Improved Estimates of the Spatial Distribution and Temporal Trends of Water Quality Parameters Using Geostatistical Data Fusion Methods

by

Yuntao Zhou

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

Doctoral Committee:

Associate Professor Anna M. Michalak, Chair
Professor Donald Scavia
Professor Peter Adriaens
Emeritus Professor Guy Meadows

ACKNOWLEDGEMENTS

First and foremost, I thank my advisor Anna M. Michalak for the support in various ways during my PhD studies. I have learned a great deal from Anna's insight, comments, and suggestions. Anna is a passionate scientist with unmatched insights. She has inspired many of the ideas in this work and helped me overcome many of the obstacles along the way. I appreciate all her contributions of time, ideas, and funding to make my Ph.D. experience productive and stimulating.

I thank Don Scavia, who was my MS advisor and is my PhD committee member. Don is a co-author of my paper during my PhD research. I really appreciate his valuable comments. I would like to thank my other committee members, Professor Peter Adriaens and Professor Guy Meadows, who have provided me helpful inputs of this dissertation.

I would also like to thank everyone in the Michalak Research Group (PUORG), both alumni and current members. They always offered insightful comments on my presentations and manuscripts. I also thank all the members for their friendship throughout this PhD. Thank you, Abhishek, Dan, Dorit, Debbie, Jeff, Kim, Sharon, Vineet, Yuanyuan and Yoichi! I really enjoy their friendship, advice and collaboration.

I am also grateful for support from National Science Foundation under Grant No. 0644648.The Rackham Graduate School has generously awarded me several Travel Grants to national and international conferences.

Last but not least, I am deeply thankful and grateful to my family for their sacrifice, patience, and encouragement. I would also like to thank my friends, especially Lulu, Qiang, Jing, Fang,

Wenjun and Yuan who all helped me keep some balance in my life. I am not sure how I would have survived the past six and half years without their support.

Portions of this dissertation were adapted from the following published studies:

- Zhou, Y., and A. M. Michalak (2009), Characterizing Attribute Distributions in Water Sediments by Geostatistical Downscaling, *Environmental Science & Technology*, *43* (24), 9267-9273, doi: 10.1021/es901431y.
- Zhou, Y., D. Obenour, D. Scavia, T. Johengen, and A. M. Michalak (2013), Spatial and Temporal Trends in Lake Erie Hypoxia, 1987-2007, *Environmental Science & Technology*, *47* (2), 899−905, doi: 10.1021/es303401b.

The co-authors of these publications are gratefully acknowledged.

Table of Contents

List of Figures

List of Tables

ABSTRACT

Improved Estimates of the Spatial Distributions and Temporal Trends of Water Quality Parameters Using Geostatistical Data Fusion Methods

by

Yuntao Zhou

Chair: Anna M. Michalak

Strategies aimed at reducing the degradation of water quality and predicting future changes in surface waters resulting from natural and anthropogenic forcing rely on the ability to track water quality changes, and to accurately quantify the distribution of water quality attributes. The three components of this dissertation focus on developing geostatistical data fusion methods that make optimal use of the available monitoring data in the Passaic River, Lake Erie, and the Chesapeake Bay, respectively.

The first component presents a method for accurately estimating the spatial distribution of the total organic carbon in the sediments of the Passaic River using a dataset with non-uniform resolution. Estimating the spatial distribution of water sediment attributes at a uniform spatial resolution is often required for site characterizations and the design of appropriate risk-based remediation alternatives. Using a pseudodata example, a noval geostatisitical downscaling approach is shown to yield better estimates with a more accurate assessment of uncertainties, relative to traditional kriging methods. When applied to the estimation of the distribution of total organic carbon, geostatistical downscaling shows that the uncertainty associated with the spatial

distribution of attribute is higher than would have been assumed if a kriging approach had been applied.

The second and third components explore the degradation of water quality in time and space. Specifically, hypoxia (low dissolved oxygen) has been observed in Lake Erie and Chesapeake Bay since the early 1900s, leading to negative impacts such as ecosystem habitat degradation, altered migration patterns, and decreased fishery production. The interannual variability in hypoxic extent since mid-1980s in these two systems is quantified by combining spatially explicit auxiliary data with *in situ* dissolved oxygen measurements. The significance of nutrient loading, weather patterns, and stratification in explaining hypoxia in these systems is also explored. This research points to strong meteorological controls on hypoxia, through impacts on stratification and nutrient loading, in addition to the impact of anthropogenic activities.

Overall, the developed geostatistical data fusion methods are shown to provide a means for producing reliable estimates of water quality attributes along with their associated uncertainties.

CHAPTER 1 Introduction

1.1 Surface water quality

The growing scarcity of fresh and clean water due to a variety of anthropogenic activities (such as irrigation, industry, recreation, and waste treatment) is one of the most important issues facing the world today [*U.S. EPA*, 2008]. Hypoxia (low dissolved oxygen), overfishing, harmful algal blooms (HABs), contaminated sediments, and physical alteration of habitats associated with coastal developments are several high-priority problems that decrease the ecological health of waters [*Committee on Environment and Natural Resources*, 2010]. This dissertation contributes to a better understanding of two of these problems, namely hypoxia and contaminated sediments.

Among the results of human activities, excessive input of nutrients (e.g., nitrogen, phosphorous) can cause the deterioration of water quality in surface water [*National Research Council*, 2000]. The increase in nutrients has led to eutrophication (defined as the increase of organic matter), a major environmental problem in estuarine and coastal waters throughout the world [*Nixon et al.*, 1995], and in inner to mid-continental shelf waters [*Rabalais*, 2005]. It is therefore necessary to identify the major anthropogenic activities impacting the eutrophication in aquatic ecosystems, and to recommend remedial actions for improving the health and viability of these systems.

Eutrophication can cause accelerated algae production, which depletes dissolved oxygen (DO) in water and further degrades water quality. When the DO concentrations drop below a certain threshold (normally 2 mg/L), this phenomenon is called hypoxia. Hypoxia can cause a loss of submerged aquatic vegetation and habitat, and can kill fish and other organisms. Tourism, boat manufacturing, fisheries, and other commercial activities that thrive on the wealth of natural

resources that estuaries supply are also affected [*NSTC*, 2003]. Therefore, reducing hypoxia and improving the quality of available surface water are necessary to benefit human health, natural ecosystems, and economic stability in the long term [*Kurunc et al.*, 2005]. To better understand and predict hypoxia, it is necessary to address the seasonal and interannual variability of hypoxia and to know how specific drivers (e.g., nutrient loading, stratification) affect the variability of hypoxia.

Sediment contaminants, including metals and organics, have been identified as a serious environmental problem in estuaries around the world [*Ridgway and Shimmield*, 2002; *Wells and Hill*, 2004]. Contaminants are introduced into the sediments from run-off, direct discharge, and atmospheric deposition [*National Research Council*, 2007]. While metals are found naturally in the natural environment, enrichment over background levels of certain trace metals can be attributed to human activities [*Valettesilver*, 1993]. Organic contaminants, which include, but are not limited to, pesticides, polychlorinated biphenyls (PCBs), and polycyclic aromatic hydrocarbons, (PAHs) come from anthropogenic sources [*National Research Council*, 2007]. These pollutants threaten the environment as they pose potentially dangerous human health risks.

Remediation of contaminated sediments (e.g., heavy metals, Polychlorobiphenyls, and pesticides) can be extremely expensive [*Adriaens et al.*, 2006]. In order to formulate effective remediation strategies, it is of great importance to develop methods capable of providing detailed knowledge about the spatial distribution of contamination in the sediment.

1.2 Importance of data fusion methods

Decisions regarding water management are difficult to make without a thorough understanding of the causes and consequences of degradation of water quality. Water management and protection authorities need extensive information about the state of aquatic

ecosystems in order to plan appropriate actions to improve water quality and the sustainability of water ecosystems [*National Research Council*, 2000].

Reliable knowledge of impact of natural processes and human activities on surface water is required to develop management strategies and to make accurate water quality forecasts for days, months or years into the future. Effective and comprehensive monitoring, data interpretation, and modeling are all needed to provide information for the development of adequate water quality management. Continuous monitoring in the long term benefits our understanding of water quality changes. Nevertheless, it is impractical and financially infeasible to implement a monitoring plan that captures comprehensive variability of parameters in surface water or sediment in both space and time. At present, our understanding of the degradation of water quality greatly depends on effective water quality models using *in situ* measurements. However, there exist many data limitations such as missing data, mixing of coarse and fine resolution of sample data, temporal and spatial gaps, and declining coverage [*Clark et al.*, 2001]. These imperfect data usually cause high levels of uncertainty in water quality models. Predictions on how unprecedented natural processes and human activities will affect the water environment are, therefore, limited.

Non-uniform sampling

In some cases, available data are obtained at a variety of spatial resolutions (i.e. support) due to the constraints of sampling instruments and natural environment. For example, information about the attributes (e.g., nutrients, organic matter, contaminants) in water sediments (i.e., benthic sediments) is typically obtained from cores. Each core is divided into several core sections of varying thickness. The reported value typically represents an average within individual segments of each core, and the sampling resolution is therefore not uniform [*Barabas*

et al., 2001]. Spatial resolution for the measurements is important because: (i) mechanisms vital to the spatial dynamics of a process at one resolution may be unimportant or inoperative at another, and (ii) relationships between variables at one resolution may be obscured or distorted when viewed at another resolution [*Gotway and Young*, 2002]. Therefore, a method for estimating the distribution of attributes at a uniform spatial resolution is of great importance.

Sparse sampling

In situ measurements of attributes are usually insufficient to gain an accurate understanding of their spatial structure in the surface water. Lack of systematic and complete data sets challenges both spatial and temporal predictions at unsampled locations. In addition to the *in situ* point measurements, more and more high-density remote sensing (e.g., satellite) data are also available for providing information on water quality attributes (e.g., chlorophyll) in surface water. To maximize the use of both *in situ* data and remote sensing data, it is necessary to develop models that can assimilate data from different sources and of different kinds.

1.3 Purpose of this dissertation

The ultimate purpose of this dissertation is to explore the causes of degrading water quality by making optimal use of sparse water quality monitoring data (Figure 1-1). To achieve this goal, this dissertation provides examples of spatial data integration techniques using geostatistical data fusion (i.e., combining different sources of information into a single final data set [*Fasbender et al.*, 2008]). Geostatistical methods were selected because they have the advantage of considering the spatial correlation of the data (to be discussed in Chapter 3). By putting the available monitoring data to optimal use, the causes of degrading water quality and the solutions to improve water quality could be found with more thorough analysis and understanding.

Specifically, this dissertation is divided into three components according to the scientific questions being addressed. The first component focuses on identifying the spatial distribution of sediment attributes (e.g., total organic carbon) in the Passaic River. The last two components involve the characterization of hypoxia in Lake Erie (the most productive lake of the five Great Lakes) and Chesapeake Bay (the most productive estuary in the United States), both of which have a long history of hypoxia [*NSTC*, 2000].

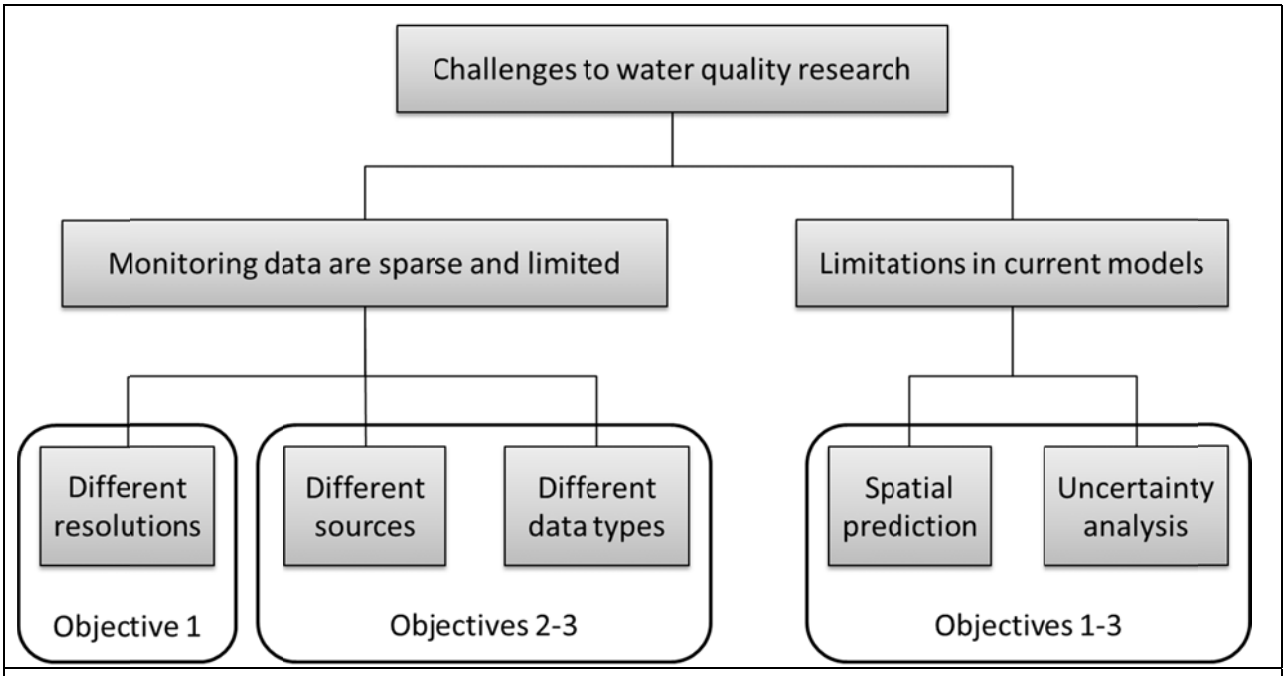

Figure 1-1 Summary of current challenges to water quality research and the objectives of this dissertation.

