphorus, and grazing on phytoplankton biomass and production in experimental lakes. *Limnol Oceanogr*, 43(1):73–80.
- Cerco, C. F. and Cole, T. (1993). Three-dimensional eutrophication model of chesapeake bay. *J Environ Eng-ASCE*, 119(6):1006–1025.
- Chan, W. H., Chung, D., and Tang, A. J. S. (1984). Precipitation concentration and wet deposition fields of pollutants in Ontario - 1982. Technical Report ARB-142-84-ARSP, Ontario Ministry of the Environment.
- Chan, W. H., Tang, A. J. S., and Lusi, M. A. (1983). Precipitation concentration and wet deposition fields of pollutants in Ontario, September 1980 to December 1981. Technical Report ARB-61-83-ARSP, Ontario Ministry of the Environment.

- Chapra, S. C. (1977). Total phosphorus model for the Great Lakes. *J Environ Eng*, 103:147–161.
- Chapra, S. C. (1997). *Surface Water-Quality Modeling*. McGraw-Hill.
- Chapra, S. C., Dove, A., and Rockwell, D. C. (2009). Great lakes chloride trends: Long-term mass balance and loading analysis. *J Great Lakes Res*, 35(2):272–284.
- Chen, C., Ji, R., Schwab, D., Beletsky, D., Fahnenstiel, G., Jiang, M., Johengen, T., Vanderploeg, H., Eadie, B., Budd, J., Bundy, M., Gardner, W., Cotner, J., and Lavrentyev, P. (2002). A model study of the coupled biological and physical dynamics in Lake Michigan. *Ecol Model*, 152:145–168.
- Chen, C. and Orlob, G. (1975). Ecological simulation for aquatic environments. In Patton, B., editor, *Systems Analysis and Simulation in Ecology*, vol. III, page 475. Academic Press.
- Cotner, J., Biddanda, B., Makino, W., and Stets, E. (2004). Organic carbon biogeochemistry of Lake Superior. *Aquat Ecosyst Health*, 7(4):451–464.
- Cox, G. M., Gibbons, J. M., Wood, A. T. A., Craigon, J., Ramsden, S. J., and Crout, N. M. J. (2006). Towards the systematic simplification of mechanistic models. *Ecol Model*, 198(1-2):240–246.
- Croley, T. E. I. and Hunter, T. S. (1993). Great Lakes monthly hydrologic data. Technical Report ERL-GLERL 902, NOAA GLERL.

- Crout, N. M. J., Tarsitano, D., and Wood, A. T. (2009). Is my model too complex? evaluating model formulation using model reduction. *Environ Modell Softw*, 24:1–7.
- Denman, K. (2003). Modelling planktonic ecosystems: parameterizing complexity. *Prog Oceanogr*, 57:429–452.
- Desai, A. R., Austin, J. A., Bennington, V., and McKinley, G. A. (2009). Stronger winds over a large lake in response to weakening air-to-lake temperature gradient. *Nature Geoscience*, 2:855–858.
- Desai, A. R., Noormets, A., Bolstad, P. V., Chen, J., Cook, B. D., Davis, K. J., Euskirchen, E. S., Gough, C., Martin, J. G., and Ricciuto, D. M. (2008). Influence of vegetation and seasonal forcing on carbon dioxide fluxes across the Upper Midwest, USA: Implications for regional scaling. *Agr Forest Meteorol*, 148(2):288–308.
- Descy, J.-P., Higgins, Harry, W., Mackey, D. J., Hurley, J. P., and Frost, T. S. (2000). Pigment ratios and phytoplankton assessment in northern Wisconsin lakes. *J. Phycol.*, 36:274–286.
- DiToro, D., Thomann, R., and O’Connor, D. (1971). A dynamic model of phytoplankton population in the Sacramento-San Joaquin delta. In Gould, R., editor, *Advances in Chemistry Series 106: Nonequilibrium Systems in Natural Water Chemistry*, page 131. American Chemical Society.
- Dole, R. (1909). *The quality of surface waters in the United States*. United States Geological Survey.

- Doney, S. C., Glover, D. M., and Najjar, R. G. (1996). A new coupled, one-dimensional biological-physical model for the upper ocean: Applications to the JGOFS Bermuda Atlantic time-series study (BATS) site. *Deep-Sea Res Pt II*, 43(2):591–624.
- Dormand, J. and Prince, P. (1980). A family of embedded Runge-Kutta formulae. *J Comput Appl Math*, 6(1):19–26.
- Droop, M. (1983). 25 years of algal growth kinetics: A personal view. *Botanica Marina*, XXVI:99–112.
- Dutkiewicz, S., Follows, M. J., and Parekh, P. (2005). Interactions of the iron and phosphorus cycles: A three-dimensional model study. *Global Biogeochem Cy*, 19(GB1021, doi:10.1029/2004GB002342).
- Feuerstein, T. P., Ostrom, P. H., and Ostrom, N. E. (1997). Isotopic biogeochemistry of dissolved organic nitrogen: A new technique and application. *Org Geochem*, 27(7-8):363–370.
- Finlay, J. C., Sterner, R. W., and Kumar, S. (2007). Isotopic evidence for in-lake production of accumulating nitrate in Lake Superior. *Ecol Appl*, 17(8):2323–2332.
- Flynn, K. J. (2001). A mechanistic model for describing dynamic multi-nutrient, light, temperature interactions in phytoplankton. *J Plankton Res*, 23(9):977–997.
- Flynn, K. J. (2003). Do we need complex mechanistic photoacclimation models for phytoplankton? *Limnol Oceanogr*, 48(6):2243–2249.



- Flynn, K. J. (2005). Castles built on sand: dysfunctionality in plankton models and the inadequacy of dialogue between biologists and modellers. *J Plankton Res*, 27(12):1205–1210.
- Flynn, K. J. (2006). Reply to horizons article ‘plankton functional type modelling: running before we can walk’ Anderson (2005): II. putting trophic functionality into plankton functional types. *J Plankton Res*, 28(9):873–875.
- Flynn, K. J. (2008). The importance of the form of the quota curve and control of non-limiting nutrient transport in phytoplankton models. *J Plankton Res*, 30(4):423.
- Follows, M., Dutkiewicz, S., Grant, S., and Chisholm, S. (2007). Emergent biogeography of microbial communities in a model ocean. *Science*, 315:1843–1846.
- Ford, D. E. and Johnson, L. S. (1986). An assessment of reservoir mixing processes. Technical Report EL-86-7, U.S. Army Corps of Engineers.
- Friedrichs, M. A. M. (2002). Assimilation of JGOFS EqPac and SeaWiFS data into a marine ecosystem model of the central equatorial Pacific Ocean. *Deep-Sea Res. II*, 49:289–319.
- Friedrichs, M. A. M., Dusenberry, J. A., Anderson, L. A., Armstrong, R. A., Chai, F., Christian, J. R., Doney, S. C., Dunne, J., Fujii, M., and Hood, R. R. (2007). Assessment of skill and portability in regional marine biogeochemical models: Role of multiple planktonic groups. *J Geophys Res*, 112:C08001.

- Friedrichs, M. A. M., Hood, R. R., and Wiggert, J. D. (2006). Ecosystem model complexity versus physical forcing: Quantification of their relative impact with assimilated Arabian Sea data. *Deep-Sea Res Pt II*, 53(5-7):576–600.
- Fuentes, M. (2007). Approximate likelihood for large irregularly spaced spatial data. *J Am Stat Assoc*, 102(477):321.
- Geider, R., MacIntyre, H., and Kana, T. (1998). A dynamic regulatory model of phytoplanktonic acclimation to light, nutrients, and temperature. *Limnol Oceanogr*, 43(4):679–694.
- Gentleman, W. (2002). A chronology of plankton dynamics in silico: how computer models have been used to study marine ecosystems. *Hydrobiologia*, 480:69–85.
- Green, P. and Mira, A. (2001). Delayed rejection in reversible jump Metropolis-Hastings. *Biometrika*, 88(4):1035–1053.
- Green, S. A. and Eadie, B. J. (2004). Introduction to special session: Transport and transformation of biogeochemically important materials in coastal waters. *J Geophys Res*, 109:C10S01.
- Haario, H., Laine, M., Mira, A., and Saksman, E. (2006). DRAM: efficient adaptive MCMC. *Stat Comput*, 16:339–354.
- Haario, H., Saksman, E., and Tamminen, J. (2001). An adaptive Metropolis algorithm. *Bernoulli*, 7(2):223–242.

- Harmon, R. and Challenor, P. (1997). A Markov Chain Monte Carlo method for estimation and assimilation into models. *Ecol Model*, 101:41–59.
- Heinen, E. and McManus, J. (2004). Carbon and nutrient cycling at the sediment-water boundary in western Lake Superior. *J Great Lakes Res*, 30:113–132.
- Hipsey, M., Romero, J., Antenucci, J., and Hamilton, D. (2006). *Computational Aquatic Ecosystem Dynamics Model (CAEDYM) v2.3 Science Manual*. Contract Research Group, Center for Water Research, University of Western Australia.
- Holmes, W. E. and Zak, D. R. (1999). Soil microbial control of nitrogen loss following clear-cut harvest in northern hardwood ecosystems. *Ecol Appl*, 9(1):202–215.
- Hunt, R., Feinstein, D., Pint, C., and Anderson, M. (2006). The importance of diverse data types to calibrate a watershed model of the Trout Lake basin, northern Wisconsin, USA. *J Hydrol*, 321(1-4):286–296.
- Hunt, R. J., Anderson, M. P., and Kelson, V. A. (1998). Improving a complex finite-difference ground water flow model through the use of an analytic element screening model. *Ground Water*, 36(6):1011–1017.
- Hurvich, C. and Tsai, C.-L. (1989). Regression and time series model selection in small samples. *Biometrika*, 76(2):297–307.

- Imberger, J. and Patterson, J. C. (1981). A dynamic reservoir simulation model - DYRESM:5. In Fischer, H. B., editor, *Transport Models for Inland and Coastal Waters*, pages 310–361. Academic Press.
- Imerito, A. (2007). *Dynamic Simulation Reservoir Model DYRESM v4: Science Manual*. Center for Water Research, University of Western Australia.
- Jakeman, A. J., Letcher, R. A., and Norton, J. P. (2006). Ten iterative steps in development and evaluation of environmental models. *Environ Modell Softw*, 21:602–614.
- Jassby, A. and Platt, T. (1976). Mathematical formulation of the relationship between photosynthesis and light for phytoplankton. *Limnol Oceanogr*, 21(4):540–547.
- Jorgensen, S. E. and Bendoricchio, G. (2001). *Fundamentals of Ecological Modelling*. Elsevier.
- Kemp, A. L. W., Dell, C. I., and Harper, N. S. (1978). Sedimentation rates and a sediment budget for Lake Superior. *J Great Lakes Res*, 4(3-4):276–287.
- Kennedy, M. C. and O’Hagan, A. (2001). Bayesian calibration of computer models. *J Roy Stat Soc B*, 63(3):425–464.
- Kirkpatrick, S., Gelatt, C. D., and Vecchi, M. P. (1983). Optimization by simulated annealing. *Science*, 220(4598):671–679.

- Klump, J. V., Paddock, R., Remsen, C. C., Fitzgerald, S., Boraas, M., and Anderson, P. (1989). Variations in sediment accumulation rates and the flux of labile organic matter in eastern Lake Superior basins. *J Great Lakes Res*, 15(1):104–122.
- Krause-Jensen, D. and Sand-Jensen, K. (1998). Light attenuation and photosynthesis of aquatic plant communities. *Limnol Oceanogr*, 43(3):396–407.
- Kumar, S., Sterner, R. W., and Finlay, J. C. (2008). Nitrogen and carbon uptake dynamics in Lake Superior. *J Geophys Res*, 113:G04003, doi:10.1029/2008JG000720.
- Kumar, S., Sterner, R. W., Finlay, J. C., and Brovold, S. (2007). Spatial and temporal variation of ammonium in Lake Superior. *J Great Lakes Res*, 33(3):581–591.
- Law, T., Zhang, W., Zhao, J., and Arhonditsis, G. B. (2009). Structural changes in lake functioning induced from nutrient loading and climate variability. *Ecol Model*, 220(7):979–997.
- Laws, E., Landry, M., Barber, R., Campbell, L., Dickson, M., and Marra, J. (2000). Carbon cycling in primary production bottle incubations: inferences from grazing experiments and photosynthetic studies using  $^{14}\text{C}$  and  $^{18}\text{O}$  in the Arabian Sea. *Deep-Sea Res Pt II*, 47(7-8):1339–1352.
- Le Quere, C. (2006). Reply to horizons article ‘plankton functional type modelling: running before we can walk’ Anderson (2005): I. abrupt changes in marine ecosystems? *J Plankton Res*, 28(9):871–872.

- Le Quere, C., Harrison, S. P., Prentice, I. C., Buitenhuis, E. T., Aumont, O., Bopp, L., Claustre, H., Cotrim Da Cunha, L., Geider, R., Giraud, X., Klaas, C., Kohfeld, K. E., Legendre, L., Manizza, M., Platt, T., Rivkin, R. B., Sathyendranath, S., Uitz, J., Watson, A. J., and Wolf-Gladrow, D. (2005). Ecosystem dynamics based on plankton functional types for global ocean biogeochemistry models. *Global Change Biol*, 11:2016–2040.
- Lenters, J. (2004). Trends in the Lake Superior water budget since 1948: A weakening seasonal cycle. *J Great Lakes Res*, 30(Supplement 1):20–40.
- Likens, G., Bormann, F., Johnson, N., Fisher, D., and Pierce, R. (1970). Effects of forest cutting and herbicide treatment on nutrient budgets in the Hubbard Brook watershed-ecosystem. *Ecological Monographs*, 40:23–47.
- Likens, G. E. and Bormann, F. H. (1974). Acid rain: a serious regional environmental problem. *Science*, 14(4142):1176–1179.
- Ljung, L. (1987). *System Identification: Theory for the User*. Prentice-Hall. 519 pp.
- Lobert, J., Scharffe, D., Hao, W., Kuhlbusch, T., Warneck, P., and Crutzen, P. (1991). Experimental evaluation of biomass burning emissions: Nitrogen and carbon containing compounds. In Levine, J., editor, *Global Biomass Burning: Atmospheric, Climatic and Biospheric Implications*. MIT Press.
- Lu, X. (2004). The carbon cycle and food web structure of Lake Superior. Master's thesis, Mich. Technol. Univ., Houghton, MI.

