ASSESSING THE CAUSES AND SEVERITY OF GULF OF MEXICO HYPOXIA USING GEOSTATISTICAL AND MECHANISTIC MODELING

by

Daniel R. Obenour

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Natural Resources & Environment and Environmental Engineering)
in the University of Michigan
2013

Doctoral Committee:

Adjunct Associate Professor Anna M. Michalak, Co-Chair
Professor Donald Scavia, Co-Chair
Assistant Professor Inés Ibáñez
Assistant Professor Valeriy Y. Ivanov
Professor R. Eugene Turner, Louisiana State University

ACKNOWLEDGEMENTS

First and foremost, I thank my advisors, Don Scavia and Anna Michalak. Their experience and expertise were invaluable to me throughout the progression of this research. In addition, Don Scavia's positive and open-minded approach to research allowed (and encouraged) me to follow my passions and intuitions, in order to make the best possible use of my time here. Furthermore, Anna Michalak's rigorous feedback and guidance assured that I would leave here feeling like a true expert in the methods employed in this research. I have greatly enjoyed working with both, and hope to continue collaborating with them in future endeavors.

I also thank my committee members: Inés Ibáñez, Valeriy Ivanov, and Gene Turner. I have been taught, challenged, and encouraged by each of them throughout this process. Working with them has greatly broadened my perspective and enhanced my ability to communicate between different disciplines.

In addition, I thank the Louisiana Universities Marine Consortium (LUMCON) staff who collected the data that made this research possible. All environmental data are prone to deficiencies, but I feel that the LUMCON data have been collected with unusual care and attention to detail, and I feel fortunate to have been able use these data in my research. I particularly thank Nancy Rabalais for helping me to sort through and understand the available datasets.

I also thank my university colleagues, both past and present, who have provided friendship and guidance. My academic journey has been positively influenced by a number of outstanding students and faculty at University of Akron, University of Texas, and University of Michigan. At Michigan, I have been very lucky to work with not just one, but two amazing lab groups: Anna Michalak's "PUORG" group and Don Scavia's water quality group. I particularly thank Yuntao Zhou and Mary Ann Evans, who introduced me to some of the key methods used in this research. I also thank my previous

employer, James Miertschin, for introducing me to the fascinating world of water quality modeling and management. By doing this, he (perhaps unwittingly) inspired my return to graduate school.

Finally, I thank my parents, James and Martha (Redd) Obenour, for their many years of guidance, support, and encouragement.

***

This work would not have been possible without the financial support of multiple institutions. The United States Environmental Protection Agency (EPA) Science To Achieve Results (STAR) fellowship program was my principle funding source. In addition, the National Oceanographic and Atmospheric Administration (NOAA) Center for Sponsored Coastal Ocean Research and the University of Michigan Rackham Graduate School both contributed substantially to this work. Throughout this process, I have been cognizant that my funding is ultimately derived from American tax payers, and I have endeavored to perform research worthy of their expense.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Assessing the Causes and Severity of Gulf of Mexico Hypoxia Using Geostatistical and Mechanistic Modeling

By

Daniel R. Obenour

Co-Chairs: Anna M. Michalak and Donald Scavia

Hypoxia, typically defined by dissolved oxygen levels below 2 mg L^{-1}, is an environmental problem common to many coastal systems. One particularly severe example of hypoxia is the large 'dead zone' that forms nearly every summer on the Louisiana-Texas shelf of the northern Gulf of Mexico. While there is considerable agreement about the primary causes of hypoxia, there remains substantial uncertainty regarding its spatial and temporal variability, such that it is difficult to predict how hypoxia will respond to management actions and other environmental changes. This research focuses on improving our understanding of Gulf hypoxia through three types of quantitative modeling. First, a geostatistical regression is developed to empirically model how water column stratification (a primary driver of hypoxia) affects bottom water dissolved oxygen (BWDO) concentrations, and to also infer the importance of other primary drivers, such as nutrient loading. Second, a geostatistical spatial estimation model is developed to simulate BWDO and hypoxic layer thickness across the Gulf shelf, providing estimates of hypoxic zone area and volume for a 27-year study period. Third, a mechanistic model, driven by nutrient loading, flow, and weather conditions is developed to predict hypoxic severity, as determined from the geostatistical model. As with all

environmental models, the models developed in this dissertation are approximations of reality, tuned to limited observational and experimental information, such that they contain significant uncertainty. Because of this, all models are developed within statistical frameworks that quantify uncertainty and allow results to be presented as ranges of likely values. Overall, this works suggests there has been considerable variability in the mid-summer hypoxic extent over the last few decades, and this variability is explained, in large part, by both nutrient loading and oceanographic conditions (i.e., stratification). Relatively parsimonious models that account for these two main drivers explain at least 70% of the year-to-year variability in hypoxic area and mean BWDO. Also, this work indicates that over the past few decades, the Gulf has not become increasingly susceptible to hypoxia formation (independent of the biophysical drivers considered), at least in terms of hypoxic area and mean BWDO.

CHAPTER 1: Introduction

Coastal and estuarine ecosystems provide numerous benefits to society, often serving simultaneously as commercial fisheries, wildlife habitats, and public recreation areas. However, it is widely understood that these beneficial uses can be degraded as a result of human activities (Lotze et al. 2006). An increasingly common cause of degradation is eutrophication, which is defined as an increase in the supply of organic matter to an ecosystem (Nixon 1995). These increases in organic matter supply are often the result of over-fertilization due to anthropogenic nutrient inputs, leading to the excessive production of phytoplankton and higher trophic-level organisms. In addition to causing poor water clarity and harmful algal blooms, a common consequence of severe eutrophication is the development of zones of low dissolved oxygen (hypoxic zones) where ecosystems and fisheries can be severely damaged (Diaz and Rosenberg 1995, Rabalais et al. 2010). Because of their deleterious effect on aquatic life, these hypoxic zones are often referred to as 'dead zones' (Rabalais and Turner 2001, Rabalais et al. 2002c).

Over the last 28 years, a large hypoxic zone has been documented nearly every summer on the Louisiana-Texas Shelf of the northern Gulf of Mexico (Rabalais et al. 2002c, Rabalais et al. 2007). It is the largest human-caused coastal hypoxic zone in the western hemisphere and the second largest worldwide, after the Baltic Sea (Rabalais et al. 2010). This hypoxic zone is caused, at least in part, by nutrient pollution, primarily related to agriculture within the Mississippi River basin (EPA 2007, Goolsby et al. 2001, Rabalais et al. 2007). Upon reaching the Louisiana-Texas shelf, these nutrients promote excessive phytoplankton production and a resulting accumulation of organic material within the system. Hypoxia occurs when the rate of oxygen depletion caused by bacterial decomposition of this organic material exceeds the rate of reaeration from above (Rabalais et al. 2002c). While Gulf surface waters are not generally susceptible to hypoxia because they are well aerated from the atmosphere, bottom waters are more susceptible due to

density stratification of the water column, which inhibits vertical mixing (Wiseman et al. 1997). Density stratification is caused by temperature and salinity differences related to river (freshwater) discharges and summer surface heating.

Gulf hypoxia has received considerable attention from scientists and the public because of concerns regarding its ecological and potential economic consequences (Justic et al. 2007, Rabalais and Turner 2001, Rabalais et al. 2002b). In response to these concerns, a task force of stakeholders and governmental agencies has developed plans to reduce its average size (EPA 2001, 2008). A primary task force goal is to reduce the hypoxic zone's mid-summer areal extent to 5000 km^2 (or less) as a five-year running average. This is considerably less than the current size of the hypoxic zone, which has averaged well over 10,000 km^2 for the last two decades (LUMCON 2012, Rabalais et al. 2007).

Despite substantial advances in hypoxia research over the last few decades, considerable uncertainties remain, confounding management of this environmental problem. For example, scientific debate still exists over the degree to which river-basin nutrient inputs reductions will result in less severe hypoxia (Bianchi et al. 2010, Boesch et al. 2009). This is partially because hypoxic variability is also controlled by oceanographic conditions (primarily stratification), and disentangling the roles of nutrients and stratification is not straightforward, as both of these factors share the same primary driver - discharge from the Mississippi River basin. As a result, there is a need for hypoxia models that can realistically represent both physical and biological processes (EPA 2007, Hetland and DiMarco 2008). Resolving ambiguities surrounding the principal drivers of hypoxia could result in more informed management decisions. And, if the models confirm that nutrient loading is the primary driver of hypoxia, this could help motivate nutrient loading reductions within the Mississippi River basin.

There is also uncertainty surrounding the potential economic consequences of Gulf hypoxia (Rabalais et al. 2010). Understanding linkages between fisheries and hypoxia is complicated, in part, by the limited availability of large-scale metrics of hypoxic severity. Traditionally, Gulf hypoxia has been measured in terms of its areal extent, but other measures, such as the thickness and volume of the hypoxic zone, may also be relevant.

Furthermore, hypoxic extent measurements from different cruises may not be directly comparable, if differences in sampling equipment and spatial coverage are not addressed. Without reliable and comprehensive hypoxic extent estimates, the linkages between hypoxia and its causes and consequences are surely obscured, ultimately complicating management of this environmental problem.

Given these current limitations in our understanding of hypoxia in the Gulf of Mexico, the overall research objectives of this dissertation (1) to estimate the extent and severity of Gulf hypoxia (2) to improve our understanding of how anthropogenic and natural factors influence the temporal variability of hypoxia, and (3) to probabilistically predict the severity of hypoxia under different scenarios based on these factors. I undertake these objectives through quantitative modeling, where geostatistical modeling is used to estimate hypoxic extent from dissolved oxygen sampling data, and empirical and mechanistic modeling are used to relate these extent estimates to biological and physical inputs (predictor variables). All of the models developed here are stochastic in the sense that model parameters and predictions are represented probabilistically. Stochastic modeling allows us to quantify the uncertainty in our predictions and present the 'prediction' as a range of likely outcomes, allowing for more informed policy and management decisions (Reckhow 2003, Reckhow 1994, Refsgaard et al. 2007, Smith and Heath 2001). Ignoring stochasticity may cause stakeholders to view modeling results with either "naïve confidence or unwarranted disbelief" (Scavia et al. 1981).

This dissertation is divided into seven chapters. Chapter 2 presents a literature review for Gulf hypoxia, discussing biophysical processes relevant to hypoxia formation, and models that have been developed to represent those processes. Chapter 3 provides background information on the quantitative modeling methods used in this dissertation. The remaining chapters focus on the dissertation research, as described below.

Chapter 4 presents a geostatistical regression developed to help disentangle the roles of stratification and nutrient loading in determining hypoxic severity. This is the first study to quantitatively model individual BWDO measurements using the intensity of stratification, as determined from measured salinity and temperature profile data. Thus, the influence of stratification is determined directly, by how it explains the spatial variability in BWDO,

rather than by comparing the temporal variability in BWDO to river discharge (a surrogate for stratification intensity, but one that is highly correlated with nutrient loading). As such, it provides a unique approach for understanding the degree to which stratification affects the severity of hypoxia, both spatially and temporally. The remaining temporal variability (not accounted for by stratification), is then be compared to other factors, such as seasonal nutrient loading.

Chapter 5 presents a geostatistical model for estimating hypoxic extent (both area and volume) using dissolved oxygen data from 27 years of mid-summer monitoring cruises. Extent estimates are determined probabilistically using a simulation (i.e., 'conditional realization') approach. In the case of hypoxic volume, a two-step simulation process is required, where BWDO is first simulated, and then hypoxic fraction (hypoxic thickness divided by total water column depth) is simulated over those locations where simulated BWDO is below the hypoxic threshold (2 mg L^{-1}). This is the first time such an approach has been used to estimate the volumetric extent of an environmental impairment; and it allows for the first rigorous estimates of Gulf hypoxic volume and thickness. In addition, this work addresses variations in cruise sampling equipment and spatial coverage that might otherwise bias the results in certain years.

Chapter 6 presents a mechanistic model developed to probabilistically predict hypoxia based on relevant biophysical processes. The model is parsimonious in terms of its resolution and the number of processes considered, but it is sufficiently detailed to allow for quantitative comparisons between modeled and field-measured biophysical rates. In addition, the model formulation is informed by numerous previous Gulf studies, including the geostatistical regression work described in Chapter 4. Spatially, the model is used to predict mean BWDO and hypoxic area (as determined by the geostatistical model, Chapter 5) on the west and east Louisiana-Texas shelf, divided at the Atchafalaya River outfall location. The shelf division allows for a richer model, as the two shelf sections sometimes have very different mean BWDO concentrations. The model is used to explore how the two shelf sections respond to variations in nutrient loading and physical oceanographic conditions. In particular, the model is used to explore how the different shelf sections will respond to nutrient loading reductions, and to analyze long-term trends in the Gulf's

susceptibility to hypoxia formation (independent of seasonal nutrient loading). Because of the model's parsimonious formulation, it can be calibrated within a Bayesian statistical framework, so that model parameters and predictions can be reported probabilistically.

Finally, Chapter 7 summarizes conclusions and future directions. Both the methodological and scientific contributions of the research are highlighted. In addition, logical next steps, building on this dissertation research, are presented. This discussion includes a proposed methodology to further integrate the geostatistical and mechanistic modeling results.

CHAPTER 2: Gulf Hypoxia Literature Review

This section provides a review of Gulf hypoxia literature that is foundational to the dissertation research presented in the following chapters. Section 2.1 covers the primary biological and physical processes that lead to hypoxia formation. Section 2.2 explains how the intensity of hypoxia has been measured, over time, through various monitoring and mapping programs. Finally, Section 2.3 focuses on the wide range of mechanistic and empirical models that have been developed to quantify how various anthropogenic and natural factors affect the severity of hypoxia, and to predict the severity of hypoxia in the future. In general, these models use the biophysical factors described in Section 2.1 to explain the measures of hypoxic severity described in Section 2.2.

2.1 Biological and physical controls on hypoxia formation

A wide array of biological and physical processes is known to influence hypoxia formation in the northern Gulf of Mexico. These processes are described in numerous meta-studies, including an Integrated Assessment report by Rabalais et al. (Rabalais et al. 1999), an assessment update report by the EPA Science Advisory Board (EPA 2007), and an extensive review paper by Dagg et al. (Dagg et al. 2007) . The following discussion is based largely on these sources, but is supplemented by other pertinent literature. A cartoon illustrating the process of hypoxia formation is provided in Figure 2-1.

2.1.1 Gulf eutrophication

Dissolved oxygen depletion is ultimately linked to the accumulation of organic matter within a system (i.e., eutrophication) through aerobic bacterial decomposition, which consumes oxygen and produces carbon dioxide, as represented below, using sugar as an example organic matter compound (Chapra 1997):

$$C_6H_{12}O_6 + 6O_2 \rightarrow 6CO_2 + 6H_2O \qquad \text{eq 2-1}$$

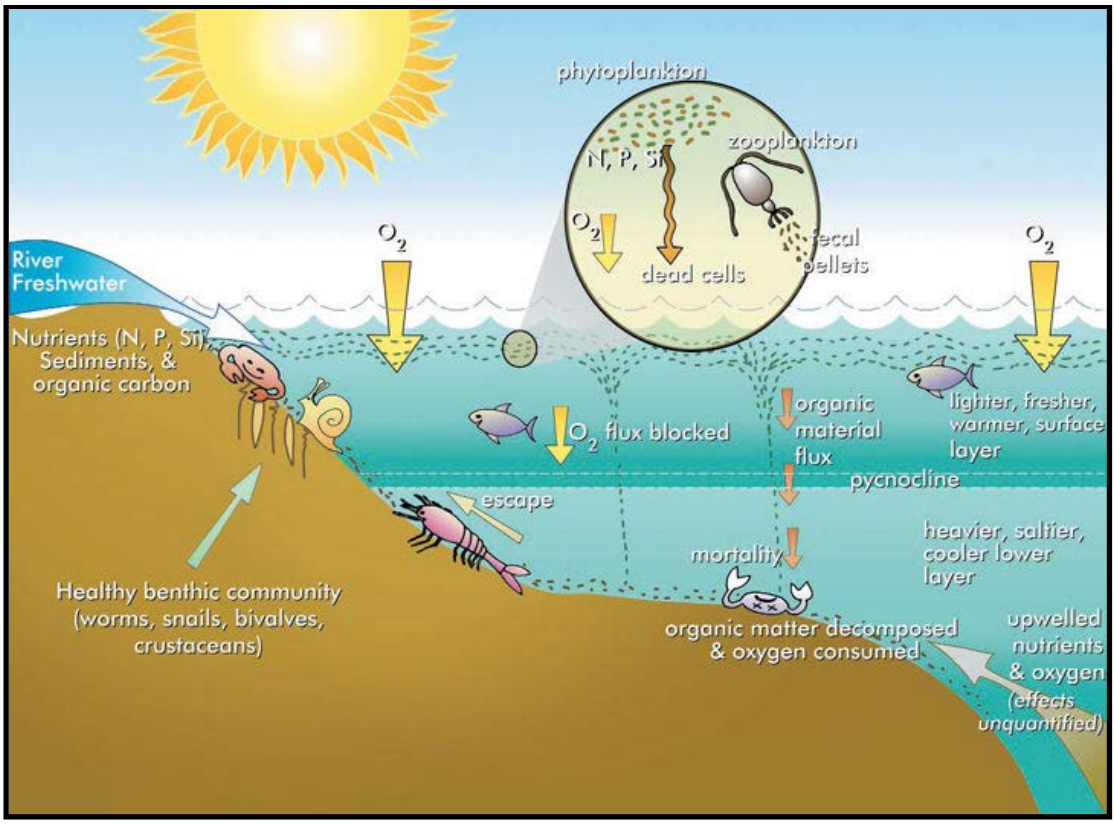

Figure 2-1: Illustration of key processes leading to hypoxia formation (EPA)

There are three plausible sources of organic matter for the Gulf shelf. The first two sources are externally (allochthonously) derived, including organic matter produced terrestrially within the Mississippi River basin and organic matter related to coastal wetland loss. The third source of organic matter is in-situ (autochthonous) production. While there is considerable uncertainty surrounding each of these sources, there are multiple lines of evidence suggesting the third source (in-situ production) is the largest and most influential with respect to hypoxia. First, externally derived organic matter is more refractory than organic matter developed in-situ, and thus more likely to be buried rather than bacterially decomposed. For example, a bacterial decomposition study of Mississippi River water demonstrated that only 34% of riverine dissolved organic carbon (DOC) was lost over a 1.7 year period (Hernes and Benner 2003). Second, even if a portion of riverine organic matter is decomposed on the shelf, the magnitude this organic matter is unlikely to account for more than 1% of total shelf decomposition (Dagg et al. 2007). In addition

DOC decomposition would likely occur in the surface layer, as dissolved materials are not subject to settling. Although some studies suggest that wetland-derived organic matter could be a more important source (Dagg et al. 2007), the general consensus is that it is not a primary contributor to hypoxia because of its refractory nature, and because wetland loss was severe prior to the period of hypoxic intensification (i.e., the 1970s) (EPA 2007). In contrast, in-situ organic matter production is controlled largely by nutrient availability, and nutrient loading from the Mississippi River basin has increased concomitantly with increasing hypoxic severity (Rabalais et al. 2002c). Finally, the dominance of in-situ nutrient-driven organic matter production is consistent with studies of other coastal and estuarine studies (NRC 2000).

In-situ organic matter production is directly related to phytoplankton production, as phytoplankton is foundational to the aquatic food web. Phytoplankton production is controlled by a variety of biophysical factors, of which nutrient limitation is often most important (NRC 2000). Key nutrients are represented in the following equation for phytoplankton ($C_{106}H_{263}O_{110}N_{16}P$) production (Chapra 1997):

$$106CO_2 + 16NO_3^- + HPO_4^{2-} + 122H_2O + 18H^+ + \text{light} \rightarrow C_{106}H_{263}O_{110}N_{16}P + 138O_2 \quad \ldots\text{eq 2-2}$$

From this equation, the molar ratio for the three primary nutrients can be obtained as 106:16:1 (C:N:P), which is commonly referred to as the Redfield Ratio (Redfield et al. 1963), and the equivalent mass ratio is 40:7.2:1. Because carbon is generally available in ample quantities (as carbon dioxide), nitrogen and phosphorus most commonly limit phytoplankton production. In addition, silicon (Si) can be a limiting nutrient for diatoms (a particular, but common type of phytoplankton), and the molecular ratio of Si:N in phytoplankton is 15:16 (Brzezinski 1985).

The molar N:P ratio of Mississippi River water has changed over the decades, and is now around 20:1 (Turner et al. 2006). While a straightforward comparison with the Redfield Ratio suggests phosphorus limitation, this is not necessarily the case, because while phosphorus is efficiently recycled within the system, nitrogen is more susceptible to

loss through denitrification (Rabalais et al. 2002a). A meta-analysis of freshwater and marine systems suggests that an N:P ratio of 50:1 or greater is more indicative of strict phosphorus limitation and that systems with ratios of 20:1 to 50:1 may be limited by either nutrient (Guildford and Hecky 2000). This is generally supported by Gulf nutrient limitation experiments showing evidence of nitrogen, phosphorus, or silicon limitation (or co-limitation) at different times and locations (EPA 2007, Rabalais et al. 2002a). Nonetheless, numerous modeling studies have demonstrated that nitrogen is the most effective predictor of Gulf hypoxia (Greene et al. 2009, Obenour et al. 2012b, Scavia and Donnelly 2007, Turner et al. 2006), suggesting that, overall, it may be the most important limiting nutrient, and thus the most suitable nutrient to target for management control.

Nutrients cycle between organic and inorganic forms at rates determined by environmental conditions. In the Gulf of Mexico, there is much uncertainty surrounding nitrogen cycling processes such as ammonification (decomposition), nitrification, denitrification, fixation, dissimilatory nitrate reduction to ammonium (DNRA), and anammox (Dagg et al. 2007, Lin et al. 2011) (Table 2-1). Ammonification and nitrification consume dissolved oxygen directly, while recycling nitrogen to a bio-available form (NO_3^- or NH_3), potentially leading to additional organic matter production. Denitrification and anammox provide a biochemical pathway for nitrogen to leave the system as N_2 gas. While N_2 gas can also be assimilated through 'fixation', only certain microorganisms (e.g., cyanobacteria) possess this capability, and it is not expected to be a major source of nitrogen to the system. DNRA provides a potential alternative pathway to denitrification and anammox, returning NO_3^- to NH_3. It should be noted that denitrification and anammox (as well as DNRA) are all anaerobic processes, such that nitrogen removal typically only occurs under anoxic conditions, often in the sediment.

Nutrient recycling is relevant to hypoxia in two important ways. First, nutrient recycling within the surface layer allows a greater portion of nutrients (and associated oxygen demands) to be transported across the shelf, rather than settling near the river outfalls (Dagg et al. 2007). Second, nutrient recycling may increase the net oxygen demand fluxed to the lower layer (Dagg et al. 2007). However, for this to happen, decomposition would need to occur in the lower layer, followed by transport of

bioavailable nutrients back into the upper layer, allowing for additional upper layer production. While vertical diffusion or upwelling could potentially accomplish this upward transport, there is currently little information on the degree to which such processes are likely to be important in the northern Gulf. Recycling that occurs solely within the bottom would not result in a net increase in oxygen demand, because the additional respiration would be offset by additional bottom-layer photosynthetic oxygen production.

Table 2-1: Common water column nitrogen cycling processes with dissolved oxygen impacts (Dagg et al. 2007, Madigan et al. 2009)

Process	Starting N species	Ending N species	O_2 effect
Assimilation	NO_3^- or NH_3	organic N	+
Fixation	N_2	organic N	+
Ammonification	organic N	NH_3	-
Nitification	NH_3	NO_3^-	-
Denitrification	NO_3^-	N_2	none
Anammox	NO_2^- + NH_3	N_2	none
DNRA	NO_3^-	NH_3	none

Light availability may also affect phytoplankton production, as solar energy is required for phytoplankton growth. As a result, phytoplankton production is lower near the Mississippi River outfall where plume turbidity inhibits light penetration (Dagg et al. 2007). Turbidity declines rapidly with distance from the river outfall, as suspended sediments settle out of surface waters, allowing phytoplankton production to increase. As such, light limitation should alter the location of phytoplankton production but have little effect on the magnitude of production.

2.1.2 Hydrodynamic transport and stratification

Coastal current patterns affect transport of freshwater and nutrients delivered to the Gulf by the Mississippi River and its distributary, the Atchafalaya River. Throughout much of the year, the dominant flow pattern is westward, and the strongest westward

currents typically occur in the spring, along the inner shelf, due to prevailing easterly winds and the buoyancy flux of the river discharge (Walker et al. 2005, Zavala-Hidalgo et al. 2003, Zhang et al. 2012). Based on previous analyses, this spring coastal current transports over 110,000 $m^3 s^{-1}$ of river and shelf water westward on average (Zavala-Hidalgo et al. 2003); and at times the transport can exceed 140,000 $m^3 s^{-1}$ (Walker et al. 2005). Because the average river discharge is only about 30,000 $m^3 s^{-1}$, the majority of this transport is ocean water. During spring months, typical surface velocities are over 0.2 m s^{-1} westward, suggesting that Mississippi River discharge can be transported 450 km to the Texas border in about 20 days (Zavala-Hidalgo et al. 2003). However, a clockwise gyre often forms directly west of the delta, which can increase travel time (Walker et al. 2005, Zhang et al. 2012). Westward transport is often interrupted in summer by a reversal in wind patterns, causing currents on the shelf to stall or to become eastward (Walker et al. 2005, Zavala-Hidalgo et al. 2003, Zhang et al. 2012).

Wind driven current patterns also affect the fraction of river discharge (and nutrients) delivered to the shelf. During periods of easterly winds, it is reported that over 75% of the Mississippi River discharge is transported westward over the shelf (Walker et al. 2005). Conversely, little Mississippi River discharge enters the shelf when wind conditions reverse, and stable isotope studies suggest Atchafalaya River discharge constitutes the majority of freshwater on the shelf by mid-summer (Strauss et al. 2012).

Water column stratification occurs due to vertical differences in temperature and salinity that affect the density profile of the water column. Surface waters are relatively fresh and warm due to river discharge and thermal heating, while bottom waters are relatively cool and salty, as they are more typical of ocean water. Typically, the water column is effectively divided into two density layers at a location of maximum density gradient known as the pycnocline. However, there are occasionally two pycnoclines that divide the water column into three distinct layers (Wiseman et al. 1997). Water column stratification is important because it inhibits mixing, so that while surface layers remain aerated from the atmosphere, bottom layers are more likely to become oxygen depleted. Studies suggest that intense wind mixing (e.g., due to tropical storms) can quickly break

down stratification and reaerate bottom waters, though the water column will quickly re-stratify in the summer (Walker et al. 2005, Wiseman et al. 1997).

2.1.3 Seasonal timing of hypoxia

The hypoxic zone is generally most expansive in summer due to the seasonal cycles of phytoplankton production and water column stratification, which are the two primary factors leading to DO depletion (EPA 2007, Obenour et al. 2012b, Wiseman et al. 1997). Both factors are related to the outflow from the Mississippi River basin, which typically peaks in March-May. Nutrient loads, coupled with ample sunlight, stimulate phytoplankton production in offshore waters, and much of the resulting organic matter eventually settles along the Louisiana-Texas shelf, where it is available for bacterial decomposition. At the same time, freshwater outflows from the Basin and surface heating promote strong water column stratification. Hypoxic conditions typically persist (with brief interruptions due to storm fronts) until fall, when cooler weather and stronger winds break down the stratification of the water column.

2.2 Documenting hypoxia in the Northern Gulf of Mexico

2.2.1 Hypoxic severity prior to 1985

Regular, comprehensive monitoring of the Louisiana-Texas shelf did not begin until the mid-1980s. Our knowledge of hypoxia prior to this time is largely based on sediment core studies that have been summarized in review papers and synthesis reports (EPA 2007, Rabalais et al. 2002b, Rabalais et al. 2007). These studies focus on the prevalence of bacterial and algal pigments (which are more prevalent and better preserved in sediment corresponding to periods of increased production and hypoxia) and the relative abundance of hypoxia-sensitive versus insensitive benthic foraminifera species. Overall, the sediment cores indicate that dissolved oxygen levels may have been in decline since around 1900, with the greatest declines occurring after 1950. The timing of these declines correspond well with the suspected anthropogenic drivers of hypoxia, including Mississippi River basin land drainage, beginning around 1900, and synthetic fertilizer use, beginning around 1950 (Rabalais et al. 2002b).

2.2.2 Monitoring Campaigns

Much of the hypoxia monitoring on the Louisiana-Texas shelf has been performed by the Louisiana Universities Marine Consortium (LUMCON), beginning around 1985. This monitoring includes intensive mid-summer 'shelfwide' cruise to document hypoxic conditions on the shelf, and to collect other water quality measurements related to hypoxia (LUMCON 2012, Rabalais et al. 2007). From these cruises, the hypoxic zone extent has been operationally defined as the area where bottom-waters have dissolved oxygen concentrations of less than 2 mg/L. Figure 2-2 shows locations commonly sampled during the shelf-wide cruises. Additional stations are often sampled further west depending upon the spatial disposition of the hypoxic zone and available ship time. As shown, sampling is typically performed along north-south transects, which are identified alphabetically (AA-K). In general, samples are collected at 5-15 kilometer intervals along these north-south transects. The transect spacing in the east-to-west direction ranges from about 20-40 kilometers.

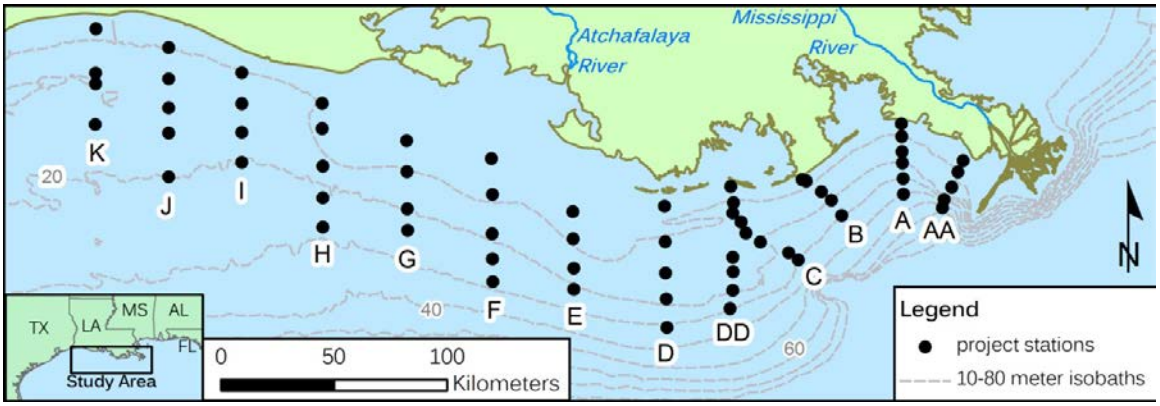

Figure 2-2: Common sampling locations for LUMCON summer shelf-wide cruises

Other types of hypoxia monitoring performed by LUMCON include 'transect sampling' and 'moored sampling' (LUMCON 2012). These types of monitoring have less spatial coverage, but are performed with greater frequency. The transect sampling typically occurs monthly or bi-monthly at roughly 16 stations (transects C and F) and has been conducted fairly regularly since 1990. There are also two moored sampling stations

(near the middles of the B and C transects) that provide virtually continuous measurements of water quality and current information. Moored sampling has been conducted since 1989, but sometimes at only one station. Also, the moored sampling equipment is occasionally fowled or taken down for maintenance, so some periods of time are missing from the record.

In addition to the LUMCON cruises, additional surveys are performed annually by the Southeast Area Monitoring and Assessment Program (SEAMAP) and Texas A&M University (TAMU). The SEAMAP cruises, conducted since 2001, are part of a Hypoxia Watch program developed by NOAA and other federal agencies (LUMCON 2012). These cruises cover an extensive area along the Gulf coast, ranging from east of the Mississippi River delta to the US/Mexico border. Because of their large extent, they take around one month to complete, generally from mid-June to mid-July. Compared to the LUMCON shelf-wide surveys, the SEAMAP sampling is coarser and is limited to depths of greater than 10 meters. TAMU began extensive monitoring of hypoxia in 2004 (TAMU 2011). Like the SEAMAP cruises, the TAMU cruises generally cover a large area but at coarser spatial resolution than the LUMCON shelf-wide cruises. The university also maintains a moored sampling station (since 2009) located near the G transect at about the 15-meter isobath.

All institutions use Conductivity, Temperature, Depth (CTD, a.k.a. 'rosette') sampling instruments to measure water quality parameters at each sampling location. Despite the name, CTDs almost always includes a dissolved oxygen probe. The CTD is typically lowered up and down throughout the water column at each station in order to provide a vertical profile of water quality measurements. However, due to concerns about damaging the instrument, the CTD is generally stopped about one meter above the seafloor. Because of this limitation, LUMCON also uses a smaller, more durable 'handheld' DO probe, which is lowered until it reaches the very bottom of the water column. The dissolved oxygen sensors on both of these devices are regularly calibrated using Winkler titrations, e.g. Rabalais et al. (Rabalais et al. 1996).

2.2.3 Mapping hypoxia

The intensity of Gulf hypoxia has been traditionally assessed in terms of its areal extent. From the inception of shelfwide cruise monitoring (1985) to present, hypoxic area has ranged from negligible in 1988 to 22,000 km^2 in 2002 (Rabalais et al. 2007). The reported hypoxic area values are determined by interpolating and hand-contouring between bottom-water dissolved oxygen (BWDO) concentrations measured at cruise sampling locations. In addition, published maps are often developed by kriging using the Surfer® geographical software package. (Rabalais et al. 1999, Rabalais et al. 2007). Kriging is a form of optimal interpolation described in greater detail in Chapter 3. Figure 2-3 shows the kriged hypoxic area for July 2002, which is reported to be the largest hypoxic area on record (LUMCON 2012).

Hypoxia Watch (SEAMAP) and Texas A&M use similar methods for mapping dissolved oxygen concentrations from their cruise data. Hypoxia Watch performs interpolation though kriging using the ArcGIS® geographical software package (NOAA 2011a). Texas A&M performs interpolation using 'objective analysis' (also referred to as 'Gauss-Markov smoothing') which is generally equivalent to kriging (Cressie 1990, DiMarco 2011, Emery and Thomson 2001). However, unlike LUMCON, Hypoxia Watch and Texas A&M have yet to publish hypoxic area estimates.

Figure 2-3: Areal extent of hypoxia determined by LUMCON for July 2002 (LUMCON 2012)

2.3 Modeling hypoxia in the Northern Gulf of Mexico

This section discusses different models that have been developed to study hypoxia in the northern Gulf. The models are divided into two primary categories: mechanistic models (Section 2.3.1) and empirical models (Section 2.3.2). While this distinction is somewhat subjective, the classification is based primarily on how the models are shown to be valid. The validity of mechanistic models is derived primarily from the adequate representation of biophysical processes, while the validity of empirical models is derived primarily from the ability to be fit to observed data. Of course, there is some overlap, as mechanistic models should still be capable of representing observed data and empirical models should be informed by known biophysical processes. Finally, Section 2.3.3 provides a comparison of simple versus complex models, in terms of their scientific and management utility.

2.3.1 Mechanistic models

Over the last two decades, a wide array of water quality models has been developed to simulate dissolved oxygen levels in the northern Gulf (see Table 2-2 for a comparison of mechanistic hypoxia models). In a review paper, Scavia et al. (2004) describe three of the earlier mechanistic Gulf hypoxia models. The simplest of these is a steady-state one-dimensional model, hereafter referred to as the "Scavia model," that simulates sub-pycnocline dissolved oxygen concentrations in the longshore dimension (Scavia et al. 2003). It is based on the classic 'Streeter-Phelps' model, commonly used to determine the DO profile of a river downstream of wastewater discharge location (Chapra 1997). The Scavia model treats the Mississippi and Atchafalaya River outlets as the two loading locations, where river May-June nitrogen loads are converted directly to oxygen demands based on stoichiometric ratios. The modeled DO results can be used to directly determine the length of the hypoxic zone, and based on an empirical relationship between hypoxic area and length (using LUMCON mapping products described in Section 2.2.3), results can be converted to hypoxic area (Scavia et al. 2003). The Scavia model includes a westward advection term that is often used for model calibration (Scavia et al. 2003, Scavia and

Donnelly 2007). If the term is held to a constant value, then the model explains 45% of the interannual variability in hypoxic area. If the term is allowed to vary, then model performance improves. However, the variability of this parameter is generally attributed to unresolved oceanographic conditions, and is thus treated stochastically for forecasting purposes (Scavia et al. 2003). Because this biophysical formulation is somewhat abstract, parameter values can only be loosely compared to true biophysical rates, and the model could arguably also be classified as empirical, as discussed in Section 2.3.2.

In more recent studies, the Scavia model has been modified to explore particular scientific questions. Scavia and Donnelly (2007) used the model to explore inter-decadal trends in hypoxia formation, through hindcasting backward in time based on historical nutrient loads; and to study whether nitrogen or phosphorus is the primary control on the size of the hypoxic zone. Based on this work, it was possible to comment on the efficacy of various nutrient control strategies. Also, Donner and Scavia (2007) used the model to explore how precipitation over the Mississippi River basin controls hypoxic variability. Finally, Liu et al. (2010) used the model to explore the legacy effect of nutrient build-up in the Gulf, suggesting that there have been two stepwise increases in the Gulf's sensitivity to nutrients since 1985.

A second early model, developed by Justić et al. (1996, 1997), simulates dissolved oxygen at a moored sampling location on the eastern Louisiana shelf (middle of the C transect, Figure 2-2). Overall, the 'Justić' model is more complex than the Scavia model because it is dynamic and because it simulates more detailed biophysical processes, both above and below the pycnocline. It includes oxygen fluxes between the different layers and the atmosphere, and biomass production and decomposition. Using observed oxygen concentrations and calculated oxygen fluxes, it was possible to infer the rate of net productivity in the surface layer and respiration in the bottom layer. The net productivity could then be related to nutrient loading, effectively linking hypoxia to nutrient loading (Justic et al. 1997). The original model was calibrated for data from 1985-1992, and verified for data from 1993. Model verification was performed in a qualitative fashion by comparing plots of observed and simulated values. In more recent studies, the Justić model has been used to examine inter-decadal trends in hypoxia formation through

hindcasting (Justic et al. 2001, 2002), and results suggest that hypoxia intensified in the 1970s due to increased nutrient loading, consistent with Scavia and Donnelly (2007). Hypothetical modeling scenarios suggest that reducing nitrate plus nitrite (NO_{2-3}) loads will substantially reduce hypoxia, but that increases in river flow (possibly due to climate change) could increase the intensity of hypoxia.

Arguably, the most complex of the early models is a steady-state, three-dimensional model that simulates nitrogen and phosphorus cycling, algal growth and decay, and oxygen dynamics. This model, developed by Bierman et al. (1994), uses the Water Analysis Simulation Program (WASP) eutrophication module, and hydrodynamics were estimated primarily from field observations and best professional judgment. Model results suggest that primary production largely controls oxygen dynamics on the shelf, though benthic oxygen demands were also significant. In addition, the model stresses the importance of underwater light attenuation as a control on primary production. For this reason, primary production may be a substantial source (as well as a sink) of BWDO on the shallower western shelf, where light attenuation is less severe. A limitation of this model is that it was only executed for a steady state scenario corresponding to July 1990. The model was calibrated in a qualitative fashion by comparing plots of observed and simulated values.

Recent years have seen the development of models with increasing hydrodynamic complexity. Wang and Justić (2009) have developed a detailed hydrodynamic model using the Finite Volume Coastal Ocean Model (FVCOM) for the Louisiana-Texas shelf. The model includes over 10,000 segments and runs at a sub-minute time step, making it computationally intensive. So far, results have only been published for the summer of 2002. The model was verified by comparing plots of simulated and observed data, and by using statistical measures of model/data agreement. While the model was forced with several climatological factors, wind velocity and freshwater discharge were found to be the dominant mechanisms controlling circulation and water column stratification in the study area. Future publications will include a coupling of FVCOM with a modified version of WASP, so that biochemical processes (and hypoxia) are also simulated (D. Justić, personal communication, 2013).

Hetland and DiMarco (2008) have also developed a detailed hydrodynamic model for the Louisiana-Texas shelf, using the Regional Ocean Modeling System (ROMS). Initial modeling results, using simple representations of oxygen demand, suggest water column respiration is the most important driver of hypoxia on the eastern shelf, while benthic respiration appears to be most important on the western shelf. In addition, results suggest that there is little cross-shelf transport of hypoxic bottom waters, such that hypoxia formation is primarily controlled by vertical water column processes, as also indicated by Bierman et al. (1994). The ROMS model was more recently expanded by Fennel et al. (2011) to included more detailed biochemical processes, such as nutrient cycling and phytoplankton growth. Simulations were performed over a fifteen-year period (1990-2004), and model results were verified by comparing plots of simulated and observed data, along with statistical measures of model/data agreement (skill metrics). Modeling results suggest the importance of phytoplankton loss mechanisms (e.g., grazing and sinking) for controlling the accumulation of organic matter within the system. Subsequent publications using this model have explored phosphorus limitation (which may shift the timing and location of phytoplankton production) and benthic oxygen demands (which were found to be an important driver of hypoxia) on the Gulf shelf (Fennel et al. 2013, Laurent et al. 2012).