1.3.1 Objective 1: Geostatistical downscaling for non-uniform resolution data sets

The first component of this dissertation develops an approach for optimal inference of the spatial distribution of sampled attributes in water sediment, using a geostatistical downscaling technique. Remediation of sediments can be extremely expensive, and often requires detailed knowledge about the distribution of attributes in the subsurface in order to formulate effective remediation strategies [*National Research Council*, 2007]. In addition, only the average values of attributes in the core sections of varying lengths are measured in benthic sediments [*Barabas et*

al., 2001]. However, an estimate of the attribute distribution at a uniform spatial resolution is often required for site characterization or the design of appropriate risk-based remediation alternatives.

The purpose of this first work is to estimate the unknown fine resolution attribute distribution. Uniform resolution (i.e. the area over which a sample averages) is the basic requirement for almost all the existing interpolation methods [*Chiles and Delfiner*, 1999]. Although frequently used to interpolate measurements, traditional kriging methods are specially designed for the data sampled at a uniform resolution. To make optimal use of the non-uniform data set, the proposed geostatistical downscaling method is applied to both pseudodata and field data (i.e., total organic carbon) in Passaic River.

1.3.2 Objective 2: Estimating areal extent of hypoxia in Lake Erie with geostatistical methods

The second component of this dissertation focuses on estimating hypoxic extent in Lake Erie by integrating water quality data from different sources, with the ultimate goal of informing fisheries and leading to better water quality management. Traditional *in situ* sampling data usually cannot produce accurate spatial and temporal information without high-density measurement campaigns that are typically expensive [*Schwab et al.*, 1999]. For Lake Erie, there are only ten regular stations measuring dissolved oxygen (DO) in its central basin. To augment these sparse *in situ* sampling DO data, auxiliary variables (e.g., data representing spatial trend, remote sensing data measuring water quality) that are correlated with the DO measurement are used. Auxiliary variables are those associated with the distribution of the target variable in time and/or space.

In order to achieve the goal, Bayesian Information Criterion was used first to select the significant auxiliary variables that explain the spatial distribution of DO concentration in the bottom of the lake. Using Universal Kriging (UK, a.k.a. kriging with an external drift), the spatial distribution of DO throughout Lake Erie for each year from 1987 to 2007 was estimated in two dimensions. In addition, conditional realizations were applied to provide quantitative estimates of the areal extent of hypoxia in the central basin of Lake Erie for August and September, and their associated uncertainties.

1.3.3 Objective 3: Impacts of nutrients and weather patterns to the temporal variability of hypoxic volume in Chesapeake Bay

The third component of this dissertation explores both the causes of variability in hypoxic volume and the temporal duration of hypoxic events in Chesapeake Bay for the years 1985 to 2010 by using the available data that are correlated with hypoxia. Both the spatial extent and temporal duration of hypoxia affect spawning grounds, migratory pathways, and feeding habitats of fish species [*Hagy et al.*, 2004]. The approach outlined in section 1.3.2 is expanded from two- to three-dimensions due to the extremely complex physical and biochemical dynamics of the Bay.

The purpose of this work is to estimate the variability of hypoxia from April to October in half-monthly resolution and to examine the effects of nutrient loading and other weather patterns (e.g., precipitation, wind) on the variability in hypoxia. In addition, this work establishes the hypoxic frequency from 1985 to 2010 at every location in the Bay.

In summary, this dissertation contributes to water quality research by developing geostatistical methods capable of providing more accurate estimates based on available data. All the methods described here could easily be applied to other aquatic ecosystems facing similar

problems or challenges (e.g., data with different resolutions or data from different sources). In addition to providing technical and methodological developments, this dissertation also answers scientific questions by using these methods. For example, this dissertation maps spatial estimates of water sediment attributes (i.e., total organic carbon) with the associated uncertainties that serve as an important method for site remediation by providing the detailed spatial distribution of these attributes. The spatial distribution is required for either full river remediation or hot spot removal (i.e., removal of small areas of highly contaminated sediments). In addition, this dissertation also provides long-term historical information about hypoxic extent in Lake Erie and Chesapeake Bay from the mid-1980s to the late 2000s. Such information can provide a basis for the research on hypoxic forecasting. With this detailed information on hypoxic extents over the last two-plus decades, this dissertation explores the causes of hypoxia, which could provide detailed suggestions and strategies (e.g., for nutrient management) to policy makers.

The remainder of this dissertation is organized as follows. Chapter 2 provides a detailed overview of the complexities in water research related to the three objectives of this dissertation and reviews previous studies relevant to the presented work. Chapter 3 briefly describes the main methods applied for this dissertation. Chapter 4 compares the spatial distribution of water attributes estimated from geostatistical downscaling with that from traditional interpolation methods, using both pseudodata and field data from the sediment of Passaic River, New Jersey. Chapter 5 derives long-term (1987-2007) hypoxic extent estimates for Lake Erie by combining *in situ* DO measurements with auxiliary data. Chapter 6 estimates the seasonal and interannual variation in hypoxic volume in Chesapeake Bay from 1985 to 2010 and explores how nutrient loads and other weather patterns affecting it. The final chapter (Chapter 7) concludes this

dissertation by summarizing the results observed and by suggesting promising avenues for future work.

CHAPTER 2 Literature Review

This chapter summarizes the background and the state of science related to the three components discussed in Chapter 1. This literature review also covers methodological studies with some discussion of scientific results and a brief introduction to the study sites.

2.1 Mapping the distribution of water attributes (e.g., organic matter, contaminants) in the sediment

Contaminated sediments in aquatic environments can pose health risks to many types of organisms, including humans. Sediments are of great importance for small creatures such as worms, crustaceans, and insect larvae that inhabit the bottom of a body of water (i.e., benthic environment). In addition to providing important habitats for these aquatic organisms, sediments play a significant role in determining the overall environmental quality of an aquatic ecosystem. Toxic sediments can kill benthic organisms, and thus reduce the quantity of food available to larger animals such as fish. Additionally, some contaminants with hydrophobicity and low degradability in the sediments are taken up by benthic organisms through a process called bioaccumulation [*Vanderoost et al.*, 1988]. When larger animals feed on these contaminated prey species, the pollutants are taken into their bodies and are passed along to other animals in the food web. In particular, high concentrations can be found in natural aquatic organisms, while at the same time concentrations found in water are very low, even below detection limits [*Vanderoost et al.*, 1988]. As a result of toxic and bioaccumulative substances, benthic organisms, fish, birds, mammals, and human being can be adversely affected by contaminated sediments [*MacDonald et al.*, 2002]. Therefore, it is necessary to develop efficient strategies for removing contaminants from the sediments (i.e., sediment remediation) [*National Research Council*, 2007].

Currently, there are three main well-developed technologies for sediment remediation: dredging (a.k.a., excavation), capping (a.k.a., isolation), and monitored natural recovery (MNR) [*Foerstner and Apitz*, 2007; *U. S. EPA*, 2005]. Dredging removes contaminated sediments from an aquatic environment with the purpose of disposing of them at a different location. In the US, about six million yards of contaminated sediment have been removed and disposed of using dredging technology as part of 71 major remediation projects, because dredging results in permanent removal of the contaminants [*Zeller and Cushing*, 2006]. However, it only shifts the problem to another place, as the removed sediment has to be deposited elsewhere and requires further management [*Perelo*, 2010]. Capping, by contrast, places clean dredged materials over a deposit of contaminants in open-water or upland locations as a means of isolating these sediments from the surrounding environment. However, contaminated sediment can be mixed with the clean material or can be resuspended during the placement of the dredged materials, and sedimentary natural attenuation processes can be altered [*Himmelheber et al.*, 2007]. The resuspension of contaminated sediments is a frequently recurring ecological threat in contaminated aquatic habitats. MNR uses naturally occurring processes (i.e., physical, biological, and chemical mechanisms) to contain, destroy, or reduce the organic matter or contaminants in sediment [*Magar and Wenning*, 2006]. However, MNR generally takes the longest time to achieve protection than dredging and *in situ* capping [*National Research Council*, 2007]. Overall, all remediation technologies have both advantages and disadvantages when applied at a particular site, and it is critical for risk management to identify a suitable technology for each site. Site conditions are of great importance in determining which remediation techniques (and combinations thereof) are appropriate [*National Research Council*, 2007].

While beneficial for qualitative site characterization and efficient management and selection of remediation technologies, sparse sampling provides insufficient information or detail about the spatial distribution of the sampled attributes throughout the sampled area. Therefore, interpolation methods, including mathematical splines, inverse distance weighted interpolation, and geostatistical methods [*French et al.*, 1995; *Kravchenko and Bullock*, 1999; *Leenaers et al.*, 1990], are typically used to map attribute distribution throughout a site.

Due to the irregular thickness of sediment above the bedrock surface, samples of contaminated stream or riverbed sediments are often obtained from cores that vary in length. Each core is subsequently subdivided into several sections of varying thickness, which are analyzed for key attributes. The reported values at each location typically represent an average within individual segments of each core, and the sampling resolution (a.k.a. support) is therefore non-uniform [*Barabas et al.*, 2001]. One method of overcoming the cross-resolution limitations is to rescale the measurements according to the requirements of the model simulation process. Most interpolation approaches (such as inverse distance weighting or point-to-point kriging) are only designed for data sets with uniform resolutions for all variables [*Chiles and Delfiner*, 1999]. Therefore, the cross-resolution issue is usually ignored during the process of applying interpolation, which will lead to errors in the estimated parameter distributions (REF). Most of the time, parameters with different resolutions are averaged up to those with the largest resolution before analysis, which results in a loss of information at small resolutions.

The direct application of traditional interpolation methods, including traditional geostatistical methods, for data with non-uniform resolution, as is typical of data sets describing contaminated sediments, is itself problematic for several reasons. First, a uniform resolution is required by a majority of interpolation approaches, including those that fall within the framework of

geostatistical interpolation (i.e. point-to-point kriging). Therefore, in order to apply point-to-point kriging methods, one needs to make some assumptions regarding the distribution of the attribute within the individual core sections. In general, this is done by performing the spatial interpolation at a resolution finer than the resolution of the core sections, and assuming either (i) that the sampled average value is representative of the attribute value at the center of each original section, or (ii) that the measured average value is representative of the attribute value at every point in the core section [*Barabas et al.*, 2001]. Violating the requirement for uniform measurement support in this way can yield inaccuracies in both the estimated field and its associated estimation error. Second, the variance of a spatially distributed attribute usually changes with the spatial support of measurements [*Chiles and Delfiner*, 1999], with apparent variance usually decreasing with coarser resolution, while the converse is true with correlation lengths [*Isaaks and Srivastava*, 1989; *Western and Bloschl*, 1999].

As a result, accurate information about the spatial structure at the resolution of the final estimates, including measures of overall variability and correlation length that are required for geostatistical analysis, is difficult to obtain using a data set with variable resolution. When data with variable support are used for estimating variability, the true variability can be underestimated due to the averaging process inherent to sampling in non-uniform core sections.

The geostatistical framework contains many unique features for making optimal use of the limited water quality data to achieve statistically rigorous estimates, and provides a broad theoretical framework for addressing the problem of water data fusion [*Chiles and Delfiner*, 1999; *Cressie*, 1993]. A general overview of data fusion theory from the spatial statistical perspective can be found in *Gotway and Young* [2002].

In order to preserve all the information provided by the measurements, estimating attribute values at a smaller and uniform resolution (i.e., downscaling) is an option for solving the mixed-resolution problem. Downscaling in a geostatistical co-kriging framework has been successfully applied in past studies [*Pardo-Iguzquiza et al.*, 2006], but requires the joint estimation of multiple attributes.

Geostatistical downscaling in the form of inverse modeling has the ability to produce estimates at a finer resolution than that of its component measurements by including a sensitivity matrix corresponding to the contribution of small resolution data to large resolution data. Geostatistical inverse modeling has been widely used in hydrogeology for characterizing contaminant sources [*Michalak and Kitanidis*, 2004; *Sun*, 2007], as well as identifying the distribution of hydraulic conductivities or transmissivities in aquifers [*Fienen et al.*, 2004; *Yeh and Zhang*, 1996; *Zimmerman et al.*, 1998]. *Fienen et al.* [2004] generated hydraulic conductivity in borehole flow-meter tests, which indicates the feasibility of downscaling by inverse modeling. *Erickson and Michalak* [2006] also produced a unified data set using different remote sensing observations of differing resolutions, grid orientations, and spatial extent from multiple sensors by applying a geostatistical inverse modeling approach.

2.2 Brief review of the study site

This dissertation provides an example of mapping the spatial distribution of attributes in the sediments of the Passaic River.

