2017

Using Water Quality Models in Management - A Multiple Model Assessment, Analysis of Confidence, and Evaluation of Climate Change Impacts

Isaac Irby
College of William and Mary - Virginia Institute of Marine Science, isaacirby@gmail.com

USING WATER QUALITY MODELS IN MANAGEMENT
–
A MULTIPLE MODEL ASSESSMENT, ANALYSIS OF CONFIDENCE, AND EVALUATION OF CLIMATE CHANGE IMPACTS

A Dissertation

Presented to

The Faculty of the School of Marine Science

The College of William and Mary in Virginia

In Partial Fulfillment

of the Requirements for the Degree of

Doctor of Philosophy

---

by

Isaac David Irby

August 2017

APPROVAL PAGE

This dissertation is submitted in partial fulfillment of

the requirements for the degree of

Doctor of Philosophy

Isaac David Irby

Approved by the Committee, July 2017

Marjorie A. M. Friedrichs, Ph.D.
Committee Chair / Advisor

Carl T. Friedrichs, Ph.D.

Carl H. Hershner, Ph.D.

Raleigh R. Hood, Ph.D.
University of Maryland Center for Environmental Sciences
Horn Point, MD, USA

Raymond G. Najjar, Ph.D.
Pennsylvania State University
State College, PA, USA

TABLE OF CONTENTS

Acknowledgements vi

List of tables vii

List of figures ix

Abstract xiii

1. INTRODUCTION 2
  1.1 Motivation 2
  1.2 The Chesapeake Bay TMDL 3
  1.3 The Role of Multiple Models 5
  1.4 Dissertation Motivating Questions and Structure 6
  Literature Cited 9

2. CHALLENGES ASSOCIATED WITH MODELING LOW-OXYGEN WATERS IN CHESAPEAKE BAY: A MULTIPLE MODEL COMPARISON 13
  Abstract 14
  2.1 Introduction 15
  2.2 Methods 17
    2.2.1 Participating Chesapeake Bay models 17
    2.2.2 Available Chesapeake Bay observations 19
    2.2.3 Calculation of stratification and mixed layer depth 20
    2.2.4 Model skill metrics 21
  2.3 Results 23
  2.4 Discussion 26
    2.4.1 How does the skill of various hydrodynamically-based DO models compare? 26
    2.4.2 How does model skill of DO compare to that of the primary drivers of DO variability? 29
    2.4.3 Why is it important for DO models to simulate the MLD_O correctly? 31
    2.4.4 How can DO simulations in the Bay be improved for management of water quality and living resources? 32
    2.4.5 What is the utility of the multi-model ensemble and Model Mean? 33
  2.5 Conclusions 34
  Tables 36
  Figures 40
  Acknowledgements 52
  Literature Cited 53

3. EVALUATING CONFIDENCE IN THE IMPACT OF REGULATORY NUTRIENT REDUCTION ON CHESAPEAKE BAY WATER QAULITY 63
  Abstract 64
  3.1 Introduction 65
  3.2 Modeling methodology 67
    3.2.1 Water quality models 68
    3.2.2 Nutrient forcing 69
    3.2.3 Model skill metrics 71

3.3 CBP procedure for assessing attainment of water quality standards 72
3.3.1 Designated uses 72
3.3.2 TMDL regulatory protocol for determining attainment of water quality standards 73
3.3.3 Study period (1991-2000) and critical period (1993-1995) 74
3.4 New methodology for assessing confidence in WQS attainment: the Confidence Index (CI) 75
3.5 Results 76
3.5.1 Model-data comparison of the standard run for both estuarine models 76
3.5.2 Comparison of the standard runs versus the TMDL-WIP scenarios 77
3.5.3 Comparison of water quality standard attainment for both models 78
3.5.4 Comparison of the impact of 3-year period 79
3.5.5 Examination of Confidence Index 80
3.6 Discussion 81
3.6.1 How do Chesapeake Bay models compare in terms of how nutrient reduction impact DO concentrations? 81
3.6.2 How do Chesapeake Bay models compare in terms of whether nutrient reductions will lead to the desired attainment of water quality? 82
3.6.3 Where is the location in the Bay with the greatest uncertainty in the impact of nutrient reduction on the desired attainment of water quality? 84
3.6.4 Within the modeling and assessment approach, what is the source of greatest uncertainty in the impact of nutrient reduction on the attainment of water quality? .. 85
3.7 Summary and Conclusions 87
Tables 89
Figures 94
Appendix 3A. Protocol for Assessing Attainment of Water Quality Standards 104
Acknowledgements 108
Literature Cited 109

4. THE COMPETING IMPACTS OF CLIMATE CHANGE AND NUTRIENT REDUCTION ON DISSOLVED OXYGEN IN CHESAPEAKE BAY 116
Abstract 117
4.1 Introduction 118
4.2 Methods 120
4.2.1 ChesROMS-ECB 120
4.2.2 Nutrient Scenarios 122
4.2.3 2050 Climate Change Scenarios 123
4.2.4 Dissolved Oxygen Analysis 126
4.3 Results 127
4.4 Discussion 131
4.4.1 How will climate change affect the impact of nutrient reduction on dissolved oxygen in the Chesapeake Bay? 131
4.4.2 How will hypoxia change as a result of climate change? 132
4.4.3 How might this impact the success of the TMDL? 134
4.4.4 How will climate change impact DO if the TMDL nutrient reductions are not met? 135
4.4.5 What are the limitations of this study and how can they be addressed in future work? 136
4.5 Conclusions 138

Tables 140
Figures 145
Appendix 4.A Modified ChesROMS-ECB Equations 154
Appendix 4.B Model skill assessment 155
Acknowledgements 167
Literature Cited 168

5. CONCLUSIONS: A LOOK AT THE RELATIONSHIP BETWEEN MULTIPLE MODELS AND THE CHESAPEAKE BAY TMDL 176
5.1 Introduction 177
5.2 Why is creating a TMDL for Chesapeake Bay so difficult? 179
5.3 Why should we use multiple models in a TMDL? 183
5.4 How are multiple models in the Chesapeake Bay currently being used? 186
5.5 How might multiple models be used for the Chesapeake Bay TMDL in the future? 188
5.6 Conclusions 189
Literature Cited 190

Vita 193

ACKNOWLEDGEMENTS

When I have been asked for advice on graduate school by undergraduates searching for the next path to take, I always start with what I see as the most important factor in ensuring success in not just research and studies but also in the mental and emotional challenge of completing a Ph.D.: find an advisor who advocates for their students and views their student's success as their own. Dr. Marjy Friedrichs is the embodiment of that ethos and I am eternally grateful for the time and energy she has spent over the last five years to ensure that I had the necessary support to thrive. During my time at VIMS I continually added to my degree program as I discovered what kind of scientist I wanted to be and how that would play out in the future. As I found my stride, Marjy was there with cautionary advice, but was always supportive and willing to take the journey with me. Not every advisor would so adeptly and willingly evolve along with their students.

I must also express thanks to my committee members for their invaluable insights and continual support. To Dr. Carl Friedrichs, I am so thankful to have had two Friedrichses from whom to derive insight and inspiration. I truly feel like part of the family. To Dr. Raleigh Hood, thank you for showing me what it means to be continually positive and engaged in other's work. To Dr. Ray Najjar, thank you for your willingness to give your time writing as many letters of support as possible. To Dr. Carl Hershner, thank you for continually pushing me to question the efficacy of my science in order to make it better.

There have also been many other individuals who have been instrumental in my research and professional development over the last five years that I would like to thank: Gary Shenk, Lew Linker, Dr. Richard Tian, Dr. Ping Wang, Kyle Hinson, and the rest of the Chesapeake Bay Program modeling group who were so willing to lend their time and energy to my research; Dr. John Gilmour for inspiring me to apply for an internship in the White House that changed the trajectory of my career; Dr. Marjory Blumenthal, Dr. Ashley Predith, and Dr. Knatokie Ford for mentoring me during my time interning with the President's Council of Advisors on Science and Technology; Dr. Linda Schaffner and Dr. Sarah Stafford for supporting me and helping me advocate for the value of the joint degree program; BioCOM lab members Dr. Danny Kaufman, Kelsey Fall, Julia Moriarity, Fei Da, and Dr. Pierre St-Laurent for your research help and friendships.

I am indebted to the amazing relationships I have fostered during my time at VIMS and William & Mary. The community at VIMS helps keep graduate school in perspective and is the institution's best asset. To my family, thank you for your love and support even all the way across the country. I am very excited to no longer be the family's "forever student." To Steven, I am so thankful that William & Mary brought us together. You have taught me what it means to be supportive and I thank you for helping me get to the finish.

LIST OF TABLES

Chapter 2. CHALLENGES ASSOCIATED WITH MODELING LOW-OXYGEN WATERS IN CHESAPEAKE BAY: A MULTIPLE MODEL COMPARISON

Table 2.1 Model Characteristics. 36

Table 2.2 Characteristics of observation stations. 37

Table 2.3 Mean and standard deviation of observations and total normalized RMSD for each model. 38

Table 2.4 Pycnocline and oxycline correlation statistics. 39

Chapter 3. EVALUATING CONFIDENCE IN THE IMPACT OF REGULATORY NUTRIENT REDUCTION ON CHESAPEAKE BAY WATER QUALITY

Table 3.1 Observation stations and segments. 89

Table 3.2 Dissolved oxygen water quality standards by designated use. 90

Table 3.3 Variances allowed in certain Maryland segments. 91

Table 3.4 Percent similarities and stoplight analysis results of green, yellow, and red percentages of the Regulatory Model and Academic Model for 1993-1995 water quality standard assessment. 92

Table 3.5 Values of the Confidence Index (CI) metrics for all 20 Bay segments. 93

Chapter 4. THE COMPETING IMPACTS OF CLIMATE CHANGE AND NUTRIENT REDUCTION ON DISSOLVED OXYGEN

Table 4.1 Characteristics of observation stations. 140

Table 4.2 List of run and scenario names. 141

Table 4.3 Monthly in freshwater flow entering the Bay used for the TMDL+riverCC, TMDL+allCC, and Base+allCC scenarios. 142

Table 4.4 Average change in bottom DO (mg L^{-1}) relative to the TMDL+noCC run for each scenario and region. 143

Table 4.5 Percent of 3-year average bottom DO change as a result of the temperature experiment due to solubility for each region at the surface and bottom of the water column. 144

Table 4A.1 Modified ChesROMS-ECB Equations. 154

Table 4B.1 Total RMSD (and observational mean) of the present model and the model version used in Chapter 2 and Chapter 3. 156

LIST OF FIGURES

Chapter 2. CHALLENGES ASSOCIATED WITH MODELING LOW-OXYGEN WATERS IN CHESAPEAKE BAY: A MULTIPLE MODEL COMPARISON

Figure 2.1 Map of the Chesapeake Bay and its watershed. 40

Figure 2.2 Location of the CBP water quality monitoring stations used in this study. 41

Figure 2.3 Denisty and dissolved oxygen profiles for a mod-bay station comparing the stratification definition used by the CBP with the 10% threshold definitions used here for density and oxygen. 42

Figure 2.4 Normalized summary target and Taylor diagrams illustrating model skill of dissolved oxygen at the surface, MLD_o, and bottom for 13 Chesapeake Bay stations in 2004-2005. 43

Figure 2.5 Normalized target diagrams for Models A-H demonstrating the temporal and spatial skill in resolving the variability of bottom dissolved oxygen concentrations. 44

Figure 2.6 Normalized summary target and Taylor disgram illustrating the skill of MLD_ρ and MLD_o, max $d\rho/dz$, and max dO/dz at 13 Chesapeake Bay stations for 2004-2005 45

Figure 2.7 Normalized target diagrams for Models A-H demonstrating the temporal and spatial skill in resolving the variability of the strength of density sratification and the depth of pycnocline initiation. 46

Figure 2.8 Notmalized summary target and Taylor diagram illustrating model skill of bottom temperature, salinity, chlorophyll, nitrate, and dissolved oxygen at 13 Chesapeake Bay stations for 2004-2005. 47

Figure 2.9 Time series of bottom dissolved oxygen concentrations for station CB4.1C... 48

Figure 2.10 Scatter plots comparing observations of the strength of stratification of the pycnocline/oxycline and the oxygen- and density-defined mixed layer depths ... 49

Figure 2.11 Time series of observations at station CB4.1C from 2003 to 2006 for bottom dissolved oxygen, dissolved oxygen at the MLD_o, and MLD_o. 50

Figure 2.12 Time series of observations of dissolved oxygen and MLD_o contours at station CB4.1C for 2004 and 2005... 51

Chapter 3. EVALUATING CONFIDENCE IN THE IMPACT OF REGULATORY NUTRIENT REDUCTION ON CHESAPEAKE BAY WATER QUALITY

Figure 3.1 Map of the Chesapeake Bay with observation station and Bay segments identified. 94

Figure 3.2 Schematic of designated uses throughout the water column for summer and non-summer seasons. 95

Figure 3.3 Example regression between hourly DO from the standard run and hourly DO from the TMDL-WIP scenario for a single grid cell in a single month. 96

Figure 3.4 Target and Taylor diagrams of the Regulatory and Academic Model demonstrating the combined spatial and temporal skill for modeled surface and bottom temperature, salinity, and DO, as well as maximum stratification and depth of maximum stratification for the 25 observation stations 97

Figure 3.5 Model results at 2 observations stations illustrating the absolute difference in summer DO concentration at the surface and bottom, and the relative change in DO concentration at the surface and bottom. 98

Figure 3.6 Time series of average modeled bottom DO concentrations across main stem stations in CB3MH and CB4MH for the Regulatory Model and the Academic Model. 99

Figure 3.7 Similarity in attainment of WQS for 1993-1995 as demonstrated by percent agreement in stoplight analysis between the Regulatory Model and the Academic Model for the four designated uses. 100

Figure 3.8 Pie charts showing attainment, attainment with buffer, and non-attainment for the 1993-1995 observations, Regulatory Model, and Academic Model for the Deep Water and Deep Channel designated uses. 101

Figure 3.9 Percent non-attainment for the Regulatory Model and the Academic Model for the Deep Channel, Deep Water, and Open Water Summer, as well as monthly Susquehanna freshwater discharge from the CBP watershed model. 102

Figure 3.10 Map of Chesapeake Bay segments color coded by Confidence Index score with green indicating highest confidence and red indicating lowest confidence.103

Chapter 4. THE COMPETING IMPACTS OF CLIMATE CHANGE AND NUTRIENT REDUCTION ON DISSOLVED OXYGEN

Figure 4.1 Map of the Chesapeake Bay with stations identified by region, based primarily on salinity. A: oligohaline, B & C: upper & lower mesohaline (with lowest observed DO concentrations), D: polyhaline.. 145

Figure 4.2 Time series of a 7-day running mean of surface and bottom oxygen concentrations computed for the average of six stations in the Chesapeake Bay upper mesohaline main stem.. 146

Figure 4.3 Time series of a 7-day running mean of difference in bottom oxygen concentrations between the TMDL climate change and TMDL+noCC scenarios computed for the average of six stations in the Chesapeake Bay upper mesohaline main stem. 147

Figure 4.4 Cumulative hypoxic volume for six ranges of DO concentrations, for each of the study years.. 148

Figure 4.5 Percent change due to climate change, relative to the improvement in CHV between the TMDL+noCC scenario and Base+noCC run................................. 149

Figure 4.6 DO differences due to climate change (between the TMDL+noCC and TMDL+tempCC scenarios) averaged for the six stations in Region B for the surface, and bottom of the water column.. 150

Figure 4.7 Bars showing duration of hypoxic volume (> 1km^3) at each DO level for the Base+noCC run and the TMDL+noCC and TMDL+allCC nutrient scenarios... 151

Figure 4.8 Along-Bay transects, with the Susquehanna River in the north and Bay mouth in the south, of DO of the Base+noCC run and TMDL+noCC scenario for July 1, 1993 and July 1, 1995, average summer (May-Sept) DO for 1993 and 1995, the difference in average summer DO between the TMDL+noCC and TMDL+allCC scenarios, and the percent difference in average summer DO between the TMDL+noCC and TMDL+allCC scenarios. 152

Figure 4.9 Percent of the entire Bay that experiences a given DO level during 1993, 1994, and 1995. 153

Figure 4B.1 Target diagram comparing the skill of the ChesROMS-ECB model version used in Chapters 2 and 3 to the one used in the present study. Statistics combine the spatial and temporal variability across 23 stations for 1993-1995................ 157

Figure 4B.2 As in Figure 4B.1, but by Region and for surface and bottom temperature. 158

Figure 4B.3 As in Figure 4B.2, but for surface and bottom salinity.............................. 159

Figure 4B.4 As in Figure 4B.2, but for surface and bottom dissolved oxygen 160

Figure 4B.5 As in Figure 4B.2, but for surface and bottom nitrate... 161

Figure 4B.6 Time series of a 7-day moving mean of temperature at the surface (blue) and bottom (red) with the associated 25th-75th percentiles of the climatological observations (1985-2014) for the surface (blue shading) and bottom (red shading) for the four regions (A, B, C, D) in Figure 4.1... 162

Figure 4B.7 As in Figure 4B.6, except for salinity... 163

Figure 4B.8 As in Figure 4B.6, except for dissolved oxygen... 164

Figure 4B.9 As in Figure 4B.6, except for ammonium... 165

Figure 4B.10 As in Figure 4B.6, except for nitrate... 166

ABSTRACT

Human impacts on the Chesapeake Bay through increased nutrient run-off as a result of land-use change, urbanization, and industrialization, have resulted in a degradation of water quality over the last half-century. These direct impacts, compounded with human-induced climate changes such as warming, rising sea-level, and changes in precipitation, have elevated the conversation surrounding the future of water quality in the Bay. The overall goal of this dissertation project is to use a combination of models and data to ***better understand and quantify the impact of changes in nutrient loads and climate on water quality in the Chesapeake Bay***. This research achieves that goal in three parts.

First, a set of eight water quality models is used to establish a model mean and assess model skill. All models were found to exhibit similar skill in resolving dissolved oxygen concentrations as well as a number of dissolved oxygen-influencing variables (temperature, salinity, stratification, chlorophyll and nitrate) and the model mean exhibited the highest individual skill. The location of stratification within the water column was found to be a limiting factor in the models' ability to adequately simulate habitat compression resulting from low-oxygen conditions.

Second, two of the previous models underwent the regulatory Chesapeake Bay pollution diet mandated by the Environmental Protection Agency. Both models exhibited a similar relative improvement in dissolved oxygen concentrations as a result of the reduction of nutrients stipulated in the pollution diet. A Confidence Index was developed to identify the locations of the Bay where the models are in agreement and disagreement regarding the impacts of the pollution diet. The models were least certain in the deep part of the upper main stem of the Bay and the uncertainty primarily stemmed from the post-processing methodology.

Finally, by projecting the impacts of climate change in 2050 on the Bay, the potential success of the pollution diet in light of future projections for air temperature, sea level, and precipitation was examined. While a changing climate will reduce the ability of the nutrient reduction to improve oxygen concentrations, that effect is trumped by the improvements in dissolved oxygen stemming from the pollution diet itself. However, climate change still has the potential to cause the current level of nutrient reduction to be inadequate. This is primarily due to the fact that low-oxygen conditions are predicted to start one week earlier, on average, in the future, with the primary changes resulting from the increase in temperature.

Overall, this research lends an increased degree of confidence in the water quality modeling of the potential impact of the Chesapeake Bay pollution diet. This research also establishes the efficacy of utilizing a multiple model approach to examining projected changes in water quality while establishing that the pollution diet trumps the impact from climate change. This work will lead directly to advances in scientific understanding of the response of water quality, ecosystem health, and ecological resilience to the impacts of nutrient reduction and climate change.

USING WATER QUALITY MODELS IN MANAGEMENT:

A MULTIPLE MODEL ASSESSMENT, ANALYSIS OF CONFIDENCE, AND EVALUATION OF CLIMATE CHANGE IMPACTS

1. INTRODUCTION

1.1 Motivation

Anthropogenic impacts, primarily in the form of excess nutrients derived from fertilizers, sewage, and storm water runoff, have resulted in the degradation of coastal water quality throughout much of the world (Diaz and Rosenberg, 2008; Rabalais et al., 2010). In the Chesapeake Bay, anthropogenic impacts have specifically resulted in decreased oxygen levels and increased volumes of low-oxygen waters, particularly exacerbated since the mid-20th century (Cooper and Brush, 1991; Boesch et al., 2001; Hagy et al., 2004). The increased volume of low-oxygen waters has negatively affected the health of the Chesapeake Bay ecosystem and economy (BRFP, 2004; Nelson, 2014). In an effort to restore the Bay's water quality to acceptable levels under the Clean Water Act, the six states and Washington, D.C. that make up the Chesapeake Bay watershed established a pollution diet for the region that is primarily focused on increasing dissolved oxygen (DO) concentrations in the waters of the central portion of the Bay (Keisman and Shenk, 2013).

The emphasis on DO in determining water quality has developed due to the observed global proliferation of hypoxic events both spatially and temporally in the world's coastal oceans (Diaz, 2001; Gooday et al., 2009). Hypoxic waters lack sufficient oxygen levels required for oxygen-dependent aquatic organisms to survive. While there is not a strict numerical definition, it is widely accepted that hypoxia is reached when DO concentrations fall below ~30% saturation, or 2 mg L^{-1}. Even though hypoxia may not be reached until DO concentrations decrease to these levels, many species can incur negative health impacts at concentrations as high as 5 mg L^{-1} (Vaquer-Sunyer and Duarte, 2008; Portner and Lanning, 2009).

The Bay's hypoxic waters are primarily caused by anthropogenic impacts such as land-use change, industrialization, and urbanization that have dramatically increased the input of nutrients to the Bay, driving increased primary production (Harding and Perry, 1997; Kemp et al., 2005). This increased primary production results in elevated organic matter throughout the water column that is decomposed by DO-utilizing microorganisms and bacteria. The resulting hypoxic events in the Bay can be intensely episodic (Kemp et al., 2009) and cause deleterious harm to commercially and ecologically valuable fish, crabs, and shellfish in the adult and larval stages (Keister et al., 2000; Breitburg, 2002; Ekau et al., 2010; Buchheister et al., 2013).

The efforts to clean up the Bay, established by the pollution diet, are estimated to cost in the tens of billions of dollars (Nelson, 2014). With such high potential costs, it is critical that the science behind the pollution diet be well founded and constrained as best as possible. As the regulation developed to clean up the Chesapeake Bay was predicated on an environmental modeling framework (USEPA, 2010), it is important that scientists removed from the regulatory process critically analyze the findings. In that vein, this research focuses on the estuarine water quality-modeling portion of the pollution diet in an effort to assess model capabilities, analyze confidence in projections, and evaluate those projections in light of a changing climate.

1.2 The Chesapeake Bay TMDL

The pollution diet developed to improve water quality in the Chesapeake Bay is called the 2010 Chesapeake Bay Total Maximum Daily Load (TMDL), which limits the loads of nitrogen, phosphorus, and sediment delivered to the Bay from the watershed. The science behind TMDLs in general is relatively young, with the Environmental Protection Agency (EPA) first drafting TMDL regulations in 1992, for the purpose of ensuring compliance with the Clean Water Act (CWA; Keisman and Shenk, 2013). In the Chesapeake Bay, the development of the TMDL was initiated in 2000 and formally implemented in 2010 as the largest and most complex TMDL system in the nation (USEPA, 2010). TMDL levels were set using a coupled watershed-estuarine modeling system developed and maintained by the EPA's Chesapeake Bay Program (CBP), together with an extensive set of CBP monitoring data. Historical observations from the CBP monitoring program, beginning in 1984 and continuing through today, serve as the basis for the regulatory model's calibrations.

Since 1983, when the first Bay agreement was signed, the Chesapeake Bay community has worked to remediate the declining health of the Bay (Batiuk et al., 2013). Despite the efforts of the restoration community, Bay health did not adequately improve in accordance with the CWA's goal that all waters of the United States be fishable and swimmable. In order to force the entire watershed to act to increase Bay health, a TMDL for the Chesapeake Bay was proposed and work began on its planning in 2000 (USEPA, 2010). Finally, in 2009, President Obama issued Executive Order 13508: Chesapeake Bay Protection and Restoration (Executive Order No. 13508, 2009). That Executive Order mandated that the federal government, under the leadership of the EPA, take the lead in the restoration of the Bay and its watershed. The mandate gave the EPA the backing, both legally and ceremoniously, that it needed to fully endorse and impose the regulatory Chesapeake Bay TMDL in accordance with the Clean Water Act.

Under the CWA, the EPA only has the authority to order a set of TMDLs if the individual states within the watershed have not adequately addressed the environmental issue. The purview of the EPA within the CWA to implement specific functions of a TMDL is very broad and resulted in a 2013 court ruling (*American Farm Bureau Federation et al., v. United States Environmental Protection Agency*, 2013) where the judge rejected all three of the plaintiffs' charges and held that (1) the EPA did have jurisdiction to set specific load allocations, (2) the EPA did allow adequate public comment, and (3) that within the modeling efforts conducted by the EPA to set TMDL levels, the EPA was granted deference. The deference finding granted by the court is based on the plaintiff's inability to satisfactorily prove harm from the EPA's selection of methods and monitoring data used in the modeling process that resulted in the specific TMDLs. By granting deference, Judge Rambo effectively punted the question of the scientific validity of the Chesapeake Bay TMDL's development because the court did not have the expertise to verify the EPA's methods.

Following the decision, the Farm Bureau appealed the District Court's ruling and was supported on the appeal by 20 states outside of the Chesapeake Bay watershed. The states were primarily concerned with the potential for the Chesapeake Bay TMDL to serve as a blueprint for other locations, such as the Mississippi River watershed. In July 2015, the Third Circuit upheld the District Court's decision to allow the EPA to move forward with TMDL implementation (*American Farm Bureau Federation et al., v. United States Environmental Protection Agency*, 2015). Soon after, the Farm Bureau looked to the Supreme Court in hopes that the nation's highest court would accept their petition for the case to be relitigated. In February 2016, the Court denied the Farm Bureau’s petition, effectively affirming the decisions of the lower courts.

While the TMDL has effectively withstood a barrage of lawsuits attempting to pick holes in the regulatory procedure and discredit the science, it is not safe for perpetuity. It is certain that the interests most negatively affected by the regulation will continue to fight against its implementation. In light of that, it is incumbent on the scientific community to continue to utilize transparency and the most current scientific techniques and understanding to ensure that the scientific basis of the regulation is on sound footing.

1.3 The Role of Multiple Models

As the EPA and CBP continue to uphold the pollution diet in the courts of law and public opinion, the use of multiple models has entered the discussion as a potential way to address future questions of confidence in the modeled scenario predictions used for TMDL development. In early 2012, the CBP requested that the Scientific and Technical Advisory Committee (STAC) conduct two workshops on Multiple Models in Management. The goal of these meetings was to discuss the utilization of a multiple model approach when evaluating the recovery of the Chesapeake Bay system in order to enhance overall confidence in model projections and to better define model uncertainty (Friedrichs et al., 2012; Weller et al., 2013).

While the use of multiple models in decision-making may be relatively new to the EPA and CBP, the Intergovernmental Panel on Climate Change (IPCC) and the National Weather Service (NWS) are already utilizing multiple models in order to contextualize their projections. The IPCC uses their set of models to establish a degree of certainty for each projection. The IPCC discusses uncertainty with a qualitative “Confidence Scale” and a quantitative “Likelihood Scale” (Mastrandrea et al., 2010). Both of these scales are meant to describe the level of agreement between models, evidence, and scientific understanding in relation to any given environmental projection. The NWS utilizes model ensembles in hurricane predictions to inform a “consensus forecast” (NWS, 2009). Depending on the situation, an NWS model ensemble may be composed of multiple models projecting identical variables over a given timeframe, or of a single model that has been run multiple times with different initial conditions or model parameters. The NWS also utilizes a "corrected" consensus technique that weights the output of different models to account for model biases since not all models are developed to examine the same processes (NWS, 2009). The methods used by the IPCC and NWS in working with multiple models can serve as frameworks for utilizing multiple models in a regulatory context for the EPA in the future.

As environmental policies continue to be developed, it is necessary that past policies be re-evaluated with improved technology and understanding, both of which can be addressed with the use of multiple models. Ultimately, a better understanding of uncertainty by examining the results of multiple models will allow the CBP to discuss the potential effectiveness of TMDL implementation with the public and individual stakeholders while also being able to better defend the methods and science behind individual nutrient allocations. To do this, the following research utilizes the expertise of water quality modelers throughout the Chesapeake Bay modeling community to assess how the different models perform across multiple variables in order to establish a range of capabilities. Using the information garnered from those multiple models, this research compares where, when, and why two of the models agree and disagree on the potential impact of the TMDL in an effort to assess confidence in the ability of the regulation to meet its goals. Additionally, this research explores the impact that climate change will have on the trajectory of Chesapeake Bay water quality under the current TMDL. The results of the project allow for an in-depth examination of the limitations of current water quality models in terms of management application, how multiple models can be used to establish a range of confidence in a regulatory setting, and why the establishment of a pollution diet for today must consider the climate impacts of tomorrow.

1.4 Dissertation Motivating Questions and Structure

This study explores the use of water quality models in management by asking three fundamental questions:

- What are the current limitations of Chesapeake Bay water quality models in terms of modeling low-oxygen waters?
- What is our confidence in the modeling of the water quality portion of the Chesapeake Bay TMDL?
- How might climate change affect the efficacy of the Chesapeake Bay TMDL?

Chapter 2 evaluates the strengths and weaknesses of eight Bay water quality models of varying complexity, model structure, and model design in terms of their ability to simulate water column oxygen dynamics, as well as the variables thought to be the main drivers of dissolved oxygen variability. The skill of a model mean was compared to the skill of the individual models and the difference in skill between the complex and more simplistic models was also analyzed. An additional analysis of observations led to the identification of a key variable that the models are poorly resolving, limiting their effectiveness for management decisions involving habitat compression.

Chapter 3 utilizes two of the models from Chapter 2 to assess confidence in the water quality modeling of the Chesapeake Bay TMDL. Both models were forced by similar base and nutrient reduction scenarios and subsequently underwent the same post-processing analysis used to initially establish the TMDL. Model results were then compared in terms of absolute and relatively changes in Bay dissolved oxygen, as well as the similarity in time and place of water quality standards. A Confidence Index was developed to identify locations in the Bay where the models were in highest agreement (high confidence) and lowest agreement (low confidence).

Chapter 4 takes one of the models from Chapter 3 to examine the potential impact changes in climate may have on the TMDL to adequately improve water quality conditions in terms of dissolved oxygen. Changes in temperature, sea level rise, and precipitation (river flow) for a 2050 time horizon were examined in order to identify the absolute and relative impacts of each projected change. To understand the difference between a reduced nutrient future and a business as usual future, climate change impacts were evaluated with both the TMDL and present day nutrient loading conditions.

Chapter 5 explores the role multiple models are, can, and should play in regulatory policies like the TMDL while establishing the complexities and limitations involved in such an endeavor.

Literature Cited

American Farm Bureau Federation et al., v. United States Environmental Protection Agency, 2013. No. 1:11-CV-0067. United States District Court for the Middle District of Pennsylvania.

American Farm Bureau Federation et al., v. United States Environmental Protection Agency, 2015. No. 13-4079. United States Court of Appeals for the Third Circuit.

Batiuk, R.A., L.C. Linker, and C.F. Cerco. 2013. Featured collection introduction: Chesapeake Bay Total Maximum Daily Load development and application. *Journal of the American Water Resources Association*, 45:5. doi: 10.1111/jawr.12114.

Boesch, D.F., R.B. Brinsfield, and R.E. Magnien. 2001. Chesapeake Bay Eutrophication: Scientific Understanding, Ecosystem Restoration, and Challenges for Agriculture. *Journal of Environmental Quality*, 30: 303-320.

Breitburg, D. 2002. Effects of Hypoxia, and the Balance between Hypoxia and Enrichment, on Coastal Fishes and Fisheries. *Estuaries*, 25: 767-781.

BRFP: Blue Ribbon Finance Panel. 2004. *Saving a National Treasure: Financing the Cleanup of the Chesapeake Bay*.

Buchheister, A., C.F. Bonzek, J. Gartland, R.J. Latour. 2013. Patterns and drivers of the demersal fish community of Chesapeake Bay. *Marine Ecology Progress Series*, 481: 161-180. doi: 10.3354/meps10253.

Cooper, S.R., and G.S. Brush. 1991. Long-Term History of Chesapeake Bay Anoxia. *Science*, 254: 992-996.

Diaz, R. J. 2001. Overview of Hypoxia around the World. *Journal of Environmental Quality*, 30: 275-281.

Diaz. R. J., and R. Rosenberg. 2008. Spreading Dead Zones and Consequences for Marine Ecosystems. *Science*, 321: 926-929, doi:10.1126/science.1156401.

Ekau, W., H. Auel, H.O. Portner, and D. Gilbert. 2010. Impacts of hypoxia on the structure and processes in pelagic communities (zooplankton, macro-invertebrates and fish). *Biogeosciences*, 7. doi:10.5194/bg-7-1669-2010.

Executive Order No. 13508, 3 C. F. R. 2009. Chesapeake Bay Protection and Restoration.

Friedrichs, M.A.M., K.G. Sellner, and M.A. Johnston. 2012. Using Multiple Models for Management in the Chesapeake Bay: A Shallow Water Pilot Project, Chesapeake Bay Program Scientific and Technical Advisory Committee Report, No. 12-003.

Gooday, A.J., F. Jorissen, L.A. Levin, J.J. Middleburg, S.W.A. Naqvi, N.N. Rabalais, M. Scranton, and J. Zhang. 2009. Historical records of coastal eutrophication-induced hypoxia. *Biogeosciences*, 6: 1707-1745.

Hagy, J.D., W.R. Boyton, C.W. Keefe, and K.V. Wood. 2004. Hypoxia in Chesapeake Bay, 1950-2001: Long-term Change in relation to Nutrient Loading and River Flow. *Estuaries*, 27: 634-658.

Harding, L.W., and E.S. Perry. 1997. Long-term increase of phytoplankton biomass in Chesapeake Bay, 1950-1994. *Marine Ecology Progress Series*, 157, 39-52.

Keisman, J., and G. Shenk. 2013. Total maximum daily load criteria assessment using monitoring and modeling data. *Journal of the American Water Resources* Association, 49, 1134-1149. doi:10.1111/jawr.12111.

Keister, J.E., E.D. Houde, and D.L. Breitburg. 2000. Effects of bottom-layer hypoxia on abundances and depth distributions of organisms in Patuxent River, Chesapeake Bay. *Marine Ecology Progress Series*, 205: 43-59.

Kemp, W.M., W. R. Boyton, J. E. Adolf, D. F. Boesch, W. C. Boicourt, G. Brush, J. C. Cornwell, T.R. Fisher, P.M. Gilbert, J.D. Hagy, L.W. Harding, E.D. Houde, D.G. Kimmel, W.D. Miller, R.I.E. Newell, M.R. Roman, E.M. Smith, and J.C. Stevenson. 2005. Eutrophication of Chesapeake Bay: Historical Trends and Ecological Interactions. *Marine Ecology Progress Series,* 303: 1-29.

Kemp, W. M., J. M. Testa, D. J. Conley, D. Gilbert, and J. D. Hagy. 2009. Temporal Responses of Coastal Hypoxia to Nutrient Loading and Physical Controls. *Biogeosciences*, 6: 2985-3008.

Mastrandrea, M. D., C. B. Field, T. F. Stocker, O. Edenhofer, K. L. Ebi, D. J. Frame, H. Held, et al. 2010: Guidance Note for Lead Authors of the IPCC Fifth Assessment Report on Consistent Treatment of Uncertainties. *Intergovernmental Panel on Climate Change (IPCC)*.

Nelson, R.H. 2014. How to Save the Chesapeake Bay TMDL: The Critical Role of Nutrient Offsets. *William & Mary Environmental Law and Policy Review*, 38.

NWS: National Weather Service. 2009. Technical Summary of the National Hurricane Center Track and Intensity Models. http://www.nhc.noaa.gov/pdf/model_summary_20090724.pdf

Portner, H.O., G. Lanning. 2009. Oxygen and capacity limited thermal tolerance. The hypoxic environment. *Fish physiology, Vol 27*, Academic Press, San Diego, CA, p 143-191.

Rabalais, N.N., R.J. Diaz, L.A. Levin, R.E. Turner, D. Gilbert, and J. Zhang. 2010. Dynamics and distribution of natural and human-caused hypoxia. *Biogeosciences*, 7: 585-619.

USEPA (U.S. Environmental Protection Agency). 2010. Chesapeake Bay Total Maximum Daily Load for Nitrogen, Phosphorus, and Sediment. U.S. Environmental Protection Agency Chesapeake Bay Program Office, Annapolis, MD.

Vaquer-Sunyer, R., and C.M. Duarte. 2008. Thresholds of Hypoxia for Marine Biodiversity. *Proceedings of the National Academies of Science*, 105: 15452-15456. doi: 10.1073/pnas.0803833195.

Weller, D., B. Benham, M.A.M. Friedrichs, N. Gardner, R. Hood, R. Najjar, M. Paolisso, P. Pasquale, K. Sellner, and G. Shenk. 2013. Multiple Models for Management in the Chesapeake Bay, Chesapeake Bay Program Scientific and Technical Advisory Committee Workshop Report, No. 14-004.

Chapter 2:

Challenges associated with modeling low-oxygen waters in Chesapeake Bay: a multiple model comparison

Isaac D. Irby[1], Marjorie A. M. Friedrichs[1], Carl T. Friedrichs[1], Aaron J. Bever[2], Raleigh R. Hood[3], Lyon W.J. Lanerolle[4,5], Ming Li[6], Lewis Linker[7], Malcolm E. Scully[8], Kevin Sellner[9], Jian Shen[1], Jeremy Testa[6], Hao Wang[3], Ping Wang[10], and Meng Xia[11]

[1]Virginia Institute of Marine Science, College of William & Mary, P.O. Box 1346, Gloucester Point, VA 23062, USA
[2]Anchor QEA, LLC, 130 Battery Street, Suite 400, San Francisco, CA 94111, USA
[3]Horn Point Laboratory, University of Maryland Center for Environmental Science, P.O. Box 775, Cambridge, MD 21613, USA
[4]NOAA/NOS/OCS Coast Survey Development Laboratory, 1315 East-West Highway, Silver Spring, MD 20910, USA
[5]ERT Inc., 14401 Sweitzer Lane Suite 300, Laurel, MD 20707, USA
[6]Chesapeake Biological Laboratory, University of Maryland Center for Environmental Science, P.O. Box 38, Solomons, MD 20688, USA
[7]US Environmental Protection Agency Chesapeake Bay Program Office, 410 Severn Avenue, Annapolis, MD 21403, USA
[8]Woods Hole Oceanographic Institution, Applied Ocean Physics and Engineering Department, Woods Hole, MA 02543, USA
[9]Chesapeake Research Consortium, 645 Contees Wharf Road, Edgewater, MD 21037, USA
[10]VIMS/Chesapeake Bay Program Office, 410 Severn Avenue, Annapolis, MD 21403, USA
[11]Department of Natural Science, University of Maryland Eastern Shore, MD, USA

Published in *Biogeosciences*

April 2016

Volume 13

Pages 2011 – 2028

doi: 10.5194/bg-13-2011-2016

2. CHALLENGES ASSOCIATED WITH MODELING LOW-OXYGEN WATERS IN CHESAPEAKE BAY: A MULTIPLE MODEL COMPARISON

Key Points

- There is no statistical difference between the ability of simple and complex models to simulate the mean and monthly variability of bottom dissolved oxygen.
- Model skill for dissolved oxygen is greater than that of the variables generally considered to drive DO variability.
- To better simulate habitat compression due to hypoxia, the ability of models to simulate the mixed layer depth must be improved.
- The mean of a set of models is more skilled than any individual model across a broad suite of variables.

Abstract

As three-dimensional (3-D) aquatic ecosystem models are used more frequently for operational water quality forecasts and ecological management decisions, it is important to understand the relative strengths and limitations of existing 3-D models of varying spatial resolution and biogeochemical complexity. To this end, 2-year simulations of the Chesapeake Bay from eight hydrodynamic-oxygen models have been statistically compared to each other and to historical monitoring data. Results show that although models have difficulty resolving the variables typically thought to be the main drivers of dissolved oxygen variability (stratification, nutrients, and chlorophyll), all eight models have significant skill in reproducing the mean and seasonal variability of dissolved oxygen. In addition, models with constant net respiration rates independent of nutrient supply and temperature reproduced observed dissolved oxygen concentrations about as well as much more complex, nutrient-dependent biogeochemical models. This finding has significant ramifications for short-term hypoxia forecasts in the Chesapeake Bay, which may be possible with very simple oxygen parameterizations, in contrast to the more complex full biogeochemical models required for scenario-based forecasting. However, models have difficulty simulating correct density and oxygen mixed layer depths, which are important ecologically in terms of habitat compression. Observations indicate a much stronger correlation between the depths of the top of the pycnocline and oxycline than between their maximum vertical gradients, highlighting the importance of the mixing depth in defining the region of aerobic habitat in the Chesapeake Bay when low-oxygen bottom waters are present. Improvement in hypoxia simulations will thus depend more on the ability of models to reproduce the correct mean and variability of the depth of the physically driven surface mixed layer than the precise magnitude of the vertical density gradient.

2.1 Introduction

Since the middle of the last century, anthropogenic impacts have dramatically decreased water quality throughout the Chesapeake Bay (Boesch et al., 2001), one of the largest estuaries in North America. Land-use change along with the industrialization and urbanization of the Chesapeake Bay watershed have caused dramatic increases in nutrient inputs to the bay (Kemp et al., 2005), spurring additional primary production and phytoplankton abundance (Harding Jr. and Perry, 1997). Because increased primary production leads to more organic matter throughout the water column that is eventually decomposed by bacteria, these increased nutrient inputs to the bay have led to a corresponding decrease in dissolved oxygen (DO) concentrations (Hagy et al., 2004). Hypoxia, generally defined as the condition in which DO concentrations are below 2 mg L^{-1}, usually initiates seasonally in the northern portion of the bay and expands southward as summer develops (Kemp et al., 2009; Testa and Kemp, 2014). Although hypoxia in the Chesapeake Bay has likely existed since European colonization (Cooper and Brush, 1991, 1993), recent studies have highlighted an accelerated rise in the number and spatial extent of hypoxic, as well as anoxic (DO concentrations <0.2 mg L^{-1}), events in the bay since the 1950s, primarily attributed to increased anthropogenic nutrient input (Hagy et al., 2004; Kemp et al., 2005; Gilbert et al., 2010). These impacts are likely to be exacerbated by future climate change (Najjar et al., 2010; Meire et al., 2013; Harding Jr. et al., 2015).

Interest in the ecological impacts of reduced DO concentrations has been elevated due to the observed proliferation of hypoxic events in the world's coastal oceans, creating vast dead zone areas that compress suitable habitats for many marine species (Diaz, 2001; Diaz and Rosenberg, 2008; Pierson et al., 2009). Low-DO waters can greatly impact the abundance and health of important ecological species, potentially resulting in suffocation and major kills of fish, crabs, and shellfish (Breitburg, 2002; Ekau et al., 2010; Levin et al., 2009). While the presence of DO concentrations < 2 mg L^{-1} have been shown to decrease the abundance of fish larvae (Keister et al., 2000), some species can incur negative health impacts and modify their behavior at significantly higher DO concentrations (Vaquer-Sunyer and Duarte, 2008). DO concentrations of ~ 4 mg L^{-1} have been found to compress demersal fish habitat as fish seek out more oxygenated waters

(Buchheister et al., 2013). Zooplankton, a crucial food source for valuable species, have also been found to exhibit changes in distribution and predation when subject to large volumes of low-DO water, potentially leading to further impacts along the food chain (Breitburg et al., 1997; Pierson et al., 2009). Invertebrates have similarly been found to alter their behavior under low-DO conditions (Riedel et al., 2014). In the Chesapeake Bay, multiple regulated fish species, such as striped bass and American shad, require oxygen restoration targets as high as 5 mg L^{-1} (USEPA, 2010). The greatest impact of low DO concentrations spatially will depend on the specific living resource; however, temporally, late spring to early fall is of most concern. As a result of the significant ecological importance of oxygen on living resources in the bay, DO concentrations are used as a primary indicator in assessing water quality for Chesapeake Bay regulations (Keisman and Shenk, 2013).

Improving the health of the Chesapeake Bay has become a priority for the Environmental Protection Agency (EPA) along with the six states and Washington, DC that make up the bay watershed (Fig. 2.1), and together they have committed to utilizing a suite of regulatory models to inform their management decisions (USEPA, 2010). The Chesapeake Bay Program (CBP), a regional partnership that has led and directed the restoration of the Chesapeake Bay since 1983, has undertaken an extensive modeling effort of the bay (Cerco and Cole, 1993; Cerco et al., 2002; Cerco and Noel, 2004, 2013). This modeling system is being used by the CBP to estimate the aggregate effect of changes in management practices, including land use, atmospheric deposition, animal populations, and fertilizer and manure application. Recently, the modeling system has been used to conduct scenario simulations to assess management actions needed to achieve desired bay water quality standards (USEPA, 2010). Ultimately this model was used to establish a regulatory set of total maximum daily loads of nutrients and sediment delivered from the watershed, with the goal of significantly improving water quality throughout the bay (USEPA, 2010).

Many 3-D hydrodynamic-oxygen models of varying complexity stemming from the academic research community have also been used to simulate DO concentrations throughout the Chesapeake Bay (Scully, 2010, 2013; Hong and Shen, 2013; Feng et al., 2015; Testa et al., 2014; Y. Li et al., 2015). Bever et al. (2013) specifically demonstrated that multiple models of varying complexity are able to generate skillful estimates of hypoxic volume in the bay. Some of these models are being used in the bay to simulate short-term and/or seasonal forecasts of DO conditions. Furthermore, some models are also being used to generate scenario forecasts, or projections, that assess the impact of changes in management practices on estuarine DO concentrations, in some cases taking into account the impacts of future changes in climate.

As ecosystem and water quality models are increasingly used for operational forecasts as well as scenario-based management decisions by the regulatory and academic research communities, it is important to understand the relative strengths and limitations of existing models of varying complexity. The ability to discern which variables must be most accurately simulated in order to adequately reproduce the temporal and spatial variability of bay oxygen concentrations is a necessary prerequisite for fully understanding how volumes of low-DO water are initiated and sustained within water quality models. The utilization of multiple models can also inform projections by providing independent confidence bounds for management decisions. To those ends, the over-arching goals of this research are to compare the relative skill of various 3-D Chesapeake Bay models characterized by different levels of biogeochemical complexity and spatial resolution, to better understand factors limiting their ability to reproduce observed DO distributions, and to suggest approaches for the continued improvement of these models.