- Magill, A., Aber, J. D., Hendricks, J., Bowden, R., Melillo, J., and Steudler, P. (1997). Biogeochemical response of forest ecosystems to simulated chronic nitrogen deposition. *Ecol Appl*, 7(2):402–415.
- Magnuson, J. J. and Bowser, C. (1990). Network for long-term ecological research in the united states. *Freshwater Biol.*, 23(1):137–143.
- McDonald, C. and Urban, N. (2010). Using a model selection criterion to identify appropriate complexity in aquatic biogeochemical models. *Ecol Model*, 221:428–432.
- McKinley, G., Follows, M., and Marshall, J. (2003). Interannual variability of air-sea  $O_2$  fluxes and the determination of  $CO_2$  sinks using atmospheric  $O_2 / N_2$ . *Geophys Res Lett*, 30(3):1101.
- McKinley, G., Follows, M., and Marshall, J. (2004). Mechanisms of air-sea  $CO_2$  flux variability in the equatorial Pacific and North Atlantic. *Global Biogeochem Cy*, 18:2011.
- McKinley, G., Takahashi, T., Buitenhuis, E., Chai, F., Christian, J., Doney, S., Le Quere, C., Lima, I., Murtugudde, R., Shi, L., and Wetzol, P. (2006). North Pacific carbon cycle response to climate variability on seasonal to decadal timescales.
- Mesinger, F., DiMego, G., Kalnay, E., Mitchell, K., Shafran, P. C., Ebisuzaki, W., Jović, D., Woollen, J., Rogers, E., and Berbery, E. H. (2006). North american regional reanalysis. *Bulletin of the American Meteorological Society*, 87(3):343–360.

- Mitra, A., Flynn, K. J., and Fasham, M. J. R. (2007). Accounting for grazing dynamics in nitrogen-phytoplankton-zooplankton models. *Limnol Oceanogr*, 52(2):649–661.
- Mroz, G., Jurgensen, M., and Frederick, D. (1985). Soil nutrient changes following whole tree harvesting on three northern hardwood sites. *Soil Sci Soc Am J*, 49:1552–1557.
- Munawar, M. and Munawar, I. (1978). Phytoplankton of Lake Superior 1973. *J Great Lakes Res*, 4(3-4):415–442.
- Myung, I. J. (2000). The importance of complexity in model selection. *J. Math. Psychol.*, 44:190–204.
- Neff, J., Holland, E., Dentener, F., McDowell, W., and Russell, K. (2002). The origin, composition and rates of organic nitrogen deposition: A missing piece of the nitrogen cycle? *Biogeochemistry*, 57/58:99–136.
- Olson, T. A. and Odlaug, T. O. (1966). Limnological observations on western Lake Superior. In *Proc 9th Conf Great Lakes Res*, pages 109–118. International Association for Great Lakes Research.
- OMNR Forests Division (2006). Forest resources of Ontario. Technical report, Ontario Ministry of Natural Resources.
- O'Neill, R. V. and Rust, B. (1979). Aggregation error in ecological models. *Ecol Model*, 7:91–105.



- Oschlies, A. (2006). On the use of data assimilation in biogeochemical modelling. In Cassignet, E. P. and Verron, J., editors, *Ocean Weather Forecasting*. Springer.
- Ostrom, N., Long, D., Bell, E., and Beals, T. (1998). The origin and cycling of particulate and sedimentary organic matter and nitrate in Lake Superior. *Chemical Geology*, 152(1-2):13–28.
- Pahlow, M. (2005). Linking chlorophyll-nutrient dynamics to the Redfield N: C ratio with a model of optimal phytoplankton growth. *Mar Ecol-Prog Ser*, 287:33–43.
- Pahlow, M. and Oschlies, A. (2009). Chain model of phytoplankton P, N and light colimitation. *Mar Ecol-Prog Ser*, 376:69–83.
- Papaioannou, G., Papanikolaou, N., and Retalis, D. (1993). Relationships of photosynthetically active radiation and shortwave irradiance. *Theoretical and Applied Climatology*, 48:23–27.
- Patterson, J. and Hamblin, P. (1988). Thermal simulation of a lake with winter ice cover. *Limnol Oceanogr*, 33(3):323–338.
- Pitt, M. and Myung, I. (2002). When a good fit can be bad. *Trends in Cognitive Sciences*, 6(10):421–425.
- Poeter, E. and Anderson, D. (2005). Multimodel ranking and inference in ground water modeling. *Ground Water*, 43(4):597–605.

- Pomeroy, L. (1974). The ocean's food web, a changing paradigm. *Bioscience*, 24(9):499–504.
- Pomeroy, L. (2001). Caught in the food web: complexity made simple? *Scientia Marina*, 65(Supp. 2):31–40.
- Pregitzer, K., Zak, D., Burton, A., Ashby, J., and MacDonald, N. (2004). Chronic nitrate additions dramatically increase the export of carbon and nitrogen from northern hardwood ecosystems. *Biogeochemistry*, 68(2):179–197.
- Rissanen, J. (1987). Stochastic complexity and the mdl principle. *Econometric Reviews*, 6(1):85 – 102.
- Saito, M., Goepfert, T., and Ritt, J. (2008). Some thoughts on the concept of colimitation: Three definitions and the importance of bioavailability. *Limnol Oceanogr*, 53(1):276–290.
- Sarmiento, J., Slater, R., Fasham, M., Ducklow, H. W., Toggweiler, J. R., and Evans, G. T. (1993). A seasonal three-dimensional ecosystem model of nitrogen cycling in the north atlantic euphotic zone. *Global Biogeochem Cy*, 7(2):417–450.
- Schartau, M. and Oschlies, A. (2003). Simultaneous data-based optimization of a 1D-ecosystem model at three locations in the North Atlantic: Part I-method and parameter estimates. *J Mar Res*, 61(6):765–793.

- Shampine, L. and Reichelt, M. (1997). The Matlab ode suite. *SIAM Journal on Scientific Computing*, 18(1):1–22.
- Shannon, J. D. and Voldner, E. C. (1992). Deposition of sulfur and nitrogen oxide (NO<sub>x</sub>)-nitrogen to the Great Lakes estimated with a regional deposition model. *Environ Sci Technol*, 26(5):970–978.
- Six, K. and Maier-Reimer, E. (1996). Effects of plankton dynamics on seasonal carbon fluxes in an ocean general circulation model. *Global Biogeochem Cy*, 10(4):559–583.
- Smith, S. and Yamanaka, Y. (2007). Quantitative comparison of photoacclimation models for marine phytoplankton. *Ecol Model*, 201(3-4):547–552.
- Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and Linde, A. V. D. (2002). Bayesian measures of model complexity and fit. *J Royal Stat Soc B*, 64(4):583–639.
- Stasio, B. D., Hill, D., Kleinhans, J., Nibbelink, N., and Magnuson, J. (1996). Potential effects of global change on small north-temperate lakes: Physics, fish, and plankton. *Limnol Oceanogr*, 41(5):1136–1149.
- Steele, J. and Henderson, E. (1981). A simple plankton model. *Am Nat*, 117(5):676–691.
- Stein, M. L. (2005). Space-time covariance functions. *J Am Stat Assoc*, 100(496):310–321.
- Sterner, R. (2010). In situ-measured primary production in Lake Superior. *J Great Lakes Res*, 36:139–149.

- Sterner, R., Smutka, T., McKay, R., Xiaoming, Q., Brown, E., and Sherrell, R. (2004). Phosphorus and trace metal limitation of algae and bacteria in Lake Superior. *Limnol Oceanogr*, 49(2):495–507.
- Sterner, R. W., Anagnostou, E., Brovold, S., Bullerjahn, G., Finlay, J., Kumar, S., McKay, R., and Sherrell, R. (2007). Increasing stoichiometric imbalance in North America's largest lake: Nitrification in Lake Superior. *Geophys Res Lett*, 34:L10406.
- Streeter, H. and Phelps, E. (1925). A study of the pollution and natural purification of the ohio river. *U.S. Public Health Service, Pub. Health Bulletin No. 146*.
- Sugiura, N. (1978). Further analysts of the data by akaike's information criterion and the finite corrections. *Communications in Statistics - Theory and Methods*, 7(1):13–26.
- Thingstad, T. F. (1987). *Analyzing the "microbial loop": Experimental and mathematical model studies of interactions between heterotrophic bacteria and their trophic neighbors in pelagic food webs*. PhD thesis, University of Bergen.
- Tierney, L. (1994). Markov chains for exploring posterior distributions. *Ann Stat*, 22(4):1701–1762.
- Upper Lakes Reference Group (1977). The waters of Lake Huron and Lake Superior. Technical report, International Joint Commission, Windsor, Ontario.

- Urban, N. (2009). Nutrient cycling in Lake Superior: A retrospective and update. In Munawar, M. and Munawar, I. F., editors, *State of Lake Superior*. Aquatic Ecosystem Health and Management Society.
- Urban, N. and Eisenreich, S. (1988). Nitrogen cycling in a forested Minnesota bog. *Can J Bot*, 66:435–449.
- Urban, N. R., Apul, D. S., and Auer, M. T. (2004). Community respiration rates in Lake Superior. *J Great Lakes Res*, 30(Supplement 1):230–244.
- Urban, N. R., Auer, M. T., Green, S. A., Lu, X., Apul, D., Powell, K. D., and Bub, L. (2005). Carbon cycling in Lake Superior. *J Geophys Res*, 110:C06S90.
- U.S.EPA (2000). National air pollutant emission trends. Technical Report EPA-454/R-00-002.
- Verry, E. and Timmons, D. (1975). Precipitation nutrients in the open and under two forests in Minnesota. *J. For. Res*, 7:112–119.
- Ward, E. (2008). A review and comparison of four commonly used Bayesian and maximum likelihood model selection tools. *Ecol Model*, 211(1-2):1–10.
- Weiler, R. (1978). Chemistry of Lake Superior. *J Great Lakes Res*, 4(3-4):370–385.

## **Appendix A**

### **CAEDYM model descriptions and estimated parameter values**

**Table A.1:** CAEDYM model parameter descriptions.

	Notation	CAEDYM variable	Units
PAR extinction coefficient	$K_e$	N/A	$m^{-1}$
Maximum phytoplankton growth rate	$\mu_{max}$	Pmax	$day^{-1}$
Light saturation for max. production	$I_S$	ISt	$\mu\text{Einsteins } s^{-1} m^{-2}$
Specific attenuation coefficient <sup>1</sup>	$K_{ep}$	Kep	$\mu g \text{ chl } a \text{ L}^{-1} m^{-1}$
Temperature multiplier for growth	$\theta_{chl \ a}$	vT	-
Standard temperature for growth	$T_{std,chl \ a}$	Tsta	$^{\circ}C$
Optimum temperature for growth	$T_{opt,chl \ a}$	Topt	$^{\circ}C$
Maximum temperature for growth	$T_{max,chl \ a}$	Tmax	$^{\circ}C$
Phosphorus half-saturation coefficient	$K_P$	KP	$mg \text{ P } L^{-1}$
Internal P:chl <i>a</i> ratio	$x_P$	IPcon	$mg \text{ P } mg^{-1} \text{ chl } a$
Respiration rate coefficient	$k_r$	kr	$day^{-1}$
Temperature multiplier for respiration	$\theta_r$	vR	-
Respiration loss:total metabolic loss	$f_r$	f1	-
Fraction of metabolic loss as DOM	$f_{DOM}$	f2	-
Phytoplankton settling velocity	$v_s$	ws	$m \text{ s}^{-1}$
Temperature multiplier for bacteria	$\theta_b$	vT	-
Standard temperature for bacteria	$T_{std,b}$	Tsta	$^{\circ}C$
Optimum temperature for bacteria	$T_{opt,b}$	Topt	$^{\circ}C$
Maximum temperature for bacteria	$T_{max,b}$	Tmax	$^{\circ}C$
Max. transfer of POP to DOP	$k_{POP}$	POP1max	$day^{-1}$
Diameter of POM <sup>2</sup>	$d$	POMDiameter	$m$
Density of POM <sup>2</sup>	$\rho$	POMDensity	$kg \text{ m}^3$
Max. mineralization of DOP to PO <sub>4</sub>	$k_{DOP}$	DOP1	$day^{-1}$
C:chl <i>a</i> ratio	$Y_{C:chl \ a}$	Ycc	$mg \text{ C } mg^{-1} \text{ chl } a$
Temperature multiplier for SOD	$\theta_{SOD}$	vOP	-
Respiration C:O <sub>2</sub>	$Y_{C:O_2}$	YOC	$mg \text{ C } mg^{-1} O_2$
Static sediment exchange rate	$S_{SOD}$	rSOs	$g \text{ m}^{-2} d^{-1}$
Half-sat. coefficient for sediment flux	$K_{SOD}$	KSOs	$mg \text{ O}_2 L^{-1}$
Photorespiration DO loss	$k_P$	prc	-
DO Half-saturation for decomposition (water)	$K_{DO,w}$	KDOB (w)	$mg \text{ L}^{-1}$
DO Half-saturation for decomposition (sed.)	$K_{DO,s}$	KDOB (s)	$mg \text{ L}^{-1}$
aerobic/anaerobic factor (water)	$x_{a,w}$	fanB (w)	-
aerobic/anaerobic factor (sed.)	$x_{a,s}$	fanB (s)	-
Temperature multiplier for sediment fluxes	$\theta_{sed}$	Theta (sed)	-
Release rate of PO <sub>4</sub>	$k_{sed,PO_4}$	SmpPO4	$g \text{ m}^{-2} d^{-1}$
Release coefficient via O	$K_{DOS,PO_4}$	KDOS-PO4	$g \text{ m}^{-3}$
Release rate of DOP	$k_{sed,DOP}$	SmpDOPL	$g \text{ m}^{-2} d^{-1}$
Release coefficient via O	$K_{DOS,DOP}$	KDOS-DOP	$g \text{ m}^{-3}$

<sup>1</sup> Light is attenuated in the model according to Beer's Law.

<sup>2</sup> POM (POP) settling,  $v_s$ , is determined using Stoke's Law.