Passaic River

The Passaic River is a tidal tributary to Newark Bay and part of the New York and New Jersey Harbor Estuary, and its flow pattern is determined by both seasonally varying freshwater input and tidal events [*Barabas et al.*, 2001]. The study area of the first component of this

dissertation is a 10 km stretch of the lower Passaic River in New Jersey, under investigation by the EPA as part of its evaluation of the Diamond Alkali Superfund site (http://www.ourpassaic.org). Nearly two centuries of industrial activities and nonpoint sources of pollution have resulted in elevated levels of dioxins in the sediments [*Gunster et al.*, 1993].

The concentrations of heavy metals (e.g., mercury, zinc, lead, and copper), Polycyclic Aromatic Hydrocarbons (PAHs), Polychlorinated Biphenyls (PCBs), and pesticides (DDT) are of particular concern in Passaic River sediment [*Crawford et al.*, 1995; *Donovan et al.*, 2008; *Walker et al.*, 1999], not only because of their high level of toxicity at low concentrations but also because of their persistence in the environment. Resident aquatic organisms are exposed to and can bioaccumulate a variety of chemical contaminants from sediments, water and other organisms.

Spatial distribution of contaminants is required to decide whether a full river remediation or hot spot removal would be required. Because the river is tidal, the water, the sediment, and most species of fish and invertebrates (i.e., blue clawed crabs) travel throughout and outside of the Passaic River. Storm events and other natural processes might disturb the sediments. Alternatives ranging from no action to dredging to capping and their combinations within different areas should be evaluated. The potential long-term risk reduction associated with these approaches requires knowledge of contaminant stability in sediment [*Adriaens et al.*, 2006]. Therefore, an interpolation method that is able to provide accurate spatial analysis of contaminants in sediment is needed.

Total organic carbon

Total organic carbon (TOC) from the lower Passaic River was used analyzed as an example, using the geostatisitical data fusion tools developed in this dissertation. TOC forms water-soluble

and water-insoluble complexes with metal ions and hydrous oxides, interact with clay minerals and bind particles together, as well as sorbs and desorbs both naturally-occurring and anthropogenically-introduced organic contaminant compounds [*Lake et al.*, 1990; *Schumacher*, 2002]. Organic carbon comes from land sources (e.g., leaves, grasses, soils, and wastewater treatment plants) that run off into water, and water sources produced in the aquatic plants (e.g., phytoplankton, seaweeds). TOC content in sediments has been used as an indicator of pollution and eutrophication rate [*Folger*, 1972; *U. S. EPA*, 2002a]. The analysis of TOC distribution is therefore essential for the further analysis of contaminant distribution knowledge.

2.3 Causes and impacts of hypoxia on surface water

Globally, hypoxia (low dissolved oxygen, DO) threatens many large bodies of water around the world, including the Baltic Sea [*Sandberg*, 1994], the Black Sea [*Daskalov*, 2003], the Yangtze River Estuary [*Chen et al.*, 2007; *Li et al.*, 2002], the Gulf of Mexico [*Rabalais et al.*, 2001], Long Island Sound [*Parker and Oreilly*, 1991; *Welsh and Eller*, 1991], Chesapeake Bay [*Breitburg*, 1990; *Dauer et al.*, 1992; *Officer et al.*, 1984; *Sanford et al.*, 1990] and Lake Erie [*Burns et al.*, 2005; *Carrick et al.*, 2005]. Hypoxia has also developed in the eastern Pacific Ocean, the south Atlantic west of Africa, the Arabian Sea, and the Bay of Bengal [*Diaz and Rosenberg*, 2008]. In the United States, increased nutrient inputs to bodies of water have led to substantial changes in the Great Lakes and two-thirds of all coastal systems [*National Research Council*, 2000].

Human activities alter the global cycling of nutrients enormously, especially the movement of nutrients to bodies of water [*National Research Council*, 2000]. Economic development, population growth, and related human activities have increased nutrient inputs to levels that are many times higher than those that occur naturally. Besides the increasing direct inputs of

agricultural fertilizer and industrial sewage, other human activities including land use change and dam construction have increased nutrient loading indirectly [*Hopkinson and Vallino*, 1995]. Nutrient enrichment stimulates algae production and biomass accumulation in freshwater and coastal marine ecosystems [*Cloern*, 1999].

It is clear that the most common single factor causing low DO in these environments is an increase in the amount of nutrient loading they receive [*Nixon et al.*, 1995]. Nutrient loading is defined as the total amount of various nutrients contributed by the upstream landscape and atmosphere [*National Research Council*, 2000]. Phosphorous and nitrogen are the two main nutrients leading to the degradation of surface water. Nitrogen is generally considered the most frequent driver of estuarine low oxygen conditions and currently is mostly the result of non-point sources (e.g., agriculture) [*NSTC*, 2003], while phosphorous is generally the most frequent driver of freshwater low oxygen conditions and most commonly comes from both non-point sources (e.g., agriculture) and point sources (e.g., wastewater treatment) [*NSTC*, 2003; *Rucinski et al.*, 2010].

In addition to increasing nutrient loading, summertime stratification of surface waters is another main cause of hypoxia. Vertical stratification isolates colder waters in the bottom layer, where low light restricts photosynthesis, from re-aeration and the diffusion of oxygen from the surface to the bottom [*National Research Council*, 2000]. Moreover, the strong vertical stratification can retain and increase the phytoplankton blooms in the upper zone where light and nutrients are more favorable, and thus encourage higher production as well as larger algal blooms and grazers [*NSTC*, 2003]. The settlement of phytoplankton and grazers within the bottom layer results in the depletion of DO from deep waters. Decomposing phytoplankton and other settled organic matter consumes oxygen, often reducing it to below the 2 mg/L hypoxic

threshold in the bottom layer. When the DO is completely consumed, anoxia, an extreme condition, occurs [*Committee on Environment and Natural Resources*, 2010].

Taking the hypoxic condition in coastal waters (e.g., Chesapeake Bay) as an example, Figure 1-1 compares a hypoxic ecosystem (left) with a healthy (i.e., non-hypoxic) ecosystem (right). Density stratification of the water column, in which a less dense layer of water floats on top of a denser bottom layer, is an almost universal characteristic of coastal systems subject to seasonal bottom water hypoxia. Stratification reduces the potential for oxygen from the atmosphere to replenish oxygen depleted at bottom layer of water. In most cases involving marine systems, a vertical gradient of salinity, creating a halocline or pycnocline (Figure 1-1), is the most important factor contributing to density stratification [*NSTC*, 2000]. While in most freshwater systems, surface heating, creating warmer surface water temperatures and thus a vertical gradient of temperature or a thermocline, is the cause of stratification [*Hawley et al.*, 2006]. In addition, wind and tidal stresses can upset the stabilizing effects and are likely to be important mechanisms for the mixing of the upper freshwater and lower seawater in estuarine ecosystems [*Committee on Environment and Natural Resources*, 2010].

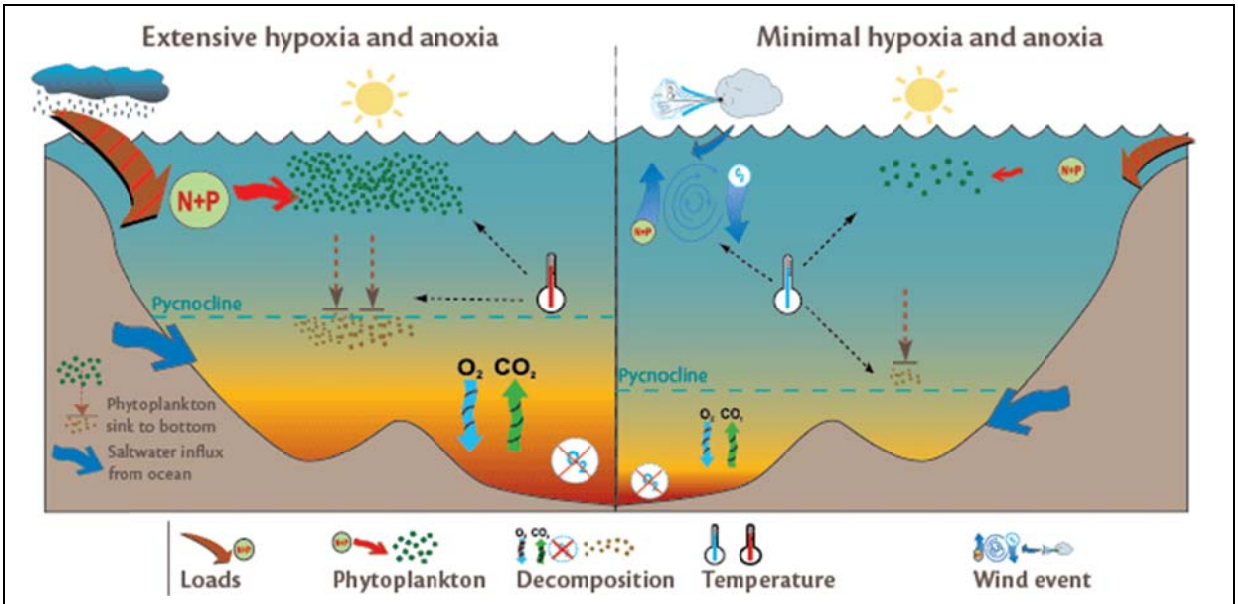

Figure 2-1 Summary of hypoxia in coastal water. Source: Ecocheck.

There is a wide range of potential negative effects resulting from hypoxia, such as decreased light availability and algal dominance changes [*NSTC*, 2003]. These primary symptoms can further lead to degradation of habitat, altered migration patterns, decreased fishery production, and subsequent economic impacts on industries dependent on ecosystem productivity [*NSTC*, 2003]. Although mobile organisms can flee the affected area when DO becomes too low, fish kills can result from hypoxia in the long term, especially when the concentration of DO drops rapidly [*Committee on Environment and Natural Resources*, 2010]. The deaths of immobile and mobile organisms due to hypoxia can result in significant impacts on aquatic food webs and the economy. Commercial activities (e.g., tourism, boat manufacturing, and fisheries) that thrive on the wealth of natural resources of lakes or estuaries may also be affected [*Nixon et al.*, 1995].

2.3.1 Methods of estimating hypoxic extent

Simple linear and nonlinear regressions have been applied to estimate the extent of hypoxia in some regions such as Chesapeake Bay (e.g., [*Hagy et al.*, 2004; *Zhang et al.*, 2006]), using parameters that include temperature, salinity, and average river flow as covariates. However,

such work only estimates the magnitude of hypoxic extent, but not its spatial location. To obtain systematic and complete data sets of hypoxic extent that can be used to better understand hypoxia, interpolation of DO concentrations at un-sampled locations (i.e., location of hypoxia) is needed. The following section discusses different types of models that can be used for estimating both the location and the extent of hypoxia.

Mechanistic models

Mechanistic models focused on hypoxia have been developed for many large bodies of surface water to provide estimates of hypoxic extent. For example, one-dimensional (e.g., [*Rucinski et al.*, 2010]) and box water quality models (e.g., [*Chapra*, 1979; *DiToro and Connolly*, 1980; *Riley*, 1946; *Vollenweider*, 1979]) have been used for estimating the extent of hypoxia in Lake Erie. *Stow and Scavia* [2009] combined a Bayesian model with a traditional water quality model to estimate DO concentration variability in Chesapeake Bay in one dimension, as well as to calculate the hypoxic volume and associated uncertainties. *Hagy and Murrell* [2007] developed a box model to analyze hypoxic extent in Pensacola Bay. Such water quality models can be used to evaluate the overall hypoxic conditions of a system, and can help decision makers to reduce the severity of hypoxia by estimating the reduction of nutrient loads. Although most multi-segment box models do estimate the spatial distribution of DO to some extent, the spatial resolution of these models is typically rather coarse. The mechanistic models are usually not designed specifically to analyze the spatial distribution of water quality parameters accurately.

In recent years, some fine-resolution mechanistic and site-specific models have been developed to estimate the spatial distribution of water quality parameters. For example, the Curvilinear-grid Hydrodynamics model in three dimensions (CH3D) and its Integrated Model

System (CH3D-IMS), developed by *Sheng* [1990], was applied to the Upper Charlotte Harbor [*Kim et al.*, 2010] to provide a measurement of DO changes in the coastal waters. In order to show the spatial extent of hypoxia in Chesapeake Bay, a sophisticated three-dimensional water quality model, CE-QUAL-ICM [*Cerco and Cole*, 1993; *Cerco and Noel*, 2004], was developed and updated over almost two decades through collaboration between the Chesapeake Bay Program (CBP) and the U.S. Army Corps of Engineers. Although CE-QUAL-ICM provides accurate estimates of hydraulic parameters, the estimates of water quality parameters (e.g., chlorophyll, DO) do not precisely match their instantaneous observations of interest [*Murphy et al.*, 2010]. The CE-QUAL-ICM model was also applied to Lake Washington to estimate DO from 1995 to 1997 [*Cerco et al.*, 2006]. Both CH3D and CE-QUAL-ICM models are eutrophication models simulating water-column and sediment processes that affect water quality at a fine spatial resolution. However, the fine-resolution outputs from these water quality models come at the expense of high computational cost and complexity, and the uncertainties associated with their estimates are difficult to quantify.

In summary, most simple mechanistic models are more useful for considering the overall nutrients or DO concentrations of an entire body of water, rather than their detailed spatial distribution. Mechanistic models capable of offering spatial details are typically site-specific and highly complicated, requiring a long time for validation [*Murphy et al.*, 2010].