2.2 Methods

2.2.1 Participating Chesapeake Bay models

Eight 3-D models were evaluated in this study (Table 2.1), each of which includes hydrodynamic and DO components. Among the eight models, there are four different hydrodynamic base models. Models B, C, D, F, and G utilize the Regional Ocean Modeling System (ROMS; Shchepetkin and McWilliams, 2005; Haidvogel et al., 2008) that employs a structured grid with sigma layers in the vertical dimension. Specifically, Models B, C, and F use a ROMS implementation developed for the Chesapeake Bay based on Xu et al. (2012; ChesROMS). Model D employs a ROMS implementation for the Chesapeake Bay based on M. Li et al. (2005), while Model G uses the ROMS-based

Chesapeake Bay Operational Forecast System (CBOFS; Lanerolle et al., 2011). Models A, E, and H each use a different hydrodynamic base model: the Curvilinear Hydrodynamics in Three Dimensions model (CH3D; Cerco et al., 2010), the Finite-Volume Community Ocean Model (FVCOM; Jiang and Xia, 2016), and the Hydrodynamic Eutrophication Model – Environmental Fluid Dynamics Code (EFDC; Park et al., 1995; Hong and Shen, 2012; Du and Shen, 2015), respectively. The only model that employs a non-sigma vertical grid is Model A and the only model utilizing an unstructured horizontal grid is Model E. While Model E contains 10 sigma vertical layers, all of the other sigma grids use 20 layers. All of the grids vary in terms of their horizontal resolution, with Models A and G utilizing the highest resolution horizontal grids.

These four hydrodynamic models are coupled to five different models used to simulate DO (Table 2.1). Models A, B, C, D, and E utilize full biogeochemical models that include various combinations of oxygen, phytoplankton, zooplankton, and multiple inorganic and organic nutrients as state variables. Specifically, Models A and E employ a version of the Integrated Compartment Model (ICM; Cerco et al., 2010; Jiang et al., 2015), Model B uses the Estuarine Carbon Biogeochemistry model (ECB; Feng et al., 2015), Model C uses the Biogeochemistry model (BGC; Brown et al., 2013), and Model D uses the Row–Column AESOP model (RCA; Testa et al., 2014). In terms of food web complexity the models vary considerably: Models B and C employ a single phytoplankton group whereas Model D uses two phytoplankton groups, Model E uses three, and Model A, the most complex of the participating models, uses five.

In contrast to the full biogeochemical models discussed above (Models A through E), Models F, G, and H represent oxygen dynamics as simply as possible and therefore do not utilize a full biogeochemical component. Rather, the models impose a biological oxygen consumption rate that is model-specific, but constant in both space and time. This component is referred to as a constant-respiration model (CRM). In this model, DO is introduced to the estuary via the river and ocean boundaries and is set to saturation at the estuarine surface. This constant-respiration oxygen parameterization (Scully, 2010) is simplistic, yet has been shown to adequately represent Chesapeake Bay oxygen dynamics (Scully, 2010, 2013; Bever et al., 2013).

The major difference in forcing between the eight model implementations is that Models A and B use riverine input derived from watershed models, whereas Models C–H used the measured flow from United States Geological Survey gauging stations, extrapolated using various techniques. Model A utilized the CBP's regulatory watershed model (Shenk and Linker, 2013), while Model B utilized the Dynamic Land Ecosystem Model (Yang et al., 2015a, b; Tian et al., 2015). At the open boundary with the Atlantic Ocean, Models B, C, D, F, G, and H utilize a sub-tidal elevation extrapolated from tidal stations on either side of the open boundary. Model E uses the TPXO tidal model, while Model A uses a mix of observational and model forcing (Cerco et al., 2010). While Model B utilizes wind forcing based on observations from the Thomas Point Light, Models C through H use wind estimates from the North American Regional Reanalysis (NARR).

The eight models used in this analysis have been developed for a variety of purposes. Model A is a governmental regulatory model developed by the CBP that has been extensively calibrated specifically to examine water quality issues in the Chesapeake Bay (Cerco and Cole, 1993; Cerco and Noel, 2004, 2013; Cerco et al., 2010) and has been used in the development of the 2010 Chesapeake Bay Total Maximum Daily Load (USEPA, 2010). The National Oceanic and Atmospheric Administration employs the hydrodynamic component of Model F for operational forecasts of a variety of physical estuarine parameters for the Chesapeake Bay (http://www.tidesandcurrents.noaa.gov/ofs/cbofs/cbofs.html). The other six models are academic models used in diverse research efforts focused on the Chesapeake Bay but not necessarily specifically on DO dynamics.

Finally, a ninth model is calculated as the mean of the results from the eight models described above, and is referred to here as Model Mean, or Model M.

2.2.2 Available Chesapeake Bay observations

Model simulations were compared to cruise data from the CBP for 2004 and 2005 from 13 stations along the main stem of the bay (Table 2.2, Fig. 2.2). The years 2004 and 2005 were selected to represent relatively wet and average years, respectively, and the 13 stations were chosen as they have been found to offer optimal estimates of bay-wide hypoxic volume (Bever et al., 2013). Stations were sampled on up to 34 cruises over the 2 years (Table 2.2), generally twice a month from April to August and once a month for the remainder of the year. Observational data can be downloaded from the CBP Water Quality Database (http://www.chesapeakebay.net/data/ downloads/cbpwaterqualitydatabase1984present). Variables downloaded from the CBP website and used in this study were temperature, salinity, DO, nitrate + nitrite (hereafter abbreviated as "nitrate"), and chlorophyll a (hereafter abbreviated as "chlorophyll"). For most cruises, observations of temperature, salinity, and DO were made at roughly 1 m intervals throughout the water column, whereas observations of chlorophyll and nitrate were generally made only at the surface, bottom, and sometimes one or two mid-water column locations. For further information on available water quality observations, please see USEPA (2012). While these observations were publicly available for model assessment during calibration of all of the models, they represent a very small subset of the 30 years of EPA observations across over 100 bay stations. The models compared here were calibrated based on access to the larger data set and for conditions in the bay in general, not specifically for the 13 stations and 2 years considered here.

2.2.3 Calculation of stratification and mixed layer depth

Stratification of the density and oxygen fields was examined to identify the maximum gradient of the pycnocline and oxycline as well as the depth of the top of the pycnocline and oxycline. In open ocean studies, the depth of the top of stratification is commonly referred to as the mixed layer depth (MLD), although this term is less frequently used in the estuarine literature. As the research presented here distinguishes between the depths of the top of the pycnocline and that of the oxycline, these will be referred to respectively as the density (ρ) mixed layer depth (MLD_ρ) and the oxygen mixed layer depth (MLD_O). Density was calculated via a classical density formula that is also utilized by the CBP for use in the Chesapeake Bay (Fofonoff and Millard, 1983; USEPA, 2004) and is a function of temperature and salinity.

The CBP defines the top and bottom of stratification in order to distinguish individual designated use areas for water quality management purposes (USEPA, 2004). They suggest that the top of the pycnocline be defined as the shallowest occurrence of a density gradient of 0.1 kg m^{-4} or greater as resolved by CBP profile observations, which are typically spaced at 0.5–2 m depth intervals. If density gradients throughout the water column are less than 0.1 kg m^{-4}, they define the water to be unstratified. The 0.1 kg m^{-4} threshold definition is designed to identify any initiation of stratification that may serve to cut off vertical mixing from a nearly perfectly well-mixed layer.

While the CBP definition described above delineates between designated use boundaries according to density, our research focuses on the relationship between the pycnocline and oxycline, requiring an alternate definition that can be applied to both the density and oxygen distributions. In addition, the CBP definition often generates estimates for the depth of the top of the pycnocline that are too shallow compared to the maximum depth of surface mixing (Fig. 2.3). As a result, a percentage threshold criterion was developed that identifies the bottom of the reasonably well-mixed layer, rather than perfectly mixed layer, and is used in this analysis. The percentage threshold method defines a density or DO profile as being stratified if a change of 10% of the difference between the profile's maximum and minimum values occurs within a single meter (Fig. 2.3). For example, if the maximum DO concentration throughout the water column on an individual sampling date is 10 mg L^{-1} and the minimum concentration is 1 mg L^{-1}, stratification is defined to be present if a difference of 0.9 mg L^{-1} is present within 1 m. As recommended by the CBP, the uppermost meter of the water column is not considered (USEPA, 2004). The mixed layer depth is therefore defined as the shallowest level (below 1 m depth) where stratification is identified. The minimum stratification criterion utilized in this analysis requiring a profile to pass the 10% threshold also ensures that observations where very little stratification exists do not bias the stratification results while also allowing for a single criterion to be used across multiple stratification variables.

2.2.4 Model skill metrics

Simulations of the Chesapeake Bay from the eight models described above were statistically compared to historical monitoring data using a variety of skill metrics including root-mean squared difference (RMSD), bias, standard deviation, and correlation coefficient. These metrics are illustrated on Taylor and target diagrams

(Taylor, 2001; Hofmann et al., 2008; Jolliff et al., 2009), which offer a compact way of assessing model skill by displaying a number of different skill metrics. Target diagrams illustrate the bias and total RMSD of model output, which Taylor diagrams do not. Taylor diagrams include quantitative information on the standard deviations and correlations between the model output and the observations, which target diagrams do not. Both diagrams, however, represent unbiased RMSD, sometimes called "centered-pattern RMSD". On target diagrams, a model symbol above the horizontal axis overestimates the mean of the observations and a model symbol to the right of the vertical axis overestimates the variability of the observations. (See Hofmann et al. (2008) and Jolliff et al. (2009) for a more detailed description of these diagrams.) On Taylor diagrams, a model symbol lying on the horizontal axis exactly correlates to the observations and a model symbol further from the origin than the observation symbol overestimates the standard deviation of the observations. (See Taylor (2001) for a more detailed description of these diagrams.)

Taylor and target diagrams presented here are normalized to the standard deviation of the observations, allowing multiple variables be represented on the same plot. This also conveniently allows the unit circle on a target diagram to represent the skill of a model defined as the mean of the observations. In effect, this means that if a model falls within the unit circle, it exhibits a skill that is greater than the skill obtained if one were to simply use the mean of the observations. The Taylor and target plots are either temporal (displaying model skill at a single station over the study period) or spatial (displaying model skill during a single month over the entire set of study stations). In addition, summary diagrams are presented which combine both temporal (examining the seasonal changes at each individual station) and spatial (examining differences across the bay during an individual month) variability.

Model skill was assessed using the hourly model output (daily for CH3D-ICM chlorophyll and nitrate) that was nearest in time to that of the observation and from the grid cell that encompassed the observation location. For months with two observations, each observation was individually matched to the model output and the skill statistics from those comparisons were averaged for that month. The native horizontal resolution and bathymetry of the individual model grids was preserved in the comparison so as not to bias the analysis through varying interpolation methodologies. For stratification variables, the models and observations were interpolated to a 1 m vertical grid that extended only as deep as the individual models' bathymetry or deepest observation in order to preserve the differences in bathymetric grids while allowing for a direct comparison of the observations to the models. Model–data comparisons at the bottom of the water column were not necessarily based on the same depths, since in many cases the modeled bathymetry was shallower (or at times, deeper) than the deepest data point at a given station. In order to avoid issues with extrapolation and/or grid stretching, data at the bottom of the water column were always compared with model estimates from the deepest grid cell provided by each particular model. Model–data comparisons for stratification and mixed layer depths only included stations and times for which stratification was defined to exist in both the observed and simulated fields.

2.3 Results

An analysis of model skill of the combined temporal and spatial variability of DO at the surface and bottom of the water column, as well as at the observed MLD_O, indicates that all models, regardless of biogeochemical complexity or spatial resolution, exhibit a high degree of skill in reproducing observed DO (Fig. 2.4). Specifically, all models produce DO concentrations at the surface and bottom that have a normalized total RMSD less than 1. The same is true for nearly all models for DO at the observed MLD_O. However, most models underestimate observed DO both at the surface and at the MLD_O (Fig. 2.4a). The correlation between the observed and modeled DO is relatively constant with depth (Fig. 2.4b), though on average slightly higher at the bottom (0.85) than at the surface (0.80). Further, on average, the models simulate DO at the surface and bottom better than they do at the MLD_O. No statistical difference exists between the skill of models that utilize a full biogeochemical component and those that utilize the simple constant-respiration oxygen parameterization. Based on an analysis of variance (ANOVA) comparing the full biogeochemical models to the CRM models, the two model types do not perform differently in terms of their ability to reproduce the combined temporal and spatial variability of bottom DO as measured by total RMSD ($p = 0.48$).

Overall, Model M (the mean of the eight models) consistently performs better than any individual model across all depths examined (Fig. 2.4).

The monthly temporal variability of bottom DO at each station over the 2 years studied is resolved similarly well by all of the models (Fig. 2.5a), but the models have difficulty simulating spatial DO variability during each month (Fig. 2.5b). Due to the stations chosen for this analysis (Fig. 2.2), the spatial variability being examined here is essentially the north to south variability. Most models exhibit a latitudinal gradient with respect to their skill in reproducing the temporal variability of bottom DO, with models overestimating DO at the more northern stations (Fig. 2.5a). Some models differ in their ability to reproduce summer (May–September) DO concentrations and winter (October–April) DO concentrations (Fig. 2.5b). Models B, F, and G all distinctively overestimate mean DO in the summer compared to the winter. In contrast, Models A and C perform similarly well in both seasons (Fig. 2.5b). In addition, all three constant respiration models, as well as Models D and E, substantially underestimate DO at several stations in the winter.

All eight models generally resolve the pycnocline and oxycline with similar skill (Fig. 2.6). All models consistently underestimate the mean and standard deviation of the maximum strength of stratification within the pycnocline and oxycline, defined herein as the maximum vertical gradients of density and oxygen (Fig. 2.6a). All models, except for Model A (see Sect. 2.4.2), also underestimate the mixed layer depth, regardless of whether it is computed in terms of density or oxygen. (Note that these model symbols in Fig. 2.6a are located above the y-axis despite this negative bias in MLD because the vertical coordinate system is oriented upwards.) Thus, the models are producing stratification that is both weaker than observed and higher (shallower) in the water column. The correlation coefficient for these metrics is low, ranging 0.1– 0.6, and indicates that all models are missing the majority of variability associated with the magnitude and location of the pycnocline and oxycline (Fig. 2.6b). However, there is slightly more consistency and better correlation coefficients among the models for the strength of stratification than the depth of the mixed layers.

All eight models are also characterized by similar skill in representing the temporal and spatial variability of density stratification and MLDρ (Fig. 2.7). There is a latitudinal difference in skill of the models in reproducing the magnitude of the pycnocline and MLDρ, with model skill generally lower at the northern stations (Fig. 2.7a). Contrary to the pattern shown for bottom DO (Fig. 2.5b), none of the models exhibit a significant seasonal pattern between summer and winter in reproducing spatial variability of dρ/dz or MLDρ (Fig. 2.7b). However, Model A differentiates itself from the rest of the models in its pattern of skill at reproducing the spatial and temporal variability of the MLDρ (see Sect. 2.4.2). Temporal and spatial patterns for oxycline stratification (dO/dz) and MLD_O closely match those of dρ /dz and MLDρ (not shown).

All eight models reproduce the variability of bottom DO better than the variables that are generally thought of as being the primary drivers of hypoxic conditions, including stratification (Fig. 2.6), salinity, chlorophyll, and nitrate (Fig. 2.8, Table 2.3). However, all models reproduce patterns in temperature across the bay and through time better than any of the other variables in this model comparison (Fig. 2.8). All eight models, as well as the Model Mean, are characterized by very low bias in modeled temperature, and correlation coefficients of approximately 0.99; this high skill results from the very strong and predictable seasonal temperature variability. Even though the five models with full biogeochemical components (Models A, B, C, D, E) are characterized by large differences in their mechanistic approaches to modeling nitrate and chlorophyll, they produce similar total RMSDs for all of the variables examined at both the surface and the bottom (Table 2.3).

The mean of the eight models (Model M) has a higher model skill (lower RMSD) than any individual model across nearly every variable examined (Table 2.3). In addition, for nearly all observations at all stations, the 95% confidence interval of all model hindcasts encapsulates the observed bottom DO concentration (Fig. 2.9), even though any individual model may overestimate or underestimate observed DO. Models generally fall into greater agreement during the summer, when DO is low, and into lesser agreement in the winter when DO is replete. While this study does not allow for a true interannual comparison, it is interesting that at station CB4.1C the model ensemble closely matches the timing of the drawdown of DO in the spring of 2004 (Fig. 2.9), whereas it produces a summer rather than spring initiation of hypoxic conditions in 2005. In addition, the model ensemble produces a premature relaxing of hypoxic conditions for both years at this observation station.

In order to better understand the impact of stratification on DO concentrations throughout the water column, the relationship between the observed pycnocline strength and MLD_ρ were compared to the observed oxycline strength and MLD_O. Observations from 1998 to 2006 demonstrate that while there is not a strong correlation between the strengths of the pycnocline and oxycline, there is a very strong correlation between MLD_ρ and MLD_O (Fig. 2.10). Depending on the criteria used for defining the existence of stratification (see Sect. 2.2.3), the correlation of the pycnocline and oxycline strengths range $r^2 = 0.18–0.26$ and the correlations of MLD_ρ and MLD_O range $r^2 = 0.51–0.82$ (Table 2.4). Furthermore, correlation of the relationship between the MLD_ρ and MLD_O is stronger for more severe stratification (Table 2.4). The relationship between the two mixed layer depths is biased towards the MLD_O being located slightly deeper in the water column than the MLD_ρ. As the cut-off criteria for the existence of stratification becomes more stringent, the relationship becomes closer to 1:1.

2.4 Discussion

2.4.1 How does the skill of various hydrodynamically-based DO models compare?

- In examining the eight 3-D models in this study, there is not a statistical difference between the ability of simple and complex models to simulate the mean and monthly variability of bottom DO; in addition, models with higher spatial resolution do not necessarily produce better estimates of DO.

Models currently simulating hypoxia throughout Chesapeake Bay compute oxygen concentrations in essentially two distinct ways: they either utilize a simple constant respiration model or a full biogeochemical model. In this study, the relative skill of both types of models is compared. Specifically, in examining results of the comparison between five biogeochemical models (A, B, C, D, E) and three simplistic constant respiration models (F, G, H), the two groups of models performed statistically similar in their skill of reproducing bottom DO concentrations (Fig. 2.3, Table 2.3). These results support those of Bever et al. (2013) who compared three constant respiration models with the CBP regulatory model (Model A) and similarly found that all four of the models were equally skillful in terms of reproducing the seasonal variability in bottom DO throughout the bay in 2004 and 2005. Consistent with the results of Scully (2013), this result implies that the seasonal variability of DO in the Chesapeake Bay is primarily dependent on underlying hydrodynamic mechanisms which are nearly identical for all eight models, rather than on aspects related to the biogeochemical cycling which vary dramatically between models and in fact are constant in three of the eight models. It should be noted, however, that the 2 years studied here were relatively wet years and an analysis of dry years may offer different results.

Many previous studies have examined the costs and benefits of adding complexity to biogeochemical models. For example, increasing biogeochemical complexity has been found to improve skill in some biogeochemical data assimilative parameter optimization studies (Friedrichs et al., 2006, 2007; Lehmann et al., 2009; Bagniewski et al., 2011; Ward et al., 2013; Xiao and Friedrichs, 2014). The additional parameters associated with increased complexity generally provide more parameters that are available for additional tuning and subsequent improved model–data agreement. This is in contrast to the results of this analysis demonstrating that increased biogeochemical complexity does not necessarily improve model–data agreement. In this case, the increase in model complexity has likely outpaced the ability of the researchers to fully tune the model to the available observations. However, even past studies that have invoked formal parameter optimization methodologies, such as genetic algorithms and variational adjoint methods (Friedrichs et al., 2007; Ward et al., 2010; Xiao and Friedrichs, 2014), have found that under certain conditions, adding too much complexity does not necessarily improve model skill and in fact can decrease model skill and portability, primarily due to artifacts resulting from overtuning. This mirrors findings from the larger ecosystem modeling community where the best-fit models are often those with intermediate complexity (Fulton et al., 2003).

In this study, horizontal grid resolution differed significantly between model implementations, with the most highly resolved grid (Model G) including more than 9 times more grid cells than the lower resolution grids (Table 2.1). A certain degree of resolution is clearly required to successfully simulate dynamic processes, and a model with 8–10 km resolution will not be able to correctly simulate the hydrodynamic processes within the bay (Feng et al., 2015). However, an increase in horizontal grid resolution from ~1.8 to ~0.6 km, which results in a run-time change of a factor of 9, or possibly of 27 if the time step is accordingly decreased by a factor of 3, does not necessarily result in a significant improvement in simulation skill of either stratification or bottom oxygen. Although not shown here, additional sensitivity experiments with Model G revealed that doubling the vertical resolution of this model had no significant effect on the model's ability to resolve the depth of stratification or the maximum magnitude of stratification. Thus, when selecting the optimal model resolution for a simulation, it is critical to weigh the advantages of increased resolution with the increased time required for simulation. With a given level of computational resources, fewer sensitivity experiments can be conducted with a model using a more highly resolved grid.

Accurately simulating the observed spatial variability of DO (Fig. 2.4b) was a greater challenge than simulating the temporal variability of DO (Fig. 2.4a) for all eight models participating in this intercomparison. This is especially true in the winter months when the vast majority of the bay is oxygen replete and the models have difficulty representing the observed variability from station to station. The majority of the models tend to slightly overestimate mean bottom DO in the summer whereas multiple models (e.g., Models D, E, F, and G) exhibit a strong negative bias during January and/or February of 2005, primarily at stations in the middle to southern portion of the bay's deep channel. Interestingly, increased biological complexity and higher grid resolution do not completely resolve this issue, as this is true for models utilizing full biogeochemical models (Models D, E) as well as those using highly resolved model grids (Model G). This is likely due to the ephemeral nature of the biological divers of DO.

The strong performance of the constant respiration models implies that these models may be excellent candidates for providing short-term bottom oxygen forecasts. The high DO skill of the CRM models primarily results from the fact that seasonal variations in physical processes (primarily wind mixing and temperature) play a dominant role in controlling the seasonal cycle of oxygen (Scully, 2013). Because the underlying hydrodynamic models all use similar physical forcing, the constant respiration models are able to simulate the seasonal cycle of DO with similar skill as the more complex biogeochemical models. As a result, these simple models that are easier to tune and require less in the way of computational resources than full biogeochemical models, may be efficiently used to produce short-term (on the order of days) DO forecasts. On the contrary, the more complex full biogeochemical models will be necessary for scenario-based and long-term (on the order of months to years) forecasting which requires that models respond to prescribed changes in the biogeochemical environment, such as increased rates of nutrient loading due to changes in land use, land cover, and/or climate.

2.4.2 How does model skill of DO compare to that of the primary drivers of DO variability?

- Overall, model DO skill is greater than that of the variables generally considered to drive DO variability, such as stratification, salinity, mixed layer depth, chlorophyll, and nitrate; only modeled temperature has higher skill than modeled DO.

Since dissolved oxygen concentrations in the Chesapeake Bay are controlled by physical processes (e.g., advection, wind mixing, heating/cooling, and stratification), as well as biological processes (e.g., photosynthesis and respiration), it is critical to understand the skill of the models in terms of how well they reproduce the many factors influencing oxygen concentrations. As expected, the five models containing a specific biogeochemical model component had more difficulty simulating the observed chlorophyll and nitrate concentrations than the physical variables (temperature and salinity), both at the surface (Table 2.3) and the bottom (Fig. 2.8). Replicating the correct location, magnitude, and timing of phytoplankton blooms and nutrient cycling is a complex issue, and as a result, these features are generally not well simulated in the models. While the models generally simulate the total amount of chlorophyll adequately, it is more uniformly spatially distributed in the models rather than in patchy blooms as in nature, leading to the underestimation of chlorophyll variability across all models. Although all models produced a relatively high correlation between observed and modeled temperature and salinity (Fig. 2.8), the correlation coefficients for chlorophyll and nitrate were much lower. The correlations for observed vs. modeled DO was more similar to that of the physical variables (temperature, salinity) than the biological variables (chlorophyll and nitrate), highlighting that the seasonal variability in bottom DO is regulated more by physical than biological factors. This also explains the success of the constant respiration models, which by definition contain no biological variability yet reproduce DO variability nearly as well as the most complex biogeochemical models.

In this study, model skill was also considerably higher for bottom oxygen than it was for the vertical gradient of stratification and mixed layer depths (Figs. 2.6, 2.8). The underestimation of the vertical gradient across all models is largely due to the numerical diffusion that characterizes all of these hydrodynamic models, but may also be partially due to an underestimation of the winds or a lack of diffuse freshwater input around the bay. Even though the models all underestimated the strength of stratification (Figs. 2.4, 2.6), modeled stratification in summer was strong enough to prevent mixing with the relatively well-oxygenated surface waters. This result suggests, somewhat surprisingly, that simulating the correct vertical gradient of stratification is not absolutely necessary for skillful bottom DO simulations. Models need only simulate enough stratification to effectively cut off vertical mixing in order to develop an isolated bottom layer that can then experience a draw down in oxygen via respiration. In addition, the models must also correctly simulate the horizontal advection of oxygen (Scully, 2013; Y. Li et al., 2015). The fact that bottom DO is simulated so well by the eight models analyzed here suggests that not only is the advection of oxygen well represented in the models but also the strength of stratification, i.e., the maximum vertical gradients of density and oxygen, produced by these models is sufficient. Thus, although novel and somewhat unexpected, these results are not contradictory to previous studies demonstrating the importance stratification plays in initiating summer hypoxia in the Chesapeake Bay (Murphy et al., 2011).

Model skill in terms of reproducing observed mixed layer depths was likewise much lower than model skill of reproducing observed oxygen concentrations. All models, except Model A, produced mixed layer depths (MLD_O and $MLD\rho$) that were generally too shallow in the water column (Fig. 2.6a). Note that Model A is a regulatory model that has been used for many years by the Chesapeake Bay Program, and has thus undergone more extensive calibration aimed at matching the mean salinity and oxygen characteristics of the bay (Cerco and Cole, 1993). Furthermore, Model A employs a z grid that matches the bathymetry in trench areas better than the sigma grids used by the other models. Although Model A produced mixed layer depths that were generally in the correct location within the water column (Fig. 2.6a), they were too variable (Fig. 2.6b). This variability may partly be a result of the 1.5 m z grid employed by Model A causing large jumps between vertical grid cells and hence resulting in overestimates of MLD variability. All other models use sigma grids typically with more highly resolved vertical resolution at the depth of maximum stratification.

The two variables for which the models have greatest skill are DO and temperature (Fig. 2.8). This is because oxygen variability is driven primarily by seasonal variability in physical processes such as solubility and wind mixing and to a lesser degree by variability in oxygen consumption (Scully, 2013). As a result, the models using a constant mean respiration rate produce as realistic hypoxia simulations as the biogeochemically complex models. Observations clearly show this strong seasonal variability in bottom DO (Fig. 2.11a) and, to a slightly lesser extent, clear seasonal variability in DO at the bottom of the bottom of the oxygen mixed layer (MLD_O; Fig. 2.11b). However, a seasonal cycle is not manifested in the MLD_O itself (Fig. 2.11c). The lack of such a strong seasonal cycle in the observed mixed layer depths makes this a more difficult variable for the models to simulate. As a result, the models can relatively skillfully simulate the combined spatial and temporal variability of DO while simultaneously missing the MLD_O.

2.4.3 Why is it important for DO models to simulate the MLD_O correctly?

- Most of the aerobic habitat in the bay during the summer is located above the MLD_O, thus it is critical for living resource managers to use models that accurately simulate this variable.

On average, the models miss the observed depth of the MLD_O by 3.4 m, which equates to roughly a 60% error in the modeled mixed layer depths. While the models have difficulty simulating the MLD_O throughout the entire year (Figs. 2.6, 2.7b), the summer months are when the mismatch has the greatest potential to impact the available habitat for oxygen-dependent species. Each year during this time period low-oxygen waters occupy nearly the entire water column below the mixed layer. At station CB4.1C, a representative mesohaline deep trough station, the contours of low-oxygen (5 mg L^{-1}) and hypoxic (2 mg L^{-1}) waters are located just below the MLD_O from late spring until late fall (Fig. 2.12). The severe depletion of oxygen below the mixed layer compresses the habitable space at this station to roughly 10 m (from a maximum of 32 m) during the annual low-oxygen event.

The impact of habitat compression can be substantial, as many bay species require DO concentrations well above the traditional hypoxic threshold (USEPA, 2010). While not all of the main stem stations develop hypoxic water each year, most mesohaline stations experience a dramatic drawdown of oxygen to levels during the summer that effectively remove a large portion of the bay from habitable space (Murphy et al., 2011; Schlenger et al., 2013). Studies have shown that some species modify their behavior based on the oxycline depth, which acts to constrict the habitable space in the water column (Prince and Goodyear, 2006; Pierson et al., 2009; Elliott et al., 2013). Since species can be negatively impacted by low-DO concentrations as high as 5 mg L^{-1} (Breitburg, 2002; Vaquer-Sunyer and Duarte, 2008; USEPA, 2010), the location of the oxycline is not only important for habitat compression in the summer months but can also be important in the winter months when an occasional lack of vertical mixing can substantially decrease bottom DO concentrations. Furthermore, in order to accurately estimate hypoxic volume, models must correctly simulate the depth of the mixed layer, since the MLD_O closely follows the depth of the 2 mg L^{-1} contour.

2.4.4 How can DO simulations in the Bay be improved for management of water quality and living resources?

- To better simulate DO conditions and summer habitat compression due to low-DO water, simulations of the depth of the top of the pycnocline ($MLD\rho$) must be improved.

Although the suite of models examined reproduce DO concentrations relatively well overall (Fig. 2.4), the models typically overestimate summer habitat compression by producing low-DO concentrations too high in the water column (Fig. 2.6). Observations from the Chesapeake Bay Program show a strong correlation between the depths of the oxygen and density-defined mixed layers (Fig. 2.10b). The models analyzed here also clearly exhibit a close relationship between their skill in simulating the depths of the oxygen and density-defined mixed layers (Fig. 2.6). These strong relationships between the depths of the oxygen and density-defined mixed layers result from the fact that the pycnocline represents the physical barrier that leads to the development of the oxycline. Therefore, the inability of the models to accurately simulate habitat compression is an artifact of their lack of skill in simulating the depth of the density-defined mixed layer. In contrast, the strength of density stratification is not well correlated to the strength of oxygen stratification. This is because a relative wide range of intensities of density stratification is still sufficient to cut off vertical mixing, leading to the observed draw-down in bottom DO. Thus, even though all models underestimate the strength of the pycnocline, they still produce enough stratification to greatly reduce mixing. The results from this paper thus indicate that to further improve DO simulations and better estimate summertime habitat compression, it is even more critical for models to accurately simulate the depth of the top of the pycnocline than to accurately simulate the absolute strength of the pycnocline.

2.4.5 What is the utility of the multi-model ensemble and Model Mean?

- The multi-model ensemble approach allows for the development of a model mean, which taken as its own model, is the most skilled model when examining the combined suite of variables analyzed in this study.

The model skill assessment presented here demonstrates that the average of all eight models, or five models in the case of chlorophyll and nitrate, does better than any individual model if looking across the suite of variables analyzed. This finding is similar to that of other studies that examined the value of the Model Mean from a multi-model ensemble (e.g., Gneiting and Raftery, 2005; Hagedorn et al., 2005). While the concept of using a multi-model ensemble has been most extensively employed by atmospheric, climatic, and global circulation modelers, such as the Intergovernmental Panel on Climate Change (e.g., Collins et al., 2013), the tool's utility for aquatic ecosystem modeling is gaining traction (Meier et al., 2012; Trolle et al., 2014; Janssen et al., 2015). As models are increasingly used in regulatory decisions regarding aquatic ecosystems, a cohort of similarly skilled models can be used to help inform a set of confidence bounds around an environmental forecast. Due to the restrictions placed on models used in regulatory actions, utilization of a multi-model ensemble may not be realistic for all environmental and resource managers; however, multiple models can be integrated into the decision-making process even when the ultimate decision must be based on a single model. For example, a confidence interval plot could help identify where regulatory model output might be acting out of sync with other skilled water quality models of the same system, thereby informing managers of the potential shortfalls associated with the regulatory model. Furthermore, if the models tend to be predicting similar DO concentrations, a cohort of models could enhance the confidence in regulatory decisions based on a single regulatory model (Friedrichs et al., 2012; Weller et al., 2013). Comparing multiple models can also help inform how to better improve models in the future, as this study has aimed to do.

2.5 Conclusions

All models analyzed here exhibited a high degree of skill in simulating dissolved oxygen concentrations within the main stem of the Chesapeake Bay in 2 years corresponding to relatively wet and average years. Their high skill results from the fact that physical processes (e.g., solubility, wind-mixing, and advection) exert a first order influence on the seasonal cycle of oxygen. As a result, the models' ability to reproduce dissolved oxygen concentrations is independent of the complexity of the biogeochemical parameterizations: the simplest constant respiration models were found to reproduce observed oxygen concentrations as well as the most biologically complex models. Essentially, all models are equally capable of respiring most of the available oxygen in the lower water column during summer.

This study also suggests that for use as management tools for water quality and living resources, it is more critical for these models to adequately resolve the depth of the mixed layer than the absolute strength of stratification (as long as modeled stratification is strong enough to limit vertical mixing). This is critical because observations show that during warmer months, oxygen-depleted water fills the water column to where stratification limits further mixing, which effectively cuts off waters below the mixed layer for use by the majority of the Chesapeake Bay's most recognized and valued living resources. These results furthermore suggest that modelers should focus their efforts on improving the hydrodynamics of their models in an effort to improve simulations of mixed layer depth dynamics and variability.

These findings have significant ramifications for short-term bottom DO forecasts, which may be successful with very simple oxygen parameterizations embedded in hydrodynamic models. In contrast, scenario-based water quality forecasts are likely to benefit from more complex models, which must adequately reproduce the longer-term response of the oxygen field to changes in nutrient and organic matter loads. This study also helps to demonstrate how multiple community models from governmental agencies and academic institutions may be used together to provide a model mean and a set of confidence bounds for regulatory model results that could be used to inform management decisions.

Tables

Table 2.1: Model characteristics.

Model	A	B	C	D	E	F	G	H
Hydrodynamic model-DO model	CH3D-ICM	ChesROMS-ECB	ChesROMS-BGC	ROMS-RCA	FVCOM-ICM	ChesROMS-CRM	CBOFS-CRM	EFDC-CRM
Grid structure	Structured	Structured	Structured	Structured	Unstructured	Structured	Structured	Structured
Average wet-cell resolution	1 km	1.8 km	1.8 km	1.89 km	1.26 km	1.8 km	0.565 km	1.2 km
Vertical grid	1.52 m	20 sigma	20 sigma	20 sigma	10 sigma	20 sigma	20 sigma	20 sigma
River forcing	CBP Watershed Model	DLEM Watershed Model	USGS Data	USGS Data	USGS Data	USGS Data	USGS Data	USGS Data
Sub-tidal elevation at open boundary	Multiple efforts	Lewes, DE to Duck, NC	Lewes, DE to Duck, NC	Wachapreague, VA to Duck, NC	TPXO Tidal Model	Lewes, DE to Duck, NC	Ocean City, MD to Duck, NC	Lewes, DE to Duck, NC
Wind forcing	Multiple efforts	Thomas Point Light	NARR	NARR	NARR	NARR	NARR & NDBC buoys	NARR
Other atmospheric forcing	Multiple efforts	NARR	NARR	NARR	NARR	NARR	NARR	Norfolk & Baltimore Airports
Biogeochemical complexity	High; 5 phytoplk. groups	High; 1 phytoplk. group	High; 1 phytoplk. group	High; 2 phytoplk. groups	High; 3 phytoplk. groups	Low; constant respiration	Low; constant respiration	Low; constant respiration
Model citation	Cerco et al., 2010	Feng et al., 2015	Brown et al., 2013	Testa et al., 2014	Jiang and Xia, 2015	Scully, 2013	Lanerolle et al., 2011	Du and Shen, 2015

Table 2.2: Characteristics of observation stations (from USEPA, 2012).

Station	Latitude	Longitude	Station Depth	# of Cruises
CB3.2	39.1634 N	76.3063 W	12.1 m	34
CB3.3C	38.9951 N	76.3597 W	24.3 m	34
CB4.1C	38.8251 N	76.3997 W	32.3 m	34
CB4.2C	38.6448 N	76.4177 W	27.2 m	34
CB4.3C	38.5565 N	76.4347 W	26.9 m	34
CB4.4	38.4132 N	76.3430 W	30.3 m	34
CB5.1	38.3185 N	76.2930 W	34.1 m	34
CB5.2	38.13678N	76.2280 W	30.6 m	34
CB5.4	37.8001 N	76.1747 W	31.1 m	26
CB6.2	37.4868 N	76.1563 W	10.5 m	30
CB6.4	37.2365 N	76.2080 W	10.2 m	29
CB7.1	37.6835 N	75.9897 W	20.9 m	27
LE2.3	38.0215 N	76.3477 W	20.1 m	34

Table 2.3: Mean and standard deviation (STD) of observations and total normalized RMSD for each model.

	Mean ± STD of Obs	Normalized RMSD								
		A	B	C	D	E	F	G	H	M
Surface Temp. (ºC)	17.44±8.82	0.13	0.13	0.12	0.09	0.13	0.13	0.16	0.19	0.10
Bottom Temp. (ºC)	15.75±8.02	0.24	0.35	0.35	0.23	0.22	0.35	0.17	0.19	0.23
Surface Salinity (PSU)	10.92±4.32	0.37	0.62	0.53	0.36	0.46	0.61	0.57	0.41	0.35
Bottom Salinity (PSU)	18.17±3.14	0.72	0.85	0.73	1.55	1.28	0.78	1.03	0.97	0.75
Max. dρ/dz (kg m^{-4})	~1.64±1.15	1.03	1.09	1.07	1.09	1.25	1.01	1.23	1.02	N/A
MLDρ (m)	~5.32±3.99	1.01	1.13	1.11	1.41	1.39	1.12	1.38	1.13	N/A
Surface DO (mg L^{-1})	9.74±2.15	0.67	0.58	0.89	0.80	1.00	0.63	0.64	0.69	0.57
DO at MLD_O (mg L^{-1})	~8.44±2.53	0.54	0.57	0.74	0.93	0.83	0.81	0.95	1.09	0.62
Bottom DO (mg L^{-1})	4.42±3.61	0.51	0.59	0.81	0.61	0.54	0.46	0.61	0.60	0.46
Max. dDO/dz (mg L^{-1} m^{-1})	~1.81±1.12	1.19	1.21	1.34	1.09	1.35	1.12	1.23	1.19	N/A
MLDo (m)	~6.62±4.01	1.24	1.01	1.10	1.33	1.33	1.05	1.30	1.29	N/A
Surface Chl *a* (mg m^{-3})	11.19±9.04	0.92	1.22	1.60	1.23	0.89	N/A	N/A	N/A	1.16
Bottom Chl *a* (mg m^{-3})	9.02±11.52	0.87	1.10	1.07	1.05	1.01	N/A	N/A	N/A	0.90
Surface Nitrate (mmolN m^{-3})	0.32±0.33	0.61	0.79	1.03	0.61	0.52	N/A	N/A	N/A	0.79
Bottom Nitrate (mmolN m^{-3})	0.12±0.13	1.08	1.38	1.38	0.92	1.46	N/A	N/A	N/A	0.85

Table 2.4: Pycnocline and oxycline correlation statistics (all correlations have p-values << 0.01).

Stratification Threshold Percentage	Max $d\rho/dz$ vs. Max dO/dz	$MLD\rho$ vs. MLD_O	Profiles with Stratification
10%	0.18	0.51	1613
15%	0.22	0.59	1303
20%	0.22	0.70	916
25%	0.26	0.82	575

Figures

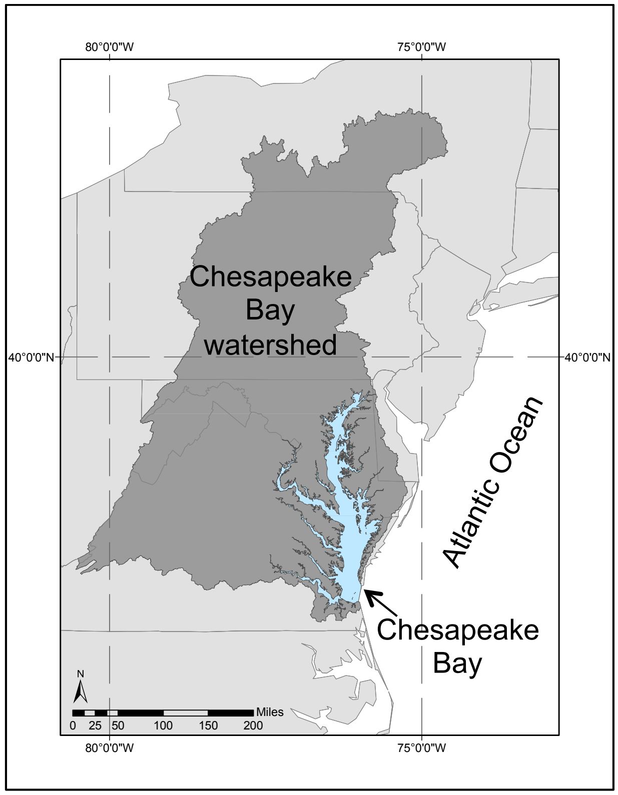


Figure 2.1: Map of the Chesapeake Bay and its watershed.

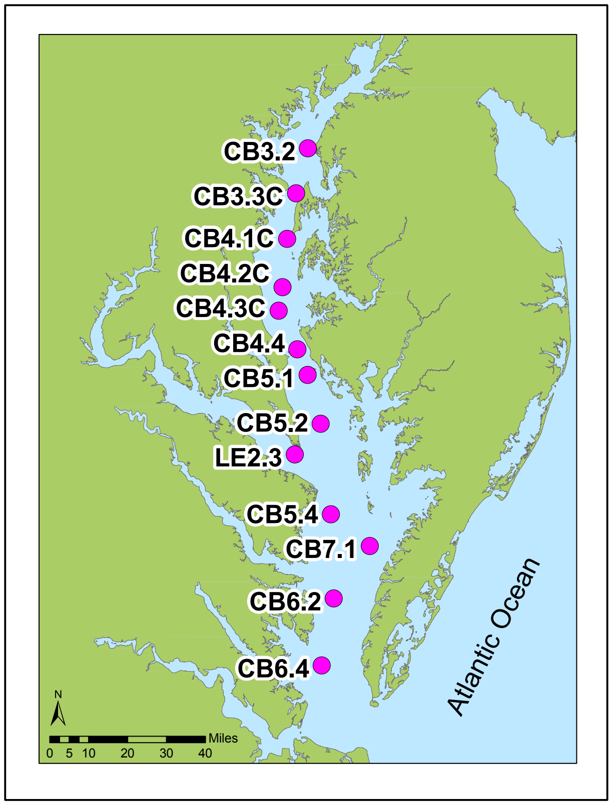


Figure 2.2: Location of the CBP Water Quality Monitoring stations used in this study.

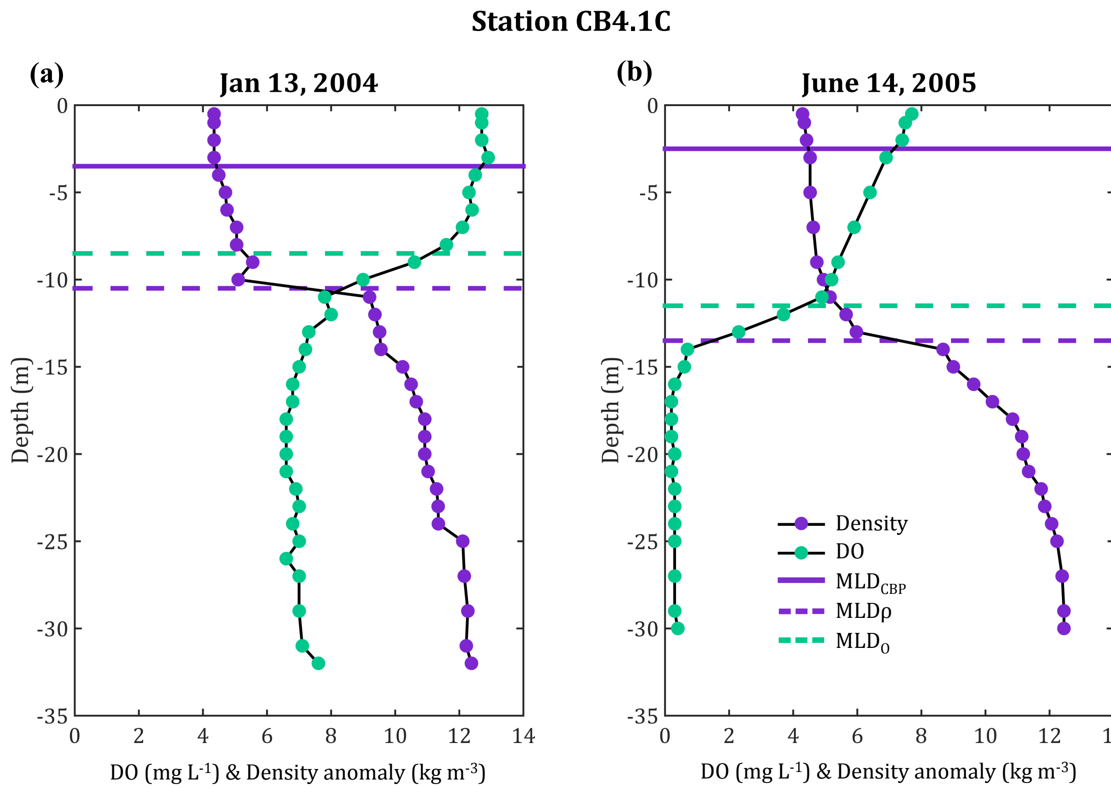
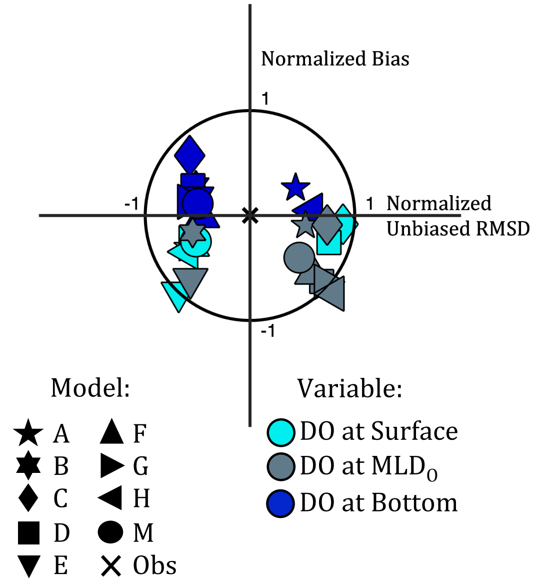


Figure 2.3: Density and dissolved oxygen profiles for a mid-Bay station (CB4.1C) on (a) January 13, 2004 and (b) June 14, 2005, comparing the 0.1 kg m^{-4} stratification definition used by the CBO (MLD_{CBP}) with the 10% threshold definitions used here for density (MLDρ) and oxygen (MLD_O).