Model 1 = Equation 1, Model 2 = Equation 2, Model 3 = Equations 3–6, and Model 4 = Equations 3 and 7–10. Equations 8–10 represent phosphorus cycling in both sediment and water. The sediment flux term,  $F_{sed,i}$ , is therefore not applicable throughout the entire water column. Phytoplankton growth rate ( $\mu$ , Equations 4, 7, & 8) is defined as the first term of Equation 3.

$$\frac{\partial \text{chl } a}{\partial t} = \left\{ \mu_{max} \left[ \frac{I}{I_S} \exp \left( 1 - \frac{I}{I_S} \right) \right] - k_r \pm \frac{v_s}{\Delta z} \right\} \text{chl } a \quad (\text{A.1})$$

$$\frac{\partial \text{chl } a}{\partial t} = \left\{ \mu_{max} \frac{I}{I_S} \exp \left( 1 - \frac{I}{I_S} \right) f_{\text{chl } a}(T) - k_r \theta_r^{T-20} \pm \frac{v_s}{\Delta z} \right\} \text{chl } a \quad (\text{A.2})$$

$$\frac{\partial \text{chl } a}{\partial t} = \left\{ \mu_{max} \min \left[ \frac{I}{I_S} \exp \left( 1 - \frac{I}{I_S} \right), \frac{\text{PO}_4}{\text{PO}_4 + K_p} \right] f_{\text{chl } a}(T) - k_r \theta_r^{T-20} \pm \frac{v_s}{\Delta z} \right\} \text{chl } a \quad (\text{A.3})$$

where  $f_i(T) = \theta_i^{T-20} - \theta_i^{k(T-a)} + b$ ;

$$a = -\ln \left[ \frac{\theta_i^{T_{opt,i}-20}}{\theta_i^{kT_{opt,i}}} \right] \frac{1}{k \ln(\theta_i)},$$

$$b = 2 + \theta_i^{k(T_{std,i}-a)} - \theta_i^{T_{std,i}-20},$$

$$0 = k\theta_i^{kT_{opt,i}} (\theta_i^{T_{std,i}-20} - \theta_i^{T_{max,i}-20} - 1) - \theta_i^{T_{opt,i}-20} (\theta_i^{kT_{std,i}} - \theta_i^{kT_{max,i}})$$

where  $i = \text{chl } a, b$ .

$$\frac{\partial \text{PO}_4}{\partial t} = k_{DOP} f_b(T) \text{DOP} - \mu(\text{chl } a) x_P \quad (\text{A.4})$$

$$\frac{\partial \text{DOP}}{\partial t} = k_{POP} f_b(T) \text{POP} - k_{DOP} \theta_b^{T-20} \text{DOP} + (1 - f_r) f_{DOM} k_r \theta_r^{T-20} (\text{chl } a) x_P \quad (\text{A.5})$$

$$\frac{\partial \text{POP}}{\partial t} = -k_{POP} f_b(T) \text{POP} \pm \frac{v_{s,P}}{\Delta z} \text{POP} + (1 - f_r)(1 - f_{DOM}) k_r \theta_r^{T-20} (\text{chl } a) x_P \quad (\text{A.6})$$



$$\begin{aligned} \frac{\partial \text{DO}}{\partial t} = & F_{O_2}^{atm} - S_{SOD} \theta_{SOD}^{T-20} \frac{\text{DO}}{K_{SOD} + \text{DO}} - [\mu(1 - k_p) - k_r \theta_r^{T-20}] (\text{chl } a) \frac{1}{Y_{C:O_2}} Y_{C:chl\ a} \\ & - k_{DOP} f_b(T) f(DO) \text{DOP} - k_{POP} f_b(T) f(DO) \text{POP} \end{aligned} \quad (\text{A.7})$$

$$\text{where } f(DO) = \left( \frac{\text{DO}}{K_{DO,i} + \text{DO}} + x_{a,i} \frac{K_{DO,i}}{K_{DO,i} + \text{DO}} \right); i = w, s.$$

$$\frac{\partial \text{PO}_4}{\partial t} = k_{DOP} \theta_b^{T-20} f(DO) \text{DOP} - \mu (\text{chl } a) x_P \pm F_{PO_4}^{sed} \quad (\text{A.8})$$

$$\begin{aligned} \frac{\partial \text{DOP}}{\partial t} = & k_{POP} \theta_b^{T-20} f(DO) \text{POP} - k_{DOP} \theta_b^{T-20} f(DO) \text{DOP} \\ & + (1 - f_r) f_{DOM} k_r \theta_r^{T-20} (\text{chl } a) x_P \pm F_{DOP}^{sed} \end{aligned} \quad (\text{A.9})$$

$$\frac{\partial \text{POP}}{\partial t} = -k_{POP} \theta_b^{T-20} f(DO) \text{POP} \pm \frac{v_{s,P}}{\Delta z} \text{POP} + (1 - f_r)(1 - f_{DOM}) k_r \theta_r^{T-20} (\text{chl } a) x_P \quad (\text{A.10})$$

$$\text{where } F_i^{sed} = \theta_{sed}^{T-20} k_{sed,i} \frac{K_{DOS,i}}{K_{DOS,i} + \text{DO}} \frac{1}{\Delta z_{bot}}; i = \text{PO}_4, \text{DOP}.$$

**Table A.2:** Estimated values of model parameters and upper and lower constraints used in parameter estimation. The parameter ranges given in bold were based on literature values (see text for citations)

Parameter	Units	Model 1	Model 2	Model 3	Model 4	min	max
$K_e$	$m^{-1}$	0.362	0.773	0.399	0.393	0.25	3.00
$\mu_{max}$	$day^{-1}$	1.53	0.453	4.66	4.65	<b>0.4</b>	<b>5.0</b>
$I_S$	$\mu\text{Einsteins } s^{-1} m^{-2}$	156	881	192	191	10	1000
$K_{ep}$	$\mu\text{g chl } a L^{-1} m^{-1}$	0.0194	0.0171	0.0199	0.0147	0.013	0.020
$\theta_{chl a}$	-		1.04	1.12	1.12	1.00	1.12
$T_{std,chl a}$	$^{\circ}C$		18.1	20.8	20.7	17.5	22.5
$T_{opt,chl a}$	$^{\circ}C$		27.3	23.3	23.4	22.5	27.5
$T_{max,chl a}$	$^{\circ}C$		29.4	28.3	28.3	27.5	36.5
$K_P$	$mg P L^{-1}$			0.00100	0.00115	<b>0.001</b>	<b>0.163</b>
$x_P$	$mg P mg^{-1} chl a$			0.0746	0.0872	0.05	0.15
$k_r$	$day^{-1}$	0.101	0.0301	0.280	0.204	<b>0.03</b>	<b>0.59</b>
$\theta_r$	-		1.140	1.136	1.132	1.00	1.14
$f_r$	-			0.803	0.845	0.5	0.9
$f_{DOM}$	-			0.642	0.659	0.5	0.9
$v_s$	$m s^{-1}$	-3.84E-05	-1.85E-06	-1.05E-05	-1.05E-05	<b>-1.98E-04</b>	<b>-5.78E-07</b>
$\theta_b$	-			1.04	1.05	1.00	1.12
$T_{std,b}$	$^{\circ}C$			22.9	22.9	16.0	24.0
$T_{opt,b}$	$^{\circ}C$			27.7	27.7	24.0	32.0
$T_{max,b}$	$^{\circ}C$			32.0	32.1	32.0	38.0
$k_{POP}$	$day^{-1}$			0.0273	0.0267	0.001	0.100
$d$	$m$			8.42E-06	8.33E-06	1.0E-06	1.0E-04
$\rho$	$kg m^3$			101	143	100	1000
$k_{DOP}$	$day^{-1}$			0.0279	0.0206	0.001	0.100
$Y_{C:chl a}$	$mg C mg^{-1} chl a$				36.0	36.0	44.0
$\theta_{SOD}$	-				1.03	0.945	1.16
$Y_{C:O_2}$	$mg C mg^{-1} O_2$				2.59	2.4	2.9
$S_{SOD}$	$g m^{-2} d^{-1}$				0.210	0.18	0.22
$K_{SOD}$	$mg O_2 L^{-1}$				0.510	0.45	0.55
$k_P$	-				0.0140	0.0126	0.0154
$K_{DO,w}$	$mg L^{-1}$				3.0	2.7	3.3
$K_{DO,s}$	$mg L^{-1}$				2.50	2.25	2.75
$x_{a,w}$	-				0.30	0.27	0.33
$x_{a,s}$	-				0.30	0.27	0.33
$\theta_{sed}$	-				0.950	0.945	1.16
$k_{sed,PO_4}$	$g m^{-2} d^{-1}$				0.00261	0.00234	0.00286
$K_{DOS,PO_4}$	$g m^{-3}$				0.491	0.45	0.55
$k_{sed,DOP}$	$g m^{-2} d^{-1}$				1.0E-05	9.0E-06	1.1E-05
$K_{DOS,DOP}$	$g m^{-3}$				1.0E-05	9.0E-06	1.1E-05



## **Appendix B**

### **CAEDYM/Simulated Annealing MATLAB Code**

The code presented here is designed to interface with the CAEDYM/DYRESM program and data files. These are freely available from the Center for Water Research at the University of Western Australia at <http://www.cwr.uwa.edu.au>.

```

function [x fval]=optimizeCAEDYManneal(Data,dates_data,dates_model,PlotData)
% optimizeCAEDYManneal
% This MATLAB routine utilizes the simulated annealing (SA) routine in
% the Optimization Toolbox. For each model run, the CAEDYM WQparameters.dat
% file is rewritten and DYRESM-CAEDYM is remotely run (on flexruns.eecn.mtu.edu;
% an ssh key pair must be generated in order for this step to work).
%
% The code shown here is for Model 4 (38 parameters); Models 1-3 omit some
% parameters and the corresponding 'fwrite' lines.
%=====

%Call optimization function

%% Define Upper and Lower Bounds, Initial Values of parameter vector
%X0=[ke Pmax IST Kep vT Tsta Topt Tmax KP IP kr vR f1 f2
% ws vT Tsta Topt Tmax POP1 POMd POMp DOP1 Ycc vOP YOC rSOs KS0s prc
% KDOBw KDOBs fanBw fanBs Theta SmpPO4 KDOS-PO4 SmpDOP1 KDOS-DOP];
lb=[0.25 0.40 10 0.013 1.00 17.5 22.5 27.5 .001 0.05 0.03 1.00 0.5 0.5,...
-198.0 1.00 16.0 24.0 32.0 0.001 .1 0.01 0.001 36 .945 2.4 0.18 0.45 .0126,...
2.700 2.250 0.27 0.27 0.945 0.00234 0.45 0.90000 0.9000 ];
ub=[3.00 5.00 1000 0.020 1.12 22.5 27.5 36.5 .163 0.15 0.59 1.14 0.9 0.9,...
-0.578 1.12 24.0 32.0 38.0 0.1 10 1.00 0.1 44 1.155 2.9 0.22 0.55 .0154,...
3.300 2.750 0.33 0.33 1.155 0.00286 0.55 1.100 1.1000 ];
X0=[0.46012 4.72263 198.99686 0.01818 1.09864 20.82427 23.38584 28.30471,...
0.00118 0.10514 0.16676 1.07358 0.79500 0.62997 -1.01340e-05 1.04818,...
22.89840 27.72764 32.20402 0.02151 8.97524e-06 4.05322e+02 0.04340 43.90552,...
1.04930 2.79025 0.18205 0.47243 0.01327 2.9 2.6 0.3 0.3 1.01 0.00268,...
0.48558 0.00001 0.00001];

%-----[ Enter SA options here ]-----%
options=saplotset('PlotFcns',{@saplotbestf,@saplotf,@saplotstopping,...
@saplottemperature},'Display','iter','ReannealInterval',10,...
'TolFun',1,'InitialTemperature',1000);

[x fval]=simulannealbnd(@runDYRESM,X0,lb,ub,options);

%-----%
%
function SSE=runDYRESM(Params)
% re-write WQparameters.dat with new parameters:
%-----[ This section of code must be carefully re-written ]-----%
% [ when estimated parameters are added or removed. ] %

fid=fopen('WQparameters_test.dat');
for i=1:1508
    lines{i}=fgets(fid);
end
fclose(fid);
fid=fopen('WQparameters_test.dat','wt');

for i=1:6
    fwrite(fid,lines{i});
end
fprintf(fid,' %15f %15f : PAR, Photosynthetically Active\n',Params(1));
for i=8:13

```

```

        fwrite(fid,lines{i});
    end
    fprintf(fid,'      %2.5f          ! DINOOF\n',Params(2));
    for i=15:21
        fwrite(fid,lines{i});
    end
    fprintf(fid,'      %2.5f          ! DINOOF\n',Params(24));
    for i=23:47
        fwrite(fid,lines{i});
    end
    fprintf(fid,'      %3.5f\n',Params(3));
    for i=49:55
        fwrite(fid,lines{i});
    end
    fprintf(fid,'      %1.5f\n',Params(4));
    for i=57:65
        fwrite(fid,lines{i});
    end
    fprintf(fid,'      %1.5f\n',Params(9));
    for i=67:201
        fwrite(fid,lines{i});
    end
    fprintf(fid,'      %1.5f\n',Params(10));
    for i=203:235
        fwrite(fid,lines{i});
    end
    fprintf(fid,'      %1.5f\n',Params(5));
    for i=237:243
        fwrite(fid,lines{i});
    end
    fprintf(fid,'      %2.5f\n',Params(6));
    for i=245:251
        fwrite(fid,lines{i});
    end
    fprintf(fid,'      %2.5f\n',Params(7));
    for i=253:259
        fwrite(fid,lines{i});
    end
    fprintf(fid,'      %2.5f\n',Params(8));
    for i=261:269
        fwrite(fid,lines{i});
    end
    fprintf(fid,'      %1.5f\n',Params(11));
    for i=271:277
        fwrite(fid,lines{i});
    end
    fprintf(fid,'      %1.5f\n',Params(12));
    for i=279:285
        fwrite(fid,lines{i});
    end
    fprintf(fid,'      %1.5f\n',Params(13));
    for i=287:293
        fwrite(fid,lines{i});
    end
    fprintf(fid,'      %1.5f\n',Params(14));
    for i=295:419