Overall, mechanistic models have been useful for testing hypotheses about how the Gulf coastal system functions, and for providing insights into which biophysical processes are most influential with regards to hypoxia formation. These models generally suggest that seasonal nutrient loads, benthic oxygen demands, and oceanographic conditions all likely contribute to hypoxia formation on the Gulf shelf, but the relative importance of these factors appears to vary among models. A primary challenge, when comparing such models, is to evaluate the degree to which each model is realistically parameterized. Model calibration and validation exercises can help to justify a model's parameterization; and advances in the spatial, temporal, and process resolution of models have allowed for more thorough model-data comparisons. However, given the large number of parameters common to most mechanistic models, there may be multiple parameterizations that produce equally satisfactory model-data comparisons, but that may have different scientific

and management implications (Beven 2006). Without rigorous methods to account for parameter uncertainty (and covariance among parameters), it is difficult to ascertain the level of confidence that should be assigned to the conclusions of mechanistic modeling studies. Parameter uncertainty has been rigorously assessed for the Scavia model using Bayesian methods (Liu et al. 2010), but because the model only includes nutrient loads as input, it has limited capacity to provide insight into other biophysical processes relevant to hypoxia formation. More complex mechanistic models have not rigorously accounted for parameter uncertainty (likely because it is computationally infeasible to do so). Instead, these studies have focused on model sensitivity to individual parameters, such as sediment and water column oxygen demands (Fennel et al. 2011, Hetland and DiMarco 2008). However, this type of analysis is less about quantifying uncertainty, and more about inferring the 'correct' value of a given parameter or parameters, under the arguably tenuous assumption that all other model parameters are realistically prescribed.

2.3.2 Empirical models

In addition to the mechanistic models described above, a number of empirical models have also been developed to assess hypoxia in the Gulf of Mexico. Turner et al. (2005, 2006, 2008, 2012) use linear regression models to analyze how nutrient loading affects the size of the hypoxic zone. These studies demonstrate that a large portion of the temporal variability in hypoxic area can be explained by the spring $NO_{2\text{-}3}$ load. In addition, these studies suggests the system is becoming increasingly susceptible to hypoxia over time, with a given $NO_{2\text{-}3}$ load now producing a larger hypoxic area than it would have in previous decades. This increase in hypoxic susceptibility appears to have been interrupted by climatological events, but even so, the trend with time is substantial (r^2=0.69). It is further suggested that changes in hypoxic susceptibility may have occurred as sudden shifts ('change points' in time) due to positive feedbacks that result in alternate states of ecosystem functioning (Turner et al. 2008, 2012).

The Scavia model, described previously as a mechanistic model, could arguably also be classified as an empirical model. Although this model includes some biophysical processes that can be loosely compared to field-measured rates, the scope of mechanisms

considered in this model is limited. Thus, the model is validated more in terms of its predictive ability, and less in terms of its representation of biophysical processes. Overall, the Scavia model was found to explain 45% of the variability in the hypoxic area when using May-June total nitrogen (TN) load to drive oxygen demand (Scavia et al. 2003).

Uncertainty quantification has always been a focus of this model. In earlier versions, uncertainty was represented by a single model parameter with a distribution determined through model calibration; and forecasts were made by sampling from this parameter's distribution using a Monte Carlo approach (Scavia et al. 2003, Scavia and Donnelly 2007). In later versions of the model, a Bayesian approach allowed uncertainty to be represented within multiple stochastic model parameters (Evans and Scavia 2011, Liu et al. 2010).

Greene et al. (2009) developed a series of linear regression models that test the predictive potential of a large suite of river load, flow, and concentration variables. May $NO_{2\text{-}3}$ load was found to explain 42% of the variability in hypoxia, more than any other single variable. However, a multiple linear regression model, using flow and nutrient concentration variables, along with a change point in hypoxic susceptibility, was found to be more effective than regressions using nutrient load alone. Forrest et al. (2011) conducted a similar linear regression study, but added wind speed data to the mix of potential variables. East-west wind velocity (from 15 June – 15 July) was found to be an effective predictor of hypoxia when used in multiple linear regressions. The importance of wind was also confirmed by Feng et al. (2012), who included westerly wind duration as an additional covariate in Gulf hypoxia regression models.

The empirical models share several common characteristics (Table 2-3). First, all models include some form of spring river flow or nitrogen load (typically May or May-June). Because flow and nitrogen load are correlated, they perform similarly in empirical models (Bianchi et al. 2010), though loads tend to perform somewhat better (Forrest et al. 2011, Greene et al. 2009). Regardless of which performs better, the correlation between these variables means that the relative influence of freshwater flows (which promote stratification) and nutrient loads (which promote organic matter production) cannot be completely disentangled based on these empirical models alone. This ambiguity can

confound hypoxia mitigation, because while nutrient loading can be reduced by watershed management practices, flows are largely outside the realm of human control.

In addition, most empirical models omit one or more 'outlier' years that are not well-represented by the model formulation and selected input variables. These outlier years sometimes include storm years (e.g., 2003), drought years (e.g., 1988), and years with unusually strong westerly winds (e.g., 2009). Removal of these types of outliers may be appropriate given the assumptions and objectives of particular models. However, this culling reduces the size and variability of the calibration dataset, and limits the range of conditions over which the model can be used to make predictions.

The empirical models also tend to include 'change points' that delineate periods with different susceptibilities to hypoxia ('alternate states'). There is some disagreement among models as to exactly when these changes occurred, but all models include at least one change point sometime in the early 1990s. Changes in system susceptibility were also indicated in an empirical study by Stow et al. (2005) that examined the degree of stratification required for hypoxia to form. However, to perform this type of long-term analysis, it is important to ensure that the monitoring practices used to develop the calibration dataset have remained consistent across the period of interest. Chapter 5 discusses how changes in monitoring practices have affected the temporal trends in the hypoxic area estimates used to calibrate these models.

In summary, the empirical models are largely predictive tools, allowing hypoxia to be forecasted (or hindcasted) based on its principle drivers, with a historical emphasis on the nutrient loading driver. Because of their computational efficiency, these models can be readily applied over long (multi-decade) time periods to assess long-term trends in hypoxic severity, and to identify potential changes in the system's susceptibility to hypoxia. However, there are several different empirical Gulf hypoxia models, each suggesting somewhat different predictor variables (Table 2-3). These differences arise, at least in part, from inconsistencies regarding what candidate predictor variables are considered, the criteria for determining outlier years, and the criteria for determining system change points.

2.3.3 Simple versus complex models

Gulf water quality models benefit scientists and natural resources managers by synthesizing knowledge about the various anthropogenic and environmental causes of hypoxia, and by predicting how the severity of hypoxia will respond as anthropogenic and environmental factors change over time (Justic et al. 2007). Existing Gulf models vary greatly in their complexity, and the more complex models generally have finer spatial resolution and represent a wider range of biophysical processes. The mechanistic models tend to be more complex, while the empirical models tend to be simpler. As discussed below, there are both advantages and disadvantages to increasing model complexity. The modeling performed in this dissertation research is not particularly complex (under this definition) but it includes sufficient spatial, temporal, and biophysical detail, to address relevant scientific and management questions.

Historically, there has been considerable debate over the appropriate level of complexity to include within environmental models, and both simple and complex models have their proponents (Canham et al. 2003). Proponents of simple models argue that they are most advantageous for exploring high-level questions and for making predictions and comparisons regarding future ecosystem states (Pace 2003). Simple models tend to avoid over-parameterization issues that can confound more complex models (Reckhow and Chapra 1999), and simple models are more transparent, so that they can be readily communicated to other scientists, as well as to policy makers and the public. However, complex models are generally more capable of capturing the spatial and temporal heterogeneity of a system, allowing more detailed model output and richer comparisons with field data (if available). Perhaps most importantly, complex models are argued to be less likely to inadvertently ignore mechanisms relevant to the scientific and management questions being investigated (DeAngelis and Mooij 2003).

Because models are used for different purposes, different levels of complexity may be appropriate for different applications. For this reason, Justic et al. (2007) argue that both simple and complex models have important roles in the science and management of Gulf hypoxia. Complex, three-dimensional models are useful for exploring the detailed

biological and physical processes related to dissolved oxygen dynamics, and for identifying the primary processes controlling hypoxia formation. Complex models can also be useful for predicting how the system will respond to future conditions, though their intensive data and computational requirements may somewhat limit their efficacy as a forecasting tool. On the other hand, simpler, more computationally efficient models are capable of being run over longer time periods, and are more readily used in hindcasting and forecasting applications. The computational efficiency of simpler models also allows them to be calibrated within statistical frameworks, so that uncertainty can be rigorously quantified – a primary goal of this dissertation research.

Table 2-2: Summary of published mechanistic models for Gulf of Mexico hypoxia

Model publication dates	Temporal character	Modeling period	Hindcasting period*	Spatial dimensions	Segments	Hydrodynamic model	Biochemical model
Bierman et al. (1994)	static	1990	-	3	21	3-D, determined by observations	nutrients/ production/ respiration/ oxygen (WASP)
Justic et al. (1996,97, 2001,02)	dynamic	1985-2000	1955-1984	1	2	wind-driven vertical mixing	production/ respiration/ oxygen
Scavia et al. (2003,07) Liu et al. (2010)	static	1985-2008	1955-1984	1	continuous	constant westward advection	oxygen demand/ oxygen
Hetland et al. (2007) Fennel et al. (2011,13)	dynamic	1990-2004	-	3	10000+	3-D, multiple forcings (ROMS)	nutrients/ production/ respiration/ oxygen
Justic et al. (2009)	dynamic	2002	-	3	10000+	3-D, multiple forcings (FVCOM)	nutrients/ production/ respiration/ oxygen (WASP)

*this period includes years when little or no calibration/verification data are available.

Table 2-3: Summary of published empirical models for Gulf of Mexico hypoxia

Model (latest publication)	Characteristics of optimum model from each study						
	Nutrients*	Flows*	Change points	Winds	R^2	Modeling period	Years omitted
Greene et al. (2009)	[$NO_{2\text{-}3}$(5+6)],[TP(2)]	Q(5)	93-94	-	81%	1985-2007	-
Liu et al. (2010)	TN(5+6)	-	91-92,97-98	-	n.r.	1985-2008	98,03
Forrest et al. (2011)	[$NO_{2\text{-}3}$(5)]	Q(5)	93-94	e-w velocity	73%	1985-2009	88
Turner et al. (2012)	$NO_{2\text{-}3}$(5)	-	89-90,99-00	-	n.r.	1979-2011	97,03,05,08-11
Feng et al. (2012)	$NO_{2\text{-}3}$(5+6)	-	92-93	w duration	69%	1985-2010	03,05

*parenthetical values represent month, unbracketed nutrients represent loads, bracketed nutrients represent concentrations

n.r. = not reported,

e/w = easterly/westerly winds

CHAPTER 3: Methods Background

This section provides a brief introduction to the two methodological approaches highlighted in this dissertation research: geostatistical and Bayesian modeling. Although other methods are also used (as described in later chapters), these are the methods most unique to this research, and are thus deserving of the additional exposition provided here. Section 3.1, describes the geostatistical modeling approach, which is applied in Chapters 4 and 5, and Section 3.2 describes the Bayesian modeling approach, which is applied in Chapter 6. The basic methodologies presented here are expanded considerably in in these following chapters.

3.1 Geostatistics modeling background

The majority of the work described in Chapters 4 and 5 uses methods from the field of geostatistics. While geostatistics is based largely on mathematical and statistical theory, it was originally developed by practitioners in the applied sciences, including mining, forestry, and agriculture (Gelfand et al. 2010). Perhaps because of its diverse and applied background, the use of geostatistics has grown rapidly since its introduction some 50 years ago. In statistical circles, geostatistics is considered to be a particular case of the larger field of 'spatial statistics' (Gelfand et al. 2010). Geostatistics focuses on the modeling of continuous spatial variation, often assuming a Gaussian error process that allows for uncertainty quantification and the use of efficient likelihood-based variable selection and model parameterization techniques (Zimmerman 2010, Zimmerman and Stein 2010).

According to Chiles and Delfiner (1999), the goal of geostatistics is to provide a quantitative description of data that vary through space (or time and space). The primary advantage of geostatistics, when compared to simpler statistical methodologies that assume independent and identically distributed (i.i.d.) data, is that the spatial correlation among the data is accounted for explicitly. Thus, geostatistics is applicable to the common situation

where observations from nearby locations tend to be more similar than observations from locations that are further apart (all other known factors being equal). Also, with its probabilistic framework, geostatistics is well-suited to describe the substantial uncertainty associated with systems that are characterized by spatial heterogeneity and limited observational data. Because many environmental variables, including dissolved oxygen in the Gulf of Mexico, have these characteristics, geostatistics is a natural choice for the study of these systems.

The effectiveness of geostatistics for studying environmental systems has been documented in numerous previous studies. For example, Goovaerts (2000) used geostatistical methods to interpolate rainfall intensities across space using elevation as a covariate. Erickson et al. (2005) developed a geostatistical model to interpolate snow depths across an alpine basin using multiple statistically-selected covariates (e.g., wind sheltering and solar radiation) that provided insight on the primary factors controlling snow accumulation. Mueller et al. (2010) used a geostatistical regression model to infer factors (covariates) influencing the temporal variability of forest CO_2 fluxes. Murphy et al. (2010) used geostatistical methods, incorporating biophysical model output, to estimate water quality indicators throughout Chesapeake Bay. Zhou et al. (2013) used geostatistical methods to interpolate dissolved oxygen concentrations and assess the extent of the hypoxic zone in Lake Erie.

3.1.1 The univariate geostatistical model

A univariate geostatistical model is implemented by representing a response variable (**z**, eq 3-1) in terms of a deterministic component and two stochastic components (Gneiting and Guttorp 2010, Zimmerman and Stein 2010). (Here, 'univariate' implies a model with a single response variable, as opposed to a 'multivariate' model with multiple response variables.) The deterministic component, $\mathbf{X\beta}$, is the portion of **z** that can be expressed in terms of known covariates, a.k.a. predictor variables. In literature, the deterministic component is sometimes referred to as the 'drift in the mean' or sometimes simply as the 'mean function'. The deterministic component represents a 'mean' in the sense that it is the central value around which stochastic model components are distributed. When the

deterministic component is represented using **Xβ**, it is implied that the mean function is a linear model (typically empirical) with covariate matrix **X** and regression coefficient vector **β,** as in a multiple linear regression. While non-linear mean functions can also be used, linear functions are much more common because their implementation is computationally efficient. In the simplest case, without covariates, **Xβ** reduces to the mean of the data (**X** is reduced to a vector of ones and **β** becomes a scalar equal to the mean).

$$\mathbf{z} = \mathbf{X\beta} + \boldsymbol{\eta} + \boldsymbol{\varepsilon} \qquad \text{eq 3-1}$$

The two stochastic components, **η** and **ε,** represent the portion of **z** not accounted for by the deterministic component (a.k.a. the 'residuals'). Here, **η** represents spatially correlated stochasticity and **ϵ** represents uncorrelated stochasticity. The inclusion of **η** is essentially what differentiates a geostatistical model from simpler statistical models.

The different model components can be illustrated graphically for an example dataset in one-dimensional space (Figure 3-1). Here, **Xβ** is a simple linear trend with space, and the stochastic portions of the model fluctuate around this trend. In this case, the magnitude of the spatially correlated stochasticity (**η**) is similar to that of the uncorrelated stochasticity (**ε**).

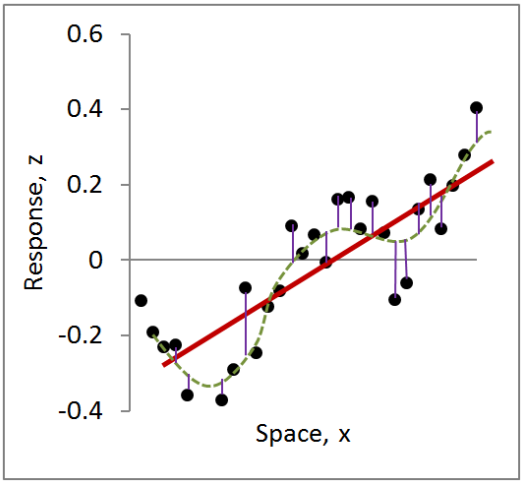

Figure 3-1: Example dataset modeled as a spatial tend (**Xβ**) in red, plus spatially correlates stochasticity (**η**) in green, plus uncorrelated stochasticity (**ε**) in purple

3.1.2 Modeling spatially correlated stochasticity

The stochastic portion of a model may or may not include spatial correlation. If the residuals from the deterministic component are i.i.d., then no spatial correlation exists ($\boldsymbol{\eta}$=0), and if in addition the variance of $\boldsymbol{\varepsilon}$ is constant in space, the geostatistical model reduces to a regular linear regression model. However, most environmental phenomena do exhibit some spatially correlated stochasticity because the deterministic component can rarely account for all of the spatial patterns in the data.

Spatial correlation can be explored through a 'variogram', which shows how the variance of the stochasticity (i.e., the residuals of the deterministic component) increases as the separation distance (h) between locations increases. The variogram is developed by determining the semivariance (γ, eq 3-2) between observation locations, where r_i and r_j are the stochastic values (residuals) at locations i and j.

$$\gamma_{ij} = (r_i - r_j)^2/2 \qquad \text{eq 3-2}$$

A 'raw' variogram is a plot of semivariances (γ_{ij}) versus separation distances (h_{ij}) for all possible pairs of residuals (Figure 3-2). (This example variogram is based on the geostatistical model for dissolved oxygen, with units of mg L^{-1}, as described in Chapter 5.) The blue line in Figure 3-2 is an experimental variogram, created by calculating the average semivariance, $\gamma(h_{ij})$, for different ranges (bins) of separation distance. Finally, the red line represents a theoretical variogram function that was fit to this stochasticity. Although multiple types of variogram functions are available (Chiles and Delfiner 1999, Zimmerman and Stein 2010), an 'exponential' variogram with 'nugget' effect (eq 3-3) is used in Figure 3-2 because it conforms well to the experimental variogram.

$$\gamma(h_{i,j}) = \begin{cases} 0, & h_{i,j} = 0 \\ \sigma_\varepsilon^2 + \sigma_\eta^2 \left[1 - \exp\left(-\frac{h_{i,j}}{l_h}\right)\right], & h_{i,j} > 0 \end{cases} \qquad \text{eq 3-3}$$

In eq 3, σ_η^2 and σ_ε^2 are parameters representing the spatially correlated and uncorrelated portions of the variance of the stochasticity, respectively. Thus, the variance is zero at a separation distance of zero. At separation distances approaching zero, the variance between locations is σ_ε^2, commonly referred to as the 'nugget', representing micro-variability and/or measurement error. As separation distances increase to infinity, the variance increases to the 'total sill' ($\sigma_\varepsilon^2 + \sigma_\eta^2$), where σ_η^2 is commonly referred to as the 'partial sill'. The term l_h is a range parameter related to how quickly the variance increases with increasing separation distance. For an exponential variogram function (eq 3-3), the variance reaches 95% of its maximum value at a separation distance of approximately $3l_h$, commonly referred to as the 'effective range,' beyond which the data are nearly uncorrelated.

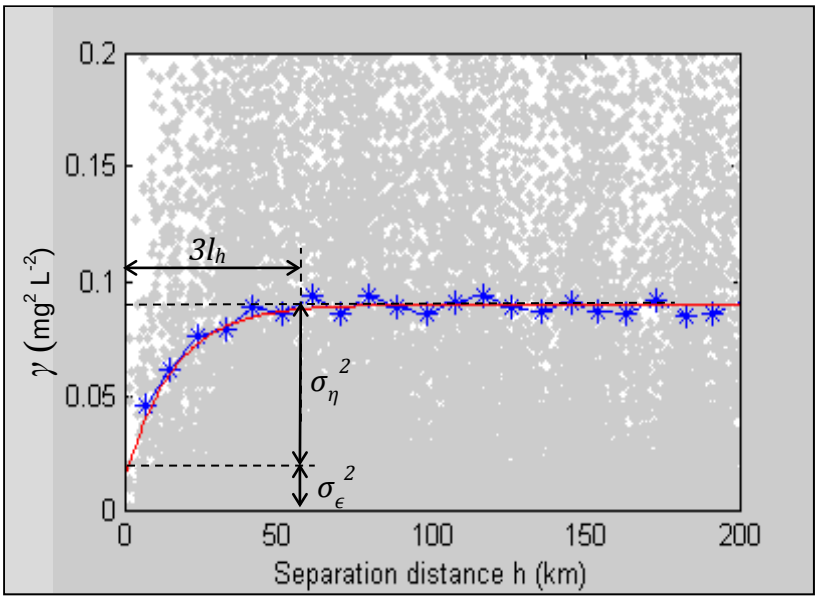

Figure 3-2: Example raw (gray), experimental (blue), and theoretical (red) variograms

When stochasticity can be characterized as second-order stationary, then a covariance function can be defined (Chiles and Delfiner 1999, Gneiting and Guttorp 2010). A variogram that approaches an asymptote as h approaches infinity, as in Figure 3-2, implies the system is second-order stationary. Here, the covariance function is simply the total sill ($\gamma(h)_{h\to\infty}$) minus the variogram function:

$$Q(h_{i,j}) = \gamma(h)_{h\to\infty} - \gamma(h_{i,j}) \qquad \text{eq 3-4}$$

Based on this relationship, the exponential variogram function can be converted to an exponential covariance function:

$$Q(h_{i,j}) = \begin{cases} \sigma_\eta^2 + \sigma_\varepsilon^2, & h_{i,j} = 0 \\ \sigma_\eta^2 \exp\left(-\frac{h_{i,j}}{l_h}\right), & h_{i,j} > 0 \end{cases} \qquad \text{eq 3-5}$$

where all parameters (σ_ε^2, σ_η^2, and l_h) are the same as those defined in eq 3-3. Note that the covariance function is often referred to as a 'model', but 'function' is used here in order to distinguish it from the overall geostatistical model (eq 3-1).

In this research, a variogram analysis was performed to examine the spatial structure of the system and to select an appropriate variogram/covariance function (in this case, an exponential function with nugget was selected). The variogram/covariance function was then parameterized using a likelihood-based method, as described in Section 4.2.3. As such, it was not necessary to visually fit the variogram/covariance parameters. However, the suitability of the likelihood-based parameterization was confirmed by comparing the resulting variogram/covariance function to the experimental variogram.

3.1.3 Geostatistical interpolation

The field of geostatistics was originally developed for the purposes of spatial interpolation, and this section discuses that aspect of geostatistical modeling. The earliest version of geostatistical interpolation was developed in the 1950s by South African mining engineer, D.G. Krige, from whom the term 'kriging' was derived. The original goal of kriging was to estimate the quantity of ore within a given area using statistically robust methods. The methodology was significantly advanced and promoted by French geologist and mathematician, Georges Mathéron, in the 1960s (Chiles and Delfiner 1999). In addition, an essentially equivalent methodology termed 'objective analysis' (or 'Gauss-Markov smoothing'), was simultaneously developed by Soviet meteorologist L.S. Gandin

(Cressie 1990, Emery and Thomson 2001). Over the last several decades, these methods have been applied to a wide array of environmental variables, including dissolved oxygen concentrations in the northern Gulf of Mexico, as discussed in Section 2.2.3.

Kriging is used to predict the value of a response variable at unobserved locations across space. For a univariate geostatistical model (eq 3-1) the predicted response ($\tilde{z}$) is determined from the vector of observed values ($\mathbf{z}$), and a vector of geostatistical weights ($\boldsymbol{\lambda}$) corresponding to these observations:

$$\tilde{z} = \boldsymbol{\lambda}^{T}\mathbf{z} \qquad \text{eq 3-6}$$

The vector of weights ($\boldsymbol{\lambda}$) is determined based on a system of equations described in detail in Chapter 5. Note that eq 3-6 applies when the mean or trend (i.e., the deterministic component) is estimated as part of the geostatistical model, as in this study. A different equation (not shown) applies if the mean or trend is pre-specified, as in the case of 'simple kriging.' Generally, the factors that influence the weights are:

1. The spatial covariance between the prediction location and observation locations, as determined by the covariance function (eq 3-5), such that observations that are closer (more strongly correlated) to the prediction location will receive more weight.
2. The spatial covariance among observation locations, as determined by the covariance function, such that clustered (i.e., non-independent) observations will receive less weight.
3. The covariate information ($\mathbf{X}$) for the prediction and observation locations, such that observation locations with covariate values more similar to the covariate values of the prediction location will receive more weight. (The use of covariate information for geostatistical interpolation is commonly referred to as 'universal kriging.' If no covariates are included in the model (mean only), then the interpolation is referred to as 'ordinary kriging' (Chiles and Delfiner 1999).)

Comparisons studies have demonstrated that geostatistical interpolation methods typically outperform simpler interpolation methods, such as inverse distance weighting (IDW) and Thiessen polygons (Goovaerts 2000, Murphy et al. 2010). This is not surprising, given that simpler methods apply a more subjective weighting system, while geostatistical methods determine weights through modeling the covariance structure of the data. Furthermore, simpler methods do not generally discount the weights of clustered observations (that provide somewhat redundant information, see item #2 in list above). The discounting of clustered data is also important when attempting to make statistical inferences about how various factors (represented by covariates) affect the response variable; and this type of statistical inference is a primary objective of Chapter 4.

Geostatisical interpolation is also advantageous because it allows for the determination of prediction uncertainty, most commonly assuming a Gaussian error process. Chapter 5 will make use of this feature of geostatistical modeling when determining confidence intervals for hypoxic extent estimates. This is accomplished through a simulation-based approach (i.e., 'conditional realizations') that essentially samples from the uncertainty in the interpolation results.

3.2 Bayesian modeling background

A common challenge in mechanistic environmental modeling is to determine the parameters that characterize the rates of various biological and physical processes (Beck and Straten 1983, Reckhow and Chapra 1999). Historically, these parameters have been determined either *a priori* based on literature review and best professional judgment, or through calibration of the model to observed conditions. While the latter approach may be preferable, it is often unrealistic when a large number of biophysical processes must be parameterized based on limited observational data. Reliance on *a priori* parameter estimation can also be problematic, however, as the range of parameter values found in literature is often quite wide, e.g., EPA (1985). Furthermore, as discussed in Section 2.3.1, there may be multiple model parameterizations that perform similarly in model calibration and validation, but which may have different scientific and management implications.

Bayesian inference helps to address the challenge of parameter estimation in two important ways. First, it provides a means of parameter estimation that systematically combines *a priori* information with information derived from model calibration (i.e., from the calibration dataset). Second, it provides a probabilistic framework for model parameterization, so that the resulting parameter estimates can be represented as a combined joint distribution that accounts for the covariance among parameters, from which probabilistic model predictions can then be made. Bayesian inference is sometimes referred to as Bayesian learning, because it represents how our initial understanding of a system is updated after considering the available observational data (Lunn et al. 2013).

The potential benefits of applying Bayesian inference to water quality models have been recognized for decades (Beck and Straten 1983). However, only recently have computational methods and hardware improved sufficiently to make the implementation of Bayesian methods attractive for water quality modeling (Reckhow and Chapra 1999). Even now, Bayesian methods are typically only applied to relatively simple water quality models, such as the one-dimensional steady-state Gulf hypoxia model originally developed by Scavia et al. (Liu et al. 2010, Scavia et al. 2003). Other examples of mechanistic water quality models developed in a Bayesian context include river dissolved oxygen models (Dilks et al. 1992, Liu et al. 2008), a bacterial decay model (Gronewold et al. 2009), and eutrophication models (Arhonditsis et al. 2007, Malve et al. 2007, Ramin et al. 2011).

Bayesian inference of model parameters ($\boldsymbol{\theta}$) from observational data ($\boldsymbol{y}$) can be represented as follows (Hoff 2009):

$$p(\boldsymbol{\theta}|\boldsymbol{y}) = \frac{p(\boldsymbol{y}|\boldsymbol{\theta})p(\boldsymbol{\theta})}{\int_{\boldsymbol{\Theta}} p(\boldsymbol{y}|\boldsymbol{\theta})p(\boldsymbol{\theta})d\boldsymbol{\theta}} \qquad \text{eq 3-7}$$

where $p(\boldsymbol{\theta})$ is the prior distribution for the parameters, as determined from *a priori* information; $p(\boldsymbol{y}|\boldsymbol{\theta})$ is the probability of $\boldsymbol{y}$, conditional on the model parameters; and $p(\boldsymbol{\theta}|\boldsymbol{y})$ is the posterior distribution of the parameters, conditional on the observational data (note that this is a joint posterior distribution and that there may be correlation among model parameters). The denominator of eq 3-7 is a normalizing constant, determined by

integrating over parameter space, Θ. However, it is usually not necessary to calculate this constant directly, and so eq 3-7 can be simplified as follows:

$$p(\boldsymbol{\theta}|\boldsymbol{y}) \propto p(\boldsymbol{y}|\boldsymbol{\theta})p(\boldsymbol{\theta}) \qquad \text{eq 3-8}$$

Further, $p(\boldsymbol{y}|\boldsymbol{\theta})$ is equivalent to the likelihood of the parameters given the observational data . And thus, eq 3-8 can be more intuitively written, with all parts expressed in terms of $\boldsymbol{\theta}$, as (Lunn et al. 2013):

$$p(\boldsymbol{\theta}|\boldsymbol{y}) \quad \propto \quad L(\boldsymbol{\theta};\boldsymbol{y})\, p(\boldsymbol{\theta}) \qquad \text{eq 3-9}$$

$$\textit{posterior} \propto \textit{likelihood} \times \textit{prior}$$

Once the joint posterior distribution of the parameters is known, making posterior model predictions, $p(\tilde{y}|\boldsymbol{y})$, is relatively straightforward:

$$p(\tilde{y}|\boldsymbol{y}) = \int_{\boldsymbol{\Theta}} p(\tilde{y}|\boldsymbol{\theta})p(\boldsymbol{\theta}|\boldsymbol{y})d\boldsymbol{\theta} \qquad \text{eq 3-10}$$

However, in all but the simplest cases, it is not possible to determine the properties of the posterior parameter distribution analytically. Instead, Markov Chain Monte Carlo (MCMC) sampling methods are commonly used to explore the posterior distribution (Lunn et al. 2013), though other types of sampling algorithms also exist. MCMC methods are iterative approaches where samples of model parameters are drawn from 'transition distributions' that are iteratively refined until the samples converge to the joint posterior parameter distribution (i.e., the 'stationary distribution'). The steps leading to convergence are commonly referred to as the 'burn-in' and are discarded. Once the MCMC process has converged, the following samples can be used to characterize the posterior distribution (Lunn et al. 2013).

Multiple MCMC sampling algorithms are available, and the effectiveness of these algorithms depends largely on the model structure. In this research, MCMC sampling is implemented using the WinBUGS program, which automatically determines the most

effective sampling algorithm (Lunn et al. 2000, Lunn et al. 2013). BUGS stands for Bayesian inference Using Gibbs Sampling, where Gibbs sampling is one of the most commonly used MCMC algorithms (Gelfand and Smith 1990). However, WinBUGS will make use of other types of sampling algorithms (e.g., Metropolis-within-Gibbs) when Gibbs sampling is infeasible.

Finally, it should be noted that geostatistical modeling and Bayesian modeling are not mutually exclusive realms. In fact, Bayesian methods are becoming increasingly common in spatial modeling (see Gelfand et al. (2010) and references therein). However, in this research, geostatistical modeling is performed using a non-Bayesian approach largely for reasons of computational efficiency. The Bayesian approach is more feasible for the mechanistic model described in Chapter 6, because the mechanistic model operates on a much smaller calibration dataset, so that MCMC can be performed relatively quickly. Moreover, the mechanistic model is nonlinear and uses *a priori* information, such that the Bayesian approach is highly advantageous, if not absolutely necessary.

CHAPTER 4: Geostatistical Regression

Stratification and nutrient loading are two primary factors leading to hypoxia in coastal systems. However, where these factors are temporally correlated, it can be difficult to isolate and quantify their individual impacts. This study provides a novel solution to this problem by determining the effect of stratification based on its spatial relationship with bottom-water dissolved oxygen (BWDO) concentration using a geostatistical regression. The text of this chapter is reproduced in part with permission from Obenour et al. (2012b), Copyright 2012 American Chemical Society.

4.1 Introduction

The relationship between nutrient loading and the interannual (i.e. year-to-year) variability of the mid-summer Gulf hypoxic area has been explored through a range of computational models. Scavia et al. (2003) use a one-dimensional, simple mechanistic model that explains 45% of the interannual variability as a function of the May-June nitrogen load. These results are generally consistent with the statistical regressions of Bianchi et al. (2010) who find that the May-June nitrogen load explains 47% of the variability, and of Greene et al. (2009) who find that the May $NO_{2\text{-}3}$ load explains 42% of this variability. In addition, studies by Turner et al. (2006, 2008) and others (Greene et al. 2009, Liu et al. 2010) have shown that model performance is improved by also accounting for the long-term cumulative effects of nutrient loading.

It is widely understood, however, that water column stratification also affects the temporal variability of hypoxia. Stratification, which inhibits the re-oxygenation of bottom waters, can be particularly strong in the northern Gulf due to warmer surface waters significantly freshened by discharges from the Mississippi and Atchafalaya Rivers overlying denser waters derived from the deep shelf. Wiseman et al. (1997), while acknowledging the importance of nutrient loading, determined that the majority of the

interannual variability in the hypoxic area (for a nine-year period) could be explained using the mean river flow from the preceding eleven months. Using a more expansive dataset, Bianchi et al. (2010) demonstrate that 41% of the interannual variability in hypoxic area can be explained in terms of the May-June river flow.

The fact that both nutrients and flow can explain a large portion of the variability in hypoxic extent is not surprising. Bianchi et al. (2010) report that these variables are highly correlated, with May-June flow accounting for 95% of the variability in May-June nitrogen load. Therefore, use of either of these variables masks the mechanistic effect of the other, leaving room for debate regarding whether eutrophication (via nutrient loading) or stratification (via freshwater flow) is the primary control on the interannual variability of hypoxia. Accordingly, Hetland and DiMarco (2008) suggest that it will be necessary to separate the physical and biological causes of hypoxia to develop models with greater predictive capability; and a recent scientific assessment (EPA 2007) stresses the need to include physical factors within hypoxia models. Moreover, a recent study by Murphy et al. (2011) demonstrates the importance of both nutrients and stratification for predicting hypoxia in the Chesapeake Bay. This need likely extends to other systems, as Diaz and Rosenberg (2008) report that there are over 400 coastal hypoxic zones around the world, and that they are controlled by both biological and physical factors (generally eutrophication and stratification, respectively). Because stratification is primarily a natural phenomenon, while nutrient loading is often an anthropogenic watershed pollution issue, the ability to distinguish between these factors is critical for managing these complex systems.

In this study, the effect of stratification on the variability of hypoxia is determined through geostatistical regression (GR) (Mueller et al. 2010, Zimmerman and Stein 2010). In the model, stratification is represented by metrics derived from salinity and temperature profiles, with regression coefficients determined based on how stratification explains the spatial variability in BWDO. While a portion of the interannual variability in BWDO is also explained by the stratification effect, the remaining portion must be explained by other factors, such as Mississippi River nutrient inputs. The methodology is evaluated using ten recent years of mid-summer cruise data (1998-2007). As such, this study generally focuses

on the interannual variability of the system within its current state; and it does not focus on long-term trends in the system.

4.2 Materials and methods

4.2.1 Data description

We use data from mid-summer hypoxia monitoring cruises conducted between 1998 and 2007 along the Louisiana-Texas shelf. These cruises are performed by the Louisiana Universities Marine Consortium (LUMCON), and the data were retrieved from the National Ocean Data Center (Rabalais 2011). To help ensure a consistent spatial envelope for this study, only locations sampled during at least nine of the ten years were included in the analysis (Figure 2-2), resulting in 61-64 monitoring locations for each year. Sampling locations were geo-referenced using the UTM Zone15 projection, and water depths were determined from a 3-arc-second digital elevation model (DEM) obtained from NOAA (NOAA 2011b). We use the dissolved oxygen, salinity, and temperature profile data collected at these locations. Typically, data were collected by two different instruments at each monitoring site: a Hydrolab and a Sea-Bird profiler (Rabalais et al. 1996). Based on a comparison with the DEM, the Hydrolab and Sea-Bird typically reached to within zero and one meters of the sea floor, respectively, with minor variability. For this study, the BWDO values were taken from whichever instrument reached the greatest recorded depth (typically the Hydrolab). To determine salinity and temperature profiles, Sea-Bird data, which have better vertical resolution, were chosen preferentially over Hydrolab data; and when the Hydrolab reached a greater depth, we appended these additional measurements to the Sea-Bird profile.

The stratification intensity at each sampling location was quantified using metrics derived from the salinity and temperature profiles. Because sampling intervals in the raw profile data were not uniform, the profiles were first resampled by linear interpolation to 0.2-meter resolution. The difference (ΔS) between the 25^{th} and 100^{th} percentiles of salinity (S_{lo} and S_{hi}, respectively) was calculated for each location, and then the maximum salinity gradient (S_g) over a 0.4-meter vertical interval, within the region of ΔS, was determined. The thickness (m) of the subhalocline (H_s), defined as the region below S_g,

was also calculated. Figure 4-1 provides an illustration of these metrics. Use of the 25th and 100th percentiles for defining S_{lo} and S_{hi} was not arbitrary; a range of percentile values was considered, and the values that optimized the GR (based on BIC score, described subsequently in Section 4.2.4) were selected. For temperature profiles, an identical analysis was performed, and the 0th and 75th percentiles were found to be optimal for defining T_{lo}, T_{hi}, and ΔT. The maximum temperature gradient and sub-thermocline thickness, T_g and H_t, respectively, were also calculated. Because temperature tends to decrease with depth, while salinity increases, the 25th percentile of salinity and the 75th percentile of temperature both generally correspond to the 25th percentile of depth, suggesting that near-surface stratification conditions are less important for predicting BWDO.

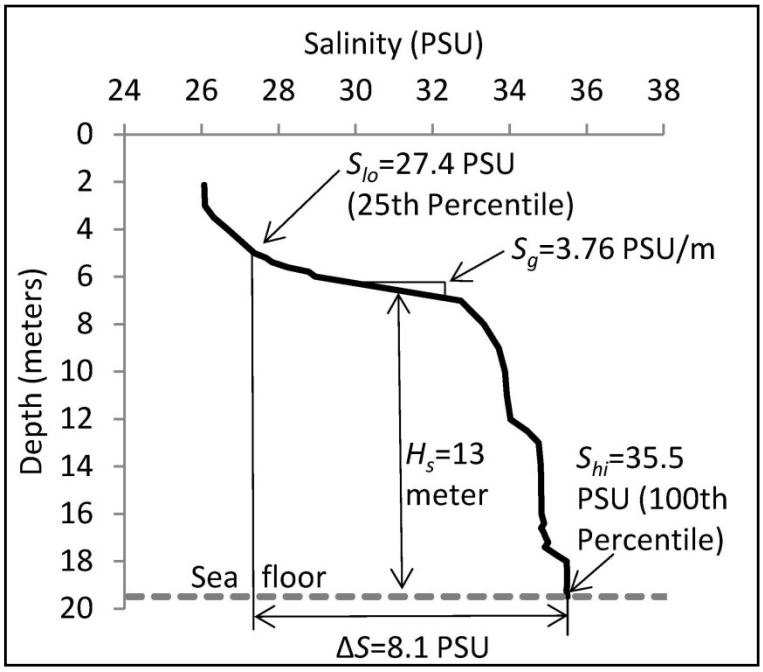

Figure 4-1: Example salinity profile with stratification metrics

The Mississippi and Atchafalaya River outfalls (Figure 2-2) provide the vast majority of fresh water and nutrients to the shelf. The Atchafalaya River is important because approximately 30% of the Mississippi River flow (and load) enters the Atchafalaya River near Simmesport, Louisiana (Scavia et al. 2003). Monthly flow and loading data were

retrieved from USGS (2010), and we used the sum of the two river inputs. The USGS uses both an Adjusted Maximum Likelihood Estimator (AMLE) and a composite (COMP) method to estimate loads (Greene et al. 2009). We used AMLE data because they generally performed better in the modeling analyses. However, the two datasets are highly correlated, and the choice does not greatly affect the results.