Simple regression and interpolation

Simple linear and nonlinear regression methods have been widely used to show the spatial distribution of DO concentration. *Burns et al.* [2005] applied a linear interpolation to estimate the maximum yearly extent of hypoxia in Lake Erie for fourteen years between 1983 and 2002, using the lowest DO measurement in each year from each of ten sampling stations. *Bahner*

[2006] applied a three-dimensional inverse distance weighted interpolation to estimate water quality parameters (including DO) in Chesapeake Bay since 1985, when cruises associated with the Chesapeake Bay Program started to collect data throughout the entire year [*Chesapeake Bay Program*, 2011]. *Coopersmith et al.* [2011] used k-nearest neighbor and regression tree algorithms to forecast the probabilities of observing hypoxia and DO levels in Corpus Christi Bay in the short term. *Pokryfki and Randall* [1987] used a simple interpolation scheme to map DO concentration in the Gulf of Mexico. These simple linear and nonlinear interpolation approaches are easily adapted to other ecosystems, and thus have gained wide application elsewhere.

Although simple techniques (e.g., linear and nonlinear interpolation [*Pokryfki and Randall*, 1987], inverse distance weighted interpolation [*Bahner*, 2006]) can provide an estimate of the spatial extent of hypoxia, they are deterministic models and, as such, can only provide limited quantitative estimates of uncertainties and cannot make adequate use of data on any other related variables to improve spatial estimates. Furthermore, although the simple statistical interpolation is easy to apply, it does not explicitly account for the spatial correlation (i.e. smoothness) of the DO distribution. However, in most environmental processes, including the distribution of DO, the residuals are generally spatially and/or temporally correlated.

Geostatistical models

Other methods that consider the spatial correlation of data, such as geostatistical Ordinary Kriging (OK), which greatly depends on the sparse *in situ* measurements, were applied to Charleston Harbor [*Rathbun*, 1998], Chesapeake Bay [*Chehata et al.*, 2007], the Gulf of Mexico [*Rabalais et al.*, 2007a], Neuse River Estuary [*Buzzelli et al.*, 2002] and Ryan Lake [*Stefanovic and Stefan*, 2002]. Although OK accounts for the spatial correlation of DO observations, it

cannot make use of information on any covariates, and therefore leads to high uncertainties if the *in situ* measurements are sparse.

The accuracy of estimating water quality parameters such as chlorophyll and DO are limited by sparse spatial sampling, uncertainty in loading, and difficulty in assigning boundary conditions. Sparse sampling of data in space makes it difficult to determine the spatial variability of DO. For a system such as Lake Erie, for example, DO is measured at only ten locations in its central basin for most years because of the difficulty and expense associated with data collection. OK cannot produce sufficiently detailed spatial information to accurately reflect the spatial and interannual variability of hypoxia [*Rathbun*, 1998]. Thus, there is a need for the application of more methods that can both account for the spatial correlation of DO and incorporate information on related ancillary data sources.

Auxiliary variables are of great help to provide more accurate estimates of the spatial distribution of DO [*Rossi and Posa*, 1990]. The amount of oxygen in the water is naturally affected by the distribution of nutrients, water temperature, water density, stratification situation and chlorophyll concentration. During the summer, the warmer and lighter water of the surface water's uppermost layer blocks the colder and denser water of the bottom layer from the atmosphere due to the availability of sunlight [*National Research Council*, 2000]. Phytoplankton produces oxygen through photosynthesis and also consumes oxygen during respiration and decomposition. As a result, the amount of DO changes with location, temperature, phytoplankton distribution, and time of year.

Whereas the majority of early applications of geostatistical methods were for interpolating sparsely sampled data from a single data set, many methodological innovations have been made in the area of geostatistical approach, which aim to infer DO distribution based on measurements

of a secondary variable. Universal Kriging (UK, a.k.a. kriging with an external drifst) uses auxiliary variables, in addition to the limited *in situ* DO measurements, to obtain a detailed spatial distribution of DO. For example, *Rossi and Posa* [1990] applied UK to interpolate the DO concentration in Mar Piccolo of Taranto, Italy using spatial trends. *Murphy et al.* [2010] used UK, among other approaches, to interpolate water quality parameters such as salinity, water temperature, and DO in Chesapeake Bay, using the output from a water quality model as auxiliary information. *Obenour et al.* [2012] also used a UK-type model to separate the impacts of stratification and nutrient loading on DO in the Gulf of Mexico.

Method	Computational cost & complexity	Consideration of spatial correlation	Estimation uncertainty	Auxiliary variables
Mechanistic model	High	No	Limited	Limited
Simple statistical model (e.g., linear or nonlinear regression)	Low	No	Yes	Yes
Simple interpolation (e.g., linear interpolation, inverse distance weighting)	Low	No	No	No
Ordinary Kriging	Low	Yes	Yes	No
Universal Kriging	Low	Yes	Yes	Yes

Table 2-1 Methods of estimating hypoxic extent.

A comparison of different types of models for estimating distribution of DO and hypoxic extent is shown in Table 2-1. Geostatistical methods, which quantify and account for spatial correlation, are more suitable for providing estimates in space and more reasonable uncertainty bounds [*Zimmerman et al.*, 1999]. All types of models have their own strengths in solving various problems. For example, mechanistic models are most effective at reflecting existing scientific understanding of the processes that control hypoxia development. Within the scope of

this dissertation, geostatisitical Universal Kriging is simple and effective for testing hypothesis regarding factors that control hypoxia, especially as long-term DO data become available.

2.3.2 Brief review of study sites

This dissertation focuses on exploring the variability of hypoxia and quantifying the potential factors causing hypoxia in Lake Erie and Chesapeake Bay. The section below briefly describes the hypoxic conditions in these two systems.

Lake Erie

Summertime hypoxia is a natural phenomenon in the central basin of Lake Erie, probably dating back thousands of years [*Delorme*, 1982]. Notably, small areas of hypoxic water were first recorded in the central basin of Lake Erie in the 1930s [*Fish*, 1960; *Herdendorf*, 1984]. However, evidence suggests that oxygen depletion rates increased in the 1950s due to anthropogenic factors [*Committee on Environment and Natural Resources*, 2010]. Although there is growing awareness of large scale, long term changes in watersheds and waters, the causes and consequences of hypoxia in Lake Erie are not yet fully understood.

As a measure of hypoxia, hypoxic areal extent is the one of the best indicators for tracking historical changes, as well as the response of hypoxia to water quality management and other ecosystem variables [*Hagy et al.*, 2004]. A comprehensive record of seasonal or interannual hypoxic extent would therefore increase knowledge of the history, variability, dynamic causes and consequences of hypoxia, as well as pertinent characteristics (e.g., the most vulnerable locations) of the lake system. Such understanding is, in turn, needed for developing efficient water quality management and conservation strategies.

Chesapeake Bay

Chesapeake Bay is the largest and most productive estuary on the East Coast of the United States, but is plagued with bottom-water hypoxia and/or anoxia due to agricultural and industrial development and population growth along its shores and headwaters [*Cerco and Cole*, 1993]. Hypoxia was first reported in Chesapeake Bay in the 1930s [*Newcombe and Horne*, 1938; *Officer et al.*, 1984], and became more common and widespread starting in the late 1950s and early 1960s, due to anthropogenic nutrient influx [*Cronin and Vann*, 2003].

Chesapeake Bay is an extremely narrow and shallow estuary, which presents a unique set of challenges for the implementation of spatial interpolation methods. As a partially stratified estuary in which fresh water and seawater are partially mixed, Chesapeake Bay has extremely complex physical and biochemical dynamics. Stratification caused by salinity differences varies geographically [*U. S. EPA*, 2002b]. Therefore, it is necessary to take auxiliary variables reflecting the stratification situation into consideration during the process of estimating DO spatial distribution.

CHAPTER 3 Methodology

This chapter describes existing geostatistical methods implemented as part of the work presented in Chapters 4 to 6. The details of the gesotatistical approach for each research component in this dissertation are presented within the discussion of each component.

3.1 Background of geostatistical methods

In order to solve the problems presented by degrading water quality, methods that predict the response of ecological communities to both anthropogenic and natural alterations are needed [*Clark et al.*, 2001]. Moreover, because of the requirement for risk benefit analysis and water management, the impact of the uncertainties associated with attribute estimation need to be taken into account throughout the analysis. Statistical methods can be formulated in a stochastic framework, yielding optimal estimates and accurate measures of uncertainty. In stochastic approaches, unknown parameters are described through statistical distributions, and meaningful uncertainty bounds can often be identified [*Burrough and McDonnell*, 1998].

In most cases, environmental data taken from locations closer to one another are likely to be more similar than the those taken from locations further apart [*Chiles and Delfiner*, 1999]. Therefore, spatial correlation is an important feature of these data. However, traditional statistical models assuming data to be independently and identically distributed (I.I.D.) usually ignore this fact, which could result in biased estimates [*Chiles and Delfiner*, 1999]. Compared to these traditional statistical models, geostatistical methods can provide more accurate analyses of the spatial distribution of environmental data by capturing their underlying spatial correlation [*Chiles and Delfiner*, 1999].

Geostatistical methods include models for characterizing the degree and the type of spatial correlation, as well as models for making use of this correlation for estimation. The original

theory of geostatistics [*Matheron*, 1963] was developed in the late 1950s and early 1960s to provide a set of inference methods for estimating ore reserves in three dimensions. Geostatistical methods, which explicitly account for spatial correlation, are normally selected as interpolation approaches [*Butcher*, 1996; *Little et al.*, 1997]. They have been extensively used and generalized for applications in many areas beyond mining, such as ground water (e.g., [*Delhomme*, 1979; *Kitanidis and Vomvoris*, 1983; *Mouser et al.*, 2005]), soil science (e.g., [*Bourennane et al.*, 2007; *Hengl et al.*, 2007; *Trangmar et al.*, 1985; *Wu et al.*, 2006]), hydrology (e.g., [*Ahmed and Demarsily*, 1987; *El Idrysy and De Smedt*, 2007; *Woodbury and Sudicky*, 1991]), ecology (e.g., [*Robertson*, 1987]), meteorology (e.g., [*Kyriakidis et al.*, 2001]), oceanography (e.g., [*Gohin and Langlois*, 1993]), atmospheric science (e.g., [*Gourdji et al.*, 2008; *Michalak et al.*, 2004; *Mueller et al.*, 2008]), and surface water (e.g., [*Murphy et al.*, 2010; *Obenour et al.*, 2012]).

In addition to spatial correlation, a trend can also be used to estimate statistical relationships between the observation variable (i.e., dependent variable) and those variables in the trend (i.e., auxiliary variables). Geostatistical data fusion methods can account for both the trend and the underlying spatial or temporal correlation in the measurements. These methods have also been used in other disciplines including oceanographic science (e.g., [*Georgakarakos and Kitsiou*, 2008]), water table mapping [*Hoeksema et al.*, 1989], and meteorological science to improve upon the predictability of short-term weather forecasting models (e.g., [*Carroll and Cressie*, 1996; *Pardo-Iguzquiza et al.*, 2005; *Phillips et al.*, 1992]).

Given that the sampled attributes (e.g., DO, total organic carbon) in the surface water or water sediment often exhibit spatial continuity [*Barabas et al.*, 2001], geostatistical data fusion approaches, which are widely used for spatially or temporally correlated data, have been selected as the main methods in this dissertation. The specific applications used in this dissertation are

geostatistical kriging and geostatistical inverse modeling. For kriging, the target variable to be estimated is sampled directly, while for inverse modeling, the variable is sampled indirectly.

3.2 Basic setup in geostatistical methods

Within a geostatistical framework, an $n \times 1$ vector of random variables, $\mathbf{z}(\mathbf{x})$, at locations $\mathbf{x}$ ($n \times 1$) can be divided into a deterministic part (a.k.a. trend) and a stochastic part (a.k.a. residual) which is dependent upon location in space:

$$\mathbf{z}(\mathbf{x}) = \mathbf{D}(\mathbf{x}) + \mathbf{z}_{res}(\mathbf{x}) \tag{3.1}$$

The deterministic part will be specified later in each research component, because it differs among the three research components. In this dissertation, the residuals are assumed to be second-order stationary, i.e. they have a constant mean over the domain and the covariance of any pair of residuals is only dependent upon the separation distance between their locations.

A variogram, $\boldsymbol{\gamma}(\mathbf{h})$, describes the variance of residuals, $\mathbf{z}_{res}(\mathbf{x})$, as a function of separation distance, $\mathbf{h}$, between any two measurements:

$$\boldsymbol{\gamma}(\mathbf{h}) = \frac{1}{2}\mathrm{E}\left[\left(\mathbf{z}_{res}(\mathbf{x}+\mathbf{h}) - \mathbf{z}_{res}(\mathbf{x})\right)\right] \tag{3.2}$$

where E[ ] denotes the expected value operator. If the residuals are second-order stationary (Figure 3-1), a covariance function, $\mathbf{Q}(\mathbf{h})$, could be linked to the variogram by the following equation [*Chiles and Delfiner*, 1999]:

$$\boldsymbol{\gamma}(\mathbf{h}) = \mathbf{Q}(\mathbf{0}) - \mathbf{Q}(\mathbf{h}) \tag{3.3}$$

A covariance function shows the correlation between a pair of data with a separation distance, while a variogram shows the dissimilarity between a pair of data. Both variogram and

covariance function are quantitative tools to characterize the spatial or temporal structure of the residuals.