(a)



(b)

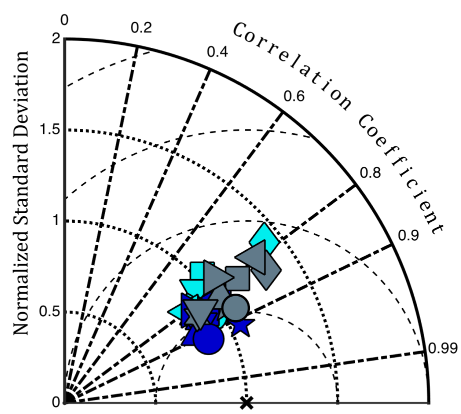


Figure 2.4: Normalized summary (a) target and (b) Taylor diagrams illustrating model skill of dissolved oxygen at the surface, MLD$_O$, and bottom for 13 Chesapeake Bay stations in 2004-2005. The “x” represents the skill of a model that perfectly reproduces the observations. The dotted, dashed-dot, and dashed lines on the Taylor diagram represent lines of constant standard deviation, correlation coefficient, and unbiased RMSD, respectively.

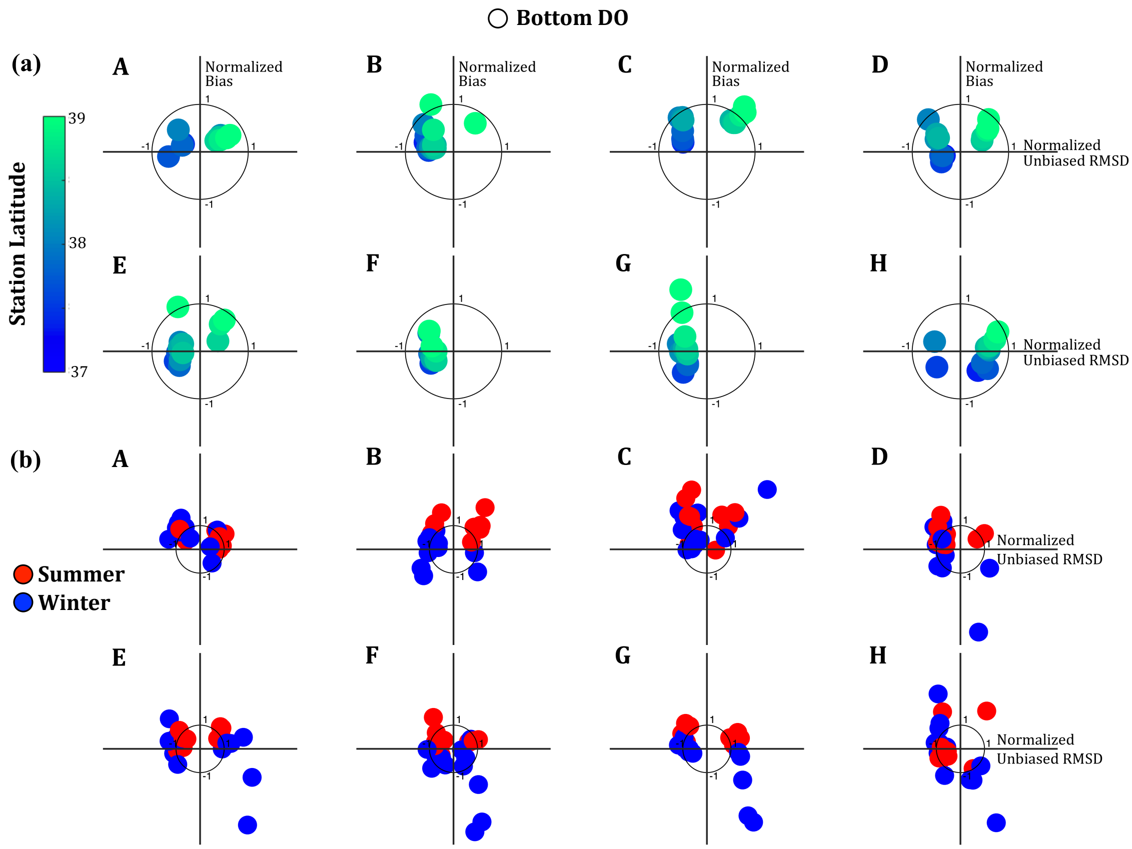
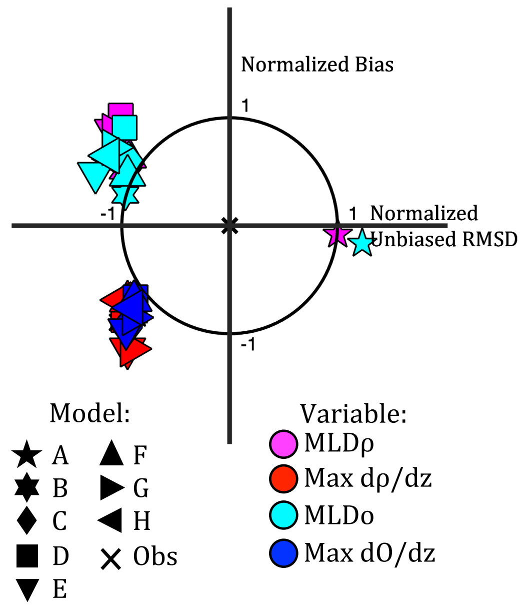


Figure 2.5: Normalized target diagrams for Models A-H demonstrating the (a) temporal and (b) spatial skill in resolving the variability of bottom dissolved oxygen concentrations. In (a), the individual dots represent the 12 stations along the main stem of the Chesapeake Bay. In (b), the dots represent the 24 months of 2004-2005 and are delineated by color: red = summer (May – September) and blue = winter (October – April).

(a)



(b)

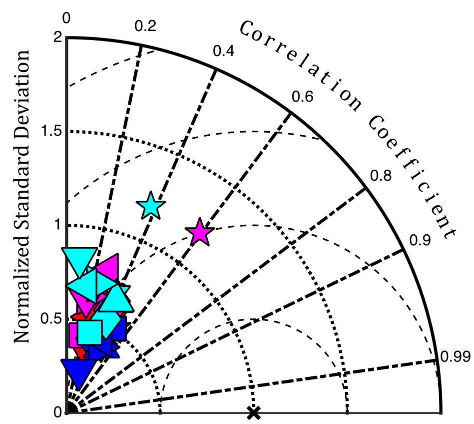


Figure 2.6: Normalized summary (a) target and (b) Taylor diagram illustrating model skill of MLDρ, and MLD_O, max dρ/dz, and max dO/dz at 13 Chesapeake Bay stations for 2004-2005. The "x" represents the skill of a model that perfectly reproduces the observations. Since RMSD of stratification is only computed at stations where both the observations and model exhibit stratification, the Model Mean is not calculable for these variables.

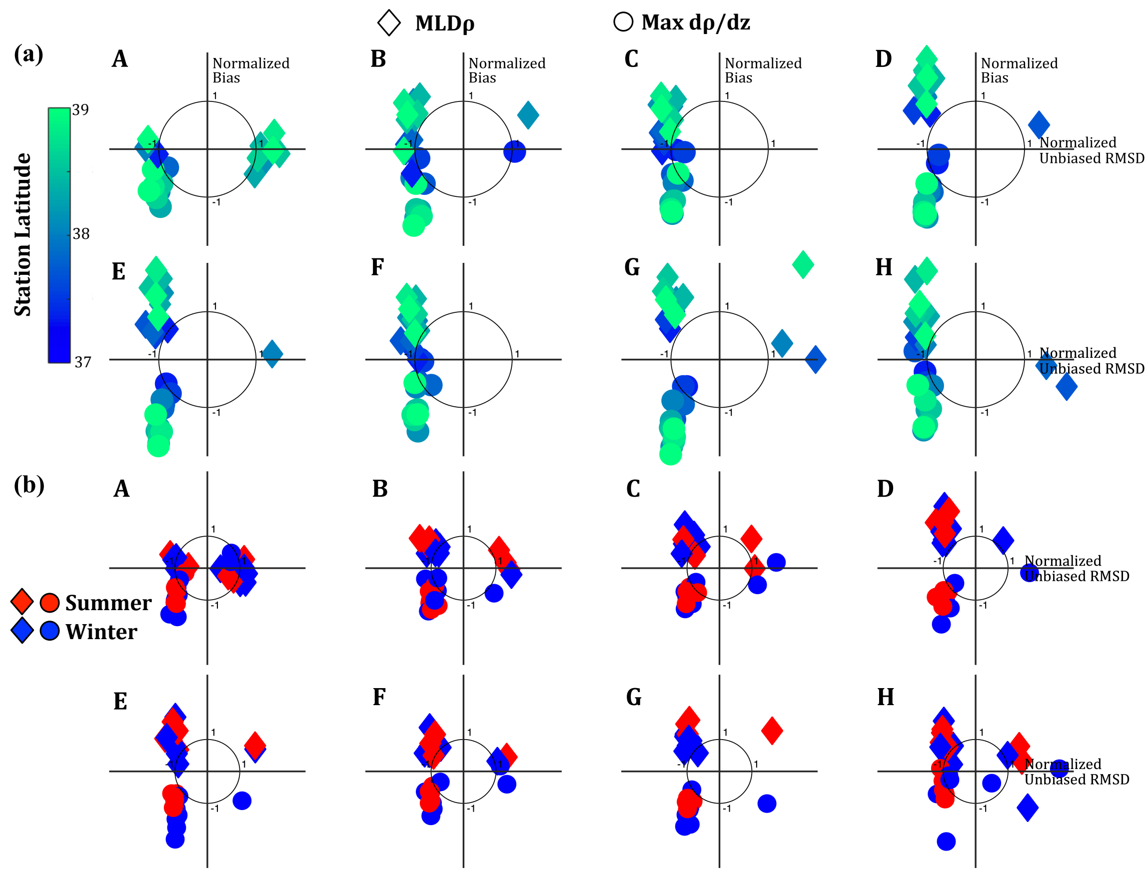


Figure 2.7: Normalized target diagrams for Models A-H demonstrating the (a) temporal and (b) spatial skill in resolving the variability of the strength of density stratification (circles) and the depth of pycnocline initiation (diamonds). In (a), the individual dots represent the 13 stations along the main stem of the Chesapeake Bay. In (b), the dots represent the 24 months of 2004-2005 and are delineated by color: red = summer (May-September) and blue = winter (October-April).

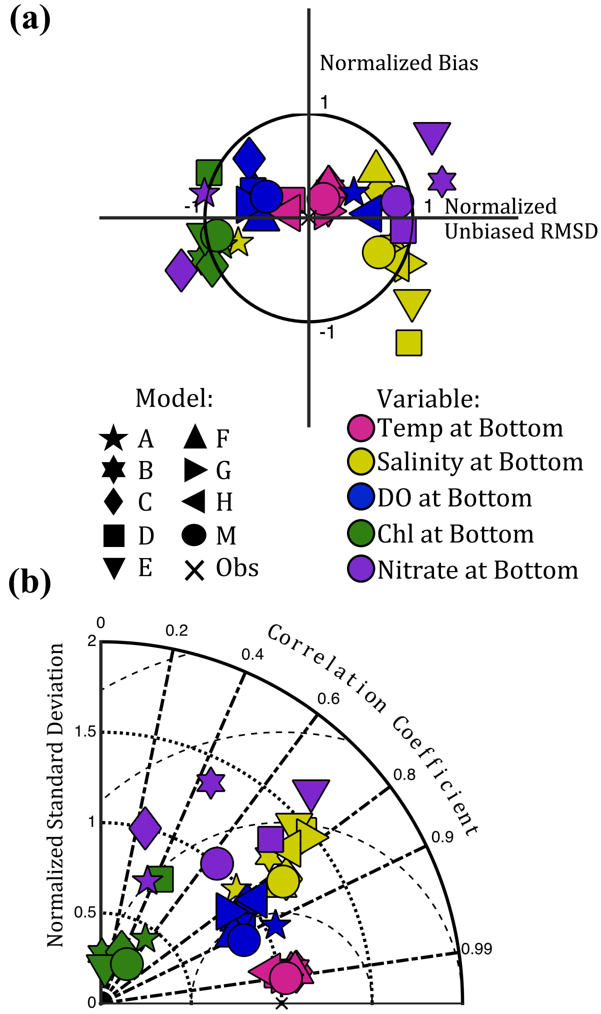


Figure 2.8: Normalized summary (a) target and (b) Taylor diagram illustrating model skill of bottom temperature, salinity, chlorophyll, nitrate, and dissolved oxygen at 13 Chesapeake Bay stations for 2004-2005. The "x" represents the skill of a model that perfectly reproduces the observations.

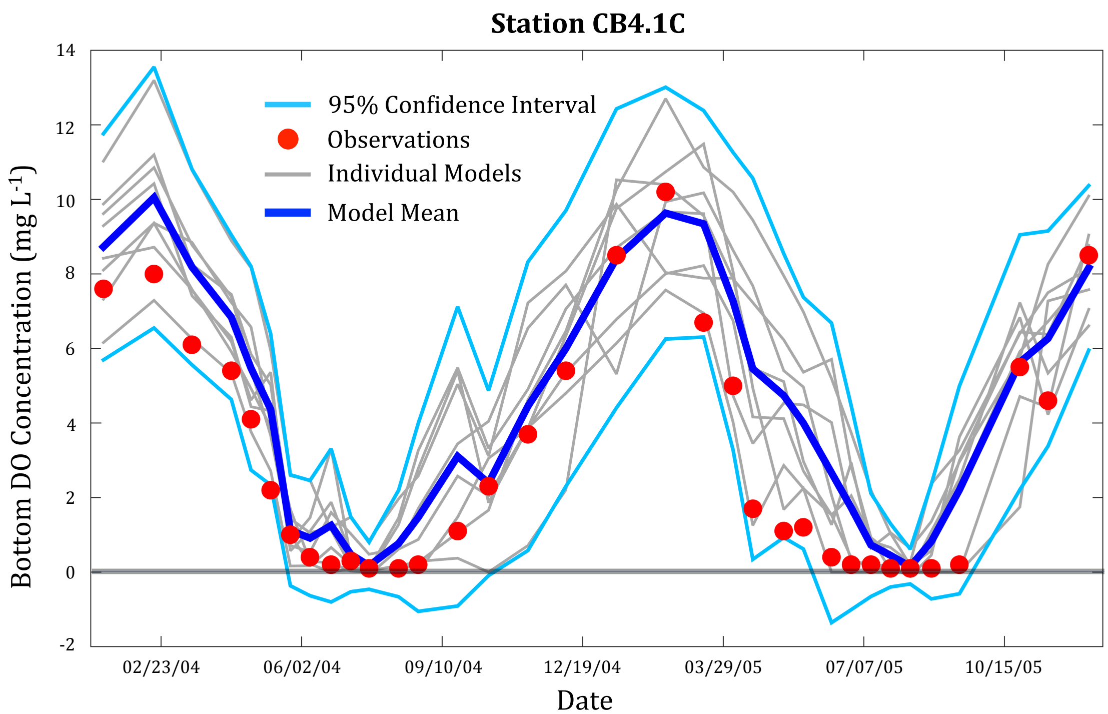


Figure 2.9: Time series of bottom dissolved concentrations for station CB4.1C. Red dots represent the 34 observations made during 2004-2005. Grey lines are the individual model simulations. The dark blue line represents the model mean while the cyan line represents the 95% confidence interval of the model simulations.

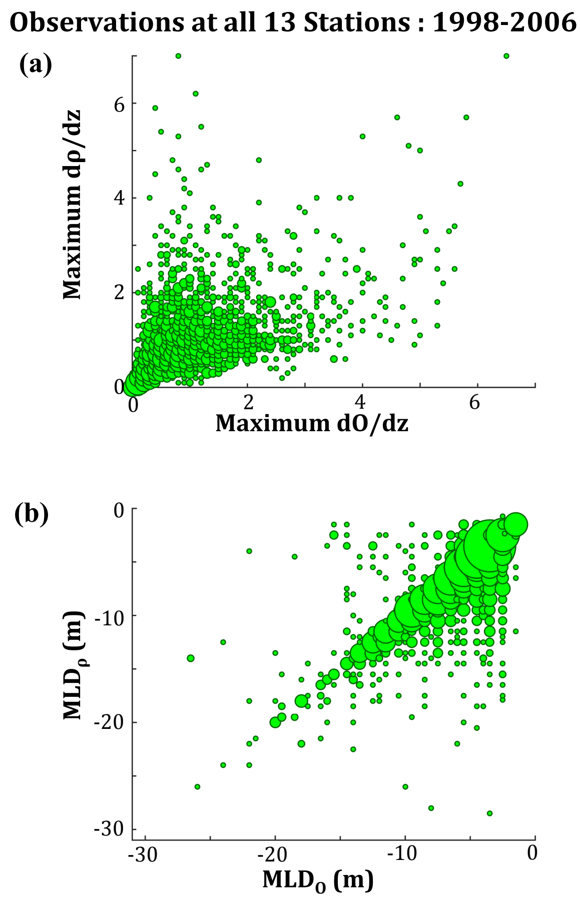


Figure 2.10: Scatter plots comparing observations of (a) the strengths of stratification of the pycnocline and oxycline and (b) the oxygen- and density-defined mixed layer depths. Size of the circles is proportional to the number of observations. Observations are from 1998-2006 at the 13 Chesapeake Bay stations shown in Figure 2.

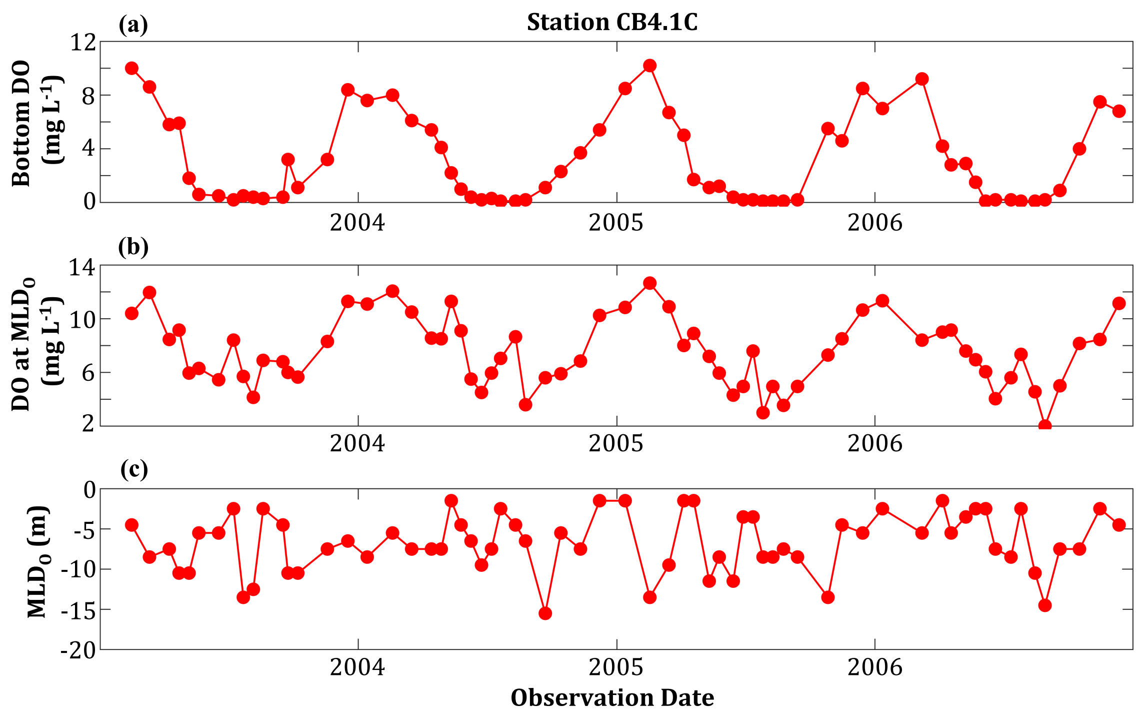


Figure 2.11: Time series of observations at Station CB4.1C from 2003 – 2006 for (a) bottom dissolved oxygen, (b) dissolved oxygen at the MLD_O, and (c) MLD_O.

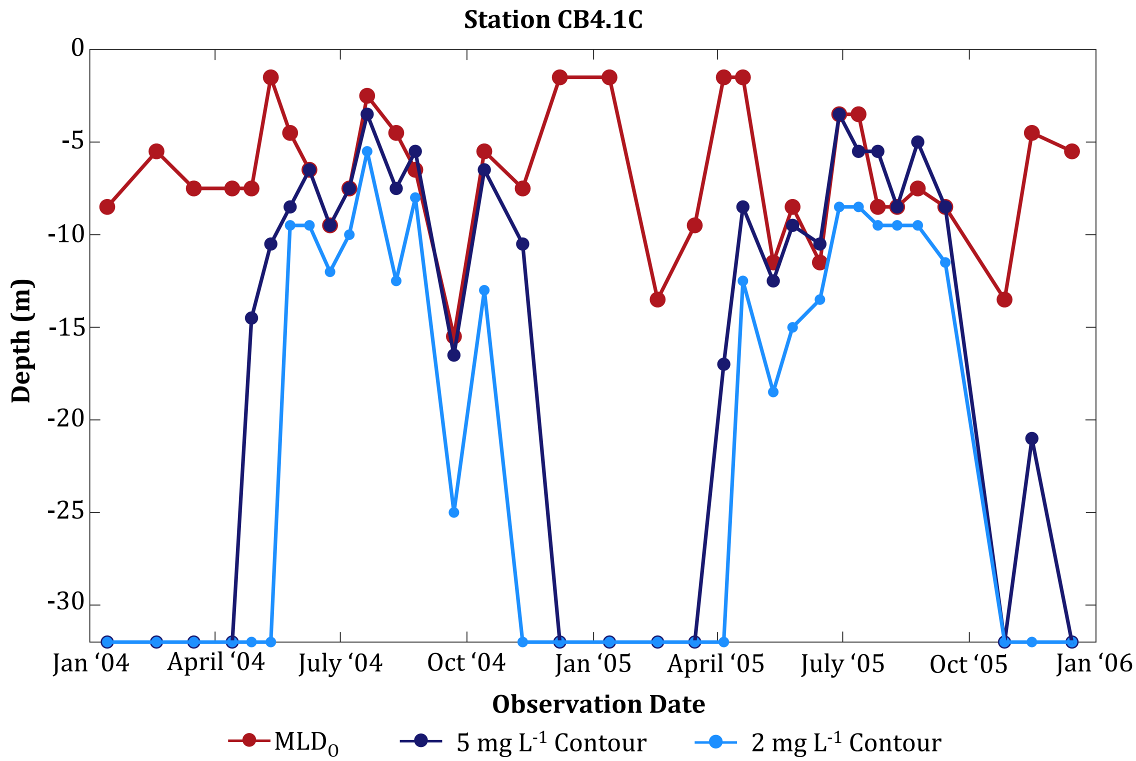


Figure 2.11: Time series of observations of dissolved oxygen and MLD$_O$ contours at Station CB4.1C for 2004 and 2005.

Acknowledgements

This work was supported by the NOAA IOOS program as part of the Coastal Ocean Modeling Testbed. We thank Yun Li and Younjoo Lee for help with the ROMS-RCA simulations used in this analysis and Ray Najjar for his insights and comments. This is VIMS contribution 3520 and UMCES contribution 5130.

Literature Cited

Bagniewski, W., Fennel, K., Perry, M. J., and D'Asaro, E. A.: Optimizing models of the North Atlantic spring bloom using physical, chemical and bio-optical observations from a Lagrangian float, Biogeosciences, 8, doi: 10.5194/bg-8-1291-2011, 2011.

Bever, A. J., Friedrichs, M. A. M., Friedrichs, C. T., Scully, M. E., and Lanerolle, L. W.: Combining observations and numerical model results to improve estimates of hypoxic volume within the Chesapeake Bay, USA, J. Geophys. Res-Oceans, 118, 4924-4944, doi:10.1002/jgrc.20331, 2013.

Boesch, D. F., Brinsfield, R. B., and Magnien, R. E.: Chesapeake Bay Eutrophication: Scientific Understanding, Ecosystem Restoration, and Challenges for Agriculture, J. Environ. Qual., 30, 303-320, 2001.

Breitburg, D.: Effects of Hypoxia, and the Balance between Hypoxia and Enrichment, on Coastal Fishes and Fisheries, Estuaries, 25, 767-781, 2002.

Breitburg, D. L., Loher, T., Pacey, C. A., and Gerstein, A.: Varying effects of low dissolved oxygen on trophic interactions in an estuarine food web, Ecol. Monogr., 67, 489-507, 1997.

Brown, C. W., Hood, R. R., Long, W., Jacobs, J., Ramers, D. L., Wazniak, C., Wiggert, J. D., Wood, R., and Xu, J.: ecological forecasting in Chesapeake Bay: Using a mechanistic-empirical modeling approach, J. Marine Syst., 125, 113-125, doi:10.1016/j.jmarsys.2012.12.007, 2013.

Buchheister, A., Bonzek, C. F., Gartland, J., and Latour, R. J.: Patterns and drivers of the demersal fish community of Chesapeake Bay, Mar. Ecol. Prog. Ser., 481, 161-180, doi:10.3354.meps10253, 2013.

Cerco, C. F., and Cole, T.: Three-dimensional eutrophication model of Chesapeake Bay, J. Environ. Eng-ASCE, 119, 1006-10025, 1993.

Cerco, C. F., and Noel, M. R.: The 2002 Chesapeake Bay Eutrophication Model, EPA 903-R-04-004, US Army Corps of Engineers, Waterways Experiment Stations, Vicksburg, MS, 2004.

Cerco, C. F., and Noel, M. R.: Twenty-one-year simulation of Chesapeake Bay water quality using the CE-QUAL-ICM eutrophication model, J. Am. Water Resour. As. 49, 1119-1133, doi:10.1111/jawr.12107, 2013.

Cerco, C., Johnson, B., and Wang, H.: Tributary refinements to the Chesapeake Bay model, ERDC TR-02-4, US Army Engineer Research and Development Center, Vicksburg, MS, 2002.

Cerco, C., Kim, S-C., and Noel, M.: The 2010 Chesapeake Bay Eutrophication Model – A Report to the US Environmental Protection Agency Chesapeake Bay Program and to The US Army Engineer Baltimore District, US Army Engineer Research and Development Center, Vicksburg MS, 2010.

National Oceanic and Atmospheric Administration: Chesapeake Bay Operational Forecast System (CBOFS), US Department of Commerce, http://tidesandcurrents.noaa.gov/ofs/cbofs/cbofs.html, last access: December 2015.

USGS: Chesapeake Bay Program Water Quality Database (1984-present): http://www.chesapeakebay.net/data/downloads/cbp_water_quality_database_1984_present, last access: December 2015.

Collins, M., Knutti, R., Arblaster, J., Dufresne, J-L., Fichefet, T., Friedlingstein, P., Gao, X., Gutowski, W. J., Johns, T., Krinner, G., Shongwe, M., Tebaldi, C., Weaver, A. J., and Wehner, M.: Long-term Climate Change: Projections, Commitments and Irreversibility. In: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change, Cambridge University Press, 2013.

Cooper, S. R., and Brush, G. S.: Long-Term History of Chesapeake Bay Anoxia, Science, 254, 992-996, 1991.

Cooper, S. R., and Brush, G. S.: A 2,500-Year History of Anoxia and Eutrophication in Chesapeake Bay, Estuaries, 16, 617-626, 1993.

Diaz, R. J.: Overview of Hypoxia around the World, J. Environ. Qual., 30, 275-281, 2001.

Diaz. R. J., and Rosenberg, R.: Spreading Dead Zones and Consequences for Marine Ecosystems, Science, 321, 926-929, doi:10.1126/science.1156401, 2008.

Du, J., and Shen, J.: Decoupling the influence of biological and physical processes on the dissolved oxygen in the Chesapeake Bay, J. Geophys. Res-Oceans, 120, doi:10.1002/2014JC010422, 2015.

Ekau, W., Auel, H., Portner, H.-O., and Gilbert, D.: Impacts of hypoxia on the structure and processes in pelagic communities (zooplankton, macro-invertebrates and fish), Biogeosciences, 7, doi:10.5194/bg-7-1669-2010, 2010.

Elliott, D. T., Pierson, J. J., Roman, M. R.: Predicting the Effects of Coastal Hypoxia on Vital Rates of the Planktonic Copepod Acartia tonsa Dana, PLoS ONE, 8, doi:10.1371/journal.pone.0063987, 2013.

Feng, Y., Friedrichs, M. A. M., Wilkin, J., Tian, H., Yang, Q., Hofmann, E. E., Wiggert, J. D., and Hood, R. R.: Chesapeake Bay nitrogen fluxes derived from a land-estuarine ocean biogeochemical modeling system: Model description, evaluation, and nitrogen budgets, J. Geophys. Res-Biogeo., 120, doi:10.1002/2015JG002931, 2015.

Fofonoff, N. P., and Millard, R. C.: Algorithms for computations of fundamental properties of seawater, Unesco Technical Papers in Marine Science, 44, pp. 53, 1083.

Friedrichs, M. A. M., Hood, R., and Wiggert, J.: Ecosystem model complexity versus physical forcing: Quantification of their relative impact with assimilated Arabian Sea data, Deep-Sea Res. Pt. II, 53, 576-600, 2006.

Friedrichs, M., Sellner, K. G., and Johnston, M. A.: Using Multiple Models for Management in the Chesapeake Bay: A Shallow Water Pilot Project, Chesapeake Bay Program Scientific and Technical Advisory Committee Report, No. 12-003, 2012.

Friedrichs, M. A. M., Dusenberry, J., Anderson, L., Armstrong, R., Chai, F., Christian, J., Doney, S. C., Dunne, J., Fujii, M., Hood, R., McGillicuddy, D., Moore, K., Schartau, M., Sptiz, Y. H., and Wiggert, J.: Assessment of skill and portability in regional marine biogeochemical models: role of multiple phytoplankton groups, J. Geophys. Res., 112, doi:10.1029/2006JC003852, 2007.

Fulton, E. A., Smith, A. D. M., and Johnson, C. R.: Effect of complexity on marine ecosystem models, Mar. Ecol. Prog. Ser., 253, 1-16, 2003.

Gilbert, D., Rabalais, N. N., Diaz, R. J., and Zhang, J.: Evidence for greater oxygen decline rates in the coastal ocean than in the open ocean, Biogeosciences, 7, 2283-2296, doi:10.5194/bg-7-2283-2010, 2010.

Gneiting, T., and Raftery, A. E.: Weather Forecasting with Ensemble Methods, Science, 310, doi: 10.1126/science.1115255, 2005.

Hagedorn, R., Doblas-Reyes, F. J., and Palmer, T. N.: The rationale behind the success of multi-model ensembles in seasonal forecasting – I. Basic concept, Tellus, 57A, 2005.

Hagy, J. D., Boyton, W. R., Keefe, C. W., and Wood, K. V.: Hypoxia in Chesapeake Bay, 1950-2001: Long-term Change in Relation to Nutrient Loading and River Flow, Estuaries, 27, 634-658, 2004.

Haidvogel, D. B., Arango, H., Budgell, W. P., Cornuelle, B. D., Curchitser, E., Di Lorenzo, E., Fennel, K., Geyer, W. R., Hermann, A. J., Lanerolle, L., Levin, J., McWilliams, J. C., Miller, A. J., Moore, A. M., Powell, T. M., Shchepetkin, A. F., Sherwood, C. R., Signell, R. P., Warner, J. C., and Wilkin, J.: Ocean forecasting in terrain-following coordinates: Formulation and skill assessment of the Regional Ocean Modeling System, J. Comput. Phys., 227, 3595-3624, doi:10.1016/j.jcp.2007.06.016, 2008.

Harding Jr., L. W., and Perry, E. S.: Long-term increase of phytoplankton biomass in Chesapeake Bay, 1950-1994, Mar. Ecol. Prog. Ser., 157, 39-52, 1997.

Harding Jr., L. W., Gallegos, C. L., Perry, E. S., Miller, W. D., Adolf, J. E., Mallonee, M. E., and Paerl, H. W.: Long-Term Trends of Nutrients and Phytoplankton in Chesapeake Bay, Estuaries. Coasts, doi:10.1007/s12237-015-0023-7, 2015.

Hofmann, E. E., Druon, J., Fennel, K., Friedrichs, M., Haidvogel, D., Lee, C., Mannino, A., McClain, C., Najjar, R., O'Reilly, J., Pollard, D., Previdi, M., Seitzinger, S., Siewert, J., Signorini, S., and Wilkin, J.: Eastern US Continental Shelf Carbon Budget: Integrating Models, Data Assimilation, and Analysis, Oceanography, 21, 86-104, doi:10.5670/oceanog.2008.70, 2008.

Hong, B., and Shen, J.: Responses of estuarine salinity and transport processes to potential future sea-level rise in the Chesapeake Bay, Estuar. Coast. Shelf S., 104-105, 33-45, doi:10.1016/j.ecss.2012.03.014, 2012.

Hong, B., and Shen, J.: Linking dynamics of transport timescale and variations of hypoxia in the Chesapeake Bay, J. Geophys. Res-Oceans, 118, 6017-6029, doi:10.1002/2013JC008859, 2013.

Janssen, A. B. G., Arhonditsis, G. B., Beusen, A., Bolding, K., Bruce, L., Bruggeman, J., Couture, R-M., Downing, A. S., Elliott, J A., Frassl, M. A., Gal, G., Gerla, D. J., Hipsey, M. R., Hu, F., Ives, S. C., Janse, J. J., Jeppsen, E., Johnk, K. D., Kneis, D., Kong, X., Kuiper, J. J., Lehmann, M. K., Lemmen, C., Ozkundakci, D., Petzoldt, T., Rinke, K., Robson, B. J., Sachse, R., Schep, S. A., Schmid, M., Scholten, H., Teurlincx, S., Trolle, D., Troost, T. A., Van Dam, A. A., Van Gerven, L. P. A., Weijerman, M., Wells, S. A., and Mooij, W. M.: Exploring, exploiting and evolving diversity of aquatic ecosystem models: a community perspective, Aquat. Ecol., doi:10.1007/s10452-015-9544-1, 2015.

Jiang, L., and Xia, M.: Dynamics of the Chesapeake Bay outflow plume: 1. Realistic plume simulations and its seasonal, interannual variability, J. Geophys. Res-Oceans, 2015, in review.

Jiang, L., Xia, M., Ludsin, S. A., Rutherford, E. S., Mason, D. M., Jarrin, J. M., and Pangle, K. L.: Biophysical modeling assessment of the drivers for plankton dynamics in dressenid-colonized western Lake Erie, Ecological Modelling, 308, 18-33, 2015.

Jolliff, J. K., Kindle, J. C., Schulman, I., Penta, B., Friedrichs, M. A. M., Helber, R., Arnone, R. A.: Summary diagrams for coupled hydrodynamic-ecosystem model skill assessment, J. Marine Syst., 76, 64-82, doi:10.1016/j.jmarsys.2008.05.014, 2009.

Keisman, J., and Shenk, G.: Total maximum daily load criteria assessment using monitoring and modeling data, J. Am. Water Resour. As., 49, 1134-1149, doi:10.1111/jawr.12111, 2013.

Keister, J. E., Houde, E. D., and Breitburg, D. L.: Effects of bottom-layer hypoxia on abundances and depth distributions of organisms in Patuxent River, Chesapeake Bay, Mar. Ecol. Prog. Ser., 205, 43-59, 2000.

Kemp, W. M., Boyton, W. R., Adolf, J. E., Boesch, D. F., Boicourt, W. C., Brush, G., Cornwell, J. C., Fisher, T. R., Gilbert, P. M., Hagy, J. D., Harding, L. W., Houde, E. D., Kimmel, D. G., Miller, W. D., Newell, R. I. E., Roman, M. R., Smith, E. M., and Stevenson, J. C.: Eutrophication of Chesapeake Bay: historical trends and ecological interactions, Mar. Ecol. Prog. Ser., 303, 1-29, 2005.

Kemp, W. M., Testa, J. M., Conley, D. J., Gilbert, D., and Hagy, J. D.: Temporal responses of coastal hypoxia to nutrient loading and physical controls, Biogeosciences, 6, 2985-3008, 2009.

Lanerolle, L. W., Patchen, R. C., and Aikman, F.: The Second Generation Chesapeake Bay Operational Forecast System (CBOFS2): Model Development and Skill Assessment, TR-NOS-CS-29, US Department of Commerce, National Oceanic and Atmospheric Administration, National Ocean Service, Office of Coast Survey, Coast Survey Development Laboratory, Silver Spring, MD, 2011.

Lehmann, M. K., Fennel, K. and He, R.: Statistical validation of a 3-D bio-physical model of the western North Atlantic, Biogeosciences, 6, 2009.

Levin, L. A., Ekau, W., Gooday, A. J., Jorissen, F., Middleburg, J. J., Naqvi, S. W. A., Neira, C., Rabalais, N. N., and Zhang, J.: Effects of natural and human-induced hypoxia on coastal benthos, Biogeosciences, 6, doi:10.5194/bg-6-2063-2009, 2009.

Li, M., Zhong, L., and Boicourt, W. C.: Simulations of Chesapeake Bay estuary: Sensitivity to turbulence mixing parameterizations and comparison with observations, J. Geophys. Res., 110, doi:10.1029/2004JC002585, 2005.

Li, Y., Li, M., and Kemp, W. M.: A Budget Analaysis of Bottom-Water Dissolved Oxygen in Chesapeake Bay, Estuar. Coast., 37, doi:10.1007/s12237-014-9928-9, 2015.

Meier, H. E. M., Andersson, H. C., Arheimer, B., Blenckner, T., Chubarenko, B., Donnelly, C., Eilola, K., Gustafsson, B. G., Hansson, A., Havenhand, J., Hoglund, A., Kuznetsov, I., MacKenzie, B. R., Muller-Karulis, B., Neumann, T., Niiranen, S., Piwowarczyk, J., Raudsepp, U., Reckermann, M., Ruoho-Airola, T., Savchuk, O. P., Schenk, F., Schimanke, A., Vali, G., Weslawski, J-M., and Zorita, E.: Comparins reconstructed past variations and future projections of the Baltic Sea ecosystem – first results from multi-model ensemble simulations, Environ. Res. Lett., 7, doi: 10.1088/1748-9326/7/3/034005, 2012.

Meire, L., Soetaert, K. E. R., and Meysman, F. J. R.: Impact of global change on coastal oxygen dynamics and risk of hypoxia, Biogeosciences, 10, doi:10.5194/bg-10-2633-2013, 2013.

Murphy, R. R., Kemp, W. M., Ball, W. P.: Long-term Trends in Chesapeake Bay Seasonal Hypoxia, Stratification, and Nutrient Loading, Estuaries. Coasts, doi:10.1007/s12237-01109413-7, 2011.

Najjar, R. G., Pyke, C. R., Adams, M. B., Breitburg, D., Hershner, C., Kemp, M., Howarth, R., Mulholland, M. R., Paolisso, M., Secor, D., Sellner, K., Wardrop, D., and Wood, R.: Potential climate-change impacts on the Chesapeake Bay, Estuar. Coast. Shelf S., 86, 1-20, doi:10.1016/j.ecss.2009.09.026, 2010.

Park, K., Kuo, A. Y., Shen, J., Hamrick, J. M.: A three-dimensional hydrodynamic eutrophication model (HEM-3D): description of water quality and sediment process submodels, Special Report in Applied Marine Science and Ocean Engineering, 327, 1995.

Pierson, J. J., Roman, M. R., Kimmel, D. G., Boicourt, W. C., and Zhang, X. S.: Quantifying changes in the vertical distribution of mesozooplankton in response to hypoxic bottom waters, J. Exp. Mar. Biol. Ecol., 381, 74-79, 2009.

Prince, E. D., and Goodyear, C. P.: Hypoxia-based habitat compression of tropical pelagic fishes, Fish. Oceanogr., 15, 451-464, doi:10.111/j.1365-2419.2005.00393.x, 2006.

Riedel, B., Pados, T., Pretterebner, K., Schiemer, L., Steckbauer, A., Haselmair, A., Zuschin, M., and Stachowitsch, M.: Effect of hypoxia and anoxia on invertebrate behavior: ecological perspectives from species to community level, Biogeosciences, 11, doi:10.5194/bg-11-1491-2014, 2014.

Schlenger, A. J., North, E. W., Schlag, Z., Li, Y., Secor, D. H., Smith, K. A., and Niklitschek, E. J.: Modeling the influence of hypoxia on the potential habitat of Atlantic sturgeon *Acipenser oxyrinchus*: a comparison of two methods, Mar. Ecol. Prog. Ser., 483, doi: 10.3354/,eps10248, 2013.

Scully, M. E.: The Importance of Climate Variability to Wind-Driven Modulation of Hypoxia in Chesapeake Bay, J. Phys. Oceanogr., 40, 1435-1440, doi:10.1175/2010JPO4321.1, 2010.

Scully, M. E.: Physical controls on hypoxia in Chesapeake Bay: A numerical modeling study, J. Geophys. Res-Oceans, 118, 1239-1256, doi:10.1001/jgrc.20138, 2013.

Shchepetkin, A. F., and McWilliams, J. C.: The Regional Ocean Modeling System (ROMS): A Split-Explicit, Free-Surface, Topography-Following-Coordinate Oceanic Model, Ocean Model., 9, 347-404, doi:10.1016/j.ocemod.2004.08.002, 2005.

Shenk, G. W., and Linker, L. C.: Development and application of the 2010 Chesapeake Bay watershed total maximum daily load model, J. Am. Water Resour. As., 49, 1-15, doi:10.1111/jawr.12109, 2013.

Taylor, K. E.: Summarizing multiple aspects of models performance in a single diagram, J. Geophys. Res., 106, 7183-7192, 2001.

Testa, J. M., and Kemp, W., M.: Spatial and Temporal Patterns of Winter-Spring Oxygen Depletion in Chesapeake Bay Bottom Water, Estuar. Coast., 37, doi:10.1007/s12237-014-9775-8, 2014.

Testa, J. M., Li, Y., Lee, Y. J., Li, M., Brady, D. C., Di Toro, D. M., Kemp, W. M., and Fitzpatrick, J. J.: Quantifying the effects of nutrient loading on dissolved O_2 cycling and hypoxia in Chesapeake Bay using a coupled hydrodynamic-biogeochemical model, J. Marine Syst., 139, 139-158, doi:10.1016/j.jmarsys.2014.05.018, 2014.

Tian, H., Yang, Q., Najjar, R., Ren, W., Friedrichs, M. A. M., Hopkinson, C. S., and Pan, S.: Anthropogenic and climatic influences on carbon fluxes from eastern North America to the Atlantic Ocean: A process-based modeling study, J. Geophys. Res-Biogeo., 120, doi: 10.1002/2014JG002760, 2015.

Trolle, D., Elliott, J. A., Mooij, W. M., Janse, J. H., Bolding, K., Hamilton, D. P., and Jeppsen, E.: Advancing projections of phytoplankton responses to climate change through ensemble modeling, Environ. Modell. Softw., 61, 371-379, doi:10.1016/j.envsoft.2014.01.032, 2014.

USEPA (U.S. Environmental Protection Agency): Ambient Water Quality Criteria for Dissolved Oxygen, Water Clarity, and Chlorophyll a for the Chesapeake Bay and its Tidal Tributaries – 2004 Addendum, EPA 903-R-03-002, USEPA Region III Chesapeake Bay Program Office, Annapolis, MD, 2004.

USEPA (U.S. Environmental Protection Agency): Chesapeake Bay Total Maximum Daily Load for Nitrogen, Phosphorus, and Sediment, U.S. Environmental Protection Agency Chesapeake Bay Program Office, Annapolis, MD, 2010.

USEPA (U.S. Environmental Protection Agency): Guide to Using Chesapeake Bay Program Water Quality Monitoring Data, EPA 903-R-12-001, U.S. Environmental Protection Agency Chesapeake Bay Program, Annapolis, MD, 2012.

Vaquer-Sunyer, R., and Duarte, C. M.: Thresholds of Hypoxia for Marine Biodiversity, P. Natl. Acad. Sci. USA, 105, 15452-15457, doi:10.1073/pnas.0803833195, 2008.

Ward, B. A., Friedrichs, M. A. M., Anderson, T. R., and Oschlies, A.: Parameter optimization techniques and the problem of underdetermination in marine biogeochemical models, J. Mar. Sys., 81, doi: 10.1016/j.marsys.2009.12.005, 2010.

Ward, B. A., Schartau, M., Oschlies, A., Martin, A. P., Follows, M. J., and Anderson, T. R.: When is a biogeochemical model too complex? Objective model reduction and selection for North Atlantic time-series sites, Prog. Oceanogr., doi: 10.1016/j.pocean.2013.06.002, 2013.

Weller, D., Benham, B., Friedrichs, M., Gardner, N., Hood, R., Najjar, R., Paolisso, M., Pasquale, P., Sellner, K., and Shenk, G.: Multiple Models for Management in the Chesapeake Bay, Chesapeake Bay Program Scientific and Technical Advisory Committee Workshop Report, No. 14-004, 2013.

Xiao, Y., and Friedrichs, M. A. M.: Using biogeochemical data assimilation to assess the relative skill of multiple ecosystem models: effects of increasing the complexity of the planktonic food web, Biogeosciences, 11, 3015-3030, doi:10.5194/bg-11-3015-2014, 2014.

Xu, J., Long, W., Wiggert, J. D., Lanerolle, L. W. J., Brown, C. W., Murtugudde, R., and Hood, R. R.: Climate forcing and salinity variability in Chesapeake Bay, USA, Estuar. Coast. Shelf S., 35, 237-261, doi:10.1007/s12237-011-9423-5, 2012.

Yang, A., Tian, H., Friedrichs, M. A. M., Hopkinson, C. S, Lu, C., and Najjar, R. G.: Increased nitrogen export from eastern North America to the Atlantic Ocean due to climatic and anthropogenic changes during 1901-2008, J. Geophys. Res-Biogeo., 120, doi: 10.1002/2014JG002763, 2015a.

Yang, Q., Tian, H., Friedrichs, M. A. M., Liu, M., Li, X., and Yang, J.: Hydrological responses to climate and land-use changes along the North American east coast: A 110-Year historical reconstruction, J. Am. Water Resour. As., 51, 47-67, doi:10.1111/jawr.12232, 2015b.

Chapter 3:

Evaluating confidence in the impact of regulatory nutrient reduction on Chesapeake Bay water quality

Isaac D. Irby[1], Marjorie A. M. Friedrichs[1]

[1]Virginia Institute of Marine Science, College of William & Mary, Gloucester Point, VA, USA.