4.2.2 Model formulation

Our geostatistical modeling approach takes into account the spatial correlation of the dependent variable (i.e. BWDO), which improves estimates of the model parameters relative to a regression based on the assumption of independent and identically-distributed residuals, and allows for a more realistic assessment of model uncertainty (Zimmerman and Stein 2010). The effectiveness of GR methods for modeling environmental phenomena has been demonstrated in studies of rainfall (Goovaerts 2000), snow depth (Erickson et al. 2005), water quality (Murphy et al. 2010), and biospheric CO_2 exchange (Mueller et al. 2010). With the exception of the last example, these works focused primarily on spatial interpolation, and less on inference of causal factors. As such, this work adds to the limited number of studies demonstrating how geostatistical modeling can be useful for confirming and quantifying causal relationships within environmental systems.

As described in Section 3.1.1 (eq 3-1), a geostatistical model represents a dependent variable in terms of its deterministic and stochastic components ($\mathbf{z} = \mathbf{X\beta} + \boldsymbol{\eta} + \boldsymbol{\varepsilon}$). In this study, the dependent variable, $\mathbf{z}$, is an $n\times1$ vector of BWDO measurements taken at different locations and times.

The deterministic component, $\mathbf{X\beta}$, is the portion of $\mathbf{z}$ that can be expressed as a function of predictor variables and annual constants (i.e. intercepts). The $k\times1$ vector $\boldsymbol{\beta}$ includes the parameters of the deterministic component, which can be divided into annual intercepts ($\boldsymbol{\beta}_a$) and regression coefficients ($\boldsymbol{\beta}_p$) (eq 4-1). Correspondingly, the $n\times k$ matrix $\mathbf{X}$ includes the annual classifiers ($\mathbf{X}_a$) and predictor variables ($\mathbf{X}_p$) for each observation. The annual classifiers are binary values that bin samples by year. Each predictor variable is normalized to a mean of zero and variance of one over the ten-year period.

$$\mathbf{X\beta} = \left[\mathbf{X}_a \mathbf{X}_p\right] \begin{bmatrix} \boldsymbol{\beta}_a \\ \boldsymbol{\beta}_p \end{bmatrix} \quad \text{eq 4-1}$$

A portion of the variability in BWDO is accounted for through stratification metrics (which are included as predictor variables). The regression coefficients for these metrics are fit to the spatial variability in BWDO, based on the assumption that sites which are strongly stratified will have lower BWDO due to lower rates of re-oxygenation. They are not fit to the interannual variability in BWDO because this could confound the role of stratification with nutrient loading, as they are temporally related, both being dependent on river flow. This approach does not, however, preclude stratification from explaining a portion of the interannual variability. As the average intensity of stratification varies from year to year, the predicted impact of stratification on BWDO varies proportionally, as a function of the regression coefficients. This is reasonable because years of intense stratification would be expected to have less re-oxygenation and thus lower BWDO. The remaining interannual variability is accounted for primarily through the annual intercepts, discussed further in Section 4.2.6.

The stochastic components, $\boldsymbol{\eta}$ amd $\boldsymbol{\varepsilon}$, were found to be well represented by the commonly used exponential covariance model with a nugget effect, as described in Section 3.1.1 (eq 3-5). However, spatial correlation is expected to be stronger in the east-west direction than in the north-south direction due to along-shore currents (Zavala-Hidalgo et al. 2003). To account for this phenomenon, $h_{i,j}$ is scaled by an anisotropy ratio, α, which is the ratio of east-west to north-south correlation ranges.

4.2.3 Parameter estimation

Restricted Maximum Likelihood (REML) (Patterson and Thompson 1971) is used to estimate the covariance model parameters and the anisotropy ratio. This method is recommended for models with strong spatial correlation and/or a large numbers of predictor variables (Zimmerman 2010), and it has been applied in previous geostatistical studies (Kitanidis and Shen 1996, Mueller et al. 2010). The parameters are optimized by

minimizing L_r (eq 4) with respect to σ_ε^2, σ_η^2, r and α, which define **Q** (an $n \times n$ covariance matrix with elements determined from eq 3-5). Because the covariance between the stochastic components of the observations collected in different years is assumed to be zero, **Q** becomes block diagonal.

$$L_r = \ln(\det[\mathbf{Q}]) + \ln(\det[\mathbf{X}^T\mathbf{Q}^{-1}\mathbf{X}]) + \mathbf{z}^T\mathbf{Q}^{-1}(\mathbf{1} - \mathbf{X}(\mathbf{X}^T\mathbf{Q}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Q}^{-1})\mathbf{z} \quad \text{eq 4-2}$$

Once **Q** is optimized, the deterministic model parameters are calculated in a straightforward manner:

$$\widehat{\boldsymbol{\beta}} = (\mathbf{X}^T\mathbf{Q}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Q}^{-1}\mathbf{z} \quad \text{eq 4-3}$$

Thus, a complete set of model parameters is uniquely defined, and the covariance of the model parameters is:

$$\mathbf{V}_\beta = (\mathbf{X}^T\mathrm{Q}^{-1}\mathbf{X})^{-1} \quad \text{eq 4-4}$$

4.2.4 Variable selection

The GR can be formulated for any subset of the available predictor variables. Candidate variables include the salinity stratification metrics ($\ln[\Delta S]$, S_{hi}, S_g, and H_s) and temperature stratification metrics ($\ln[\Delta T]$, T_{lo}, T_g, and H_t). Note that S_{lo} and T_{hi} were omitted to avoid issues of colinearity; and the salinity and temperature differences are log-transformed to account for an apparent non-linearity in their relationships with BWDO. Variables for water column depth, UTM easting, and UTM northing (D, E, N) were also included, along with their squares (D^2, E^2, N^2) to allow for potential spatial and bathymetric trends. Overall, there are fourteen candidate predictor variables, yielding over 16,000 combinations of one to fourteen variables to be considered. The challenge is to determine an optimal combination of predictor variables that best describes the phenomenon of interest, without resulting in over-parameterization of the model (Mueller

et al. 2010). In general, this is achieved by including only those variables that have a sufficiently high degree of explanatory ability, as defined by a statistical criterion. In this study, the Bayesian Information Criterion (BIC) (Schwarz 1978) is used, where models with different sets of predictor variables ($\mathbf{X}_s$) are compared based on their BIC scores:

$$BIC_s = -2 \ln(L_s) + k_s \ln(n) \qquad \text{eq 4-5}$$

In eq 4-5, the first term is twice the negative log likelihood function (L) of the estimated model. This term represents the goodness-of-fit of the model, with lower numbers indicating a better fit. The second term is a penalty term which addresses the loss in degrees of freedom that occurs as more predictor variables are added to the model. If the spatially correlated model residuals are assumed to follow a Gaussian distribution, then the likelihood of the model can be determined as:

$$L_j(\mathbf{X}_j, \hat{\boldsymbol{\beta}}_j | \mathbf{y}) = \frac{1}{(2\pi)^{n/2} |\mathbf{Q}|^{1/2}} \exp\left[-\frac{1}{2} (\mathbf{y} - \mathbf{X}_j \hat{\boldsymbol{\beta}}_j)^T \mathbf{Q}^{-1} (\mathbf{y} - \mathbf{X}_j \hat{\boldsymbol{\beta}}_j) \right] \qquad \text{eq 4-6}$$

Eq 4-3 can be substituted into eq 4-6, and eq 4-6 can then be substituted into eq 4-5. After simplifying and removing the constant term, the equation for the BIC score becomes (Mueller et al. 2010):

$$BIC_s = \ln(\det[\mathbf{Q}]) + \mathbf{z}^T \mathbf{Q}^{-1} \left(1 - \mathbf{X}_s (\mathbf{X}_s^T \mathbf{Q}^{-1} \mathbf{X}_s)^{-1} \mathbf{X}_s^T \mathbf{Q}^{-1}\right) \mathbf{z} + k_s \ln(n) \qquad \text{eq 4-7}$$

The preceding is a 'geostatistical' version of the BIC. Note that if there is no correlation among residuals (i.e., $\boldsymbol{\eta}$ is zero), then the covariance matrix, $\mathbf{Q}$, can be simplified to $\sigma_r^2 \mathbf{I}$, where $\mathbf{I}$ is an nxn identity matrix. Here, σ_r^2 is the variance of the residuals, which is equal to $\mathbf{y}^T(\mathbf{I}-\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T)\mathbf{y}/n$. In this case, eq 4-7 can be simplified to the more common form:

$$BIC = n \ln(\sigma_r^2) + k_j \ln(n) \qquad \text{eq 4-8}$$

The BIC score quantifies the explanatory power of a model relative to its complexity, with the lowest score indicating the model with optimal balance. The GR model should be optimized in terms of both its covariance parameters (Section 4.2.3) and its selection of predictor variables. Because the covariance parameters of the stochastic component depend on the selected predictor variables and vice versa, an iterative approach is used.

4.2.5 Annual mean values

The interannual variability of mid-summer hypoxic intensity can be assessed by comparing the mean BWDO concentrations between years. Annual mean BWDO values were determined using geostatistical kriging of the mean (Wackernagel 2003), which assigns lower weights to observations from clustered sampling locations. The mean BWDO for year *i* is calculated as:

$$\overline{BWDO}_i = \boldsymbol{\lambda}_i^T \mathbf{z}_i \qquad \text{eq 4-9}$$

with weights, $\boldsymbol{\lambda}_i$, corresponding to observations, $\mathbf{z}_i$. The weights are determined from the system of linear equations presented in eq 4-10, where $\mathbf{Q}_i$ is the $n_i \times n_i$ covariance matrix for the observations from year *i*, with elements determined from eq 3-5. The scalar, v_i, is a Lagrange multiplier, its square root being the standard error of the estimated mean value.

$$\begin{bmatrix} \boldsymbol{\lambda}_i \\ \nu_i \end{bmatrix} = \begin{bmatrix} \mathbf{Q}_i & \mathbf{1}_{n_i x 1} \\ \mathbf{1}_{n_i x 1}^T & 0 \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{0}_{n_i x 1} \\ 1 \end{bmatrix} \qquad \text{eq 4-10}$$

The annual mean impacts of the predictor variables (on BWDO) are also determined geostatistically. For a given year *i*, the mean impact of each predictor variable is calculated by applying the vector of location-specific weights, $\boldsymbol{\lambda}_i$, to the vector of location-specific impacts, $\mathbf{x}_{i,j}\boldsymbol{\beta}_j$, where *j* refers to the j^{th} predictor variable.

4.2.6 Model for the annual intercepts

Linear regression models were developed to explore potential relationships between the annual intercepts and nutrient and flow data. The response variable is the set of ten annual intercepts, $\widehat{\boldsymbol{\beta}}_a$, determined from the GR. Candidate predictor variables include the April, May, and June monthly loads and concentrations for $NO_{2\text{-}3}$, total kjeldahl nitrogen (TKN), and total phosphorus (TP), as well as monthly flows (Q). These concentrations are flow-weighted averages, calculated by dividing monthly load by monthly flow. To account for potential changes in the system's susceptibility to hypoxia, a linear trend across years is also included as a candidate variable.

Because of the large number of candidate variables (p=22), many of which are correlated, and relatively small sample size (n=10), the 'elastic-net' method (Zou and Hastie 2005) was used to identify important predictor variables. This method constrains the magnitude of the model parameters (and shrinks some parameters to zero, thereby eliminating them) to prevent problems commonly associated with the over-fitting of models (Faraway 2005). This method is a compromise between 'ridge regression' and the 'Lasso' method (Tibshirani 1996). Ridge regression improves the predictive capacity of the model through parameter shrinkage (by limiting the parameter sum of squares, $\boldsymbol{\beta}^T\boldsymbol{\beta}$), but does not eliminate any predictor variables from the model entirely (Faraway 2005). The Lasso method performs both parameter shrinkage and variable selection (by limiting the parameter sum of absolute values $||\boldsymbol{\beta}||$). Elastic net produces results similar to Lasso, but is more tolerant of the inclusion of multiple correlated predictor variables (Zou and Hastie 2005). Because concentrations and loads from different months are likely correlated, this method was found to be most appropriate.

The elastic-net method was implemented using the 'glmnet' package for the R statistical computing program (Friedman et al. 2010). The degree to which model parameters are constrained is optimized through a leave-one-out cross validation routine included in this package. We used the elastic-net as an exploratory tool and to select one or more primary variables from which to create a more parsimonious regression model. An alpha value, α, supplied by the user determines whether the results are more similar to

ridge regression or Lasso. A value of zero indicates a ridge regression model, whereas a value of one indicates a Lasso model. In this application, an alpha of 0.5 is used, indicating no particular preference for ridge regression or Lasso (i.e., an elastic-net model).

4.2.7 Hypoxic area model and nutrient reduction scenarios

The information developed above can be used to predict how the intensity of mid-summer hypoxia will respond (within the system's current state) to changes in nutrient loading. However, while this study focuses on BWDO concentration, policy goals have centered on the areal extent of hypoxia (EPA 2001, 2008). Thus, a regression model for mid-summer hypoxic area (LUMCON 2012, Rabalais et al. 2007) using the annual mean stratification effects (Section 4.2.5) and one or more nutrient variables (Section 4.2.6) is developed. Using this model, the effects of hypothetical nutrient reduction scenarios are explored.

The goal of the Intergovernmental Task Force Action Plan (EPA 2001, 2008) is to reduce the five-year running average of hypoxic area to less than 5,000 km^2. Thus, modeling results are developed for all possible consecutive five-year periods across the ten-year study period. The consecutive five-year period which requires the greatest nutrient reduction to achieve the Action Plan goal is defined as the 'critical' period.

Because of the non-negativity constraint on hypoxic area (any predictions of negative hypoxic area are treated as zero), probabilistic results are best determined through Monte Carlo simulation (eq 4-11) (Rencher 1998). Here, $\tilde{\mathbf{y}}_{\mathbf{sim}}$ is a 5×1 vector of simulated hypoxic areas for a given five-year period, $\tilde{\mathbf{y}}$ is the predicted (i.e. mean) response, $\mathbf{u}$ is a vector of independent samples from a standard normal distribution, and $\mathbf{C}(\widetilde{\boldsymbol{\Sigma}})$ is the upper-triangular matrix obtained by Cholesky decomposition of the covariance matrix for the predictions (eq 4-12). In eq 10, $\mathbf{W}$ is the matrix of predictor variables used to develop the regression (corresponding to the observed conditions), $\widetilde{\mathbf{W}}$ is a matrix of predictors with reduced nutrient levels, σ_r^2 is the variance of the model residuals, and $\mathbf{I}$ is the identity matrix.

$$\tilde{\mathbf{y}}_{\mathbf{sim}} = \tilde{\mathbf{y}} + \mathbf{C}(\widetilde{\boldsymbol{\Sigma}})^{\mathrm{T}}\mathbf{u} \qquad \text{eq 4-11}$$

$$\tilde{\boldsymbol{\Sigma}} = \left[\tilde{\mathbf{W}}(\mathbf{W}^{\mathrm{T}}\mathbf{W})^{-1}\tilde{\mathbf{W}}^{\mathrm{T}} + \mathbf{I}\right]\sigma_r^2 \qquad \text{eq 4-12}$$

For each nutrient reduction scenario, 10,000 simulations of $\tilde{\mathbf{y}}_{\mathbf{sim}}$ were generated in accordance with eq 4-11, using different random draws of the vector **u**. Any simulations of negative hypoxic area were replaced with a value of zero, and the mean of each $\tilde{\mathbf{y}}_{\mathbf{sim}}$ was then determined, resulting in 10,000 simulations of five-year average hypoxic area. Summary statistics for each nutrient reduction scenario were then calculated based on this ensemble of results.

4.3 Results and discussion

4.3.1 Model parameters

The deterministic component of the GR (eq 2) was optimized in terms of the regression coefficients and annual intercepts. The regression coefficients ($\widehat{\boldsymbol{\beta}}_p$) indicate how BWDO is related to the various predictor variables; all parameters are significantly different from zero (p<0.05, Table 4-1). Because the predictor variables were normalized, the magnitudes of the parameter values demonstrate their relative impacts on BWDO. The regression coefficients for northing (N) and depth (D) suggest that BWDO concentrations tend to be lower at locations that are shallower and farther north (i.e. closer to shore). This is likely because the near-shore waters mix less with the more oxygenated and less eutrophic deep-shelf waters. As expected, the regression coefficients for the stratification metrics all indicate BWDO tends to be lower in areas of intense water column stratification. Annual intercepts ($\widehat{\boldsymbol{\beta}}_a$) ranged from 1.66 to 3.50 mg L^{-1}, and are discussed further in Section 3.2. For the ten-year study period, the deterministic component of the model ($\mathbf{X}\widehat{\boldsymbol{\beta}}$) explains 52% of the variability in BWDO across space and time. For individual years, the deterministic component of the model explains from 27% to 61% of the spatial variability in BWDO.

The remaining spatial variability may be due largely to the effects of varying coastal current patterns that can influence transport of fresh water, nutrients, and organic matter.

The stochastic portion of the model, which accounts for this remaining variability, has substantial spatial correlation, consistent with current patterns acting across large spatial scales. The parameters σ_ε^2 and σ_η^2 (often referred to as the nugget and the partial sill in geostatistics literature) were determined to be 0.53 and 1.29 mg^2L^{-2}, respectively. The range parameter, *r*, was determined to be 30.5 km along the east-west direction, which means that the effective range of spatial correlation is 91.5 km in that direction. The anisotropy ratio was determined to be 1.56, which means that the effective range of spatial correlation in the north-south direction is 58.7 km. These effective ranges are considerably longer than the typical spacing between sampling locations (Figure 2-2). Overall, these results verify that there is substantial spatial correlation in the stochastic portion of the model.

Table 4-1: Regression coefficients $(\hat{\beta}_p)$ with standard errors $(\sigma_{\hat{\beta}})$ for GRa.

Variable	$\widehat{\boldsymbol{\beta}}_{\boldsymbol{p}}$	$\boldsymbol{\sigma}_{\widehat{\boldsymbol{\beta}}}$
northing, N	-0.77	0.11
depth, D	0.83	0.12
salinity diff., ln(ΔS)	-0.43	0.11
max. salinity, S_{hi}	-0.45	0.10
salinity gradient, S_g	-0.21	0.07
temp. diff. ln(ΔT)	-0.62	0.13
min. temp. T_{lo}	0.69	0.15

[a]units are mg L^{-1}.

4.3.2 Test of linearity

The geostatistical regression (GR) model assumes that BWDO concentrations can be modeled using linear relationships with the examined predictor variables (or transformations of the predictor variables, e.g. ln[ΔS]). Figure 4-2 provides a visual test of this assumption by plotting the BWDO residuals versus each of the selected predictor variables. These residuals are referred to as 'e*' because while a normal residual is

calculated by subtracting the observed value from the deterministic portion of the model ($e = X\beta - y$), these residuals are calculated by also removing the effect of the variable of interest ($e = X\beta - x_j\beta_j - y$). In this way, the relationship between the variable of interest, x_j, and BWDO is clearly illustrated, with the slope of the trend line approximately equal to the regression coefficient, β_j. In general, these plots suggest that the assumption of linearity between the predictor variables and BWDO is reasonable.

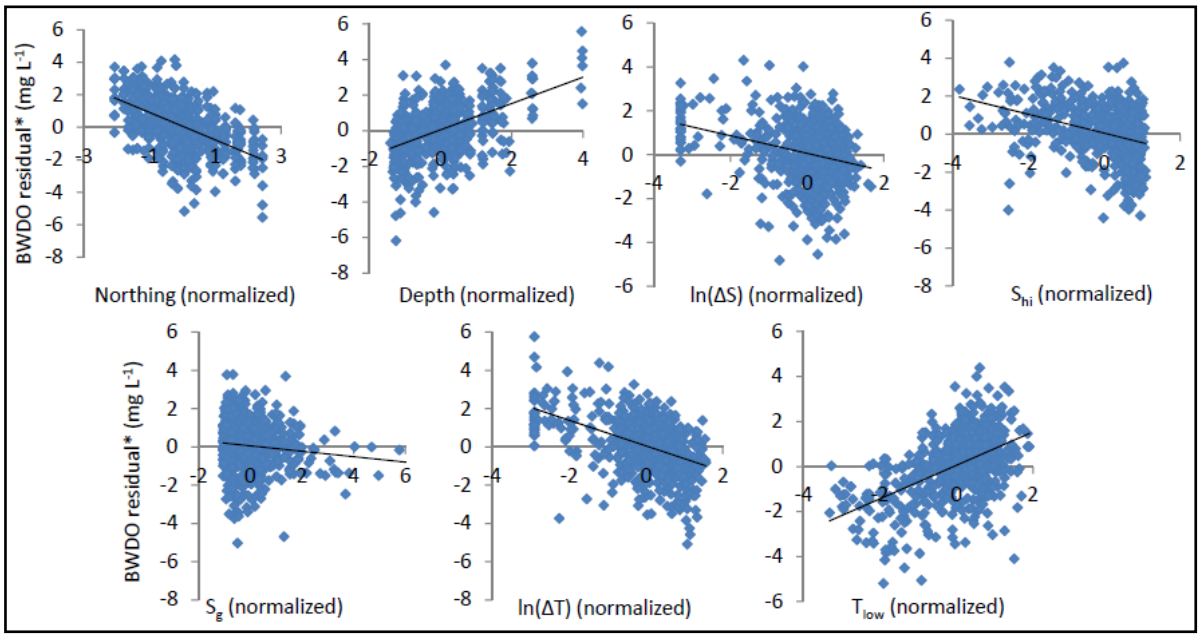

Figure 4-2: BWDO residuals (e*) versus normalized predictor variables (Note: $e^* = X\beta - x_j\beta_j - y$ where x_j is the variable represented on the horizontal axis of each graph.)

4.3.3 Site-specific model results

Model predictions, based on the deterministic component (eq 4-1) can be compared with observed values at each sampling location (Figure 4-3). The deterministic model component appears to capture much of the spatial variability in BWDO concentration. For example, it captures the above-average BWDO concentrations on the western shelf in 1998 and 2000.

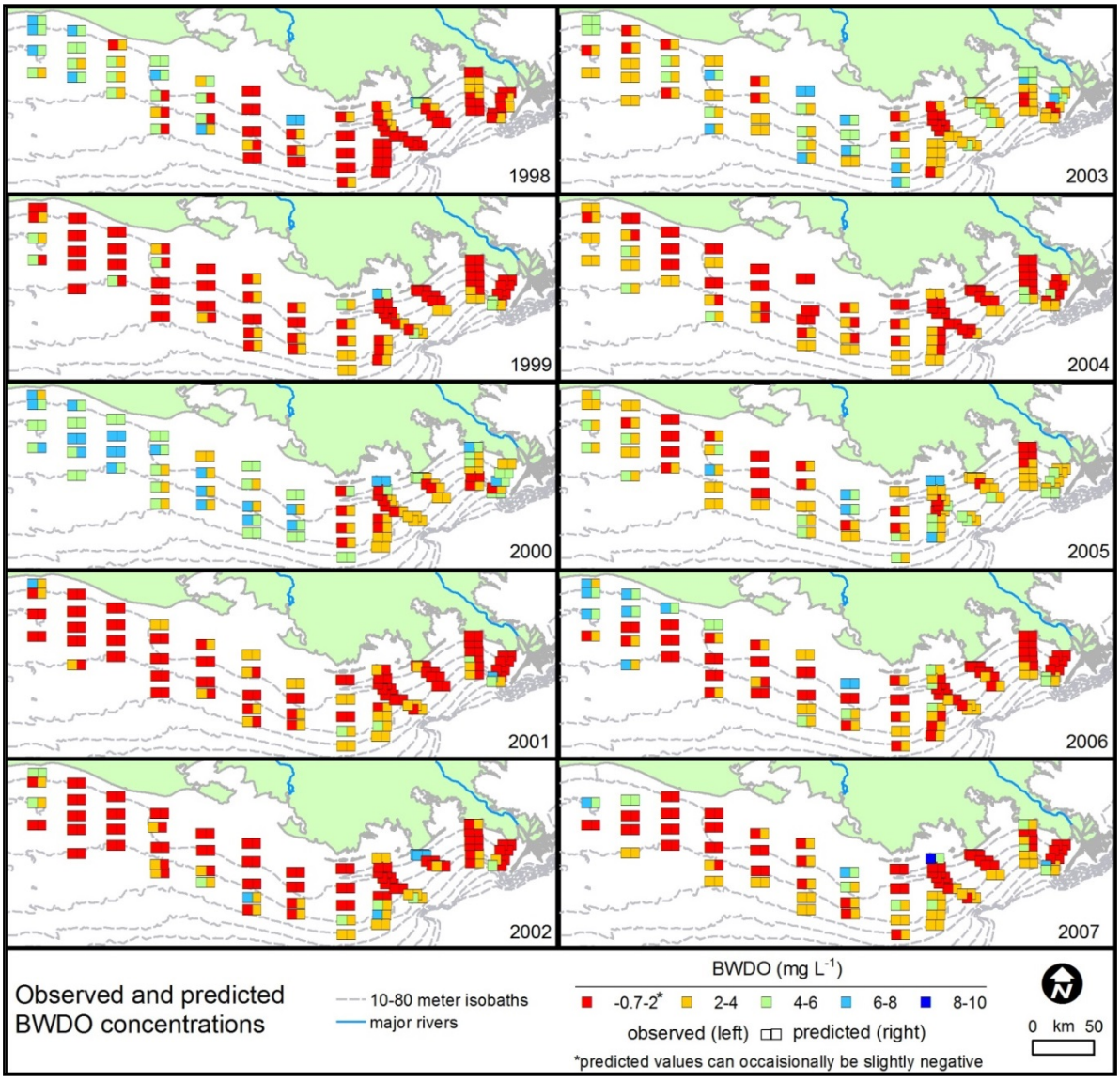

Figure 4-3: Observed and model-predicted BWDO concentrations for ten-year study period. Predicted values are from deterministic component of GR model (eq 4-2).

4.3.4 Analysis of interannual variability

The results of the GR can be summarized on an annually-averaged basis to analyze the interannual variability of the factors affecting mean BWDO concentration. Figure 4-4 illustrates how mean BWDO is affected by northing and bathymetry; salinity stratification ($\ln(\Delta S)$, S_{hi}, and S_g); temperature stratification ($\ln(\Delta T)$ and T_{lo}); and annual intercepts. The combination of these effects can be used to exactly predict the estimated mean BWDO (eq

4-9). For illustrative purposes, however, each factor is presented as a BWDO depletion, calculated by determining each effect relative to its year of minimum impact; for example, temperature stratification was least severe in 2006 and most severe in 2004. For each factor, the standard deviation of the ten annual impacts is calculated, providing a metric for assessing the degree to which each factor contributes to the interannual variability of BWDO, with higher standard deviations indicating greater contributions. From the figure, it is clear that stratification has a substantial impact on the interannual variability in BWDO depletion, and the standard deviation of the net stratification effect (salinity plus temperature) is 0.51 mg L^{-1} BWDO. The standard deviation of the northing/bathymetry effect is only 0.05 mg L^{-1} BWDO, which is expected because there is little interannual variation in the sampling locations used in this study. The greatest portion of the interannual variability is accounted for through the annual intercepts, with a standard deviation of 0.71 mg L^{-1} BWDO. While these intercepts do not have intrinsic explanatory value, they can be modeled in terms of other factors, such as nutrient inputs.

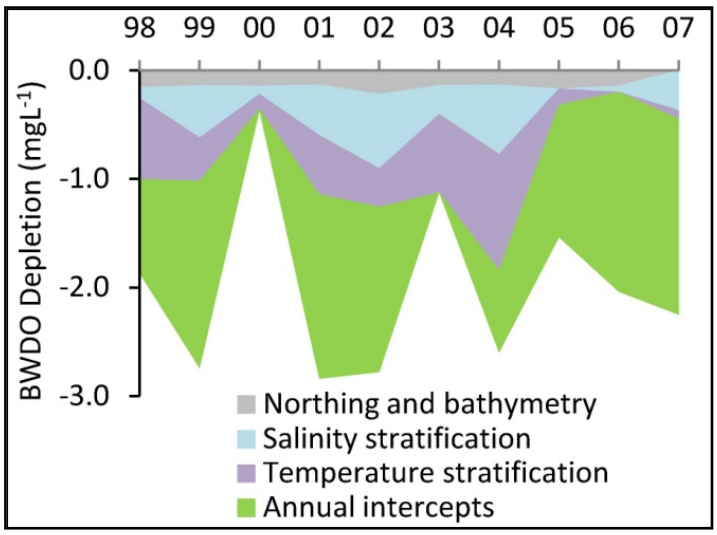

Figure 4-4: BWDO depletions attributed to different factors from GR model; each factor presented relative to its year of minimum impact.

The annual intercepts ($\widehat{\boldsymbol{\beta}}_a$) were first related to candidate predictor variables through development of an elastic-net model (Section 4.2.6). The resulting model (eq 4-13) includes six selected variables. Here, variables enclosed in brackets represent nutrient

concentrations, while un-bracketed variables represent loads, and subscripts represent months. Because variables were normalized, the regression coefficients indicate their relative importance.

$$\hat{\beta}_a = 2.36 - 0.348 * [NO_{May}] - 0.138 * NO_{2-3,May} - 0.107 * Year \ldots$$
$$- 0.073 * [NO_{Jun}] - 0.062 * TKN_{Apr} - 0.027 * NO_{Apr} \quad \text{eq 4-13}$$

The elastic-net model indicates May NO_{2-3} concentration has the largest impact, followed by May NO_{2-3} load, which is discussed more below. The trend with time (*Year*) also has a relatively large effect, suggesting that hypoxia is becoming more intense irrespective of seasonal nutrient inputs and stratification. This is consistent with previous studies indicating increasing hypoxia irrespective of seasonal nutrient loading (Greene et al. 2009, Liu et al. 2010, Turner et al. 2006, 2008) or stratification (Stow et al. 2005), but this is the first study to suggest this trend while considering both factors in combination. June NO_{2-3} concentration and April TKN and NO_{2-3} loads are also included, but appear to have relatively small effects. The selection of TKN only for the month of April is consistent with this more refractory nitrogen fraction requiring additional time to become biologically available. The elastic-net model did not select any of the variables for TP or flow, suggesting that these factors are less important for predicting mid-summer hypoxia. Overall, the elastic-net model explains 85% of the variability in the annual intercepts. Figure 4-5 graphically illustrates how the various factors in eq 4-13 affect the annual intercepts.

We also developed a simple linear regression (SLR) for the annual intercepts, using May NO_{2-3} concentration alone (eq 12). The NO_{2-3} concentrations were normalized so that the model coefficients are comparable to those in eq 11. This model explains 76% of the variability in the annual intercepts, indicating that its performance is similar to the elastic-net model, despite its relative simplicity.

$$\hat{\beta}_a = 2.36 - 0.620 * [NO_{2-3,May}] \quad \text{eq 4-14}$$

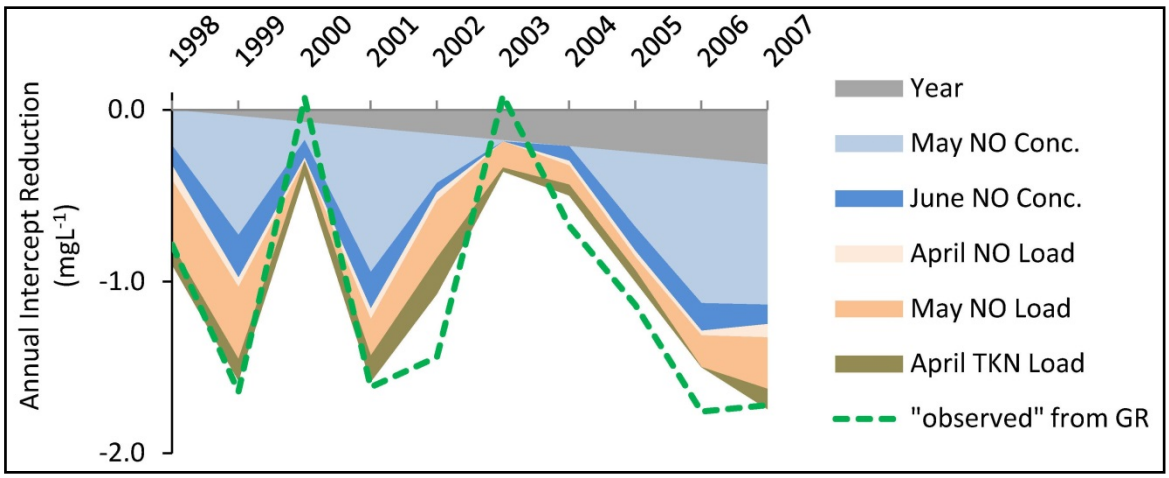

Figure 4-5: Annual intercept reductions attributed to different predictor variables from elastic-net model. Each factor presented relative to its year of minimum impact.

The use of May (or May-June) nitrate data to predict hypoxia has precedent in multiple previous studies (Bianchi et al. 2010, Greene et al. 2009, Liu et al. 2010, Scavia et al. 2003, Turner et al. 2006, 2008), and is reasonable given the known processes of phytoplankton growth and bacterial decomposition that link nitrogen inputs to BWDO depletion over time(Rabalais et al. 2007, Wiseman et al. 1997). However, while previous studies generally indicate that load is the best single nutrient predictor of hypoxia, this study indicates concentration (Greene et al. (2009) also consider models using flow and concentration, but the confounding correlation between flow and load was not addressed.) To explore this distinction further, Figure 4-6 illustrates how May NO_{2-3} loads and concentrations correlate with mean BWDO (eq 7) and the annual intercepts. As shown, mean BWDO is somewhat correlated with NO_{2-3} load (Figure 4-6.a), and as expected, the strength of this correlation (r^2=0.41, p=0.045) is similar to that of previous studies relating nitrogen loads to hypoxic area.(Bianchi et al. 2010, Greene et al. 2009, Liu et al. 2010, Scavia et al. 2003, Turner et al. 2006, 2008) The correlation between the annual intercepts and May NO_{2-3} concentrations (Figure 4-6.d), corresponding to the SLR (eq 12), is much stronger (r^2=0.76, p=0.001). This improvement can be attributed to the fact that here we use nutrients to predict only the variability in hypoxia that remains after accounting for stratification (i.e. the annual intercepts), rather than the total variability in hypoxia (i.e. mean BWDO or hypoxic area). If we do not remove stratification effects, nutrient load is a

better predictor of hypoxia than concentration (Figure 4-6.a versus 4-6.b). This is as expected given load is correlated with flow, and higher flow tends to increase stratification, making load partially reflective of both stratification and nutrient concentration.

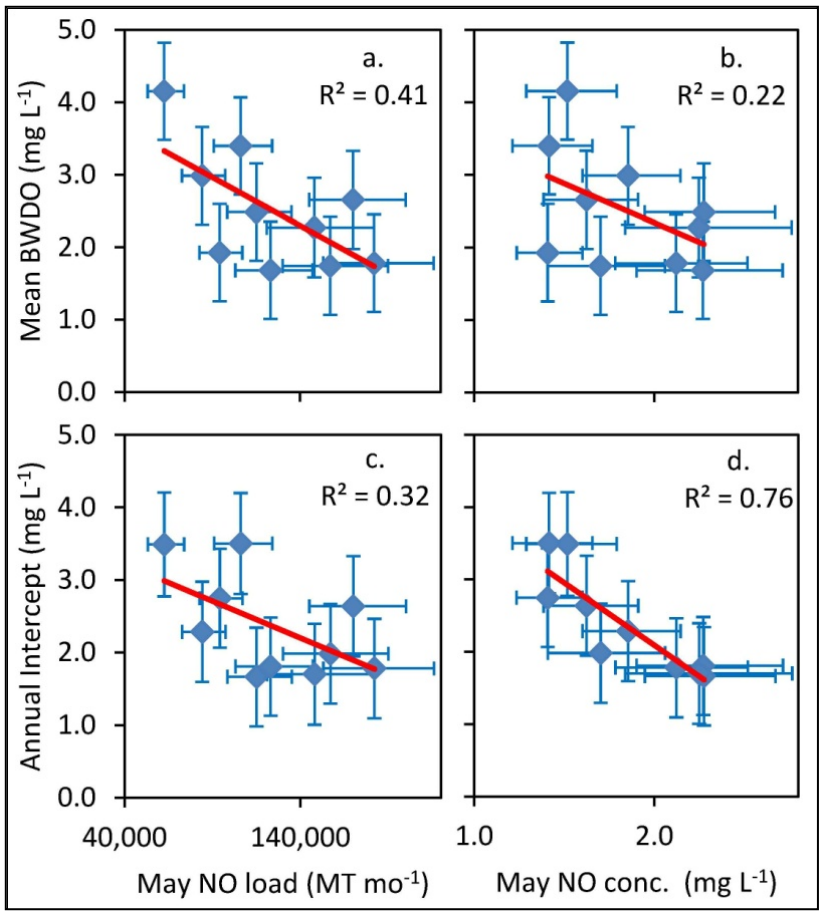

Figure 4-6: Annual mean BWDO (top) and annual intercept from GR (bottom) versus May $NO_{(2-3)}$ loads (left) and concentrations (right), with 95% confidence intervals.

While these results suggest that nutrient concentration is the best predictor of hypoxia (after accounting for stratification), nutrient load may also be of some predictive importance. The elastic-net model (eq 4-13) did include multiple nutrient loading terms, though they generally had less impact than the concentration terms. Similarly, Figure 4-6.c indicates that there is some correlation between the annual intercepts and May NO_{2-3} load, though markedly less than with concentration.

From a mechanistic perspective, nutrient load is an intuitive predictor of hypoxia based on the assumption that it is proportional to the primary production and oxygen demand that develops within the system. However, the importance of concentration (as suggested by this study) indicates that the amount of oxygen demand generated on the shelf is also affected by the degree to which that load is diluted by flow. The exact reasons for this are not readily apparent and could benefit from further research. However, loads are highly correlated with flows, and high flows can intensify coastal current velocities (Wang and Justic 2009), thus reducing the time available to produce and settle organic matter within the study area. Also, because biochemical process rates are typically concentration dependent, dilution can reduce the rates of oxygen demand formation and nutrient recycling.

The approach outlined in this study made it possible to isolate the effects of stratification through GR, and then determine the effects of nutrients based on the interannual variability in BWDO that remained. Overall, the results suggest that both river nutrient concentration and stratification play large and comparable roles in explaining the interannual variability of hypoxia in the northern Gulf of Mexico. The standard deviation of the net stratification effect is 0.51 mg L^{-1} BWDO, while the standard deviation of the May $NO_{2\text{-}3}$ effect (from the SLR) is 0.62 mg L^{-1} BWDO, indicating nutrients may play a slightly larger role. When the annual intercepts predicted by the SLR are used in combination with the net stratification and northing/bathymetry effects, the overall model explains a large majority (82%) of the interannual variability in mean BWDO.

Because of its explanatory power, this approach provides a more precise way to evaluate the response of Gulf hypoxia to changes in Mississippi River nutrients. Scavia et al. (2003) note that when hypoxia is modeled using nutrient loads alone, the interannual variability in oceanographic conditions can mask the effects of changes in nutrient loading, at least in the short term. This can confound management of the system by obscuring the effect of watershed management practices aimed at curtailing nutrient inputs. However, by using the approach outlined here, we separate the effects of stratification from the effects of nutrients, allowing for a more detailed assessment of how each is impacting hypoxia over time. Although the intensity of stratification can be difficult to predict in advance

because it varies over short time scales due to wind forcing (Wang and Justic 2009), it can likely be approximated using river flow and climate data. For example, the May river flow alone explains 27% of the variability in the net stratification effect determined by this study. The stratification effect can also be determined following each mid-summer monitoring cruise using the measured salinity and temperature profiles, as was done here.

While May $NO_{2\text{-}3}$ concentration and load were not well correlated over the period of this study (r^2=0.14); longer-term trends in nitrate concentration and load have been shown to be parallel (Goolsby and Battaglin 2001). Thus, if the existing Action Plan is successful at reducing $NO_{2\text{-}3}$ load, it will likely also be successful in reducing concentration. So, while we think that concentration should receive more attention in future studies and in evaluating year-to-year changes in hypoxic severity, we do not think this emphasis on concentration is in discordance with the long-term Action Plan goal of reducing nutrient loads.