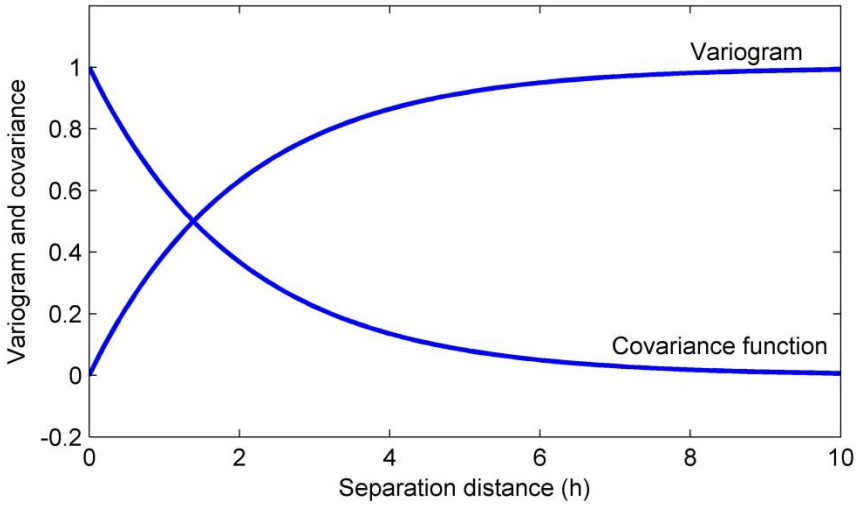

Figure 3-1 Variogram and covariance function for an exponential model.

There are many types of variogram models describing the expected behavior of the residuals, such as nugget effect, linear, exponential and Gaussian, to describe the spatial behavior of residuals [*Cressie*, 1993; *Kitanidis*, 1997]. For example, an exponential variogram together with a nugget effect is:

$$\boldsymbol{\gamma}\left(\mathbf{h}|\sigma^2, l, \sigma_Q^2\right) = \sigma^2\left(1 - \exp\left(-\frac{\mathbf{h}}{l}\right)\right) + \sigma_Q^2 \cdot (1 - \delta[\mathbf{h}]) \qquad 3.4$$

where σ^2 is the part of variance that is spatially correlated, σ_Q^2 is the micro-variability which is independent (i.e., uncorrelated) in space, and $3l$ is the practical range. δ[ ] is a Kronecker delta function, equal to one when the separation distance is zero and equal to zero otherwise. The corresponding covariance function is:

$$\mathbf{Q}\left(\mathbf{h}|\sigma^2, l, \sigma_Q^2\right) = \sigma^2 \exp\left(-\frac{\mathbf{h}}{l}\right) + \sigma_Q^2 \cdot \delta[\mathbf{h}] \quad 3.5$$

3.3 Ordinary Kriging and Universal Kriging

Ordinary Kriging (OK) has been extensively used for interpolation when only the target parameter to be estimated is available or the parameter does not suggest trend [*Chiles and Delfiner*, 1999; *Cressie*, 1993; *Isaaks and Srivastava*, 1989]. Universal Kriging (UK) is a geostatistical method that can include some other covariates (i.e. auxiliary variables) that are related to the estimates [*Cressie*, 1993]. Both OK and UK require the data to have the same resolution at each measurement and estimation location within the entire domain [*Chiles and Delfiner*, 1999]. The procedural steps for the OK and UK models followed in this research are described in *Kitanidis* [1997]. Only the main steps were summarized in this chapter.

As mentioned in Section 3.1, the measurement $\mathbf{z}$ can be divided into a deterministic part and a stochastic part. For both OK and UK, the deterministic part represents the large-scale mean structure:

$$\mathbf{z} = \mathbf{X_z}\boldsymbol{\beta}_\mathbf{z} + \mathbf{z}_{res} \quad 3.6$$

where $\boldsymbol{\beta}_\mathbf{z}$ is a $p \times 1$ vector of unknown drift coefficients that scale the components in $\mathbf{X_z}$ ($n \times p$). $\mathbf{X_z}\boldsymbol{\beta}_\mathbf{z}$ is the deterministic part described in equation 3.1. For UK, $\mathbf{X_z}$ defines a trend of $\mathbf{z}$ usually including a constant (i.e., intercept) and auxiliary variables. For OK, $\mathbf{X_z}$ only includes the constant and does not include the auxiliary variables. The expected value of measurement $\mathbf{z}$ could be represented as:

$$\mathrm{E}[\mathbf{z}] = \mathbf{X_z}\boldsymbol{\beta}_\mathbf{z} \quad 3.7$$

3.4 Geostatistical inverse modeling

In this dissertation, geostatistical inverse modeling (GIM) is based on the assumption of a linear relationship between the estimated field (**s**) and the available measurement (**z**):

$$\mathbf{z} = \mathbf{Hs} + \mathbf{z}_{res} \tag{3.8}$$

where **z** is an $n \times 1$ vector of observations (in Chapter 4, average concentrations) and **s** is an $m \times 1$ vector of unknowns (in Chapter 4, concentrations at a fine and uniform resolution). The known $n \times m$ Jacobian matrix **H** (i.e., $H_{ij} = \partial z_i / \partial s_j$) describes the relationship between measurement **z** and estimated concentrations **s**. $\mathbf{z}_{res}$ is the model data mismatch error modeled as an $n \times 1$ vector of independent normally distributed random variables with zero mean and variance σ_R^2. The covariance matrix **R** ($n \times n$) of $\mathbf{z}_{res}$ is shown as:

$$\mathbf{R} = \sigma_R^2 \mathbf{I}_n \tag{3.9}$$

where $\mathbf{I}_n$ is an $n \times n$ identity matrix.

The estimated field **s** is modeled as a random vector with its expected value:

$$\mathrm{E}[\mathbf{s}] = \mathbf{X_s}\boldsymbol{\beta_s} \tag{3.10}$$

where $\boldsymbol{\beta_s}$ is a $p \times 1$ vector of unknown drift coefficients that scale the components in auxiliary variables $\mathbf{X_s}$, and $\mathbf{X_s}$ therefore becomes an $m \times p$ matrix. $\mathbf{X_s}$ defines known zonation [*Fienen et al.*, 2004] or spatial trend of **s** [*Michalak et al.*, 2004; *Snodgrass and Kitanidis*, 1997].

3.5 Solutions

After obtaining the variogram model (to be discussed in each research component), all the estimation results **s**, including the total organic carbon in Chapter 4 and DO concentration in Chapters 5 and 6, in this work can be solved from the equation below:

$$\begin{bmatrix} \boldsymbol{\Sigma}_{\mathbf{zz}} & \mathbf{F}_{\mathbf{zz}} \\ \mathbf{F}_{\mathbf{zz}}{}^{T} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \boldsymbol{\Lambda}^{T} \\ \mathbf{M} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\Sigma}_{\mathbf{zs}} \\ \mathbf{F}_{\mathbf{ss}}^{T} \end{bmatrix} \tag{3.11}$$

Table 3-1 explains the terms in equation 3.11 (i.e., $\boldsymbol{\Sigma}_{\mathbf{zz}}$, $\boldsymbol{\Sigma}_{\mathbf{zs}}$, $\mathbf{F}_{\mathbf{zz}}$, $\mathbf{F}_{\mathbf{ss}}$ and $\boldsymbol{\Sigma}_{\mathbf{ss}}$) as well as their dimensions for OK, UK, and GIM, respectively:

Method	$\boldsymbol{\Sigma}_{\mathbf{zz}}$ $(n \times n)$	$\boldsymbol{\Sigma}_{\mathbf{zs}}$ $(n \times m)$	$\mathbf{F}_{\mathbf{zz}}$ $(n \times p)$	$\mathbf{F}_{\mathbf{ss}}$ $(m \times p)$	$\boldsymbol{\Sigma}_{\mathbf{ss}}$ $(m \times m)$
OK	$\mathbf{Q}_{\mathbf{zz}}$	$\mathbf{Q}_{\mathbf{zs}}$	$\mathbf{X}_{\mathbf{z}}$	$\mathbf{X}_{\mathbf{s}}$	$\mathbf{Q}_{\mathbf{ss}}$
UK	$\mathbf{Q}_{\mathbf{zz}}$	$\mathbf{Q}_{\mathbf{zs}}$	$\mathbf{X}_{\mathbf{z}}$	$\mathbf{X}_{\mathbf{s}}$	$\mathbf{Q}_{\mathbf{ss}}$
GIM	$\mathbf{HQ}_{\mathbf{ss}}\mathbf{H}^{T}+\mathbf{R}$	$\mathbf{HQ}_{\mathbf{ss}}$	$\mathbf{HX}_{\mathbf{s}}$	$\mathbf{X}_{\mathbf{s}}$	$\mathbf{Q}_{\mathbf{ss}}$

Table 3-1 Parameters in equation 3.11 for OK, UK, and GIM.

where T represents matrix transposition. $\mathbf{Q}_{\mathbf{zz}}$ is an $n \times n$ covariance matrix of measurements, $\mathbf{Q}_{\mathbf{zs}}$ is an $n \times m$ covariance matrix between measurements and estimates, and $\mathbf{Q}_{\mathbf{ss}}$ is an $m \times m$ covariance matrix between estimates. $\mathbf{X}_{\mathbf{z}}$ and $\mathbf{X}_{\mathbf{s}}$ are auxiliary variables at measurement and estimate locations, respectively. $\boldsymbol{\Lambda}$ is an $m \times n$ matrix of weights assigned to measurements for estimates, $\mathbf{M}$ is a $p \times m$ vector of Lagrange multipliers. $\boldsymbol{\Lambda}$ and $\mathbf{M}$ are parameters to be solved in the three approaches.

The geostatistical approaches determine the estimation precision along with the best estimates. The estimated concentrations and their associated covariance matrix in OK, UK and GIM are determined by:

$$\hat{\mathbf{s}} = \boldsymbol{\Lambda}\mathbf{z} \tag{3.12}$$

$$\mathbf{V}_{\hat{\mathbf{s}}} = -\mathbf{F}_{\mathbf{ss}}\mathbf{M} + \boldsymbol{\Sigma}_{\mathbf{ss}} - \boldsymbol{\Sigma}_{\mathbf{zs}}^{T}\boldsymbol{\Lambda}^{T} \tag{3.13}$$

where $\hat{\mathbf{s}}$ is the best estimate at fine resolution, and the square roots of the diagonal elements of $\mathbf{V}_{\hat{\mathbf{s}}}$ are standard deviations (i.e., estimation error).

The best estimate of drift coefficients, $\boldsymbol{\beta_z}$ of kriging and $\boldsymbol{\beta_s}$ of inverse modeling, are [*Chiles and Delfiner*, 1999]:

$$\widehat{\boldsymbol{\beta}} = (\mathbf{F}_{\mathbf{zz}}^{T}\boldsymbol{\Sigma}_{\mathbf{zz}}^{-1}\mathbf{F}_{\mathbf{zz}})^{-1}\mathbf{F}_{\mathbf{zz}}^{T}\boldsymbol{\Sigma}_{\mathbf{zz}}^{-1}\mathbf{z} \tag{3.14}$$

and their associated variance and covariance are:

$$\mathbf{V}_{\widehat{\boldsymbol{\beta}}} = (\mathbf{F}_{\mathbf{zz}}^{T}\boldsymbol{\Sigma}_{\mathbf{zz}}^{-1}\mathbf{F}_{\mathbf{zz}})^{-1} \tag{3.15}$$

CHAPTER 4 Geostatistical Downscaling for Non-uniform Resolution Data

Information about attributes such as contaminant concentrations or hydraulic properties in benthic sediments is typically obtained in core sections of varying lengths, and only the average value is measured in each section. However, an estimate of the attribute distribution at a uniform spatial resolution is often required for site characterization and the design of appropriate risk-based remediation alternatives. Because attributes exhibit spatial autocorrelation, geostatistical methods have become an essential tool for estimating their spatial distribution. The purpose of this chapter is to optimally infer the spatial distribution of sampled attributes at a uniform resolution from fluvial core sampling data, using a downscaling technique formulated as a geostatistical inverse problem.

4.1 Introduction

Remediation of contaminated sediments can be very expensive, and often requires detailed knowledge about the distribution of contamination in the subsurface in order to formulate effective remediation strategies [*Barabas et al.*, 2001; *Wang et al.*, 2004]. Measuring attributes important for remediation (e.g., contaminant concentrations, hydraulic properties, and microbial activities) everywhere in a system is not practical or financially feasible, however. The knowledge of subsurface properties of water sediment is often limited to a set of incomplete information obtained from sparse sampling. Due to the irregular thickness of sediment above the bedrock surface, samples of contaminated stream or riverbed sediments are often obtained from cores that vary in length. Each core is subsequently sub-divided into several sections of varying thickness, which are analyzed for key attributes. The reported value at each location typically represents an average within individual segments of each core, and the sampling resolution (a.k.a. support) is therefore not uniform.

In order to map the attribute distribution everywhere in the subsurface, interpolation methods, such as linear interpolation, geostatistical methods, and inverse distance weighting, are often used. Because the sampled attributes often exhibit spatial continuity [*Barabas et al.*, 2001], geostatistical kriging methods, which explicitly account for spatial correlation, are commonly employed for interpolating parameters and estimating their spatial distribution [*Butcher*, 1996; *Little et al.*, 1997]. In addition to an estimation of the distribution of attribute concentrations, geostatistical methods also provide an indication of the errors and uncertainties associated with the interpolated values [*Burrough and McDonnell*, 1998].