Submitted to *Estuaries and Coasts*
May 2017

3. EVALUATING CONFIDENCE IN THE IMPACT OF REGULATORY NUTRIENT REDUCTION ON CHESAPEAKE BAY WATER QAULITY

Key Points

- Two estuarine water quality models simulate a similar relative improvement in dissolved oxygen as a result of nutrient reduction.
- The models predict very similar levels of water quality standards attainment throughout most of the Bay.
- The greatest uncertainty in the impact of nutrient reduction on the attainment of water quality standards is generally in those areas with historically large hypoxic issues.
- The regressions derived form the raw model output are the greatest source of uncertainty in the process of evaluating water quality standards attainment.

Abstract

Excess nutrients derived from anthropogenic activity have resulted in the degradation of coastal water quality and an increase in low-oxygen events worldwide. In an effort to curb these impacts and restore water quality in the Chesapeake Bay, a Total Maximum Daily Load (TMDL) of nutrients and sediment has been established based on a framework of regulatory standards and models. This research aims to evaluate the uncertainty in projected changes in water quality resulting from TMDL implementation by applying the methodology used to establish the regulatory loads to two models: one developed and utilized in an academic research setting, and one developed for use in the regulatory process. Results of this comparison demonstrate that although the two models differ structurally and in biogeochemical complexity, they project a similar relative improvement in water quality along the mainstem of the Chesapeake Bay and the lower reaches of the tributaries. Furthermore, the models largely agree on the attainment of regulatory water quality standards as a result of nutrient reduction, while also establishing that meeting water quality standards is relatively independent of hydrologic conditions. By developing a Confidence Index, this research identifies the locations and causes of greatest uncertainty in modeled projections of water quality. Although there are specific locations and times where the models disagree, overall this research lends an increased degree of confidence in the appropriateness of the TMDL levels and in the general impact nutrient reductions will have on Chesapeake Bay water quality under current environmental conditions.

3.1 Introduction

Degraded coastal water quality as a result of anthropogenic nutrient enrichment can have a large negative impact on local economies (Rabotyagov et al. 2014). Low-oxygen events have been linked to declines in catch of commercially valuable species, biodiversity, and ecosystem value (Diaz 2001; Rabalais and Turner 2001; Brietburg et al. 2009; Buchheister et al. 2013). These impacts can eventually lead to regulatory action by the affected communities in an effort to restore water quality to a sustainable standard. Governments have worked across state, territory, and country lines to address water quality issues in areas such as the Baltic Sea (HELCOM 2007), the Black Sea (BSC 2009), the Gulf of Mexico (HTF 2011), and the Chesapeake Bay (USEPA 2010a). The regulatory response can take many forms but commonly involves the reduction of nutrients derived from agriculture practices, wastewater treatment, and urban runoff in an effort to improve water quality and the associated positive externalities. The regulatory structure used to define the level of nutrient reduction is often referred to as a Total Maximum Daily Load (TMDL), which sets the maximum level of pollutant delivery a particular jurisdiction can deliver to a waterway.

As the largest estuary in the United States with a watershed supporting a growing population of over 17 million people, the Chesapeake Bay is particularly prone to water quality degradation as a result of human activity. The Bay has experienced decreased dissolved oxygen (DO) concentrations and a degradation of other water quality metrics over the last 200 years with degradation amplifying since the 1950s (Cooper and Brush 1993; Curtin et al. 2001; Hagy et al. 2004). Increased volumes of low-oxygen waters in the Bay (Bever et al. 2013) have had a negative effect on the health of the Bay ecosystem and economy (Phillips and McGee 2014). Volumes of low-DO water compress fish habitat and impact the catch per unit effort across the Bay (Buchheister et al. 2013). Blue crab harvest in the Bay is also negatively affected by low-DO conditions (Mistiaen et al. 2003).

In the 1980’s nutrient pollutants such as phosphorus and nitrogen were identified as the main sources of the Bay’s water quality issues (USEPA 1982; USEPA 1983). Multiple efforts by state, federal, and private partners to restore the Bay eventually resulted in the 2010 Chesapeake Bay TMDL, the largest and most complex TMDL in the

United States to date (USEPA 2010a). Under the authority of the Clean Water Act Section 303(d), the TMDL established location-specific mandated pollutant reductions throughout the six states and Washington, D.C. that make up the Chesapeake Bay watershed (Fig. 3.1). The specific level of reduction was set to ensure that water quality standards as defined by the Environmental Protection Agency (EPA) would be eventually met. The timeline instituted by the regulation mandates that all pollutant reduction efforts be in place by 2025 and that a mid-point assessment be conducted in 2017 (USEPA 2010a).

The Chesapeake Bay Program (CBP) under the authority of the EPA established allocations of mandated nutrient load reductions by utilizing a complex modeling system, including an airshed model, watershed model, and estuarine water quality model (USEPA 2010a). Much of the focus of the research in developing the Bay TMDL was on the complex watershed modeling system, as that was the model that would indicate how much reduction each section of the watershed would have to incur. However, the CBP has also had a long history of estuarine water quality modeling in the Chesapeake Bay. The model currently used by the CBP for estuarine water quality modeling was first developed in 1983 and has been continually in use and undergone frequent updates (Cerco et al. 2010).

Recently, many other governmental and academic groups have also committed themselves to modeling the Chesapeake Bay as exemplified by the Chesapeake Community Modeling Program (http://ches.communitymodeling.org/). The combination of multiple marine science institutions and robust observational datasets as a result of the EPA's extensive water quality monitoring of the Bay (http://www.chesapeakebay.net/data/downloads/cbp_water_quality_database_1984_present) has provided the research community with the means necessary for interdisciplinary modeling of the Bay ecosystem. Studies examining the relative strengths of various modeling strategies and methods have increased the caliber of Bay water quality modeling and understanding (e.g., Bever et al. 2013; Chapter 2). From relatively simple water quality models (e.g., Hong and Shen 2013; Lake and Brush 2015; Li et al., 2015; Scully 2016) to more complex coupled circulation-biogeochemical models (e.g., Xu and Hood 2006; Testa et al. 2014; Feng et al. 2015; Xia and Jiang 2016) to ecosystem models that incorporate higher trophic levels (e.g., Ihde et al. 2016), the Chesapeake Bay research community has committed to furthering our understanding of the Bay ecosystem as a whole and Bay water quality in particular.

The research presented here harnesses the strength of the governmental and academic research on Chesapeake Bay water quality modeling to evaluate the uncertainty in projections of changes in water quality resulting from regulatory nutrient reduction. Regulations aimed at improving the water quality of the Chesapeake Bay are estimated to cost in the tens of billions of dollars (Nelson 2014). With such astounding potential costs, it is crucial for regulatory efforts to be targeted and successful. It is also critical that the uncertainties associated with projected future conditions be well characterized and quantified. This research utilizes a comparison of two model simulations for the Chesapeake Bay to evaluate confidence in the impact of regulatory nutrient reduction on water quality. In this context, low uncertainty, or high confidence, does not mean that water quality standards will be met; rather, it means that both models agree on the impact of nutrient reductions on water quality regardless of whether or not the standards themselves are met.

3.2 Modeling methodology

In this study, simulations from two water quality models were assessed and compared; one model was developed and used in a regulatory context, hereafter referred to as the Regulatory Model, and the other was used in an academic research context, hereafter referred to as the Academic Model. Both models were forced by the same nutrient scenarios and underwent the same model skill assessment. To facilitate model comparisons, the Academic Model output was vertically mapped to the grid of the Regulatory Model. Where the Regulatory Model grid depths were deeper (shallower) than those of the Academic Model, the Academic Model output profiles were linearly stretched (compressed) to match those of the Regulatory Model output.

3.2.1 Water quality models

3.2.1.1 Regulatory Model: CH3D-ICM

The CBP's water quality and sediment transport model, the Curvilinear Hydrodynamics in Three Dimensions – Integrated Compartment Model (CH3D-ICM; Cerco et al. 2010), is a coupled hydrodynamic-biogeochemical model used to help set regulatory policies for the Chesapeake Bay TMDL. The model employs a curvilinear boundary-fitted horizontal grid consisting of 11,604 horizontal cells with an average wet cell resolution of 1 km and a 1.52 m vertical z-grid. The ecological component of the model uses 24 state variables throughout the water column that interact with a complete sediment diagenesis module. CH3D-ICM has been extensively calibrated for the Chesapeake Bay and has been in use and development since the 1980s (Johnson et al. 1993). Model output was provided by the CBP and is from the version of the model used in the development of the TMDL.

3.2.1.2 Academic Model: ChesROMS-ECB

The Chesapeake Bay Regional Ocean Modeling System – Estuarine Carbon Biogeochemistry model (ChesROMS-ECB; Feng et al. 2015) is a coupled circulation-biogeochemical model based on the Regional Ocean Modeling System (ROMS; Shchepetkin and McWilliams 2005) with the specific model domain and curvilinear horizontal grid based on the Chesapeake Bay modeling community's ChesROMS model (Xu et al. 2012). The vertical framework follows a sigma-grid with 20 layers and a stretching parameter that condenses the grid at the air-water and sediment-water interfaces while expanding it mid-water column. The average wet cell resolution inside the Bay for the Academic Model (1.7 km) is larger than that of the Regulatory Model and also employs a more smoothed bathymetry. The differences in bathymetry are most apparent on the shoals and mouth of the Bay where the Academic Model is generally shallower than the Regulatory Model and true bathymetry. Another difference between the two water quality models is the locations at which watershed loads are delivered to the Bay. In the Regulatory Model, watershed loads are input continually along the entire land-water interface. While efforts are underway to improve the Academic Model boundary so that it can also receive watershed loads along the entire land-water interface, it currently only receives loads at ten major rivers (Feng et al. 2015).

The biogeochemical component of the Academic Model is modified from Fennel et al. (2006), Druon et al. (2010), and Hofmann et al. (2011) in order to be specifically applicable for estuarine applications (Feng et al. 2015). These modifications include the addition of an inorganic suspended solids state variable, water column denitrification, oxygen limitation of nitrification, and a new parameterization for light attenuation that is a function of suspended particulate matter and salinity (as a proxy for colored dissolved organic matter; Xu et al. 2005). The biogeochemical component of the Academic Model is mid-complexity relative to other less complex models that do not employ a full biogeochemical component but have been shown to adequately simulate seasonal DO dynamics in the Chesapeake Bay utilizing a parameterized constant oxygen consumption rate (Scully 2010; Bever et al. 2013; Chapter 2) and more complex models such as the Regulatory Model (Cerco et al. 2010).

While the TMDL mandates the reduction of nitrogen, phosphorus, and sediment, the Academic Model used in this study does not include phosphorus. While not including phosphorus may be a limitation for realistically simulating marine nutrient cycles, the Academic Model has been calibrated with nitrogen as the only nutrient available for biologic uptake and is able to successfully simulate DO variability especially in the deep mid-Bay during the hypoxic summer season (Feng et al. 2015; Chapter 2), since this time and region is more limited by nitrogen than by phosphorous (Testa et al. 2014).

3.2.2 Nutrient forcing

3.2.2.1 CBP Watershed Model (WSM) inputs

To ensure consistency between model runs, both models were forced at the land-water interface by the CBP regulatory watershed model version 5.3.2 (WSM; Shenk and Linker 2013). This is the version used in the development of the TMDL. The WSM, based on Hydrologic Simulation Program – Fortran (HSPF), utilizes multiple components to simulate land use, river flow, and the loading of nutrients and sediment to the Bay and has been in continual development with the input of multiple stakeholders since 1982 (Linker et al. 2002; Shenk and Linker 2013). The WSM was used in establishing the Bay

TMDL by developing and running multiple nutrient and sediment load reduction scenarios. Once an appropriate level of reduction was established, the WSM was used to allocate the requisite location-specific reductions required by each jurisdiction throughout the watershed (Shenk and Linker 2013).

A transfer of variables was required from the WSM to both water quality models as the watershed and estuarine models express organic nutrient constituents differently. To match the relative contribution of the organic nitrogen constituents in the Academic Model (refractory dissolved organic nitrogen, semi-labile dissolved organic nitrogen, and particulate organic nitrogen) to the WSM (lysed freshwater phytoplankton nitrogen, total refractory organic nitrogen, and the biological oxygen demand from organic nitrogen), the components from the watershed were divided up and input as the necessary components to the estuary. To achieve this, the refractory dissolved organic nitrogen was set to be 20% of the total refractory organic nitrogen, the semi-labile dissolved organic nitrogen was the total biological oxygen demand from organic nitrogen and 80% of the phytoplankton nitrogen, and the particulate organic nitrogen was 80% of the refractory organic nitrogen and 20% of the phytoplankton nitrogen. These assumptions are consistent with the partitioning used in the development of the Academic Model (Feng et al. 2015).

3.2.2.2 Standard run and TMDL-WIP scenario

Two WSM nutrient load scenarios were applied to both estuarine models used in this study. The first was the calibration scenario for the WSM, which the CBP used as a baseline for the TMDL. This scenario represents a realistic simulation of the observed watershed loads for 1991-2000. Estuarine model results utilizing the WSM calibration scenario will hereafter be referred to as the "standard run." The second was the TMDL Watershed Implementation Plan (WIP) nutrient reduction scenario. This scenario assumes all nutrient reduction strategies as prescribed by the WIP have been implemented, and is very similar to the actual scenario used to establish the TMDL. Estuarine model results utilizing the WSM WIP nutrient reduction scenario will hereafter be referred to as the "TMDL-WIP scenario." Specific information on the nutrient reduction loads can be found in Shenk and Linker (2013) and include a percent reduction from the study period average of 33% for total nitrogen, 27% for total phosphorus, and 27% for sediment across the entire watershed.

The 2010 TMDL was the first instance where atmospheric nitrogen deposition was included in an agreement to improve water quality of the Bay. Direct atmospheric nitrogen deposition for both models is derived from the CBP airshed model, which combines output from the Community Multiscale Air Quality model (CMAQ) for dry deposition and a regression model for wet deposition (Linker et al. 2013a). The TMDL allocates a maximum annual direct deposition of 7.1 million kg N to the tidal waters of the Bay (USEPA 2010a; Linker et al. 2013a). This maximum allowable loading represents a 30% reduction of direct atmospheric nitrogen deposition from the study period average. To keep the atmospheric nitrogen deposition loads consistent between the models, the 30% reduction in direct deposition was also applied to the Academic Model during the nutrient reduction scenario.

The impact of the TMDL-WIP reduced nutrient scenario is compared for both estuarine models by examining the absolute as well as relative changes in DO concentrations spatially and temporally. By dividing the absolute change by the concentration from the standard run, the relative change in DO concentration as a result of the TMDL-WIP scenario is determined. This relative change in DO allows for a comparison of the impact of nutrient reduction between the models.

3.2.3 Model skill metrics

Before comparing the impact of nutrient reductions on the DO concentrations from the two models, it is critical to compare their relative skill when they are both forced by the same realistic watershed inputs, i.e. the standard runs. The Academic and Regulatory Models have been previously shown to exhibit similar skill in simulating seasonal DO concentrations along the main-stem of the Bay in 2004 and 2005 (Chapter 2). Here, the skill assessment of these models is extended to include a longer time frame (1991-2000) and a larger variety of stations encompassing both the main stem and the lower portions of the major tributaries of the Bay (Fig. 3.1, Table 3.1). Model skill is statistically compared via target and Taylor diagrams (Taylor 2001; Hofmann et al. 2008; Jolliff et al. 2009). Target diagrams allow for a comparison of the total root-mean squared difference (RMSD), bias, and unbiased RMSD on a single diagram. Taylor diagrams similarly offer information on the unbiased RMSD, but additionally include quantitative information on the standard deviations and correlations between the model output and observations. Both diagrams are normalized to the standard deviation of the observations in order to facilitate plotting multiple variables on the same diagram. The normalization of the target diagram allows the unit circle to represent the skill of a model as defined by the mean of the observations. (See Hofmann et al. (2008) and Joliff et al. (2009) for a more detailed description of target diagrams and Taylor (2001) for a more detailed description of Taylor diagrams.) Full methods for the comparison follow Chapter 2.

3.3 CBP procedure for assessing attainment of water quality standards

3.3.1 Designated uses

The CBP evaluates water quality for multiple habitats across the Bay and its tributaries (Tango and Batiuk 2013). These habitats are termed "designated uses" by the CBP and are characterized by ecological use (USEPA 2003; USEPA 2010b). Each designated use (Fig. 3.2) has a specific mandated minimum DO criterion, otherwise known as a DO "Water Quality Standard" (Table 3.2). Because of the importance of seasonal differences in DO, the TMDL specifies different Water Quality Standards (WQS) for the summer (June – September) and non-summer (October – May) seasons. During the non-summer, Open Water encompasses the entire water column. During the summer, the Open Water designation incorporates all surface water environments across the Bay and extends down to the bottom of the mixed surface layer, if there is one. The Deep Water designated use represents the summer transitional zone of the water column that is influenced by the pycnocline, incorporating all water below the well-mixed surface layer and above the well-mixed deep layer. The Deep Channel designated use encompasses the deep summer waters of the main stem trench and deep tributaries where physical characteristics limit the elevation of DO concentrations regardless of controls on water quality during the summer months. The depths delineating each designated use are defined by the observed physical characteristics for each month and are therefore non-uniform across time and space. The migratory fish spawning and nursery designated use and the shallow-water Bay grass designated use were not individually evaluated in this study because they follow the Open Water DO criteria for the summer months.

3.3.2 TMDL regulatory protocol for determining attainment of water quality standards

To examine whether nutrient reductions will result in DO WQS being met, i.e. attained, the output from the standard run and TMDL-WIP scenario of both estuarine models underwent the published process for identifying attainment of WQS as established by the EPA and CBP (USEPA 2010a). While a brief synopsis of the methodology is below, a more complete documentation of this process can be found in the Supplemental Material for this manuscript and throughout the Chesapeake Bay TMDL literature (USEPA 2010a; Linker et al. 2013b; Tango and Batiuk 2013).

The TMDL regulations were not based off the absolute DO concentrations simulated by the models; rather, the regulations use the change in DO between the standard run and TMDL-WIP scenario. To quantify the change in DO due to the nutrient reduction at a specific location and specific time, the hourly output for each month at each vertical cell at each station for the standard and scenario runs were regressed against each other (Fig. 3.3). (Also see USEPA Appendix H 2010). The resulting linear regression was used to create a scenario-modified dataset by inputting an actual observed DO concentration from 1991-2000 into the regression equation as the independent variable and obtaining a projected DO concentration as a result of the nutrient reduction scenario. By utilizing this regression method, the errors in the true predictive capabilities of the models are minimized since the models are not being used to predict an explicit future DO concentration, but rather they are predicting the relative change in DO concentrations that can be expected as a result of decreased nutrient availability.

Once the regressions were applied and the set of future "observations" were generated, the CBP applied a stoplight analysis to identify the percent time and space that the volume of water in question met or exceeded the mandated WQS. This process uses the CBP Interpolator (USEPA 2012) to interpolate scenario-modified DO concentrations throughout the Bay. The regulations were designed with the flexibility to allow water quality levels to exceed minimum standards, i.e. fail, for a specific percent time and space while still being granted an overall passing grade based on a cumulative distribution function reference curve (USEPA 2010b; Tango and Batiuk 2013). In the stoplight analysis, green represents the percent time and space that WQS are met, yellow represents the percent time and space that WQS are not met but are still within the buffer allowed by the individual reference curve, and red represents the percent time and space that WQS are not met and are beyond the buffer. Therefore, while all three colors are utilized in the stoplight analysis, only red signifies a segment that has failed, or exceeded, the regulatory standards. All percentages are rounded to the nearest whole percent and an exceedance of 1% red is deemed allowable due to impacts from rounding and computational uncertainty. Furthermore, in order to account for the fact that some locations in the Bay might exhibit low DO concentrations even under pristine conditions, an extra allowance was given in specific cases (Table 3.3). These extra allowances, defined as "variances," are allocated at the state level and therefore the stoplight analyses presented here do not include the variances that have been granted; however their significance will be considered in the Discussion.

3.3.3 Study period (1991-2000) and critical period (1993-1995)

The Chesapeake Bay TMDL was established using a 10-year hydrologic study period from 1991-2000 (USEPA 2010a; Linker et al. 2013b). This study period was chosen because it characterized a representative 10-year variability of freshwater flow and was fully encompassed within the 1985-2005 period for which model results were available. Within the 10-year study period, the 3-year critical period of 1993-1995 was used as the basis for the TMDL assessment. These three years were selected based on being representative of a relatively high-flow period, because higher stream flow has been found to result in larger nutrient fluxes from the watershed and ultimately worse water quality conditions (Murphy et al. 2011). Whereas the study period encompasses 1991-2000, the majority of the research presented in this study is focused on the three-year critical period (1993-1995) that was used in the TMDL regulations. (Note that although the Regulatory Model was specifically calibrated for the ten-year study period (Cerco et al. 2010; Cerco and Noel 2013), the Academic Model (Feng et al. 2015) was calibrated for 2001-2005.) In order to explore the impact of the choice of 3-year period on projected water quality, the results of every 3-year period between 1991-2000 (8 total periods) were put through the regulatory WQS protocol.

3.4 New methodology for assessing confidence in WQS attainment: the Confidence Index (CI)

To evaluate the degree of uncertainty in the projected impact of nutrient reductions on water quality, a confidence index (CI) is introduced, which incorporates multiple forms of information regarding how similarly the models react to nutrient reduction for each segment of the Bay. Specifically, the CI is the average of three metrics of similarity that are each given a fractional percent similarity score with a score of 1.0 representing perfect similarity between the model results. A score of 0.0 is different for each metric and is defined as the level at which a dissimilarity of the models would be concerning. Assessing confidence at the segment level allows for an easily digestible framework for visualizing the metric.

The first metric included in the CI is the average similarity of the stoplight analysis for 1993-1995 for the designated uses in each segment, hereafter referred to as the designated use metric. The similarity of the two models is calculated as the total percent identical between the two stoplight analyses. For example, if the Regulatory Model for a given segment and a given designated use results in a stoplight analysis of 85% green, 10% yellow, 5% red, and the Academic Model is 83% green, 8% yellow, 9% red, then there is a 96% similarity between the two models (83% + 8% + 5% = 96%). For some segments the average of all four designated uses was calculated, but for shallow segments, like CB1TF, the average was calculated only from the two Open Water designated uses (summer and non-summer). A score of 0.0 for the designated use metric was set at 75% similarity and a score of 1.0 is given for 100% similarity.

The second metric included in the CI is the similarity in the impact of the 3-year critical period across all designated uses that pertain to any given segment, hereafter referred to as the critical period metric. In order to explore the similarity between the models' stoplight analyses as a result of the choice of 3-year period, every 3-year period between 1991-2000 (8 total periods) were put through the regulatory WQS protocol. This essentially changes the baseline hydrologic conditions used in the assessment. For ease in examining across multiple segments, the percent exceedance, or red stoplight, for each segment was weighted by the total volume of the specific designated use in question for that segment and then all of the segments for a given designated use were added together. Using this method, even though a model may exhibit a large exceedance of 20% in the Deep Water for a relatively small segment like the Patapsco (PATMH), if the rest of the segments display no exceedance, then the total exceedance for Deep Water will be much less that 20%. To establish the similarity between the two models, the average exceedance value (red) for each segment across the eight potential 3-year periods was calculated for each model. A score of 0.0 for the critical period metric was set at a difference in average exceedance between the two models of 2% since a difference of that amount would determine whether or not a segment was in attainment.

The last metric is taken from the average similarity of three sub-metrics that are generated as part of the regressions used to determine attainment of WQS: the correlation (r), y-intercept, and slope. This metric is hereafter referred to as the regression statistic metric. A score of 1.0 represents a perfect match in correlation, y-intercept, and slope, while a score of 0.0 was set at 0.5 for the correlation metric, 2.0 for the y-intercept metric, and 1.0 for the slope metric. These values were chosen because they represent values that would denote a fundamental dissimilarity between the model results. The regression statistic is only based on the summer regressions, as that is when the models are most different. Finally, the full CI is then computed as the average of the three metrics for each segment (Table 3.1).

3.5 Results

3.5.1 Model-data comparison of the standard run for both estuarine models

Before comparing the two models in terms of how they respond to nutrient reductions, it is important to assess the overall skill of both models. Although both models have been previously found to have similar skill along the main stem for 2004-2005 (Chapter 2), it is important to assess skill for the time period used in this analysis

(1991-2000) and expand that analysis beyond the deep mainstem observation stations to incorporate the lower tributary segments.

Both models exhibit a similar level of skill across 25 observation stations (Fig. 3.1) and 10 years (1991-2000) for temperature, salinity, and DO despite differences in model structures, model complexity, and calibration years. In terms of both total RMSD (Fig. 3.4a) and correlation (Fig. 3.4b), the models perform very similarly: they both perform best in terms of temperature, then salinity, then DO, and then stratification. Both models also exhibit a slightly higher degree of skill at the surface relative to the bottom for all variables. The one exception is for DO: the Regulatory Model performs slightly better at the bottom (lower bias and higher correlation) than at the surface. In contrast, the Academic Model tends to overestimate mean DO at both the surface and bottom and thus the Academic Model generally produces DO concentrations that are slightly higher than those generated by the Regulatory Model and the observations.

In terms of stratification (defined as maximum dS/dz), both models similarly underestimate the mean (strength) and variability (variation of strength) while also having a poor correlation with the observations (r = 0.4). Although the Academic Model places the location of maximum stratification too high in the water column (positive bias) whereas the Regulatory Model simulates the correct mean location, the total RMSD for the magnitude and depth of stratification, i.e. the distance from the center of the target diagram to the green symbols (Fig. 3.4a), is roughly 1.0 for both models.

3.5.2 Comparison of the standard runs versus the TMDL-WIP scenarios

When the TMDL-WIP nutrient reduction scenario is applied to both models for 1993-1995, they produce similar changes in summer DO concentrations; however, especially at the bottom, the relative changes are more similar than the absolute changes (Fig. 3.5). Although throughout most of the Bay the nutrient reductions result in decreases in surface DO for both models, at the northernmost stations the models simulate an increase in surface DO. Whereas the Academic Model simulates slightly larger decreases in the southern half of the Bay compared to the Regulatory Model, the relative change in DO between the models (Fig. 3.5e-f), defined as the change in DO divided by the DO concentration in the standard run, is extremely similar across the entire Bay surface.

At the bottom, the TMDL-WIP nutrient reduction scenarios result in the absolute increase in summer DO in the Academic Model being higher than the Regulatory Model along the central main stem (Fig. 3.5c,d); however again the relative increase is remarkably similar in magnitude between the two models (Fig. 3.5g-h). The differences in relative impact between the models is accentuated in some of the tributaries where the Academic Model generally produces small relative changes in summer bottom DO while the Regulatory Model simulates relatively large increases. This is most evident in the Chester River (CHSMH in Fig. 3.1) and Eastern Bay (EASMH in Fig. 3.1) where the Regulatory Model simulates a large relative increase in DO of ~160% while the Academic Model simulates a modest relative increase of ~5%.

The nutrient reduction scenarios cause both models to exhibit a larger increase in DO concentrations during the summer than in the winter (Fig. 3.6). During the summer in the mesohaline main stem (CB3MH and CB4MH), the Academic Model simulates a slightly larger increase in bottom DO (Fig. 3.6a) than the Regulatory Model (Fig. 3.6b); however, the difference between the two standard runs (Fig. 3.6c; black line) is larger than the difference between their changes in DO (Fig. 3.6c; orange line). In other words, the change in DO simulated by the Academic Model is more similar to that of the Regulatory Model than the absolute DO concentration of the Academic Model is to that of the Regulatory Model.

3.5.3 Comparison of water quality standard attainment for both models

When the results of the model simulations are put through the CBP protocol for determining whether or not WQS would be met with the TMDL-WIP nutrient reduction, the two models both predict surprisingly similar results (Table 3.4; Fig. 3.7). With 0% red in the stoplight analysis for Open Water Summer and Open Water Non-Summer, both models simulate a complete attainment of WQS in these habitats. While the Open Water Non-Summer was widely in attainment before nutrient reduction, the Open Water Summer designated use was not. The models begin to diverge in Deep Water where the percent agreement in the stoplight analysis falls below 95% for four of the segments. In only two of these segments do the models disagree on percent non-attainment (red) by >1%. Differences are larger in the Deep Channel waters where four of the eight segments disagree by >1%. In general, the greatest differences occur in the mid-Bay main stem and mid-Bay tributaries.

A further examination of the Deep Water and Deep Channel results highlights that while the model simulations of WQS attainment differ most in these portions of the water column, the increase in WQS attainment compared to the 1993 – 1995 levels is still quite similar between the models (Fig. 3.8). Based on observations from 1993 – 1995, all Deep Channel and the majority of Deep Water volumes were out of attainment with many segments exhibiting substantial percentages of red (Fig. 3.8a,d). In the Deep Channel, both models simulate a considerable reduction in non-attainment (Fig. 3.8e,f). In the Deep Water, the pattern of non-attainment diverges between the models particularly in the tributaries with the Academic Model generating non-attainment in the Rappahannock and Patuxent Rivers and the Regulatory Model generating non-attainment in the Patapsco and Chester Rivers (Fig. 3.8b,c). Both models identify issues in the Eastern Bay. Along the main stem, the non-attainment simulated in the Academic Model is isolated to CB4MH. In the Regulatory Model, the non-attainment spans the entire mid-Bay from CB3MH to CB5MH. Overall, however, both models simulate a dramatic improvement in WQS attainment compared to the 1993 – 1995 levels (Fig. 3.8b,c compared to Fig. 3.8a; Fig. 3.8e,f compared to 3.Fig. 8d) and many of the differences between the models are due to non-attainments of << 1%.

3.5.4 Comparison of the impact of 3-year period

The two models behaved similarly across all eight 3-year periods examined, with both models exhibiting higher non-attainment of WQS during wet periods (Fig. 3.9). As was seen for the 1993 – 1995 period (Fig. 3.7, 3.8), the Deep Water generally exhibits the largest percent non-attainment for both models regardless of which years are examined. Deep-water non-attainment in the Academic Model is below ~0.4% for each time period examined except for 1996 – 1998. This corresponds to a 3-year period of prolonged high flows. While the Regulatory Model also simulates the highest percent Deep Water non-attainment for 1996 – 1998, it is not as large of a difference between that time period and the other wet periods (1993 – 1995, 1994 – 1996) as is seen for the Academic Model. Non-attainment in the Deep Channel for the Academic Model is also particularly high in 1996 – 1998, resulting in one of the few instances where the Academic Model exhibits greater non-attainment than the Regulatory Model. However, even considering the variability of non-attainment, the non-attainment for any given designated use across all eight potential time periods for both models does not go above 1.6% non-attainment and averages much less than 1.0% non-attainment. As a result, while there is variability between the 3-year periods, the specific time period chosen does not have a major impact on the total non-attainment.

3.5.5 Examination of Confidence Index

Calculation of the Confidence Index (CI), based on the three similarity metrics described in the previous section, reveals a high degree of confidence for the majority of Bay segments (Table 3.5, Fig. 3.10). Exceptions include the Chester River (CHSMH), which received a negative score in the critical period metric, the Eastern Bay (EASMH), which has the lowest degree of similarity in the regression statistics metric, and the central mid-Bay (CB4MH), which scored low for all three metrics. Across the rest of the Bay, the two models are very similar for the three metrics examined (CI > 0.75) lending a high degree of confidence in their projections of whether or not WQS are attained. The CI for Mobjack Bay (MOBPH), Pocomoke Sound (POCMH), lower James River (JMSPH), Potomac River (POTMH), tidal fresh main stem (CB1TF), Patuxent River (PAXMH), and Tangier Sound (TANMH) are particularly high, leading to a high confidence in these WQS attainment projections. Overall, the models are most similar in terms of the designated use metric, exhibit the largest spread among scores in the critical period metric, and produce the lowest average score in the regression statistics metric.

3.6 Discussion

3.6.1 How do Chesapeake Bay models compare in terms of how nutrient reduction impact DO concentrations?

- The Academic Model simulates a larger absolute improvement in DO compared to the Regulatory Model, but both models simulate a similar relative improvement in DO.

Along the main stem of the Chesapeake Bay, the Academic Model simulates a higher summer absolute increase in bottom DO as a result of nutrient reduction than the Regulatory Model. This difference continues up the water column, attenuating to the surface where the models perform quite similarly. The difference in the absolute change in bottom DO between the models is potentially due to the positive bias of DO concentrations in the Academic Model, differences in parameter tuning, and the relative simplicity of the water column biogeochemistry in the Academic Model compared to the Regulatory Model. At the surface, the decreased input of nutrients causes both models to predict a decrease in DO. This decrease in DO is a result of decreased production in the surface layer of the water column and is consistent with other modeling studies exploring the impact of nutrients on water quality (e.g. Testa et al., 2014). The prediction of both models of an increase in DO at the surface for the northern-most stations is likely a result of the decrease in sediment in the TMDL-WIP scenario, which alleviates light limitation on production. This area has the highest turbidity, and consequently benefits most from the reduction in sediment delivered to the Bay.

While the models disagree somewhat on the absolute change in bottom DO as a result of the nutrient reductions, they are surprisingly similar in terms of the relative change in DO at the bottom and throughout the water column. The only important difference between their simulated relative changes in DO is in the middle of the main stem of the Bay at depth, where the magnitude of the relative change is similar between the models but the Regulatory Model places the maximum impact further north than does the Academic Model. This has important ramifications for the assessment of water quality standards since the Regulatory Model simulates the greatest impact in segment CB3MH, while the Academic Model places the largest impact squarely in CB4MH.

3.6.2 How do Chesapeake Bay models compare in terms of whether nutrient reductions will lead to the desired attainment of water quality?

- The models predict very similar levels of water quality standards attainment throughout most of the Bay, with all but one segment exhibiting a >90% similarity between the models. Furthermore, the impact of different baseline hydrologic conditions as a result of the choice of 3-year study period did not significantly impact the overall attainment of water quality standards for either model.

Water quality observations from 1993 – 1995 demonstrate that there were large areas throughout the Deep Channel and Deep Water of the Bay where water quality standards were not being met (Fig. 3.8). Both models predict that the vast majority of those exceedances will be alleviated once the TMDL-WIP nutrient reduction is in place. The two segments where the models disagree most are CB4MH and CHSMH (Table 3.4, Fig. 3.7). The former is a direct result of the spatial dissimilarity in where the largest relative impact of the nutrient reduction is located. Since the Regulatory Model simulates its largest impact in CB3MH, CB4MH does not pass the WQS attainment as it does for the Academic Model, which simulates the largest impact in CB4MH. In the Chester River (CHSMH), the Regulatory Model predicts that even with the required nutrient load reductions in place, 16% of the Deep Channel will not meet the required water quality levels, whereas the Academic Model is fully in attainment. This is potentially due to both a mischaracterization in the Regulatory Model of oxygen concentrations in the lateral freshwater flow entering the Chester River as well as the bathymetric grid of the Academic Model being far too shallow along the central river channel. The Regulatory Model issue in the Chester River has been identified and is currently being remedied (C. Cerco, pers. comm.) and both models are currently being used in a shallow water study of the Chester River to improve simulations (Friedrichs et al. 2012).

The difference between the model results (Fig. 3.7, Table 3.4) is exaggerated in terms of the true attainment of WQS because the similarity comparison is evaluated across green, yellow, and red, while the actual TMDL combines green and yellow together as the passing grade and only the red is identified as exceedance, or non-attainment. Because a large fraction of the stoplight results are green, small differences in percent similarity can have important ramifications. Examining only the red columns, there are only two segments where the models disagree by more than two percent: CB4MH and CHSMH.

In order for water quality levels to pass the regulatory minimums as mandated by the TMDL regulation, all areas must pass the WQS with no exceedances. To account for some numerical errors in calculating the volumes and percent space and time of attainment, the TMDL allows for a 1% buffer for all segments and all designated uses. Therefore, a stoplight analysis that exhibits 1% or less of “red” can still be considered in attainment. Unfortunately, even with the 1% rule, some segments and designated uses still do not meet WQS for both models. In the development of the TMDL, the Regulatory Model was tested using progressively stringent nutrient reduction scenarios to explore just how much of a potential impact nutrient reductions could have. In some segments, the model never went to full attainment even with aggressive nutrient reductions. Since all of the problem segments were located in the Maryland portion of the Bay, Maryland was able to account for these segments that would not fall into traditional attainment by allotting a “variance.” The variances (Table 3.3) are defined by Maryland state regulation rather than in the TMDL regulation and only impact those segments identified as unable to meet WQS with the mandated nutrient reduction. The regulation states that the variances must be reviewed every three years as the modeling and understanding of the ecosystem are continually improving.

Only in the CB3MH Deep Water does the Academic Model require a variance in order to fall within attainment. The Regulatory Model, on the other hand, requires variances in five of the segments/designated uses. As discussed previously, the Chester River (CHSMH) is a special case that is currently being studied by the CBP. However, the iteration of the Regulatory Model used in this analysis results in the CHSMH Deep Water falling out of attainment even with the variance. The difference in whether or not the models need the variances to meet WQS is important to note. The results of the Academic Model potentially indicate that some of the variances are the result of modeling artifacts and not the environment. This is critical to note, since the WQS are biologically based and exceedances of 16% or even 7%, as allowed by the variances, could prove biologically detrimental considering there are many important Bay species unable to tolerate low-DO conditions (USEPA 2003).

While there are differences between the models as to the level of WQS attainment, neither model exhibited a large sensitivity to the choice of study period when examining non-attainment relative to the entire volume of a designated use. The EPA underwent a complex process to identify the best 3-year period on which to base the hydrologic conditions of the TMDL. While there are certainly individual segments that exhibit sensitivity to 3-year period, the research performed in this study (Fig. 3.9) indicates that when looking at an entire designated use, the models are relatively insensitive to the baseline hydrologic conditions in terms of attainment of WQS. In most cases, both models simulate an exceedance of less than 1% across entire designated uses. The models also similarly exhibit changes between 3-year periods with a higher percent exceedance in the wetter 3-year periods than in the drier ones. This is to be expected, as there is an observed correlation between years with high freshwater flow and years with large hypoxic volumes (Murphy et al. 2011) and it therefore is beneficial to use a wetter 3-year period in an effort to employ a more conservative approach.

3.6.3 Where is the location in the Bay with the greatest uncertainty in the impact of nutrient reduction on the desired attainment of water quality?

- The greatest uncertainty in the impact of nutrient reduction on the attainment of water quality standards is in the Chester River, Eastern Bay, and upper mid-Bay main stem. These locations historically exhibit some of the Bay's lowest summer DO concentrations.

The overall goal of this research was to establish a level of confidence in the attainment of WQS resulting from required nutrient reductions. The Confidence Index developed here offers insight into the segments where the models behave most similarly, i.e. where we have high relative confidence in their projection of the impact of reduced nutrient inputs, and where they behave least similarly, i.e. where we have relatively low confidence in the impact of reduced nutrient inputs.

The two segments with the lowest CI were the Chester River and Eastern Bay, with both segments exhibiting some of the lowest scores for all three CI metrics. The Academic Model never simulated a Deep Water exceedance in the Chester River, while the Regulatory Model Deep Water exceedance fluctuated between 3-22% exceedance. This is by far the location and metric where the models were most dissimilar and is likely a result of deficiencies in both of the models in the Chester River. In the Eastern Bay, the bathymetric grid of the Academic Model is far too shallow, which limits the development of a true delineation between the surface mixed layer and the bottom layer. As a result, the mean y-intercept of the regressions differs by 2.16 mg L^{-1}, a substantial difference when examining hypoxic waters.

The upper mid-Bay CB4MH segment exhibits the next lowest confidence. The low CI score for CB4MH is a result of the spatial difference in the location of maximum relative bottom DO increase for the two models, which impacts all three of the metrics. Because this difference lies at the boundary of CB4MH and CB3MH, it is possible that if the delineation between the two segments were slightly further south, the CI would be higher. This raises the question of which model, if either, is correctly simulating the location of greatest impact.

3.6.4 Within the modeling and assessment approach, what is the source of greatest uncertainty in the impact of nutrient reduction on the attainment of water quality?

- The regressions derived from the raw model output are the greatest source of uncertainty in the process of evaluating water quality standards attainment.

Three main sources of uncertainty in estimating the impact of nutrient reduction on the attainment of water quality were analyzed at the segment level. The designated use metric evaluated the average percent similarity across the applicable designated uses for each segment. The critical period metric compared the non-attainment between the models for each of the eight three-year periods. The regression statistic metric compared the average slope, y-intercept, and correlation of the regressions used in the WQS attainment methodology.

The high degree of similarity between how the two models perform in the various designated uses and critical periods results in a relatively high degree of confidence when examining these two metrics. The high values for the designated use and critical period metrics are partially due to the fact that both metrics utilize a threshold for pass/fail categorization. To attempt to account for this, the designated use metric compares the green/yellow/red individually but since they must all add to 100%, there is only so much possible spread in the scores. In the critical period metric, only the red is examined but weighting the percent non-attainment by volume gives the large mid-Bay mainstem segments much more influence on the metric.

Unlike the designated use and critical period metrics, the regression statistics differ considerably between the two models, thus leading to low values in this component of the CI. This is partially because the regression statistics do not have an upper or lower bound, nor do they weight segments by volume. Because the regression statistics metric utilizes the slope, y-intercept, and correlation, the scores are sensitive to large differences between the models in any one of the three sub-metrics. These large differences occur throughout the water column and across the Bay, but there are some locations and times more prone to large discrepancies between the models. The most important distinction is that the regression statistic scores are only for the summer months and the models are much more similar in terms of all three regression statistic sub-metrics during the non-summer.

Of the three individual sub-metrics that go into the regression statistic CI score, the y-intercept exhibited the greatest difference between the two models. While the majority of the largest discrepancies between the models in the slope and correlation were in deep areas that generally experience annual hypoxia that causes the normality assumption of the regressions to be violated, the discrepancies in the y-intercept are throughout the water column. While some of these large differences occur near the DO-replete surface and therefore do not have a large impact on the pass/fail nature of the stoplight analysis, there are many deeper locations where large differences in y-intercept (for example, > 3mg L^{-1}) could potentially impact the stoplight analysis since the y-intercept value is added to the observed DO in the regression. While it is difficult to determine isolated problems with any of the >1,100 individual regressions from a single model, the ability to compare regression statistics between multiple models can isolate the regressions that are most different between the models and thus help to identify problem locations and times.

3.7 Summary and Conclusions

Both the Regulatory and Academic Models analyzed in this study simulate a similar level of attainment of Chesapeake Bay water quality standards as a result of regulatory nutrient reduction. While the models differ in their simulated absolute change in dissolved oxygen concentrations resulting from the nutrient reduction scenario, the relative change in DO between the models is quite similar. Since the methodology for evaluating the impact of nutrient reduction is based on a relative change within each model between the standard run and the nutrient reduction scenario, the models can differ in their simulation of the absolute change in DO while still simulating a similar level of water quality standards attainment.

Although the predicted attainment of water quality standards between the models is similar, there are locations in the Bay where there is relatively low uncertainty (high confidence), and locations where there is relatively high uncertainty (low confidence) in these projections. The parts of the Bay where uncertainty is greatest are the Chester River (CHSMH) and Eastern Bay (EASMH). The area of the main stem (CB4MH) between Annapolis, MD and the Patuxent River is also identified as a low confidence area, albeit slightly higher than for the Chester River and Eastern Bay. While specific modeling issues can potentially explain the particularly high uncertainties in the Chester River and Eastern Bay, the high uncertainty in the mid-Bay main stem is primarily a result of the models differing in the location of greatest impact from the nutrient reduction scenario with the Regulatory Model placing the greatest impact slightly further north than the Academic Model. The greatest source of uncertainty identified in the process of evaluating whether or not water quality standards will be met was the regressions derived from the raw model output. Although individual outlier regressions will not severely impair the overall analysis of water quality standard attainment, these regressions are critical to the methodology and their occasional lack of normality should be examined and modified in future updates of the TMDL.

Although this study identified locations and sources of uncertainty in estimates of the attainment of water quality resulting from nutrient reductions, overall the results presented here highlight that the similarities between the two sets of model results far outweighed the differences. This lends greater confidence in the anticipated impact of the regulated nutrient reduction of the Chesapeake Bay TMDL. Furthermore, the framework for assessing confidence in model predictions of water quality standard attainment, via the Confidence Index, can be expanded beyond the two models evaluated in this research.

While this study utilizes a multiple model approach to evaluate confidence in model projections of the future, the future examined is one with similar environmental and climatological conditions as the present day. This leads to the question of whether or not these results would stand if climate change impacts were added to the analysis. Although this study demonstrated that the TMDL is likely to eventually produce the required DO improvements under the current climate, it is not clear whether the established nutrient loads will be adequate under near-term future climate conditions that include rising temperature and sea level along with changes in precipitation patterns.

Tables

Table 3.1: Observation stations and segments shown in Figure 1. Asterisk (*) indicates the 25 Stations used in skill assessment.

Segment	Stations used in CI analysis
CB1TF	CB1.1, CB2.1
CB2OH	*CB2.2, CB3.1
PATMH	WT5.1
CB3MH	*CB3.2, *CB3.3C
CHSMH	*ET4.2
CB4MH	*CB4.1C, *CB4.2C, *CB4.3C, *CB4.4
EASMH	*EE1.1
CHOMH1	*EE2.1
PAXMH	LE1.1, LE1.2, *LE1.3
CB5MH	*CB5.1, *CB5.2, *CB5.3, *CB5.4
TANMH	EE3.1, *EE3.2
POTMH	RET2.4, *LE2.2, LE2.3
POCMH	*EE3.4
RPPMH	LE3.1, *LE3.2, LE3.4
CB6PH	CB6.1, *CB6.2, CB6.3, *CB6.4
CB7PH	*CB7.1, CB7.2, CB7.3, *CB7.4
YRKPH	LE4.2, *LE4.3
MOBPH	WE4.1
JMSPH	LE5.4, LE5.5
CB8PH	*CB8.1, CB8.1E

Table 3.2: Dissolved oxygen Water Quality Standards (WQS) by designated use (adapted from Tango and Batiuk, 2013).

Designated Use	Dissolved Oxygen Water Quality Standard	Rationale	Timeframe
Open Water	30-day mean ≥ 5.0 mg/L (tidal habitats with salinity ≥ 0.5 PSU)	Protects growth of larval, juvenile, and adult fish and shellfish as well as threatened/endangered species	Year-round
	7-day mean ≥ 4.0 mg/L	Protects survival of open water fish larvae	
	Instantaneous minimum ≥ 3.2 mg/L	Protects survival of threatened/endangered sturgeon species	
Deep Water	30-day mean ≥ 3.0 mg/L	Protects survival and recruitment of Bay anchovy eggs and larvae	June 1 – September 30
	1-day mean ≥ 2.3 mg/L	Protects survival of open water juvenile and adult fish	
	Instantaneous minimum ≥ 1.7 mg/L	Protects survival of Bay anchovy eggs and larvae	
Deep Channel	Instantaneous minimum ≥ 1.0 mg/L	Protects survival of bottom-dwelling worms and clams	June 1 – September 30

Table 3.3: Variances allowed in certain Maryland segments.