4.3.5 Nutrient reduction scenarios

Stratification metrics and river nutrient concentrations were also used to develop a regression model for hypoxic area (y, km^2), as described in Section 4.2.7. In the model (eq 4-15), S is the aggregated BWDO depletion due to salinity and temperature stratification (as illustrated in Figure 4-4); and $[NO_{2\text{-}3,May}]$ is the May $NO_{2\text{-}3}$ concentration. Both of the variables were normalized so that the regression coefficients indicate their relative importance. Overall, the model explains 79% of the variability in observed hypoxic area.

$$y = 15300 + 3800*S + 5500*[NO_{2\text{-}3,May}] \qquad \text{eq 4-15}$$

We used the model to estimate the average hypoxic area for different nutrient reduction scenarios, along with associated uncertainty determined through Monte Carlo simulation (eq 4-11); and to estimate the percent nutrient reduction required to consistently (over the ten-year study period) reduce the five-year running average of hypoxic area to less than 5,000 km^2 in accordance with Action Plan goals (EPA 2001, 2008). Figure 4 illustrates how nutrient reductions will affect the five-year running average of hypoxic area for the

critical period (1998-2002 is the critical period because it requires the greatest nutrient reduction to achieve the Action Plan goal). The model suggests that a 42% reduction in May NO_{2-3} concentration is required, with a 90% confidence interval of 29-62% reduction. The non-negativity constraint (Section 4.2.7) had only a minor impact, resulting in a recommended nutrient reduction of 42% instead of 41%. These results support multiple previous modeling studies suggesting nutrient reductions of around 40-45% (Greene et al. 2009, Scavia et al. 2003, Scavia and Donnelly 2007), but are somewhat lower than the 71% reduction recommended by Liu et al (2010).

Note that these predictions assume the linear relationships in eq 4-15 remain valid outside of the range of observed conditions. If the effects of stratification become less severe under scenarios of reduced nutrient inputs, this would make the estimated loading reductions somewhat conservative. It should also be noted that this study used recent data (1998-2007), so the results generally reflect the system at its current state. If the system becomes more or less susceptible to hypoxia over time, as a result of long-term trends in nutrient loading, climate change, or other factors, then the reductions required to meet the Action Plan goal would also change.

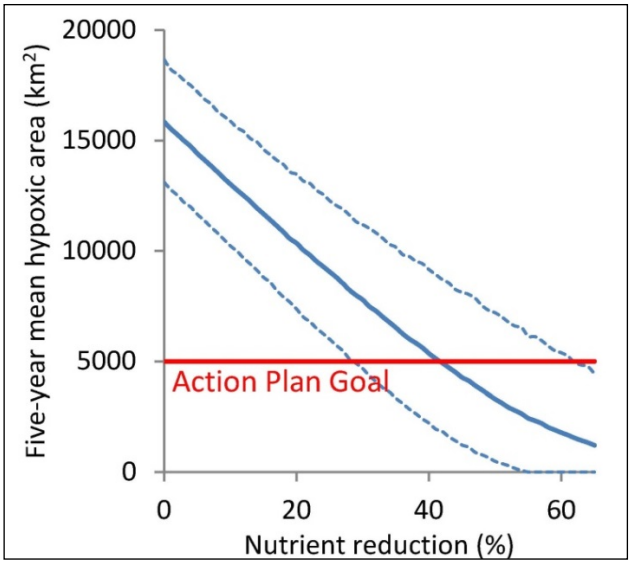

Figure 4-7: Predicted average hypoxic area for the critical five-year period (1998-2002) under different levels of nutrient reduction, with 90% confidence intervals.

CHAPTER 5: Geostatistical Spatial Estimation

Robust estimates of hypoxic extent (both area and volume) are important for assessing the impacts of low dissolved oxygen on aquatic ecosystems at large spatial scales. Such estimates are also important for calibrating models linking hypoxia to causal factors, such as nutrient loading and stratification, and for informing management decisions. In this study, we develop a rigorous geostatistical modeling framework to estimate the hypoxic extent in the northern Gulf of Mexico from data collected during midsummer, quasi-synoptic monitoring cruises (1985-2011). Instead of a traditional interpolation-based approach, we use a simulation-based approach that yields more robust extent estimates and quantified uncertainty. The modeling framework also makes use of covariate information (i.e., trend variables such as depth and spatial position), to reduce estimation uncertainty. Furthermore, adjustments are made to account for observational bias resulting from the use of different sampling instruments in different years. The text of this chapter is reproduced in part from unpublished work (in review):

Obenour, D. R., D. Scavia, N. N. Rabalais, R.E. Turner, and A. M. Michalak. A retrospective analysis of mid-summer hypoxic area and volume in the northern Gulf of Mexico, 1985-2011.

5.1 Introduction

The hypoxic extent, operationally defined as the region (area or volume) where DO concentrations are below 2 mg L^{-1}, is often assessed using data from 'shelfwide' sampling cruises. The cruises are considered 'quasi-synoptic' because changes in weather conditions (e.g., tropical storms) preceding or during the cruises can disrupt typical seasonal patterns in hypoxia. Cruises to document the occurrence of hypoxia, as well as

related physical and biological parameters, have been performed by the Louisiana Universities Marine Consortium (LUMCON) beginning in 1985 (Rabalais et al. 2002c). The estimated bottom-water hypoxic area (LUMCON 2012) has been determined by interpolating between sampling locations and hand-contouring parallel to isobaths over a calibrated (planimeter) grid. These estimates have been used in multiple modeling studies linking hypoxia to nutrient loads from the Mississippi River basin and other environmental factors (Forrest et al. 2011, Greene et al. 2009, Scavia et al. 2003, Turner et al. 2006) and for setting policy goals to reduce the severity of hypoxia (EPA 2001, 2008). However, these estimates are generally conservative in that the most inshore, most offshore, and most western extents of hypoxia are not always captured because of logistical constraints (Rabalais et al. 1999, Rabalais et al. 2007). Also, these estimates do not quantify the uncertainty inherent in making estimates from limited observations (Chiles and Delfiner 1999, Zhou et al. 2013). Quantified uncertainties are useful for assessing the adequacy of existing sampling programs and for improving models that link hypoxic extent to environmental causes and effects.

Hypoxic area estimates only partially characterize the hypoxic extent. The thickness and volume of hypoxia are also important, as they relate to how hypoxia affects pelagic organisms (Kimmel et al. 2009, Zhang et al. 2009). In addition, hypoxic volume, rather than area, should be more closely related to the total oxygen deficit of the system, potentially making it a more useful metric for biogeochemical modeling studies (Rabalais et al. 2010). While hypoxic volume measurements are available for other systems, such as Chesapeake Bay (Murphy et al. 2010, Zhou et al. (submitted manuscript)), estimates for the Gulf have only been available for a subset of years, using an un-published methodology (Rabalais et al. 2010).

The primary goal of this study is to improve our knowledge of the Gulf's hypoxic extent over the 27-year study period (1985-2011) by systematically estimating midsummer hypoxic area and volume. These are the first Gulf hypoxic extent estimates to include 'instrument bias adjustments' that account for the use of different sampling instruments, capable of being lowered to within different proximities of the sea floor, in different years (Turner et al. 2012). Our approach also uses trends between DO, hypoxic fraction, depth,

and spatial position to account for consistent, large-scale spatial patterns in DO, thereby improving the explanatory power of the model (Zhou et al. 2013). Our extent estimates are developed using a Monte Carlo-type simulation approach, rather than traditional interpolation, allowing for uncertainty quantification (i.e., confidence intervals) (Chiles and Delfiner 1999, Zhou et al. 2013). These confidence intervals reflect the spatial stochasticity of the system, uncertainties in trends among variables, and uncertainties in instrument bias adjustments. Hypoxic volume is estimated using a novel two-step approach where DO is simulated first, and hypoxic fraction (i.e., the fraction of the water column that is hypoxic) is simulated second using DO as a trend variable.

5.2. Materials and methods

5.2.1 Data and study boundaries

We use DO data from LUMCON mid-summer sampling cruises conducted between 1985 and 2011. Data for 1998-2008 are retrieved from the National Ocean Data Center, while data for other years are obtained directly from LUMCON (Rabalais 2011). Sampling locations are geo-referenced using the Universal Transverse Mercator (UTM) Zone 15 projection, and bathymetry is determined from a 3-arc-second digital elevation model obtained from the National Oceanic and Atmospheric Administration (NOAA 2011b).

During cruises, DO is sampled using one or two types of instruments (Figure 5-1): a rosette-mounted DO probe and a handheld DO probe, with the latter capable of being lowered closer to the sea floor (Obenour et al. 2012b, Turner et al. 2012). (Hereafter, we refer to these instruments as the “rosette sampler” and “handheld sampler”.) While instrument technology has changed over time, all instruments were calibrated against Winkler titrations, and post-cruise corrections were made as necessary. For sampling events where both instruments are used, data are combined into synthesized profiles (Obenour et al. 2012b). (Here, a ‘sampling event’ refers to the depth profile of data collected at a specific latitude/longitude and time.)

We extract bottom water dissolved oxygen (BWDO) and minimum dissolved oxygen (MinDO) concentrations from the DO profiles. BWDO and MinDO concentrations greater

than 7 mg L^{-1} (less than 1% of total samples), are treated as 7 mg L^{-1}, because higher concentrations are outliers representing super-saturation conditions that are not of interest in this study. For sampling events where hypoxia was present, we also extract the fraction of the water column that is within the hypoxic bottom layer (hereafter, bottom water hypoxic fraction, or BWHF) and, where other hypoxic layers are present, the total hypoxic fraction (THF) of all hypoxic layers. Bottom and upper layers of hypoxia are observed at 49% and 12% of sampling events, respectively. Here, we focus on the models for BWDO and BWHF because bottom layer results are most comparable to previous studies (LUMCON 2012, Rabalais et al. 2007). Example BWDO and BWHF values from an observed DO profile are illustrated in Figure 5-2. The models for MinDO and THF are formulated in the same way and yield comparable results, but for brevity are relegated to the Appendix.

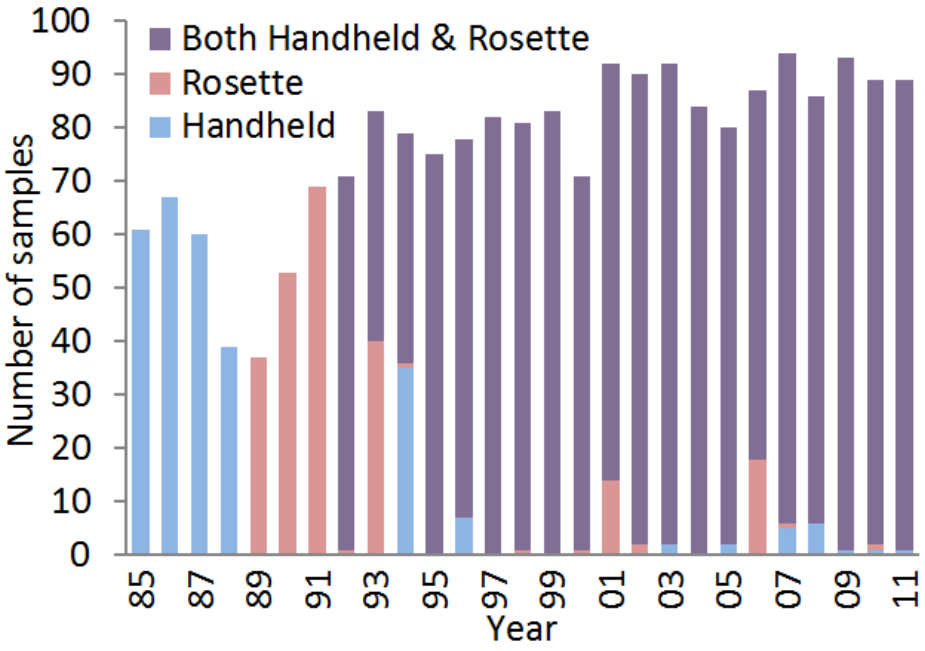

Figure 5-1: Number of locations sampled during the annual midsummer shelfwide cruises using handheld and rosette instruments

Geostatistical modeling is performed on a 5 km × 5 km grid of estimation points (Figure 5-3), covering 342.5-837.5 km UTM easting and 3122.5-3292.5 km UTM

northing, reflecting the general extent of sampling. The estimation grid is limited to depths between 3 and 80 m, which are typical of the Louisiana-Texas shelf region where hypoxia occurs. Observed values of BWDO and BWHF for the entire study are presented in Figures 5-4 and 5-5, respectively.

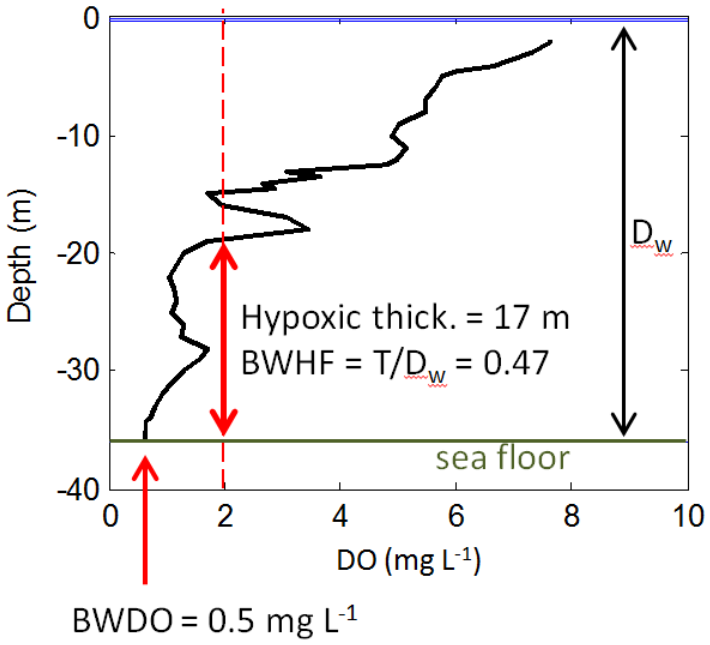

Figure 5-2: Example dissolved oxygen profile with calculated BWDO and BWHF

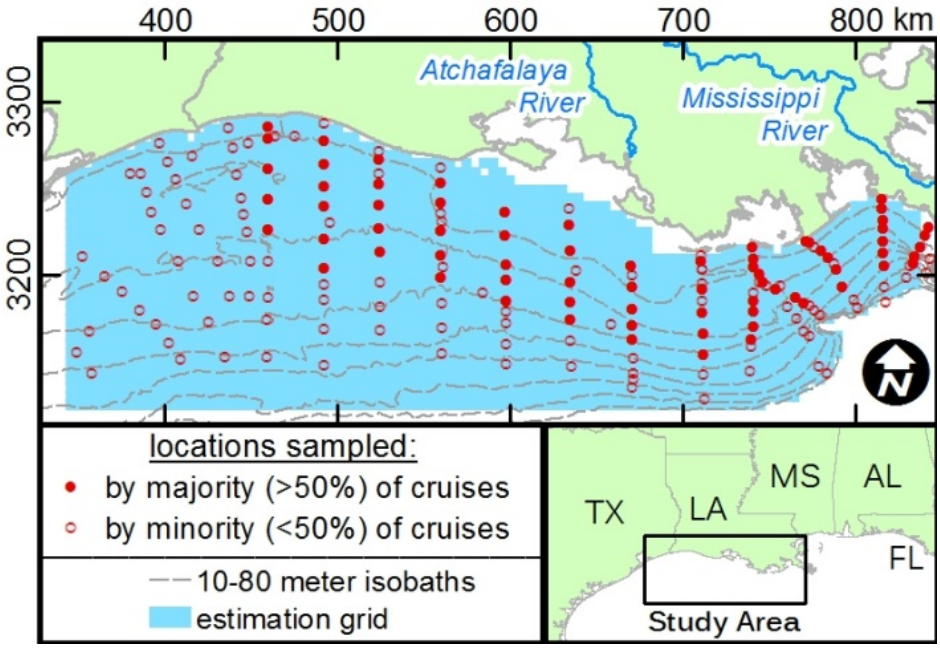

Figure 5-3: Study area bathymetry, sampling, and estimation locations

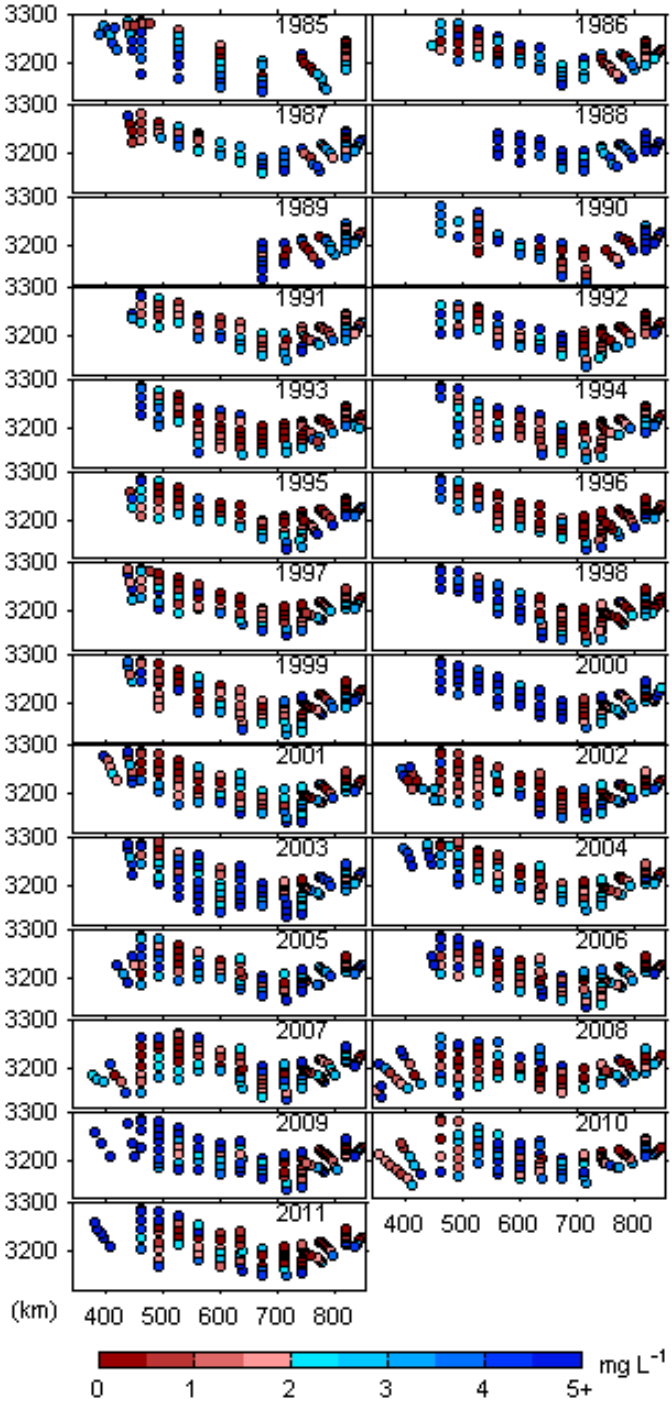

Figure 5-4: Maps of observed BWDO concentrations, 1985-2011

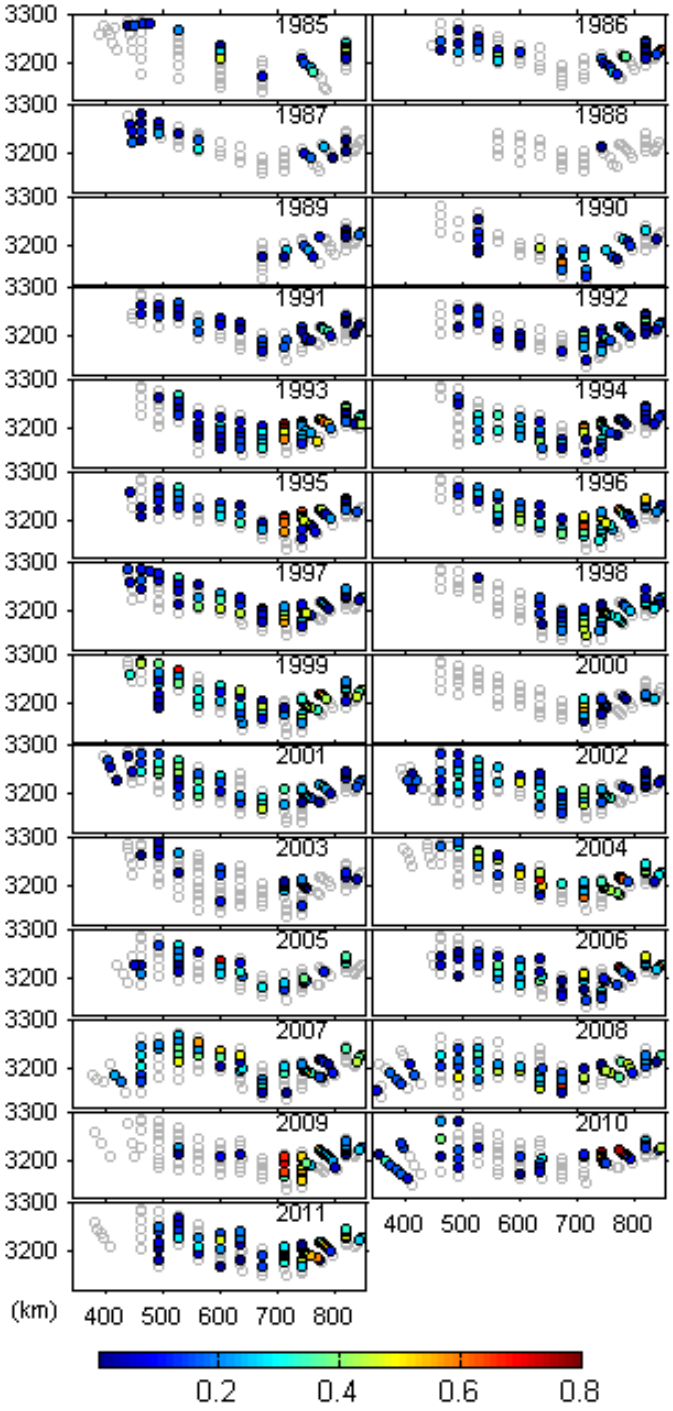

Figure 5-5: Maps of observed BWHF, 1985-2011

5.2.2 Model formulation

Geostatistical methods provide an effective means to model data that exhibit spatial correlation (Chiles and Delfiner 1999, Zimmerman and Stein 2010). The efficacy of these methods has been demonstrated in previous environmental analyses of rainfall, snow depth, soil phosphorus, and water quality indicators (Erickson et al. 2005, Goovaerts 2000, Murphy et al. 2010, Qian 1997, Zhou et al. 2013). Many of these studies have focused on mapping through spatial interpolation, and comparison studies have demonstrated that geostatistical interpolation outperforms simpler interpolation methods, such as inverse distance weighting (Goovaerts 2000, Murphy et al. 2010).

As described in Section 3.1.1 (eq 3-1), a geostatistical model represents a dependent variable in terms of its deterministic and stochastic components ($\mathbf{z} = \mathbf{X\beta} + \boldsymbol{\eta} + \boldsymbol{\varepsilon}$). This formulation is similar to linear regression, but includes an additional term, $\boldsymbol{\eta}$, representing spatially correlated stochasticity, in addition to the more commonly modeled uncorrelated stochasticity, $\boldsymbol{\varepsilon}$. As in linear regression, $\mathbf{X\beta}$ is the portion of $\mathbf{z}$ that can be expressed as a deterministic function of categorical and/or trend variables ($\mathbf{X}$) and their corresponding regression coefficients ($\boldsymbol{\beta}$). If the $\mathbf{X}$ term includes only cruise-specific categorical variables (vectors of zeros and ones that bin the data by cruise), then the model can essentially be used to perform ordinary kriging (OK). If $\mathbf{X}$ also includes trend variables, then it can be used to perform universal kriging (UK) (Chiles and Delfiner 1999).

In this study, the stochastic components ($\boldsymbol{\eta}$ and $\boldsymbol{\varepsilon}$) were found to be well represented by the commonly used exponential covariance function with a nugget effect as defined in eq 3-5. As in Section 4.2.2, anisotropy was accounted for by scaling $h_{i,j}$ using parameter α, which represents the ratio of east-west to north-south correlation ranges. Covariance function parameters are estimated using REML and deterministic component parameters are estimated using generalized least squares, as described in Section 4.2.3.

Geostatistical models are developed for BWDO and BWHF to determine bottom layer hypoxic area and volume, respectively. The model for BWDO uses all observations, while the model for BWHF uses only observations from locations where the BWDO concentration was hypoxic. Both models use the UK formulation to take advantage of

potential relationships between the response variables and available trend variables (Zhou et al. 2013).

The deterministic components of the model include two types of trends. First, 'constant trends' represent relationships between response and trend variables using regression coefficients that are the same for all cruises. For the BWDO model, potential constant trends include linear and quadratic trends with depth (Depth and $Depth^2$), easting (Easting and $Easting^2$), and northing (Northing and $Northing^2$). Based on a preliminary examination of model residuals, we noted that the overall trend between depth and BWDO does not continue for depths greater than 40 meters, and so depths of greater than 40 meters were treated as 40 meters in the final BWDO model development. This approach is justified because model residuals are evenly distributed across all depths when depths greater than 40 m are modified as described here (see Figure 5-10). For BWHF, potential constant trends include linear and quadratic trends with easting, northing, and BWDO.

Second, 'cruise-specific trends' (with cruise-specific regression coefficients for Easting) represent relationships that are specific to individual cruises. The inclusion of these trends was motivated by previous studies (Rabalais et al. 2007, Turner et al. 2012) indicating that the east-west distribution of hypoxia is influenced by alongshore current velocity, which can vary inter-annually in response to prevailing winds. When selected, these cruise-specific trends effectively modify the constant trend.

To prevent over-parameterization of the model, we use only trend variables selected through a geostatistical adaptation of BIC, as described in Section 4.2.4. Because of the large number of variables considered in this study, an initial optimal model is first selected among those formed from constant trend variables only. Cruise-specific trend variables are then individually tested relative to that model, and only cruise-specific trends that improve the BIC score are included in the final model.

As in Section 4.2.1, the deterministic portion of the model also includes categorical variables that bin the data by cruise. These variables reflect the fact that the mean BWDO varies from cruise to cruise. In UK, the categorical variables essentially allow for different 'intercepts' (as in linear regression), that shift the trend up or down to best fit the

observations from each cruise. In OK, these categorical variables simply allow for a different mean for each cruise.

The covariance model, selected deterministic variables, and categorical variables are used to determine a set of geostatistical weights, $\boldsymbol{\Lambda}$, that are applied to observations when performing geostatistical interpolation and simulation. For each cruise, *y*, a unique set of weights, $\boldsymbol{\Lambda}_y$, are determined by solving a system of linear equations:

$$\begin{bmatrix} \mathbf{Q}_{oo} & \mathbf{X}_o \\ \mathbf{X}_o{}^T & \mathbf{0} \end{bmatrix} \begin{bmatrix} \boldsymbol{\Lambda}_y \\ -\mathbf{G}_y \end{bmatrix} = \begin{bmatrix} \mathbf{Q}_{oe,y} \\ \mathbf{X}_e{}^T \end{bmatrix} \qquad \text{eq 5-1}$$

where $\mathbf{Q}_{oo}$ is an $n \times n$ covariance matrix for the n observation locations (all cruises), with elements determined from eq 3-5. Because we assume no correlation among stochasticity from different cruises, inter-cruise covariances are assigned a value of zero. Similarly, $\mathbf{Q}_{oe,y}$ is an $n \times m$ covariance matrix of n observation locations and m estimation locations, and the rows of $\mathbf{Q}_{oe,y}$ that correspond to observations from cruises other than cruise y are assigned a value of zero. The matrix $\mathbf{X}_o$ is $n \times p$ and includes the p deterministic variables (trend and categorical) for the observation locations, and the matrix $\mathbf{X}_e$ ($m \times p$) includes the same variables for the estimation locations. The trend variables are normalized to a mean of zero and variance of one. Note that terms with a '*y*' subscript are cruise-specific.

In eq 5-1, $\boldsymbol{\Lambda}_y$ is an $n \times m$ matrix of cruise-specific weights that reflect both the spatial correlation structure and the deterministic trends. Also, $\mathbf{G}_y$ is a $p \times m$ matrix of Lagrange multipliers that can be used with $\boldsymbol{\Lambda}_y$ to determine location-specific estimation uncertainties. Using the weights, along with the observations $\mathbf{z}_o$ (an $n \times 1$ vector), one can develop estimates of the response variable across the estimation grid:

$$\mathbf{z}_{e,y}^0 = \boldsymbol{\Lambda}_y^T \mathbf{z}_o \qquad \text{eq 5-2}$$

where $\mathbf{z}_{e,y}^0$ is an $m \times 1$ vector of interpolated BWDO or BWHF values for cruise *y*. Eq 5-2 is equivalent to eq 3-6 solved for a vector of estimation locations.

Although we use a UK model formulation, hypoxic area and volume are not determined by kriging (i.e., spatial interpolation), but instead by developing conditional realizations (CRs), which are essentially "*spatially consistent* Monte Carlo simulations" (Chiles and Delfiner 1999). While both kriging and CRs provide equivalent information for individual estimation locations, CRs are necessary to estimate spatially aggregated quantities (e.g., area and volume) probabilistically. A CR is performed by first creating an 'unconditional realization' and then 'conditioning' it to the observed data and deterministic trends (Chiles and Delfiner 1999). Unconditional realizations (eq 5-3) include simulated values at the estimation locations ($\mathbf{z}^{\mathrm{u}}_{e,y}$) as well as at the observation locations ($\mathbf{z}^{\mathrm{u}}_{o,y}$):

$$\begin{bmatrix}\mathbf{z}^{\mathrm{u}}_{e,y}\\ \mathbf{z}^{\mathrm{u}}_{o,y}\end{bmatrix} = \mathbf{C}\left(\begin{bmatrix}\mathbf{Q}_{ee} & \mathbf{Q}^{T}_{oe,y}\\ \mathbf{Q}_{oe,\boldsymbol{y}} & \mathbf{Q}_{oo}\end{bmatrix}\right)^{T}\mathbf{u} \qquad \text{eq 5-3}$$

Note that the vector $\mathbf{z}^{\mathrm{u}}_{o,y}$ includes simulated values corresponding to the observations from all cruises, but only observations from cruise *y* have their stochasticity correlated with that of the estimation locations. Here, $\mathbf{Q}_{ee}$ is the $m \times m$ covariance matrix between estimation locations, and $\mathbf{u}$ is an $(m+n)\times 1$ vector of random independent samples from the standard normal distribution. The operator $\mathbf{C}(\,)$ returns the triangular matrix resulting from Cholesky decomposition of the subject matrix.

The unconditional realizations are then conditioned to the observed data and deterministic trends through:

$$\mathbf{z}^{\mathrm{c}}_{\mathrm{e,y}} = \boldsymbol{\Lambda}^{T}_{y}\left(\mathbf{z}_{o} - \mathbf{z}^{\mathrm{u}}_{o,y}\right) + \mathbf{z}^{\mathrm{u}}_{e,y} \qquad \text{eq 5-4}$$

where $\mathbf{z}_{o}$ is the $n\times 1$ vector of observed values, and $\mathbf{z}^{\mathrm{c}}_{e,y}$ is the resulting cruise-specific CR, an $m\times 1$ vector of values corresponding to the estimation locations.

The CRs are performed in two steps. First, BWDO concentration is simulated across the entire estimation grid. Then, BWHF is simulated over those locations where the simulated BWDO is below the hypoxic threshold. At hypoxic estimation locations, the

simulated BWDO values are used as a trend variable in the BWHF model. Simulated values are limited to within realistic ranges (determined to be 0-7 mg L^{-1} for BWDO and 0.001-0.8 for BWHF, based on a review of the observed data). However, these constraints did not have a large effect on results, because the vast majority of simulated values fell within these ranges. Without these constraints, the estimated hypoxic volume would be 6% smaller, on average, as a result of (unrealistic) negative hypoxic thicknesses.

The two-step simulation process is repeated 1,000 times for each cruise, resulting in 1,000 realizations of both BWDO and BWHF. For each realization of BWDO, the hypoxic area is calculated as the number of estimation locations simulated to be hypoxic multiplied by the grid cell area (25 km^2); and for each realization of BWHF, the hypoxic volume is calculated as the vector of simulated BWHF values multiplied by the corresponding vector of water column depths, all multiplied by the grid cell area. From this ensemble of results, we determine the mean and 95% confidence intervals for the hypoxic area and volume for each cruise.

5.2.3 Instrument adjustment

Adjustments were made to address biases that arise from the use of different sampling instruments. In particular, for cruises when only the rosette sampler was used (Figure 5-1), one would expect the hypoxic extent to be underestimated because the rosette sampler does not reach as close to the sea floor as the handheld sampler. We quantified this bias by comparing data from sampling events where both instruments were used. For these cases, BWDO and BWHF were calculated for both the synthesized profile (both instruments) and the rosette-only profile. Probabilistic relationships were then developed between the synthesized results and the rosette-only results. When performing the conditional realizations (Section 5.2.2), rosette-only observations were adjusted by sampling from these relationships. The process of developing these relationships is described in the following paragraphs.

Using data from sampling events where both the rosette and handheld samplers were deployed, we developed relationships between the hypoxic conditions (BWDO and hypoxic thickness) derived from the synthesized data (both instruments) and the hypoxic

conditions derived from rosette-only data. Figure 5-6 presents BWDO values derived from the synthesized data (S_{BO}) versus BWDO values from the rosette-only data (R_{BO}). We divided the data into two different categories (blue and red, Figure 5-6), where the blue data meet the following criterion:

$$S_{BO} - R_{BO} > 2\sigma_{\varepsilon} \qquad \text{eq 5-5}$$

where σ_{ε} is the standard deviation of the stochasticity that is not spatially correlated (i.e. microvariability), as described in the primary text. For the blue data, the relationship between S_{BO} and R_{BO} can be represented approximately using a simple linear regression with normally distributed residuals (ϵ):

$$\hat{S}_{BO,blue} = 0.973 R_{BO,blue} + \epsilon \qquad \text{eq 5-6}$$

The remaining (red) data can be modeled as a uniform distribution between zero and the threshold criterion used in eq 5-5:

$$\hat{S}_{BO,red} \sim \mathrm{U}(0, [R_{BO,blue} - 2\sigma_{\varepsilon}]) \qquad \text{eq 5-7}$$

Using these relationships, we can simulate values of S_{BO} for rosette-only sampling events. (Each conditional realization is assigned a unique set of simulated values.) For R_{BO} less than $2\sigma_{\varepsilon}$ mg L^{-1}, eq 5-6 always applies. For R_{BO} greater than $2\sigma_{\varepsilon}$ mg L^{-1}, eq 5-6 is applied at an 88.9% probability and eq 5-7 at an 11.1% probability. These percentages reflect the actual partitioning of the data as presented in Figure 5-6. From a physical perspective, application of eq 5-7 represents situations where there is a thin, high-density, bottom layer that is not reached by the rosette. Conversely, the application of eq 5-6 implies that the rosette did reach the bottom-most layer of water. When performing simulations, we do not sample from the error term (ϵ) in eq 5-6, as this variability is expected to be primarily reflective of the microvariability already accounted for within the covariance model.

For hypoxic thickness, the relationship between the synthesized and rosette-only data is somewhat simpler (Figure 5-7). Here, the relationship between synthesized thickness (S_{Th}) and rosette thickness (R_{Th}) can be approximately modeled using a simple linear regression with normally distributed residuals (the units of the equation are meters):

$$\hat{S}_{Th} = R_{Th} + 0.82 + \epsilon \qquad \text{where ... } \epsilon \sim \mathrm{N}(0,0.36) \qquad \text{eq 5-8}$$

When performing simulations, we do sample from the error term (ϵ) of eq 5-8, as this error is expected to be primarily reflective of the variability in the maximum rosette sampling depth, rather than the natural variability in the thickness of the hypoxic layer. For observations that are not hypoxic based on the rosette measured BWDO, but become hypoxic when performing the instrument adjustment for BWDO ($R_{BO} > 2$ mg L^{-1} and $\hat{S}_{BO} <$ 2 mg L^{-1}), we multiply $\hat{S}_{Th}$ (which is 0.82 + ϵ in this case) by a sample from a standard uniform distribution because it is unclear what portion of the offset is hypoxic. Although more realistic, this additional step has a negligible impact on results.

In 1991 a larger bias adjustment is required for the first 38 sampling events because the ship's fathometer was not functioning correctly, causing the rosette sampler to be lowered 1.5 meters less than it would have been otherwise (N.N. Rabalais, cruise records). For these observations, the probabilistic relationships between synthesized and rosette-only results are again determined based on sampling events where both instruments were functioning properly (as described above) but with the bottom 1.5 m of the rosette-only profiles removed. Here, eqs 5-9, 5-10 and 5-11 are analogous to eqs 5-6, 5-7, and 5-8, respectively. For $R_{1.5,BO}$ greater than $2\sigma_\varepsilon$ mg L^{-1}, eq 5-9 is applied at an 67.5% probability and eq 5-10 at an 32.5% probability.

$$\hat{S}_{BO,blue} = -0.163 + 0.967R_{1.5,BO,blue} + \epsilon \qquad \text{eq 5-9}$$

$$\hat{S}_{BO,red} \sim \mathrm{U}(0, (R_{1.5,BO,blue} - 2\sigma_\varepsilon)) \qquad \text{eq 5-10}$$

$$\hat{S}_{Th} = R_{1.5,Th} + 2.3 + \epsilon \qquad \text{where ... } \epsilon \sim \mathrm{N}(0,0.39) \qquad \text{eq 5-11}$$

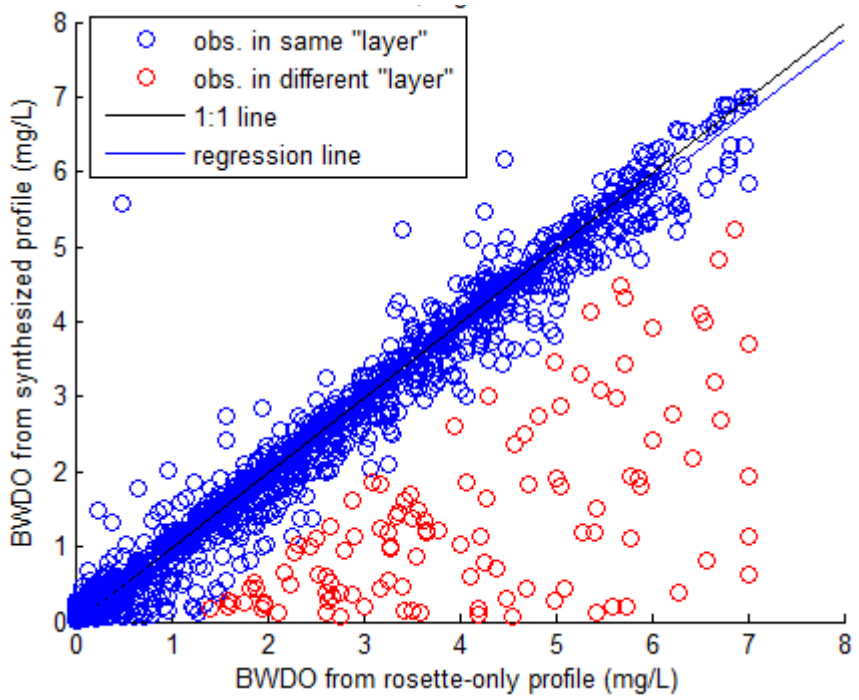

Figure 5-6: BWDO from synthesized data (S_{BO}) vs. BWDO from rosette instrument only (R_{BO})

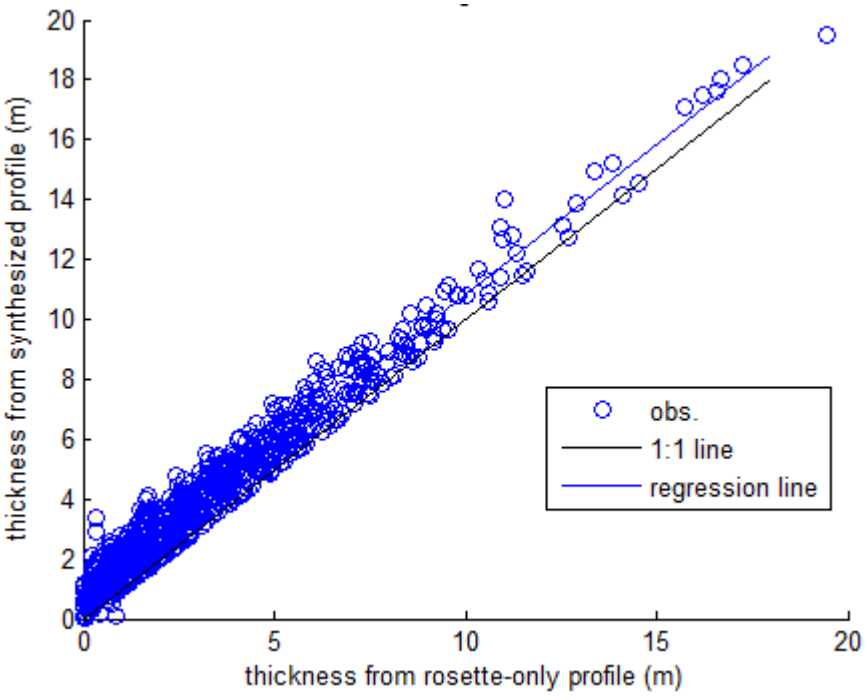

Figure 5-7: Hypoxic thickness from synthesized data (S_{Th}) vs. hypoxic thickness from rosette instrument (R_{Th})

5.3 Results

5.3.1 Model parameters

The BWDO and BWHF models include several parameters that characterize the deterministic and stochastic model components. Regression coefficients for the BIC-selected trend variables, explaining a portion of the spatial variability in observed BWDO and BWHF, are provided in Table 5-1. The standard errors of these coefficients are low (i.e., $p<0.05$), suggesting these trends are statistically significant. The coefficients for the cruise-specific categorical variables, accounting for year-to-year variability in the responses, are included in Table 5-2. Overall, the deterministic model components explain 28% and 32% of the total (spatial plus inter-annual) variability in BWDO and BWHF, respectively, while the stochastic components of the models explains the remainder of the spatial variability. If trend variables are omitted (i.e., the OK formulation), then the deterministic components only account for inter-annual variability, and only 12% and 11% of the total variability in BWDO and BWHF is explained, respectively.