The idea of accounting for the connections between fine and coarse resolution data but ignoring the resolution issue is comparable to the area-to-point kriging method [*Kyriakidis*, 2004]. In that work, however, the estimation of covariance parameters of the fine resolution data is not included, which prevents the application to the field data.

The approach in this part of work for estimating the statistical model describing variability at the uniform resolution is further developed. Unlike point-to-point Ordinary Kriging, the presented approach does not assume a uniform resolution for the sampled data. As a result, geostatistical downscaling as presented here is designed to preserve the true, underlying, variability of the sampled attributes during the downscaling process. The primary objective of this work is to investigate the potential advantages of implementing a geostatistical downscaling method for inferring attribute values at a consistent resolution, using data from measured varying-resolution core sections using geostatistical downscaling. Taking sediment total organic carbon concentration observations as an example, geostatistical downscaling was compared to the more traditional approach of Ordinary Kriging.

This chapter is organized as follows: Section 4.2 presents the data used in this study. Section 4.3 presents a brief overview of the inputs to the Ordinary Kriging and geostatistical downscaling models. Section 4.4 presents the results of the analysis, including the comparison of the results between the two methods for both pseudodata and field data studies. Section 4.5 summarizes the main conclusions of the study.

4.2 Data description

This section describes the two applications used to evaluate the proposed approaches. The first application is based on pseudodata generated to be representative of variability typically observed in contaminated river sediments. The second application uses total organic carbon (TOC) data from the Passaic River in New Jersey.

The data for the first application were generated at a fine, uniform resolution, using a pre-specified spatial covariance matrix (Figure 4-2a), and were then averaged to represent core sections of differing lengths (Figure 4-2b).). For simplicity, the data were generated using an isotropic model (i.e., same correlation length along different directions), such that only one correlation range parameter l is required. These average concentrations were assumed known, and were used as the basis for recovering the original fine-resolution spatial distribution of the attribute and its associated spatial covariance parameters.

The field data used in the second application represent the TOC weight percentage of sediment in a ten-kilometer stretch of the lower Passaic River in New Jersey. The river is part of the Hudson Raritan tidal estuary, the sediments of which are cohesive and fine grained [*Barabas et al.*, 2004]. The data were acquired from the US EPA [*U. S. EPA*, 2002c] and a database of New Jersey sediment data compiled by NOAA [*NOAA*, 2002]. A two-dimensional section along flow and vertical directions of the sediment bed is analyzed, which includes 27 cores and a total

of 153 core sections as shown in Figure 4-1, as well as over 5500 estimation locations (Figure 3b). The TOC measurements represent the average weight percent within the core sections, all of which are larger or equal to 30 cm in length. TOC measurement represents the average concentrations of the core sections that are 30 cm and greater in length. Due to the meandering nature of the river in an s-shape, physical distances between observations cannot be identified with a mere Euclidean distance as some of them would necessarily be measured over land. All the separation distances between two locations are of the shortest in-water path to avoid including incorrect paths over intervening land [*Barabas et al.*, 2001; 2004]. In addition, to ensure that all distances are measured within the water, sample coordinates were transformed prior to analysis with Gridgen, a grid generating software [*Barabas et al.*, 2001]. Geostatistical interpolation methods are applied afterward.

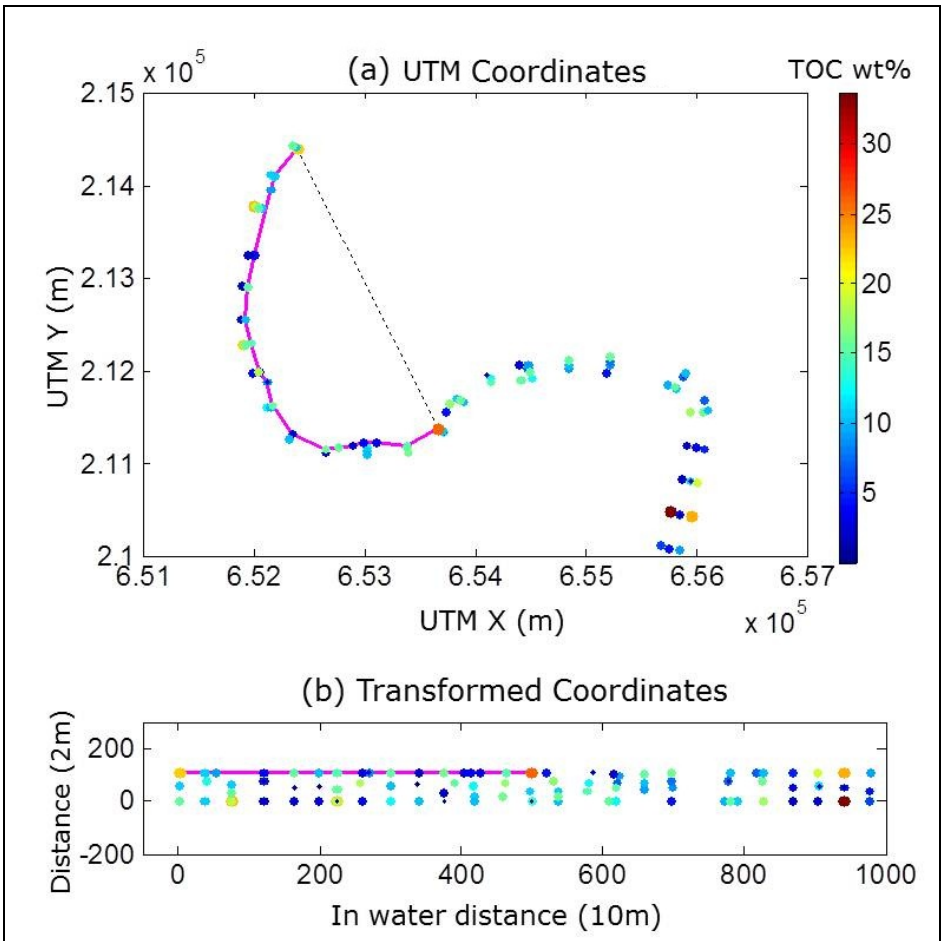

Figure 4-1 Location of field data in the Passaic River, New Jersey, UTM coordinates (a) and in-water distance (b).

4.3 Methodology

This section describes the proposed geostatistical downscaling framework, and briefly reviews the Ordinary Kriging approach that is used in Section 4.4 as a comparison to the obtained results.

4.3.1 Geostatistical downscaling

The geostatistical downscaling (GD) approach is formulated as a geostatistical inverse modeling (GIM) problem. For the interpolation of TOC distribution in water sediment case, the forward model refers to the process of estimating average concentrations at the sampled

resolution from concentrations at a fine and uniform resolution, while the inverse model involves the estimation of concentrations at the fine and uniform distribution given sampled values representative of variable depth increments. Detailed descriptions of linear geostatistical inverse modeling are available in *Snodgrass and Kitanidis* [1997] and *Michalak et al.* [2004], and only the key equations are reproduced here.

Detailed GD method steps are described in the methodology section 3.2. The known $n \times m$ Jacobian matrix $\mathbf{H}$ (i.e., $H_{ij} = \partial z_i / \partial s_j$) describes the averaging relationship between measurements $\mathbf{z}$ and estimated attribute values $\mathbf{s}$. In the case examined here, $\mathbf{H}$ is formulated as:

$$H_{ij} = \begin{cases} \frac{1}{N_i}, & d_{Ui} < d_j < d_{Li} \\ 0, & otherwise \end{cases} \quad 4.1$$

where N_i is the number of fine-resolution estimation locations contained in the i^{th} measured core section, d_j is the depth of estimation location j, and d_{Ui} and d_{Li} are the upper and lower boundaries of the core section corresponding to measurement i, respectively.

The estimated field $\mathbf{s}$ is modeled as a random vector with its expected value:

$$\mathrm{E}[\mathbf{s}] = \mathbf{X_s}\beta_\mathbf{s} \quad 4.2$$

In this application, the distribution of attributes in sediment are assumed to have a constant but unknown mean $\beta_\mathbf{s}$ over the whole domain, and $\mathbf{X_s}$ therefore becomes an $m \times 1$ vector of ones.

The corresponding variogram describing the spatial autocorrelation of the estimated attribute is assumed to follow an exponential model, based on an examination of the data from the Passaic River, New Jersey.

$$\boldsymbol{\gamma}_{\mathrm{ss}}\left(\mathbf{h}|\sigma^2, l, \sigma_Q^2\right) = \sigma^2\left(1 - \exp\left(-\frac{\mathbf{h}}{l}\right)\right) + \sigma_Q^2 \cdot (1 - \delta[\mathbf{h}]) \qquad 4.3$$

where **h** is the separation distance between two estimate locations. The parameters used in Equation 4.3 are optimized using REML given the available variable-resolution data, as described in Section **Error! Reference source not found.**.1. Once all relevant vectors and matrices have been defined, the GIM estimation problem is solved as described in Section **3.1**.

4.3.2 Ordinary Kriging

Ordinary Kriging (OK) has been extensively used for parameter interpolation when only limited samples are available. It requires the data to have the same or at least compatible resolution at each measurement and estimation location within the entire domain [*Chiles and Delfiner*, 1999; *Cressie*, 1993; *Isaaks and Srivastava*, 1989]. Therefore, if OK is chosen as the interpolations method, the observations or measurements must be assumed as a uniform resolution data set. To achieve this goal, all the varying depth core sections are divided into uniform core subsections and the measured value in the core section is assumed to represent the value at the center of the core section (e.g. Figure 4-2b). Hence, the spatial covariance structure of the fine resolution estimates and variable resolution measurements are implicitly assumed to be the same. In the presented applications, the portion of the observed variability that is not spatially correlated is assumed to be due to measurement errors, and OK is therefore implemented in the continuous part kriging form [*Chiles and Delfiner*, 1999; *Kitanidis*, 1997]. Methodology section 3.3 outlines the main steps of OK.

4.3.3 Covariance parameter optimization

There are many approaches of estimating the covariance parameters required for the application of geostatistical methods, including least squares (LS) fits to observed variability

[*Bogaert and Russo*, 1999], maximum likelihood methods [*Kitanidis and Lane*, 1985; *Michalak et al.*, 2005; *Pannone and Kitanidis*, 2001], Restricted Maximum Likelihood (REML) [*Kitanidis*, 1997], composite likelihood [*Schabenberger and Gotway*, 2005], and generalized estimation equations [*Schabenberger and Gotway*, 2005]. REML has the advantage of eliminating the unknown drift coefficient or mean $\boldsymbol{\beta}_s$ that reduces the bias in the estimated variogram parameters [*Kitanidis and Lane*, 1985]. Major steps are summarized in this section.

For kriging

In OK, the probability distribution function (*pdf*) of measurements $\mathbf{z}$ (i.e., the likelihood part) was shown in equation 4.4 [*Kitanidis*, 1995]:

$$p(\mathbf{z}, \boldsymbol{\beta}_\mathbf{z}) \propto exp\left(-\frac{1}{2}(\mathbf{z}-\mathbf{X}_\mathbf{z}\boldsymbol{\beta}_\mathbf{z})^T \mathbf{Q_{zz}}^{-1}(\mathbf{z}-\mathbf{X}_\mathbf{z}\boldsymbol{\beta}_\mathbf{z})\right) \quad 4.4$$

Maximizing the probability in the equation is equivalent to minimizing its negative logarithm:

$$L(\mathbf{z}, \boldsymbol{\beta}_\mathbf{z}) \propto -\frac{1}{2}(\mathbf{z}-\mathbf{X}_\mathbf{z}\boldsymbol{\beta}_\mathbf{z})^T \mathbf{Q_{zz}}^{-1}(\mathbf{z}-\mathbf{X}_\mathbf{z}\boldsymbol{\beta}_\mathbf{z}) \quad 4.5$$

For OK, the structural parameters in $\mathbf{Q_{zz}}$ are obtained by minimizing the cost function:

$$L_{\boldsymbol{\theta}\mathrm{OK}} = \frac{1}{2} ln|\mathbf{Q_{zz}}| + \frac{1}{2} ln\left|\mathbf{X_z}^T \mathbf{Q_{zz}^{-1}} \mathbf{X_z}\right| + \frac{1}{2}\mathbf{z}^T \boldsymbol{\Xi}_{\mathrm{OK}} \mathbf{z} \quad (1)$$

$$\boldsymbol{\Xi}_{\mathrm{OK}} = \mathbf{Q_{zz}^{-1}} - \mathbf{Q_{zz}^{-1}} \mathbf{X_z} \left(\mathbf{X_z}^T \mathbf{Q_{zz}^{-1}} \mathbf{X_z}\right)^{-1} \mathbf{X_z}^T \mathbf{Q_{zz}^{-1}} \quad (2)$$

where $\mathbf{X_z}$ is an $n \times 1$ vector of ones.