```

```

        fwrite(fid,lines{i});
    end
    fprintf(fid,'    %1.5e\n',Params(15)*1E-6);
    for i=421:1262
        fwrite(fid,lines{i});
    end
    fprintf(fid,'    %1.5f\n',Params(25));
    for i=1264:1264
        fwrite(fid,lines{i});
    end
    fprintf(fid,'    %1.5f\n',Params(26));
    for i=1266:1269
        fwrite(fid,lines{i});
    end
    fprintf(fid,'    %1.5f\n',Params(27));
    fprintf(fid,'    %1.5f\n',Params(28));
    fprintf(fid,'    %1.5f\n',Params(29));
    for i=1273:1276
        fwrite(fid,lines{i});
    end
    fprintf(fid,'    %1.1f\n',Params(30));
    fprintf(fid,'    %1.1f\n',Params(31));
    for i=1279:1279
        fwrite(fid,lines{i});
    end
    fprintf(fid,'    %1.1f\n',Params(32));
    fprintf(fid,'    %1.1f\n',Params(33));
    for i=1282:1283
        fwrite(fid,lines{i});
    end
    fprintf(fid,'    %1.5f\n',Params(16));
    for i=1285:1285
        fwrite(fid,lines{i});
    end
    fprintf(fid,'    %2.5f\n',Params(17));
    for i=1287:1287
        fwrite(fid,lines{i});
    end
    fprintf(fid,'    %2.5f\n',Params(18));
    for i=1289:1289
        fwrite(fid,lines{i});
    end
    fprintf(fid,'    %2.5f\n',Params(19));
    for i=1291:1313
        fwrite(fid,lines{i});
    end
    fprintf(fid,'    %2.5f\n',Params(20));
    for i=1315:1318
        fwrite(fid,lines{i});
    end
    fprintf(fid,'    %1.5e\n',Params(21)*1e-5);
    for i=1320:1321
        fwrite(fid,lines{i});
    end
    fprintf(fid,'    %1.5e\n',Params(22)*1e4);
    for i=1323:1333

```

```

        fwrite(fid,lines{i});
    end
    fprintf(fid,'      %2.5f\n',Params(23));
    for i=1335:1368
        fwrite(fid,lines{i});
    end
    fprintf(fid,'      %1.2f\n',Params(34));
    for i=1370:1370
        fwrite(fid,lines{i});
    end
    fprintf(fid,'      %1.5f\n',Params(35));
    fprintf(fid,'      %1.5f\n',Params(36));
    for i=1373:1391
        fwrite(fid,lines{i});
    end
    fprintf(fid,'      %1.5f\n',Params(37)*1E-5);
    for i=1393:1393
        fwrite(fid,lines{i});
    end
    fprintf(fid,'      %1.5f\n',Params(38)*1E-5);
    for i=1395:1508
        fwrite(fid,lines{i});
    end
    fclose(fid);

% Transmit edited *.cfg file:
unix('scp WQparameters_test.dat cpmcdona@flexruns.eecn.mtu.edu:~/dat7/
WQparameters_test.dat');

%-----%

Params

%run DYRESM
[null null2]=unix('ssh -l cpmcdona flexruns.eecn.mtu.edu "cd dat7 && createdYref
troutlake.stg troutlake_noice.met troutlake.inf troutlake.wdr DYref.nc"');
[null null2]=unix('ssh -l cpmcdona flexruns.eecn.mtu.edu "cd dat7 && createDYsim
troutlake.pro troutlake.par troutlake_1.con DYsim.nc"');
[null null2]=unix('ssh -l cpmcdona flexruns.eecn.mtu.edu "cd dat7 && extractDYinfo
DYref.nc DYsim.nc troutlake.cfg"');
[null output]=unix('ssh -l cpmcdona flexruns.eecn.mtu.edu "cd dat7 && dycd"');

%if-else loop checks to make sure DYRESM-CAEDYM executed successfully;
%if not, a 'high' value is assigned to the fitness function.
[null sizeOutput]=size(output);
output(sizeOutput-50:sizeOutput-1)
if sum(output(sizeOutput-50:sizeOutput-1))==...
    '===== DYRESM-CAEDYM simulation completed =====')==50;
%get CHLOR data from DYRESM output file
unix('scp cpmcdona@flexruns.eecn.mtu.edu:~/dat7/DYsim.nc /Users/corymcdonald/Documents/
Research/Trout\ Lake/DYRESM.CAEDYM.v2/dat\ 7');
chlor=nc_varget('DYsim.nc','dyresmDINOF_Var');
heights=nc_varget('DYsim.nc','dyresmLAYER_HTS_Var');
[days depth]=size(chlor);
for i=1:depth

```



```

        xi(i)=i-1;
        xi=xi';
    end
    for i=1:days
        % pull out each row of chl & depth data, discard NaNs
        Y=chl(i,:);
        x=heights(i,:);
        k = find(~isnan(Y));
        Y = Y(k);
        Y=Y';
        x=x(k);
        x=x';
        % convert heights to depths
        top=x(1);
        x=-(x-top);
        % interpolate to 1-m grid
        yi=interp1q(x,Y,xi);
        Model(:,i)=yi;
    end

    % remove NaN-containing rows
    Model(any(isnan(Model),2),:) = [];

    SSE=0;
    e=[];
    y=1;
    z=1;

    [dates null]=size(dates_data);
    [modeldepths null]=size(Model);
    for i=1:dates
        if dates_data(i)>=min(dates_model) && dates_data(i)<=max(dates_model)
            plotboundary(z)=i;
            z=z+1;

            j=find(dates_model==dates_data(i));

            [depths null]=size(Data{i});

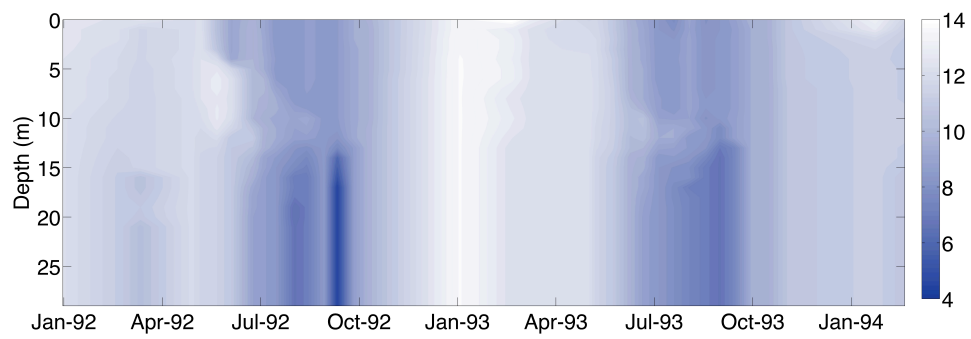
            for k=1:depths
                if Data{i}(k,1)<=modeldepths-1
                    SSE=SSE+(Data{i}(k,2)-Model(Data{i}(k,1)+1,j))^2;
                end
            end
        end
    end

    else
        SSE=1E6
    end
end
end
end

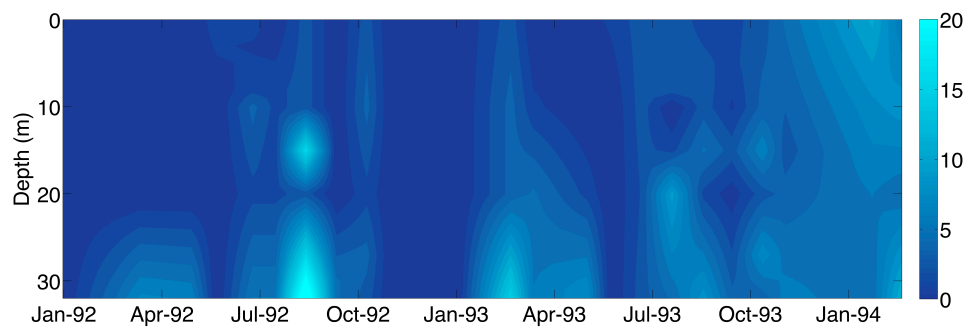
```

## **Appendix C**

### **Trout Lake dissolved oxygen and total dissolved phosphorus data**



(a) DO



(b) TDP

**Figure C.1:** Trout Lake DO and TDP profiles for 1992–1993. Note that elevated dissolved phosphorus in the hypolimnion is observed when DO is depleted (especially in 1992), suggesting sediment phosphorus release is occurring during the stratified period

## **Appendix D**

### **KITES calibration data**

The data presented here are compiled from the KITES project (Auer and Johnson, 2004; Green and Eadie, 2004) 1999–2000, and represent only measurements taken at site HN 210 (47 ° 24.33' N, 88 ° 44.29' W).

**Table D.1:** Chlorophyll *a* data ( $\mu\text{g L}^{-1}$ ).

Depth	Julian Day (11/199 = 1)												659									
	127	137	155	169	183	195	209	221	236	281	484	495										
0	0.807	0.684	0.784	0.513	0.357	0.37	0.347	0.506	0.3155	0.891	0.646	0.795	NaN	0.796	0.266	0.313	0.346	0.843	0.842	0.994	0.803	1.0465
5	NaN	0.7265	0.63	0.374	0.409	0.4105	0.172	0.512	0.517	0.96	0.502	0.843	0.5923	0.454	0.298	0.336	0.412	0.901	0.886	1.115	0.839	1.177
10	NaN	0.871	0.72	0.348	0.707	NaN	0.196	0.423	NaN	0.947	NaN	NaN	NaN	NaN	0.291	NaN	0.482	0.887	NaN	1.25	NaN	1.289
15	NaN	NaN	0.577	0.474	0.474	NaN	0.158	0.972	NaN	0.998	NaN	NaN	NaN	NaN	0.533	NaN	0.8	1.068	NaN	1.318	NaN	1.13
18	NaN	NaN	NaN	0.577	0.474	NaN	NaN	0.972	NaN	0.998	NaN	NaN	NaN	NaN	0.533	NaN	0.8	1.068	NaN	1.265	NaN	1.13
20	NaN	NaN	NaN	0.463	0.672	NaN	0.174	1.369	NaN	1.032	NaN	NaN	NaN	NaN	0.937	NaN	1.562	1.13	NaN	1.287	NaN	1.164
22	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	1.504	NaN	1.193	NaN	NaN
24	NaN	0.839	NaN	NaN	0.911	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	1.833	NaN	NaN	1.598	NaN	NaN	NaN	NaN
25	NaN	NaN	NaN	0.458	NaN	NaN	NaN	1.972	NaN	0.975	NaN	NaN	NaN	NaN	NaN	NaN	NaN	1.124	NaN	1.334	NaN	NaN
26	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	2.045	NaN	2.048	1.815	NaN	NaN	NaN	NaN
28	NaN	NaN	NaN	NaN	0.851	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	2.403	NaN	1.879	1.75	NaN	NaN	NaN	NaN
30	NaN	NaN	0.57	0.468	0.773	NaN	NaN	1.599	NaN	0.776	NaN	NaN	NaN	NaN	2.38	NaN	1.81	1.841	NaN	NaN	NaN	0.959
32	NaN	NaN	NaN	NaN	0.937	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	2.468	NaN	1.788	NaN	NaN	NaN	NaN	NaN
34	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	2.262	NaN	1.878	1.691	NaN	NaN	NaN	NaN
35	NaN	NaN	NaN	0.498	NaN	NaN	0.087	1.068	NaN	0.596	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	1.132	NaN	NaN
36	NaN	NaN	NaN	NaN	0.783	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	1.843	NaN	NaN	NaN	NaN	NaN	NaN	NaN
38	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	1.108	NaN	NaN
40	NaN	NaN	NaN	0.559	0.722	NaN	0.092	0.606	NaN	0.488	NaN	NaN	NaN	NaN	1.776	NaN	1.356	1.724	NaN	1.092	NaN	NaN
42	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	0.917	NaN	NaN
50	NaN	0.878	0.505	0.531	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
60	NaN	NaN	NaN	NaN	0.477	NaN	0.043	NaN	NaN	0.404	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
75	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	0.419	NaN	0.276	NaN	NaN	NaN	NaN	NaN
80	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	0.329	NaN	0.185	NaN	0.143
100	NaN	0.86	NaN	0.514	0.344	NaN	NaN	NaN	NaN	0.188	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
152	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	0.188	NaN	0.191	NaN	NaN	NaN	NaN	NaN
155	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	0.158	NaN	NaN	NaN	NaN
156	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	0.122	NaN	NaN
158	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	0.065

**Table D.2:** Total dissolved phosphorus data ( $\mu\text{M}$ ).

Depth	Julian Day (1/1/99 = 1)													
	126	136	194	235	280	482	496	538	576	602	635	658		
0	0.0366	0.0103	0.1262	NaN	0.1176	0.0274	0.0343	0.0325	NaN	0.0285	0.0375	0.0836		
10	0.0589	0.0177	0.0509	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN		
25	NaN	NaN	NaN	NaN	0.0706	NaN	NaN	NaN	NaN	NaN	NaN	NaN		
28	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	0.0132	NaN	NaN	NaN		
30	0.0149	0.0883	0.0078	0.0078	NaN	0.0147	0.0144	0.0314	NaN	0.042	NaN	NaN		
40	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	0.0115	NaN		
80	NaN	NaN	NaN	NaN	NaN	0.0274	0.0277	0.0314	0.0132	0.0285	NaN	NaN		
153	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	0.0132	NaN	NaN	NaN		
154	NaN	NaN	0.0186	NaN	0.0119	NaN	0.0115	NaN	NaN	NaN	NaN	NaN		
155	NaN	NaN	NaN	NaN	NaN	0.0274	NaN	NaN	NaN	0.0555	NaN	NaN		
156	0.0388	NaN	NaN	NaN	NaN	NaN	NaN	0.0455	NaN	NaN	NaN	NaN		
157	NaN	0.042	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN		
158	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	0.0718		