The BWDO model includes trends with easting, northing, and depth. The trend between BWDO and easting is quadratic with a minimum at 794 km UTM easting, between the Mississippi and Atchafalaya river outfalls, which is reasonable given that these rivers provide the freshwater flows and nutrients that are important for hypoxia formation. In 1998, the only year for which a cruise-specific east-west trend with BWDO is selected, the minimum is shifted to 1111 km easting, outside of the study area, indicating that BWDO concentrations decrease monotonically with easting within the study area. The unique spatial distribution in 1998 has been noted previously, and is generally attributed to unusually persistent eastward currents (Rabalais et al. 2007). A similar BWDO pattern has been noted for 2009 (Turner et al. 2012) but it was not sufficiently strong to result in a cruise-specific trend in this analysis. The trend between BWDO and northing is linear, and suggests that BWDO concentrations are higher to the south. The trend between BWDO and bathymetry (depth) is quadratic with a minimum at 22 m.

Table 5-1: Regression coefficients ($\hat{\beta}_p$) with standard errors ($\sigma_{\hat{\beta}}$) for normalized trend variables selected in the BWDO and BWHF models

Variable	BWDO (mg L^{-1})		BWHF (-)	
	$\hat{\beta}$	$\sigma_{\hat{\beta}}$	$\hat{\beta}$	$\sigma_{\hat{\beta}}$
Easting	-0.62	0.09	0.018	0.007
Easting^2	0.25	0.07	-0.020	0.006
Northing	-0.36	0.09	n.s.	
Depth	-2.31	0.18	n.a.	
Depth^2	2.45	0.17	n.a.	
BWDO	n.a.		-0.065	0.005
c.s.E 1998	-1.35	0.45	n.s.	

c.s.E=cruise specific trend for Easting,
n.s.=not selected,
n.a.=not available

Table 5-2: Coefficients (a.k.a. annual intercepts, $\hat{\beta}_a$) for categorical variables in the BWDO and BWHF models

year	BWDO (mg L^{-1})	BWHF (-)	year	BWDO (mg L^{-1})	BWHF (-)
1985	2.43	0.19	1998	2.71	0.18
1986	2.71	0.18	1999	1.82	0.31
1987	3.13	0.16	2000	4.19	0.24
1988	5.40	0.10	2001	2.13	0.22
1989	3.30	0.11	2002	1.97	0.19
1990	2.84	0.16	2003	3.63	0.15
1991	2.71	0.15	2004	2.22	0.32
1992	2.76	0.14	2005	2.99	0.23
1993	1.91	0.20	2006	2.51	0.22
1994	2.30	0.21	2007	2.01	0.29
1995	1.94	0.18	2008	1.84	0.26
1996	1.65	0.21	2009	3.82	0.30
1997	2.03	0.19	2010	2.10	0.27
			2011	2.39	0.20

The linear and quadratic trends with depth, easting, and northing can be aggregated, and the overall trend in BWDO can be illustrated graphically (Figure 5-8). The overall trend should realistically represent the spatial pattern in BWDO, and this is evaluated by

comparing the trend to a hypoxic frequency map developed by LUMCON (2012) (Figure 5-9). As shown, there is remarkable agreement between these maps, with the areas of low BWDO in Figure 5-8 generally coinciding with the areas of high hypoxic frequency in Figure 5-9.

Because BWHF was modeled as a function of BWDO, it effectively inherits the trends from the BWDO model. The trend with BWDO is linear and suggests that the BWHF is larger where BWDO concentrations are lower. The model for BWHF also includes a quadratic trend with easting, suggesting a maximum at 701 km, about 30 km east of the Atchafalaya outfall location (though the overall trend in BWHF is also affected by the trends in BWDO). No cruise-specific east-west trends were selected for the BWHF model.

The models also include covariance parameters that characterize their stochastic components. For BWDO, σ_ε^2 and σ_η^2, commonly referred to as the nugget and partial sill, were 0.39 and 2.33 ($mg^2\ L^{-2}$), respectively. Since σ_ε^2 is smaller than σ_η^2, the majority of the stochasticity in the model is spatially correlated. The approximate range of spatial correlation ($3r$, per eq 3-5) is 94 km in the east-west direction and 53 km in the north-south direction. The greater correlation distance in the east-west direction was expected due to the dominant east-west current pattern (Wiseman and Kelly 1994, Zavala-Hidalgo et al. 2003). For BWHF, the spatial correlation of the stochasticity is somewhat weaker with σ_ε^2 and σ_η^2 similar in magnitude, having values of 0.009 and 0.011 (unit-less), respectively. Also, the correlation range was 63 km in all directions (anisotropy was negligible). Overall, spatially correlated stochasticity accounts for the greatest portion of the variability in both BWDO and BWHF. As discussed in Section 4.3.1, the spatial correlation of the stochasticity is consistent with the effects of varying coastal current patterns, influencing the distribution of hypoxia over the spatial scales described above. In general, the correlation ranges are considerably longer than the typical distances between sampling locations (Figure 5-3), especially on the eastern shelf, suggesting the sampling network is adequate for resolving the spatially correlated stochasticity of the system.

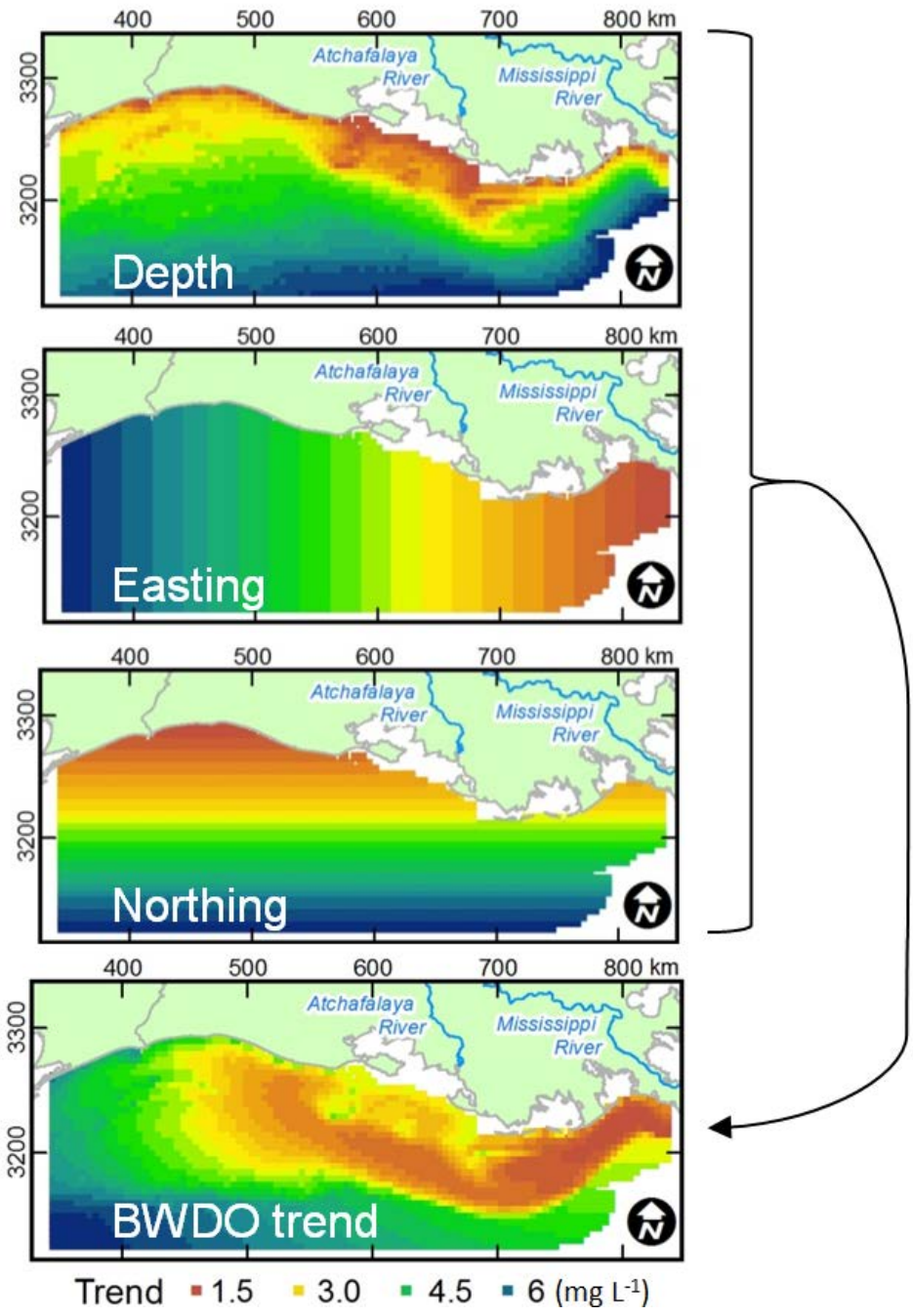

Figure 5-8: Covariate maps and resulting deterministic trend (for an average year)

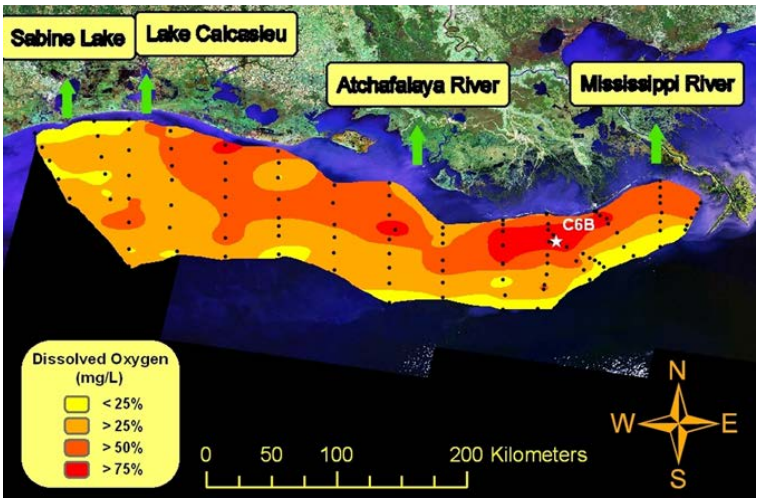

Figure 5-9: Frequency of hypoxia, 1985-2008, station locations as black dots (LUMCON 2012)

5.3.2 Tests of linearity

In Figures 5-10 and 5-11, the residuals (stochastic portion) of the UK models for BWDO and hypoxic fraction, respectively, are plotted versus each of the trend variables used in these models. Because the residuals are generally evenly distributed around zero throughout the ranges of the trend variables, the linear model formulation appears reasonable. Note that the models do include nonlinear transformations of the trend variables (i.e. depth-squared) but they are incorporated within a linear modeling framework.

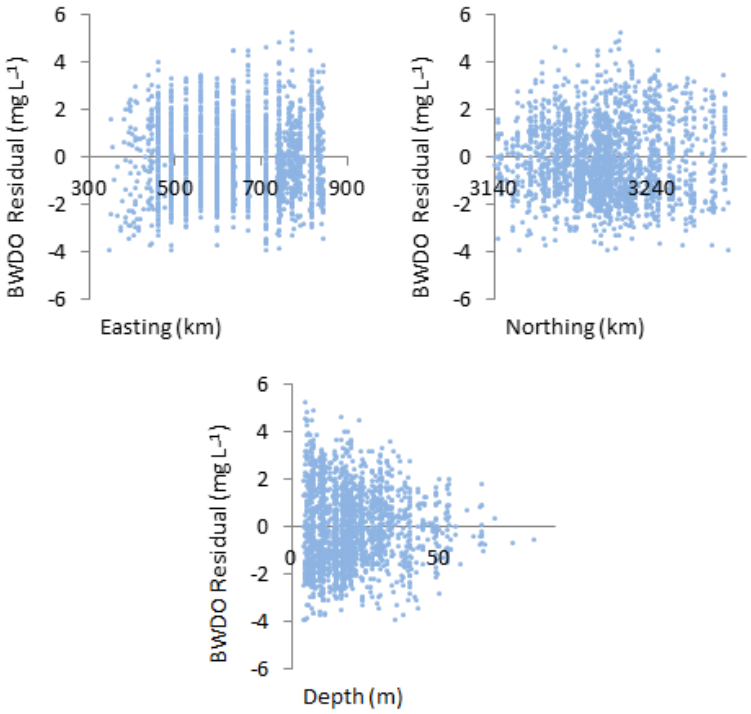

Figure 5-10: BWDO residuals (stochastic portion of UK model) vs. covariates

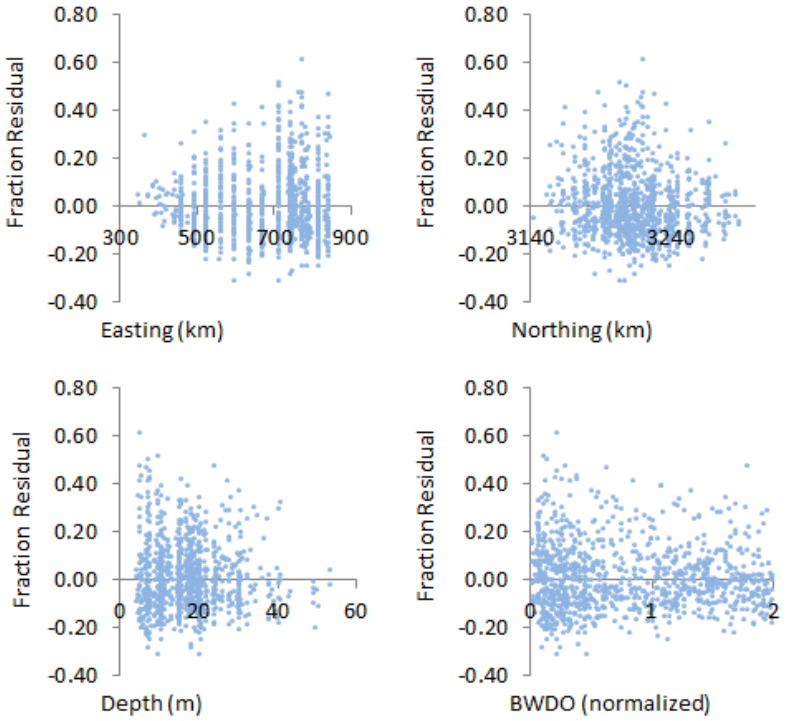

Figure 5-11: BWHF residuals (stochastic portion of UK model) vs. covariates

5.3.3 Hypoxic extent

Using CR, we determined the mean and 95% confidence intervals for hypoxic area and volume of each cruise (Figure 5-12). The largest estimated hypoxic area was for 1996, but it was not significantly different from 1991, 1993, 1995, 1997, 1999, 2001, 2002, 2007, or 2008 ($p>0.05$), given the uncertainty in the area estimates. The largest hypoxic volume estimate was for 2008, but it was not significantly different from 1993, 1996, 1999, 2004, or 2007 ($p>0.05$). Both hypoxic area and volume were lowest in 1988, a drought year (Rabalais et al. 2007).

Two sets of hypoxic area estimates, determined by LUMCON, are also included in Figure 5-12 for comparison. The first set are the original LUMCON estimates, determined by hand contouring, as described above. The second set are revised estimates developed as part of this study; also by hand contouring, but using the updated BWDO values for this study (which are superior because they reflect post-cruise DO calibrations). In most cases, the original LUMCON estimates (Figure 5-12, open squares) overlap the revised estimates (solid squares), and the majority of the LUMCON estimates fall within the 95% confidence intervals determined by this study. However, for the first third of the study period (1985-1993), the new estimates are consistently higher than the LUMCON estimates.

LUMCON did not make an estimate for 1989, because only a portion of the shelf was sampled that year. Our method does allow for extent estimates for 1989, and the wide confidence intervals (Figure 5-12) reflect the relatively large uncertainties. Even though data for 1989 were limited, the overall spatial trends (based on the data from all years) help to constrain the variability in the CRs to within realistic ranges, across the study area. We note that the 1989 estimate presented here is consistent with previous estimates developed from nutrient loading models (Scavia and Donnelly 2007, Turner et al. 2005).

There was considerable interannual variability in the spatial distribution of hypoxia and in the thickness of the hypoxic bottom layer (Figure 5-13 through 5-15). Years with similar hypoxic areas may have very different average hypoxic thicknesses and thus

volumes. For example, 2002 and 2008 have similar hypoxic areas, but they are very different in terms of hypoxic thickness, such that 2008 has approximately twice the hypoxic volume of 2002 (Table 5-3). The results indicate that the average bottom layer hypoxic thickness for the 27-year study period was 3.9 m, with the thickest average layers of approximately 6.2 and 6.3 m occurring in 2008 and 2009, respectively. Over the 27-year study period, hypoxic volume is correlated with area (r^2=0.77). Interestingly, hypoxic area and thickness are also somewhat positively correlated over most years (r^2=0.37, not including 1998 and 2009, which were subject to unusually strong eastward currents (Rabalais et al. 2007, Turner et al. 2012)), such that volume increases exponentially relative to area. Additional maps, showing example CRs, are included in Figure 5-16

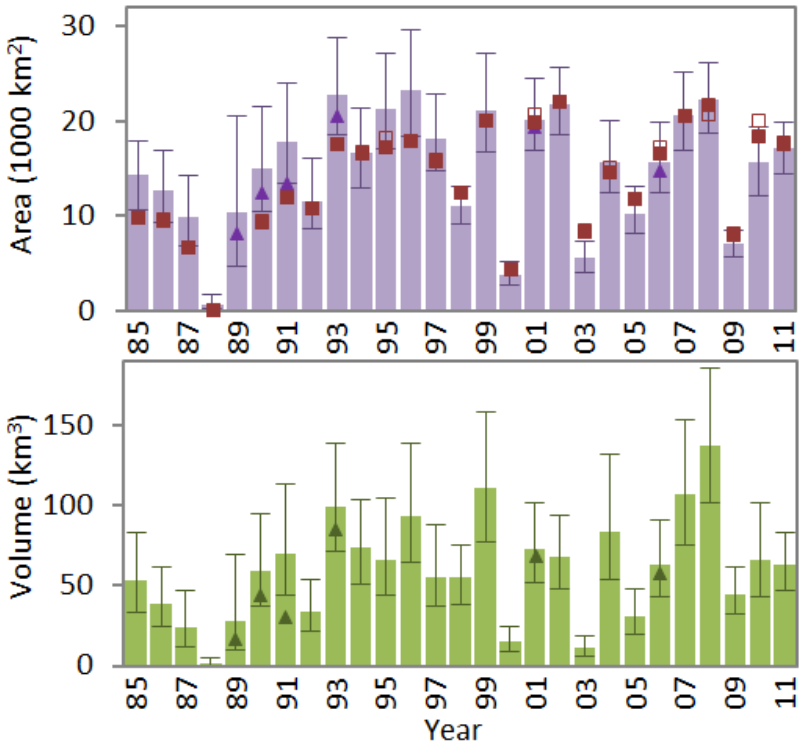

Figure 5-12: Bottom layer hypoxic extent estimates with 95% confidence intervals by year; estimates prior to making adjustments for instrument bias as triangles; previous LUMCON area estimates as open squares; revised LUMCON area estimates as solid squares.

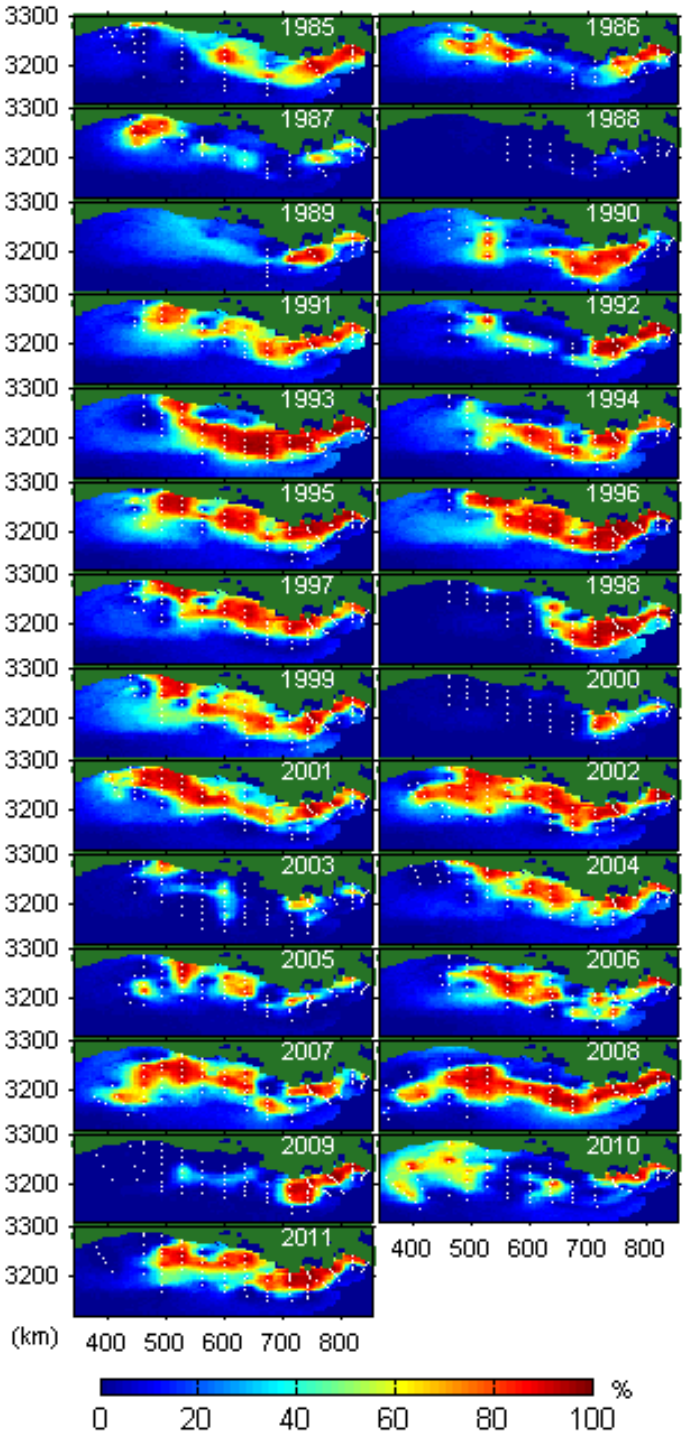

Figure 5-13: Maps showing modeled probability of hypoxia, 1985-2011

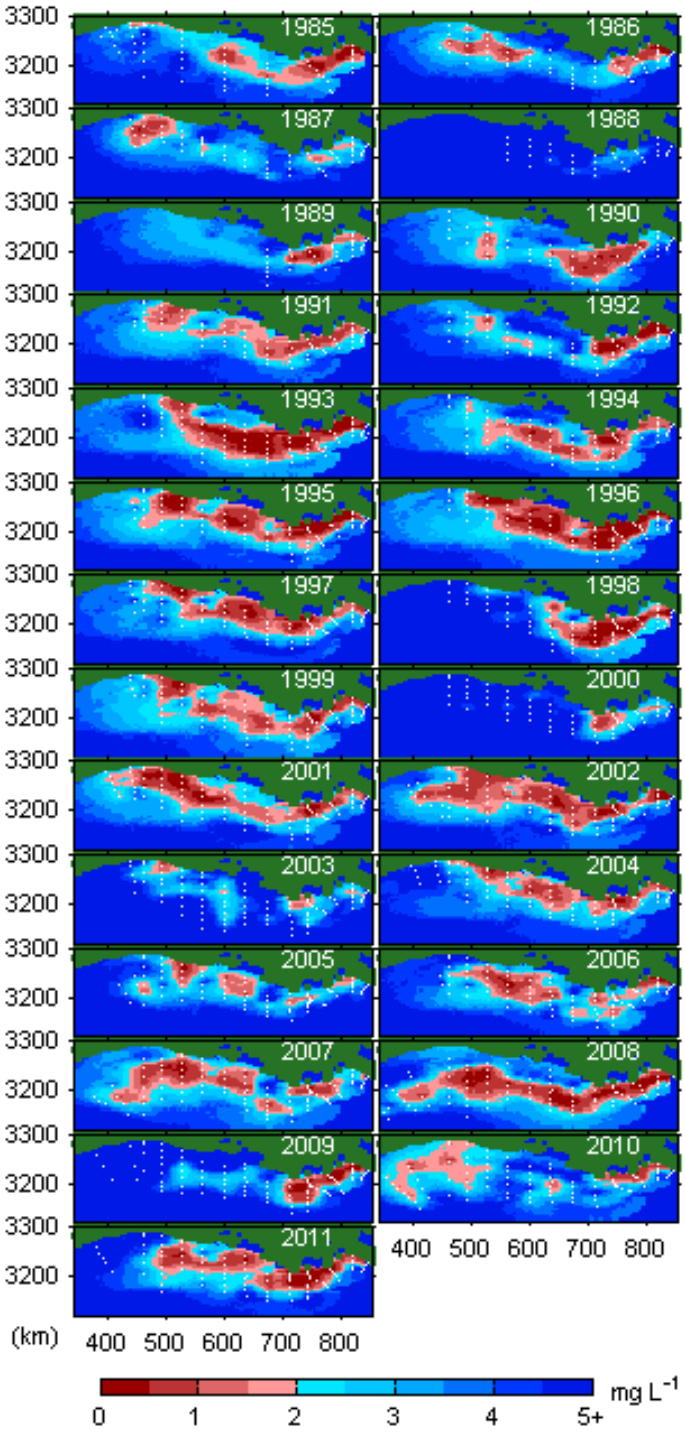

Figure 5-14: Maps showing modeled median BWDO concentration, 1985-2011

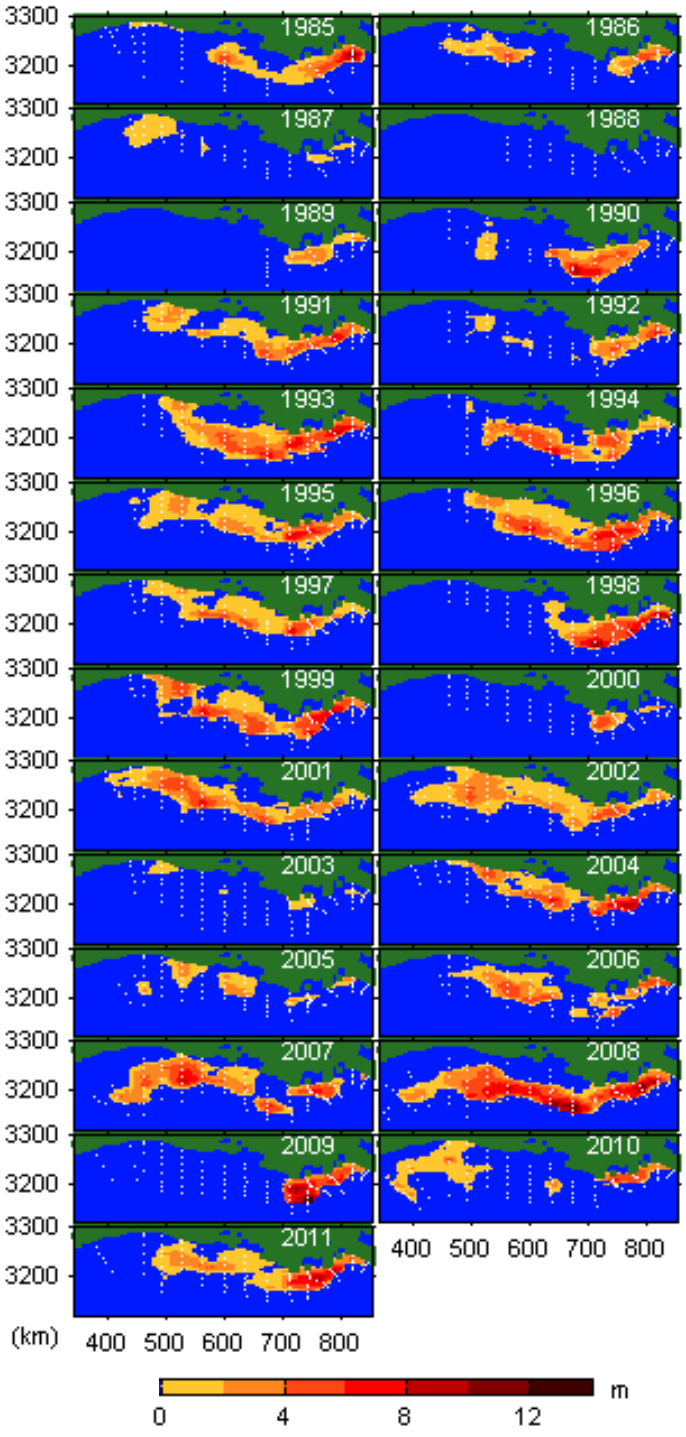

Figure 5-15: Maps showing modeled median BW hypoxic thickness, 1985-2011

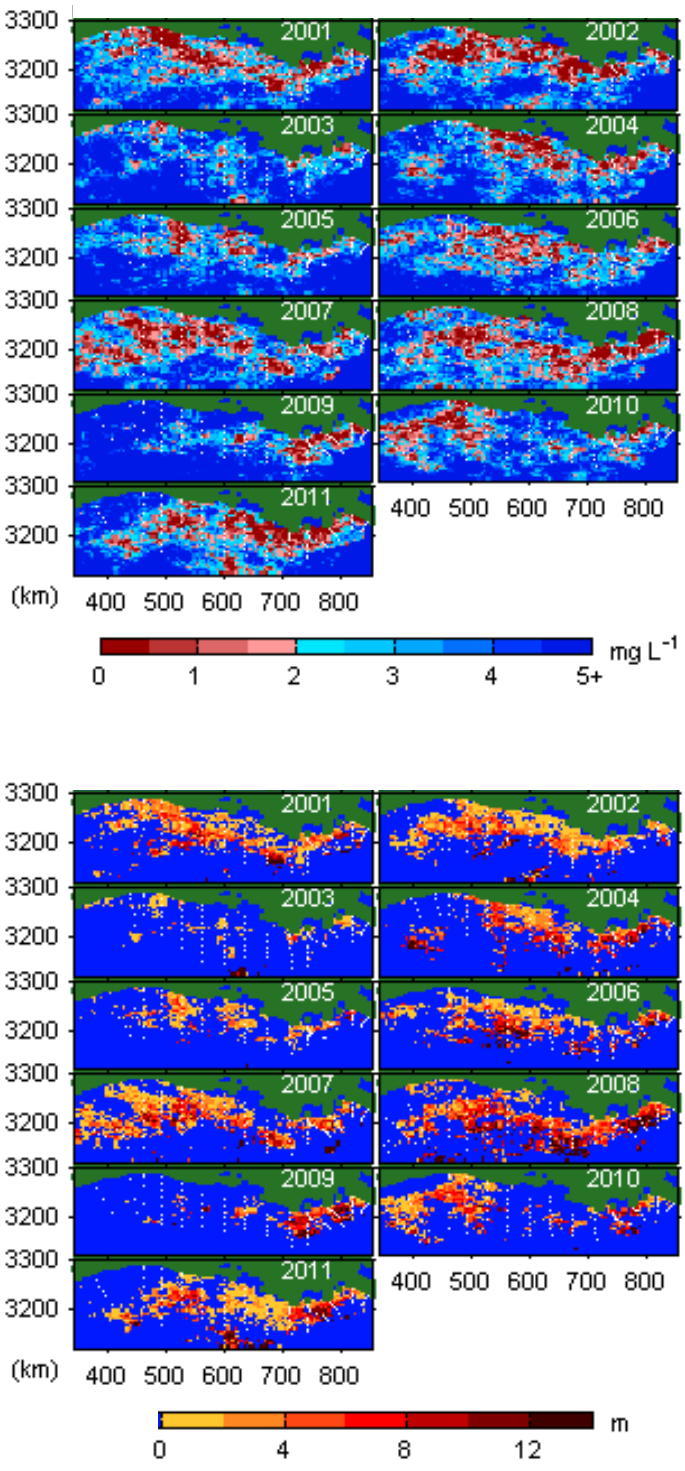

Figure 5-16: Maps showing example conditional realizations for BWDO (top) and hypoxic thickness (bottom), 2001-2011

Table 5-3: Tabulated hypoxic area and volume estimates, 1985-2011

Year	Area (1000 km^2)				Volume (km^3)			
	mean	median	2.5 perc	97.5 perc	mean	median	2.5 perc	97.5 perc
1985	14.3	14.4	10.6	18.0	53.6	52.2	32.8	83.4
1986	12.7	12.5	9.4	17.0	39.0	37.5	24.3	61.7
1987	9.8	9.6	6.8	14.3	24.2	22.5	12.1	46.7
1988	0.7	0.6	0.2	1.7	1.3	0.8	0.0	5.3
1989	10.4	9.7	4.7	20.6	28.4	24.0	10.2	69.6
1990	15.0	14.6	10.5	21.6	59.3	57.2	37.1	94.7
1991	17.9	17.6	13.4	23.9	70.0	67.4	44.3	112.9
1992	11.6	11.3	8.7	16.1	33.7	32.5	22.0	54.2
1993	22.7	22.4	18.6	28.7	99.5	97.0	71.2	139.2
1994	16.6	16.4	12.9	21.4	73.8	72.4	51.2	103.9
1995	21.3	20.9	17.0	27.2	66.3	64.1	44.0	105.0
1996	23.2	22.9	18.5	29.6	92.9	89.5	64.3	139.0
1997	18.2	17.9	14.7	22.9	54.8	52.1	36.8	88.2
1998	11.1	11.0	9.2	13.2	54.9	53.6	38.5	74.8
1999	21.2	21.0	16.8	27.1	111.3	108.9	77.2	158.6
2000	3.8	3.7	2.7	5.2	15.0	14.4	8.5	24.2
2001	20.1	19.9	16.8	24.5	73.0	71.6	52.2	101.9
2002	21.7	21.6	18.5	25.7	67.6	65.6	48.3	94.3
2003	5.5	5.5	4.1	7.3	11.3	11.0	6.1	18.5
2004	15.7	15.4	12.5	20.0	83.6	80.0	54.2	131.6
2005	10.2	10.2	8.1	13.1	30.7	29.4	19.7	47.9
2006	15.6	15.5	12.5	19.9	62.6	60.9	42.9	90.9
2007	20.6	20.6	17.0	25.1	107.0	104.9	75.1	153.4
2008	22.3	22.3	18.8	26.2	137.5	135.2	101.3	185.5
2009	7.1	7.0	5.7	8.5	44.3	43.6	32.5	61.3
2010	15.6	15.5	12.1	19.4	65.6	63.9	42.8	102.1
2011	17.1	17.1	14.4	19.9	62.6	61.5	47.0	83.1
average	14.9	14.7	11.7	19.2	60.1	58.3	40.6	90.1

The hypoxic extent can be determined using a variety of different interpolation and simulation-based methods. Figure 5-17 compares the preferred area estimates (Table 5-3) to areas inferred from less-optimal methods. This comparison includes interpolation ('kriged') estimates, which are consistently lower than the results determined from other methods, as discussed further in the next section.

This comparison (Figure 5-17) also includes estimates developed using CRs from an OK formulation (without trend variables). The 'OK' estimates have average confidence intervals more than twice as wide as those from UK; and the OK hypoxic area and volume estimates are 53% and 121% greater than the UK estimates, respectively. OK tends to over-estimate the extent of hypoxia outside of the sampling cruise envelope, because unlike UK, OK does not use trend variables to represent large-scale spatial patterns in DO and hypoxic fraction. These trends (e.g., Figure 5-8) generally indicate that conditions become less hypoxic as one moves away from the most intensively sampled areas of the shelf.

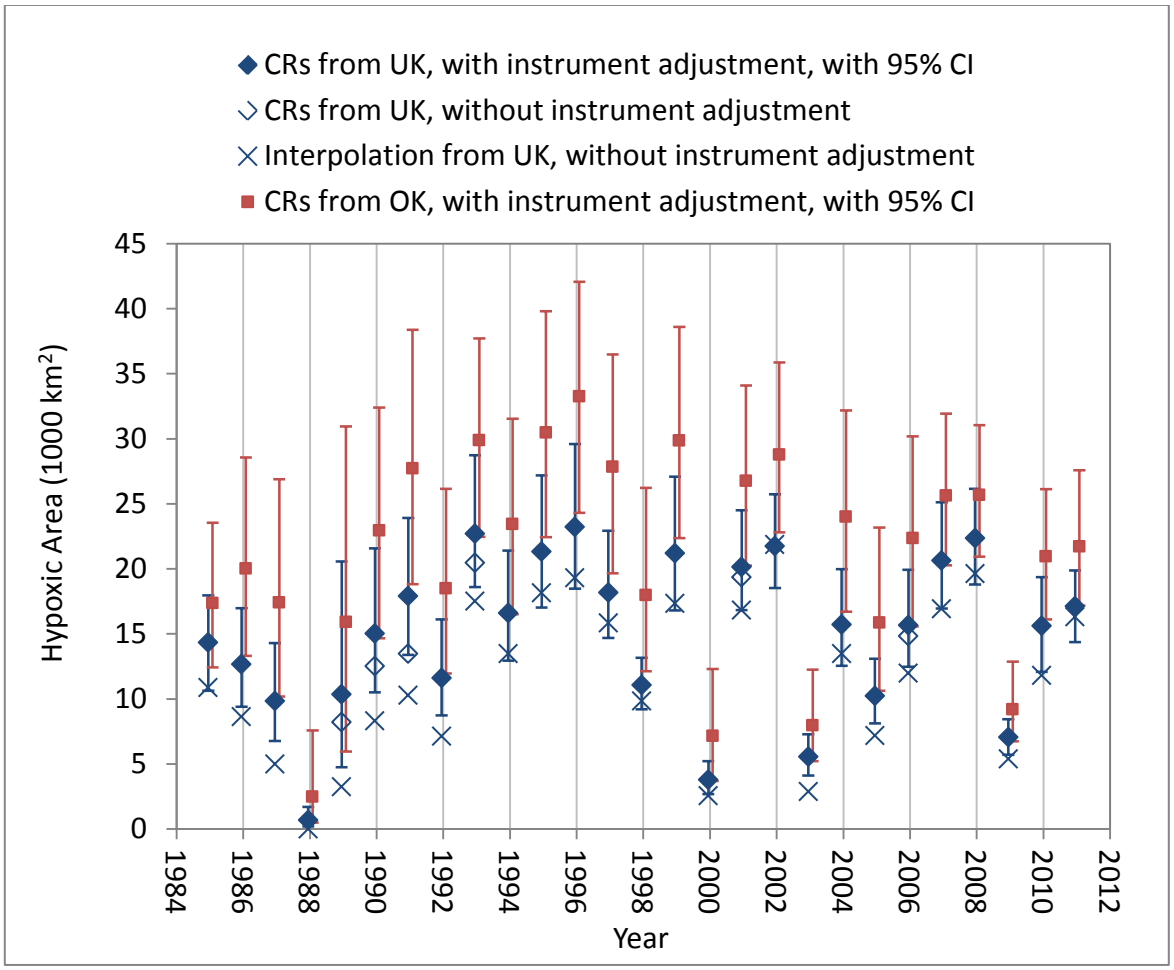

Figure 5-17: Hypoxic area estimates developed using different geostatistical methods (solid blue diamonds represent preferred estimates)

5.4. Discussion

The geostatistical modeling results can be used to assess temporal trends in hypoxic zone size for the 27-year study period. Hypoxic volume increased by an average of 2.3% per year as a linear trend with time (1.4 $km^3\ yr^{-1}$, percent increase determined by dividing by the 27-year mean hypoxic extent), but this trend was not significant (p=0.12). Hypoxic area increased at a lesser rate of 0.9% per year (140 $km^2\ yr^{-1}$) and was also not significant (p=0.42). However, the relatively large increase in volume relative to area reflects a significant increasing trend in hypoxic layer thickness (1.8% per year, 0.069 m yr^{-1}, p=0.05). Note that trend significance is affected by the uncertainty in the geostatistical estimates; trend significance is determined using a Monte Carlo approach where temporal trend coefficients are developed for each of the 1000 sets of CRs, and an overall probability distribution is developed by sampling from the uncertainty in these trend coefficients (note p-values are based on a two-sided test). Without accounting for uncertainty, the volume, area, and thickness trends would have p-values of 0.09, 0.38, and 0.01, respectively.