For geostatistical inverse modeling

In GIM, the *pdf* of estimate (**s**) based on measurements (**z**) can be expressed as [*Michalak and Kitanidis*, 2004; *Snodgrass and Kitanidis*, 1997]:

$$p(\mathbf{s},\boldsymbol{\beta}|\mathbf{z}) \propto exp\left(-\frac{1}{2}(\mathbf{z}-\mathbf{Hs})^T\mathbf{R}^{-1}(\mathbf{z}-\mathbf{Hs}) - \frac{1}{2}(\mathbf{s}-\mathbf{X_{ss}}\boldsymbol{\beta_s})^T\mathbf{Q_{ss}}^{-1}(\mathbf{s}-\mathbf{X_{ss}}\boldsymbol{\beta_s})\right) \quad 4.6$$

The first part of equation 4.6 is the *pdf* of likelihood, and the second half is the *pdf* of *a priori* information between measurements and estimates. Maximizing the probability in the equation is equivalent to minimizing its negative logarithm:

$$L(\mathbf{s},\boldsymbol{\beta}|\mathbf{z}) \propto \frac{1}{2}(\mathbf{z}-\mathbf{Hs})^T\mathbf{R}^{-1}(\mathbf{z}-\mathbf{Hs}) + \frac{1}{2}(\mathbf{s}-\mathbf{X_s}\boldsymbol{\beta_s})^T\mathbf{Q_{ss}}^{-1}(\mathbf{s}-\mathbf{X_s}\boldsymbol{\beta_s}) \quad 4.7$$

After integrating over all values of the $\boldsymbol{\beta}_\mathbf{z}$ and $\boldsymbol{\beta}_\mathbf{s}$ [*Fienen et al.*, 2004; *Kitanidis*, 1995], the cost function becomes:

$$L_{\boldsymbol{\theta}\mathrm{GD}} = \frac{1}{2}ln|\boldsymbol{\Omega}| + \frac{1}{2}ln\left|\mathbf{X_s}^T\mathbf{H}^T\boldsymbol{\Omega}^{-1}\mathbf{H}\mathbf{X_s}\right| + \frac{1}{2}\mathbf{z}^T\boldsymbol{\Xi}_{\mathrm{GD}}\mathbf{z} \quad 4.8$$

$$\boldsymbol{\Omega} = \mathbf{H}\mathbf{Q}_{\mathrm{ss}}\mathbf{H}^T + \mathbf{R} \quad 4.9$$

$$\boldsymbol{\Xi}_{\mathrm{GD}} = \boldsymbol{\Omega}^{-1} - \boldsymbol{\Omega}^{-1}\mathbf{H}\mathbf{X_s}\left(\mathbf{X_s}^T\mathbf{H}^T\boldsymbol{\Omega}^{-1}\mathbf{H}\mathbf{X_s}\right)^{-1}\mathbf{X_s}^T\mathbf{H}^T\boldsymbol{\Omega}^{-1} \quad 4.10$$

where | | denotes matrix determinant.

4.4 Results and discussion

This section describes the two applications used to test the GD approach presented, followed by a discussion of key results. The first application is based on pseudodata generated to be representative of variability typically observed in contaminated river sediments. The second application uses total organic carbon data from the lower Passaic River in New Jersey (Figure 4-1).

To validate our initial hypothesis that GD is a more appropriate method for reproducing the true variability of the subsurface than OK, a two-dimensional pseudodata case is presented here. The pseudodata involving the estimation of an attribute distribution with a constant mean in water sediment are generated according to the structural characteristics (i.e., covariance matrix) of the field data. However, anisotropy (i.e., different correlation lengths along different directions) is not included in the pseudodata case for simplification. Pseudodata are generated everywhere at uniform, fine resolution in a domain and followed by averaging up to the concentrations of different-length core sections in the vertical direction. The depth-averaged concentration samples are used, which is typically only available in field, as known measurement to infer the fine resolution concentrations as well as estimation error and compare with the original pseudodata.

4.4.1 Pseudodata application

Two-dimensional pseudodata application, presented in Figure 4-2, is designed to suggest the advantage of GD over OK at recovering estimation in fine resolution. The fine resolution pseudodata and coarse resolution sample experimental variograms, as well as corresponding variograms at fine resolution scale generated from GD and OK based on coarse resolution sample are all illustrated in Figure 4-2g. The blue lines in Figure 4-2g present the experimental variogram obtained using the samples from Figure 4-2b, together with the theoretical variogram fit based on the parameters estimated using the OK version of the REML algorithm. The red lines in Figure 4-2g represent the experimental variogram at the estimation resolution (Figure 4-2a), and the theoretical variogram fit based on the parameters estimated using the GD version of the REML algorithm. Ordinary Kriging variogram matches the average concentration variogram very well, while the variogram generated from GD matches the fine-resolution

concentration variogram. As interpolation results (Figure 4-2c and Figure 4-2d) are derived based on the variogram parameters estimated, the estimated concentration distribution strongly depends on the variogram obtained. As long as the variogram parameters are identified, estimation results can be interpolated accurately to the true distribution and its associated errors. Because of the difference between the resolutions of two data sets, the sills of variograms between the fine resolution data set and the average concentration are quite different. During the arithmetical averaging process that changes the resolution of the data set, extreme values are smoothed out which causes the inadequate description of variability by the variogram at the original resolution. When two resolutions are getting close, the sills of variograms are closer. Instead, when two resolutions are far away from each other, the sills of variograms are far away from each other as well. Additionally, the correlation length predicted from GD is shorter than that from OK. Because the variability of the data is smoothed out during the average process, spatial continuity becomes longer.

Figure 4-2c and Figure 4-2d present the best estimates of the attribute distribution at the estimation resolution for GD and OK, respectively. Visually, it is clear that the best estimates from GD are closer to the fine resolution pseudodata than those from OK. It is also clear that the OK best estimates show less variability when compared to GD. This can also be seen in Figure 4-3a, which shows a one-dimensional vertical slice from the two-dimensional results presented in Figure 4-2. Figure 4-3a again confirms that the fine resolution pseudodata display more variability than the measured core sections. Although neither of the applied interpolation methods can reproduce the true fine resolution attribute distribution precisely due to the lack of fine scale observations, the GD results are much closer to the true distribution of fine-scale pseudodata, with the OK estimates showing insufficient variability.

Figure 4-2e and Figure 4-2f present the estimation uncertainties associated with the GD and OK estimated attribute distribution. A good uncertainty estimate is one that correctly quantifies the errors associated with the estimates provided by a particular method. Figure 4-3a also presents the upper and lower 95% confidence intervals for one vertical slice from the 2D results. As can be seen from Figure 4-2e, Figure 4-2f, and Figure 4-3a, the uncertainty estimates from the GD approach are higher relative to those from OK. Looking at the results presented in Figure 4-3a, however, it is clear that OK underestimates the uncertainty associated with its estimates. This can be seen from the fact that the true attribute distribution at the estimation scale (which is known in this case because this is a pseudodata application) lies outside of the uncertainty bounds for a large portion of the estimation locations. The two-standard-deviation uncertainty bounds estimated using the GD approach, on the other hand, correctly encompass the true attribute value at the estimation resolution approximately 95% of the time. Although the uncertainties from OK are narrower, they are not adequately representative for the true uncertainties of the fine-resolution data. Since OK assumes that the measured average attribute values are representative of variability at the finer estimation resolution, it underestimates variability at fine resolutions leading to the observed underestimates of the estimation uncertainty.

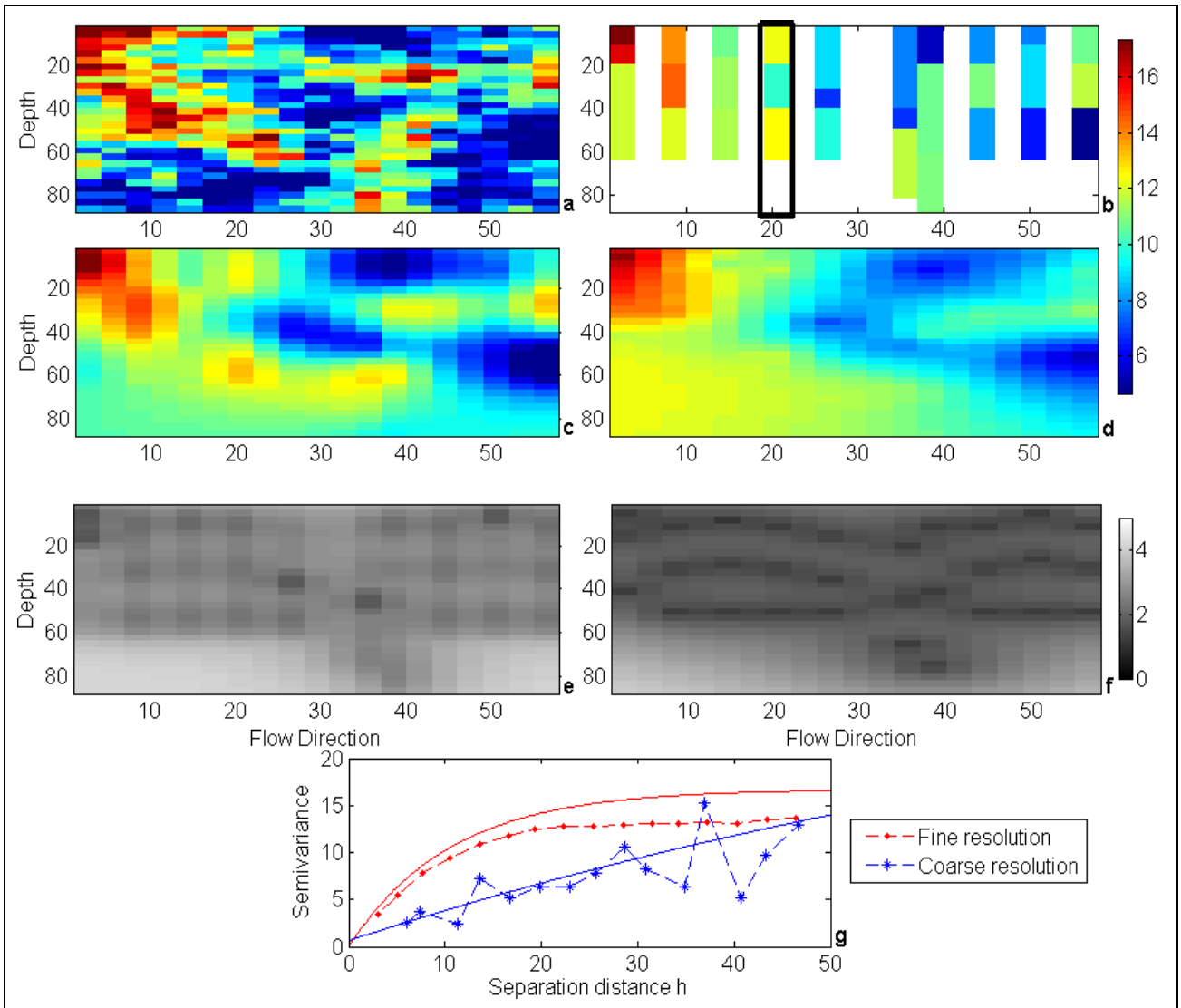

Figure 4-2 Data and estimates for pseudodata example: (a) fine-resolution attribute distribution, (b) available coarse-resolution (average) data, (c) GD best estimates, (d) OK best estimates, (e) GD estimation uncertainty standard deviation, (f) OK estimation uncertainty standard deviation, (g) experimental (dashed lines) and theoretical (solid lines) variograms. The coarse-resolution experimental variogram (blue dashed) was derived using available data in (b). The fine-resolution experimental variogram (red dashed) was derived using the true fine-resolution data in (a). Both theoretical variograms were derived using REML and the data in (b).

Figure 4-2e and Figure 4-2f also confirm that the uncertainty for both methods grows as the distance between the sampled cores and estimation locations increases, as would be expected. In addition, these figures and Figure 4-3a also suggest that the uncertainty from OK is lowest at the

center of each core section, while the GD the uncertainties within core sections are relatively uniform. This again reflects the different assumptions made by OK and GD about the location where the measurements reflect the true attribute value.

In summary, the results are consistent with our initial hypothesis that the variogram generated through GD shows better agreement with the actual variance and correlation lengths observed at the fine-resolution relative to the covariance parameters obtained through OK. GD also provides better estimates of the spatial distribution of the attribute and its associated uncertainty relative to OK for the pseudodata case.