Designated Use	Segment	Variance
Deep Water	CB4MH	7%
	PATMH	7%
Deep Channel	CB4MH	2%
	EASMH	2%
	CHSMH	16%

Table 3.4: Percent similarities (%Sim, bolded) and stoplight analysis results of green (%Gn), yellow (%Ye), and red (%Rd) percentages of the Regulatory Model (top, gray shading) and Academic Model (bottom) for 1993-1995 water quality standard assessment.

	Deep Channel				**Deep Water**				**Open Water Summer**				**Open Water Non-Summer**			
Segment	**% Sim**	%Gn	%Ye	%Rd	**% Sim**	%Gn	%Ye	%Rd	**% Sim**	%Gn	%Ye	%Rd	**% Sim**	%Gn	%Ye	%Rd
		Regulatory / Academic				Regulatory / Academic				Regulatory / Academic				Regulatory / Academic		
CB1TF									**100**	96	4	0	**100**	98	2	0
										96	4	0		98	2	0
CB2OH									**99**	94	6	0	**100**	98	2	0
										95	5	0		98	2	0
PATMH	**100**	95	5	0	**94**	87	13	0	**100**	95	5	0	**100**	97	3	0
		95	5	0		93	7	0		95	5	0		97	3	0
CB3MH	**96**	93	7	0	**99**	91	9	0	**100**	96	4	0	**100**	97	3	0
		89	9	2		90	10	0		96	4	0		97	3	0
CHSMH	**79**	73	11	16	**91**	79	18	3	**100**	96	4	0	**100**	98	2	0
		94	6	0		88	12	0		96	4	0		98	2	0
CB4MH	**95**	88	9	3	**94**	79	16	5	**99**	95	5	0	**99**	95	5	0
		93	7	0		84	15	1		96	4	0		96	4	0
EASMH	**92**	88	10	2	**99**	87	12	1	**100**	96	4	0	**100**	97	3	0
		96	4	0		88	12	0		96	4	0		97	3	0
CHOMH1									**99**	94	6	0	**100**	98	2	0
										93	7	0		98	2	0
PAXMH					**98**	92	8	0	**97**	95	5	0	**100**	98	2	0
						90	10	0		92	8	0		98	2	0
CB5MH	**99**	97	3	0	**94**	88	11	1	**100**	96	4	0	**99**	97	3	0
		98	2	0		94	6	0		96	4	0		98	2	0
TANMH									**99**	93	7	0	**100**	98	2	0
										94	6	0		98	2	0
POTMH	**100**	98	2	0	**97**	92	8	0	**100**	96	4	0	**100**	98	2	0
		98	2	0		95	5	0		96	4	0		98	2	0
POCMH									**100**	96	4	0	**100**	98	2	0
										96	4	0		98	2	0
RPPMH	**100**	97	3	0	**96**	91	9	0	**100**	96	4	0	**100**	98	2	0
		97	3	0		87	13	0		96	4	0		98	2	0
CB6PH					**98**	94	6	0	**99**	93	7	0	**100**	98	2	0
						96	4	0		94	6	0		98	2	0
CB7PH					**100**	96	4	0	**97**	90	10	0	**100**	98	2	0
						96	4	0		93	7	0		98	2	0
YRKPH					**98**	96	4	0	**97**	96	4	0	**100**	98	2	0
						94	6	0		93	7	0		98	2	0
MOBPH									**99**	93	7	0	**100**	98	2	0
										92	8	0		98	2	0
JMSPH									**100**	95	5	0	**100**	98	2	0
										95	5	0		98	2	0
CB8PH									**99**	96	4	0	**100**	98	2	0
										95	5	0		98	2	0

Table 3.5: Values of the Confidence Index (CI) metrics for all 20 Bay segments.

Segment	**Designated Use metric**	**Critical Period metric**	**Regression Statistics metric**	**Confidence Index**
CB1TF	1	1	.47	**.82**
CB2OH	.98	1	.68	**.89**
PATMH	.94	.99	.57	**.84**
CB3MH	.95	.85	.69	**.83**
CHSMH	.70	-.34	.53	**.30**
CB4MH	.88	.55	.60	**.68**
EASMH	.91	.80	.12	**.61**
CHOMH1	.98	.99	.60	**.86**
PAXMH	.93	.97	.71	**.87**
CB5MH	.92	.96	.70	**.86**
TANMH	.98	1	.71	**.90**
POTMH	.97	.99	.67	**.88**
POCMH	1	1	.68	**.89**
RPPMH	.96	.98	.65	**.86**
CB6PH	.96	.99	.66	**.87**
CB7PH	.96	.99	.64	**.86**
YRKPH	.93	.68	.78	**.80**
MOBPH	.98	1	.75	**.91**
JMSPH	1	1	.68	**.89**
CB8PH	.98	1	.54	**.84**

Figures

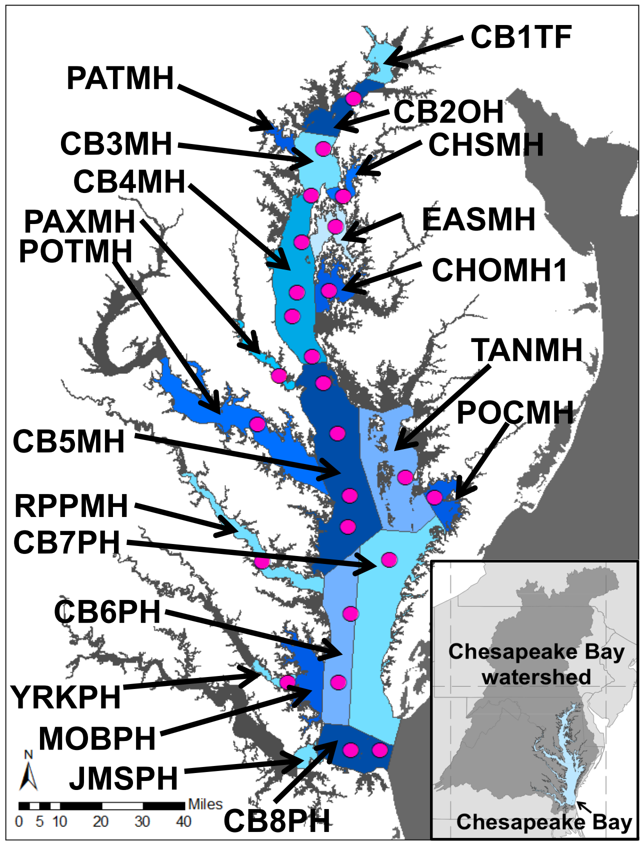


Figure 3.1: Map of the Chesapeake Bay showing CBP observation stations (pink circles) and Bay segments (blue regions) used in this analysis (Table 1). For those segments that were split in half for regulatory purposes by the TMDL along the Virginia/Maryland border, the combined segment was utilized in this analysis.

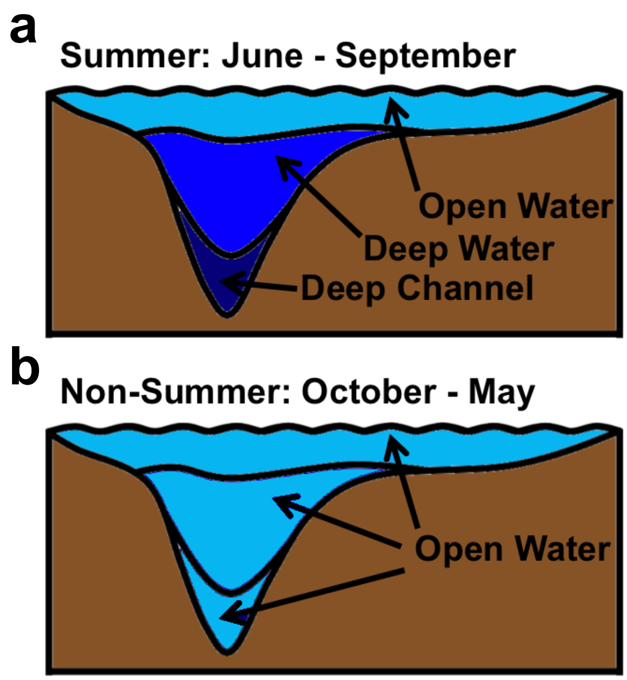


Figure 3.2: Schematic of designated uses throughout the water column for (a) summer and (b) non-summer seasons.

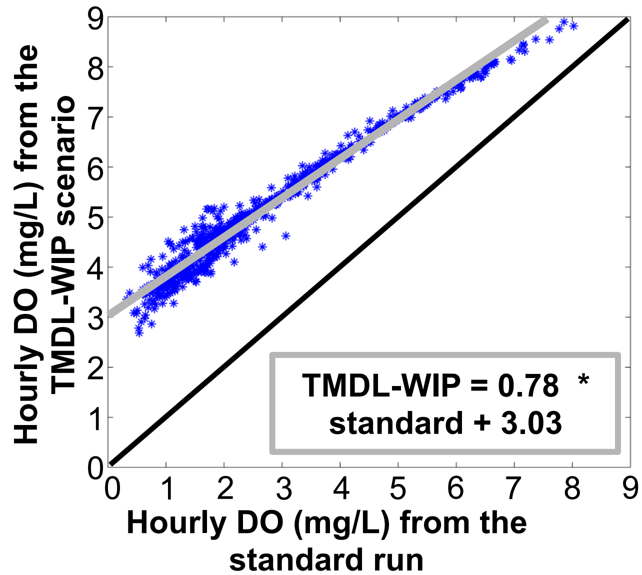


Figure 3.3: Example regression between hourly DO from the standard run and hourly DO from the TMDL-WIP scenario for a single depth at a single observation station in a single month. Black line represents the 1:1 line; gray line represents best-fit regression.

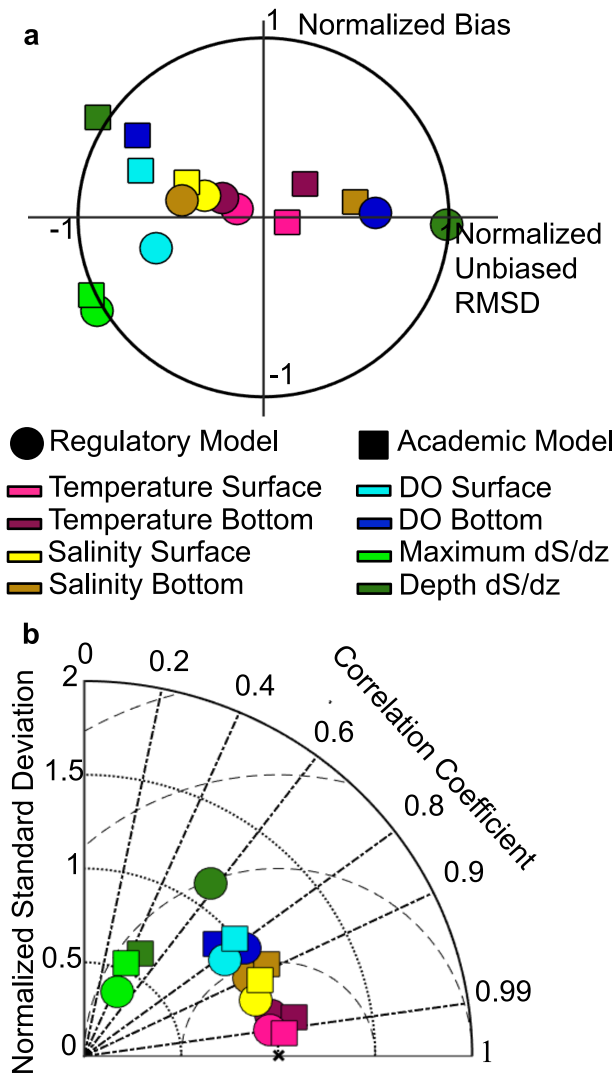


Figure 3.4: (a) Target and (b) Taylor diagrams of the Regulatory and Academic Model demonstrating the combined spatial and temporal skill for modeled surface and bottom temperature, salinity, and DO, as well as maximum stratification and depth of maximum stratification for the 25 observation stations (Table 1). A positive bias in depth of stratification means the location of stratification is too high in the water column.

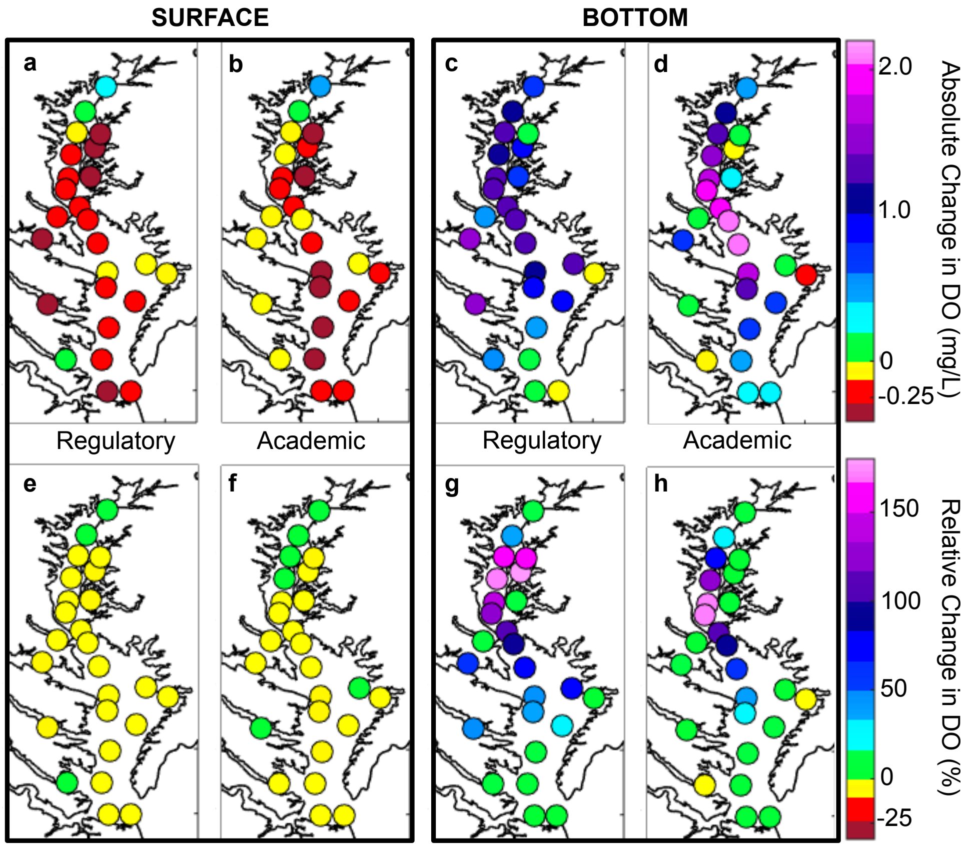


Figure 3.5: Model results at 25 observation stations illustrating the absolute difference in summer DO concentration (Scenario Run – Standard Run) at the surface (a, b) and bottom (c, d), and the relative change in DO concentration ((Scenario Run – Standard Run)/Standard Run) at the surface (e, f) and bottom (g, h). Regulatory Model results are displayed in a, c, e, and g. Academic Model results are displayed in b, d, f, and h.

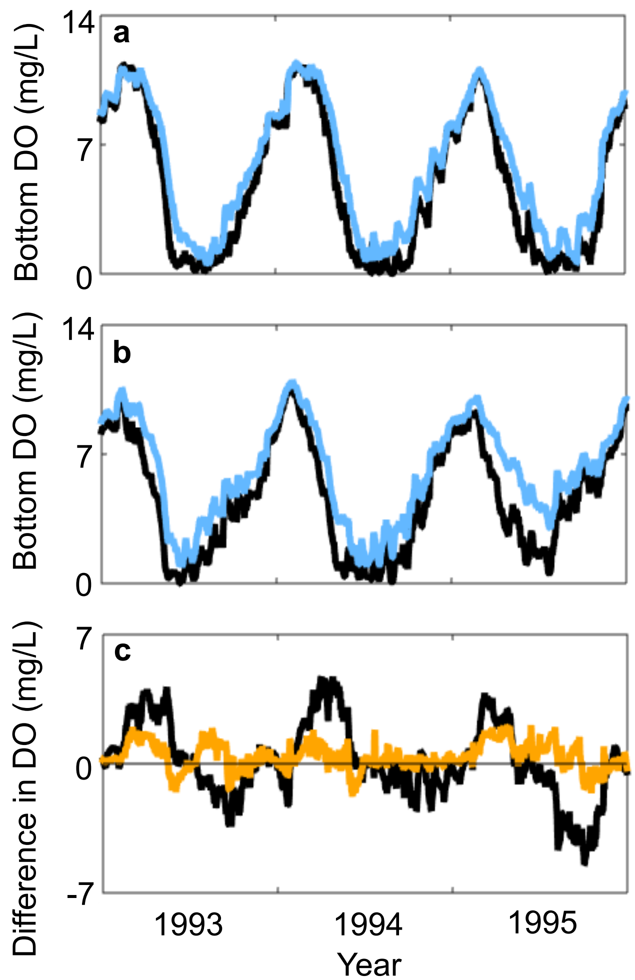


Figure 3.6: Times series of average modeled bottom DO concentrations across main stem stations in CB3MH and CB4MH for (a) the Regulatory Model and (b) the Academic Model at a representative mid-Bay deep channel station. In (a) and (b), the black line represents the standard run and the blue line represents the TMDL-WIP scenario. In (c), the black line is the standard run of the Regulatory Model minus the standard of the Academic Model; the orange line is the difference between the standard run and TMDL-WIP scenario for the Regulatory Model minus the difference between the standard run and TMDL-WIP scenario for the Academic Model.

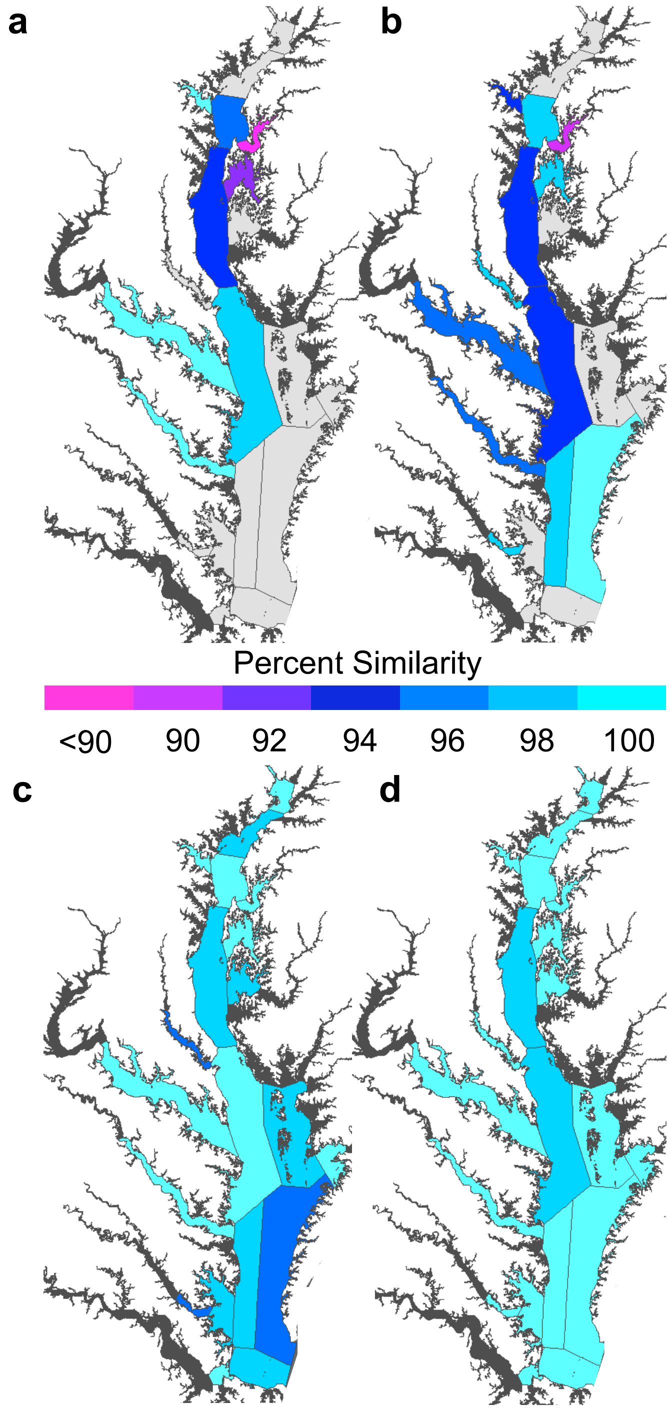


Figure 3.7: Similarity in attainment of WQS as demonstrated by percent agreement between the stoplight analysis the Regulatory Model and the Academic model for the four designated uses: Deep Channel (a), Deep Water (b), Open Water Summer (c), Open Water Non-Summer (d). Colors represent the percent of agreement between the stoplight analyses with cyan demonstrating the highest agreement and magenta demonstrating the lowest agreement.

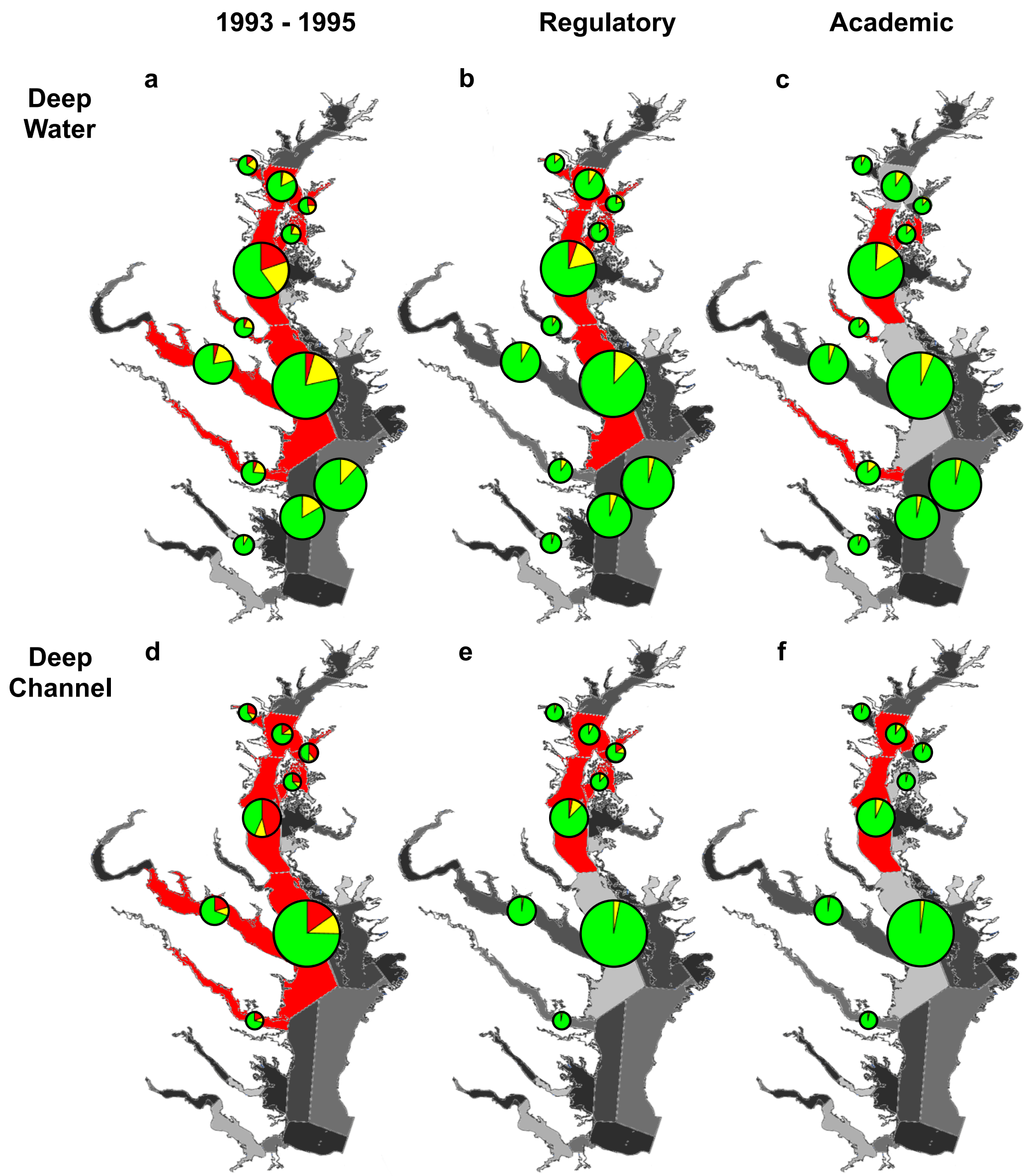


Figure 3.8: Pie charts showing attainment (green), attainment with buffer (yellow), and non-attainment (red) for the 1993-1995 observations (a, d), Regulatory Model (b, e), and Academic Model (c, f) for the Deep Water (a, b, c) and Deep Channel (d, e, f) designated uses. Size of the pies is relative to the volume of applicable water for that given segment. Segments coded in red exhibit a stoplight analysis of red that is greater than 0%.

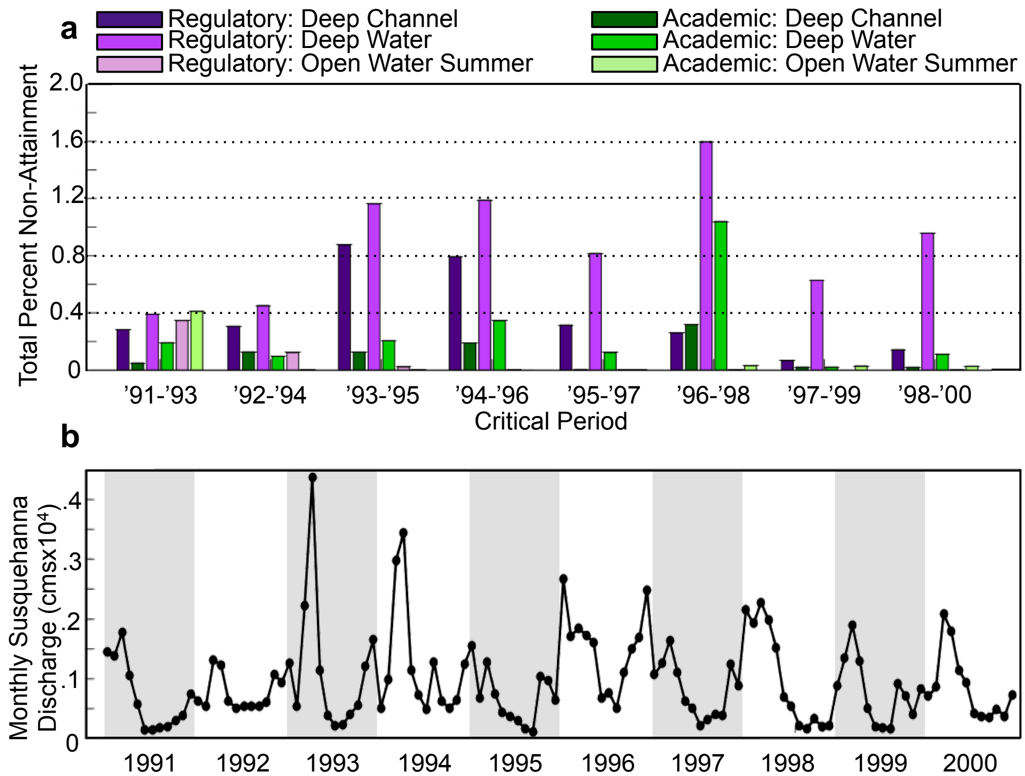


Figure 3.9: (a) Total percent non-attainment for the Regulatory Model and the Academic Model for the Deep Channel, Deep Water, and Open Water Summer. (b) Monthly Susquehanna freshwater discharge from the CBP watershed model. Open Water Non-Summer is in near full attainment and therefore is not shown.

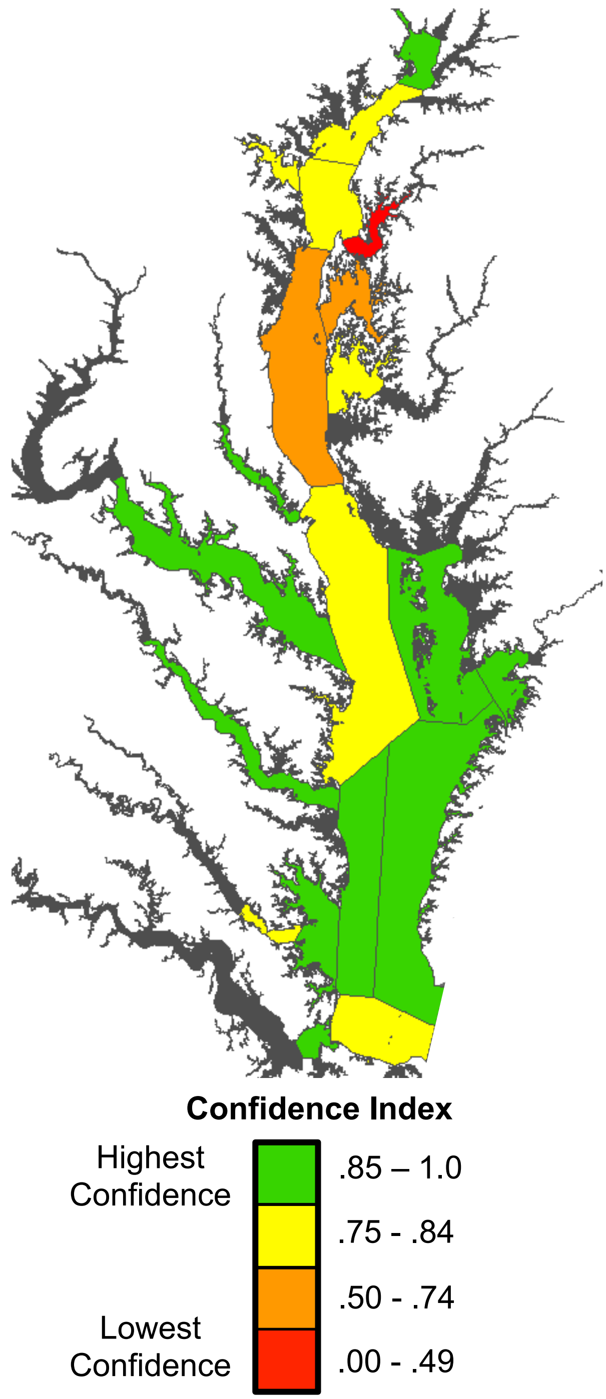


Figure 3.10: Map of Chesapeake Bay segments color coded by Confidence Index score with green indicating highest confidence and red indicating lowest confidence.

Appendix 3A. Protocol for Assessing Attainment of Water Quality Standards

The Environmental Protection Agency (EPA) and the Chesapeake Bay Program (CBP) have established a protocol for assessing the degree of attainment of the regulatory water quality standards as laid out in the Chesapeake Bay Total Maximum Daily Load (TMDL). While the official TMDL document (USEPA, 2010a) and the associated appendices outline the specific steps taken to go from the raw output of the water quality model to the pass/fail assessment of water quality standards, the protocol is complex and generalized in the text making it difficult for researchers to replicate. Further documentation in the *Journal of the American Water Resources Association*'s October 2013 "Featured Collection on the Chesapeake Bay Total Maximum Daily Load Development and Application" adds necessary insight into the methodology described in the official TMDL document. However, even with these resources, it is still quite difficult to garner a general understanding of the process used to manipulate the raw model output to determine the potential success of the TMDL. In light of that, while the documented methodology below is not perfectly comprehensive, it is meant to provide a colloquial account of the steps necessary to complete the research presented in the accompanying manuscript. Steps 1-4 involve the pre-processing of model output while steps 5-7 utilize the assessment code developed by the CBP.

1 – Standard and Scenario Runs

The first step is to conduct two model runs that will be compared to each other in order to examine the difference in dissolved oxygen (DO) concentrations between them. The "standard run" should use all forcing from the base period (1993-1995 for the TMDL). The "scenario run" will use the same forcing as the standard run expect with a set nutrient reduction applied. In the TMDL, the nutrient reduction is applied to the nutrients derived from the watershed as well as from the airshed. For the best results in the following steps, it is necessary to catalogue hourly model output.

2 – Transfer output from native grid

The code currently used by the CBP in assessing the attainment of water quality standards (WQS) is based on the vertical grid of the Regulatory Model. While the potential exists for this grid and code to be updated in the future to allow for multiple vertical grids to be used with the code, the best way for the code to currently be used with model output not from the Regulatory Model is to map the raw output to the Regulatory Model grid. This mapping only needs to be done at the grid cells that contain the 304 observation stations used in assessing WQS. Because the WQS are not constant with depth in the summer, how the output is mapped to the new grid is important to consider. For this study, we decided that it was best to linearly interpolate the Academic Model grid to the Regulatory Model grid so as to keep the integrity of the vertical profile. However, simply extending the bottom grid cell down or cutting the bottom grid cell(s) off in order to match the depth of the Regulatory Model grid is also justifiable.

3 – Pull regressions

Once the hourly output at each station has been mapped to the Regulatory Model grid, the hourly output at each grid cell for the two model runs are regressed against each other for each month. There are 1,104 grid cells for the 304 stations. This means that for each month, there will be 1,104 individual regressions. It is important to note that each of these regressions may not be unique since some of the stations will likely fall in the same horizontal grid cell. While the CBP code requires an MS Excel file with 11 values (or columns) for each regression (or row), only the first four columns need to be populated. Columns 5 – 11 can be populated with zeroes. The first column identifies which cell in the model grid the regression is for. The Regulatory Model numbers each cell in its grid and that unique identifier is used to track the regressions through time. The second column is the slope of the Ordinary Least Squares regression found by regressing the hourly DO output from the standard run (X) by the scenario run (Y). The third column is the y-intercept of that regression. The fourth column is the correlation (r) of the regression. The correlation is not used in the protocol. Rather, it is simply used as a reference for the quality of regression.

4 – Format regressions

Before the regressions can be used in the CBP code, they need to be formatted as .csv files for each month. The files must be labeled in a specific format and contain all of the regressions for all 1,104 cells for the specific month (ex. 1993_1_DO_regression_stats.csv). All of the monthly files should then be zipped together and transferred to the CBP computing network. Once this step is complete, the rest of the protocol is accomplished via the scripts developed by the CBP with minor edits required based on the individual experiment.

5 – Locate scenario-modified dataset in space

The first step of the CBP code takes the 1104 regressions for each month and applies the observed value (if there was one) for that month and grid cell to the regression to get the scenario-modified "observation" (SMO). The SMO is essentially the DO concentration that the model predicts would have been observed if the nutrient reductions had been in place when the actual observation was taken. Therefore, the SMO is the simulated future DO as a result of nutrient reduction. This SMO dataset is then mapped in space for each month. The space between observation stations is interpolated using inverse distance weighting. For months with two observations (primarily during the summer), a straight average of the two time points is used.

6 – Determine exceedance

The TMDL allows each segment/designated-use combination, referred to in the regulation as an "assessment unit", to exceed the DO minimum criteria for a certain percentage of time and space. For those assessment units where sufficient observations were available, specific biological reference curves were used to determine how much space and time an assessment unit must meet WQS. Otherwise, the assessment units were given a 10% allowable exceedance in space and time. The exceedance is measured using a cumulative frequency distribution (CDF) plotting percent time versus percent space. The percent space that exceeds the regulatory standards is found using the interpolation from the previous step. The percent time is found by ranking the applicable months (for a summer designated use that would constitute June-September for all three study years) by the percent space exceedence. From these ranks, the percent time that a given percent space exceedence can be expected is calculated using: Percent Time Exceedence = rank/(n+1). The resulting plot of Percent Space Exceeding Criteria versus Percent Time a Specified Space Exceeds Criteria gives the CDF for that given assessment unit. The assessment unit-specific curve is then compared to the allowable exceedence curve. If the assessment unit-specific curve falls beyond the allowable exceedence curve then the assessment unit is considered out of attainment of WQS. This procedure is described in Section 3.3 of the TMDL document (USEPA, 2010a).

7 – Catalog Stoplight Analysis

The final step determines how much percent space/time an assessment unit meets the criteria (green), fails the criteria but is within the allowable exceedence (yellow), and fails the criteria and is beyond the allowable exceedence (red). Ideally, all assessment units would meet the criteria and therefore fall in the green category 100% of space/time. However, since the regulations allow each assessment unit to fail minimum DO standards for a set percent space/time, the green and yellow categories can effectively be combined and recognized as meeting the WQS criteria. As a result, only those assessment units that have a percent space/time that fall into the red category are considered out of attainment and fail to meet the regulatory minimums.

Acknowledgements

This paper is the result of research funded in part by NOAA's National Centers for Coastal Ocean Science under award NA16NOS4780207 to the Virginia Institute of Marine Science (VIMS) ad by NOAA's U.S. Integrated Ocean Observing System Program Office as a subcontract to VIMS under award NA13NOS0120139 to the Southeastern University Research Association. Model output and guidance on TMDL methodology from the Chesapeake Bay Program were offered by G. Shenk, R. Tian, and P. Wang. Thank you to C. Hershner, G. Shenk, R. Hood, L. Linker, C. Friedrichs, A. Bever and J. Moriarty for comments on an initial version of this manuscript. This work was performed in part using computing facilities at the College of William and Mary, which were provided by contributions from the National Science Foundation, the Commonwealth of Virginia Equipment Trust Fund, and the Office of Naval Research.

Literature Cited

Bever, A.J., M.A.M. Friedrichs, C.T. Friedrichs, M.E. Scully, and L.W. Lanerolle. 2013. Combining observations and numerical model results to improve estimates of hypoxic volume within the Chesapeake Bay, USA. *Journal of Geophysical Research: Oceans*, 118: 4924-4944. doi: 10.1002/jgrc.20331.

Breitburg, D.L., D.W. Hondorp, L.A. Davias, and R.J. Diaz. 2009. Hypoxia, Nitrogen, and Fisheries: Integrating Effects Across Local and Global Landscapes. *Annual Review of Marine Science*, 1: 329-349. doi: 10.1146/annurev.marine.010908.163754.

BSC. 2009. Strategic Action Plan for the Environmental Protection and Rehabilitation of the Black Sea. *Black Sea Commission*, Sofia, Bulgaria.

Buchheister, A., C.F. Bonzek, J. Gartland, and R.J. Latour. 2013. Patterns and drivers of the demersal fish community of Chesapeake Bay. *Marine Ecology Progress Series*, 481: 161-180. doi: 10.3354/meps10253.

Cerco, C., S.C. Kim, and M. Noel. 2010. The 2010 Chesapeake Bay Eutrophication Model – A Report to the US Environmental Protection Agency Chesapeake Bay Program and to The US Army Engineer Baltimore District. US Army Engineer Research and Development Center, Vicksburg, MS.

Cerco, C.F., and M.R. Noel. 2013. Twenty-one-year simulation of Chesapeake Bay water quality using the CE-QUAL-ICM eutrophication model. *Journal of the American Water Resources Association*, 49: 1119-1133. doi: 10.1111/lawr.12107.

Cooper, S.R., and G.S. Brush. 1993. A 2,500-year history of anoxia and eutrophication in the Chesapeake Bay. *Estuaries*, 16: 617-626.

Curtin, P.D., G.S. Brush, and G.W. Fisher. 2001. Discovering the Chesapeake: the history of an ecosystem. Johns Hopkins University Press, Baltimore, MD.

Diaz, R.J. 2001. Overview of hypoxia around the world. *Journal of Environmental Quality*, 30: 275-281.

Druon, J.N., A. Mannino, S. Sigorini, C. McClain, M. Friedrichs, J. Wilkin, and K. Fennel. 2010. Modeling the dynamics and export of dissolved organic matter in the Northeastern U.S. continental shelf. *Estuarine Coastal Shelf Science*, 88: 488-507. doi: 10.1016/j.ecss.2010.05.010.

Feng, Y., M.A.M. Friedrichs, J. Wilkin, H. Tian, Q. Yang, E.E. Hofmann, J.D. Wiggert, and R.R. Hood. 2015. Chesapeake Bay nitrogen fluxes derived from a land-estuarine ocean biogeochemical modeling system: model description, evaluation, and nitrogen budgets. *Journal of Geophysical Research: Biogeochemistry*, 120: 1666-1695. doi: 10.1002/2015JG002931.

Fennel, K., J. Wilkin, J. Levin, J. Moisan, J. O'Reilly, and D. Haidvogel. 2006. Nitrogen cycling in the Middle Atlantic Bight: Results from a three-dimensional model and implications for the North Atlantic nitrogen budget. *Global Biogeochemical Cycles*, 20. doi: 10.1029/2005GB002456.

Friedrichs, M., K.G. Sellner, and M.A. Johnston. 2012. Using Multiple Models for Management in the Chesapeake Bay: A shallow Water Pilot Project. Chesapeake Bay Program Scientific and Technical Advisory Committee Report. No. 12-003, Edgewater, MD, 11 pp.

Hagy, J.D., W.R. Boyton, C.W. Keefe, and K.V. Wood. 2004. Hypoxia in Chesapeake Bay, 1950-2001: long-term change in relation to nutrient loading and river low. *Estuaries*, 27: 634-658.

HELCOM. 2007. HELCOM Baltic Sea Action Plan. *Helsinki Commission: Baltic Marine Environment Protection Commission*, Krakow, Poland.

Hofmann, E.E., J. Druon, K. Fennel, M. Friedrichs, D. Haidvogel, C. Lee, A. Mannino, C. McClain, R. Najjar, J. O'Reilly, D. Pollard, M. Previdi, S. Seitzinger, J. Siewert, S. Sigorini, and J. Wilkin. 2008. Eastern US continental shelf carbon budget: integrating models, data assimilation, and analysis. *Oceanography*, 21: 86-104. doi: 10.5670/oceanog.2008.70.

Hofmann, E.E., B. Cahill, K. Fennel, M.A.M. Friedrichs, K. Hyde, C. Lee, A. Mannino, R.G. Najjar, J.E. O'Reilly, J. Wilkin, and J. Xue. 2011. Modeling the Dynamics of Continental Shelf Carbon. *Annual Review of Marine Science*, 3: 93-122. doi: 10.1146/annurev-marine-120709-142740.

Hong, B., and J. Shen. 2013. Linking dynamics of transport timescale and variations of hypoxia in the Chesapeake Bay. *Journal of Geophysical Research: Oceans*, 118: 6017-6029. doi: 10.1002/2013JC008859.

HTF. 2011. Moving Forward on Gulf Hypoxia. *Gulf of Mexico Hypoxia Task Force*, U.S. Environmental Protection Agency, Washington, DC.

Ihde, T.F., I.C. Kaplan, E.A. Fulton, I.A. Gray, M. Hasan, D. Bruce, W. Slacum, and H.M. Townsend. 2016. Design and parameterization of the Chesapeake Bay Atlantis Model: A spatially explicit end-to-end ecosystem model. U.S. Department of Commerce, NOAA. NOAA Technical Memorandum NMFS-F/SPO-166.

Irby, I.D., M.A.M. Friedrichs, C.T. Friedrichs, A.J. Bever, R.R. Hood, L.W.J. Lanerolle, M. Li, L. Linker, M.E. Scully, K. Sellner, J. Shen, J. Testa, H. Wang, P. Wang, and M. Xia. 2016. Challenges associated with modeling low-oxygen waters in Chesapeake Bay: a multiple model comparison. *Biogeosciences*, 13: 2011-2028. doi: 10.5194/bg-13-2011-2016.

Johnson, B.H., K.W. Kim, R.E. Heath, B.B. Hsieh, and H.L. Butler. 1993. Validation of a three-dimensional hydrodynamic model of Chesapeake Bay. *Journal of hydraulic engineering*, 119: 1.

Jolliff, J.K., J.C. Kindle, I. Schulman, B. Penta, M.A.M. Friedrichs, R. Helber, and R.A. Arnone. 2009. Summary diagrams for coupled hydrodynamic-ecosystem model skill assessment. *Journal of Marine Systems*, 76: 64-82. doi: 10.1016/j.jmarsys.2008.05.014.

Lake, S.J. and M.J. Brush. 2015. Modeling estuarine response to load reductions in a warmer climate: York River Estuary, Virginia, USA. *Marine Ecology Progress Series*, 538: 81-98. doi: 10.3354/meps11448.

Li, Y., M. Li, and W.M. Kemp. 2015. A Budget Analysis of Bottom-Water Dissolved Oxygen in Chesapeake Bay. *Estuaries and Coasts*, doi: 10.1007/s12237-014-9928-9.

Linker, L.C., R. Dennis, G.W. Shenk, R.A. Batiuk, J. Grimm, and P. Wang. 2013a. Computing Atmospheric Nutrient Loads to the Chesapeake Bay Watershed and Tidal Waters. *Journal of the American Water Resources Association*, 49: 1025-1041. doi: 10.1111/jawr.12112.

Linker, L.C., R.A. Batiuk, G.W. Shenk, and C.F. Cerco. 2013b. Development of the Chesapeake Bay Watershed Total Maximum Daily Load Allocation. *Journal of the American Water Resources Association*, 1-21. doi: 10.1111/jawr.12105.

Linker, L.C., G.W. Shenk, P. Wang, C.F. Cerco, A.J. Butt, P.J. Tango, and R.W. Savidge. 2002. A Comparison of Chesapeake Bay Estuary Model Calibration With 1985 – 1994 Observed Data and Method of Application to Water Quality Criteria. *A Report of the Modeling Subcommittee Chesapeake Bay Program Office, Annapolis, MD*.

Mistiaen, J.A., I.E. Strand, and D. Lipton. 2003. Effects of Environmental Stress on Blue Crab (Callinectus sapidua) Harvests in Chesapeake Bay Tributaries. *Estuaries*, 26(2): 316-322.

Murphy, R.R., W.M. Kemp, and W.P. Ball. 2011. Long-term trends in Chesapeake Bay seasonal hypoxia, stratification, and nutrient loading. *Estuaries and Coasts*, 34: 1293-1309. doi: 10.1007/s12237-011-9413-7.

Nelson, R.H. 2014. How to Save the Chesapeake Bay TMDL: The Critical Role of Nutrient Offsets. *William & Mary Environmental Law and Policy Review*, 38.

Phillips, S., and B. McGee. 2014. The Economic Benefits of Cleaning Up the Chesapeake: A Valuation of the Natural Benefits Gained by Implementing the Chesapeake Clean Water Blueprint. Chesapeake Bay Foundation. http://cbf.org/ECONOMICBENEFITS

Rabalais, N.N., and R.E. Turner. 2001. Coastal Hypoxia: Consequences for Living Resources and Ecosystems. *Coastal and Estuarine Studies*, 58.