The new hypoxic area estimates can be compared to the previous hypoxic area estimates (LUMCON 2012, Rabalais et al. 2007). For the 27-year study period, the previous area estimates increased at a highly significant rate of 2.6% per year (360 $km^2\ yr^{-1}$, p=0.01), substantially greater than the 0.9% rate for the new estimates. This dissimilarity is due primarily to differences in estimates for the earlier years of the study period. For 1985-1987, the new estimates are consistently higher than the previous estimates primarily because the geostatistical methodology accounts for the possibility of hypoxia occurring outside the envelopes of the cruises, which were relatively small in these years. For 1990, 1991, and 1993, the new estimates are also higher because of the instrument bias adjustments developed in this study. For 1985-1993 (except 1989, for which there is no previous estimate), the mean geostatistically-determined hypoxic area is 39% (3650 km^2) greater than that of the previous estimates.

For 1994-2011, the new area estimates are in general agreement with the previous area estimates. Over this time period, the two sets of estimates are highly correlated (r^2=0.88),

the means of the two datasets are of negligible difference, and only three of the previous estimates (1996, 2003, and 2010) fall outside the 95% confidence intervals of the new estimates (Figure 5-12). This suggests that when the biasing issues noted above are avoided, there is approximate agreement between the geostatistical and hand-contouring estimates.

The new extent estimates have implications for our understanding of how hypoxia is changing over long temporal scales. Multiple studies using nutrient loads to predict the previous hypoxic area estimates have suggested the Gulf is becoming increasingly susceptible to hypoxia, based on increasing hypoxic area relative to nutrient loading (Forrest et al. 2011, Greene et al. 2009, Liu et al. 2010, Turner et al. 2008). (Nutrient loading increased greatly in the 1970s but remained relatively stable throughout the study period (Goolsby and Battaglin 2001, Scavia and Donnelly 2007).) Increases in hypoxic susceptibility have also been suggested in other studies of the Gulf and other coastal systems (Conley et al. 2009, Kemp et al. 2009, Stow et al. 2005). The new hypoxic area estimates, however, exhibit relatively little increase over time, especially when compared to the previous estimates, potentially suggesting less system change during the study period than previously thought. But, the new hypoxic thickness and volume estimates do increase to a greater degree, potentially suggesting a more vertically-oriented increase in hypoxic extent. Future studies could focus on re-calibrating existing nutrient-loading models to these new hypoxic extent estimates to develop a refined understanding of how Gulf hypoxia may be changing over time.

Uncertainties in the new hypoxic extent estimates, represented by the 95% confidence intervals (Figure 5-12), reflect the limited spatial scope and resolution of shelfwide cruise sampling. Uncertainties are generally greatest in the earlier years when cruises were smaller and did not always use instruments that reached the sea floor. From 1985-1993, the mean relative standard error for hypoxic area was 23%, but it decreased to 11% for 1994-2011. Uncertainties also appear to be larger in years with relatively severe hypoxia (i.e., lower average BWDO and larger hypoxic area), likely because more of the estimation grid is subject to the possibility of hypoxia (in years with higher average BWDO, most of the estimation grid is determined to be well above the hypoxic threshold, such that

simulated values rarely fall below the hypoxic threshold). In the future, the modeling framework presented here could be used to evaluate different sampling designs based on how well they constrain estimate uncertainty. The modeling approach could also be used to compare hypoxic extent results and associated uncertainties using hypoxic threshold choices other than 2 mg L^{-1}.

Regardless of the precision of the estimates, it is important to remember that they represent conditions at only one point in time (i.e., during the shelfwide cruises), and that hypoxic extent can vary substantially throughout the summer due to changes in organic matter production, wind-driven mixing events, and fluctuating current patterns (Hetland and DiMarco 2008, Wang and Justic 2009, Wiseman et al. 1997). As a result, these estimates do not necessarily reflect hypoxic conditions over the entire summer (though back-to-back cruises have demonstrated fairly consistent hypoxic areas under stable weather conditions (Rabalais et al. 1999)). More detailed monitoring and biophysical modeling are needed to better understand the short- and long-term dynamics of hypoxia formation, and to mechanistically interpret the temporal variability in the extent estimates presented here.

The primary feature of the geostatistical approach, when compared to more traditional, interpolation-based approaches, is the use of simulations (i.e., CRs) (Zhou et al. 2013). The most obvious benefit is the ability to quantify the uncertainty in extent estimates. A second benefit is that CRs provide more realistic extent estimates than can be derived from kriging (or Gauss-Markov smoothing (Emery and Thomson 2001)) alone. In this study, hypoxic area estimates derived from kriged maps (Figure 5-17) were substantially lower than both our CR estimates and the LUMCON estimates. This is because the kriged maps tend to characterize large portions of the estimation grid as slightly above the hypoxic threshold. However, due to the stochasticity of the system, these locations still have some probability of being hypoxic. The CR approach, which samples from the uncertainty in the system, accounts for this possibility, and thus results in larger hypoxic area estimates. This is consistent with Chiles and Delfiner (1999), who argue that CR, rather than kriging, is the more appropriate approach for determining spatially aggregated quantities. A third advantage of the CR approach is that it provides a framework for performing instrument

bias adjustments probabilistically, such that adjustment uncertainty is propagated to the hypoxic extent estimates. Finally, CR was fundamental to probabilistically determining hypoxic volume (in addition to area). The two-step CR approach, developed here, can be compared to other volume estimation methods, such as multi-layer kriging (Murphy et al. 2010) and three-dimensional CR (Zhou et al. (submitted manuscript)), which have been applied in Chesapeake Bay. A benefit of our approach is that it allows uncertainty quantification of hypoxic volume within a relatively simple (two-dimensional) geostatistical framework.

This study also demonstrates the benefits of including trend variables within the geostatistical model (i.e., the UK formulation) because the deterministic trends help reduce model uncertainty and result in more realistic extent estimates (Figure 5-17). The advantages of UK have been demonstrated previously for Lake Erie and Chesapeake Bay hypoxia (Murphy et al. 2010, Zhou et al. 2013), but are perhaps even more salient for an open system such as the Gulf shelf. When trends in DO are not modeled, it has been necessary to limit the estimation grid around the bounds of the sampling cruise (NOAA 2012, Rabalais et al. 1999), such that the size of the cruise can potentially bias the inferred hypoxic area. By including trend variables that explain the large-scale spatial patterns in BWDO and BWHF, it is possible to develop realistic CRs of DO across the entire study area, so the same estimation grid can be used for all cruises.

Finally, the results confirm that hypoxic area on the Louisiana-Texas shelf greatly exceeds the Hypoxia Task Force goal of 5000 km^2 as a five-year running average (EPA 2001, 2008). The most recent five-year period from our study, 2007-2011, has a mean hypoxic area of 16,600 km^2 with a 95% confidence range of 15,100-18,000 km^2. Clearly, additional management measures are required if the hypoxic extent is to be reduced to comply with the Task Force goal.

CHAPTER 6: Mechanistic Model

A mechanistic model is developed to predict mid-summer bottom water dissolved oxygen (BWDO) concentration and hypoxic area on the Louisiana-Texas shelf of the northern Gulf of Mexico (1985-2011). Because of its parsimonious formulation, the model possesses many of the benefits of simpler, more empirical models, in that it is computationally efficient and can rigorously account for uncertainty through Bayesian inference. At the same time, it incorporates enough biophysical realism that its parameterization can be informed by field-measured biological and physical rates. The model is used to explore how freshwater flows, nutrient loads, and winds affect hypoxia on western and eastern sections of the shelf, delineated by the Atchafalaya River outfall. The text of this chapter is reproduced in part from unpublished work (in review):

Obenour, D. R., A. M. Michalak, and D. Scavia. Probabilistic prediction of dissolved oxygen on the east and west Louisiana-Texas shelf.

6.1 Introduction

Quantitative modeling plays an important role in the management of Gulf hypoxia by synthesizing knowledge about the causes of hypoxia, and by predicting how the severity of hypoxia is affected by changing nutrient loads and other environmental factors (Justic et al. 2007, Scavia et al. 2004). Most existing Gulf hypoxia models can either be categorized as 'simple' in that they are more empirical and have little (or no) spatial detail (Greene et al. 2009, Turner et al. 2012), or 'complex' in that they are more mechanistic and provide richer spatial information (Bierman et al. 1994, Fennel et al. 2011). The purpose of this study is to develop and validate a model of intermediate complexity that is parsimoniously mechanistic and spatially two-dimensional (though at low resolution). The resulting model

possesses many of the benefits of simpler models, in that it can rigorously account for uncertainty, is computationally efficient, and can be readily applied to forecasting scenarios relevant to water quality management. At the same time, it incorporates enough biophysical realism that its parameterization can be informed by field-measured biological and physical rates.

Another difference between this study and previous Gulf modeling studies is that it makes use of new estimates of mean bottom water dissolved oxygen (BWDO) concentration and hypoxic area, as determined from a geostatistical model (Chapter 5). The geostatistical estimates were developed for mid-summer conditions, based on dissolved oxygen measurements collected by the Louisiana Universities Marine Consortium (LUMCON) during extensive monitoring cruises (Rabalais et al. 1999, Rabalais et al. 2007). The geostatistical approach circumvents biasing issues that were shown to affect previous hypoxic area estimates derived from these data. In addition, the geostatistical estimates include measures of uncertainty, related primarily to sampling coverage and types of sampling instruments used. These uncertainties are used in the present study, effectively giving more weight to estimates with lower uncertainty.

This work builds on findings from previous studies that used largely empirical models to identify key anthropogenic and environmental factors related to Gulf hypoxia formation. Many of these earlier models confirmed the relationship between mid-summer hypoxia and spring nutrient load (Scavia et al. 2003, Scavia and Donnelly 2007, Turner et al. 2006). However, other studies have shown that river flow, which is significantly correlated with nutrient load, is also an effective predictor of hypoxia, such that the relative roles of nutrients and flows cannot be completely disentangled using empirical models alone (Forrest et al. 2011, Wiseman et al. 1997). Flows affect the degree of water column stratification, which was found to be another important predictor of hypoxia, along with nutrients, in a recent geostatistical modeling study (Obenour et al. 2012b). The importance of both nutrients and stratification has also been demonstrated in mechanistic modeling of an intensely monitored location on the eastern shelf (Justic et al. 1996, 2002). In addition, two recent studies have found wind metrics to be useful in improving predictors of hypoxic area, likely because of wind's influence on transport and stratification (Feng et al. 2012,

Forrest et al. 2011). The present study provides a means of testing these relationships over a 27-year study period (1985-2011) on the east and west Louisiana-Texas shelf.

6.2 Materials and methods

6.2.1 Dissolved oxygen and hypoxic extent

Mid-summer mean dissolved oxygen concentration and hypoxic area are the primary response variables considered in this study. Results from a geostatistical model (Chapter 5) provide estimates of these variables, with associated uncertainties, for two sections of the Louisiana-Texas shelf, divided at the Atchafalaya River outfall location (Figure 6-1). The west shelf section extends from 342.5-672.5 km Universal Trans Mercator (UTM) easting, and the east shelf section extends from 672.5-837.5 km UTM easting. The shelf sections represented in the model are limited to depths of 3 to 80 meters, and to UTM northings greater than 3122.5 km. Overall, the study area covers the regions where mid-summer hypoxia is most commonly observed. The geostatistical model inputs include only LUMCON DO measurements, spatial coordinates, and bathymetry. No biophysical processes are represented in the geostatistical model, so the results provide an independent dataset for calibrating mechanistic models. The uncertainties in the geostatistical estimates are approximately normally distributed, and are represented as such in this study.

6.2.2 River flow and load data

The study uses United States Geological Survey (USGS) monthly flow and nutrient loading data for the Atchafalaya and Mississippi Rivers (http://toxics.usgs.gov/hypoxia/mississippi/nutrient_flux_yield_est.html). Although USGS provides loading data based on two different estimation methods, only the Adjusted Maximum Likelihood Estimator (AMLE) (Runkel et al. 2004) method includes results for the entire period of this study, and thus only AMLE results are used here. The flow and load estimates are available for each calendar month, and linear interpolation is used to recalculate the loads for 30-day averaging periods leading up to the starts of the annual shelfwide cruises. Previous studies have generally focused on TN or $NO_{2\text{-}3}$ loading data for modeling hypoxia (Scavia et al. 2004). While TN loads are naturally larger in

magnitude than NO_{2-3} loads, they tend to be highly correlated and yield similar results in more empirical models (Forrest et al. 2011). However, because this model is mechanistically derived, the magnitude of loading is important. TN includes NO_{2-3}, NH_3, and organic nitrogen (ON); and while NO_{2-3} and NH_3 are highly bioavailable, organic nitrogen is more recalcitrant. Based on studies suggesting that about 60% of Mississippi River ON is dissolved (Goolsby and Battaglin 2001), and about 20% of dissolved ON is photo-chemically converted to NH_3 on the shelf (Bushaw et al. 1996), 12% of river ON is represented as bioavailable in this study. Thus, total 'bioavailable' model load is calculated as $NO_{2-3} + NH_3 + 0.12 \times ON$.

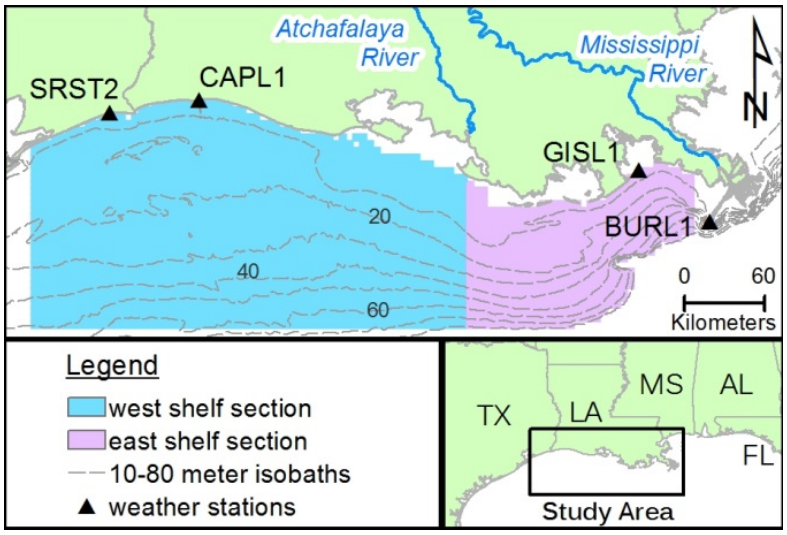

Figure 6-1: Louisiana-Texas Shelf study area

6.2.3 Coastal wind data

The study uses coastal wind data from the National Data Buoy Center (http://www.ndbc.noaa.gov/). There are two stations with fairly complete records for the period of interest: Sabine Pass (SRST2) and Southwest Pass (BURL1). Data from Calcasieu Pass (CAPL1) and Grand Isle (GISL1) are used to augment missing data for Sabine Pass and Southwest Pass, respectively, after applying appropriate bias corrections (station locations are shown in Figure 6-1). For brevity, we refer to the combined SRST2+CAPL1 dataset as the 'west' dataset, and the combined BURL1+GISL1 dataset as

the ‘east’ dataset. Both datasets cover about 95% of the examined period, and mean monthly values are substituted for missing data. Datasets are used to determine daily east-west wind velocities that are then averaged over 30-day consecutive periods leading up to the starts of the annual shelfwide cruises. In addition, weighted mean wind stresses (wind speed-squared) are determined for 14-day periods prior to the starts of the annual shelfwide cruises. Continuous monitoring data suggest that it takes approximately two weeks for DO to be re-depleted following wind mixing events (Rabalais et al. 2007, Walker and Rabalais 2006), and linearly decreasing weights (14 down to 1) are therefore assigned to wind stresses for each of the fourteen days preceding a cruise. (Seven and 21-day periods were also tested, but did not significantly alter model performance.) The wind stresses from the west and east weather stations are applied to the west and east shelf sections using inverse distance weighting to shelf section centroids.

6.2.4 Model formulation

Model segmentation is based on dividing the shelf east-west and vertically into four mixed reactors. The east-west division is at 672.5 km UTM easting (as described previously). The west and east shelf sections have areas of 48.5 and 14.0 Gm^2, and mean depths of 28 and 31 m, respectively. The shelf sections are vertically divided at the pycnocline into upper and lower layers. The exact depth of the vertical division is unimportant, given the steady-state modeling assumption used here.

The upper layers of the model receive flows and nitrogen loads from the Mississippi and Atchafalaya Rivers, partitioned between shelf sections using a transport submodel (described below). Nitrogen loads are converted to organic matter and fluxed to the lower layer based on an effective settling velocity (v_s). Overall, the differential equation for nitrogen in the surface layers is as follows:

$$\frac{dN}{dt} = Q_r C_{rN} + Q_u C_{uN} - \left(Q_r + Q_u + Q_g\right) C_N - A v_s C_N \qquad \text{eq 6-1}$$

Here, *dN/dt*, is the time rate of change for nitrogen mass in the shelf section, and variables Q_r, Q_u, and Q_g are flows ($Gm^3\ d^{-1}$) entering the system from the rivers, upstream

model segment, and Gulf, respectively. The variables C_{rN}, C_{uN}, and C_N represent nitrogen concentrations (mg L^{-1}) in the river, upstream model segment, and the subject model segment, respectively. Note that $Q_r C_{rN}$ is the river load (L_{rN}). A is the area of the shelf section.

The nitrogen that settles to the lower layer is generally associated with organic matter developed through primary production. The differential equation for organic matter (represented by carbon) in the lower layers is as follows:

$$\frac{dC}{dt} = Av_s R_{C:N} C_N - Vk_C C_C \qquad \text{eq 6-2}$$

Here, *dC/dt*, is the time rate of change for carbon mass in the lower shelf segment, V is segment volume (Gm^3), k_C is the first-order decay rate for organic carbon (d^{-1}), $R_{C:N}$ is the ratio of organic carbon to nitrogen, and C_C is the concentration (mg L^{-1}) of organic carbon in the segment. Note that it is assumed that there is no advective flux laterally or longitudinally because bottom water velocities are small relative to the temporal and spatial scales of this model (Rabalais et al. 1999).

Dissolved oxygen (DO) in the lower layer is lost through bacterial decomposition of organic matter and regained through diffusion and mixing of dissolved oxygen from the surface layer. The differential equation for DO in the lower layers is as follows:

$$\frac{dO}{dt} = Ak_a(C_{OS} - C_O) - Vk_C C_C R_{O:C}\omega - AC_O B/C_{OB} \qquad \text{eq 6-3}$$

Here, *dO/dt*, is the time rate of change for oxygen mass in the shelf section, k_a is the reaeration rate (m d^{-1}), C_{OS} is the oxygen concentration of the overlying surface layer (mg L^{-1}), $R_{O:C}$ is the ratio of oxygen demand to organic carbon, ω is an oxygen demand adjustment factor related to processes such as photosynthetic oxygen production, B is the benthic oxygen demand (g $m^{-2}d^{-1}$) at reference DO concentration C_{OB} (mg L^{-1}), and C_O is the DO concentration (mg L^{-1}) in the lower layer. As shown, both the reaeration and benthic oxygen demand terms are affected by the segment's dissolved oxygen

concentration. Reaeration increases as the gradient between surface and lower layer DO concentrations increase (Justic et al. 1996). Benthic oxygen demand increases as DO increases, and this relationship is approximated as linear, as suggested by Lehrter et al. (2012).

Equations 6-1, 6-2, and 6-3 can be solved together for C_O under steady state conditions ($dN/dt = dC/dt = dO/dt = 0$), yielding:

$$C_O = \frac{1}{k_a + B/C_{OB}} \left(K_a C_{OS} - \frac{R_{O:C} R_{N:C} \omega [Q_r C_{rN} + Q_u C_{uN}]}{([Q_r + Q_u + Q_g]/v_s + A)} \right) \quad \text{eq 6-4}$$

Eq 6-4 is the primary mechanistic model formulation used in this study. However, the flows and loads used in eq 6-4 are determined by a transport submodel; and k_a is determined by a reaeration submodel, both described below. Note that terms V and k_C (eq 6-3) cancel out when the system is solved for steady state conditions (eq 6-4), and water column DO demand is proportional to the organic matter flux from the surface layer. Also note that eq 6-4 predicts the DO concentration representative of the entire lower layer, whereas the geostatistical estimates are for BWDO (DO at the very bottom of the water column). Based on an analysis of DO profile data, the 20th percentile DO concentration (expected to be typical of the lower layer) is 0.5 mg L^{-1} higher than the BWDO concentration, on average, and this adjustment is included within the model (i.e., $C_{BWDO} = C_O - 0.5$).

6.2.5 Transport submodel

Coastal current patterns affect transport of freshwater and nutrients delivered to the Gulf by the Mississippi and Atchafalaya Rivers. Throughout much of the year, the dominant flow pattern is westward, and the strongest westward currents typically occur in spring, along the inner shelf, due to prevailing easterly winds and the buoyancy flux of river discharge (Walker et al. 2005, Zavala-Hidalgo et al. 2003, Zhang et al. 2012). In the model, transport of water and load is determined using a simple 'flow partitioning' equation:

$$F_e = 0.5 + \beta_e W_e \quad \text{eq 6-5}$$

Here, F_e is the fraction of water and load transported to the east, W_e is the mean east-west wind velocity (westerly winds are positive, easterly winds are negative), and β_e is a transport coefficient determined through model calibration. To ensure that F_e is constrained to within the range [0,1], W_e is constrained to a range of [-2,2] m s^{-1}, and β_e is constrained to be calibrated within a range of [0,0.25]. The fraction of water and load transported west is simply $(1 - F_e)$.

Spring flows and loads, partitioned by spring easterly wind velocity, are used to drive the mechanistic model (eq 6-4). In the spring, 80% of the Mississippi River discharge (flow and load) is assumed available for partitioning using eq 6-5, while the remaining 20% of Mississippi River discharge is assumed lost to the south or east under any wind condition (Walker et al. 2005, Zhang et al. 2012). Mississippi discharge partitioned to the east leaves the study area, while discharge partitioned to the west enters the east shelf section. The discharge from the Atchafalaya River and the gulf dilution flow (Q_g) are both also partitioned, such that the westward partition enters the west shelf section, and the eastward partition enters the east shelf section. Finally, flow and load within the east shelf section is partitioned, such that the westward partition enters the west shelf section and the eastward partition exits the model. Transport from the west shelf section back to the east shelf section could not be represented within a steady-state model. However, this type of eastward transport is expected to be rare in spring, as river flows become entrained in the dominantly westward spring shelf current (Wang and Justic 2009, Zhang et al. 2012).

6.2.6 Reaeration submodel

Reaeration is represented using a quasi-mechanistic formulation based on flow and wind stress:

$$k_a = \beta_{k0} + \beta_{k1}\tau/(Q_s/A) \quad \text{eq 6-6}$$

Here, k_a is the reaeration rate (m d^{-1}), τ is the 14-day weighted mean wind stress for the shelf section ($m^2 s^{-2}$), A is the area of the shelf section (Gm^2), and Q_s is the summer freshwater flow onto the shelf section ($Gm^3 d^{-1}$). The model follows the logic that reaeration should increase due to wind stress-induced mixing, but this mixing is inhibited by the density of freshwater on the shelf section (represented by Q_s/A). The terms β_{k0} and β_{k1} are essentially empirical parameters determined through model calibration.

The summer freshwater flow, Q_s, is determined using the Atchafalaya River discharge partitioned between the east and west shelf sections based on the transport submodel. Mississippi River discharge is not included in Q_s because relatively little Mississippi River flow enters the shelf in summer due to reversal of the westward coastal current (Walker et al. 2005, Zhang et al. 2012). Stable isotope studies suggest Atchafalaya River discharge constitutes the majority of freshwater in the surface waters of the shelf by mid-summer (Strauss et al. 2012).

6.2.7 Spring versus summer conditions

The preceding descriptions make reference to 'spring' and 'summer' flows and loadings. The spring flows and loadings are expected to control the production of organic matter on the shelf, as represented in eq 6-4. Here, spring flows, loads, and east-west winds are determined as weighted averages of 60-90 and 30-60 day periods prior to the starts of the shelfwide cruises, with the earlier period receiving twice the weight of the later period. Because most cruises start in late July, this roughly corresponds to a period of late April-late June, but more heavily weighted toward the beginning of this period. This period approximately coincides with previous modeling studies suggesting May or May-June nutrient loads correlate best with hypoxic zone size (Greene et al. 2009, Scavia et al. 2004, Turner et al. 2012).

Summer flows are expected to be related to the intensity of stratification on the shelf at the time of the shelfwide cruises, and are thus incorporated into the reaeration submodel (eq 6-6). Here, summer flows and east-west winds are determined as weighted averages of the 30-60 and 0-30 day periods prior to the starts of the shelfwide cruises, with the later period receiving twice the weight of the earlier period. Because most cruises start in late

July, this roughly corresponds to a period of late May-late July (but more heavily weighted toward the end of this period). This is reasonable because the total flow from this period correlates well (r^2=0.8) with the intensity of stratification determined from a previous study by Obenour et al. (2012b).

6.2.8 Prior information and model calibration

The mechanistic model includes several parameters determined through calibration or specified as known. The parameters listed in Table 6-1 are calibrated within the model by Bayesian inference, implemented using the WinBUGS program (Lunn et al. 2000) called from R via the R2WinBUGS software package (Gelman and Hill 2007). The Bayesian model calibration approach has multiple benefits, in that it can incorporate prior information, rigorously account for parameter and data uncertainty, and be applied to non-linear model formulations (e.g., eq 6-4). Prior information for model parameters is represented as probability distributions, as shown in Table 6-1. For most of these parameters, the priors are wide uniform distributions that are nearly non-informative (they are only narrow enough to prevent the calibration from occasionally reverting to unrealistic local minima). However, an informative normal prior distribution is used for benthic oxygen demand, based on recent research by Lehrter et al. (2012), who performed an extensive study of shelf sediment fluxes, and determined a mean benthic oxygen flux of 0.28 g $m^{-2}d^{-1}$ with standard error 0.06 g $m^{-2}d^{-1}$. A less informative normal prior was used for ω, which represents an adjustment in water column oxygen demand related to a number of factors. The prior mean (0.5 [unitless]) reflects studies suggesting that upwards of 40% of sub-pycnocline oxygen demands are offset by photosynthetic oxygen production (Lehrter et al. 2009, Rowe 2001). Other factors that could be reflected in ω include off-shelf losses, nutrient recycling between layers, and inaccuracies in assumptions regarding load utilization (fraction of load that enters the shelf, fraction that is bioavailable, etc.). There is no strong evidence for the importance of these other factors, but the uncertainty in this prior reflects their potential relevance. Finally, a moderately informative uniform prior is used for r_{Qg}, the ratio of Gulf flow to mean Mississippi River flow ($Q_g = r_{Qg} \times \text{mean}[Q_{Miss}]$). A previous study suggests this ratio is approximately 5-6 [unitless] under

easterly winds (Walker et al. 2005). However, because river flow does not mix completely with Gulf flow, the effective dilution may be lower, and thus a uniform prior of [1,6] is used.

Some modeling parameters can be reasonably specified as known because the uncertainty associated with these parameters is relatively small, and calibrating them would be computationally expensive while providing little additional scientific insight. The ratio of carbon to nitrogen, $R_{C:N}$, is based on the Redfield Ratio (5.68 gC gN^{-1}) (Redfield et al. 1963). The ratio of oxygen demand to carbon, $R_{O:C}$, is based on stoichiometric relationships for the aerobic decomposition of organic matter (3.5 g O gC^{-1}) (Chapra 1997, Justic et al. 1996). A surface layer oxygen concentration, C_{OS}, of 7.5 mg L^{-1} is used based on an examination of surface layer DO data. The reference oxygen concentration for benthic oxygen demand, C_{OB}, is simply the DO concentration (3 mg L^{-1}) corresponding to the prior information for B (Lehrter et al. 2012).

The model also makes use of prior information for modeling output, particularly the vertical (downward) organic matter flux ($v_s R_{C:N} C_N$ per eq 6-2). Sediment trap experiments by Redalje et al. (1994), suggest east shelf summer carbon fluxes of 0.18-0.40 gC $m^{-2}d^{-1}$. Thus, east shelf carbon flux is calibrated to an 'observed' mean carbon flux, represented as a normal distribution, N(0.29,0.05) gC $m^{-2}d^{-1}$.

Both mechanistic model uncertainty and geostatistical 'observation' uncertainty are accounted for within the Bayesian calibration framework using the following relationship:

$$C_{geo(i,j)} \sim N\left(C_{mech(i,j)}, \sqrt{\sigma^2_{geo(i,j)} + \sigma^2_{mech(j)}}\right) \qquad \text{eq 6-7}$$

In eq 6-7, i and j represent the cruise year and shelf section, respectively. Terms $C_{geo(i,j)}$ and $C_{mech(i,j)}$ represent the mean BWDO concentrations predicted by the geostatistical model and mechanistic model, respectively. Mechanistic model (residual) uncertainties for the two shelf sections, $\sigma_{mech(j)}$, are parameters determined through the Bayesian calibration process, using an effectively uninformative prior distribution, U(0.1,3) mg L^{-1}.

Geostatistical uncertainty, $\sigma_{geo(i,j)}$, is specific to each year and shelf section, as determined by the geostatistical model.

Table 6-1: Prior information for mechanistic model parameters to be calibrated by Bayesian inference

symbol	description	prior	unit
v_s	effective settling velocity	U(0.01,1.0)	m d^{-1}
ω	oxygen demand adjustment	N(0.5,0.2)	-
r_{Qg}	ratio of Gulf flow to mean Mississippi flow	U(1.0,6.0)	-
B	benthic oxygen demand	N(0.28,0.06)	g m^{-2} d^{-1}
β_e	flow partitioning	U(0,0.25)	-
β_{k0}	reaeration intercept term	U(0,0.5)	m d^{-1}
β_{k1}	reaeration term modifying $\tau/(Q_s/A)$	U(0.05,0.5)	-

Note: U(lower bound, upper bound) ~ uniform distribution
N(mean, standard deviation) ~ normal distribution

6.2.9 Predicting hypoxic area from DO

Mean BWDO results can be converted to hypoxic area using a linear regression between the BWDO and hypoxic area values from the geostatistical model. Because the relationships between mean BWDO and hypoxic area are nonlinear, both mean BWDO and mean BWDO-squared are used as predictors. The resulting regressions for west and east shelf hypoxic area are as follows:

$$A_w = 58230 - 18390(C_{BWDO,w}) + 1460(C_{BWDO.w})^2 \quad \text{eq 6-8}$$

$$A_e = 16950 - 5530(C_{BWDO,e}) + 440(C_{BWDO.e})^2 \quad \text{eq 6-9}$$

These regressions explain 98.5% and 98.7% of the variability in hypoxic area on the west and each shelf section, respectively, as estimated using the geostatistical model (Chapter 5).

6.3 Results

6.3.1 Model calibration and validation results

Seven mechanistic model parameters are calibrated probabilistically through Bayesian inference. The likely range for each parameter is represented by its posterior distribution (Figure 6-2). Three of these parameters, v_s, ω, and B, primarily control oxygen dynamics within the model. The best estimate (i.e., mean of the posterior distribution) for effective settling velocity, v_s, is 0.19 m d^{-1}. This rate corresponds to 11-17 meters of settling over 2-3 months, suggesting a substantial flux of organic material through the pycnocline, consistent with the hypothesis that spring nutrient loads affect mid-summer hypoxia. The best estimate for the oxygen demand adjustment parameter, ω, is 0.37, suggesting that factors such as photosynthesis or off-shelf losses substantially reduce the effective oxygen demand of fluxed organic material. The best estimate for benthic oxygen demand, B, is 0.33 g $m^{-2}d^{-1}$, shifted slightly higher than the prior distribution (mean of 0.28 g $m^{-2}d^{-1}$) determined from the extensive field study by Lehrter et al. (2012). This is consistent with other field studies also documenting higher mean benthic oxygen demands (Dortch et al. 1994, McCarthy et al. 2013, Rowe 2001).

The parameters β_e and r_{Qg} are related primarily to model hydrodynamics. The best estimate for the flow partitioning parameter, β_e, is 0.22, such that transport ranges from 94% eastward to 94% westward under mean east-west winds ranging from 2 to -2 m s^{-1}, respectively. (Note that β_e was constrained by prior information to a maximum of 0.25, because higher values would unrealistically result in flow partitions greater than 100%). The best estimate for the ratio of Gulf dilution flow to mean Mississippi River flow, r_{Qg}, is 3.8, but the posterior distribution is very similar to the prior distribution (in terms of both mean and variance), suggesting this parameter could not be well resolved within the model. While r_{Qg} could be set to a fixed value without significantly affecting model

performance, allowing the parameter to vary acknowledges a source of mechanistic uncertainty within the model.

The reaeration parameters, β_{k0} and β_{k1}, are used to determine the reaeration rate, k_a. Under conditions of no wind (or infinite freshwater flow), k_a is equal to β_{k0}. As freshwater flows decrease and wind stresses increase, k_a increases as a function of β_{k1} (eq 6-6). Based on the calibrated model, reaeration rates (by cruise) range from 0.16-0.53 and 0.14-0.37 m d^{-1} for the west and east shelf sections, respectively. Overall, mean reaeration rates for the west and east shelf sections are 0.23 and 0.17 m d^{-1}, respectively, and their distributions are highly right-skewed. The highest reaeration rates are for 1988, due to the combined effects of low freshwater flows (drought year for the Mississippi River basin) and high winds (tropical storm Beryl). The mean rates from this study are higher than the 0.1 m d^{-1} mean July reaeration rate from a modeling study by Justic et al. (1996) for a location near the center of the east shelf, where stratification would be expected to be particularly severe.

The statistical modeling framework allows for examination of correlation among parameters. The largest correlation is between ω and β_{k0} (r^2=0.65). This correlation is expected because these parameters have similar but opposite effects on DO levels. Other parameter pairings with notable correlations (r^2>0.5) are v_s and r_{Qg}, B and β_{k0}, and B and β_{k1}. Correlation among parameters could be avoided by setting some parameters to fixed values, but by allowing these correlations, uncertainty about the relative importance of mechanistic drivers is represented within the model.

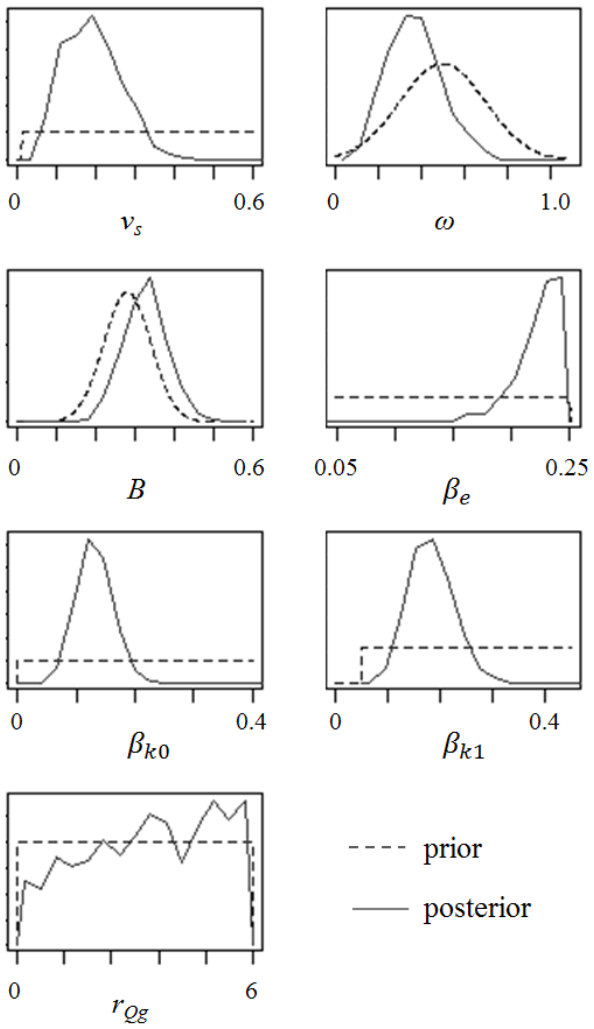

Figure 6-2: Prior and posterior probability distributions for calibrated mechanistic model parameters (as described in Table 6-1)

Overall, the model explains approximately 75% of the variability in BWDO concentration on each of the two shelf sections (Figure 6-3). Through the Bayesian calibration process, mechanistic model residual standard deviations, $\sigma_{mech(j)}$, were determined to be 0.35 and 0.31 mg L^{-1} for the west and east shelf sections, respectively. To test the robustness of the model, a leave-one-out cross validation (CV) was also performed. In CV, the mean BWDO for each cruise is predicted after removing that cruise from the calibration dataset and re-calibrating the model to the remaining data. Thus, the model's CV performance is a better measure of how well it will perform when predicting future conditions. In CV mode, the model explains 72% of the variability in BWDO on

each of the two shelf sections (Figure 6-4). Thus, performance is not greatly diminished relative to the full-model (Figure 6-3), suggesting the model is robust (i.e., not over-parameterized).

The 27-year mean vertical carbon fluxes determined by the model are 0.14 and 0.27 gC $m^{-2}d^{-1}$ for the west and east shelf sections, respectively. The modeled east shelf carbon flux conforms well with prior information for east shelf carbon flux, N(0.29,0.05) gC $m^{-2}d^{-1}$, as described previously. The carbon flux on the west shelf is expected to be lower, as it is further removed from the river outfalls, on average. Modeled carbon fluxes can also be compared to measured water column respiration rates. An extensive study by Murrell et al. (2013) reports lower-layer water column respiration rates averaging 0.22 gO $m^{-3}d^{-1}$ over the entire shelf. Assuming a 10-m thick lower layer and an $R_{O:C}$ of 3.5 gO gC^{-1}, this is equivalent to an areal rate of 0.63 gC $m^{-2}d^{-1}$, considerably higher than the modeled carbon flux rates. This may be due to lower-layer photosynthesis, resulting in higher measured rates of both oxygen production and respiration in the sub-pycnocline (but no net increase in oxygen demand). Also, respiration rates and the relative contributions of surface and benthic respiration have varied substantially across different Gulf field studies (Dortch et al. 1994, McCarthy et al. 2013, Quinones-Rivera et al. 2007), suggesting uncertainty related to experimental design or analytical methods.

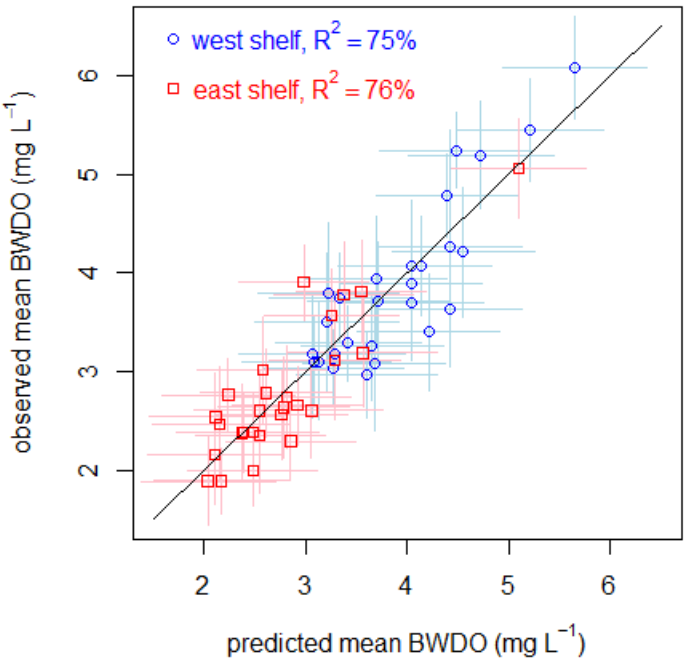

Figure 6-3: Observed (geostatistical) mean BWDO versus full-model mechanistic model predicted BWDO for the west and east shelf sections, with 95% prediction intervals

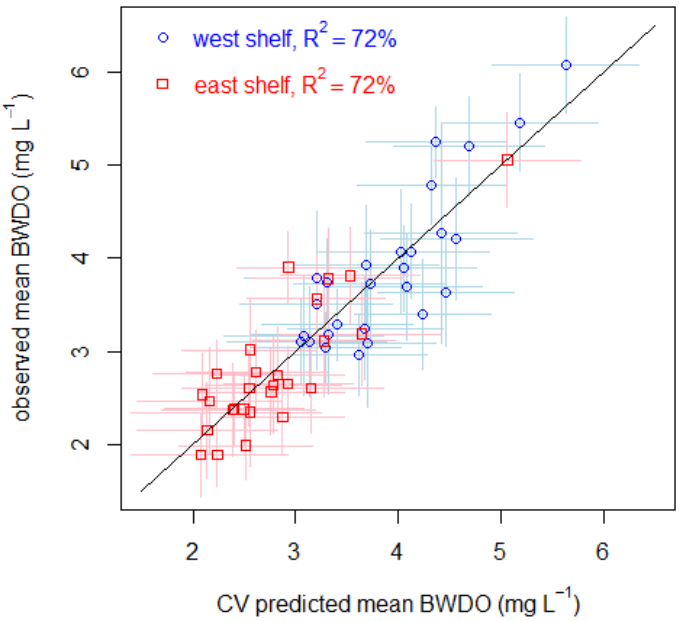

Figure 6-4: Observed (geostatistical) mean BWDO versus CV mechanistic model predicted BWDO for the west and east shelf sections, with 95% prediction intervals

6.3.2 Model sensitivity to stratification and nutrient loading

The model includes 'spring' inputs related to seasonal nutrient loading and 'summer' inputs related to stratification. By holding one of these input sets constant (at 27-year mean values), while allowing the other to vary, it is possible to examine the relative roles of these two drivers of hypoxia. As shown in Figure 6-5, both drivers substantially impact the year-to-year variability in mean BWDO. An 'influence metric' for quantifying the impact is determined by calculating the standard deviation of the 27 predicted BWDO values, under the different model input conditions. The influence metrics for nutrient inputs (holding summer flows and winds constant) and stratification inputs (holding spring flows, loads, and winds constant) are 0.24 and 0.53 mg L^{-1} for the west shelf, respectively; and 0.33 and 0.46 mg L^{-1} for the east shelf, respectively. These results suggest a somewhat larger role for stratification in explaining year-to-year variability in BWDO. However, when the uncertainty in the mechanistic model parameters is accounted for, the differences between the nutrient and stratification influence metrics are only significant for the west shelf (there is a 13% probability that the role of nutrient exceeds that of stratification on the east shelf). Stratification on the west shelf appears heavily influenced by weather conditions resulting in minimal stratification in 1988 due to drought, and in 1998, 2000, and 2009, due to unusually strong westerly winds, as has been noted in previous studies (Rabalais et al. 2007, Turner et al. 2012).