**Table D.3:** Particulate organic phosphorus data ( $\mu\text{M}$ ).

Depth	Julian Day (1/1/99 = 1)										
	126	136	194	280	483	496	576	602	635	658	
0	0.0309	0.0358	0.0445	0.0486	NaN	0.05	0.0209	0.0396	0.0318	NaN	
10	NaN	NaN	0.0422	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
28	NaN	NaN	NaN	NaN	NaN	NaN	0.0419	NaN	NaN	NaN	
30	NaN	NaN	NaN	NaN	0.032	0.0323	NaN	0.0477	NaN	NaN	
40	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	0.0509	NaN	
80	NaN	NaN	NaN	NaN	NaN	0.0318	0.036	0.0373	0.0289	NaN	
153	NaN	NaN	NaN	NaN	NaN	NaN	0.0248	NaN	NaN	NaN	
154	NaN	NaN	0.0294	0.0291	NaN	0.0289	NaN	NaN	NaN	NaN	
155	NaN	NaN	NaN	NaN	NaN	NaN	NaN	0.0198	NaN	NaN	
156	0.0301	NaN	NaN	NaN	NaN	NaN	NaN	NaN	0.0309	NaN	
157	NaN	0.0497	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
158	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	0.0236	

**Table D.4:** Dissolved organic carbon data ( $\mu\text{M}$ ).

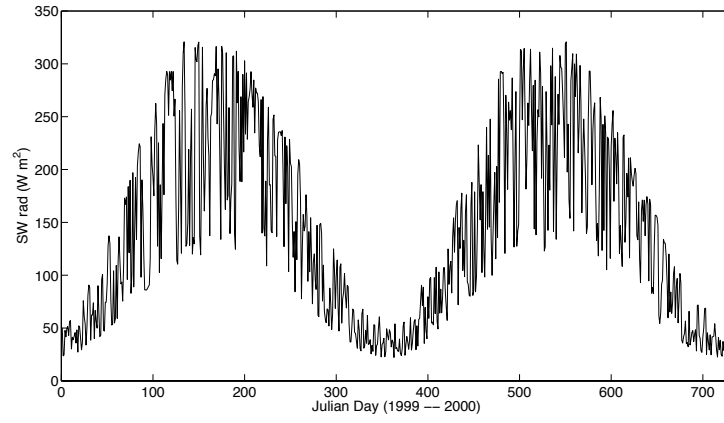
Depth	Julian Day (1/1/99 = 1)															
	137	155	183	195	236	281	484	497	526	539	577	603	604	636	659	
0	113.91	121.45	109.82	118.8	116.23	123.58	NaN	101.79	106.33	95.17	178.915	126.83	147.25	136.08	146.54	
5	111.8	NaN	109.04	118.49	118.98	123.17	76.25	108.72	NaN	NaN	110.58	NaN	NaN	NaN	NaN	
10	111.7	NaN	NaN	117.69	NaN	122.5	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
15	NaN	NaN	109.88	NaN	NaN	122.82	NaN	NaN	NaN	NaN	141.25	NaN	NaN	NaN	NaN	
20	112.31	NaN	110.98	NaN	NaN	118.03	NaN	NaN	NaN	NaN	162.08	NaN	NaN	NaN	NaN	
24	NaN	NaN	111.57	NaN	NaN	NaN	NaN	NaN	NaN	NaN	119.67	NaN	NaN	NaN	NaN	
25	NaN	NaN	NaN	NaN	NaN	119.53	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
28	NaN	NaN	110.33	119.41	NaN	NaN	NaN	NaN	NaN	NaN	113.045	NaN	NaN	NaN	NaN	
30	113.88	NaN	NaN	NaN	111.82	114.45	NaN	105.655	NaN	NaN	110.33	NaN	NaN	NaN	NaN	
32	NaN	NaN	110.17	NaN	NaN	NaN	NaN	NaN	NaN	NaN	119.33	NaN	NaN	NaN	NaN	
34	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	95.42	NaN	NaN	NaN	NaN	
35	NaN	NaN	NaN	NaN	NaN	113.56	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
36	NaN	NaN	107.83	NaN	NaN	NaN	NaN	NaN	NaN	NaN	85.58	NaN	NaN	NaN	NaN	
40	NaN	NaN	109.49	NaN	NaN	110.74	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
50	112.65	NaN	NaN	NaN	NaN	NaN	NaN	105.51	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
60	NaN	NaN	108.06	NaN	NaN	107.11	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
80	NaN	NaN	NaN	NaN	NaN	NaN	NaN	101.1	NaN	NaN	230.83	NaN	NaN	NaN	NaN	
100	110.41	NaN	NaN	NaN	NaN	106.76	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
134	NaN	NaN	NaN	NaN	106.16	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
140	NaN	NaN	NaN	116.37	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
142	111.85	NaN	110.05	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
145	NaN	NaN	NaN	117.96	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
147	111.64	NaN	107.55	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
149	NaN	NaN	NaN	NaN	NaN	NaN	NaN	106.08	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
150	NaN	NaN	NaN	115.5	NaN	107.18	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
152	112.07	NaN	108.31	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
153	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	130.5	NaN	NaN	NaN	NaN	
154	NaN	NaN	NaN	NaN	108.6	NaN	NaN	104.08	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
156	NaN	NaN	109.23	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	



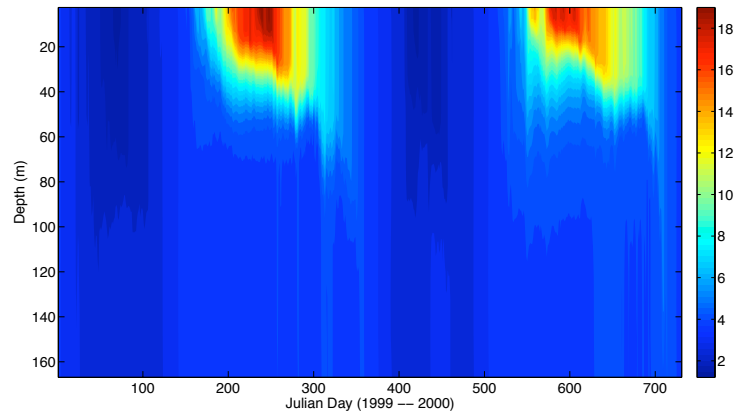


## **Appendix E**

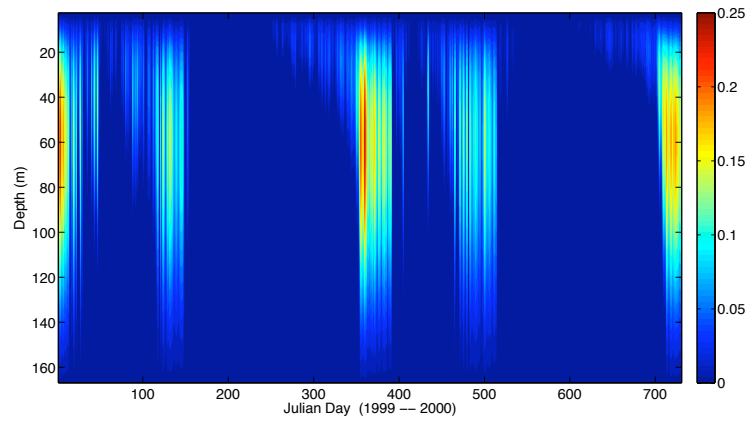
### **Lake Superior hydrodynamic emulator input data**



(a) Shortwave radiation,  $\text{W m}^{-2}$



(b) Temperature,  $^{\circ}\text{C}$



(c) Diffusivity,  $\text{m}^2 \text{s}^{-1}$

**Figure E.1:** Forcing data for 1-D emulator. Temperature and diffusivity data are outputs from the 3-D hydrodynamic model. Solar radiation data are from NARR (Mesinger et al., 2006).

**Table E.1:** 1-D emulator grid characteristics (equivalent to the 3-D grid at KITES site HN 210).

Layer Number	Midpoint Depth	Thickness	Bottom Depth
1	-2.5	5	5
2	-7.5	5	10
3	-12.5	5	15
4	-17.5	5	20
5	-22.5	5	25
6	-27.5	5	30
7	-32.5	5	35
8	-37.5	5	40
9	-42.5	5	45
10	-47.5	5	50
11	-52.75	5.5	55.5
12	-58.55	6.1	61.6
13	-64.95	6.7	68.3
14	-72	7.4	75.7
15	-79.8	8.2	83.9
16	-88.4	9	92.9
17	-97.85	9.9	102.8
18	-108.3	11	113.8
19	-119.85	12.1	125.9
20	-133.1	14.4	140.3
21	-148.7	16.8	157.1
22	-166.9	19.6	176.7



## **Appendix F**

### **Lake Superior biogeochemical model code**

The Matlab code presented in this appendix describes the NPD, NPZD, and simplified carbon cycle models described in Chapters 3 and 4, processes the model output, and computes the cost function. These routines are designed to be compatible with Marko Laine's MCMC toolbox for Matlab, available at <http://www.helsinki.fi/~mjlane/mcmc/>.

```

function ydot = BGCSys(t,y,theta,xdata)
% ode system function for CyCLEs ecological model clone.
%
% This is the model function for the NPD model presented in chapters 3/4.
%

% control variables are assumed to be saved
% at each time unit interval (1 day)
SWin = xdata.x(ceil(t),2); % daily incident sw rad (W m^-2)
layers = size(xdata.layers,1); % number of vert. layers in GCM output
Temp = xdata.x(ceil(t),3:layers+2); % surface temp, deg. C
Vvert = xdata.x(ceil(t),layers+3:2*layers+2); % vertical water movement (m/s)
Vvert = Vvert.*86400;
Dvert = xdata.x(ceil(t),2*layers+3:3*layers+2); % turb. diffusion (m^2/s)
Dvert = Dvert.*86400;
zmid = xdata.layers(:,1).*-1; % mid-layer depth (m)
zheight = xdata.layers(:,2); % layer thickness (m)
xyarea = xdata.area; % cell area (m^2)

% Preallocate arrays for speed:
Nut=zeros(1,layers);
Phy=zeros(1,layers);
DOP=zeros(1,layers);
DOC=zeros(1,layers);
POP=zeros(1,layers);
POC=zeros(1,layers);
CHL=zeros(1,layers);
atten=zeros(1,layers);
ydot=zeros(layers,1);

GPP=zeros(1,layers);

% Define state variables
for i=1:layers
    Nut(i) = y(8*(i-1)+1); %
    Phy(i) = y(8*(i-1)+2); %
    DOP(i) = y(8*(i-1)+3); %
    DOC(i) = y(8*(i-1)+4); %
    POP(i) = y(8*(i-1)+5); %
    POC(i) = y(8*(i-1)+6); %
    CHL(i) = y(8*(i-1)+7); %
    GPP(i) = y(8*(i-1)+8); %
end

% Define biochemical parameters
k0 = theta(1); % attenuation coefficient water (m^-1)
kc = theta(2); % attenuation coefficient chl (m^2 (mg chl)^-1)
phygrow = theta(3); % max growth rate for phy1 (d^-1)
mort = theta(4); % phy1 mortality rate (d^-1)
ilim = theta(5); % phy1 light 1/2 sat. (W/m2)
nlim = theta(6); % phy1 po4 1/2 sat. (uM)
reminp = theta(7); % remin rate of DOP/POP (d^-1)
reminc = theta(8); % remin rate of DOC/POC (d^-1)
dopfracm1 = theta(9); % fraction of mortality to DOP/POP
dopfracmc1 = theta(10); % fraction of mortality to DOP/POF

```

```

Rcp = theta(11);          % ratio C to P in phy1
% minimum populations
  phymin = theta(12);    % seed population of phy (uM)
wp_sink = theta(13); % sinking rate of POP (m/d)
wc_sink = theta(14); % sinking rate of POC (m/d)
% for chl
  chlpmax = theta(15);
  chlpmin = theta(16);
  istar = theta(17);      % uEin/m2/d
% for temperature function of growth rate
  temp_coeff_base = theta(18);          % unitless, Eppley
  temp_coeff_norm = theta(19);

  pym(1:4) = max((Phy(1:4)-phymin),0);
  pym(5:layers) = Phy(5:layers);

  Nutup(1) = 0;
  Phyup(1) = 0;
  DOPup(1) = 0;
  DOCup(1) = 0;
  POPup(1) = 0;
  POCup(1) = 0;
  Nutup(2:layers) = Nut(1:layers-1);
  Phyup(2:layers) = Phy(1:layers-1);
  DOPup(2:layers) = DOP(1:layers-1);
  DOCup(2:layers) = DOC(1:layers-1);
  POPup(2:layers) = POP(1:layers-1);
  POCup(2:layers) = POC(1:layers-1);
  zup(1) = 0;
  zup(2:layers) = zmid(1:layers-1);

  Nutdown(layers) = 0;
  Phydown(layers) = 0;
  DOPdown(layers) = 0;
  DOCdown(layers) = 0;
  POPdown(layers) = 0;
  POCdown(layers) = 0;
  Nutdown(1:layers-1) = Nut(2:layers);
  Phydown(1:layers-1) = Phy(2:layers);
  DOPdown(1:layers-1) = DOP(2:layers);
  DOCdown(1:layers-1) = DOC(2:layers);
  POPdown(1:layers-1) = POP(2:layers);
  POCdown(1:layers-1) = POC(2:layers);
  zdown(layers) = 0;
  zdown(1:layers-1) = zmid(2:layers);
  Ddown(layers) = 0;
  Ddown(1:layers-1) = Dvert(2:layers);

% % Photosynthetically active radiation
PAR0 = SWin* 0.45; % Papaioannou et al. 2004 = 0.473
atten(1)=(k0+kc*CHL(1))*zmid(1);
for i=2:layers
  atten(i) = atten(i-1)+(k0+kc*CHL(i-1))*zheight(i-1)/2+...
    (k0+kc*CHL(i))*zheight(i)/2;
end
PAR=PAR0*exp(-atten);