Rabotyagov, S.S., C.L. Kling, P.W. Gassman, N.N. Rabalais, and R.E. Turner. 2014. The Economics of Dead Zones: Causes, Impacts, Policy Challenges, and a Model of the Gulf of Mexico Hypoxic Zone. *Review of Environmental Economics and Policy*, 8: 58-79. doi: 10.1093/reep/ret024.

Scully, M.E. 2010. The importance of climate variability to wind-driven modulation of hypoxia in Chesapeake Bay. *Journal of Physical Oceanography*, 40: 1435-1440. doi: 10.1175/2010JPO4321.1.

Scully, M.E. 2016. Mixing of dissolved oxygen in Chesapeake Bay driven by the interaction between wind-driven circulation and estuarine bathymetry. *Journal of Geophysical Research – Oceans*, 121, doi: 10.1002/2016JC011924.

Shchepetkin, A.F., and J.C. McWilliams. 2005. The Regional Ocean Modeling System (ROMS): a split-explicit, free-surface, topography-following-coordinate oceanic model. *Ocean Modeling*, 9: 347-404. doi: 10.1016/j.ocemod.2004.08.002.

Shenk, G.W., and L.C. Linker. 2013. Development and application of the 2010 Chesapeake Bay watershed total maximum daily load model. *Journal of the American Water Resources Association*, 49: 1-15. doi: 10.1111/lawr.12109.

Tango, P.J., and R.A. Batiuk. 2013. Deriving Chesapeake Bay water quality standards. *Journal of the American Water Resources Association*, 1-18. doi: 10.1111/jawr.12108.

Taylor, K.E. 2001. Summarizing multiple aspects of models performance in a single diagram. *Journal of Geophysical Research*, 106: 7183-7192.

Testa, J.M., Y. Li, Y.J. Lee, M. Li, D.C. Brady, D.M. Di Toro, W.M. Kemp, and J.J. Fitzpatrick. 2014. Quantifying the effects of nutrient loading on dissolved O_2 cycling and hypoxia in Chesapeake Bay using a coupled hydrodynamic-biogeochemical model. *Journal of Marine Systems*, 139: 139-158. doi: 10.1016/j.jmarsys.2014.05.018.

USEPA. 1982. Chesapeake Bay Program Technical Studies: A Synthesis. U.S. Environmental Protection Agency, Washington, DC.

USEPA. 1983. Chesapeake Bay: A Framework for Action. U.S. Environmental Protection Agency, Philadelphia, PA.

USEPA. 2003. Ambient Water Quality Criteria for Dissolved Oxygen, Water Clarity and Chlorophyll *a* for the Chesapeake Bay and Its Tidal Tributaries. *U.S. Environmental Protection Agency Chesapeake Bay Program Office*, Annapolis, MD. EPA 903-R-03-002.

USEPA. 2010a. Chesapeake Bay Total Maximum Daily Load for Nitrogen, Phosphorus, and Sediment. *U.S. Environmental Protection Agency Chesapeake Bay Program Office*, Annapolis, MD.

USEPA. 2010b. Ambient Water Quality Criteria for Dissolved Oxygen, Water Clarity and Chlorophyll *a* for the Chesapeake Bay and Its Tidal Tributaries: 2010 Technical Support for Criteria Assessment Protocols Addendum. *U.S. Environmental Protection Agency Chesapeake Bay Program Office*, Annapolis, MD. EPA 903-R-10-002.

USEPA. 2012. Guide to Using Chesapeake Bay Program Water Quality Monitoring Data. *U.S. Environmental Protection Agency Chesapeake Bay Program Office*, Annapolis, MD. EPA 903-R-12-001.

USEPA Appendix H. 2010. Chesapeake Bay Total Maximum Daily Load for Nitrogen, Phosphorus, and Sediment: Appendix H – Criteria Assessment Procedures using Model Scenario Output with Bay Monitoring Data. *U.S. Environmental Protection Agency Chesapeake Bay Program Office*, Annapolis, MD.

Xia, M., and L. Jiang. 2016. Application of an Unstructured Grid-Based Water Quality Model to Chesapeake Bay and Its Adjacent Coastal Ocean. *Journal of Marine Science and Engineering*, 4, 52, doi: 10.3390/jmse4030052.

Xu, J. R.R. Hood, and S.Y. Chao. 2005. Modeling the light field in a partially mixed estuary. *Estuaries*, 28: 572-580.

Xu, J., and R.R. Hood. 2006. Modeling biogeochemical cycles in Chesapeake Bay with a coupled physical-biogeochemical model. *Coastal, Shelf, and Estuarine Science*, 69(1-2):19-46.

Xu, J., W. Long, J.D. Wiggert, L.W.J. Lanerolle, C.W. Brown, R. Murtugudde, and R.R. Hood. 2012. Climate forcing and salinity variability in Chesapeake Bay, USA. *Estuarine, Coastal, and Shelf Science*, 35: 237-261. doi: 10.1007/s12237-011-9423-5.

Chapter 4:

The competing impacts of climate change and nutrient reduction on dissolved oxygen in Chesapeake Bay

4. THE COMPETING IMPACTS OF CLIMATE CHANGE AND NUTRIENT REDUCTION ON DISSOLVED OXYGEN IN CHESAPEAKE BAY

Key Points

- Climate change will decrease dissolved oxygen conditions in the Chesapeake Bay, but this effect is smaller than the positive impacts resulting from mandated nutrient reductions
- Climate change is expected to move the onset of hypoxia ~7 days earlier by 2050
- Increased temperature is the strongest driver of future reductions in dissolved oxygen due to a combination of decreased solubility and increased biological oxygen demand

Abstract

The Chesapeake Bay region is projected to experience changes in temperature, sea level, and precipitation as a result of climate change. This research uses an estuarine hydrodynamic-biogeochemical model along with projected changes in temperature, freshwater flow, and sea level rise for a 2050 scenario to explore the impact climate change may have on future Chesapeake Bay dissolved oxygen (DO) concentrations and the potential success of nutrient reductions in attaining mandated estuarine water quality improvements. Results indicate that warming Bay waters will decrease oxygen solubility year-round, while also increasing oxygen utilization via respiration and remineralization, primarily impacting bottom oxygen in the spring. Rising sea level will increase the volume of the Bay and push saline water northward. Changes in precipitation are projected to deliver higher winter and spring freshwater flow and nutrient loads, fueling increased spring primary production. Together, these multiple climate impacts will negatively affect DO throughout the Chesapeake Bay and impact progress towards meeting water quality standards associated with the Chesapeake Bay Total Maximum Daily Load (TMDL). However, this research shows that the potential impacts of climate change will be significantly smaller than improvements in DO expected in response to the required nutrient reductions established by the TMDL, especially at the anoxic and hypoxic levels. Overall, increased temperature exhibits the strongest control on the change in future DO concentrations, primarily due to decreased solubility, while sea level rise is expected to exert a small positive impact and increased river flow is anticipated to exert a small negative impact.

4.1 Introduction

Global climate change is projected to alter the world's marine environments with coastal and estuarine systems bearing exacerbated impacts. Rising temperatures and sea levels, along with changes in precipitation patterns, have the potential to dramatically alter water quality conditions in these highly productive and increasingly human-influenced systems (Najjar et al., 2010; Altieri and Gedan, 2015). While there are multiple metrics with which to evaluate water quality, dissolved oxygen (DO) concentrations are widely used to identify systems in distress. Large volumes of hypoxic water (generally considered to be waters with DO < 2 mg L^{-1}), commonly referred to as dead zones, can be found in many coastal oceans and estuaries around the world (Diaz and Rosenberg, 2008). As the climate continues to change, it is important to evaluate the impact these changes will have on DO concentrations in critical coastal environments like the Chesapeake Bay.

Climate change is generally predicted to have a net negative effect on DO in coastal waters (Meier et al., 2011; Altieri and Gedan, 2015). Higher temperatures impact both the timing and rates of biological functions, while potentially driving long-term shifts in phytoplankton composition (Winder and Sommer, 2012). Although increased temperature is not anticipated to have a major affect on estuarine stratification, which is primarily controlled by salinity in systems such as the Chesapeake Bay (Murphy et al., 2011), the increased temperature will act to reduce the amount of oxygen a given volume of water can hold via decreased solubility. Sea level rise (SLR) can act to increase estuarine circulation (Chua and Xu, 2014), water column stratification and residence time (Hong and Shen, 2012), and water body volume. These impacts are possibly counteractive, as increasing volume and circulation can bring in high-oxygen water from the coastal ocean, while increased stratification inhibits downward mixing of the high-DO water from the surface waters. In addition, over much of the mid-Atlantic region annual precipitation is projected to increase, with increased precipitation most likely to occur during the winter/spring and in the northern part of the region (Najjar et al., 2009; IPCC Annex I, 2013). This will deliver higher river flows and nutrient loads that fuel spring productivity and produce more organic matter available for summer decomposition (Najjar et al., 2010). Changes in nutrient loading and hydrologic conditions can also alter the Bay's phytoplankton composition, changing the biomass available for eventual decomposition (Harding et al., 2015). Historical reconstructions of the mid-Atlantic region support these future climatological projections, demonstrating that the largest increases in precipitation over the last century have occurred over the northern half of the U.S. East Coast, although this is potentially due to natural variability (Yang et al., 2015a; Yang et al., 2015b).

Compounding the complicated process of projecting future water quality conditions are nutrient management efforts necessitated by anthropogenic pollution such as the Chesapeake Bay 2010 Total Maximum Daily Load (TMDL; USEPA, 2010) that was developed to improve water quality conditions in the Bay by decreasing nutrient and sediment loads. The TMDL is intended to be fully implemented by 2025 with the ultimate goal of reducing summer hypoxia (Keisman and Shenk, 2013). Examining the potential impact of climate change in light of mandated nutrient reductions is important, because the multiple impacts of climate change have the potential to render nutrient reduction levels inadequate (Altieri and Gedan, 2015).

While much of the discussion around water quality regulations focuses on hypoxia (DO < 2 mg L^{-1}), studying low-DO water that encompasses concentrations greater than hypoxic levels (DO concentrations up to 5 mg L^{-1}) is also important due to the impact of climate change on economically important fisheries. For example, not only do temperature increases impact DO concentrations, but they also increase metabolic rates in fish. This increase causes fish to experience adverse health impacts at higher and higher DO concentrations (Portner and Knust, 2007; Vaquer-Sunyer and Duarte, 2011). Further, the TMDL mandates multiple levels of minimum DO concentrations at various times and locations throughout the Chesapeake Bay (USEPA, 2010; Tango and Batiuk, 2013). While much of the regulation targets traditional hypoxia, the TMDL mandates a monthly mean DO $\geq$ 3 mg L^{-1} in the deep water of the Bay to protect the survival and recruitment of Bay anchovy eggs and larvae, and a monthly mean of DO $\geq$ 5 mg L^{-1} above the pycnocline to protect the growth of larval, juvenile, and adult fish and shellfish (Tango and Batiuk, 2013).

This study examines the impact of climate change on oxygen concentrations in the Chesapeake Bay (Fig. 1) by utilizing a coupled hydrodynamic-biogeochemical model that has previously been compared to other models (Chapter 2) and used to study the impact of the 2010 Chesapeake Bay TMDL (Chapter 3). As the TMDL stipulates a time horizon of 2025, this research assumes that the required nutrient management strategies would be in place and would be limiting nutrient delivery to their full potential by 2050. With that in mind, the present study employs projections of 2050 temperature, SLR, and watershed precipitation to examine the individual and combined impacts each variable has on anoxic (< 0.2 mg L^{-1}), hypoxic (< 2 mg L^{-1}) and low-DO (2 – 5 mg L^{-1}) water in the Chesapeake Bay. This study is structured as an initial exploration of the potential ramifications of various climate change variables on future DO concentrations in light of these nutrient reduction efforts. Future studies through the Chesapeake Hypoxia Analysis and Modeling Program (CHAMP) will build on the results from this research to include sensitivity analysis and a multiple model framework. Limitations of this research that will need to be addressed in the future are discussed in Section 4.4.5.

4.2 Methods

4.2.1 ChesROMS-ECB

The estuarine model is based on the Regional Ocean Modeling System (ROMS; Shchepetkin and McWilliams, 2005) and uses the Chesapeake Bay curvilinear horizontal grid (ChesROMS) of Xu et al. (2012) with an average wet cell resolution inside the Bay of 1.7 km. As in Feng et al. (2015), the model is configured to use the recursive MPDATA 3-D advection scheme for tracers, third-order upstream advection scheme for horizontal momentum and fourth-order centered difference for momentum in the vertical, with a 20-layer vertically stretched sigma grid. The Estuarine-Carbon-Biogeochemistry (ECB) component of the model (Feng et al., 2015) was developed originally from a continental shelf application (Hofmann et al., 2011), using dissolved organic matter cycling similar to that described in Druon et al. (2010). With only single phytoplankton and zooplankton classes and only one limiting nutrient (nitrogen), the ECB model is simpler than that employed by the Chesapeake Bay Program (Cerco et al., 2010), but is more complex than simple dissolved oxygen models that utilize a constant oxygen consumption rate (e.g. Scully, 2010; Bever et al., 2013). ChesROMS-ECB has been previously shown to adequately resolve the spatial and temporal variability of key physical and biological variables such as temperature, salinity, nitrogen, and DO (Feng et al., 2015; Chapter 2; Chapter 3).

Before using ChesROMS-ECB to determine the impact of changes in temperature on water quality parameters, the temperature dependence of the biogeochemical formulations within the model required a careful evaluation. Several biogeochemical formulations within ChesROMS-ECB did not previously include a dependence on temperature, and temperature dependence was added as part of this study (a complete list of model changes is provided in Appendix 4A). For example, temperature-dependence was introduced to the rates for maximum phytoplankton growth, zooplankton grazing/growth, nitrification, detrital solubilization, and detrital remineralization. All modifications introduce an exponential relationship between temperature and maximum rate, except for maximum phytoplankton growth. The function for phytoplankton growth is based on Lomas et al. (2002) and employs a constant growth rate below 15°C of 2.15/day, with an exponential maximum growth curve only for temperatures above 15°C. Remineralization of the dissolved organic constituents previously included temperature dependence, but to ensure consistency, these rates were modified to match the Chesapeake-specific community respiration Q_{10} values from Lomas et al. (2002).

In addition, two changes were made to improve the light attenuation parameterization in ChesROMS-ECB. First, a minimum value of 0.6/m was applied to the diffuse attenuation coefficient, based on model-data comparisons (Wang et al., 2009; Son and Wang, 2015). Second, the organic portion of the total suspended solids term in the light attenuation formulation of Feng et al. (2015) was multiplied by two, since carbon is generally considered to be half of the total weight of organic matter.

To assess the relative skill of the revised model, the skill in reproducing water quality observations at 23 stations along the Bay is compared to the skill of the earlier version of the model used in Chapters 2 and 3. The 23 stations (Table 4.1, Fig. 4.1) are assigned to four regions that are functionally delineated by salinity characteristics, with Region A representing the oligohaline, Regions B and C representing the mesohaline (and generally the lowest DO concentrations), and Region D representing the polyhaline. The updated model retained its gross skill in terms of total root mean squared difference

(RMSD) compared to the version of the model evaluated in Chapter 2 and the updated model particularly improved in skill for bottom DO in Regions C and D, primarily due to the modification for calculating light attenuation as mentioned above (see Appendix 4B).

4.2.2 Nutrient Scenarios

Consistent with Chapter 3, this study utilizes freshwater output and riverine nutrient concentrations from the Chesapeake Bay Program's Watershed Model that was used in the development of the TMDL (Shenk and Linker, 2013). The Watershed Model has undergone continual improvements with the input of stakeholders from both government and academia since 1982 (Linker et al., 2002). It should be noted that for the 2017 Mid-Point Assessment of the TMDL, a new version of the Watershed Model recently has been developed. However, to ensure comparability with Chapter 3 and to provide a reference for future climate change studies, this research uses model version 5.3.2, which is the version used in the development of the 2010 TMDL. Unless otherwise stated, hereafter, "TMDL" refers to the specific nutrient reduction associated with the application of this version of the Watershed Model.

This research assumes that the management practices required to meet the nutrient reductions mandated by the Chesapeake Bay TMDL in the absence of climate change (Shenk and Linker, 2013) are fully realized by 2050. Because of this assumption, the climate change scenarios in this research are imposed on a nutrient-reduced future. However, a brief examination of the potential impact of climate change without nutrient reduction is also explored. Because the TMDL is based on a reference time period of 1993-1995 (USEPA, 2010), these are the reference years used in this study. Fortuitously, this period includes both relatively wet years (1993, 1994) and a dry year (1995). Simulations using the TMDL reduction in nutrient concentrations are hereafter referred to as the TMDL scenarios while the base 1993 to 1995 simulations will hereafter be referred to as the Base run (Table 4.2).

4.2.3 2050 Climate Change Scenarios

A 2050 climate change time horizon was chosen because it is far enough in the future to assume full implementation of the TMDL could be realized (including nutrient transport lag effects) while also being soon enough for relatively constrained projections of climate change impacts. The climate change scenarios used in this research are primarily based on Coupled Model Intercomparison Phase 5 projections for Representative Concentration Pathway (RCP) 4.5, a mid-severity future climate scenario used in the 5th Assessment of the Intergovernmental Panel on Climate Change (IPCC), that projects a peak in emissions around mid-century combined with a stabilization of radiative forcing by 2100 (IPCC Summary, 2013). It should be noted that for 2050 projections, studies have demonstrated that the difference between RCP scenarios is smaller than the spread of individual global climate models that utilize the RCP emission scenarios (e.g., Goberville et al., 2015). The projected regional impacts for three aspects of climate change (temperature, SLR, and precipitation/rivers) have been included and are discussed below.

4.2.3.1 Temperature

By 2050, the Chesapeake Bay region is expected to experience air temperature increases greater than the global average. Specifically, the IPCC projection of median annual average atmospheric temperature increase for 2046-2065 relative to 1986-2005 for the Chesapeake Bay region is about 2°C (~0.036°C/y; IPCC Annex I, 2013), whereas the analogous global increase is projected to be 1.4°C (~0.025°C/y; IPCC Summary, 2013). Further research from the IPCC establishes that ocean warming tends to be 20 to 40% lower than the rate of atmospheric warming (Collins et al., 2013). As the Chesapeake Bay is a relatively shallow, well-mixed estuary and there has recently been an observed increase in the rate of Chesapeake Bay warming (Ding and Elmore, 2015), this research utilizes a ratio between atmospheric and ocean warming that is slightly lower than the open ocean range. A 1.75°C (~0.032°C/y) increase in Bay water temperature for 2050 relative to the mid-1990s is used in this study (Table 4.2). This value is higher than observed Chesapeake Bay warming between 1949 and 2002 of ~1°C, or ~0.02°C/y (Preston, 2004). However, Preston (2004) found evidence of increased warming in the late 1990s. The rate of warming used in this analysis is consistent with projected increases by the end of the century from downscaled global climate models (Muhling et al., 2017), and less than the average satellite derived rate of Bay surface water warming of 0.005-0.175°C/y from 1984 to 2007 (Ding and Elmore, 2015).

The 1.75°C water temperature increase was applied uniformly across time and space to biogeochemical process and oxygen solubility throughout the Bay, but the temperature increase was not applied to other physical properties or processes, such as water density gradients or meteorological forcing. Thus, increased temperature affects do not impact stratification or other physical dynamics of the Bay within the model. This approach implicitly assumes that the Bay is shallow enough that climatic warming will occur uniformly over time. Supporting this assumption, Preston (2004) found that the surface and subsurface waters of the Bay warmed at relatively similar rates, even finding that, on average, the subsurface waters warmed slightly faster than surface waters. In addition, recent trends in the intensification of early summer stratification have been found not to be due to water column temperature changes, but rather are primarily due to changes in salinity as a result of SLR and altered freshwater inputs (Murphy et al., 2011). The temperature increase scenario will hereafter be referred to as the TMDL+tempCC scenario since the increase in temperature is applied to the TMDL nutrient scenario (Table 4.2).

4.2.3.2 Sea Level Rise (SLR)

The Chesapeake Bay is also expected to incur a greater increase in sea level than the global average, and the Bay has experienced a recent acceleration in SLR along with the majority of the Mid-Atlantic coast (Sallenger et al., 2012). Boon and Mitchell (2015) found a roughly 0.1m increase in sea level in Norfolk, Virginia between 1993 and 2014. Assuming a linear extrapolation of that rate (~5mm/y), by 2050 Norfolk would expect a SLR of 0.3m relative to the mid-1990s. However, the linear extrapolation ignores the projected, and recently observed, acceleration in SLR. Incorporating anticipated acceleration, Boon and Mitchell (2015) estimate an average increase in SLR by 2050 of ~0.5m (~10mm/y) in the Chesapeake Bay relative to the relative mean sea level between 1969-2014. Using downscaled global models, Sweet et al. (2017) estimate a similar SLR in the Mid-Atlantic for 2050 under an intermediate emissions scenario. Adjusted for the accelerated relative SLR expected in the Chesapeake Bay, this research employs a 2050 SLR of 0.5m (~9mm/y) relative to the mid-1990s, which is consistent with recent regional projections (Boon and Mitchell, 2015; Sweet et al., 2017).

Model implementation of SLR follows that of Hong and Shen (2012). The 0.5m increase was added to the free water surface layer at the outer boundary. The vertical grid stretching parameters were not altered and the simulation required less than six months for the Bay to equilibrate to the SLR. The SLR scenario will hereafter be referred to as the TMDL+slrCC scenario since the 0.5m increase is applied to the TMDL scenario (Table 4.2).

4.2.3.3 River Flow

The Chesapeake Bay watershed spans a range of projected precipitation changes with the southern portion of the watershed expected to experience a lower intensity change than the northern portion of the watershed, complicating projections of precipitation change, and as a result, river flow (Najjar et al., 2009). While precipitation exerts a first order control on river flow, the projected changes in river flow derived from a watershed model can be greatly influenced by different modeling approaches to evapotranspiration. The watershed model evapotranspiration used in this research is based off of the Hargreaves-Samani equation (Hargreaves and Samani, 1982) and increased stomatal resistance due to elevated CO_2 was also included. The Hargreaves-Samani equation is a simplistic representation of evapotranspiration dynamics as it only explicitly accounts for solar radiation and temperature while not accounting for advective processes and only implicitly representing relative humidity by including the difference in maximum and minimum temperature.

The river flow projections used here are derived from average precipitation estimates from 32 Global Climate Models downscaled to a 1/8° resolution using a bias-corrected spatial disaggregation (Reclamation, 2013) and have been run through the CBP Watershed Model. Projections are based on the RCP4.5 scenario and are similar to those being used to for the CBP climate change analysis used in the 2017 Mid-Point Assessment of the TMDL (CBP, pers. comm.). Table 4.3 displays the ratio of monthly freshwater delivery to the Bay from the Susquehanna River as calculated by the CBP Watershed Model for the climate change scenario relative to the Watershed Model's base case. For simplicity, this same climate change discharge factor was applied to all rivers in ChesROMS-ECB. This is a reasonable approach given that the Susquehanna watershed accounts for > 80% of the Bay watershed area that drains directly to the main stem and is the primary source of the nutrients that drive the summer hypoxic region of the Bay between the Patapsco River in the north and the Rappahannock River in the south (Hagy et al., 2004). Overall, there is an increase in river flow applied to the model. This increase in river flow results in both an increase in freshwater discharge and an increase in nutrient delivery. The combined impact of increased freshwater flow and nutrient loads will hereafter be referred to as the TMDL+riverCC scenario (Table 4.2).

4.2.3.4 Combined Climate Change Scenario

A final scenario that combines all three of the climate change impacts was run for both the TMDL scenario and Base run. The climate change impacts applied to the TMDL nutrient reductions will hereafter be referred to as the TMDL+allCC scenario, since the combined impact of all of the climate change variables (temperature, SLR, and rivers) was applied. To establish the sensitivity of these results to the assumption that the effects of the TMDL would be fully realized by 2050, the full set of combined climate change impacts were also run on the 1993-1995 Base run (Table 4.2).

4.2.4 Dissolved Oxygen Analysis

To examine the impact climate change has on DO concentrations throughout the Chesapeake Bay in space and time, two metrics are addressed: hypoxic volume (HV) and hypoxic duration (HD). Hypoxic volume (HV) is a commonly used metric to quantify the amount of water that experiences a given level of DO concentration over a specific time (e.g. Murphy et al., 2011; Bever et al., 2013). Specifically, this study will focus on cumulative HV (CHV), calculated as the sum of each day's hypoxic volume over a year (Bever et al., 2013). Hypoxic duration (HD) is measured in days of hypoxia with a volume of >1 km^3. While traditional DO concentration levels of hypoxia (< 2 mg L^{-1}) and anoxia (< 0.2 mg L^{-1}) will be utilized, this research will also consider impacts of low-DO, defined here as DO < 5 mg L^{-1}. This level is consistent with the highest DO concentrations stipulated in the Chesapeake Bay TMDL (USEPA, 2010) and is a conservative upper bound on DO concentration found to initiate stress on marine fish (Vaquer-Sunyer and Duarte, 2008; Buchheister et al., 2013).

4.3 Results

For bottom DO concentrations, especially in the deep main stem, the impact of nutrient reduction is greater than the impact of climate change. In Region B, the biggest differences in DO due to both nutrient reduction and climate change generally occur during the draw down of bottom oxygen in the spring and early summer (Fig. 4.2). The reduction of nutrients causes a general increase in DO concentrations, which is largest in spring and early summer (April to June) during the initial drawdown of oxygen. This impact is most obvious during 1995. In contrast, the differences in these scenarios are much smaller at the surface for all three years. While not shown here, the time series results for bottom and surface DO are similar in Regions C and D, albeit slightly diminished at the bottom. The entire water column in Region A, however, responds most similarly to the Region B surface, given the shallow well-mixed waters of the northern Bay. Overall, across all regions at both the surface and bottom of the water column, the changes in DO that result from the TMDL nutrient reduction are larger than those that result from the impacts of climate change.

In examining the individual climate change factors, it is evident that the largest impacts from climate change are due to the increase in temperature and that the overall impacts are nearly additive (Fig. 4.3). As a result, the TMDL+allCC scenario is most similar to the TMDL+tempCC scenario with both scenarios exhibiting a decrease in winter/spring bottom DO in Region B of ~0.5mg L^{-1} compared to the TMDL+noCC scenario. The individual climate change effects are largest during the summer of 1995. Both the TMDL+slrCC and the TMDL+riverCC scenarios have a relatively minimal impact on bottom DO during the wet years of 1993 and 1994; however, in the dry year of 1995, the impact of SLR increases bottom DO during the spring and summer, while changes in the rivers (increased seasonality and nutrient load) suppress DO. These two essentially equal and opposite effects largely cancel each other out (Fig. 4.3).

The magnitude of the individual impacts of the climate change scenarios differed by region with Region A exhibiting the largest overall change (Table 4.4). The average change in bottom DO for Region B across the entire three-year period for a TMDL plus climate change scenario compared to the TMDL+noCC scenario was most positive for the TMDL+slrCC scenario (+0.09 mg L^{-1}) and most negative for the TMDL+tempCC scenario (-0.40 mg L^{-1}; Table 4.4). In the TMDL+allCC scenario, bottom DO decreased compared to the TDML+noCC run in all four regions with Region A exhibiting the highest total change. This is primarily due to the large negative change in the TMDL+slrCC scenario in Region A relative to its small (mostly positive) changes in the other regions. Compared to the TMDL+noCC scenario, the TMDL+allCC scenario is most similar to the TMDL+tempCC scenario in terms of bottom DO, particularly in Regions B and C. Overall, the impact of all three of the climate change scenarios is nearly linearly additive at the bottom of the water column (Table 4.4).

Examining DO throughout the entire water column, results indicate that the CHV for all of the TMDL scenarios (both with and without climate change) is less than the CHV from the Base+noCC run (Fig. 4.4). This pattern holds true for all DO levels examined (< 0.2 mg L^{-1} to < 5 mg L^{-1}). At higher DO levels (DO < 3mg L^{-1} to DO < 5mg L^{-1}) the impact of the TMDL+tempCC scenario begins to separate from the other TMDL scenarios, exerting greater influence on the TMDL+allCC scenario. At each DO level, the CHV for the dry year (1995) is much less than for the wet years (1993 and 1994) for each TMDL scenario. Furthermore, the CHV for the TMDL scenarios in the wet years is generally higher than the CHV from the Base run for the dry year. The CHV in the TMDL+slrCC and TMDL+riverCC scenarios tend to track closely to the TMDL+noCC scenario, while the TMDL+tempCC scenario is again most similar to the TMDL+allCC scenario (Fig. 4.4).

The percent change in CHV relative to the progress, or gains, made in CHV by applying the TMDL nutrient reductions varies across DO level and by scenario (Fig. 4.5). In general, the TMDL+slrCC scenario resulted in a ~0-10% increase in the improvement made by the TMDL scenario (here, an increase of gains is actually a decrease in CHV) across all DO levels and all years. In contrast, the TMDL+riverCC and TMDL+tempCC scenarios resulted in a degradation of the system, compared to the TMDL+noCC scenario. The TMDL+riverCC scenario consistently causes a loss of ~0-5% of the gains, with slightly larger losses in 1994 and 1995 at higher DO levels. The TMDL+tempCC scenario was the strongest function of DO level, with a relatively small loss of ~5% at the < 2mg L^{-1} level and a large ~40% loss at the < 5mg L^{-1} level. The combined effect of climate change (TMDL+allCC) was a net increase in CHV of more than 50% over the TMDL+noCC scenario in the wet years of 1993 and 1994 for DO < 5mg L^{-1}.

Both the TMDL+slrCC and TMDL+riverCC scenarios result in small (< 10%) changes to the improvement in CHV as a result of nutrient reduction (Fig. 4.5). In general, the TMDL+riverCC scenario exerts the smallest impact on bottom DO and CHV. The TMDL+riverCC scenario combines two separate, but linked, climate change impacts: increased freshwater flow (particularly in the winter) and increased nutrient loads (as a result of increased freshwater flow). While not shown, separate experiments isolating the impacts of flow and load demonstrated that the increase in nutrient load caused the degradation of DO concentrations in the TMDL+riverCC scenario. The TMDL+slrCC is the only scenario to consistently improve CHV (except for 1994 at DO < 5 mg L^{-1}; Fig. 4.5). However, the improvement is not consistent across DO levels or hydrologic conditions.

An increase in temperature generally maintains the greatest control on the TMDL+allCC scenario (Figs. 4.3, 4.4). The impact of temperature on DO in this analysis is due to two controls: chemical solubility and biological oxygen demand. To isolate the impact on DO of rising water temperature causing a decrease in oxygen solubility and an increase in biological oxygen demand, the differences in modeled DO computed with and without warming are computed considering only solubility effects and considering both solubility and biological oxygen demand (Fig. 4.6). Since oxygen saturation is more sensitive to changes in temperature at low temperatures, there is a larger change in DO as a result of changes in solubility during the winter even though the change in temperature is constant in time. Deviations from the change in DO due to solubility can be attributed to changes in biological oxygen demand. Overall, 65-85% of the change in DO expressed in the TMDL+tempCC scenario compared to the TMDL+noCC scenario is a result of temperature's impact on solubility (Table 4.5). The impact of biological oxygen demand is consistently negative at depth during spring and early summer, enhancing the initiation of hypoxic conditions (Fig. 4.6b). In general, solubility plays a slightly greater role at the surface than at depth, and in the northern and southern portions of the Bay than in the central main-stem (Table 4.5).

In examining the number of days the Bay experiences hypoxic and low-oxygen conditions each year, climate change acts to reduce the positive impact of the nutrient reduction (Fig. 4.7). While there is a large decrease in hypoxic duration resulting from the nutrient reduction, the TMDL+allCC scenario demonstrates that when climate change is included all levels of low-DO and hypoxia initiate an average of ~7 days earlier. This trend is not evident in the cessation of hypoxia and low-DO with climate change not always causing hypoxia to last later in the year (e.g., 1994 DO < 1 mg L^{-1}). While all three years exhibit a similar pattern and timeline of cessation of low-DO with < 0.2 mg L^{-1} ceasing 3-4 months before < 5 mg L^{-1}, each year is different in terms of initiation timing. In 1993 for the Base+noCC run, all levels of DO initiate within 2 weeks of each other. This timing holds true for the TMDL scenarios as well, but with anoxia lagging behind. In 1994 in the Base+noCC run, there is a steady progression from low-DO to anoxia over ~6 weeks. In the TMDL scenarios, that is extended to ~3 months. In 1995, the TMDL nutrient reduction results in no DO < 1 mg L^{-1} and significantly delays the onset of low-DO by up to ~3 months compared to the Base run.

Examining a north-south transect along the main stem of the Bay for July 1st, 1993 (Fig. 4.8a,c) and 1995 (Fig. 4.8b,d) reveals that nutrient reduction acts to compress the southern extent of the hypoxic zone. One similarity between all four subplots (a-d) is the vertical extent of the low-oxygen waters, which are capped by the pycnocline at ~ 5m depth. The extent and severity of anoxia and hypoxia on July 1st is much greater than the summer (May-September) average for both the Base+noCC run and TMDL+noCC scenario for both years (Fig 4.8e-h). In general, the impact of climate change is greater in the dry year (1995; Fig. 4.8j,l) than in the wet year (1993; Fig. 4.8i,k). The location of the greatest magnitude change is near the pycnocline depth (Fig. 4.8i,j) but the location of greatest percent change is below the pycnocline (Fig. 4.8k,l).

Climate change will likely cause a larger volume of the Bay to experience low-DO concentrations in both wet and dry years and under both the Base+allCC and TMDL+allCC scenarios (Fig. 4.9). While climate change does not greatly exacerbate the volume of the Bay that experiences anoxic and hypoxic conditions, climate change increases the percent of the Bay experiencing conditions of DO < 5mg L^{-1} by ~3-6 %, regardless of whether or not the TMDL nutrient reductions have occurred. Similarly, regardless of whether or not climate change occurs, the volume of the Bay experiencing low-DO under nutrient reduction is considerably lower than that in the 1993-1995 Base run nutrient conditions. Overall, the dry year (1995) results in roughly half as much of the Bay experiencing low-DO and hypoxic waters as compared to the wet years (1993, 1994).

4.4 Discussion

4.4.1 How will climate change affect the impact of nutrient reduction on dissolved oxygen in the Chesapeake Bay?

- In general, the impact of climate change will be much smaller than the impact of the TMDL nutrient reduction. However, the combined impacts of climate change will reduce the increase in DO concentrations derived from nutrient reduction with temperature being the strongest driver of this change.

In examining the individual and combined impacts of projected temperature, SLR, and river flow in 2050 on Chesapeake Bay DO concentrations, temperature exhibits a large negative impact, and river flow exhibits a small negative impact, while SLR exhibits a mixed impact depending on region but is generally positive (Figs. 4.4, 4.5; Table 4.4). The large impact of increased temperature on DO in light of nutrient reduction is consistent with other modeling research focused on the York River estuary, a tributary of the Chesapeake Bay (Lake and Brush, 2015). The present research demonstrates the importance of solubility on temperature, as the annual average impact of temperature on oxygen saturation outpaced the impact of temperature on biological functions by roughly 2:1 in the region of the Bay that experiences hypoxia. This ratio is decreased to roughly 1:1 during the spring/early summer drawdown of bottom DO in the main stem channel. Murphy et al. (2011) similarly found that increased respiration due to increased temperature potentially plays a smaller role on changes in hypoxia than the physical and chemical changes. However, it is possible that as temperature continues to increase, the ratio of impact between solubility and biological oxygen demand may shift toward a greater influence by biological oxygen demand. This is because the additional impact of further reductions in solubility will decrease as temperatures continue to rise, while biological respiration at depth and production at the surface may continue to steadily increase as temperatures continue to rise.

Both SLR and changes in river flow exert their greatest relative impact during the driest year considered (1995). The increase in winter precipitation will deliver both increased freshwater flow and increased nutrient loads and accounts for a larger percentage of the overall change in DO during the dry year of 1995 because the low-flow conditions cause the Bay to be more sensitive to changes in freshwater flow and nutrient loading. SLR also exhibits its greatest influence during 1995, causing a decrease in CHV likely influenced by an influx of high-DO water from the shelf and an overall increase in Bay volume acting to reduce the unit consumption of DO per volume given a consistent loading of organic matter. The larger impact of SLR during dry years is consistent with a study from the Delaware Bay showing that high flow dampens the salinity intrusion that results from SLR (Ross et al., 2015) and with a study in San Francisco Bay finding that the impact of SLR is limited under high flow conditions (Chua and Xu, 2014).

4.4.2 How will hypoxia change as a result of climate change?

- Hypoxic and low-DO conditions can be expected to begin about one week earlier due to climate change, with changes in volume and extent being largest at the margins and at the southern extent. Significant impacts will be felt on water with DO concentrations in the range of 2-5mg L^{-1}, and not only on hypoxic waters.

The most consistent impact across all levels of low-DO waters due to climate change is an earlier onset of hypoxic and low-DO conditions by an average of ~7 days.

While an earlier onset was identified, there was no trend in the cessation of hypoxic and low-DO conditions with climate change sometimes causing an earlier and sometimes a later cessation. Furthermore, an earlier onset of conditions is projected to occur under both nutrient-reduced and nutrient-replete futures. The pattern of earlier onset is primarily due to the additive impacts of an increase in spring biological oxygen utilization at depth and decreased solubility, both the result of the increase in temperature (Fig. 4.6b). An analysis of climate change impact on DO of an estuarine tributary of the Chesapeake Bay similarly found that hypoxic duration is likely to be extended in the future (Lake and Brush, 2015).

In terms of a change in the volume of low-DO waters, the relative impact of climate change increases with DO concentration (Figs. 4.4, 4.5). The loss of gains made by the TMDL as a result of climate change range from ~5% for DO < 0.2 mg L^{-1} to ~45% for DO < 5 mg L^{-1}. The difference between impact at anoxic versus low-DO waters is accentuated during the dry year of 1995 due to the fact that the TMDL results in no modeled DO < 1 mg L^{-1} during this year (Fig. 4.7), regardless of climate change. Even assuming base 1995 nutrient inputs, the volume and duration of anoxia under climate change in 1995 is very small.

Throughout the water column, the greatest change in DO will be at the edges of the low-DO and hypoxic zones, particularly at the southern and vertical extents (Fig. 4.8). Conversely, the smallest changes will occur in the anoxic waters where DO cannot be decreased further (Fig. 4.8). As hypoxia is capped by the pycnocline (Fig. 4.8a-h; Chapter 2), the magnitude of DO change (~ 0.5 mg L^{-1}) is not great enough to extend low-DO conditions to the DO-replete surface waters. Laterally, the largest changes in bottom DO will be in the southern extent of hypoxia and the degree of east-west compression along the main stem of the Bay. Such a change would be likely to detrimentally impact demersal fish and shellfish communities along the shallow flanks of the Bay and its tributaries.

4.4.3 How might this impact the success of the TMDL?

- Climate change may cause the current TMDL to be insufficient to meet the required water quality improvements in the Chesapeake Bay. Increased duration of low-oxygen waters is the greatest impact at anoxic and hypoxic levels.

As discussed above, this research demonstrates that the improvements in Chesapeake Bay water quality due to the TMDL nutrient reductions are much greater than the deleterious impacts of 2050 climate change; however, results also indicate that by 2050 climate change will likely decrease oxygen levels and increase both hypoxic volume and hypoxic duration. Because some locations in the Bay barely pass TMDL standards and others require special allowances to meet the standards (Chapter 3), even these small increases in anoxic and hypoxic conditions can cause locations that previously passed the water quality standards to fail under a changing climate. The DO minima in the TMDL regulations are based on both space and time criteria. Although the spatial dimension may not be greatly impacted at the anoxic and hypoxic levels, this research suggests that the temporal dimension will be. This could cause locations in the Bay that are currently projected to pass the minimum standards to fail them in light of climate change, simply due to an extension of the hypoxic season without an expansion of hypoxic volume.

With increased temperature being the primary cause of the impact of climate change on DO concentrations, it is important to consider other potential impacts increased temperature may have on the ecosystem in the context of the success of the TMDL. Temperature increases in the Chesapeake Bay are anticipated to produce temperatures outside of previously observed extremes (Muhling et al., 2017), lending increased pertinence to understanding the impact of temperature changes on meeting water quality goals. In light of this, the impact on the TMDL of a decrease in oxygen concentrations due to climate change should be viewed in conjunction with the impact increased temperature is likely to have on the species the DO levels in the TMDL were predicated on. Multiple studies have established that increasing water temperature increases metabolic rates in fish that cause them to experience negative health impacts at higher DO concentrations than they do at lower temperatures (Breitburg, 2002; Portner and Lanning, 2009; Lapointe et al., 2014). Due to those compounding impacts and the large role temperature is expected to play in regulating future DO, it may be prudent for the TMDL to elevate the mandated minimum DO levels in an effort to protect the target species. If this occurred, the impacts of climate change would likely cause a larger failure rate of TMDL standards than the current analysis demonstrates.

4.4.4 How will climate change impact DO if the TMDL nutrient reductions are not met?

- Although the relative impact of climate change is similar on a reduced nutrient future and a high nutrient future, the degree of interannual variability in hypoxia may change in a reduced versus high nutrient future due to differences in the responses of oxygen to fluctuations between dry and wet years.

The relative impact of climate change on a reduced nutrient versus a high nutrient future is similar in terms of hypoxic volume and duration. In both a low and high nutrient future, the percent of the Bay that experiences a given DO level is increased with climate change (Fig. 4.9). Further, in both cases, the impact of climate change at low-DO concentrations (< 5 mg L^{-1}) is greater than that at hypoxic levels (< 2 mg L^{-1}). In terms of relative change in DO along the main stem of the Bay, a high nutrient future is expected to experience a higher (~9-15%) change in DO concentration than a low nutrient future (~6-9%), with the largest changes occurring at the southern end of the hypoxic zone (Fig. 4.8).

The largest potential ecological difference between the two futures is in the dry year of 1995. In this year TMDL scenarios exhibited no anoxia in the Bay, regardless of whether or not climate change was occurring. This suggests that during dry years, when the nutrient reduction may be sufficient to alleviate anoxia, climate change impacts may not be large enough to overcome the hysteric or threshold level of DO initiation similar to what has been observed with hypoxic responses to nutrient loading (Kemp et al., 2009). It may seem counterintuitive, but this suggests that the interannual variability of anoxic conditions may be exacerbated in a future with nutrient reduction because the interannual percent change in anoxic conditions will be relative to ~0% in the very dry years. Because of this, when climate change is added to the TMDL nutrient reductions, there is likely to be greater disparity in terms of anoxic volume between wet and dry years. Further intensifying the difference between wet and dry years is the potential impact of nutrient storage in the watershed during dry years that is delivered to the Bay in a successive wet year, amplifying hypoxia and anoxia (Lee et al., 2016).

4.4.5 What are the limitations of this study and how can they be addressed in future work?

- Limitations include: a single watershed and estuarine model, a single change in temperature, river flow, and SLR, a neglect of lag effects and continuous/contemporary change, no changes to meteorological forcing, a simplistic approach to changes in temperature and river flow. These limitations will be addressed via the CHAMP project.

This research is a first look at the potential impacts that changes in climate may have on the efficacy of nutrient reduction efforts in the Chesapeake Bay. While this first order look identified that climate change has the potential to negatively impact DO concentrations and limit the effectiveness of current nutrient reduction regulations, more robust examinations of the problem are needed in order to adequately aid in the regulatory decision making process going forward. With that in mind, this section addresses some of the limitations of this research. Many of these limitations will be addressed by a new multiple model project called the Chesapeake Hypoxia Analysis and Modeling Program (CHAMP). The utilization of a multiple model approach to assessing the interactions of climate change and nutrient reduction will greatly enhance understanding of the potential future water quality conditions of the Chesapeake Bay and provide information to enhance management decisions.

As the present research has identified increased temperature as the largest contributor to changes in DO, future efforts should work to incorporate the impact of increased air temperature and changes in meteorological forcing on the air-sea interface and Bay hydrodynamics. However, caution should be exercised when relying on increased air temperature to increase the temperature of the Bay as there is evidence that the current rates of Bay warming can not be fully explained by the observed increase in regional air temperature (Ding and Elmore, 2015). The increased complexity can also be applied to the simulated changes in river flow derived from a watershed model. While the presented research utilized changes in monthly river flow applied to hydrologic conditions in 1993-1995, projections of future precipitation indicate changes in storm intensification and extreme events that could have dramatic effects on nutrient delivery to the Bay (Sinha et al., 2017).

Due to the uncertainty in projected changes in temperature, river flow, and SLR, constraining the sensitivity of DO to multiple levels of climatic changes will be important. This research establishes that the increase in temperature has the strongest control on DO, but that does not mean that DO concentrations are most sensitive to the bounds of potential 2050 temperature changes. While the high computational expense of running multiple sensitivity tests through complex coupled hydrodynamic-biogeochemical models can be prohibitive, establishing a range of uncertainty is critical to informed adaptive management decision-making.

The limitations that will be some of the most difficult to address are both related to timing. The first is that the present research assumes a discontinuity between the reduction or nutrients and the changes in climate. This is an unrealistic assumption due to the fact that nutrient reduction and climate change are occurring and will continue to occur contemporaneously. These changes are also not immediate but manifest over time in a continuously evolving environment. The second is that the current approach simply identifies the potential ramifications of climate change on nutrient reduction efforts but does not establish a timeline for the water quality changes as a result of nutrient reduction to occur. This means that climate change has the potential to further limit the effectiveness of nutrient reduction efforts because the impacts of climate change may be more immediate than the impacts of nutrient reduction.

4.5 Conclusions

Overall, the most striking result of this research is that the potential impact of climate change in 2050 is much smaller than the impact of the previously planned TMDL nutrient reductions, particularly at anoxic and hypoxic levels. However, the decrease in DO concentrations resulting from the combined impacts of climate change may cause portions of the Bay that are currently expected to meet water quality standards under the TMDL, to fail them. At the most stringent DO standards, this is primarily due to an increase in hypoxic duration rather than hypoxic volume, as under climate change, the onset of hypoxic conditions is projected to initiate ~7 days earlier on average across all DO concentrations 0.2 – 5 mg L^{-1}.

Changes in DO as a result of the increase in temperature dominate the combined climate change impact. While the influence of solubility on DO concentrations is the primary control on decreased DO throughout the year, the impact of increased biologic oxygen demand is most prevalent at depth in the spring to early summer, contributing to the initiation of hypoxic conditions. The impact of temperature is likely to affect low-oxygen tolerance of higher trophic levels as well by increasing metabolic rates, making species less tolerant at higher DO levels. This may result in the DO minimums mandated in the TMDL to be insufficiently able to protect key species even if the goals are met.