It should be emphasized that the 'nutrient effects' presented in this analysis (i.e., Figure 6-5) are related only to how spring nutrient loads regulate the year-to-year variability in BWDO for the 27-year study period, not how nutrient loads may further regulate hypoxia under potential nutrient loading reduction (or intensification) scenarios. In addition, long-term changes in nutrient loading would be expected to affect benthic oxygen demands, and thus produce larger impacts on BWDO. An analysis of how changing benthic oxygen demands could affect BWDO is included in the subsequent section on nutrient loading reduction scenarios.

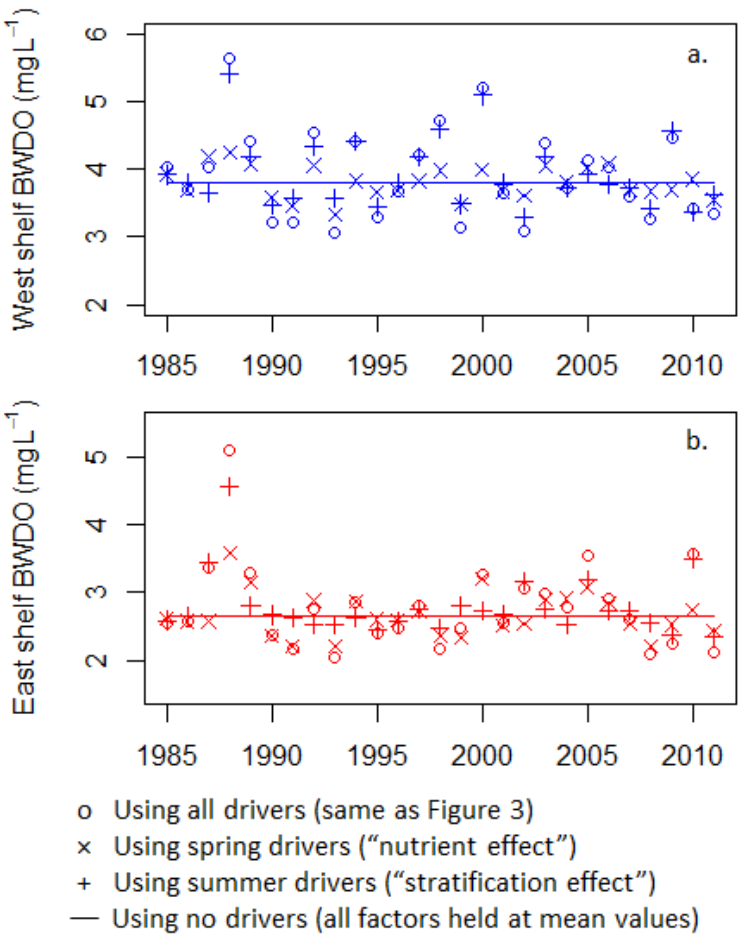

Figure 6-5: Model predicted mean BWDO for the (a.) west and (b.) east shelf sections based on different mechanistic drivers (holding other factors at 27-year mean value).

6.3.3 Temporal trends in hypoxia

Model residuals were analyzed for signs of change in the system's susceptibility to hypoxia over the 27-year study period, where a negative temporal trend in residuals would indicate increasing susceptibility. Temporal trends in residuals were found to be far from significant on both shelf sections, and a visual examination of residuals did not indicate any abrupt temporal transitions (Figure 6-6). This is perhaps not surprising, given that nutrient loading has remained relatively stable throughout the 27-year study period (Goolsby and Battaglin 2001). Model residuals can also be compared with nutrient loads from the preceding year (July-June loads), as shown in Figure 6-6. Here, the residuals are area-weighted averages of the two shelf sections (west shelf receives more weight).

However, no pattern between residuals and these loads (or multi-year averages of these loads) was identified.

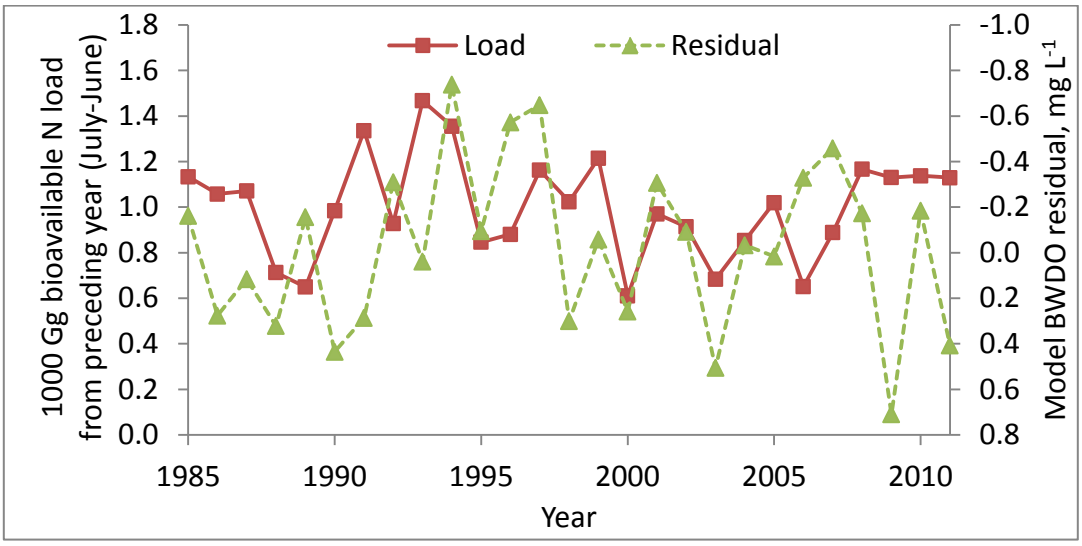

Figure 6-6: Area-weighted mechanistic model residuals (observed-predicted) and nutrient loads from preceding year

6.3.4 Hypoxic area prediction and nutrient reduction scenarios

Predicted BWDO concentrations can be converted to predicted hypoxic areas using equations 6-7 and 6-8. Based on these relationships, the model explains 68% and 73% of year-to-year variability in hypoxic area on the west and east shelf sections, respectively. Aggregating the results, the model explains 70% of the variability in total hypoxic area. Model performance is greater for BWDO than hypoxic area because of the nonlinear relationship between these variables, such that model errors for years of relatively low BWDO are amplified when converted to hypoxic area.

Using the model and the relationship between BWDO and hypoxic area, it is possible to examine how nutrient load reductions would affect the average areal extent of hypoxia (for the 27-year study period). The reductions are relative to the historical bioavailable spring nitrogen load, averaging 133 Gg mo^{-1} for the study period. An important consideration in this analysis is whether the nutrient loading reductions will also result in reductions in benthic oxygen demand. If benthic oxygen demand remains constant, then

even an 80% reduction in nutrient loading (Figure 6-7.a) will still result in a mean total hypoxic area greater than 5000 km^2. However, if nutrient loading reductions are accompanied by a proportional reductions in benthic oxygen demand (Figure 6-7.b), then a 45% (+/- 5%) reduction in nutrient loading would achieve a mean hypoxic area of 5000 km^2. While one would expect benthic oxygen demands to decline as nutrient loading to the system is diminished, the degree of this decline and the time scale over which it would occur remain subjects for future research.

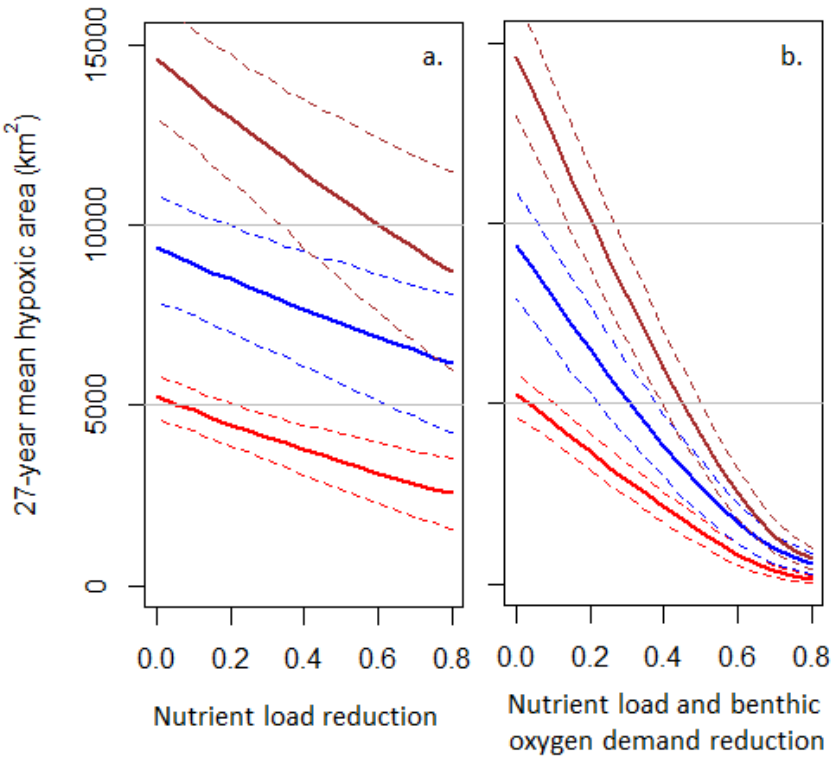

Figure 6-7: 27-year mean hypoxic areas (with 95% credible intervals) for total shelf (brown), west shelf (blue) and east shelf (red) under (a.) spring nutrient load reductions alone and (b.) nutrient load reductions with proportional benthic oxygen demand reductions

6.3.5 Comparison to linear regression modeling

Linear regression (LR) is a modeling approach that has been used in several previous Gulf hypoxia studies (Forrest et al. 2011, Greene et al. 2009, Turner et al. 2012). These previous modeling studies are not directly comparable to this study, as they focus on

somewhat different hypoxia metrics. However, the effectiveness of LR and parsimonious mechanistic modeling for predicting mean BWDO on the east and west shelf sections can be compared here. To enable this comparison, LRs were developed for the east and west shelf sections using the candidate predictor variables from this study (i.e., flow, load, concentration, wind velocity, and wind stress). To help avoid over-parameterization, only variables selected from the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) are used; where BIC results in more parsimonious models than does AIC (Faraway 2005). Models are compared in terms of percent variance explained (R^2), based on both full-model and CV predictions.

Results suggest the parsimonious mechanistic model, developed in this study, outperforms LR models, especially in CV mode (Table 6-2). There are two primary reasons why this is likely to be the case. First, plots of observed versus predicted values for the LRs (not shown) demonstrate a non-linear pattern in residuals, indicating the system is not well represented by linear combinations of the available predictor variables. The mechanistic model performs better in this respect (Figure 6-3), likely because its non-linear relationships better approximate the true functioning of the system. Second, the CV performance of the LRs tends to decrease when more predictor variables are included within the model (i.e., BIC results are better than AIC), suggesting over-parameterization issues. The mechanistic model includes more predictor variables than either of the LR models, but the impact of these variables is regulated by the prior information and biophysical structure of the model. Furthermore, in the mechanistic model, the differential response in BWDO for the two shelf sections is determined based on biophysical properties (e.g., the relative position and areal extent of each shelf section), such that the model can be calibrated to both shelf sections (44 geostatistical 'observations') without relying on shelf-specific calibration parameters. Because the LR models do not benefit from these mechanistic relationships, it is generally necessary to fit separate regressions to the different shelf sections, such that each LR is based on only 27 observations. This smaller sample size makes the LR model parameterizations more sensitive to exclusion of individual data points, which can be problematic in CV.

Table 6-2: Variance explained (R^2) by mechanistic and LR models, based on full model and CV predictions

Model	R^2	CV R^2
West shelf results:		
Mechanistic model (all variables)	75%	72%
West shelf LR w/AIC variables: L_{rN}*(Miss),* W_e*(spring),* W_e*(sum.),* τ*(east)*	66%	31%
West shelf LR w/BIC variables: L_{rN}*(Miss),* W_e*(spring),* W_e*(summer)*	62%	43%
East shelf results:		
Mechanistic model (all variables)	76%	72%
West shelf LR w/AIC variables: L_{rN}*(Miss),* W_e*(summer),* τ*(west),* τ*(east)*	72%	51%
East shelf LR w/BIC variables: L_{rN}*(Miss),* W_e*(summer),* τ*(east)*	69%	54%

6.4 Discussion

Methodologically, this study demonstrates the benefits of using Bayesian inference for calibrating mechanistic environmental models. First, the Bayesian approach readily allowed for probabilistically estimating the parameters of a non-linear model, such as was developed here. Second, the approach provided a systematic means of incorporating prior information about biophysical rates (with associated uncertainties), as determined from previous studies. Third, compared to more traditional approaches, model uncertainties were not constrained to follow normal (Gaussian) distributions, allowing more flexibility in how parameters are represented (Fig 6-2). The benefits of the Bayesian modeling framework have been demonstrated in a previous hypoxia modeling study by Liu et al. (2010), and this study builds on that work by applying the Bayesian framework to a richer mechanistic model, capable of integrating a larger suite of environmental inputs.

In addition, this is the first Gulf hypoxia modeling study to systematically test the model's predictive performance for observations not included within the calibration dataset (using CV). While the mechanistic model developed here performed well in CV, linear regression models developed from the same input variables, performed substantially less well (Table 6-2). This is noteworthy, given that none of the previous Gulf hypoxia regression models have been formally validated. However, models used to make annual hypoxia forecasts (Liu et al. 2010, Turner et al. 2012) have received some degree of validation by comparing these blind annual forecasts to observed values (Evans and Scavia 2011). In the future, CV or other systematic validation exercises could potentially be used to test and lend additional credibility to such models.

The model developed in this work provides new insights into the relative roles of benthic and water column oxygen demands. In the model, benthic oxygen demand is represented by a constant value, such that it is not related to seasonal nutrient loading; whereas water column oxygen demand is directly related to the spring nitrogen load. The calibrated model suggests benthic oxygen demand is approximately 0.33 g $m^{-2}d^{-1}$. In comparison, net water column oxygen demands are estimated to be 0.18 g $m^{-2}d^{-1}$ and 0.34 g $m^{-2}d^{-1}$ for the west and east shelf sections, respectively. The relatively low water column oxygen demand on the western shelf section suggests the western shelf is less responsive to the year-to-year variability in spring nitrogen load, as also suggested by Hetland and DiMarco (2008). For example, the model indicates a 50% reduction in spring nitrogen produces approximately 22% and 34% reductions in west and east shelf hypoxic area, respectively (Figure 6-7.a). If benthic oxygen demands are also reduced, then a much larger reduction in hypoxic area can be expected for both shelf sections (Figure 6-7.b). While benthic oxygen demand would be expected to decrease under sustained nutrient loading reductions, this cannot be verified by the model, because Mississippi River nutrient loads have remained, on average, near historically high levels throughout the study period (Goolsby and Battaglin 2001).

Contrary to previous long-term modeling studies (Greene et al. 2009, Liu et al. 2010, Turner et al. 2008), this study does not indicate a change in the shelf's susceptibility to hypoxia over the 27-year study period. This is likely due, in part, to the use of revised

hypoxia metrics that increase less over the study period (Chapter 5). (The model is not calibrated directly to hypoxic area, but it is calibrated to geostatistical mean BWDO estimates that are highly correlated with the new geostatistical area estimates, as described previously.) It is noted, however, that model residuals were largely negative from 1994-1997 on both shelf sections (Figure 6-6). An intriguing hypothesis is that the large loads of 1993 and 1994 resulted in an accumulation of organic matter that persisted in following years as additional benthic oxygen demand. While this hypothesis cannot be verified by this study alone, the unusually severe impact of the 1993 flood on Gulf water quality has been noted previously (Rabalais et al. 2007). It is also noted that new estimates of hypoxic volume (Chapter 5) suggest a somewhat larger increase over time, and extending this model to predict hypoxic volume may be beneficial future research.

Multiple previous Gulf modeling studies (Greene et al. 2009, Liu et al. 2010, Turner et al. 2012) have suggested nutrient loading reductions ranging from around 40-70% to achieve the Task Force goal of reducing hypoxic area to 5000 km^2 (EPA 2008). This study generally supports that range, but with the caveat that it is dependent on the degree to which benthic oxygen demand (i.e., oxygen demand not related to seasonal nutrient loading) will be diminished as a result of sustained, long-term nutrient loading reductions (Figure 6-7.a versus 6-7.b). As such, this study indicates the need for further research into factors that control the intensity of benthic oxygen demand on the Gulf shelf.

CHAPTER 7: Conclusions and Future Directions

This chapter discusses the scientific and methodological contributions of the dissertation research, as well as new questions raised. In addition, it highlights potential future extensions to this research that would further advance our understanding of Gulf hypoxia. The following three sections address contributions and future directions specific to research performed in Chapters 4, 5, and 6; and a final section discusses overall contributions and future directions.

7.1 Geostatistical Regression

The geostatistical regression (Chapter 4) offers a new way to understand the role stratification plays in regulating the severity of Gulf hypoxia. This is the first shelf-wide study to develop relationships between measured BWDO concentrations and water column profile data (i.e., salinity and temperature). As such, it provides a relatively direct way to understand the relationship between BWDO and stratification, when compared to models that use flow and wind data to represent stratification. Based on this work, stratification metrics, along with depth and northing variables, explain 27% - 61% of the spatial variability in BWDO for each annual cruise.

The stratification effects were spatially aggregated in order to determine their overall impact on mean BWDO for each annual shelfwide cruise. Because the impact of stratification is determined relatively directly, rather than as a function of river flow, this analysis avoids the confounding correlation between river flow and load, which has made it difficult to disentangle the roles of these two factors (Hetland and DiMarco 2008). Stratification is found to explain about half of the interannual variability in BWDO. The remaining variability can then be explained in terms of other factors, of which May nitrate concentration is found to be most influential. Thus, the study not only addresses the role

of stratification, but also nutrient loading, suggesting that both factors have similar levels of influence on the year-to-year variability of hypoxia.

The geostatistical regression also informs our understanding of the time scale over which river flows correlate with stratification intensity. Stratification was found to be highly correlated (r^2>0.8) with the flow from the two preceding months (Section 6.2.7). This information was important to the development of the mechanistic model (Chapter 6), by suggesting an appropriate temporal averaging period for the river flow inputs used to model reaeration (which is primarily controlled by the intensity of stratification).

Methodologically, this research demonstrates how geostatistical modeling can be used to improve our understanding of how an environmental system functions. The geostatistical approach was found to be necessary because of the significant correlation among model residuals (i.e., correlated stochasticity). Multiple linear regression, for example, would not be valid in this case, as the assumption of i.i.d. residuals would be violated, and highly correlated observations (in close proximity to each other) would receive undue wait in determining the model parameterization (as described in Section 3.1). Despite this, geostatistical modeling is not commonly used for the purposes of making statistical inferences about environmental systems. Two studies, by Mueller et al. (2010) and Yadav et al. (2009) used a geostatistical regression models to infer factors influencing the temporal variability of terrestrial CO_2 fluxes and gross primary productivity, respectively. However, most geostatistical studies focuses on spatial estimation (interpolation, etc.) rather than inference, and if a deterministic component is included, it is primarily used to improve model estimates (as in Chapter 5). Thus, this study adds to the relatively limited literature demonstrating how geostatistical models can be effective inferential tools.

This study raises the question of whether Mississippi/Atchafalaya River nutrient concentration or load is more influential in determining the severity of Gulf hypoxia. The study indicates that both May nitrate load and concentration are correlated with the interannual variability of hypoxia (i.e., the portion of the variability not accounted for by stratification), but the correlation with concentration (r^2=0.76) is substantially greater than with load (r^2=0.32). While two other studies have also indicated that concentration (along

with flow), is an effective predictor of hypoxia (Forrest et al. 2011, Greene et al. 2009), multiple studies demonstrate that load alone is an effective predictor (Justic et al. 1997, Liu et al. 2010, Turner et al. 2012). In fact, the mechanistic model (Chapter 6) generally indicates that load is the more important predictor of BWDO depletion. One limitation of the geostatistical regression is that it only covers ten years of data. As such, future research could expand the study to cover the entire period of record (currently 27 years), allowing for more robust inference, and more extensive comparison with the mechanistic model.

The modeling approach developed in this study can be applied to other environmental variables as well. In the Gulf, geostatistical regression could be used to study the spatial and temporal variability of nutrients and other water quality variables related to hypoxia formation. A particularly intriguing future research direction would be to model benthic and water column oxygen demands, using data collected in recent extensive field studies of the Gulf shelf (Lehrter et al. 2012, Murrell et al. 2013). Both benthic and water column oxygen demands are expected to be important to hypoxia formation, as indicated by the mechanistic model (Chapter 6) and previous modeling studies (Bierman et al. 1994, Fennel et al. 2013, Rowe 2001). These oxygen demands could be modeled geostatistically (as a response variable) to better understand their temporal and spatial variability. For example, the model could be used to test whether oxygen demands (benthic or water column) are higher nearer to the river outfalls, and whether they are higher following large inputs of river flow and nutrient load. The oxygen demands could also be used as covariates in a geostatistical regression to predict BWDO. Assuming that oxygen demands are related to nutrient loads, such an approach could provide new insights into how nutrient loading controls hypoxia formation, both spatially and temporally. Also, if both oxygen demands and stratification metrics are included within a geostatistical regression for BWDO, it may be possible to further refine our understanding of how both factors control hypoxia formation.

7.2 Geostatistical Spatial Estimation

The geostatistical spatial estimation model (Chapter 5), provides new estimates of hypoxic area and volume for the entire 27-year record of available LUMCON shelfwide cruise data (1985-2011). Accurate hypoxic extent estimates are important to hypoxia management, as existing Action Plans call for a reduction in the areal extent of hypoxia (EPA 2001, 2008), and for calibrating models that predict hypoxic extent based on nutrient loading and other factors (DiMarco 2011, Forrest et al. 2011, Greene et al. 2009, Liu et al. 2010, Turner et al. 2012).

The new hypoxic area estimates compare well with previous area estimates (LUMCON 2012, Rabalais et al. 2007) in most years. However, in the earlier years of the study period (1985-1993), the new estimates are significantly higher than the previous estimates. As discussed in Section 5.4, this is likely due to smaller cruises and inconsistent sampling equipment, combined with the limitations of the interpolation approach used to develop the previous estimates. Because of this discrepancy, the new hypoxic area estimates indicate that there has been less of an increase in hypoxic area over time, such that the existing models for predicting hypoxic area may require recalibration. This should affect our understanding of how hypoxic area is changing over large time scales, as discussed further in Section 7.3.

This research provides the first rigorous estimates of hypoxic volume for the northern Gulf, offering a new perspective into how the severity of hypoxia has changed from year to year. For example, 2002 and 2008 both have similar hypoxic areas (they are not significantly different, $p>0.05$), but they are very different in terms of hypoxic thickness, such that 2008 has approximately twice the hypoxic volume of 2002. In addition, while hypoxic area was not found to increase significantly over they study period ($p=0.42$), hypoxic thickness was found to be increasing over time ($p=0.05$). Thus, the thickness and volume estimates provide new information for understanding the temporal variability of hypoxia, and assessing its large-scale consequences.

This research uses many of the same geostatistical methodologies developed to estimate hypoxic area in Lake Erie (Zhou et al. 2013). The modeling framework makes

use of covariate information (i.e., trend variables such as depth and spatial position), to improve estimates and reduce uncertainty. In addition, results are determined using a simulation-based approach (i.e., conditional realizations), rather than an interpolation-based approach, yielding more robust extent estimates and quantified uncertainty. This research extends the previous work of Zhou et al. by developing a new methodology for determining hypoxic volume, based on a two-step simulation approach, where BWDO is first simulated, followed by simulation of BWHF at locations where simulated BWDO is below the hypoxic threshold. In addition, this work provides a way to test for changes in the spatial pattern of hypoxia, by introducing candidate variables that allow for cruise-specific east-west trends. Finally, this work provides a new mechanism for performing BWDO measurement adjustments that account for inconsistent sampling equipment. These adjustments are performed as part of the simulation process, so that resulting extent uncertainties reflect these adjustments.

There is considerable potential to expand this work in the future. While hypoxic volume provides a unique new metric to gage the severity of hypoxia, neither area nor volume completely characterize the degree of dissolved oxygen depletion in the northern Gulf. Future research could focus on developing other relevant dissolved oxygen metrics, such as the total oxygen deficit on the shelf. Here, oxygen deficit would be defined as the mass of oxygen that has been depleted relative to saturation conditions. Such metrics could potentially be more useful to biogeochemical modeling studies for which the 2 mg L^{-1} hypoxic threshold does not have particular scientific relevance.

The geostatistical model could also be used to help optimize hypoxia sampling network design. Some sampling network evaluation, based on this dissertation research, was presented at a Hypoxia Coordination Workshop (Obenour et al. 2012a), but because the geostatistical model has been improved considerably since this workshop, this work should be revisited and expanded. The geostatistical model could be used to determine the most effective sampling network (in terms of sampling interval and envelope) for minimizing uncertainty in hypoxic extent estimates, subject to limitations in sampling budget and ship time. The model could also potentially be used to study the effectiveness of new sampling

technologies, such as remotely controlled vehicles (e.g., 'gliders'), that may allow for more economical sampling, but that cannot be lowered all of the way to the sea floor.

For sampling network design studies, it may also be useful to develop BWDO pseudo-data covering the entire shelf at fine spatial resolution. This pseud-data could be considered as 'truth', to which geostatistical modeling results can be compared. In this type of analysis, the BWDO 'observations' would be taken from the pseudo-data based on a proposed sampling network design, and then used within the geostatistical model to develop hypoxic extent estimates that can then be compared to the 'true' hypoxic extents determined from the pseudo-data. In this way, it would be possible to consider sampling locations not included within historical sampling networks. Pseudo-data could also be used to validate the estimates and uncertainties determined by the geostatistical model. However, such a validation exercise may provide somewhat overly-favorable results if the model used to generate the pseudo-data is the same model used to determine the hypoxic extent estimates. Thus, uncertainty in model structure would need to be considered.

Another future research direction could be to re-develop the geostatistical model within a hierarchical framework, where model parameters would have the flexibility to vary by annual cruise. Currently, all model parameters are constant across all 27 years, except for the cruise-specific east-west trends. The current methodology for allowing unique east-west trends could be applied to other model parameters as well. However, this methodology is rigid, in the sense that parameters are either determined solely from a specific cruise, or from the data from all regular (non-specific) cruises. Hierarchical modeling, on the other hand, would allow parameters to be determined by both the total dataset and the cruise-specific data, simultaneously. This approach could also be applied to the covariance parameters, as well, such that if the variance and spatial correlation varies from year-to-year, this could be more accurately represented within the model. A drawback to this approach is that hierarchical modeling would require a more complex and computationally expensive Bayesian implementation, and it would be important to explore whether this expense is justified through significantly improved hypoxic extent estimates with reduced uncertainty.

7.3 Mechanistic Model

The mechanistic model (Chapter 6) links the severity of hypoxia on the west and east Louisiana-Texas shelf to its primary biophysical drivers, including nutrient loading, river flow, and weather. The model provides insight into the relative importance of these different factors and allows for prediction of hypoxic severity under different future scenarios. The model is rigorously calibrated and validated over the entire 27-year record of available mid-summer shelfwide cruise data (1985-2011). It is the first such model to mechanistically incorporate flow and wind data (in addition to load), and to predict the severity of hypoxia on multiple shelf segments.

The mechanistic model is calibrated to output from the spatial estimation model (Chapter 5). Specifically, the model predicts the (geostatistical) mean BWDO values for the two shelf sections. In addition, modeling results can easily be converted to hypoxic area, through linear regression. Modeling results have not yet been developed for hypoxic volume, though this could be beneficial future research because the temporal pattern in hypoxic volume is somewhat different than that of hypoxic area (Figure 5-12), and they are only moderately correlated (r^2=0.77). Plots of hypoxic area and volume per spring nitrogen load versus time (Figure 7-1) suggest that volume increases more over the study period relative to nutrient loading (though neither of these metrics increase significantly based on linear trends with time, p>0.05). Mechanistic modeling could be used to study the temporal variability of hypoxic volume in more detail, and if long-term trends exist, explore whether they are related to eutrophication or changing oceanographic conditions (e.g., stratification).

The model results confirm that both seasonal nutrient loading and stratification contribute substantially to the year-to-year variability in hypoxia, as also suggested by the geostatistical regression study in Chapter 4 (2012b). Stratification is represented by the reaeration submodel (eq 6-6), which is a function of summer river discharge, summer east-west wind velocity, and wind stress. The role of stratification is found to be larger on the west shelf than on the east shelf, which is not surprising given that seasonal nutrient loading contributes less to west shelf oxygen demand. The previous study by Obenour et

al. indicated that stratification and seasonal nutrient loading have approximately equal influence on the year-to-year variability in hypoxia. The larger role of stratification in this study is likely due, in part, to the shelf division, because summer east-west wind velocity impacts stratification oppositely on the west and east shelf sections, such that these impacts compensate for each other when studying the shelf as a whole. (Nutrient loading can also be distributed to the east or west, based on spring wind velocity, but spring winds are more consistently easterly.) Also, the previous study focused only on the most commonly sampled locations, and thus did not extend as far west, where the role of stratification is indicated to be larger.

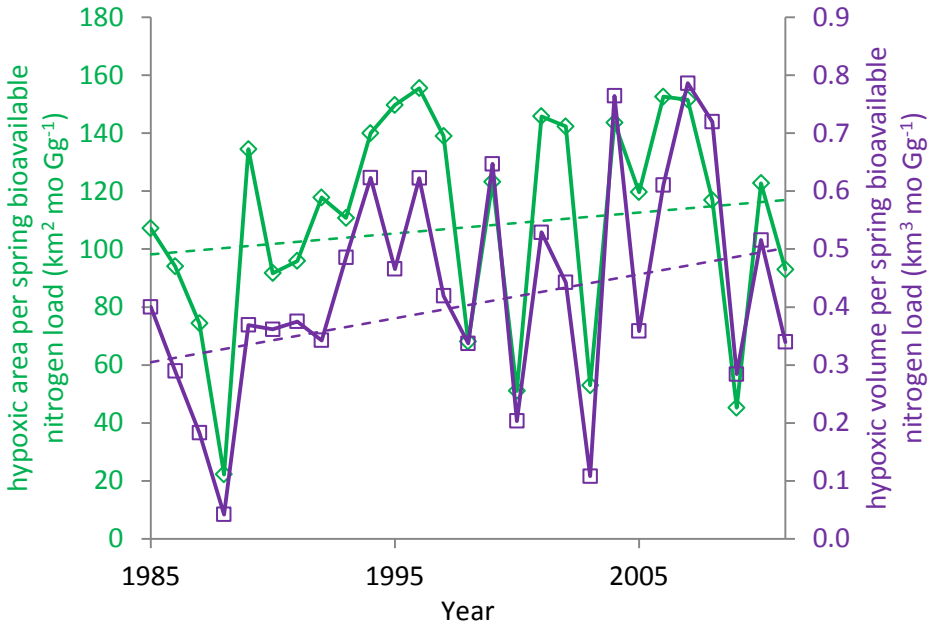

Figure 7-1: Hypoxic area and volume relative to nutrient loading, versus time. Dashed lines show linear tends with time (p=0.48 for area and p=0.12 for volume, accounting for extent estimate uncertainty)

Future efforts could focus on exploring assumptions made in model development. For example, it could be useful to explore whether benthic oxygen demands vary across the Gulf shelf. In the current model, a single value for benthic oxygen demand is used for both shelf sections, as Murrell and Lehrter (2011) report no clear spatial trends in benthic oxygen demands (or water column oxygen demands). However, one would expect benthic

oxygen demands to be higher nearer to river outfalls, and this could be explored more rigorously using a geostatistical regression, as suggested in Section 7.1. Based on the results of such an analysis, the two shelf sections could be calibrated to different benthic oxygen demands constrained by prior information from the geostatistical regression. In addition, the submodels for hydrodynamic transport and reaeration could potentially be refined based on comparisons with more complex hydrodynamic models (Hetland and DiMarco 2008, Wang and Justic 2009). This would require close collaboration with the developers of these other models, in order to extract appropriate comparison data.

Future efforts could also expand the scientific scope of the model. For example, the model could be used to hindcast the severity of hypoxia in past decades (prior to the beginning of the shelfwide cruise sampling program). This type of analysis has been performed in previous studies using historical loading data (Scavia et al. 2004), but the mechanistic model developed here could also make use of historical flow and weather data. The model could also be used to evaluate climate change scenarios, particularly those related to changes in Mississippi River basin hydrology that would affect flows and loads delivered to the Gulf shelf.

7.4 Overall Contributions and Path Forward

As indicated in the previous sections, this research provides new insights into how the severity of Gulf hypoxia varies over time and space. Particularly, it characterizes the spatial and interannual variability of Gulf hypoxia through geostatistical spatial estimation, and explains this variability in terms of biophysical drivers through geostatistical regression and mechanistic modeling. A primary contribution of this research has been to quantify the roles of stratification and seasonal nutrient loading in regulating the year-to-year variability of Gulf hypoxia. Overall, these models suggest that stratification and seasonal nutrient loading are of similar importance, and taken together, these factors explain at least 70% of hypoxic variability throughout the study period (1985-2011).

In addition to accounting for the uncertainty in various types of Gulf hypoxia models, as discussed in the previous sections, this research demonstrates how uncertainty can be propagated between models. For example, the uncertainty associated with the instrument

bias adjustment (Section 5.2.3) is propagated all of the way to the biophysical inferences of the mechanistic model (e.g., Section 6.3.2). This is because the geostatistical model samples from the uncertainty in the instrument bias adjustment 'model' when performing conditional realizations; and the mechanistic model samples from the uncertainty in the geostatistical model estimates during the Bayesian calibration process.

A potentially rewarding future research path would be to further integrate the mechanistic and geostatistical modeling results. These models provide information about hypoxic severity (mean BWDO or hypoxic area) based on different types of information. The mechanistic model predictions are best estimates of hypoxic severity based on environmental drivers (loads, flows, and winds) and the biophysical model formulation prescribed. The geostatistical model provides best estimates of mean BWDO based on the sampled dissolved oxygen concentrations and static covariate information such as depth and spatial position. As such, it may be desirable to synthesize the geostatistical and mechanistic modeling results in order to obtain overall best estimates of BWDO. This can be accomplished using a Bayesian approach, where the 'prior' geostatistical modeling estimate is updated (i.e., multiplied) by the likelihood of BWDO, given the biophysical controls on hypoxia, as determined from the mechanistic model (per eq 3-9).

This approach is demonstrated in Figure 7-2, where geostatistical, mechanistic, and synthesized model estimates for mean BWDO on each shelf section are compared side-by-side for the 27-year study period. Here, all model predictions are represented by 95% prediction intervals. While mechanistic model uncertainties are fairly uniform across all cruises, geostatistical estimate uncertainties are more variable due to factors described in Section 5.4. In almost all cases, the geostatistical estimates have less uncertainty, and this is not surprising, given the relatively dense sampling performed during these mid-summer cruises. However, in 1989, the west shelf was not sampled, resulting in greater uncertainty in the geostatistical estimate (Figure 7-2.a). In all cases, the synthesized estimate more closely resembles whichever model estimate contains less uncertainty. Also, the synthesized estimate always has the lowest uncertainty, and may be the best metric for future studies exploring the large-scale ecological and economic consequences of hypoxia.

A potential concern regarding this approach is that because the mechanistic model is calibrated to the geostatistical estimates, the geostatistical estimates are effectively used twice. Thus, it would be more defensible to use blind mechanistic model predictions (as determined from CV) in this type of analysis. However, this approach would be expected to yield similar results to those shown in Figure 7-2, because the CV analysis (Section 6.3.1) indicates that a mechanistic model prediction for a given year is largely independent from its corresponding geostatistical estimate. It should also be noted that the geostatistical model is independent of the temporal processes considered in the mechanistic model; the only temporally varying inputs to the geostatistical model are the DO measurements themselves.

In the future, it would also be beneficial to expand the mechanistic and geostatistical modeling over the entire summer season. Currently, model results are developed only for the times of the LUMCON shelfwide cruises (usually in late July). However, as discussed in Section 5.4, hypoxic conditions may vary considerably throughout the summer based on changing weather and river inputs. The geostatistical model could be used to develop hypoxic extent estimates using data from other periodic hypoxia monitoring cruises (Section 2.2.2). These other cruises typically have less dense sampling than the LUMCON shelfwide cruises, but this would be reflected in increased estimation uncertainty. Also, the mechanistic model could be applied across the entire summer by allowing model inputs to vary in time, and by potentially modifying the model formulation (if the model is not found to be representative of the entire summer season). The results synthesis presented above (Figure 7-2) could then be applied at times when a geostatistical estimate is available. Furthermore, if the differences between the geostatistical estimates and mechanistic predictions are correlated in time, then synthesized predictions could be temporally interpolated, between the times when geostatistical estimates are available. This temporal interpolation could potentially be performed using an adaptation of the geostatistical methods presented in this research.

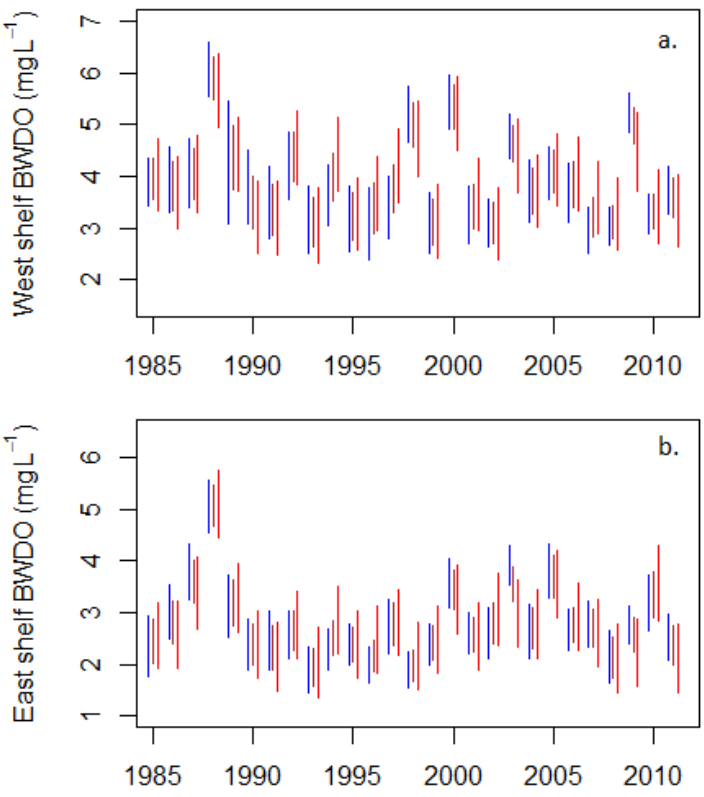

Figure 7-2: BWDO 95% prediction intervals for (a.) west and (b.) east shelf sections, from geostatistical model (blue, left), mechanistic model (red, right), and synthesized model results (brown, middle)

APPENDIX: Geostatistical Estimates for total hypoxic area and volume

Chapter 5 describes the models for BWDO and BWHF, which yield the area and volume of the hypoxic bottom layer. However, models can also be developed for the minimum dissolved oxygen (MinDO) and total hypoxic fraction (THF) which yield the total hypoxic area and volume, as described in Section 5.2.1. These models account for layers of hypoxia existing higher (i.e., suspended) in the water column. The parameterization and results of these models are provided below. Table A-1 corresponds to Table 5-1 in the primary text. Table A-2 presents the total hypoxic area and volume results (CR+UK methodology), corresponding to Table 5-3. Figures A-1 and A-2 present maps of expected MinDO and total hypoxic thickness, corresponding to Figures 5-14 and 5-15.