```



```

% Phytoplankton growth rates
mu = phygrow.*(PAR./(ilim+PAR)).*(Nut./(nlim+Nut));
tempfunc=1/temp_coeff_norm*temp_coeff_base.^Temp;
mu = mu.*tempfunc;

reminp_new = reminp.*tempfunc;
reminc_new = reminc.*tempfunc;

%Calculate derivatives...

dotNut = -mu.*Phy+reminp_new.*DOP+reminp_new.*POP;
dotPhy = mu.*Phy-mort.*pym;
dotDOP = dopfracm1.*mort.*pym-reminp_new.*DOP;
dotDOC = Rcp.*(dopfracm1.*mort.*pym) - reminc_new.*DOC;
dotPOP = (1-dopfracm1).*mort.*pym - reminp_new.*POP;
dotPOC = (1-dopfracm1).*Rcp.*mort.*pym - reminc_new.*POC;

% not GPP but sp. growth rate.
dotGPP = mu.*Phy.*Rcp;

%Sinking:
dotPOP(1:layers-1) = dotPOP(1:layers-1) + wp_sink.*(POPup(1:layers-1)./(zmid(1:layers-1)'-zup
(1:layers-1))-POP(1:layers-1)./(zdown(1:layers-1)-zmid(1:layers-1)'));
dotPOP(layers) = dotPOP(layers) + wp_sink.*POPup(layers)./(zmid(layers)'-zup(layers));
dotPOC(1:layers-1) = dotPOC(1:layers-1) + wc_sink.*(POCup(1:layers-1)./(zmid(1:layers-1)'-zup
(1:layers-1))-POC(1:layers-1)./(zdown(1:layers-1)-zmid(1:layers-1)'));
dotPOC(layers) = dotPOC(layers) + wc_sink.*POCup(layers)./(zmid(layers)'-zup(layers));

%Diffusion:
dotNut = dotNut + (Dvert.*xyarea./(zmid'-zup).*(Nutup-Nut))./(xyarea.*(zmid'-zup))+
(Ddown.*xyarea./(zdown-zmid').*(Nutdown-Nut))./(xyarea.*(zdown-zmid'));
dotPhy = dotPhy + (Dvert.*xyarea./(zmid'-zup).*(Phyup-Phy))./(xyarea.*(zmid'-zup))+
(Ddown.*xyarea./(zdown-zmid').*(Phydown-Phy))./(xyarea.*(zdown-zmid'));
dotDOP = dotDOP + (Dvert.*xyarea./(zmid'-zup).*(DOPup-DOP))./(xyarea.*(zmid'-zup))+
(Ddown.*xyarea./(zdown-zmid').*(DOPdown-DOP))./(xyarea.*(zdown-zmid'));
dotDOC = dotDOC + (Dvert.*xyarea./(zmid'-zup).*(DOCup-DOC))./(xyarea.*(zmid'-zup))+
(Ddown.*xyarea./(zdown-zmid').*(DOCdown-DOC))./(xyarea.*(zdown-zmid'));
dotPOP = dotPOP + (Dvert.*xyarea./(zmid'-zup).*(POPup-POP))./(xyarea.*(zmid'-zup))+
(Ddown.*xyarea./(zdown-zmid').*(POPdown-POP))./(xyarea.*(zdown-zmid'));
dotPOC = dotPOC + (Dvert.*xyarea./(zmid'-zup).*(POCup-POC))./(xyarea.*(zmid'-zup))+
(Ddown.*xyarea./(zdown-zmid').*(POCdown-POC))./(xyarea.*(zdown-zmid'));

% Chlorophyll
t1 = chlpmx-(chlpmax-chlpmin).*min(PAR./istar,1);
t3 = max(chlpmin, t1);
cpratio1 = min(chlpmax,t3);
newCHL = cpratio1.*(Phy+dotPhy);
dotCHL = newCHL - CHL;

for i=1:layers
    ydot((i-1)*8+1,1) = dotNut(i);
    ydot((i-1)*8+2,1) = dotPhy(i);
    ydot((i-1)*8+3,1) = dotDOP(i);

```

```
    ydot((i-1)*8+4,1) = dotDOC(i);  
    ydot((i-1)*8+5,1) = dotPOP(i);  
    ydot((i-1)*8+6,1) = dotPOC(i);  
    ydot((i-1)*8+7,1) = dotCHL(i);  
    ydot((i-1)*8+8,1) = dotGPP(i);  
end
```

```

function ydot = BGCSys(t,y,theta,xdata)
% ode system function for CyCLEs ecological model clone.
%
% This is the model function for the NPZD model presented in chapter 4.

% control variables are assumed to be saved
% at each time unit interval (1 day)
SWin = xdata.x(ceil(t),2); % daily incident sw rad (W m-2)
layers = size(xdata.layers,1); % number of vert. layers in GCM output
Temp = xdata.x(ceil(t),3:layers+2); % surface temp, deg. C
Vvert = xdata.x(ceil(t),layers+3:2*layers+2); % vertical water movement (m/s)
Vvert = Vvert.*86400;
Dvert = xdata.x(ceil(t),2*layers+3:3*layers+2); % turb. diffusion (m2/s)
Dvert = Dvert.*86400;
zmid = xdata.layers(:,1).*-1; % mid-layer depth (m)
zheight = xdata.layers(:,2); % layer thickness (m)
xyarea = xdata.area; % cell area (m2)

% Preallocate arrays for speed:
Nut=zeros(1,layers);
Phy=zeros(1,layers);
Zoo=zeros(1,layers);
DOP=zeros(1,layers);
ZoC=zeros(1,layers);
DOC=zeros(1,layers);
POP=zeros(1,layers);
POC=zeros(1,layers);
CHL=zeros(1,layers);
atten=zeros(1,layers);
ydot=zeros(layers,1);

GPP=zeros(1,layers);

% Define state variables
for i=1:layers
    Nut(i) = y(9*(i-1)+1); %
    Phy(i) = y(9*(i-1)+2); %
    DOP(i) = y(9*(i-1)+3); %
    DOC(i) = y(9*(i-1)+4); %
    POP(i) = y(9*(i-1)+5); %
    POC(i) = y(9*(i-1)+6); %
    CHL(i) = y(9*(i-1)+7); %
    GPP(i) = y(9*(i-1)+9); %
    Zoo(i) = y(9*(i-1)+8); %
end

% Define biochemical parameters
k0 = theta(1); % attenuation coefficient water (m-1)
kc = theta(2); % attenuation coefficient chl (m2 (mg chl)-1)
phygrow = theta(3); % max growth rate for phy1 (d-1)
mort = theta(4); % phy1 mortality rate (d-1)
ilim = theta(5); % phy1 light 1/2 sat. (W/m2)
nlim = theta(6); % phy1 po4 1/2 sat. (uM)
reminp = theta(7); % remin rate of DOP/POP (d-1)
reminc = theta(8); % remin rate of DOC/POC (d-1)
dopfracm1 = theta(9); % fraction of mortality to DOP/POP
dopfracmc1 = theta(10); % fraction of mortality to DOP/POF

```

```

Rcp = theta(11);          % ratio C to P in phy1
% minimum populations
phymin = theta(12);      % seed population of phy (uM)
% make sure wn_sink*dt<delz
wp_sink = theta(13); % sinking rate of POP (m/d)
wc_sink = theta(14); % sinking rate of POC (m/d)
% for chl
chlpmax = theta(15);
chlpmin = theta(16);
istar = theta(17);      % uEin/m2/d
% for temperature function of growth rate
temp_coeff_base = theta(18);      % unitless, Eppley
temp_coeff_norm = theta(19);
% Zooplankton parameters
zoograz = theta(20); % max zoo grazing rate (d^-1)
mortz = theta(21); % zoo mortality rate (d^-1)
plim = theta(22); % zoo phy 1/2 sat (uM)
gampn = theta(23); % zoo phy1 assimilation coeff
dopfracmn = theta(24);
dopfracgn = theta(25); % fraction of grazing to DOP/POP
dopfracmc = theta(26);
dopfracgc = theta(27); % fraction of grazing to DOP/POF
zoomin = theta(28); % egg population of zoo (uM)
grazlim = theta(29);

pym(1:4) = max((Phy(1:4)-phymin),0);
pym(5:layers) = Phy(5:layers);
zom(1:4) = max((Zoo(1:4)-zoomin),0);
zom(5:layers) = Zoo(5:layers);

Nutup(1) = 0;
Phyup(1) = 0;
Zooup(1) = 0;
DOPup(1) = 0;
DOCup(1) = 0;
POPup(1) = 0;
POCup(1) = 0;
Nutup(2:layers) = Nut(1:layers-1);
Phyup(2:layers) = Phy(1:layers-1);
Zooup(2:layers) = Zoo(1:layers-1);
DOPup(2:layers) = DOP(1:layers-1);
DOCup(2:layers) = DOC(1:layers-1);
POPup(2:layers) = POP(1:layers-1);
POCup(2:layers) = POC(1:layers-1);
zup(1) = 0;
zup(2:layers) = zmid(1:layers-1);

Nutdown(layers) = 0;
Phydown(layers) = 0;
Zoodown(layers) = 0;
DOPdown(layers) = 0;
DOCdown(layers) = 0;
POPdown(layers) = 0;
POCdown(layers) = 0;
Nutdown(1:layers-1) = Nut(2:layers);
Phydown(1:layers-1) = Phy(2:layers);
Zoodown(1:layers-1) = Zoo(2:layers);

```

```

DOPdown(1:layers-1) = DOP(2:layers);
DOCdown(1:layers-1) = DOC(2:layers);
POPdown(1:layers-1) = POP(2:layers);
POCdown(1:layers-1) = POC(2:layers);
zdown(layers) = 0;
zdown(1:layers-1) = zmid(2:layers);
Ddown(layers) = 0;
Ddown(1:layers-1) = Dvert(2:layers);

% % Photosynthetically active radiation
PAR0 = SWin* 0.45; % Papaioannou et al. 2004 = 0.473
atten(1)=(k0+kc*CHL(1))*zmid(1);
for i=2:layers
    atten(i) = atten(i-1)+(k0+kc*CHL(i-1))*zheight(i-1)/2+...
        (k0+kc*CHL(i))*zheight(i)/2;
end
PAR=PAR0*exp(-atten);

% Phytoplankton growth rates
mu = phygrow.*(PAR./(ilim+PAR)).*(Nut./(nlim+Nut));
tempfunc=1/temp_coeff_norm*temp_coeff_base.^Temp;
mu = mu.*tempfunc;

reminp_new = reminp.*tempfunc;
reminc_new = reminc.*tempfunc;

zp = max(Phy-grazlim,0);
graz = zoograz.*zp./(zp+plim);

%Calculate derivatives...
dotNut = -mu.*Phy+reminp_new.*DOP+reminp_new.*POP;
dotPhy = mu.*Phy-mort.*pym-graz.*Zoo;
dotZoo = gampn.*graz.*Zoo - mortz.*zom;
dotDOP = dopfracm1.*mort.*pym-reminp_new.*DOP+dopfracmn.*mortz.*zom...
    +dopfracgn.*((1-gampn).*graz.*Zoo);
dotDOC = Rcp.*(dopfracm1.*mort.*pym+dopfracgc.*(1-gampn).*graz.*Zoo)...
    +dopfracmc.*mortz.*zom.*Rcp- reminc_new.*DOC;
dotPOP = (1-dopfracm1).*mort.*pym - reminp_new.*POP+(1-dopfracmn).*mortz.*zom...
    +(1-dopfracgn).*(1-gampn).*graz.*Zoo;
dotPOC = (1-dopfracm1).*Rcp.*mort.*pym - reminc_new.*POC+(1-dopfracmc).*mortz.*zom.*Rcp...
    +Rcp.*(1-dopfracgc).*(1-gampn).*graz.*Zoo;

% not GPP but sp. growth rate.
dotGPP = mu.*Phy.*Rcp;

% Need to add REMINERALIZATION of bottom material

%Sinking:
dotPOP(1:layers-1) = dotPOP(1:layers-1) + wp_sink.*(POPup(1:layers-1)./(zmid(1:layers-1)'-zup
(1:layers-1))-POP(1:layers-1)./(zdown(1:layers-1)-zmid(1:layers-1)'));
dotPOP(layers) = dotPOP(layers) + wp_sink.*POPup(layers)./(zmid(layers)'-zup(layers));
dotPOC(1:layers-1) = dotPOC(1:layers-1) + wc_sink.*(POCup(1:layers-1)./(zmid(1:layers-1)'-zup
(1:layers-1))-POC(1:layers-1)./(zdown(1:layers-1)-zmid(1:layers-1)'));
dotPOC(layers) = dotPOC(layers) + wc_sink.*POCup(layers)./(zmid(layers)'-zup(layers));

%Diffusion:
dotNut = dotNut + (Dvert.*xyarea./(zmid'-zup).*(Nutup-Nut))./(xyarea.*(zmid'-zup))+

```

```

(Ddown.*xyarea./(zdown-zmid').*(Nutdown-Nut))./(xyarea.*(zdown-zmid'));
dotPhy = dotPhy + (Dvert.*xyarea./(zmid'-zup).*(Phyup-Phy))./(xyarea.*(zmid'-zup))+
(Ddown.*xyarea./(zdown-zmid').*(Phydown-Phy))./(xyarea.*(zdown-zmid'));
dotZoo = dotZoo + (Dvert.*xyarea./(zmid'-zup).*(Zooup-Zoo))./(xyarea.*(zmid'-zup))+
(Ddown.*xyarea./(zdown-zmid').*(Zoodown-Zoo))./(xyarea.*(zdown-zmid'));
dotDOP = dotDOP + (Dvert.*xyarea./(zmid'-zup).*(DOPup-DOP))./(xyarea.*(zmid'-zup))+
(Ddown.*xyarea./(zdown-zmid').*(DOPdown-DOP))./(xyarea.*(zdown-zmid'));
dotDOC = dotDOC + (Dvert.*xyarea./(zmid'-zup).*(DOCup-DOC))./(xyarea.*(zmid'-zup))+
(Ddown.*xyarea./(zdown-zmid').*(DOCdown-DOC))./(xyarea.*(zdown-zmid'));
dotPOP = dotPOP + (Dvert.*xyarea./(zmid'-zup).*(POPup-POP))./(xyarea.*(zmid'-zup))+
(Ddown.*xyarea./(zdown-zmid').*(POPdown-POP))./(xyarea.*(zdown-zmid'));
dotPOC = dotPOC + (Dvert.*xyarea./(zmid'-zup).*(POCup-POC))./(xyarea.*(zmid'-zup))+
(Ddown.*xyarea./(zdown-zmid').*(POCdown-POC))./(xyarea.*(zdown-zmid'));

% Chlorophyll
t1 = chlpmx-(chlpmx-chlpmn).*min(PAR./istar,1);
t3 = max(chlpmn, t1);
cpratio1 = min(chlpmx,t3);
newCHL = cpratio1.*(Phy+dotPhy);
dotCHL = newCHL - CHL;

for i=1:layers
    ydot((i-1)*9+1,1) = dotNut(i);
    ydot((i-1)*9+2,1) = dotPhy(i);
    ydot((i-1)*9+3,1) = dotDOP(i);
    ydot((i-1)*9+4,1) = dotDOC(i);
    ydot((i-1)*9+5,1) = dotPOP(i);
    ydot((i-1)*9+6,1) = dotPOC(i);
    ydot((i-1)*9+7,1) = dotCHL(i);
    ydot((i-1)*9+9,1) = dotGPP(i);
    ydot((i-1)*9+8,1) = dotZoo(i);
end