Both sea level rise and changes in river flow exert a greater influence on change in DO during dry, low-river flow years. Changes in river flow are likely to deliver higher freshwater flows during the winter and spring that will both deliver higher nutrient loads and increase estuarine circulation. These two effects act to impact DO concentrations oppositely, with higher loads stimulating increased primary productivity and increased estuarine circulation delivering more oxygen-rich ocean water: however, the impact of increased loads out competes greater circulation. Sea level rise exerts the only consistently positive impact of climate change on DO concentrations, increasing the effectiveness of the TMDL nutrient reductions by ~5%. However, this positive impact is undermined overall by the large negative impact of temperature.

The relative effects of climate change are similar whether the DO concentrations stipulated in the TMDL are achieved or not. In both cases, there is a slight increase in anoxic conditions, and the relative impact of climate change intensifies up to DO concentrations < 5 mg L^{-1}. The impact of the TMDL on dry years is accentuated compared to the business as usual dry years due to the greater moderating influence sea level rise exerts during low-flow conditions. This results in anoxic and hypoxic conditions to be depressed with nutrient reduction plus climate change in the dry year of 1995, but not when climate change is combined with no nutrient reduction.

Overall, this study demonstrates that climate change has the potential to limit the effectiveness of the TMDL. However, those impacts are not likely to exacerbate hypoxic conditions beyond what they were before TMDL implementation. In the assessment of the relative impact of contemporary anthropogenic nutrient and future climatic influences on DO concentrations in Chesapeake Bay, it is evident that the reduction of nutrients plays a greater role. Given that this analysis only considers a 2050 time horizon and climate impacts are expected to intensify with time, it is critical to continue to examine how the Bay may evolve in the future.

Tables

Table 4.1 Characteristics of observation stations.

Station	Latitude (^oN)	Longitude (^oW)	Station Depth (m)	Region
CB1.1	39.54794	-76.08481	6.1	A
CB2.1	39.44149	-76.02599	6.3	A
CB2.2	39.34873	-76.17579	12.4	A
CB3.1	39.2495	-76.2405	13	A
CB3.2	39.16369	-76.30631	12.1	B
CB3.3C	38.99596	-76.35967	24.3	B
CB4.1C	38.82593	-76.39945	32.2	B
CB4.2C	38.64618	-76.42127	27.2	B
CB4.3C	38.55505	-76.42794	26.9	B
CB4.4	38.41457	-76.34565	30.3	B
CB5.1	38.3187	-76.29215	34.1	C
CB5.2	38.13705	-76.22787	30.6	C
CB5.3	37.91011	-76.17137	26.9	C
CB5.4	37.80013	-76.17466	31.1	C
CB5.5	37.6918	-76.18967	17	C
CB6.1	37.58847	-76.16216	12.5	D
CB6.2	37.4868	-76.15633	10.5	D
CB6.3	37.41153	-76.15966	11.3	D
CB6.4	37.23653	-76.20799	10.2	D
CB7.1	37.68346	-75.98966	20.9	D
CB7.2	37.41153	-76.07966	20.2	D
CB7.3	37.11681	-76.12521	13.6	D
CB7.4	36.9957	-76.02048	14.2	D

Table 4.2 List of run and scenario names.

Name	Nutrients	Climate Change
Base+noCC	Realistic 1993 – 1995 conditions	None
TMDL+noCC	TMDL nutrient reductions	None
TMDL+riverCC	TMDL nutrient reductions	Rivers only (Table 4.3)
TMDL+tempCC	TMDL nutrient reductions	1.75°C increase
TMDL+slrCC	TMDL nutrient reductions	0.5m increase in sea level
TMDL+allCC	TMDL nutrient reductions	All three above changes
Base+allCC	Realistic 1993 – 1995 conditions	All three above changes

Table 4.3 Monthly in freshwater flow entering the Bay used for the TMDL+riverCC, TMDL+allCC, and Base+allCC scenarios.

Month	Freshwater change factor*
January	1.165
February	1.168
March	1.035
April	0.964
May	1.034
June	1.015
July	0.965
August	1.042
September	0.986
October	0.984
November	1.093
December	1.158

*Fractional change factor = freshwater inputs in 2050 divided by freshwater inputs in Base run

Table 4.4 Average change in bottom DO (mg L^{-1}) relative to the TMDL+noCC run for each scenario and region.

Scenario	Region A	Region B	Region C	Region D
TMDL+allCC	-0.58	-0.37	-0.44	-0.44
TMDL+slrCC	-0.21	0.09	0.04	-0.04
TMDL+riverCC	-0.01	-0.05	-0.03	-0.01
TMDL+tempCC	-0.36	-0.40	-0.44	-0.38
Additive impact of slrCC+riverCC+tempCC	-0.58	-0.36	-0.43	-0.43

Table 4.5 Percent* of 3-year average bottom DO change as a result of the temperature experiment due to solubility for each region at the surface and bottom of the water column.

Region	Surface	Bottom
A	75%	75%
B	72%	66%
C	77%	69%
D	85%	79%

*Percent calculated as the expected change in bottom DO as predicted by solubility divided by the modeled change in bottom DO.

Figures

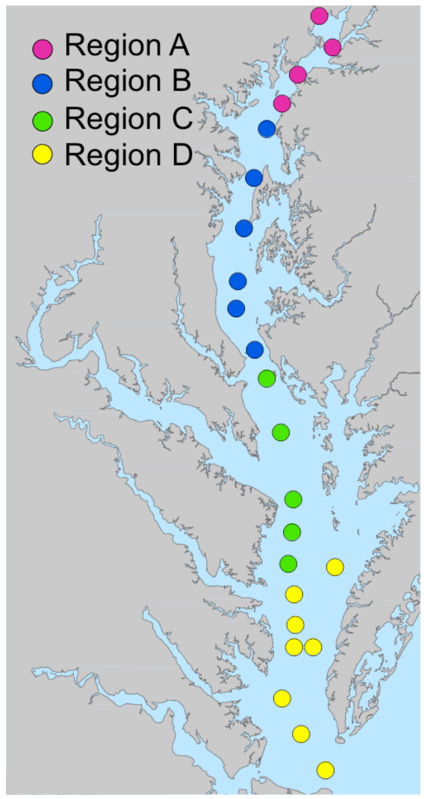


Figure 4.1 Map of the Chesapeake Bay with stations (Table 4.1) identified by region, based primarily on salinity. A: oligohaline, B & C: upper & lower mesohaline (with lowest observed DO concentrations), D: polyhaline.

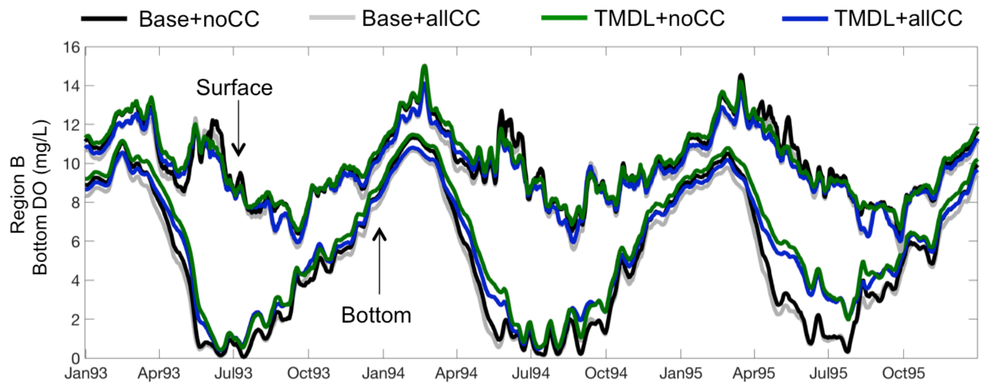


Figure 4.2 Time series of a 7-day running mean of surface and bottom oxygen concentrations computed for the average of six stations in the Chesapeake Bay upper mesohaline main stem (Region B; Table 4.1).

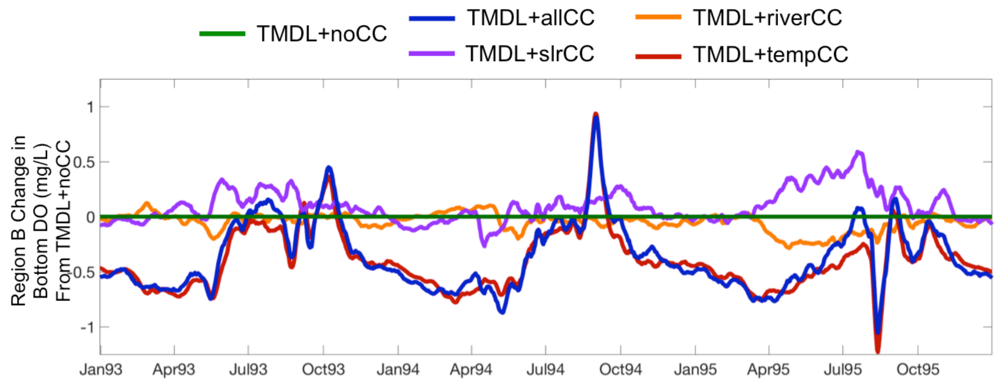


Figure 4.3 Time series of a 7-day running mean of difference in bottom oxygen concentrations between the TMDL climate change and TMDL+noCC scenarios computed for the average of six stations in the Chesapeake Bay upper mesohaline main stem (Region B; Table 4.1).

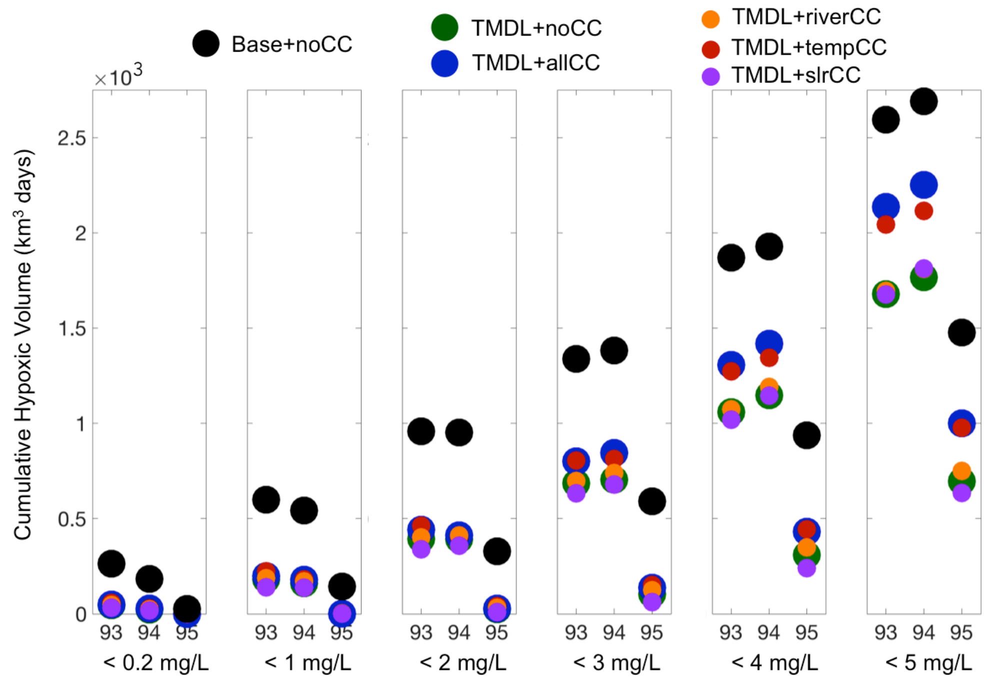


Figure 4.4 Cumulative hypoxic volume for six ranges of DO concentrations, for each of the study years.

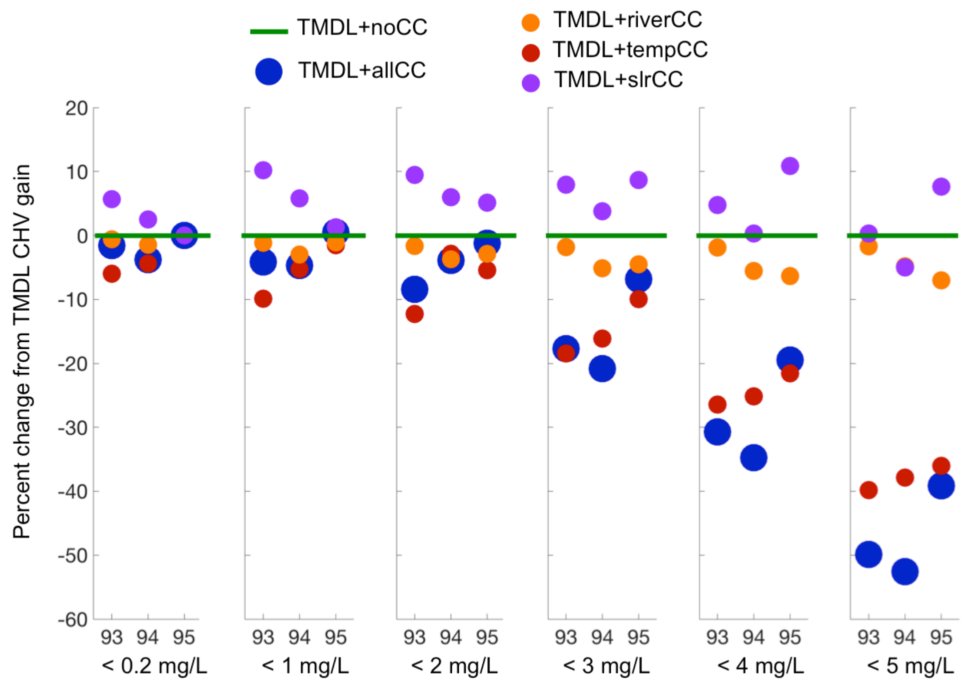


Figure 4.5 Percent change due to climate change, relative to the improvement in CHV between the TMDL+noCC scenario and Base+noCC run. Percent change in CHV gain is defined as: (TMDL+xx – TMDL+noCC)/(TMDL+noCC – Base run+noCC)).

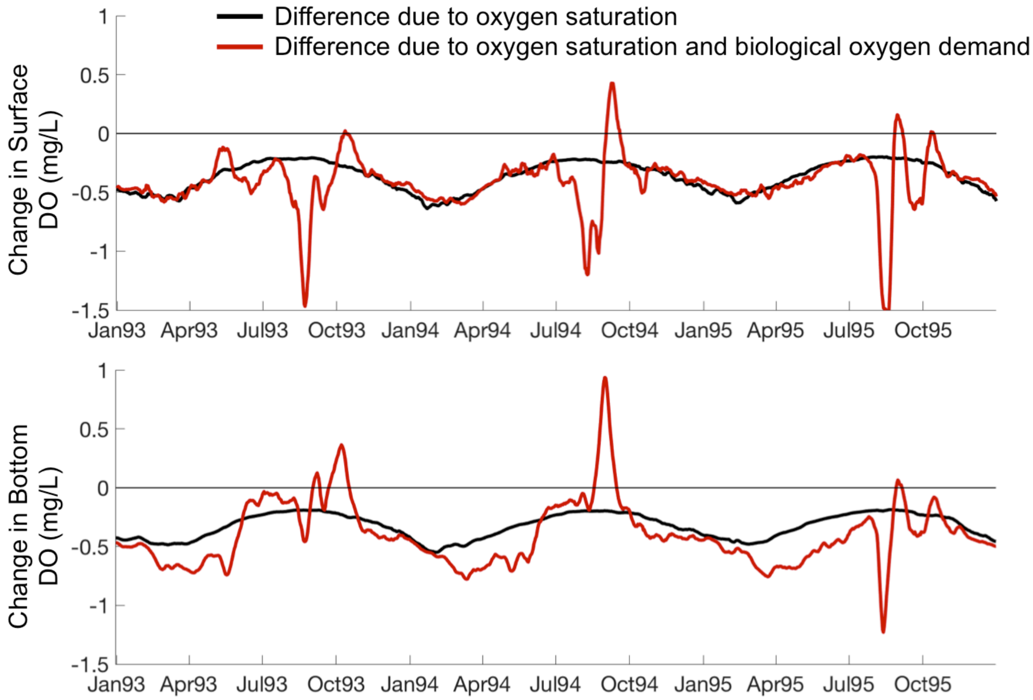


Figure 4.6 DO differences due to climate change (between the TMDL+noCC and TMDL+tempCC scenarios) averaged for the six stations in Region B (Fig. 4.1; Table 4.1) for the (a) surface, and (b) bottom of the water column. The black lines are the average change expected if only solubility was impacted by an increase in temperature. The red lines are the modeled change in DO as a result of the increase in temperature affecting both solubility and biological oxygen demand.

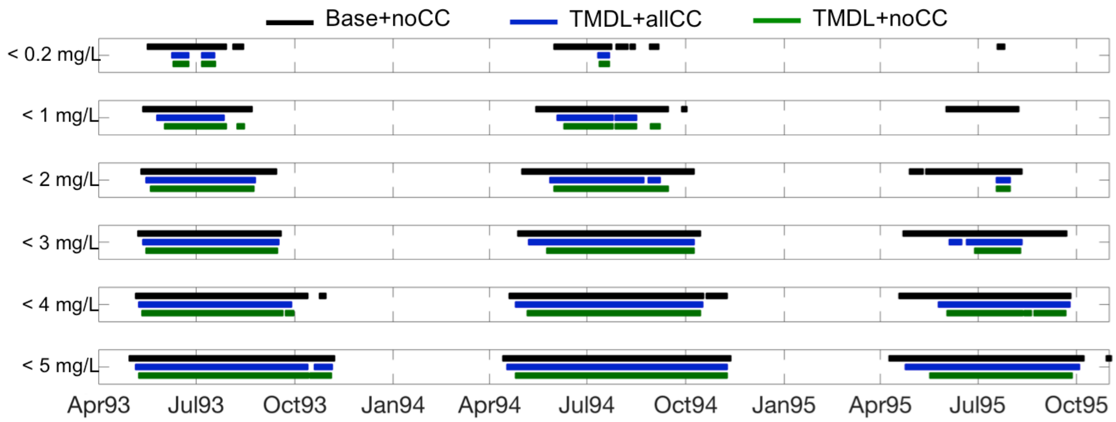


Figure 4.7 Bars showing duration of hypoxic volume (> 1km^3) at each DO level for the Base+noCC run and the TMDL+noCC and TMDL+allCC nutrient scenarios.

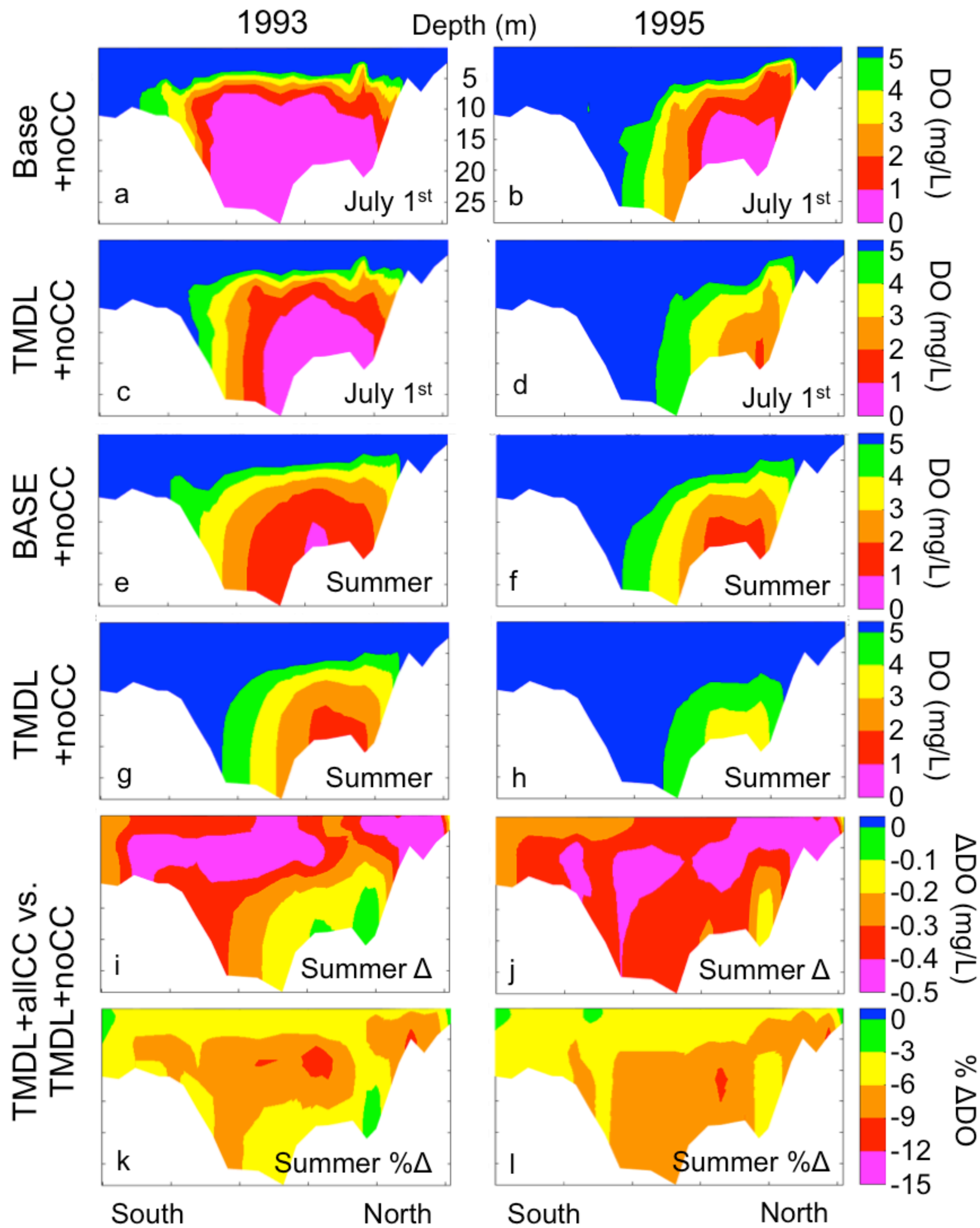


Figure 4.8 Along-Bay transects, with the Susquehanna River in the north and Bay mouth in the south, of DO of the Base+noCC run and TMDL+noCC scenario for July 1, 1993 (a,c) and July 1, 1995 (b,d), average summer (May-Sept) DO for 1993 (e,g) and 1995 (f,h), the difference in average summer DO between the TMDL+noCC and TMDL+allCC scenarios (i,j), and the percent difference in average summer DO between the TMDL+noCC and TMDL+allCC scenarios (k,l)..

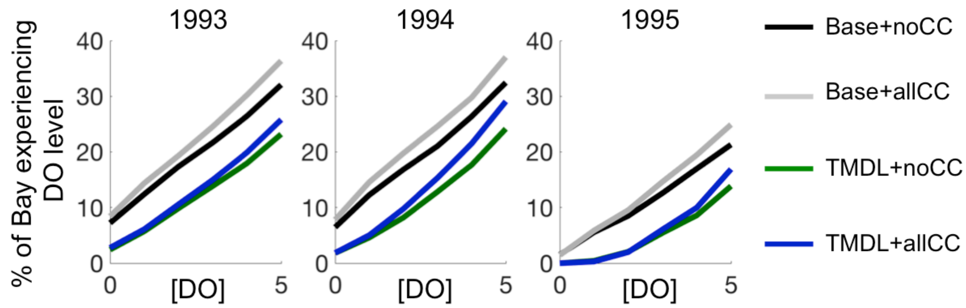


Figure 4.9 Percent of the entire Bay that experiences a given DO level during 1993 (a), 1994 (b), and 1995 (c).

Appendix 4.A Modified ChesROMS-ECB Equations

Modifications of biological functions from the model version published in Feng et al. (2015) are presented below. Temperature dependence was added to the zooplankton maximum growth rate, the remineralization rates of large and small detritus, and the phytoplankton growth rate at temperatures above 20°C. The maximum rate of nitrification, the temperature dependency on remineralization of semi-labile DON, and the remineralization rate of DOC at 0°C were also modified to fit with current understanding.

Symbol	Description	Feng et al. (2015)	Chapter 4	Units
g_{max}	*Zooplankton maximum growth rate	0.3	$0.05*e^{0.0742*T}$	d^{-1}
n_{max}	Maximum rate of nitrification	0.05	0.2	d^{-1}
r_{D_L}	*Remineralization of large detritus	0.2	$0.05*e^{0.0742*T}$	d^{-1}
r_{D_S}	*Remineralization of small detritus	0.2	$0.05*e^{0.0742*T}$	d^{-1}
$\kappa_{[DON]_{SL}}$	*Temperature dependency remineralization of semi-labile DON	0.07	0.0742	$(°C)^{-1}$
a_{0C}	Remineralization rate of DOC at 0 °C	0.003835	0.008	d^{-1}
μ_0	^Phytoplankton growth rate	2.15	<20°C, 2.15 T ≥20°C, $1.81 + e^{0.16*T-4.27}$	d^{-1}

*Community respiration and zooplankton grazing temperature dependent functions are based on a Q_{10} of 2.1 (Lomas et al., 2002)

^Phytoplankton growth rate at low temperatures (T < 20°C) is constant with higher temperatures following a rate based on Lomas et al. (2002) with a Q_{10} from 20°C to 40°C of 2.18.

Appendix 4.B Model skill assessment.

A skill assessment of the model is presented below. Skill was assessed via total RMSD (Table 4B.1), normalized target diagrams (Joliff et al., 2009), and time series analysis. Model results are compared to observational data from the Chesapeake Bay Program Water Quality Database (http://www.chesapeakebay.net/data/downloads/cbpwaterqualitydatabase1984present). For the total RMSD and target diagrams, the model results were compared to monthly/bi-monthly observations at the stations and regions shown in Figure 4.1. For the time series comparison, the model results were compared to the mean historical observations from (1985-2014). Model is also compared to the pervious iteration of the model evaluated in Irby et al., 2016 (Fig. 4B.1).

Table 4B.1 Total RMSD (and observational mean) of the present model and the model version used in Chapter 2 and Chapter 3.

	Model Version used in Chapters 2 and 3					Model Version used in Chapter 4				
Variable	Total	A	B	C	D	Total	A	B	C	D
Surface	1.23	1.69	0.91	1.17	1.15	1.23	1.73	0.92	1.19	1.17
Temp (^{o}C)	(17.15)	(16.86)	(16.71)	(17.25)	(17.62)	(17.15)	(16.86)	(16.71)	(17.25)	(17.62)
Bottom	2.27	1.68	2.92	2.28	1.88	2.22	1.72	2.84	2.22	1.86
Temp (^{o}C)	(15.98)	(16.74)	(15.42)	(15.89)	(16.09)	(15.98)	(16.74)	(15.42)	(15.89)	(16.09)
Surface	2.14	2.32	1.96	1.95	2.30	1.86	2.11	1.56	1.54	2.11
Salinity	(12.64)	(1.84)	(10.54)	(14.20)	(19.15)	(12.64)	(1.84)	(10.54)	(14.20)	(19.15)
Bottom	2.09	2.10	1.78	2.04	2.34	2.17	2.31	2.04	1.83	2.40
Salinity	(18.18)	(3.92)	(17.59)	(20.88)	(24.52)	(18.18)	(3.92)	(17.59)	(20.88)	(24.52)
Surface DO	1.60	1.89	1.61	1.56	1.43	1.48	1.73	1.61	1.28	1.34
(mg/L)	(9.35)	(9.24)	(9.65)	(9.39)	(9.13)	(9.35)	(9.24)	(9.65)	(9.39)	(9.13)
Bottom DO	2.51	2.62	1.63	2.49	2.95	1.82	2.47	1.63	1.58	1.72
(mg/L)	(5.78)	(8.00)	(3.82)	(5.01)	(6.71)	(5.78)	(8.00)	(3.82)	(5.01)	(6.71)
Surface NO_3	0.21	0.36	0.18	0.17	0.13	0.20	0.35	0.18	0.13	0.11
(mmolN/m^3)	(0.32)	(0.93)	(0.34)	(0.17)	(0.07)	(0.32)	(0.93)	(0.34)	(0.17)	(0.07)
Bottom NO_3	0.17	0.34	0.28	0.07	0.07	0.23	0.37	0.28	0.14	0.06
(mmolN/m^3)	(0.21)	(0.81)	(0.14)	(0.08)	(0.04)	(0.21)	(0.81)	(0.14)	(0.08)	(0.04)

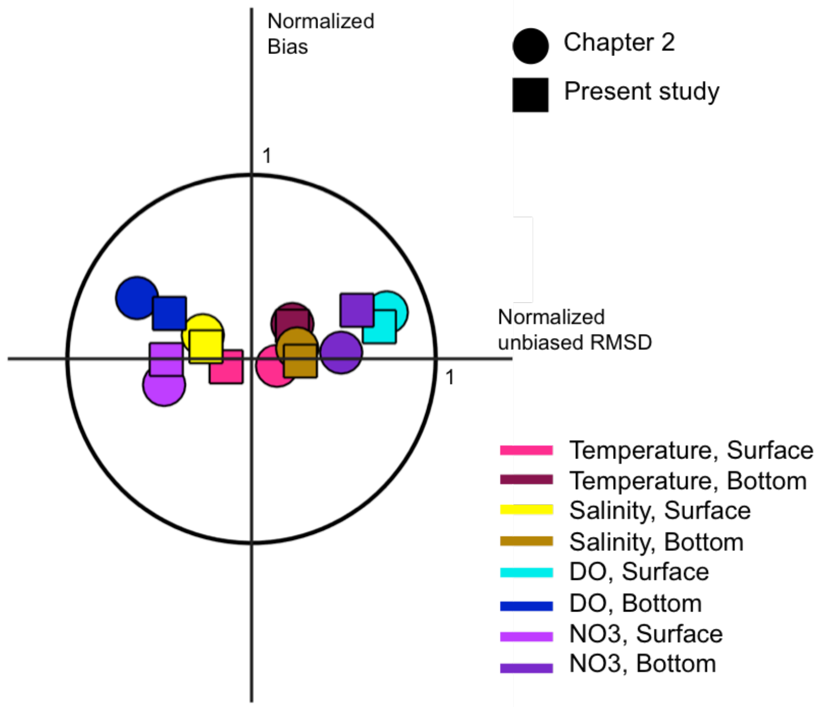


Figure 4B.1 Target diagram comparing the skill of the ChesROMS-ECB model version used in Chapters 2 and 3 to the one used in the present study. Statistics combine the spatial and temporal variability across 23 stations (Table 4.1) for 1993-1995.

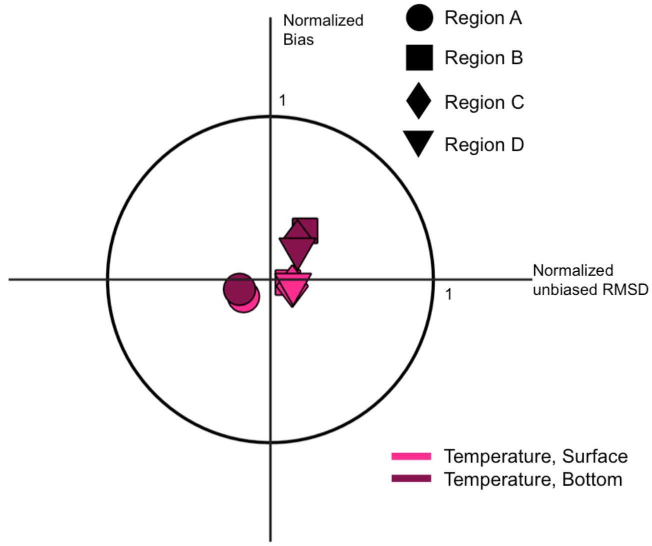


Figure 4B.2 As in Figure 4B.1, but by Region and for surface and bottom temperature.

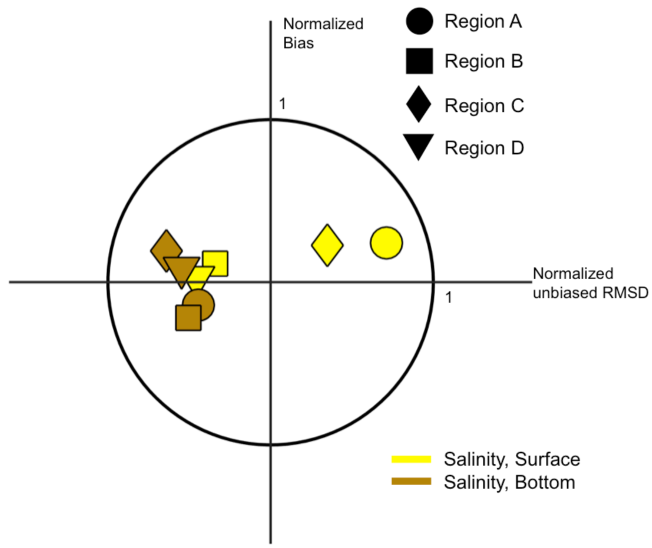


Figure 4B.3 As in Figure 4B.2, but for surface and bottom salinity.

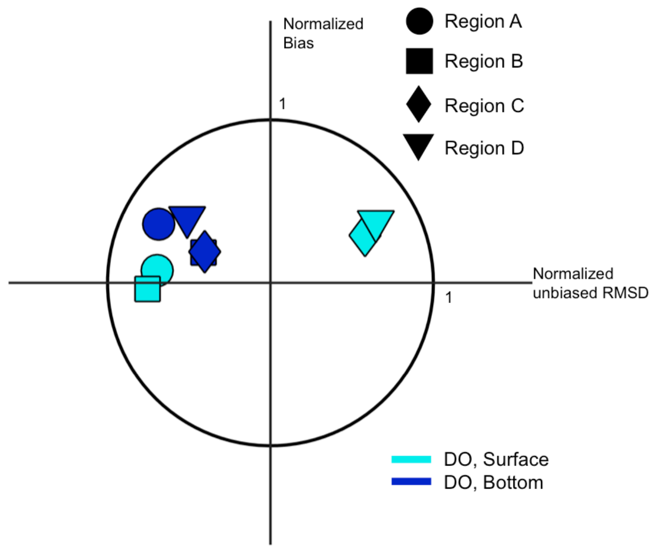


Figure 4B.4 As in Figure 4B.2, but for surface and bottom dissolved oxygen.

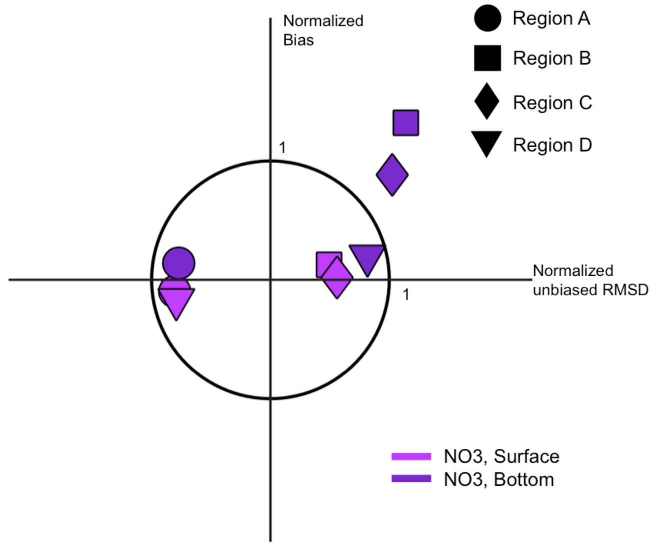


Figure 4B.5 As in Figure 4B.2, but for surface and bottom nitrate.

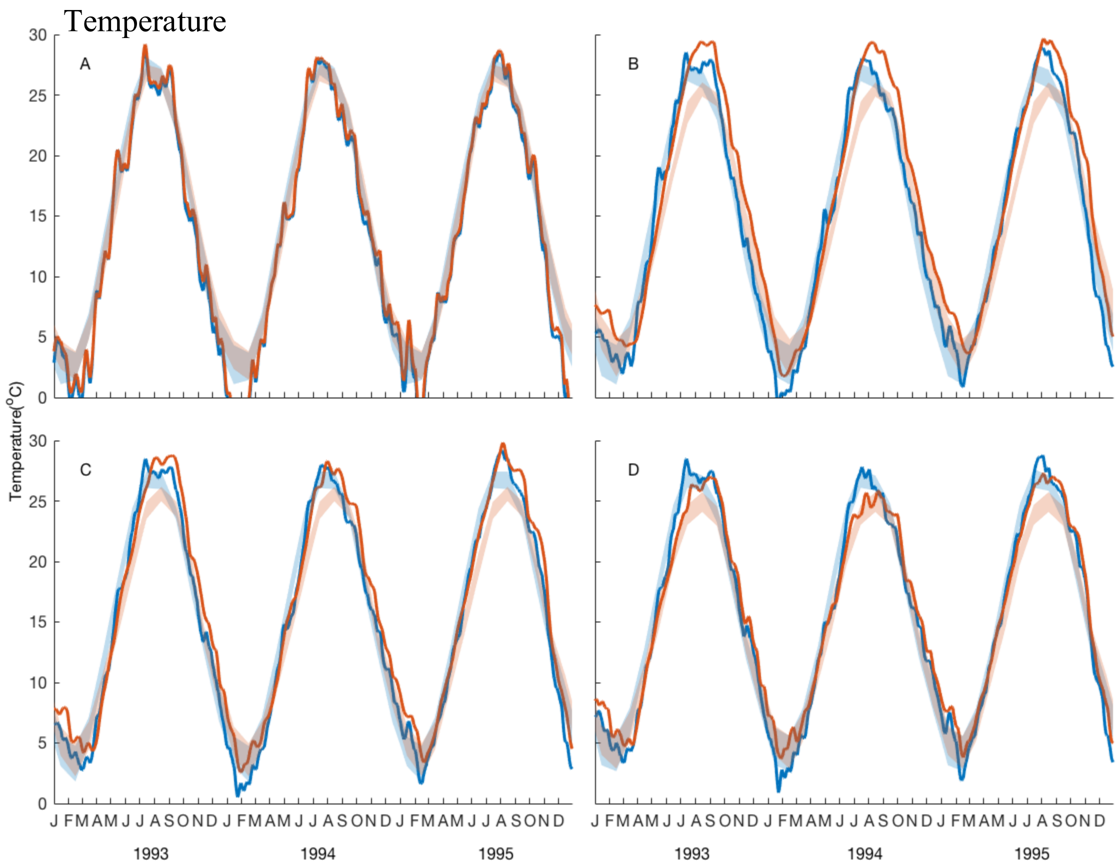


Figure 4B.6 Time series of a 7-day moving mean of temperature at the surface (blue) and bottom (red) with the associated 25th-75th percentiles of the climatological observations (1985-2014) for the surface (blue shading) and bottom (red shading) for the four regions (A, B, C, D) in Figure 1.

Figure 4B.7 As in Figure 4B.6, except for salinity.

Dissolved Oxygen

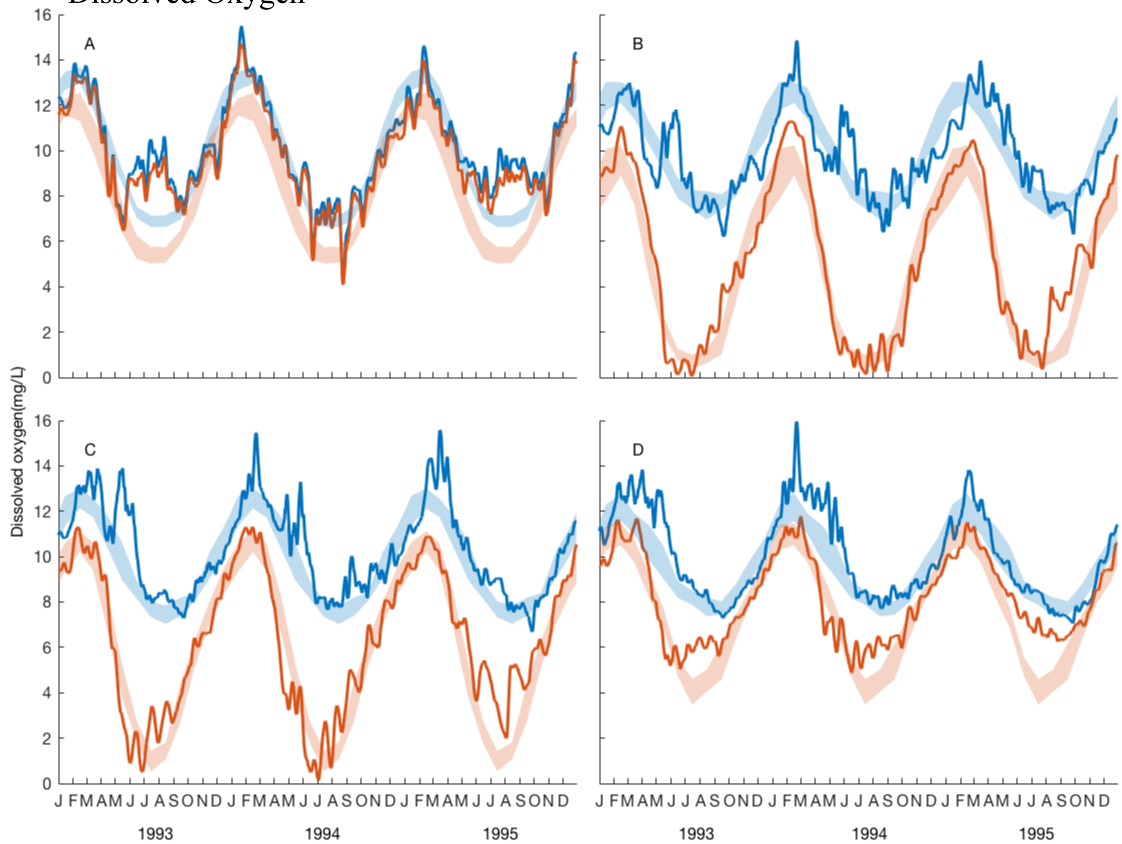


Figure 4B.8 As in Figure 4B.6, except for dissolved oxygen.

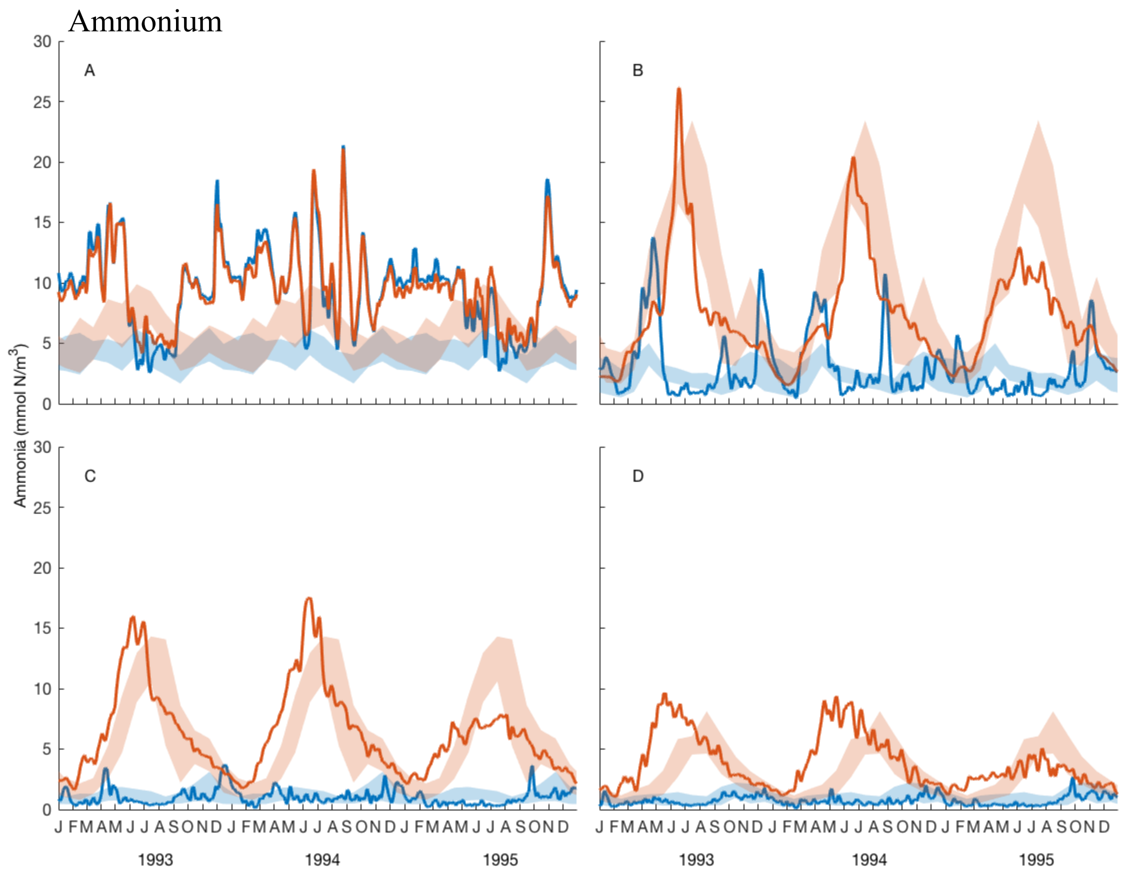


Figure 4B.9 As in Figure 4B.6, except for ammonium.

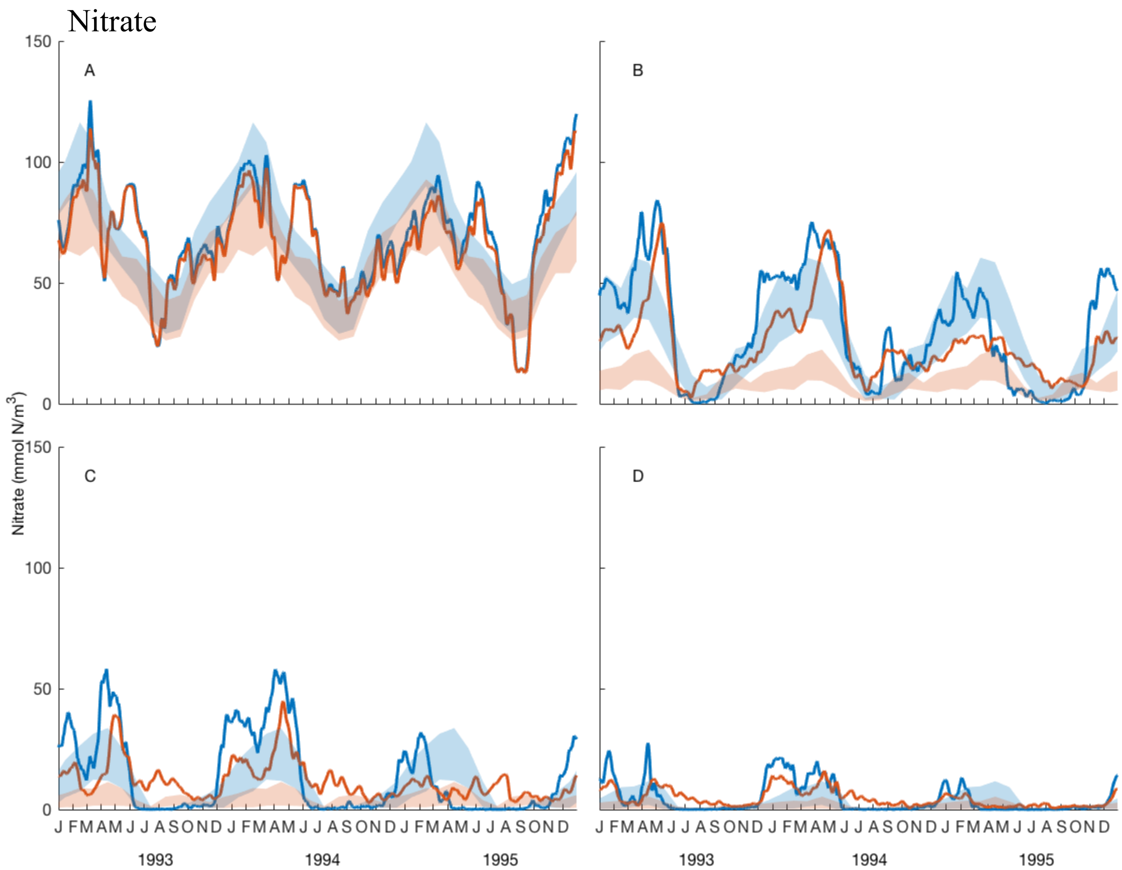


Figure 4B.10 As in Figure 4B.6, except for nitrate.