In general, the results for total hypoxic area and volume are similar to the results for bottom layer hypoxic area and volume (as presented in the main text). On average, the total hypoxic area is 14% larger than the bottom layer hypoxic area, and the two sets of estimates are highly correlated (r^2=0.97). Similarly, the total hypoxic volume is 18% larger than the bottom layer hypoxic volume, on average, and they are also highly correlated (r^2=0.97). This indicates that the traditionally reported bottom layer hypoxic extent is also proportionally representative of the total hypoxic extent.

Table A-1: Regression coefficients ($\hat{\beta}_p$) with standard errors ($\sigma_{\hat{\beta}}$) for normalized trend variables selected in the MinDO and THF models

Variable	MinDO ($mg\ L^{-1}$)		THF (-)	
	β	σ_β	β	σ_β
E	-0.74	0.09	n.s.	
E^2	0.32	0.07	n.s.	
N	-0.45	0.09	-0.005	0.007
N^2			-0.016	0.006
D	-2.45	0.18	n.a.	
D^2	2.39	0.16	n.a.	
O	n.a.		-0.080	0.004
O^2	n.a.		n.s.	
c.s. E 1998	-1.29	0.43	n.s.	
c.s. E 2010	1.04	0.33	n.s.	

c.s.=cruise specific,
n.s. = not selected,
n.a. = not available

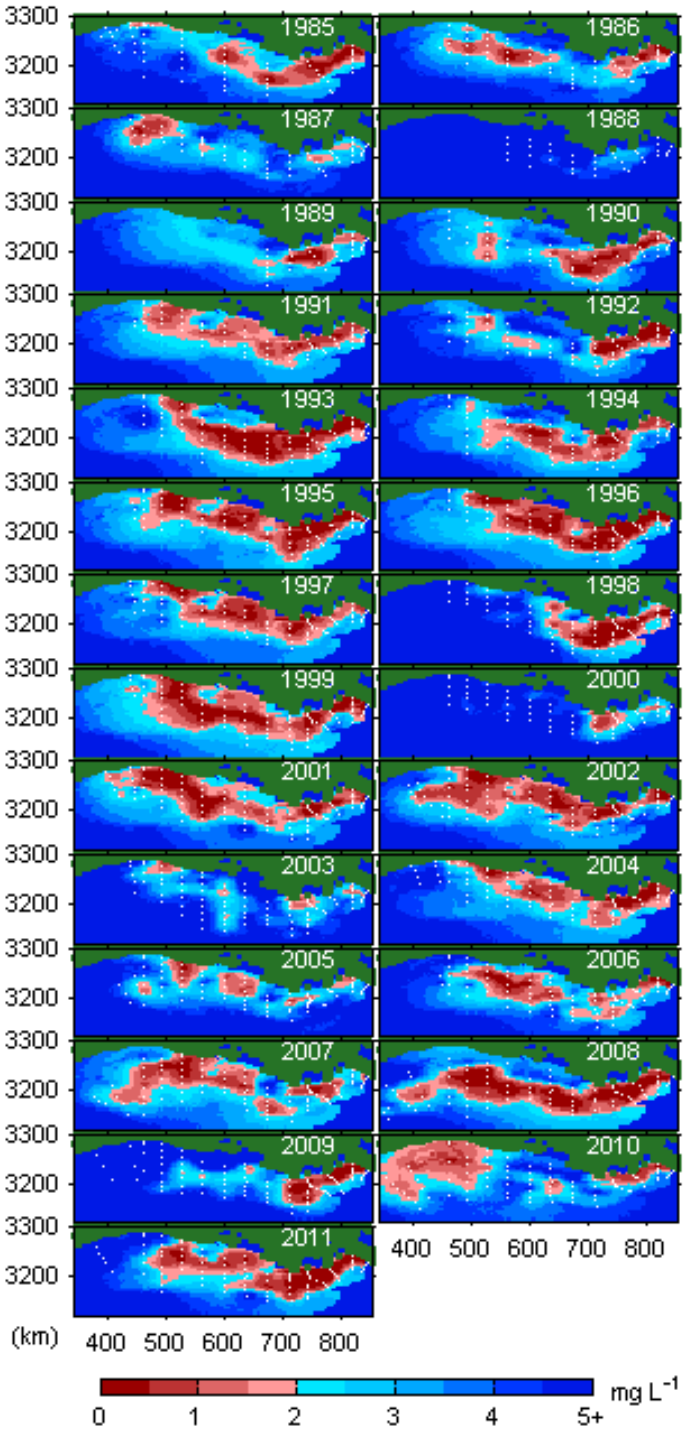

Figure A-1: Maps showing modeled median MinDO concentration, 1985-2011

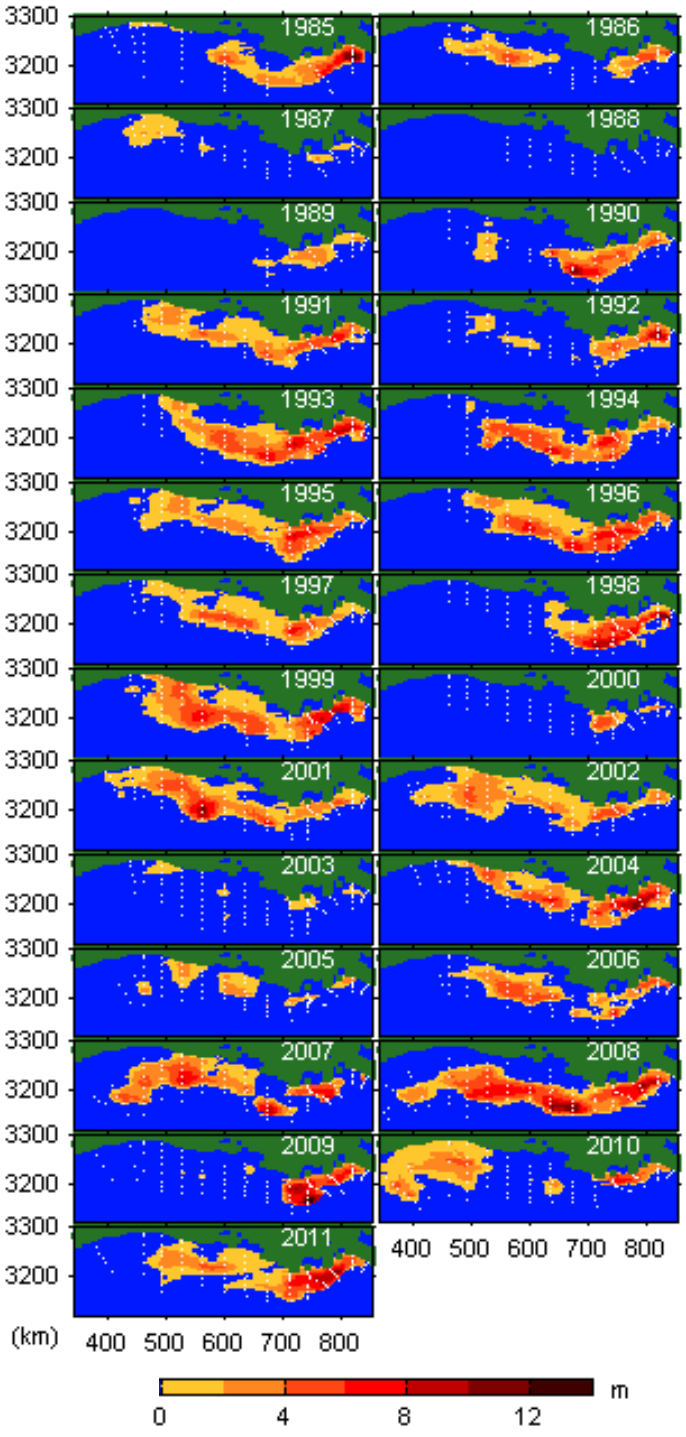

Figure A-2: Maps showing modeled median total hypoxic thickness, 1985-2011

Table A-2: Tabulated total (multi-layer) hypoxic area and volume estimates, 1985-2011

Year	Area (1000 km^2)				Volume (km^3)			
	mean	median	2.5 perc	97.5 perc	mean	median	2.5 perc	97.5 perc
1985	15.8	15.8	12.3	19.6	63.7	62.5	39.5	92.9
1986	14.6	14.4	11.1	19.0	45.1	43.3	27.9	71.9
1987	10.3	10.1	7.1	14.3	25.4	23.9	12.5	46.4
1988	0.6	0.6	0.2	1.4	1.2	0.8	0.0	4.5
1989	13.7	13.1	6.5	23.6	41.6	37.1	15.4	92.2
1990	15.8	15.6	11.7	21.3	61.6	59.3	39.3	95.7
1991	21.1	20.9	16.1	27.6	86.0	82.7	52.6	139.2
1992	12.7	12.4	9.8	16.9	42.0	40.4	27.8	66.0
1993	24.4	24.2	19.9	30.5	112.6	110.1	82.5	160.0
1994	18.4	18.2	14.8	22.9	89.4	87.6	62.1	130.6
1995	25.3	25.2	20.5	31.3	88.1	85.6	58.4	131.6
1996	25.0	24.8	20.0	31.6	109.3	105.2	74.2	163.2
1997	21.2	20.9	17.3	26.5	71.0	67.9	45.9	114.8
1998	11.7	11.7	9.9	13.8	58.6	57.4	42.6	81.8
1999	28.2	27.9	23.0	34.6	140.4	138.3	96.4	197.5
2000	3.9	3.8	2.8	5.3	15.9	15.3	9.1	26.5
2001	23.9	23.8	20.0	29.0	88.8	86.9	64.2	125.8
2002	24.3	24.1	20.7	28.6	74.7	72.5	53.3	105.0
2003	6.6	6.6	5.1	8.5	14.2	13.4	7.6	24.5
2004	19.7	19.4	15.8	24.9	107.6	104.5	71.9	161.7
2005	10.8	10.6	8.5	13.5	34.3	33.0	21.4	54.5
2006	16.6	16.4	13.6	20.6	69.1	67.1	47.6	102.9
2007	23.3	23.2	19.3	27.8	123.6	120.9	88.8	173.8
2008	26.1	26.2	22.4	30.3	142.8	140.6	110.2	190.8
2009	8.9	8.9	7.4	10.7	51.9	50.9	38.7	71.9
2010	20.2	20.1	16.2	24.4	78.9	77.6	54.8	110.5
2011	20.7	20.6	17.8	24.0	82.1	81.1	61.6	110.4
average	17.2	17.0	13.7	21.6	71.1	69.1	48.4	105.4

REFERENCES

Arhonditsis, G. B.; Qian, S. S.; Stow, C. A.; Lamon, E. C.; Reckhow, K. H. (2007). Eutrophication risk assessment using Bayesian calibration of process-based models: Application to a mesotrophic lake. *Ecological Modelling 208* (2-4), 215-229.

Beck, M. B.; Straten, G. v. (1983). *Uncertainty and forecasting of water quality*. Springer-Verlag: Berlin ; New York.

Beven, K. (2006). A manifesto for the equifinality thesis. *Journal of Hydrology 320* (1-2), 18-36.

Bianchi, T. S.; DiMarco, S. F.; Cowan, J. H.; Hetland, R. D.; Chapman, P.; Day, J. W.; Allison, M. A. (2010). The science of hypoxia in the Northern Gulf of Mexico: A review. *Science of the Total Environment 408* (7), 1471-1484.

Bierman, V. J.; Hinz, S. C.; Zhu, D. W.; Wiseman, W. J.; Rabalais, N. N.; Turner, R. E. (1994). A preliminary mass-balance model of primary productivity and dissolved-oxygen in the Mississippi River plume inner Gulf Shelf region. *Estuaries 17* (4), 886-899.

Boesch, D. F.; Boynton, W. R.; Crowder, L. B.; Diaz, R. J.; Howarth, R. W.; Mee, L. D.; Nixon, S. W.; Rabalais, N. N.; Rosenberg, R.; Sanders, J. G.; Scavia, D.; Turner, R. E. (2009). Nutrient Enrichment Drives Gulf of Mexico Hypoxia. *Eos, Transactions American Geophysical Union 90* (14), 117-118.

Brzezinski, M. A. (1985). The Si-C-N Ratio of Marine Diatoms - Interspecific Variability and the Effect of Some Environmental Variables. *Journal of Phycology 21* (3), 347-357.

Bushaw, K. L.; Zepp, R. G.; Tarr, M. A.; Schulz-Jander, D.; Bourbonniere, R. A.; Hodson, R. E.; Miller, W. L.; Bronk, D. A.; Moran, M. A. (1996). Photochemical release of biologically available nitrogen from aquatic dissolved organic matter. *Nature 381* (6581), 404-407.

Canham, C. D.; Cole, J. J.; Laurenroth, W. K. (2003). *Models in ecosystem science*. Princeton University Press: Princeton, NJ.

Chapra, S. C. (1997). *Surface water-quality modeling*. McGraw-Hill: New York.

Chiles, J.-P.; Delfiner, P. (1999). *Geostatistics : Modeling spatial uncertainty*. Wiley: New York.

Conley, D. J.; Carstensen, J.; Vaquer-Sunyer, R.; Duarte, C. M. (2009). Ecosystem thresholds with hypoxia. *Hydrobiologia 629* (1), 21-29.

Cressie, N. (1990). The Origins of Kriging. *Mathematical Geology 22* (3), 239-252.

Dagg, M. J.; Ammerman, J. W.; Amon, R. M. W.; Gardner, W. S.; Green, R. E.; Lohrenz, S. E. (2007). A review of water column processes influencing hypoxia in the northern Gulf of Mexico. *Estuaries and Coasts 30* (5), 735-752.

DeAngelis, D. L.; Mooij, W. M. (2003). In praise of mechanistically rich models. In *Models in ecosystem science*, Canham, C. D.; Cole, J. J.; Laurenroth, W. K., Eds. Princeton University Press: Princeton, NJ.

Diaz, R. J.; Rosenberg, R. (1995). Marine benthic hypoxia: A review of its ecological effects and the behavioural responses of benthic macrofauna. *Oceanography and Marine Biology - an Annual Review, Vol 33 33*, 245-303.

Diaz, R. J.; Rosenberg, R. (2008). Spreading dead zones and consequences for marine ecosystems. *Science 321* (5891), 926-929.

Dilks, D. W.; Canale, R. P.; Meier, P. G. (1992). Development of Bayesian Monte-Carlo Techniques for Water-Quality Model Uncertainty. *Ecological Modelling 62* (1-3), 149-162.

DiMarco, S. F. (2011). Gauss-Markov optimal interpolation and spatial scales of dissolved oxygen in the northern Gulf of Mexico hypoxic zone. In *ASLO Aquatic Sciences Meeting*, San Juan, Puerto Rico.

Donner, S. D.; Scavia, D. (2007). How climate controls the flux of nitrogen by the Mississippi River and the development of hypoxia in the Gulf of Mexico. *Limnology and Oceanography 52* (2), 856-861.

Dortch, Q.; Rabalais, N. N.; Turner, R. E.; Rowe, G. T. (1994). Respiration rates and hypoxia on the Louisiana Shelf. *Estuaries 17* (4), 862-872.

Emery, W. J.; Thomson, R. E. (2001). *Data analysis methods in physical oceanography*. 2nd and rev. ed.; Elsevier: Amsterdam ; New York.

EPA Mississippi River/Gulf of Mexico Watershed Nutrient Task Force http://water.epa.gov/type/watersheds/named/msbasin/learn.cfm. (Accessed 2013.)

EPA (1985). *Rates, constants, and kinetics formulations in surface water quality modeling (Second Edition)*. Tetra Tech Inc. and Environmental Protection Agency, Office of Research and Development: Athens, GA.

EPA (2001). *Mississippi River/Gulf of Mexico Watershed Nutrient Task Force Action Plan for reducing, mitigating, and controlling hypoxia in the northern Gulf of Mexico*; Office of Wetlands, Oceans, and Watersheds, U.S. Environmental Protection

Agency: Washington, DC, http://water.epa.gov/type/watersheds/named/msbasin/history.cfm.

EPA (2007). *Hypoxia in the Northern Gulf of Mexico, An Update by the EPA Science Advisory Board*; Washington, DC, EPA-SAB-08-003.

EPA (2008). *Mississippi River/Gulf of Mexico Watershed Nutrient Task Force Gulf Hypoxia Action Plan 2008 for reducing mitigating, and controlling hypoxia in the northern Gulf of Mexico and improving water quality in the Mississippi River Basin*; U.S. Environmental Protection Agency, Office of Wetlands, Oceans, and Watersheds: Washington, D.C., http://water.epa.gov/type/watersheds/named/msbasin/actionplan.cfm.

Erickson, T. A.; Williams, M. W.; Winstral, A. (2005). Persistence of topographic controls on the spatial distribution of snow in rugged mountain terrain, Colorado, United States. *Water Resources Research 41* (4).

Evans, M. A.; Scavia, D. (2011). Forecasting hypoxia in the Chesapeake Bay and Gulf of Mexico: model accuracy, precision, and sensitivity to ecosystem change. *Environmental Research Letters 6* (1), 015001.

Faraway, J. J. (2005). *Linear models with R*. Chapman & Hall/CRC: Boca Raton, Fla.

Feng, Y.; DiMarco, S. F.; Jackson, G. A. (2012). Relative role of wind forcing and riverine nutrient input on the extent of hypoxia in the northern Gulf of Mexico. *Geophysical Research Letters 39*, L09601.

Fennel, K.; Hetland, R. D.; Feng, Y.; DiMarco, S. F. (2011). A coupled physical-biological model of the Northern Gulf of Mexico shelf: model description, validation and analysis of phytoplankton variability. *Biogeosciences 8* (7), 1881-1899.

Fennel, K.; Hu, J.; Laurent, A.; Marta-Almeida, M.; Hetland, R. (2013). Sensitivity of hypoxia predictions for the northern Gulf of Mexico to sediment oxygen consumption and model nesting. *Journal of Geophysical Research-Oceans 118*.

Forrest, D. R.; Hetland, R. D.; DiMarco, S. F. (2011). Multivariable statistical regression models of the areal extent of hypoxia over the Texas-Louisiana continental shelf. *Environmental Research Letters 6* (4), 045002.

Friedman, J.; Hastie, T.; Tibshirani, R. (2010). Regularization Paths for Generalized Linear Models via Coordinate Descent. *Journal of Statistical Software 33* (1), 1-22.

Gelfand, A. E.; Smith, A. F. M. (1990). Sampling-Based Approaches to Calculating Marginal Densities. *Journal of the American Statistical Association 85* (410), 398-409.

Gelfand, A. E.; Diggle, P. J.; Fuentes, M.; Guttorp, P. (2010). *Handbook of spatial statistics*. CRC Press: Boca Raton, FL.

Gelman, A.; Hill, J. (2007). *Data analysis using regression and multilevel/hierarchical models*. Cambridge University Press: Cambridge ; New York.

Gneiting, T.; Guttorp, P. (2010). Continuous parameter stochastic process theory. In *Handbook of spatial statistics*, Gelfand, A. E.; Diggle, P. J.; Fuentes, M.; Guttorp, P., Eds. CRC Press: Boca Raton, FL.

Goolsby, D. A.; Battaglin, W. A. (2001). Long-term changes in concentrations and flux of nitrogen in the Mississippi River Basin, USA. *Hydrological Processes 15* (7), 1209-1226.

Goolsby, D. A.; Battaglin, W. A.; Aulenbach, B. T.; Hooper, R. P. (2001). Nitrogen input to the Gulf of Mexico. *Journal of Environmental Quality 30* (2), 329-336.

Goovaerts, P. (2000). Geostatistical approaches for incorporating elevation into the spatial interpolation of rainfall. *Journal of Hydrology 228* (1-2), 113-129.

Greene, R. M.; Lehrter, J. C.; Hagy, J. D. (2009). Multiple regression models for hindcasting and forecasting midsummer hypoxia in the Gulf of Mexico. *Ecological Applications 19* (5), 1161-1175.

Gronewold, A. D.; Qian, S. S.; Wolpert, R. L.; Reckhow, K. H. (2009). Calibrating and validating bacterial water quality models: A Bayesian approach. *Water Research 43* (10), 2688-2698.

Guildford, S. J.; Hecky, R. E. (2000). Total nitrogen, total phosphorus, and nutrient limitation in lakes and oceans: Is there a common relationship? *Limnology and Oceanography 45* (6), 1213-1223.

Hernes, P. J.; Benner, R. (2003). Photochemical and microbial degradation of dissolved lignin phenols: Implications for the fate of terrigenous dissolved organic matter in marine environments. *Journal of Geophysical Research-Oceans 108* (C9).

Hetland, R. D.; DiMarco, S. F. (2008). How does the character of oxygen demand control the structure of hypoxia on the Texas-Louisiana continental shelf? *Journal of Marine Systems 70* (1-2), 49-62.

Hoff, P. D. (2009). A First Course in Bayesian Statistical Methods. In Springer-Verlag New York: New York, NY.

Justic, D.; Rabalais, N. N.; Turner, R. E. (1996). Effects of climate change on hypoxia in coastal waters: A doubled CO2 scenario for the northern Gulf of Mexico. *Limnology and Oceanography 41* (5), 992-1003.

Justic, D.; Rabalais, N. N.; Turner, R. E. (1997). Impacts of climate change on net productivity of coastal waters: implications for carbon budgets and hypoxia. *Climate Research 8* (3), 225-237.

Justic, D.; Rabalais, N. N.; Turner, R. E. (2001). Future perspectives for hypoxia in the northern Gulf of Mexico. *Coastal and Estuarine Sciences, Vol 58 58*, 435-449.

Justic, D.; Rabalais, N. N.; Turner, R. E. (2002). Modeling the impacts of decadal changes in riverine nutrient fluxes on coastal eutrophication near the Mississippi River Delta. *Ecological Modelling 152* (1), 33-46.

Justic, D.; Bierman, V. J.; Scavia, D.; Hetland, R. D. (2007). Forecasting Gulf's hypoxia: The next 50 years? *Estuaries and Coasts 30* (5), 791-801.

Kemp, W. M.; Testa, J. M.; Conley, D. J.; Gilbert, D.; Hagy, J. D. (2009). Temporal responses of coastal hypoxia to nutrient loading and physical controls. *Biogeosciences 6* (12), 2985-3008.

Kimmel, D. G.; Boicourt, W. C.; Pierson, J. J.; Roman, M. R.; Zhang, X. S. (2009). A comparison of the mesozooplankton response to hypoxia in Chesapeake Bay and the northern Gulf of Mexico using the biomass size spectrum. *Journal of Experimental Marine Biology and Ecology 381*, S65-S73.

Kitanidis, P. K.; Shen, K. F. (1996). Geostatistical interpolation of chemical concentration. *Advances in Water Resources 19* (6), 369-378.

Laurent, A.; Fennel, K.; Hu, J.; Hetland, R. (2012). Simulating the effects of phosphorus limitation in the Mississippi and Atchafalaya River plumes. *Biogeosciences 9* (11), 4707-4723.

Lehrter, J. C.; Murrell, M. C.; Kurtz, J. C. (2009). Interactions between freshwater input, light, and phytoplankton dynamics on the Louisiana continental shelf. *Continental Shelf Research 29* (15), 1861-1872.

Lehrter, J. C.; Beddick, D. L.; Devereux, R.; Yates, D. F.; Murrell, M. C. (2012). Sediment-water fluxes of dissolved inorganic carbon, O-2, nutrients, and N-2 from the hypoxic region of the Louisiana continental shelf. *Biogeochemistry 109* (1-3), 233-252.

Lin, X. A.; McCarthy, M. J.; Carini, S. A.; Gardner, W. S. (2011). Net, actual, and potential sediment-water interface NH4+ fluxes in the northern Gulf of Mexico (NGOMEX): Evidence for NH4+ limitation of microbial dynamics. *Continental Shelf Research 31* (2), 120-128.

Liu, Y.; Yang, P. J.; Hu, C.; Guo, H. C. (2008). Water quality modeling for load reduction under uncertainty: A Bayesian approach. *Water Research 42* (13), 3305-3314.

Liu, Y.; Evans, M. A.; Scavia, D. (2010). Gulf of Mexico hypoxia: exploring increasing sensitivity to nitrogen loads. *Environmental Science & Technology 44* (15), 5836-5841.

Lotze, H. K.; Lenihan, H. S.; Bourque, B. J.; Bradbury, R. H.; Cooke, R. G.; Kay, M. C.; Kidwell, S. M.; Kirby, M. X.; Peterson, C. H.; Jackson, J. B. C. (2006). Depletion, degradation, and recovery potential of estuaries and coastal seas. *Science 312* (5781), 1806-1809.

LUMCON (2012). Hypoxia in the Northern Gulf of Mexico: Research; http://www.gulfhypoxia.net/Research/. (Accessed October, 2012.)

Lunn, D. J.; Thomas, A.; Best, N.; Spiegelhalter, D. (2000). WinBUGS - A Bayesian modelling framework: Concepts, structure, and extensibility. *Statistics and Computing 10* (4), 325-337.

Lunn, D. J.; Jackson, C.; Best, N.; Thomas, A.; Spiegelhalter, D. (2013). *The BUGS book: A practical introduction to Bayesian analysis*. Chapman & Hall/CRC: Boca Raton, Fla.

Madigan, M. T.; Madigan, M. T.; Brock, T. D. (2009). *Brock biology of microorganisms*. 12th ed.; Pearson/Benjamin Cummings: San Francisco, CA.

Malve, O.; Laine, M.; Haario, H.; Kirkkala, T.; Sarvala, J. (2007). Bayesian modelling of algal mass occurrences - using adaptive MCMC methods with a lake water quality model. *Environmental Modelling & Software 22* (7), 966-977.

McCarthy, M. J.; Carini, S. A.; Liu, Z.; Ostrom, N. E.; Gardner, W. S. (2013). Oxygen consumption in the water column and sediments of the northern Gulf of Mexico hypoxic zone. *Estuarine, Coastal and Shelf Science 123*, 46-53.

Mueller, K. L.; Yadav, V.; Curtis, P. S.; Vogel, C.; Michalak, A. M. (2010). Attributing the variability of eddy-covariance CO2 flux measurements across temporal scales using geostatistical regression for a mixed northern hardwood forest. *Global Biogeochemical Cycles 24*.

Murphy, R. R.; Curriero, F. C.; Ball, W. P. (2010). Comparison of spatial interpolation methods for water quality evaluation in the Chesapeake Bay. *Journal of Environmental Engineering-Asce 136* (2), 160-171.

Murphy, R. R.; Kemp, W. M.; Ball, W. P. (2011). Long-Term Trends in Chesapeake Bay Seasonal Hypoxia, Stratification, and Nutrient Loading. *Estuaries and Coasts 34* (6), 1293-1309.

Murrell, M. C.; Lehrter, J. C. (2011). Sediment and Lower Water Column Oxygen Consumption in the Seasonally Hypoxic Region of the Louisiana Continental Shelf. *Estuaries and Coasts 34* (5), 912-924.

Murrell, M. C.; Stanley, R. S.; Lehrter, J. C.; Hagy, J. D. (2013). Plankton community respiration, net ecosystem metabolism, and oxygen dynamics on the Louisiana continental shelf: Implications for hypoxia. *Continental Shelf Research 52*, 27-38.

Nixon, S. W. (1995). Coastal Marine Eutrophication - a Definition, Social Causes, and Future Concerns. *Ophelia 41*, 199-219.

NOAA (2011a). Gulf of Mexico Hypoxia Watch; http://ecowatch.ncddc.noaa.gov/hypoxia. (Accessed 2011.)

NOAA (2011b). National Geophysical Data Center (NGDC) Coastal Relief Model; http://www.ngdc.noaa.gov/mgg/coastal/startcrm.htm. (Accessed June, 2010.)

NOAA (2012). Gulf of Mexico Hypoxia Watch: Hypoxia Contour Process; http://www.ncddc.noaa.gov/hypoxia/. (Accessed October, 2012.)

NRC (2000). *Clean coastal waters : understanding and reducing the effects of nutrient pollution*. National Research Council (U.S.) Ocean Studies Board, Water Science and Technology Board, National Academy Press: Washington, D.C.

Obenour, D. R.; Michalak, A. M.; Zhou, Y.; Scavia, D. (2012a). Sizing up the gulf's dead zone through geostatistical modeling (Poster available at http://www.ncddc.noaa.gov/activities/healthy-oceans/gulf-hypoxia-stakeholders/workshop-2012/proceedings/). In *3rd Annual Hypoxia Coordination Workshop*, Bay St. Louis, MS.

Obenour, D. R.; Michalak, A. M.; Zhou, Y.; Scavia, D. (2012b). Quantifying the impacts of stratification and nutrient loading on hypoxia in the northern Gulf of Mexico. *Environmental Science & Technology 46* (10), 5489-96.

Pace, M. L. (2003). The utility of simple models in ecosystem science. In *Models in ecosystem science*, Canham, C. D.; Cole, J. J.; Laurenroth, W. K., Eds. Princeton University Press: Princeton, NJ.

Patterson, H. D.; Thompson, R. (1971). Recovery of inter-block information when block sizes are unequal. *Biometrika 58* (3), 545-554.

Qian, S. S. (1997). Estimating the area affected by phosphorus runoff in an Everglades wetland: A comparison of universal kriging and Bayesian kriging. *Environmental and Ecological Statistics 4* (1), 1-26.

Quinones-Rivera, Z. J.; Wissel, B.; Justic, D.; Fry, B. (2007). Partitioning oxygen sources and sinks in a stratified, eutrophic coastal ecosystem using stable oxygen isotopes. *Marine Ecology Progress Series 342*, 69-83.

Rabalais, N. N.; Wiseman, W. J.; Turner, R. E.; SenGupta, B. K.; Dortch, Q. (1996). Nutrient changes in the Mississippi River and system responses on the adjacent continental shelf. *Estuaries 19* (2B), 386-407.

Rabalais, N. N.; Turner, R. E.; Justic, D.; Dortch, Q.; Wiseman, W. J. (1999). *Characterization of hypoxia*. National Oceanic and Atmospheric Administration (NOAA) Coastal Ocean Program Decision Analysis Series No. 15: Silver Spring, MD.

Rabalais, N. N.; Turner, R. E. (2001). *Coastal Hypoxia : consequences for living resources and ecosystems*. Coastal and estuarine studies 58; American Geophysical Union: Washington, D.C.

Rabalais, N. N.; Turner, R. E.; Dortch, Q.; Justic, D.; Bierman, V. J.; Wiseman, W. J. (2002a). Nutrient-enhanced productivity in the northern Gulf of Mexico: past, present and future. *Hydrobiologia 475* (1), 39-63.

Rabalais, N. N.; Turner, R. E.; Scavia, D. (2002b). Beyond science into policy: Gulf of Mexico hypoxia and the Mississippi River. *Bioscience 52* (2), 129-142.

Rabalais, N. N.; Turner, R. E.; Wiseman, W. J. (2002c). Gulf of Mexico hypoxia, aka "The dead zone". *Annual Review of Ecology and Systematics 33*, 235-263.

Rabalais, N. N.; Turner, R. E.; Sen Gupta, B. K.; Boesch, D. F.; Chapman, P.; Murrell, M. C. (2007). Hypoxia in the northern Gulf of Mexico: Does the science support the plan to reduce, mitigate, and control hypoxia? *Estuaries and Coasts 30* (5), 753-772.

Rabalais, N. N.; Diaz, R. J.; Levin, L. A.; Turner, R. E.; Gilbert, D.; Zhang, J. (2010). Dynamics and distribution of natural and human-caused hypoxia. *Biogeosciences 7* (2), 585-619.

Rabalais, N. N. (2011). Louisiana Hypoxia Surveys, available from NOAA National Oceanographic Data Center (NODC); http://www.nodc.noaa.gov/General/getdata.html. (Accessed January, 2011.)

Ramin, M.; Stremilov, S.; Labencki, T.; Gudimov, A.; Boyd, D.; Arhonditsis, G. B. (2011). Integration of numerical modeling and Bayesian analysis for setting water quality criteria in Hamilton Harbour, Ontario, Canada. *Environmental Modelling & Software 26* (4), 337-353.

Reckhow, K. (2003). On the need for uncertainty assessment in TMDL modeling and implementation. *Journal of Water Resources Planning and Management-Asce 129* (4), 245-246.

Reckhow, K. H. (1994). Importance of Scientific Uncertainty in Decision-Making. *Environmental Management 18* (2), 161-166.

Reckhow, K. H.; Chapra, S. C. (1999). Modeling excessive nutrient loading in the environment. *Environmental Pollution 100* (1-3), 197-207.

Redalje, D. G.; Lohrenz, S. E.; Fahnenstiel, G. L. (1994). The relationship between primary production and the vertical export of particulate organic-matter in a river-impacted coastal ecosystem. *Estuaries 17* (4), 829-838.

Redfield, A. C.; Ketchum, B. H.; Richards, F. A. (1963). The influence of organisms on the composition of seawater. In *The Sea*, Hill, M. N., Ed. John Wiley: New York; pp 26-77.

Refsgaard, J. C.; van der Sluijs, J. P.; Hojberg, A. L.; Vanrolleghem, P. A. (2007). Uncertainty in the environmental modelling process - A framework and guidance. *Environmental Modelling & Software 22* (11), 1543-1556.

Rencher, A. C. (1998). *Multivariate statistical inference and applications*. Wiley: New York.

Rowe, G. T. (2001). Seasonal hypoxia in the bottom water off the Mississippi River delta. *Journal of Environmental Quality 30* (2), 281-290.

Runkel, R. L.; Crawford, C. G.; Cohn, T. A. (2004). Load estimator (LOADEST) a FORTRAN program for estimating constituent loads in streams and rivers. In U.S. Dept. of the Interior, U.S. Geological Survey: Reston, Virginia; p 69.

Scavia, D.; Canale, R. P.; Powers, W. F.; Moody, J. L. (1981). Variance Estimates for a Dynamic Eutrophication Model of Saginaw Bay, Lake Huron. *Water Resources Research 17* (4), 1115-1124.

Scavia, D.; Rabalais, N. N.; Turner, R. E.; Justic, D.; Wiseman, W. J. (2003). Predicting the response of Gulf of Mexico hypoxia to variations in Mississippi River nitrogen load. *Limnology and Oceanography 48* (3), 951-956.

Scavia, D.; Justic, D.; Bierman, V. J. (2004). Reducing hypoxia in the Gulf of Mexico: Advice from three models. *Estuaries 27* (3), 419-425.

Scavia, D.; Donnelly, K. A. (2007). Reassessing hypoxia forecasts for the Gulf of Mexico. *Environmental Science & Technology 41* (23), 8111-8117.

Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics 6* (2), 461-464.

Smith, J. E.; Heath, L. S. (2001). Identifying influences on model uncertainty: An application using a forest carbon budget model. *Environmental Management 27* (2), 253-267.

Stow, C. A.; Qian, S. S.; Craig, J. K. (2005). Declining threshold for hypoxia in the Gulf of Mexico. *Environmental Science & Technology 39* (3), 716-723.

Strauss, J.; Grossman, E. L.; DiMarco, S. F. (2012). Stable isotope characterization of hypoxia-susceptible waters on the Louisiana shelf: Tracing freshwater discharge and benthic respiration. *Continental Shelf Research 47*, 7-15.

TAMU (2011). Mechanisms Controlling Hypoxia; http://hypoxia.tamu.edu/. (Accessed 2011.)

Tibshirani, R. (1996). Regression shrinkage and selection via the Lasso. *Journal of the Royal Statistical Society Series B-Methodological 58* (1), 267-288.

Turner, R. E.; Rabalais, N. N.; Swenson, E. M.; Kasprzak, M.; Romaire, T. (2005). Summer hypoxia in the northern Gulf of Mexico and its prediction from 1978 to 1995. *Marine Environmental Research 59* (1), 65-77.

Turner, R. E.; Rabalais, N. N.; Justic, D. (2006). Predicting summer hypoxia in the northern Gulf of Mexico: Riverine N, P, and Si loading. *Marine Pollution Bulletin 52* (2), 139-148.

Turner, R. E.; Rabalais, N. N.; Justic, D. (2008). Gulf of Mexico hypoxia: Alternate states and a legacy. *Environmental Science & Technology 42* (7), 2323-2327.

Turner, R. E.; Rabalais, N. N.; Justic, D. (2012). Predicting summer hypoxia in the northern Gulf of Mexico: Redux. *Marine Pollution Bulletin 64* (2), 319-324.

USGS (2010). Nutrient Flux for the Mississippi River Basin and Subbasins; http://toxics.usgs.gov/hypoxia/mississippi/nutrient_flux_yield_est.html. (Accessed Sept, 2010.)

Wackernagel, H. (2003). *Multivariate geostatistics : an introduction with applications*. 3rd, completely rev. ed.; Springer: Berlin ; New York.

Walker, N. D.; Wiseman, W. J.; Rouse, L. J.; Babin, A. (2005). Effects of river discharge, wind stress, and slope eddies on circulation and the satellite-observed structure of the Mississippi River plume. *Journal of Coastal Research 21* (6), 1228-1244.

Walker, N. D.; Rabalais, N. N. (2006). Relationships among satellite chlorophyll a, river inputs, and hypoxia on the Louisiana continental shelf, gulf of Mexico. *Estuaries and Coasts 29* (6B), 1081-1093.

Wang, L. X.; Justic, D. (2009). A modeling study of the physical processes affecting the development of seasonal hypoxia over the inner Louisiana-Texas shelf: Circulation and stratification. *Continental Shelf Research 29* (11-12), 1464-1476.

Wiseman, W. J.; Kelly, F. J. (1994). Salinity variability within the Louisiana coastal current during the 1982 flood season. *Estuaries 17* (4), 732-739.

Wiseman, W. J.; Rabalais, N. N.; Turner, R. E.; Dinnel, S. P.; MacNaughton, A. (1997). Seasonal and interannual variability within the Louisiana coastal current: stratification and hypoxia. *Journal of Marine Systems 12* (1-4), 237-248.

Zavala-Hidalgo, J.; Morey, S. L.; O'Brien, J. J. (2003). Seasonal circulation on the western shelf of the Gulf of Mexico using a high-resolution numerical model. *Journal of Geophysical Research-Oceans 108* (C12), C123389.

Zhang, H. Y.; Ludsin, S. A.; Mason, D. M.; Adamack, A. T.; Brandt, S. B.; Zhang, X. S.; Kimmel, D. G.; Roman, M. R.; Boicourt, W. C. (2009). Hypoxia-driven changes in the behavior and spatial distribution of pelagic fish and mesozooplankton in the northern Gulf of Mexico. *Journal of Experimental Marine Biology and Ecology 381*, S80-S91.

Zhang, X. Q.; Hetland, R. D.; Marta-Almeida, M.; DiMarco, S. F. (2012). A numerical investigation of the Mississippi and Atchafalaya freshwater transport, filling and

flushing times on the Texas-Louisiana Shelf. *Journal of Geophysical Research-Oceans 117*, C11009.

Zhou, Y.; Obenour, D. R.; Scavia, D.; Johengen, T. H.; Michalak, A. M. (2013). Spatial and temporal trends in Lake Erie hypoxia, 1987-2007. *Environmental Science & Technology 47* (2), 899-905.

Zhou, Y.; Scavia, D.; Michalak, A. M. ((submitted manuscript)). Nutrient loading and meteorological conditions explain interannual variability of hypoxia in the Chesapeake Bay.

Zimmerman, D. L. (2010). Likelihood-based methods. In *Handbook of spatial statistics*, Gelfand, A. E.; Diggle, P. J.; Fuentes, M.; Guttorp, P., Eds. CRC Press: Boca Raton, FL.

Zimmerman, D. L.; Stein, M. (2010). Classical geostatistical methods. In *Handbook of spatial statistics*, Gelfand, A. E.; Diggle, P. J.; Fuentes, M.; Guttorp, P., Eds. CRC Press: Boca Raton, FL.

Zou, H.; Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society Series B-Statistical Methodology 67*, 301-320.