```

```

function ydot = BGCSys2(t,y,theta,xdata)
% ode system function for the extended simplified model (ch.4)
%
% note that parameters estimated from literature are hard-coded,
% and therefore not included in the 'theta' vector.
%

SWin = xdata.x(ceil(t),2); % daily incident sw rad ( $W m^{-2}$ )
layers = size(xdata.layers,1); % number of vert. layers in GCM output
Temp = xdata.x(ceil(t),3:layers+2); % surface temp, deg. C
Vvert = xdata.x(ceil(t),layers+3:2*layers+2); % vertical water movement (m/s)
Vvert = Vvert.*86400;
Dvert = xdata.x(ceil(t),2*layers+3:3*layers+2); % turb. diffusion ( $m^2/s$ )
Dvert = Dvert.*86400;
zmid = xdata.layers(:,1).*-1; % mid-layer depth (m)
zheight = xdata.layers(:,2); % layer thickness (m)
xyarea = xdata.area; % cell area ( $m^2$ )
waterdens = xdata.density(ceil(t),:);

% Preallocate arrays for speed:
NPP=zeros(1,layers);
atten=zeros(1,layers);
ydot=zeros(layers,1);

GPP=zeros(1,layers);
CHL=zeros(1,layers);
AIC=zeros(1,layers);
DOC=zeros(1,layers);
POC=zeros(1,layers);
Resp=zeros(1,layers);

% Define state variables
for i=1:layers
    GPP(i) = y(6*(i-1)+1); % No data.
    AIC(i) = y(6*(i-1)+2);
    CHL(i) = y(6*(i-1)+3);
    DOC(i) = y(6*(i-1)+4);
    POC(i) = y(6*(i-1)+5);
    Resp(i) = y(6*(i-1)+6);
end

% Define biochemical parameters
k0 = 0.15; % attenuation coefficient water ( $m^{-1}$ )
kc = 0.0149; % attenuation coefficient chl ( $m^2 (mg chl)^{-1}$ )
C = theta(1);
Ea = theta(2);
Popt = theta(3);
a = theta(4);
mort = theta(5);
Qa = theta(6);
ChlCmin = theta(7);
ChlCmax = theta(8);
Istar = theta(9);
dens = theta(10);
diam = theta(11);

night = 0.77;

```

```

fres = 0.75;%1.35;
fexc = 0.13;%0.35;
fdom = 0.14;
fpom = 0.50;%0.57;
reminD = 0.01;
Qd = 1.06;
reminP = 0.01;
Qp = 1.06;
vp = 0.2;

AICup(1) = 0;
P04up(1) = 0;
AIPup(1) = 0;
DOCup(1) = 0;
POCup(1) = 0;
CHLup(1) = 0;

AICup(2:layers) = AIC(1:layers-1);
P04up(2:layers) = P04(1:layers-1);
AIPup(2:layers) = AIP(1:layers-1);
DOCup(2:layers) = DOC(1:layers-1);
POCup(2:layers) = POC(1:layers-1);
CHLup(2:layers) = CHL(1:layers-1);

zup(1) = 0;
zup(2:layers) = zmid(1:layers-1);

AICdown(layers) = 0;
P04down(layers) = 0;
AIPdown(layers) = 0;
DOCdown(layers) = 0;
POCdown(layers) = 0;
CHLdown(layers) = 0;

AICdown(1:layers-1) = AIC(2:layers);
P04down(1:layers-1) = P04(2:layers);
AIPdown(1:layers-1) = AIP(2:layers);
DOCdown(1:layers-1) = DOC(2:layers);
POCdown(1:layers-1) = POC(2:layers);
CHLdown(1:layers-1) = CHL(2:layers);

zdown(layers) = 0;
zdown(1:layers-1) = zmid(2:layers);
Ddown(layers) = 0;
Ddown(1:layers-1) = Dvert(2:layers);

%% Photosynthetically active radiation
PAR0 = SWin* 0.47; % Papaioannou et al. 2004 = 0.473
atten(1)=(k0+kc*CHL(1))*zmid(1);
for i=2:layers
    atten(i) = atten(i-1)+(k0+kc*CHL(i-1))*zheight(i-1)/2+...
        (k0+kc*CHL(i))*zheight(i)/2;
end
PAR=PAR0*exp(-atten);

% Productivity, umol C per L per day (following Sterner 2010)
P = C.*exp(-Ea./((0.0000862.*(Temp+273.15)))).*Popt.*(1-exp(-a*PAR/Popt))./12.*night;

```



```

% Gross productivity
G = (1+fres+fexc).*P;

m = mort.*Qa.^Temp.*AIC;

%Calculate chl concentration based on last time step:
dotCHL = (ChlCmax-(ChlCmax-ChlCmin)*min(PAR./Istar,1)).*AIC - CHL;

%Calculate derivatives...
dotGPP = G - GPP;
dotAIC = P - m; %G - m - P.*fres - P.*fexc;
dotDOC = P.*fexc + m.*fdom - reminD.*DOC.*Qd.^Temp;
dotPOC = m.*fpom - reminP.*POC.*Qp.^Temp;
dotResp = reminD.*DOC.*Qd.^Temp + reminP.*POC.*Qp.^Temp + fres.*P + (1-fdom-fpom).*m - Resp;

wc_sink = 0.033634.*(dens-waterdens).*diam.^2;

% %Sinking:
dotAIC(1:layers-1) = dotAIC(1:layers-1) + wc_sink(1:layers-1).*(AICup(1:layers-1)./(zmid(1:layers-1)'-zup(1:layers-1))-AIC(1:layers-1)./(zdown(1:layers-1)-zmid(1:layers-1)'));
dotAIC(layers) = dotAIC(layers) + wc_sink(layers).*(AICup(layers)./(zmid(layers)'-zup(layers)));
dotCHL(1:layers-1) = dotCHL(1:layers-1) + wc_sink(1:layers-1).*(CHLup(1:layers-1)./(zmid(1:layers-1)'-zup(1:layers-1))-CHL(1:layers-1)./(zdown(1:layers-1)-zmid(1:layers-1)'));
dotCHL(layers) = dotCHL(layers) + wc_sink(layers).*(CHLup(layers)./(zmid(layers)'-zup(layers)));

dotPOC(1:layers-1) = dotPOC(1:layers-1) + vp.*(POCup(1:layers-1)./(zmid(1:layers-1)'-zup(1:layers-1))-POC(1:layers-1)./(zdown(1:layers-1)-zmid(1:layers-1)'));
dotPOC(layers) = dotPOC(layers) + vp.*POCup(layers)./(zmid(layers)'-zup(layers)));

% %Diffusion:
dotAIC = dotAIC + (Dvert.*xyarea./(zmid'-zup).*(AICup-AIC))./(xyarea.*(zmid'-zup))+
(Ddown.*xyarea./(zdown-zmid').*(AICdown-AIC))./(xyarea.*(zdown-zmid'));
dotCHL = dotCHL + (Dvert.*xyarea./(zmid'-zup).*(CHLup-CHL))./(xyarea.*(zmid'-zup))+
(Ddown.*xyarea./(zdown-zmid').*(CHLdown-CHL))./(xyarea.*(zdown-zmid'));
dotDOC = dotDOC + (Dvert.*xyarea./(zmid'-zup).*(DOCup-DOC))./(xyarea.*(zmid'-zup))+
(Ddown.*xyarea./(zdown-zmid').*(DOCdown-DOC))./(xyarea.*(zdown-zmid'));
dotPOC = dotPOC + (Dvert.*xyarea./(zmid'-zup).*(POCup-POC))./(xyarea.*(zmid'-zup))+
(Ddown.*xyarea./(zdown-zmid').*(POCdown-POC))./(xyarea.*(zdown-zmid'));

for i=1:layers
    ydot((i-1)*6+1,1) = dotGPP(i);
    ydot((i-1)*6+2,1) = dotAIC(i);
    ydot((i-1)*6+3,1) = dotCHL(i);
    ydot((i-1)*6+4,1) = dotDOC(i);
    ydot((i-1)*6+5,1) = dotPOC(i);
    ydot((i-1)*6+6,1) = dotResp(i);
end

```

```

function [ss ymodel] = BGCSs(theta,data)
% biogeochemical sum of squares function for CyCLEs ecological model clone.
% created 5/27/09 by Cory McDonald
% updated for 0-D data 6/25/09

timeCHL = data.ydata.chl(:,1);
timeTDP = data.ydata.tdp(:,1);
timeDOC = data.ydata.doc(:,1);
timePOP = data.ydata.pop(:,1);

timeNPP = data.ydata.npp(:,1);

dataCHL = data.ydata.chl(:,2:end); % ug/L
dataTDP = data.ydata.tdp(:,2:end); % uM
dataDOC = data.ydata.doc(:,2:end); % uM
dataOxy = data.ydata.oxy(:,2:end); % mg/L
dataPOP = data.ydata.pop(:,2:end); % uM

dataNPP = data.ydata.npp(:,2:end); % umol C/L/d

xdata = data.xdata;
layers = size(xdata.layers,1);
height = xdata.layers(:,2);
bottom = xdata.layers(:,3);

% 7 last parameters are the initial states
y0 = theta(end-6:end);
y0(end+1) = 0;
% replicating for a uniform distribution through water column
y0 = repmat(y0,1,layers);

% estimated parameter for allocthonous DOC
allocDOC = theta(20);

%Preallocate arrays for speed in 'for' loops:
dataCHLgrid = zeros(size(dataCHL,1),layers);
dataTDPgrid = zeros(size(dataTDP,1),layers);
dataDOCgrid = zeros(size(dataDOC,1),layers);
dataPOPgrid = zeros(size(dataPOP,1),layers);
dataNPPgrid = zeros(size(dataNPP,1),layers);

ymodel = BGCFun(data.xdata.x(:,1),theta,y0,xdata);

modelCHL = ymodel(timeCHL,7:8:end); % in ug/L!!
modelTDP = (ymodel(timeTDP,1:8:end)+ymodel(timeTDP,3:8:end));
modelDOC = ymodel(timeDOC,4:8:end)+allocDOC;
modelPOP = (ymodel(timePOP,2:8:end)+ymodel(timePOP,5:8:end));

temp = ymodel(timeNPP,8:8:end); %cumulative C fixation
temp2 = ymodel(timeNPP-1,8:8:end); % @ prev. day

modelNPP=temp-temp2;

% Add NaN's to end of data if shorter than model output:
if size(dataCHL,2) < floor(bottom(end))
    dataCHL(:,size(dataCHL,2)+1:floor(bottom(end))+1)=NaN;

```

```

end
if size(dataTDP,2) < floor(bottom(end))
    dataTDP(:,size(dataTDP,2)+1:floor(bottom(end))+1)=NaN;
end
if size(dataDOC,2) < floor(bottom(end))
    dataDOC(:,size(dataDOC,2)+1:floor(bottom(end))+1)=NaN;
end
if size(dataPOP,2) < floor(bottom(end))
    dataPOP(:,size(dataPOP,2)+1:floor(bottom(end))+1)=NaN;
end
if size(dataNPP,2) < floor(bottom(end))
    dataNPP(:,size(dataNPP,2)+1:floor(bottom(end))+1)=NaN;
end

% Bin data into model cells:
% USING UNWEIGHTED AVERAGE OF ALL POINTS IN GRID CELL (INCL. EDGES)
for i=1:layers
    low = ceil(bottom(i) - height(i)+1);
    high = floor(bottom(i)+1);
    dataCHLgrid(:,i) = nanmean(dataCHL(:,low:high),2);
    dataTDPgrid(:,i) = nanmean(dataTDP(:,low:high),2);
    dataDOCgrid(:,i) = nanmean(dataDOC(:,low:high),2);
    dataPOPgrid(:,i) = nanmean(dataPOP(:,low:high),2);
    dataNPPgrid(:,i) = nanmean(dataNPP(:,low:high),2);
end

%% New objective function (all data combined) 5/25

dataCHL=reshape(dataCHLgrid,size(dataCHLgrid,1)*size(dataCHLgrid,2),1);
dataTDP=reshape(dataTDPgrid,size(dataTDPgrid,1)*size(dataTDPgrid,2),1);
dataDOC=reshape(dataDOCgrid,size(dataDOCgrid,1)*size(dataDOCgrid,2),1);
dataPOP=reshape(dataPOPgrid,size(dataPOPgrid,1)*size(dataPOPgrid,2),1);
dataNPP=reshape(dataNPPgrid,size(dataNPPgrid,1)*size(dataNPPgrid,2),1);

modelCHL=reshape(modelCHL,size(modelCHL,1)*size(modelCHL,2),1);
modelTDP=reshape(modelTDP,size(modelTDP,1)*size(modelTDP,2),1);
modelDOC=reshape(modelDOC,size(modelDOC,1)*size(modelDOC,2),1);
modelPOP=reshape(modelPOP,size(modelPOP,1)*size(modelPOP,2),1);
modelNPP=reshape(modelNPP,size(modelNPP,1)*size(modelNPP,2),1);

CHLvector = horzcat(modelCHL,dataCHL);
CHLvector(any(isnan(CHLvector),2),:) = [];
CHLn = size(CHLvector,1);
CHLy = CHLvector(:,1)-CHLvector(:,2);
iqrCHL = iqr(CHLvector(:,2));
TDPvector = horzcat(modelTDP,dataTDP);
TDPvector(any(isnan(TDPvector),2),:) = [];
TDPn = size(TDPvector,1);
TDPy = TDPvector(:,1)-TDPvector(:,2);
iqrTDP = iqr(TDPvector(:,2));
DOCvector = horzcat(modelDOC,dataDOC);
DOCvector(any(isnan(DOCvector),2),:) = [];
DOCn = size(DOCvector,1);
DOCy = DOCvector(:,1)-DOCvector(:,2);
iqrDOC = iqr(DOCvector(:,2));
POPvector = horzcat(modelPOP,dataPOP);