Acknowledgements

This paper is the result of research funded in part by NOAA's National Centers for Coastal Ocean Science under award NA16NOS4780207 to the Virginia Institute of Marine Science (VIMS) and by NOAA's U.S. Integrated Ocean Observing System Program Office as a subcontract to VIMS under award NA13NOS0120139 to the Southeastern University Research Association. Thank you to K. Hinson and F. Da for their assistance and to R. Najjar and C. Friedrichs for comments on an initial version of this manuscript. This work used the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by National Science Foundation grant number ACI-1548562.

Literature Cited

Altieri, A.H., and K.B. Gedan. 2015. Climate change and dead zones. *Global Change Biology*, 21, 1395-1406. doi: 10.1111/bcg.12754.

Bever, A.J., M.A.M. Friedrichs, C.T. Friedrichs, M.E. Scully, and L.W. Lanerolle. 2013. Combining observations and numerical model results to improve estimates of hypoxic volume within the Chesapeake Bay, USA. *Journal of Geophysical Research: Oceans*, 118: 4924-4944. doi: 10.1002/jgrc.20331.

Breitburg, D. 2002. Effects of Hypoxia, and the Balance between Hypoxia and Enrichment, on Coastal Fishes and Fisheries. *Estuaries*, 25, 767-781.

Boon, J.D, and M. Mitchell. 2015. Nonlinear Change in Sea Level Observed at North American Tide Stations. *Journal of Coastal Research*, 31: 1295-1305. doi: 10.2112/JCOASTRES-D-15-00041.1.

Buchheister, A., C.F. Bonzek, J. Gartland, and R. Latour. 2013. Patterns and drivers of the demersal fish community of Chesapeake Bay. *Marine Ecology Progress Series*, 481: 161-180. doi: 10.3354/meps10253.

Cerco, C., S.C. Kim, and M. Noel. 2010. The 2010 Chesapeake Bay Eutrophication Model – A Report to the US Environmental Protection Agency Chesapeake Bay Program and to The US Army Engineer Baltimore District. US Army Engineer Research and Development Center, Vicksburg, MS.

Chua, V.P., and M. Xu. 2014. Impacts of sea-level rise on estuarine circulation: An idealized estuary and San Francisco Bay. *Journal of Marine Systems*, 139: 58-67. doi: 10.1016/j.jmarsys.2014.05.012.

Church, J.A., P.U. Clark, A. Cazenave, J.M. Gregory, S. Jevrejeva, A. Levermann, M.A. Merrifeld, G.A. Milne, R.S. Nerem, P.D. Nunn, A.J. Payne,W.T. Pfeffer, D. Stammer, and A.S. Unnikrishnan, 2013: Sea Level Change. In: *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, Cambridge University Press.

Collins, M., Knutti, R., Arblaster, J., Dufresne, J-L., Fichefet, T., Friedlingstein, P., Gao, X., Gutowski, W. J., Johns, T., Krinner, G., Shongwe, M., Tebaldi, C., Weaver, A. J., and Wehner, M. 2013. Long-term Climate Change: Projections, Commitments and Irreversibility. In: *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, Cambridge University Press.

Diaz, R. and R. Rosenberg. 2008. Spreading Dead Zones and Consequences for Marine Ecosystems. *Science* 321.

Ding, H., and A.J. Elmore. 2015. Spatio-temporal patterns in water surface temperature from Landsat time series data in the Chesapeake Bay, U.S.A. *Remote Sensing of Environment*, 168: 335-348. doi: 10.1016/j.rse.2015.07.009.

Druon, J.N., A. Mannino, S. Sigorini, C. McClain, M. Friedrichs, J. Wilkin, and K. Fennel. 2010. Modeling the dynamics and export of dissolved organic matter in the Northeastern U.S. continental shelf. *Estuarine Coastal Shelf Science*, 88: 488-507. doi: 10.1016/j.ecss.2010.05.010.

Feng, Y., M.A.M. Friedrichs, J. Wilkin, H. Tian, Q. Yang, E.E. Hofmann, J.D. Wiggert, and R.R. Hood. 2015. Chesapeake Bay nitrogen fluxes derived from a land-estuarine ocean biogeochemical modeling system: model description, evaluation, and nitrogen budgets. *Journal of Geophysical Research: Biogeochemistry*, 120: 1666-1695. doi: 10.1002/2015JG002931.

Goberville, E., G. Beaugrand, N-C. Hautekeete, Y. Piquot, and C. Luczak. 2015. Uncertainties in the projection of species distributions related to general circulation models. *Ecology and Evolution*, 5: 1100-1116. doi: 10.1002/ece3.1411.

Hagy, J.D., W.R. Boyton, C.W. Keefe, and K.V. Wood. 2004. Hypoxia in Chesapeake Bay, 1950-2001: long-term change in relation to nutrient loading and river low. *Estuaries*, 27: 634-658.

Harding, L.W., J.E. Adolf, M.E. Mallonee, W.D. Miller, C.L. Gallegos, E.S. Perry, J.M. Johnson, K.G. Sellner, and H.W. Paerl. 2015. Climate effects on phytoplankton floral composition in Chesapeake Bay. *Estuarine, Coastal and Shelf Science*, 162: 53-68. doi: 10.1016/j.ecss.2014.12.030.

Hargreaves, G.H, and Samani, Z.A. 1982. Estimating potential evapotranspiration. *Journal of the Irrigation and Drainage Division*, 108: 225-230.

Hofmann, E.E., B. Cahill, K. Fennel, M.A.M. Friedrichs, K. Hyde, C. Lee, A. Mannino, R.G. Najjar, J.E. O'Reilly, J. Wilkin, and J. Xue. 2011. Modeling the Dynamics of Continental Shelf Carbon. *Annual Review of Marine Science*, 3: 93-122. doi: 10.1146/annurev-marine-120709-142740.

Hong, B., and Shen, J. 2012. Responses of estuarine salinity and transport processes to potential future sea-level rise in the Chesapeake Bay. *Estuarine, Coastal, and Shelf Science*, 104-105, 33-45, doi:10.1016/j.ecss.2012.03.014.

IPCC, 2013: Annex 1: Atlas of Global and Regional Climate Projections [van Oldenborgh, G.J., M. Collins, J. Arblaster, J.H. Christensen, J. Marotzke, S.B. Power, M. Rummukainen, and T. Zhou (eds.]. In: *Climate Change 2013: The Physical Science Basis. Controbutions of the Working Group 1 to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* [Stocker, T.F., D. Qin, G-K. Plattner, M. Tignor, S.K. Allen, J.Boschung, A. Nauels, Y. Xia, V. Bex, and P.M. Midgley (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

IPCC, 2013: Summary for Policymakers. In: *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* [Stocker, T.F., D. Qin, G-K. Plattner, M. Tignor, S.K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex, and P.M. Midgley (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

Jolliff, J.K., J.C. Kindle, I. Shulman, B. Penta, M.A.M. Friedrichs, R. Helber, and R.A. Arnone. 2009. Summary diagrams for coupled hydrodynamic-ecosystem model skill assessment. *Journal of Marine Systems*, 76: 64-82. doi: 10.1016/j.jmarsys.2008.05.014.

Kemp, W. M., Testa, J. M., Conley, D. J., Gilbert, D., and Hagy, J. D. 2009. Temporal responses of coastal hypoxia to nutrient loading and physical controls. *Biogeosciences*, 6, 2985-3008.

Lake, S.J. and M.J. Brush. 2015. Modeling estuarine response to load reductions in a warmer climate: York River Estuary, Virginia, USA. *Marine Ecology Progress Series*, 538: 81-98. doi: 10.3354/meps11448.

Lapointe, D., W.K. Vogelbein, M.C. Fabrizio, D.T. Gauthier, and R.W. Brill. 2014. Temperature, hypoxia, and mycobacteriosis: effects on adult striped bass *Morone saxatilis* metabolic performance. *Diseases of Aquatic Organisms*, 108: 113-127. doi: 10.3354/dao02693.

Lee, M., E. Shevliakova, S. Malyshev, P.C.D. Milly, and P.R. Jaffe. 2016. Climate variability and extremes, interacting with nitrogen storage, amplify eutrophication risk. *Geophysical Research Letters*, 43: 7520-7528. doi: 10.1002/2016GL069254.

Linker, L.C., G.W. Shenk, P. Wang, C.F. Cerco, A.J. Butt, P.J. Tango, and R.W. Savidge. 2002. A Comparison of Chesapeake Bay Estuary Model Calibration With 1985 – 1994 Observed Data and Method of Application to Water Quality Criteria. *A Report of the Modeling Subcommittee Chesapeake Bay Program Office, Annapolis, MD*.

Lomas, M.W., P.M. Gilbert, F.K. Shiah, and E.M. Smith. 2002. Microbial processes and temperature in Chesapeake Bay: current relationships and potential impacts of regional warming. *Global Change Biology*, 8: 51-70.

Meier, H.E.M., H.C. Anderson, K. Eilola, B.G. Gustafsson, I. Kuznetsov, B. Muller-Karulis, T. Neumann, and O.P. Savchuk. 2011. Hypoxia in future climates: A model ensemble study for the Baltic Sea. *Geophysical Research Letters*, 38. doi: 10.1029/2011GL049929.

Muhling, B.A., C.F. Gaitan, C.A. Stock, V.S. Saba, D. Tommasi, and K.W. Dixon. 2017. Potential salinity and temperature futures for the Chesapeake Bay using a statistical downscaling spatial disaggregation framework. *Estuaries and Coasts*. doi: 10.1007/s12237-017-0280-8.

Murphy, R.R., W.M. Kemp, and W.P. Ball. 2011. Long-term trends in Chesapeake Bay seasonal hypoxia, stratification, and nutrient loading. *Estuaries and Coasts*, 34: 1293-1309. doi: 10.1007/s12237-011-9413-7.

Najjar, R.G., L. Patterson, and S. Graham. 2009. Climate Simulations of Major Estuarine Watersheds in the Mid-Atlantic Region of the US. *Climate Change* 95.

Najjar, R.G., Pyke, C.R., Adams, M.B., Breitburg, D., Hershner, C., Kemp, M., Howarth, R., Mulholland, M.R., Paolisso, M., Secor, D., Sellner, K., Wardrop, D., and Wood,

R. 2010. Potential climate-change impacts on the Chesapeake Bay. *Estuarine, Coastal, and Shelf Science*, 86, 1-20, doi:10.1016/j.ecss.2009.09.026.

Portner, H.O., and R. Knust. 2007. Climate change affects marine fishes through the oxygen limitation of thermal tolerance. *Science*, 315: 95-97.

Portner, H.O., G. Lanning. 2009. Oxygen and capacity limited thermal tolerance. The hypoxic environment. *Fish physiology, Vol 27*, Academic Press, San Diego, CA, p 143-191.

Preston, B.L. 2004. Observed Winter Warming of the Chesapeake Bay Estuary (1949-2002): Implications for Ecosystem Management. *Environmental Assessment*, 34: 125-139. doi: 10.1007/s00267-004-0159-x.

Reclamantion. 2013. Downscaled CMIP3 and CMIP5 Climate and Hydrology Projections: Release of Downscaled CMIP5 Climate Projections, Comparison with preceding Information, and Summary of User Needs. Prepared by the U.S. Department of the Interior, Bureau of Reclamation, Technical Services Center, Denver, Colorado. 47pp.

Ross, A.C., R.G. Najjar, M. Li, M.E. Mann, S.E. Ford, and B. Katz. 2015. Sea-level rise and other influences on decadal-scale salinity variability in a coastal plain estuary. *Estuarine, Coastal and Shelf Science*, 157: 79-92. doi: 10.1016/j.ecss.2015.01.002.

Sallenger, A.H., K.S. Doran, and P.A. Howd. 2012. Hotspot of accelerated sea-level rise on the Atlantic coast of North America. *Nature Climate Change*. doi: 10.1038/NCLIMATE1597.

Scully, M.E. 2010. The importance of climate variability to wind-driven modulation of hypoxia in Chesapeake Bay. *Journal of Physical Oceanography*, 40: 1435-1440. doi: 10.1175/2010JPO4321.1.

Shchepetkin, A.F., and J.C. McWilliams. 2005. The Regional Ocean Modeling System (ROMS): a split-explicit, free-surface, topography-following-coordinate oceanic model. *Ocean Modeling*, 9: 347-404. doi: 10.1016/j.ocemod.2004.08.002.

Shenk, G.W., and L.C. Linker. 2013. Development and application of the 2010 Chesapeake Bay watershed total maximum daily load model. *Journal of the American Water Resources Association*, 49: 1-15. doi: 10.1111/lawr.12109.

Sinha, E., A.M. Michalak, and V. Balaji. 2017. Eutrophication will increase during the 21st century as a result of precipitation changes. *Science*, 357: 405-408. doi: 10.1126/science.aan2409.

Son, S.H., and M. Wang. 2015. Diffuse attenuation coefficient of the photosynthetically available radiation K_d(PAR) for global open ocean and coastal waters. *Remote Sensing of Environment*, 159: 250-258. doi: 10.1016/j.rse.2014.12.011.

Sweet, W.V., R.E. Kopp, C.P. Weaver, J. Obeysekera, R.M. Horton, E.R. Thieler, and C. Zervas. Global and Regional Sea Level Rise Scenarios for the United States. 2017. NOAA Technical Report NOS CO-OPS 083.

Tango, P.J., and R.A. Batiuk. 2013. Deriving Chesapeake Bay water quality standards. *Journal of the American Water Resources Association*, 1-18. doi: 10.1111/jawr.12108.

USEPA. 2010. Chesapeake Bay Total Maximum Daily Load for Nitrogen, Phosphorus, and Sediment. *U.S. Environmental Protection Agency Chesapeake Bay Program Office*, Annapolis, MD.

Vaquer-Sunyer, R., and C.M. Duarte. 2008. Thresholds of Hypoxia for Marine Biodiversity. *Proceedings of the National Academies of Science*, USA, 105: 15452-15457, doi:10.1073/pnas.0803833195.

Vaquer-Sunyer, R., and C.M. Duarte. 2011. Temperature effects on oxygen thresholds for hypoxia in marine benthic organisms. *Global Change Biology*. doi: 10.1111/j.1365-2486.2010.02343.x

Wang, M., S.H. Son, and L.W. Harding. 2009. Retrieval of diffuse attenuation coefficient in the Chesapeake Bay and turbid ocean regions for satellite ocean color applications. *Journal of Geophysical Research – Oceans*, 114. doi: 10.1029/2009JC005286.

Winder, M., and U. Sommer. 2012. Phytoplankton response to a changing climate. *Hydrobiologia*, 698: 5-16.

Xu, J., W. Long, J.D. Wiggert, L.W.J. Lanerolle, C.W. Brown, R. Murtugudde, and R.R. Hood. 2012. Climate forcing and salinity variability in Chesapeake Bay, USA. *Estuarine, Coastal and Shelf Science*, 35: 237-261. doi: 10.1007/s12237-011-9423-5.

Yang, Q., H. Tian, M.A.M. Friedrichs, C. Hopkinson, and C. Lu. 2015a. Increased nitrogen export from eastern North America to the Atlantic Ocean due to climatic and anthropogenic changes during 1901-2008. *Journal of Geophysical Research: Biogeosciences*, 120: 1046-1068. doi: 10.1002/2014JG002763.

Yang, Q., H. Tian, M.A.M. Friedrichs, M. Liu, L. Xia, and J. Yang. 2015. Hydrological responses to climate and land-use changes along the North American East Coast: a 110-year historical reconstruction. *Journal of the American Water Resources Association*, 51: 47-67. doi: 10.1111/jawr.12232.

Chapter 5:

Conclusions: a look at the relationship between multiple models and the Chesapeake Bay TMDL

5. CONCLUSIONS: A LOOK AT THE RELATIONSHIP BETWEEN MULTIPLE MODELS AND THE CHESAPEAKE BAY TMDL

5.1 Introduction

The Chesapeake Bay Total Maximum Daily Load (TMDL) is touted as the most extensive and complex TMDL in the nation (Batiuk et al., 2013). While projected costs run in the tens of billions of dollars (BRFP, 2004; Nelson, 2014), the associated positive impacts on the environment, industry, and health are projected to easily compensate for the costs (BRFP, 2004; CBF, 2014). The complexity of the endeavor combined with the high upfront, concentrated costs and the even higher long term, disperse benefits make it critical that the TMDL is positioned for success. A large part of the responsibility for that success falls on the science at the foundation of the regulation as well as the regional partnership responsible for the TMDL, the Chesapeake Bay Program (CBP).

As with many environmental regulations, the Chesapeake Bay TMDL has endured an intense and lengthy set of legal battles, primarily at the hand of the American Farm Bureau. In 2016, the Supreme Court denied the Farm Bureau's petition to the court, effectively affirming the decision of the lower courts that in part granted deference to the science conducted by the CBP in developing the TMDL (*American Farm Bureau Federation et al., v. United States Environmental Protection Agency*, 2013; *American Farm Bureau Federation et al., v. United States Environmental Protection Agency*, 2015). With the courts upholding the scientific basis of the regulation, the CBP continues to work towards ensuring the use of the best science available.

To address the need to ensure that the science behind the TMDL is the best available, in early 2012, the Environmental Protection Agency's (EPA) director of the CBP asked its Science and Technical Advisory Committee (STAC) to explore the possibility and efficacy of utilizing a multiple model approach in the TMDL. Out of this request grew two workshops on multiple models (Friedrichs et al, 2012; Weller et al., 2013), a shallow water multiple modeling pilot research project (USEPA, 2013), and much of the catalyst for this dissertation. In this context, a multiple model approach encompasses a variety of ways to utilize the information from more than one model or modeling system, such as multi-model ensembles, model inter-comparisons, and modular community modeling. The Chesapeake Bay research community is uniquely poised to take up such a charge with multiple academic institutions focused on Bay research, as well as the resources of scientists involved with the Chesapeake Community Modeling Program (http://ches.communitymodeling.org/).

As the CBP continues to take advantage of the large group of academic research scientists throughout the Chesapeake Bay region, it is important to not keep the academic scientists behind closed doors and to allow them to be fully integrated as stakeholders. A 2016 study by the Pew Research Center (PRC, 2016) found that scientists were one of the groups most trusted to act in the public interest, with 76% of respondents stating that they have "a great deal" or "a fair amount" of confidence in their actions. This is in stark contrast to elected officials at only 27% and business leaders at 41%. While the CBP scientists themselves are non-partisan civil servants, their position as government employees may affect the public's perception of their work. Furthermore, while 75% of Americans think that protecting the environment is a major role of government, only 59% think that the government is doing a good job at protecting the environment (PRC, 2015a). This makes it clear that public buy-in can be enhanced with the visible incorporation of widely trusted academic research scientists.

This process works both ways as only 27% of scientists polled by the American Association for the Advancement of Science (AAAS) think that the best science guides regulations on clean air and water (PRC, 2015). If the scientists do not trust the efficacy of a regulation purportedly based on science, how can the public be expected to buy in? While it is certainly true that the best scientific advice can often make for lousy regulatory policy without the incorporation of political acumen, the political spin does not have to denigrate the scientific basis. Incorporating scientists as stakeholders along the entire process can help ensure the scientific foundation is not lost as the policy is developed while also helping the scientific community understand the role their science plays in regulatory policy.

In the end, the Chesapeake Bay TMDL is a good case study in the value of utilizing an academic research community to enhance the scientific basis and public understanding of a complex, costly, and contentious environmental regulation. One way to do this is through the incorporation of a multiple modeling system that leads to increased scientific confidence, stakeholder engagement, and regulation effectiveness.

Fortunately, the CBP is readily looking for ways to enhance the TMDL through multiple models and the rest of this chapter examines the basic questions behind such a task.

5.2 Why is creating a TMDL for Chesapeake Bay so difficult?

Even prior to the incorporation of multiple models, formulating a regulation to clean up the Chesapeake Bay is immensely complex. The issues arise primarily due to three aspects of the problem: location, impacts, and time. The Bay's watershed is 14 times larger than the Bay itself and extends from New York to Virginia encompassing a population of more than 18 million people and land uses from major metropolitan centers like Washington, D.C., to the forested mountains of the Shenandoah National Park. Impacts from activity in the watershed have the potential to make their way to the Bay via the vast network of upland streams and waterways that eventually connect to the Bay's large tributary rivers. Additionally, these human impacts have been occurring in the watershed for hundreds of years and intensified with colonization and then industrialization (Cooper and Brush, 1993). These three aspects combine to create an issue that is vast in size, complicated in origin, and compounded over time. As a result, it is not hard to understand why solving the problem with a single regulation and in a relatively short time frame is so challenging.

The concept of regulating pollution that is coming from a large watershed encompassing six states but that is delivered to a Bay that only two of those states border, forces a whole host of government, industry, and resident stakeholders into the equation, each with a different sense of responsibility and motivation for action (or inaction). Further, not all pollutants are equal in terms of both source and type, and distance from the Bay itself also plays a role. For source, excess nitrogen in a forested area will impact downstream waters differently than the same amount of nitrogen applied in an urban or suburban setting. For type, phytoplankton preferentially utilize ammonium as a nutrient source relative to nitrate, but both are considered a pollutant in the TMDL. Therefore, with these two impacts combined, ammonium applied to a farm in upstate New York will impact the Bay much differently than nitrate applied to a golf course in southern Virginia.

A further logistical and scientific complication is that the goal of the TMDL is not singular. While the primary end goal is to increase dissolved oxygen concentrations (Keisman and Shenk, 2013), the TMDL also contains minimum water quality standards aimed at limiting chlorophyll *a* concentrations and increasing water clarity. All three of these water quality metrics are ecologically linked, but they manifest themselves in different locations of the Bay, at different times of the year, and are controlled by different mechanisms.

As with any regulation, there are costs incurred in compliance. Part of the issue with the Chesapeake Bay TMDL is that the costs are so high and fall on particular industries such as agriculture and local government entities like municipal wastewater facilities. While it is projected that the high upfront costs will be more than offset by the long-term benefits (BRFP, 2004; CBF, 2014), resources and methods to alleviate the costs are necessary (CBC, 2012; Nelson, 2014). In an attempt to increase opportunities to share the costs of TMDL implementation, the regulation allowed for nutrient trading markets to be adopted by the individual states. While the current markets are likely too restrictive to be truly economically viable options, it is possible for the states to open up the market-based system in the future (Nelson, 2014).

Further complicating the TMDL is the evaluation of progress made towards the goal of increasing water quality. It is critically important for the CBP to define and track progress, and they do that in a variety of different ways via their new Chesapeake Progress initiative (http://www.chesapeakeprogress.com/). Chesapeake Progress is technically focused on progress towards meeting the goals listed in the 2014 Chesapeake Bay Watershed Agreement, an agreement signed by the watershed states that commits to broader goals than just the reduction of nutrients and sediment in the TMDL. These goals include metrics, such as the size of the blue crab stock, that will be aided by successful attainment of the water quality standards in the TMDL, but they also go further and include issues such as environmental literacy. The transparency and attempt at communicating tangible effects that is offered by Chesapeake Progress is good for the TMDL, but many of the metrics that are used to show progress are fundamentally affected by water quality.

Breaking it down as simply as possible, the fundamental goal of the TMDL is to reduce nutrient inputs in an effort to improve water quality. Therefore, measuring progress in water quality itself is critically important for the TMDL. On the surface, measuring both sides of the equation should be relatively simple since we have already established that the amount of nutrients entering the Bay is too high and water quality is too low, while also establishing the requisite amount of nutrients that would lead to an acceptable level of water quality. Unfortunately, simply because the before and after have been identified (albeit with some uncertainty), it does not mean that the incremental steps from beginning to end are easy to resolve. There are four fundamental reasons for this: interannual variability, lag-time, changing futures, and adaptive management.

In terms of simply measuring progress each year of either nutrients delivered to the Bay or the water quality of the Bay itself, the issue of interannual variability is difficult to overcome. Each year, the University of Maryland Center for Environmental Science publishes a report card for overall ecosystem health of the Bay (https://ecoreportcard.org/report-cards/chesapeake-bay/). The report card gives grades to different portions of the Bay while also offering an overall grade for the health of the Bay for that individual year. One would hope that the actions taken to reduce nutrient inputs to the Bay since the TMDL was implemented in 2010 would be manifested in that report card with an overall trend of better grades each year. However, environmental and climate differences between years can easily mask a trend of progress. For example, Chapters 3 and 4 established that years with wet winters experience worse water quality conditions in the subsequent summer than dry winters. So if 2010 was relatively dry while 2016 was relatively wet, one could reasonably expect that the grade in 2010 would be higher than in 2016 even if multiple management practices to reduce nutrient runoff had been implemented in the interim years.

Compounded with interannual variability is the issue of lag-time in response to nutrient reduction. The idea that the Bay may not immediately respond to actions taken in the watershed is based on the fact that at any given time, water quality in the Bay is a consequence of contemporary human impacts, but also the history of past human impacts throughout the watershed (STAC, 2013). For example, Staver and Brinsfield (1998) found that after the incorporation of a cover crop in Maryland, shallow groundwater nitrate concentrations dropped after a 2-3 year lag but did not reach maximum reduction until ~10 years after application. As mentioned earlier, human impacts have greatly degraded the health of the Bay since industrialization (Cooper and Brush, 1993). As a result, some of those impacts have stored up in the sediment of the watershed and slowly find their way into the Bay via mechanisms such as groundwater, land-use changes, erosion, and major storm events.

Changing futures also involves a time component, but instead of issues of the past impacting the present, changing futures involve the present catching up with the future. This primarily plays out in the context of climate change. The TMDL was developed for implementation in 2010 based on conditions from 1993 to 1995. As a result, the TMDL states that if the mandated nutrient reductions take place, then the water quality of the Bay will eventually meet minimum standards assuming that the environmental conditions in the future are similar to those experienced in the 1990s. Climate change, as was seen in Chapter 4, invalidates this assumption. Land-use change also falls into the changing futures category. It is not difficult to assume that urbanization and the resulting pattern of land-use in the watershed in 2050 will be different from what it was in 1993. Adequately accounting for how the climate and the use of land have and will change in the future is necessary for identifying contemporary progress towards meeting water quality standards.

Lastly, the incorporation of adaptive management further complicates identifying progress. Adaptive management can take many forms, but the CBP concisely defines their approach to adaptive management as "learning by doing" (http://www.chesapeakebay.net/about/how/management). The primary purpose behind adaptive management is that science is not definitive but that should not stop policy makers from making decisions based on current, sound scientific evidence. Rather, policies should be formulated and implemented that are adaptable so that as scientific understanding evolves, the regulation can evolve with it. While the CBP has committed to working towards an adaptive management framework, the structure of a regulation like the TMDL makes this difficult because in a truly adaptive management system each iteration of new knowledge has the potential to change the nutrient reduction necessary to achieve the water quality goal. This makes building stakeholder confidence difficult since there is the potential that nutrient limits may change.

The fact that adaptive management is challenging in a regulation like the TMDL should not dissuade the CBP and partner states and agencies from moving towards an adaptive management framework. Section 10 of the TMDL (USEPA, 2010) is titled "Implementation and Adaptive Management," a true indication of the CBP's commitment to moving towards adaptive management. But it is telling that the words "adaptive management" only occur once in the eight pages of the section. In that sentence, the CBP states a commitment to taking an adaptive management approach by incorporating new scientific understanding in the 2017 Mid-Point Assessment. The cooperation of the CBP with the research presented in this dissertation, along with the changes in the updated CBP watershed model, demonstrate that the CBP has followed through on their commitment. However, the CBP can and should continue to incorporate more adaptive management strategies, such as utilizing a framework, like multiple models, to constrain uncertainty in an effort to refine a TMDL forecast over time (see NRC, 2011).

Incorporated in all of the complexity involved in a regulation like the Chesapeake Bay TMDL is uncertainty. Uncertainty exists at every level of the regulatory process, from how much pollution must be reduced, to what individual actions actually work best at reducing nutrient runoff, to how long it will take to see results, to what the climate impacts will be in the future. The complexity of a regulation like the TMDL makes it even more important to understand the uncertainty surrounding the science fundamental to the regulation's development. It was in that light that the CBP requested its STAC to explore the viability of incorporating multiple models.

5.3 Why should we use multiple models in a TMDL?

The two workshops that STAC hosted on multiple models elucidated some fundamental advantages to utilizing multiple models in the TMDL process in a variety of ways. The reports that followed these workshops are filled with words such as uncertainty, accountability, and confidence (Friedrichs et al., 2012; Weller et al., 2013).

While the application of multiple models can address those (very important) aspects of the science, the social aspects of incorporating a multiple model framework should not be overlooked. In that vein, multiple models should be utilized in the TMDL for three basic reasons: to help assess uncertainty, to incorporate the best science and methodology, and to assist in moving the regulation forward.

The incorporation of multiple models allows for an assessment of uncertainty by framing a boundary of possibilities. This is similar to how the spaghetti plots form the cone of uncertainty in hurricane forecasts. Observations provide information as to where the hurricane is at present, and multiple models from multiple institutions that incorporate a variety of underlying assumptions project where the hurricane will be in the future (NWS, 2009). One day out, the models all generally agree in the trajectory of the hurricane, but further into the future, the models begin to disagree due to the differences in the models such as their configuration, tuning, assumptions, and forcing. Therefore, while a hurricane may be 5 days out from landfall, and the scientists do not know exactly where landfall will occur, they can have a high confidence that the hurricane is headed towards New Jersey rather than Florida. In that example, it is easy to see how important decisions, such as a warning for New Jersey but not Florida, can be made without perfect information. While projecting future water quality may be more complicated than predicting a hurricane's path, the fundamental advantage of multiple lines of evidence that allow an assessment of uncertainty still holds.

One important aspect of the hurricane example is that the models originated from different institutions (be they government or academic). This is important because it generally allows for a greater diversity of models than if they had all come from a single institution. Model diversity enhances the projection of uncertainty since no model is individually correct, but incorporating a more diverse group can increase the odds that future reality falls within the cone of uncertainty (Janssen et al., 2015). The inclusion of academic research labs and institutions also helps ensure that the best modeling science and methodology are being used, since academic research partially exists to continually redefine the cutting edge.

Including the academic scientific community can also build necessary stakeholder confidence while also taking advantage of the difference in approach between academic and regulatory scientists. As previously mentioned, scientists are broadly trusted to act in the public interest. Adding to this is the recent move in academia (and to some degree, government) to open source science. These open source methods not only increase transparency for stakeholders, they also allow for the sharing of scientific advances between research groups much more rapidly than in the past. The academic science community also approaches questions differently from government scientists engaged in developing regulations. This is not a knock on those regulatory scientists. Their job is to identify a scientific basis for a regulation and they work hard at doing that properly. But academic research scientists approach problems with flexibility that allows them to ask questions about their understanding of the environment and a model's representation of that environment. Furthermore, academic research models themselves are developed and utilized in a different way from a regulatory model. Many regulatory models, like the one developed for the Chesapeake Bay TMDL (Cerco et al., 2010), are essentially designed to answer a particular question. On the other hand, academic research models, particularly those developed using open source software, are designed to be flexible so that they can be used to ask multiple questions.

The incorporation of more scientists and models into the process also increases the diversity of expertise and experiences that can help ensure that the best science is being utilized. This can help to move the science behind the regulation forward and towards a truly adaptive management framework even if it must exist within the rigid standards of a TMDL. Part of the feedback problem inherent in applying adaptive management to a TMDL is that if new information necessitates that load restrictions are altered for a given jurisdiction then the entire load allocation for the whole watershed must be recalculated and redistributed. This may result in a given location's required load reduction to fluctuate with each iteration of the adaptive management cycle. Multiple models can be used to hedge against major fluctuations in each iteration by establishing an understanding of uncertainty to begin with, laying out the potential bounds of possibility. Multiple models can also be used to assess confidence in particular locations of the Bay, directing targeted research efforts at those locations exhibiting the highest uncertainty.

Overall, a multiple modeling framework applied to the TMDL would take into account multiple lines of evidence to increase scientific understanding of the fundamental processes underlying the TMDL goals. This would allow for an assessment of uncertainty in both model results and the eventual attainment of water quality standards while also promoting an open and transparent scientific and regulatory process. The regulation would also benefit from incorporating the diverse expertise of Chesapeake Bay research community that is already focused on understanding the natural and anthropogenic impacts on Chesapeake Bay's past, present, and future.

5.4 How are multiple models in the Chesapeake Bay currently being used?

The shallow water multiple model pilot project that grew out of one of the STAC workshops (Friedrichs et al., 2012) is still underway (http://www.chesapeakebay.net/groups/group/evaluation_of_multiple_shallow_water_systems_analysis). The project has brought researchers from multiple institutions together to utilize multiple models developed across academia and government to better understand how to overcome the obstacles water quality models face in the shallow and ecologically critical reaches of the Bay. While this project was derived directly from the CBP's request to study the possibility of using multiple models in the TMDL, it is not the only project to do so. Other federal agency projects, such as the National Oceanic and Atmospheric Administration's Integrated Ocean Observing System's Coastal Ocean Modeling Testbed (NOAA-IOOS-COMT), have also started projects to understand how to promote the use of multiple models in our understanding of coastal ecosystems. Partially as a result of the robust Chesapeake Bay modeling community, these types of projects and others have utilized and plan to utilize multiple models in the watershed, the Bay, and a mix of both, to further our understanding of the ecosystem, to explore the potential impacts of the TMDL, and to create a framework from which the CBP could move forward in implementing in the regulatory process.

In both the watershed (e.g., Boomer et al., 2013) and the Bay (e.g., Bever et al., 2013; Chapter 2), multiple models studies are being used to evaluate the skill of models relative to each other by comparing model output to the extensive historical observations available throughout the Chesapeake Bay region. Primary outcomes of these studies include the identification of strengths and weakness among the models overall and the research results point towards future research needs while also attempting to quantify uncertainty and establish a framework for the future use of multiple models. Furthermore, these studies include as one of the models in the study the relevant CBP model used in the regulatory process to create the TMDL. This allows for a direct comparison of the regulatory model to other models and can act as a gut check to identify whether or not the models perform similarly. Studies of both the watershed (e.g., Sharifi et al., 2016) and the Bay (e.g., Chapter 3) have also utilized multiple models, including the CBP model, to run scenario experiments to identify if models project a similar response in variables as a result of different scenario forcing.

The research presented in Chapter 3 went further than a comparison of scenarios and utilized the multiple model approach to assess confidence in the impact of TMDL nutrient reduction on water quality standard attainment. Via a Confidence Index, each segment of the Bay was given a confidence score between 0 and 1. This isolated the portions of the Bay where the models are in least agreement, and are therefore locations where the impact of nutrient reduction is least certain. Because the Confidence Index is composed of multiple indices, the reason for a low score could be isolated. These results pointed to oddities in the regression methodology used in TMDL development. The research was presented to the CBP and the information is currently being incorporated into the 2017 TMDL Mid-Point Assessment.

The newest version of the CBP watershed model (referred to as Phase 6) that will be used in the Mid-Point Assessment takes a necessary step toward integrating multiple models in the TMDL process. The new model is designed to be more transparent and easier to understand and incorporates multiple models at various steps within the model structure. For example, Phase 6 utilizes the average of three watershed models to identify the average nutrient loads delivered to the Bay from different land-use types. The three models are diverse in terms of how they work and where and why they were developed. Further, the use of a multi-model mean has been found to be of higher predictive skill than any individual model used in the mean (Gneiting and Raftery, 2005; Chapter 2).

5.5 How might multiple models be used for the Chesapeake Bay TMDL in the future?

While the CBP has continually emphasized their support of efforts to incorporate multiple models into the scientific foundation of the TMDL and they have begun the process of incorporating multiple models into the watershed portion of the regulatory development, there is much more to be done. While the 2017 Mid-Point Assessment will include aspects of multiple modeling in the watershed, the modeling of the Bay will continue to be based on a single model. As this discussion has demonstrated, the research community will continue to move forward along with the CBP towards the adoption of more multiple model frameworks used in TMDL decision making.

One such endeavor that has recently gotten off the ground aims to fully incorporate two watershed models (Phase 6 from the CBP and one from academia) with two Bay models (the one used by the CBP and one from academia used in Chapter 4) to examine the potential impacts of climate change on the TMDL nutrient reductions. This multiple model framework will enable the uncertainty that stems from both the nutrient reduction and future climate change to be examined and constrained. The incorporation of more models into the Confidence Index (Chapter 3) will further allow the CBP and research community to understand how the modeling results compare to each other beyond the raw model output and to the assessment of water quality standard attainment. Utilizing researchers from multiple academic institutions alongside the CBP, this project also has the potential to promote transparency and take advantage of the diversity of expertise.

Future multiple modeling efforts could expand on the chain of multiple models to incorporate economic models that can help identify options for offsets that can reduce the overall implementation costs without hampering TMDL progress. With costs in mind, the funding of these research projects is imperative but the costs of the modeling efforts are vastly outweighed by the costs of the full TMDL implementation and trumped even further by the total cost of making poor management decisions. Incorporating more adaptive management would further hedge against potentially ineffective management decisions.

5.6 Conclusions

The water quality regulatory process in the United States is established in the Clean Water Act Section 303(d), which mandates the development of a TMDL for impaired waterways. Since the TMDL is the regulatory infrastructure that the EPA has to work under to repair the nation's waterways, it is important to evaluate how multiple models can be used to enhance the scientific foundation of a TMDL. In that vein, the CBP has solicited the advice of the scientific community and has committed to move forward with incorporating multiple model approaches into the Chesapeake Bay TMDL development process.

Incorporating multiple models in the TMDL can establish a degree of confidence in the regulation on multiple fronts. The science itself can quantitatively identify a range of uncertainty that can be used in adaptive management of the ecosystem. The use of multiple models and the incorporation of academic scientists and their open source research can enhance the confidence of the scientific community that the environmental regulations are based on sound science. The use of academic scientists and transparent methodologies can also enhance the confidence of the public in the efficacy of an environmental regulation. The increased confidence and buy-in from all fronts can help keep the TMDL on track to achieve the water quality goals.

The success of the Chesapeake Bay TMDL is imperative for the future of water quality regulations across the United States. As the largest and most complex TMDL to date, it is critical that the science at the foundation of the regulation is sound. Multiple models must be at the heart of understanding how to ensure the future success of regulatory efforts to save the Bay.

Literature Cited

American Farm Bureau Federation et al., v. United States Environmental Protection Agency, 2013. No. 1:11-CV-0067. United States District Court for the Middle District of Pennsylvania.

American Farm Bureau Federation et al., v. United States Environmental Protection Agency, 2015. No. 13-4079. United States Court of Appeals for the Third Circuit.

Batiuk, R.A., L.C. Linker, and C.F. Cerco. 2013. Featured collection introduction: Chesapeake Bay Total Maximum Daily Load development and application. *Journal of the American Water Resources Association*, 45:5. doi: 10.1111/jawr.12114.

Bever, A.J., M.A.M. Friedrichs, C.T. Friedrichs, M.E. Scully, and L.W. Lanerolle. 2013. Combining observations and numerical model results to improve estimates of hypoxic volume within the Chesapeake Bay, USA. *Journal of Geophysical Research: Oceans*, 118: 4924-4944. doi: 10.1002/jgrc.20331.

Boomer, K.M.B, D.E. Weller, T.E. Jordan, L. Linker, Z.-J. Liu, J. Reilly, G. Shenk, and A.A. Voinov. 2012. Using Multiple Watershed Models to Predict Water, Nitrogen, and Phosphorus Discharges to the Patuxent Estuary. *Journal of the American Water Resources Association*, 1-25. doi: 10.1111/j.1752-1688.2012.00689.x.

BRFP: Blue Ribbon Finance Panel. 2004. *Saving a National Treasure: Financing the Cleanup of the Chesapeake Bay*.

CBC: Chesapeake Bay Commission. 2012. Nutrient Credit Trading for the Chesapeake Bay: An Economic Study. http://www.chesbay.us/Publications/nutrient-trading-2012.pdf.

CBF: Chesapeake Bay Foundation. 2014. The Economic Benefits of Cleaning Up the Chesapeake. www.cbf.org/economicbenefits.

Cerco, C., S.C. Kim, and M. Noel. 2010. The 2010 Chesapeake Bay Eutrophication Model – A Report to the US Environmental Protection Agency Chesapeake Bay Program and to The US Army Engineer Baltimore District. US Army Engineer Research and Development Center, Vicksburg, MS.

Cooper, S.R., and G.S. Brush. 1993. A 2,500-year history of anoxia and eutrophication in the Chesapeake Bay. *Estuaries*, 16: 617-626.

Friedrichs, M.A.M., K.G. Sellner, and M.A. Johnston. 2012. Using Multiple Models for Management in the Chesapeake Bay: A Shallow Water Pilot Project, Chesapeake Bay Program Scientific and Technical Advisory Committee Report, No. 12-003.

Gneiting, T., and A.E. Raftery. 2005. Weather Forecasting with Ensemble Methods. *Science*, 310. doi: 10.1126/science.1115255.

Janssen, A.B.G., G.B. Arhonditsis, A. Beusen, K. Bolding, L. Bruce, J. Bruggeman, R-M. Couture, A.S. Downing, J.A. Elliott, M.A. Frassl, G. Gal, D.J. Gerla, M.R. Hipsey, F. Hu, S.C. Ives, J.J. Janse, E. Jeppsen, K.D. Johnk, D. Kneis, X. Kong, J.J. Kuiper, M.K. Lehmann, C. Lemmen, D. Ozkundakci, T. Petzoldt, K. Rinke, B.J. Robson, R. Sachse, S.A. Schep, M. Schmid, H. Scholten, S. Teurlincx, D. Trolle, T.A. Troost, A.A.Van Dam, L.P.A. Van Gerven, M. Weijerman, S.A. Wells, and W.M. Mooij. 2015. Exploring, exploiting and evolving diversity of aquatic ecosystem models: a community perspective. *Aquatic Ecology*, doi:10.1007/s10452-015-9544-1.

Keisman, J., and G. Shenk. 2013. Total maximum daily load criteria assessment using monitoring and modeling data. *Journal of the American Water Resources* Association, 49, 1134-1149. doi:10.1111/jawr.12111.

Nelson, R.H. 2014. How to Save the Chesapeake Bay TMDL: The Critical Role of Nutrient Offsets. *William & Mary Environmental Law and Policy Review*, 38.

NRC: National Research Council. 2011. Achieving nutrient and sediment reduction goals in the Chesapeake Bay: An evaluation of program strategies and implementation. Committee on the Evaluation of Chesapeake Bay Program Implementation for Nutrient Reduction to Improve Water Quality, Water Science and Technology Board, National Research Council. The National Academies Press, Washington, DC.

NWS: National Weather Service. 2009. Technical Summary of the National Hurricane Center Track and Intensity Models. http://www.nhc.noaa.gov/pdf/model_summary_20090724.pdf

PRC: Pew Research Center. 2015a. An Elaboration of AAAS Scientists' Views.

PRC: Pew Research Center. 2015b. Beyond Distrust: How Americans View Their Government.

PRC: Pew Research Center. 2016. The Politics of Climate.

Sharifi, A., H. Yen, K.M.B. Boomer, L. Kalin, X. Li, and D.E. Weller. 2016. Using multiple watershed models to assess the water quality impacts of alternate land development scenarios for a small community. *Catena*, 150: 87-99. doi: 10.1016/j.catena.2016.11.009.

STAC: Chesapeake Bay Program Scientific and Technical Advisory Committee. 2013. Incorporating Lag-Times into the Chesapeake Bay Program. *STAC Publication #13-004*.

Staver, K.W., and R.B. Brinsfield. 1998. Using cereal grain winter cover crops to reduce groundwater nitrate contamination in the mid-Atlantic coastal plain. *Journal of Soil and Water Conservation*, 1998: 53 (230-240).

USEPA (U.S. Environmental Protection Agency). 2010. Chesapeake Bay Total Maximum Daily Load for Nitrogen, Phosphorus, and Sediment. U.S. Environmental Protection Agency Chesapeake Bay Program Office, Annapolis, MD.

USEPA (U.S. Environmental Protection Agency). 2013. Chesapeake Bay Program Office fiscal year 2014 request for proposals for evaluation of multiple shallow-water systems analysis to improve the assessment of Chesapeake Bay water clarity and Submerged Aquatic Vegetation water quality standards. RFP number EPA-R3-CBP-14-02. Catalog of Federal Domestic Assistance (CFDA) number: 66.466.

Weller, D., B. Benham, M.A.M. Friedrichs, N. Gardner, R. Hood, R. Najjar, M. Paolisso, P. Pasquale, K. Sellner, and G. Shenk. 2013. Multiple Models for Management in the Chesapeake Bay, Chesapeake Bay Program Scientific and Technical Advisory Committee Workshop Report, No. 14-004.

Vita

Isaac (Ike) Irby was born in California on December 25th 1986. Before starting kindergarten, he moved to Colorado with his family and lived there until his junior year of high school when the family moved back to California. Ike earned his B.A. from Bowdoin College in Brunswick, Maine with a major in Geology and a minor in Physics and Astronomy in 2009. After college, Ike spent three years teaching high school physics and middle school earth science at John Burroughs School in Saint Louis, Missouri before enrolling in the M.S. program at the College of William & Mary, School of Marine Science in the fall of 2012. Under Dr. Marjorie Friedrichs, Ike bypassed to the doctoral program and in 2013 he enrolled in the joint Ph.D./M.P.P. degree program between the School of Marine Science and the William & Mary Program in Public Policy. In 2014, Ike was awarded a prestigious internship with the President's Council of Advisors on Science and Technology in the Office of Science and Technology Policy, The White House. In 2017 Ike was awarded the William & Mary Thatcher Prize for Excellence in Graduate and Professional Study, the University's top graduate student award. Ike will graduate from William & Mary in August 2017 with two degrees, a Ph.D. and an M.P.P.