```

```

POPvector(any(isnan(POPvector),2),:) = [];
POPn = size(POPvector,1);
POPy = POPvector(:,1)-POPvector(:,2);
iqrPOP = iqr(POPvector(:,2));
NPPvector = horzcat(modelNPP,dataNPP);
NPPvector(any(isnan(NPPvector),2),:) = [];
NPPn = size(NPPvector,1);
NPPy = NPPvector(:,1)-NPPvector(:,2);
iqrNPP = iqr(NPPvector(:,2));

CHLnmse = sqrt(sum(CHLy.^2)/CHLn)/iqrCHL
TDPnmse = sqrt(sum(TDPy.^2)/TDPn)/iqrTDP
DOCnmse = sqrt(sum(DOCy.^2)/DOCn)/iqrDOC
POPnmse = sqrt(sum(POPy.^2)/POPn)/iqrPOP
NPPnmse = sqrt(sum(NPPy.^2)/NPPn)/iqrNPP
%
ss = (CHLnmse+TDPnmse+DOCnmse+POPnmse+3*NPPnmse)

```

```

function [ss ymodel] = BGCss2(x,theta,data)
% biogeochemical sum of squares function for extended simplified model.
%

timeCHL = data.ydata.chl(:,1);
timeDOC = data.ydata.doc(:,1);

timeNPP = data.ydata.npp(:,1);

dataCHL = data.ydata.chl(:,2:end); % ug/L
dataDOC = data.ydata.doc(:,2:end); % uM

dataNPP = data.ydata.npp(:,2:end); % umol C/L/d

xdata = data.xdata;
layers = size(xdata.layers,1);
height = xdata.layers(:,2);
bottom = xdata.layers(:,3);

y0 = [0 2 0.5 0 0 0];
y0 = repmat(y0,1,layers);

%Preallocate arrays for speed in 'for' loops:
dataCHLgrid = zeros(size(dataCHL,1),layers);
dataDOCgrid = zeros(size(dataDOC,1),layers);
dataNPPgrid = zeros(size(dataNPP,1),layers);

ymodel = BGCfun2(data.xdata.x(:,1),theta,y0,xdata);

modelCHL = ymodel(timeCHL,3:6:end); % in ug/L!!
modelDOC = ymodel(timeDOC,4:6:end)+x;
modelNPP = ymodel(timeNPP,1:6:end);

% Add NaN's to end of data if shorter than model output:
if size(dataCHL,2) < floor(bottom(end))
    dataCHL(:,size(dataCHL,2)+1:floor(bottom(end))+1)=NaN;
end
if size(dataDOC,2) < floor(bottom(end))
    dataDOC(:,size(dataDOC,2)+1:floor(bottom(end))+1)=NaN;
end
if size(dataNPP,2) < floor(bottom(end))
    dataNPP(:,size(dataNPP,2)+1:floor(bottom(end))+1)=NaN;
end

% Bin data into model cells:
% USING UNWEIGHTED AVERAGE OF ALL POINTS IN GRID CELL (INCL. EDGES)
for i=1:layers
    low = ceil(bottom(i) - height(i)+1);
    high = floor(bottom(i)+1);
    dataCHLgrid(:,i) = nanmean(dataCHL(:,low:high),2);
    dataDOCgrid(:,i) = nanmean(dataDOC(:,low:high),2);
    dataNPPgrid(:,i) = nanmean(dataNPP(:,low:high),2);
end

%% New objective function (all data combined) 5/25

```

```

dataCHL=reshape(dataCHLgrid,size(dataCHLgrid,1)*size(dataCHLgrid,2),1);
dataDOC=reshape(dataDOCgrid,size(dataDOCgrid,1)*size(dataDOCgrid,2),1);
dataNPP=reshape(dataNPPgrid,size(dataNPPgrid,1)*size(dataNPPgrid,2),1);

modelCHL=reshape(modelCHL,size(modelCHL,1)*size(modelCHL,2),1);
modelDOC=reshape(modelDOC,size(modelDOC,1)*size(modelDOC,2),1);
modelNPP=reshape(modelNPP,size(modelNPP,1)*size(modelNPP,2),1);

```

```

CHLvector = horzcat(modelCHL,dataCHL);
CHLvector(any(isnan(CHLvector),2),:) = [];
CHLn = size(CHLvector,1);
CHLy = CHLvector(:,1)-CHLvector(:,2);
iqrCHL = iqr(CHLvector(:,2));
DOCvector = horzcat(modelDOC,dataDOC);
DOCvector(any(isnan(DOCvector),2),:) = [];
DOCn = size(DOCvector,1);
DOCy = DOCvector(:,1)-DOCvector(:,2);
iqrDOC = iqr(DOCvector(:,2));
NPPvector = horzcat(modelNPP,dataNPP);
NPPvector(any(isnan(NPPvector),2),:) = [];
NPPn = size(NPPvector,1);
NPPy = NPPvector(:,1)-NPPvector(:,2);
iqrNPP = iqr(NPPvector(:,2));

```

```

CHLnrms = sqrt(sum(CHLy.^2)/CHLn)/iqrCHL
DOCnrms = sqrt(sum(DOCy.^2)/DOCn)/iqrDOC
NPPnrms = sqrt(sum(NPPy.^2)/NPPn)/iqrNPP

```

```

% The output of 'ss' can be defined as one of the above parameters, or as a
% weighted sum of a combination of parameters:
ss= CHLnrms;

```



## **Appendix G**

### **Lake Superior nitrogen Simulink model**





## **Appendix H**

### **Lake Superior nitrogen data**

Table H.1: Lake Superior total nitrogen model inputs: riverine total nitrogen loading and atmospheric nitrate deposition. Atmospheric deposition of  $\text{NH}_4$  and organic N is held constant in the model (see Chapter 5).

Year	River total N ( $\text{mol yr}^{-1}$ )	Atmospheric $\text{NO}_3^-$ ( $\text{mol yr}^{-1}$ )
1900	1.470E+09	1.770E+08
1908	2.240E+09	2.474E+08
1909	1.871E+09	2.562E+08
1910	7.426E+08	2.650E+08
1911	1.309E+09	2.738E+08
1912	1.187E+09	2.826E+08
1913	1.731E+09	2.914E+08
1914	1.758E+09	3.002E+08
1915	1.511E+09	3.090E+08
1916	2.413E+09	3.178E+08
1917	1.004E+09	3.266E+08
1918	1.036E+09	3.354E+08
1919	1.470E+09	3.442E+08
1920	2.027E+09	3.530E+08
1921	1.468E+09	3.648E+08
1922	1.535E+09	3.766E+08
1923	1.173E+09	3.884E+08
1924	1.266E+09	4.002E+08
1925	1.450E+09	4.120E+08
1926	1.743E+09	4.238E+08
1927	2.430E+09	4.356E+08
1928	2.852E+09	4.474E+08
1929	1.436E+09	4.592E+08
1930	1.752E+09	4.710E+08
1931	1.300E+09	4.828E+08
1932	1.735E+09	4.946E+08
1933	1.593E+09	5.064E+08
1934	1.902E+09	5.182E+08
1935	2.209E+09	5.300E+08
1936	2.010E+09	5.418E+08
1937	1.979E+09	5.536E+08
1938	2.375E+09	5.654E+08
1939	2.313E+09	5.772E+08
1940	1.745E+09	5.890E+08
1941	2.494E+09	6.008E+08
1942	2.438E+09	6.126E+08

Table H.1: (continued)

Year	River total N (mol yr <sup>-1</sup> )	Atmospheric NO <sub>3</sub> <sup>-</sup> (mol yr <sup>-1</sup> )
1943	2.489E+09	6.244E+08
1944	2.603E+09	6.362E+08
1945	2.487E+09	6.480E+08
1946	2.429E+09	6.598E+08
1947	2.632E+09	6.716E+08
1948	1.824E+09	6.834E+08
1949	1.956E+09	6.952E+08
1950	3.201E+09	7.070E+08
1951	3.220E+09	7.342E+08
1952	2.613E+09	7.613E+08
1953	2.716E+09	7.885E+08
1954	2.983E+09	8.156E+08
1955	2.271E+09	8.428E+08
1956	2.268E+09	8.699E+08
1957	2.329E+09	8.971E+08
1958	2.183E+09	9.242E+08
1959	2.511E+09	9.514E+08
1960	2.516E+09	9.785E+08
1961	2.322E+09	1.006E+09
1962	2.122E+09	1.033E+09
1963	2.091E+09	1.060E+09
1964	2.914E+09	1.180E+09
1965	2.841E+09	1.300E+09
1966	3.056E+09	1.420E+09
1967	2.736E+09	1.540E+09
1968	3.329E+09	1.660E+09
1969	3.046E+09	1.780E+09
1970	2.917E+09	1.900E+09
1971	3.185E+09	1.887E+09
1972	2.919E+09	1.875E+09
1973	2.604E+09	1.862E+09
1974	2.754E+09	1.849E+09
1975	2.471E+09	1.836E+09
1976	2.185E+09	1.824E+09
1977	2.321E+09	1.811E+09
1978	2.273E+09	1.798E+09
1979	3.110E+09	1.785E+09
1980	2.343E+09	1.773E+09
1981	2.254E+09	1.760E+09

Table H.1: (continued)

Year	River total N (mol yr <sup>-1</sup> )	Atmospheric NO <sub>3</sub> <sup>-</sup> (mol yr <sup>-1</sup> )
1982	2.639E+09	1.747E+09
1983	2.653E+09	1.734E+09
1984	2.648E+09	1.722E+09
1985	3.017E+09	1.709E+09
1986	2.695E+09	1.696E+09
1987	1.607E+09	1.683E+09
1988	2.310E+09	1.671E+09
1989	2.517E+09	1.658E+09
1990	2.232E+09	1.645E+09
1991	2.386E+09	1.632E+09
1992	2.813E+09	1.620E+09
1993	2.755E+09	1.607E+09
1994	2.322E+09	1.594E+09
1995	2.150E+09	1.581E+09
1996	3.239E+09	1.569E+09
1997	2.351E+09	1.556E+09
1998	1.539E+09	1.543E+09
1999	2.552E+09	1.530E+09
2000	2.002E+09	1.518E+09
2001	2.470E+09	1.505E+09
2002	2.416E+09	1.492E+09
2003	2.104E+09	1.479E+09
2004	2.498E+09	1.467E+09
2005	2.094E+09	1.454E+09
2010		1.390E+09

Table H.2: Lake Superior nitrate concentrations used for model calibration (Chapter 5), listed by source.

Year	$\text{NO}_3^-$ ( $\mu\text{mol}$ )
Weiler, 1978	
1906	5.4
1940	10.4
1948	13.9
1959	16.9
1970	19.1
1971	19.4
1968	19.4
1969	19.7
1973	22
1976	22.2
Environment Canada	
1968.63	17.8
1971.46	17.9
1971.54	18.1
1970.88	18.6
1973.71	19.3
1970.79	19.4
1973.54	19.4
1973.63	19.4
1973.79	19.7
1973.46	20.5
1973.88	20.7
1983.79	22.1
1976.46	22.1
1983.71	22.9
1987.38	23
1973.38	23
1983.54	23.3
1983.46	23.7
1986.38	23.9
1983.38	24
1990.38	24.1
1989.38	24.3
1985.38	24.5
1992.38	25.2
1991.38	25.4
KITES	

Table H.2: (continued)

Year	$\text{NO}_3^-$ ( $\mu\text{mol}$ )
1998.38	24.3
1998.54	22.8
1999.38	24.5
1999.46	24
1999.54	23.4
1999.63	23.3
1999.79	21.2
2000.29	25.2
2000.38	22.9
2000.46	23.9
2000.54	22.8
2000.63	23.3
2000.71	19.8
2000.79	23.1
EPA	
1992.38	24
1993.38	25.8
1993.63	22.9
1996.38	24.5
1996.63	21.4
1997.38	23.1
1997.63	21.6
1998.38	22.7
1998.63	22.1
1999.38	21.6
1999.63	22.4
2000.38	22.7
2000.63	21.2
2001.38	23.1
2001.63	22.5
2002.38	23.5
2002.63	23.6
2003.38	24.3
2003.63	23.1
2004.38	24.54
2004.63	22.31
2005.38	23.74
2005.63	20.37
2006.38	25

Table H.2: (continued)

Year	$\text{NO}_3^-$ ( $\mu\text{mol}$ )
2006.63	21.65





# **Appendix I**

## **Elsevier copyright policy**

The following is an excerpt of the Elsevier copyright policy, accessed online (1:00 pm on October 29, 2010) at <http://www.elsevier.com/wps/find/authorsview.authors/copyright#whatrights>. This policy clearly states that journal authors retain the right to include published articles in a thesis or dissertation (item 8, emphasis added).

[...]

### **What rights do I retain as a journal author\*?**

As a journal author, you retain rights for a large number of author uses, including use by your employing institute or company. These rights are retained and permitted without the need to obtain specific permission from Elsevier. These include:

- the right to make copies (print or electric) of the journal article for their own personal use, including for their own classroom teaching use;
- the right to make copies and distribute copies (including via e-mail) of the journal article to research colleagues, for personal use by such colleagues (but not for Commercial Purposes\*\*, as listed below);
- the right to post a pre-print version of the journal article on Internet web sites including electronic pre-print servers, and to retain indefinitely such version on such servers or sites (see also our information on electronic preprints for a more detailed discussion on these points);
- the right to post a revised personal version of the text of the final journal article (to reflect changes made in the peer review process) on the author's personal or institutional web site or server, incorporating the complete citation and with a link to the Digital Object Identifier (DOI) of the article;
- the right to present the journal article at a meeting or conference and to distribute copies of such paper or article to the delegates attending the meeting;
- for the author's employer, if the journal article is a 'work for hire', made within the scope of the author's employment, the right to use all or part of the information in (any version of) the journal article for other intra-company use (e.g. training), including by posting the article on secure, internal corporate intranets;
- patent and trademark rights and rights to any process or procedure described in the journal article;
- *the right to include the journal article, in full or in part, in a thesis or dissertation;*
- the right to use the journal article or any part thereof in a printed compilation of works of the author, such as collected writings or lecture notes (subsequent to publication of the article in the journal); and
- the right to prepare other derivative works, to extend the journal article into book-length form, or to otherwise re-use portions or excerpts in other works, with full acknowledgement of its original publication in the journal.

[...]