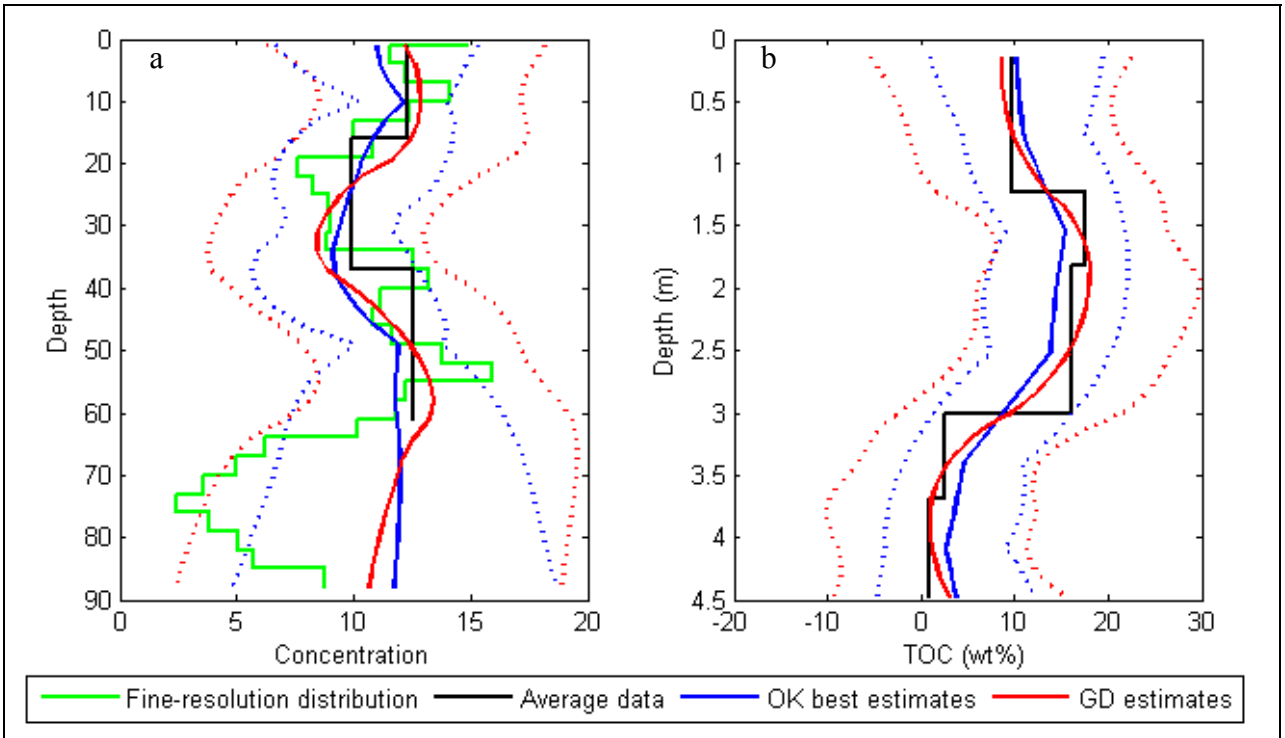

Figure 4-3 One-dimensional slices of estimates for (a) pseudodata and (b) field data examples. The locations of the two slices are indicated with black frame in Figure 4-2b and Figure 4-4a, respectively. Solid blue and red lines represent best estimates. Dotted lines represent 95% confidence intervals. Note that the true fine-resolution distribution is unknown for the field data example.

4.4.2 Field data application

Sediment TOC determinations are typically involved with contaminant analyses which is an important part of ecological risk assessment [*Schumacher*, 2002]. TOC forms water-soluble and water-insoluble complexes with metal ions and hydrous oxides, interact with clay minerals and bind particles together, as well as sorbs and desorbs both naturally-occurring and anthropogenically-introduced organic contaminant compounds [*Schumacher*, 2002]. The analysis of TOC distribution is therefore essential for the further analysis of contaminant distribution knowledge. The spatial analysis is needed to fully understand and quantify these correlations requires sensitive probabilistic techniques [*Adriaens et al.*, 2006]. The analysis of the TOC distribution is therefore an essential component of the analysis of contaminant distributions. TOC data from the lower Passaic River (Figure 4-4a) are analyzed here, using the same tools applied for the pseudodata application in Section 4.4.1.

The blue lines in Figure 4-4f and Figure 4-4g present the experimental variogram (dashed line) obtained using the samples from Figure 4-4a, together with the theoretical variogram fit (solid line) based on the parameters estimated using the OK version of the REML algorithm. The experimental variogram at the fine-scale resolution cannot be obtained in this case, because the attribute distribution at a fine resolution is unknown. However, the GD version of REML is used to derive the fine-scale variability. This approach was shown in Section 4.3 to yield a good representation of the true fine-scale variability. Based on the results from the pseudodata application, it is expected that the estimated variogram at the estimation scale would display more variability than the variogram at the measurement resolution, and this is indeed the case. Moreover, correlation lengths along the vertical direction (Figure 4-4f) from OK and GD were found to be much shorter than those along the flow direction (Figure 4-4g), which is consistent

with earlier work [*Barabas et al.*, 2001]. This anisotropy of sediment structure is caused by the fact that the streambed sediments are known to show a strong horizontal layered structure [*Salehin et al.*, 2004].

Figure 4-4a and Figure 4-4c present the best estimates of the attribute distribution at the estimation resolution for GD and OK, respectively. Consistent with the pseudodata example, it is also clear that the OK best estimates show less variability when compared to GD. This can also be seen in Figure 4-3b, which shows a one-dimensional vertical slice from the two-dimensional results presented in Figure 4-4. Although the "true" TOC distribution at the fine-scale resolution is not known in this case, the pseudodata example presented in Section 4.4.1 showed that the GD estimates are more representative of the expected variability at the fine-scale resolution. The significant differences between the OK (Figure 4-4c) and GD (Figure 4-4b) estimates confirm that explicitly accounting for scale differences has a large impact on the estimated spatial distribution of field attributes.

Figure 4-4d and Figure 4-4e present the estimation uncertainties associated with the GD and OK. Figure 4-3b also presents the 95% confidence intervals for one vertical slice from the two-dimensional results. The uncertainty estimates from the GD approach are again higher relative to those from OK. Because OK assumes that the measured average attribute values are representative of the variability at the finer estimation resolution, it underestimates the variability and the uncertainties at fine resolutions. This indicates that, although the estimated uncertainties from OK are lower, they cannot accurately represent the true uncertainties of the OK estimates. Note that, although not implemented here, a numerical approach such as a Gibbs sampling algorithm could be implemented to enforce nonnegativity in the uncertainty bounds [*Michalak*, 2008].

The GD method presented in this chapter provides a rigorous approach for using data sampled at a non-uniform resolution to characterize the spatial distribution of attributes in sediments. GD infers the degree of spatial variability at a target resolution, and uses this information, together with information about the resolution of individual available samples, to yield an estimate of the attribute distribution at the target resolution and its associated uncertainty. The pseudodata example demonstrated that the GD approach can more accurately represent the true degree of spatial variability in the underlying distribution, and yield better estimates with a more accurate assessment of uncertainties, relative to traditional OK. The field application confirms that these differences have an impact on estimates at the field scale. For the Passaic River application, the GD estimates of TOC suggest more continuity in the flow direction at this site relative to OK. In addition, the estimation uncertainty is very high in many areas of the domain, due to the uneven and sparse sampling in the field. The high uncertainties in Figure 4-4d indicate that the prediction precision is limited, and that additional sampling would be beneficial, especially along the flow direction.

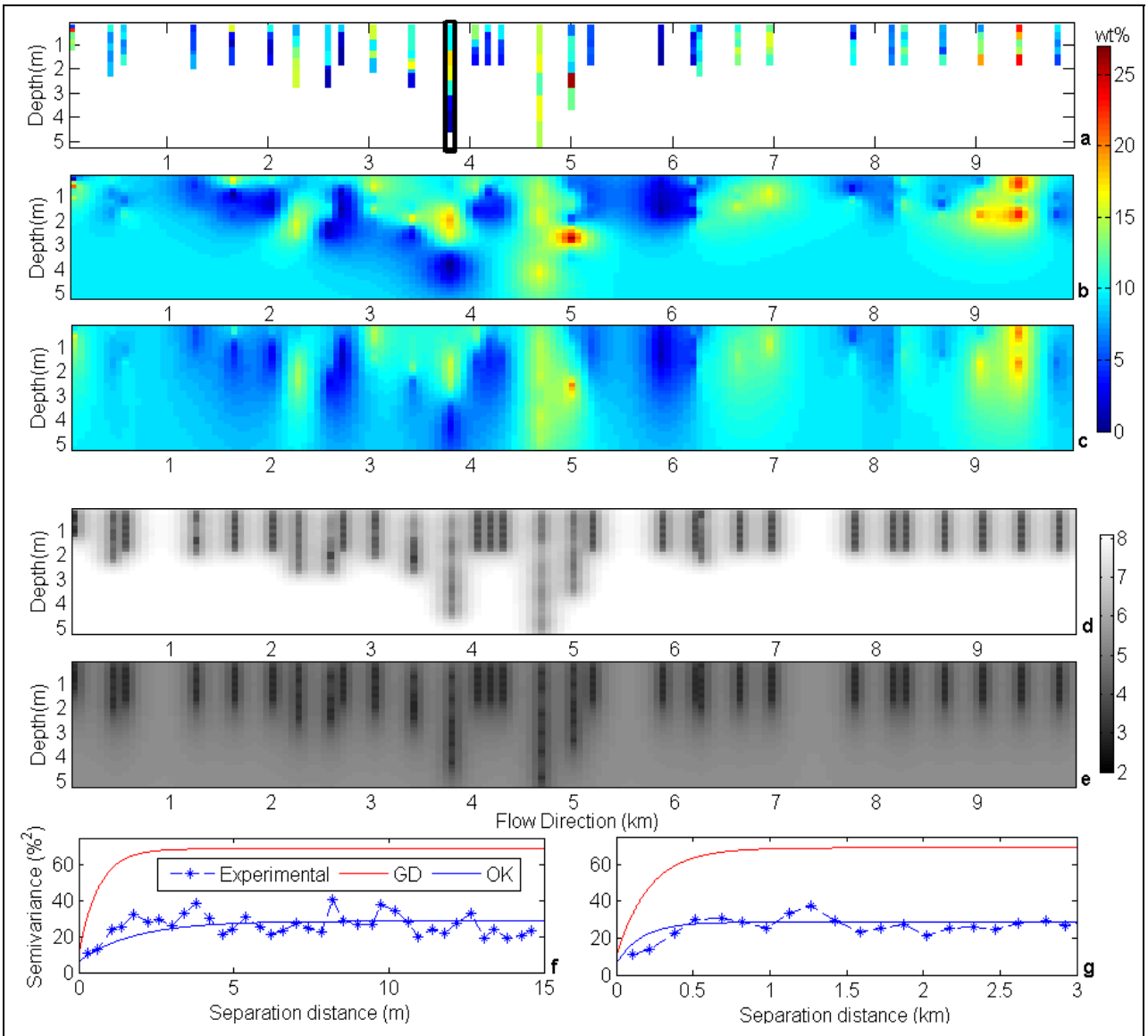

Figure 4-4 Data and estimates for field data TOC (a) available coarse-resolution (average) data, (b) GD best estimates, (c) OK best estimates, (d) GD estimation uncertainty standard deviation, (e) OK estimation uncertainty standard deviation, (f) experimental (dashed line) and theoretical (solid lines) variograms along vertical direction, (g) experimental (dashed line) and theoretical (solid lines) variograms along flow direction. The experimental variograms were derived using available data in (a). All theoretical variograms were derived using REML and the data in (a). Units: weight percentage.

The results from the pseudodata case and the field data case consistently suggest that GD is better able to represent the true fine-resolution variability and characterize the uncertainty

associated with estimates relative to OK, one of the most widely-used interpolation methods, while honoring the average concentrations measured at larger, inconsistent resolutions.

4.5 Conclusions

Samples of contaminated stream or riverbed sediments are often obtained from cores that vary in length because of the irregular thickness of sediment above the bedrock surface and tidal influence from the river flow. Such non-uniform resolution data sets create a challenge for traditional interpolation methods in accurately estimating the contaminant distribution at un-sampled locations. In this work, geostatistical downscaling was compared to the more traditional approach of point-to-point Ordinary Kriging for a hypothetical case study, and for total organic carbon observations from the Passaic River, New Jersey.

The geostatistical downscaling method presented in this chapter (i.e., *Zhou and Michalak* [2009]) provides a rigorous approach for using data sampled at a non-uniform resolution to characterize the spatial distribution of attributes in contaminated sediments. The proposed method infers the degree of spatial variability at a target resolution, and uses this information, together with information about the resolution of individual available samples, to yield an estimate of the attribute distribution at the target resolution and its associated uncertainty. Although frequently used to interpolate measurements, Ordinary Kriging is shown not to be able to estimate the spatial distribution of attributes accurately, because this approach assumes that data are sampled at a uniform resolution. Geostatistical downscaling, on the other hand, is shown to resolve this problem by explicitly accounting for the relationship between the known average measurements and the unknown fine-resolution attribute distribution to be estimated.

Using a pseudodata example, the approach is shown to more accurately represent the true degree of spatial variability in the underlying constituent distribution, and to yield better estimates with a more accurate assessment of uncertainties, relative to traditional point-to-point kriging methods. When the applied to the estimation of the total organic carbon distribution in the Passaic River, GD shows that the uncertainty associated with the spatial distribution of TOC is higher than would have been assumed if a kriging approach had been applied. While the application presented here involves water sediments, the methodology can also be adapted to other areas where sampling is conducted at multiple resolutions.

CHAPTER 5 Estimating Extent of Hypoxia in Lake Erie with Geostatistical Methods

The work described in this chapter focuses on applying improved methods for merging *in situ* water quality parameters and other auxiliary variables (e.g., spatial trend related data, and related remote sensing observations) to obtain more accurate spatial distribution of dissolved oxygen. In addition, this work developed a novel method of obtaining uncertainties of hypoxic extent (i.e., areas of dissolved oxygen below a certain threshold). The proposed methods are applied to estimate the historical hypoxic extent in the central basin of Lake Erie during August and September from 1987 to 2007.

5.1 Introduction

Hypoxia refers to an abnormally low level of dissolved oxygen (DO) in a body of water, which leads to a large range of potentially negative effects [*NSTC*, 2000]. Hypoxia is generally defined as occurring when DO concentrations drop below 2 mg/L [*Officer et al.*, 1984; *Pearson and Rosenberg*, 1978; *Pihl et al.*, 1991; *Rosenberg and Loo*, 1988], and can lead to degraded habitat, altered migration patterns of aquatic wildlife, and decreased water quality. Thus, industries that depend on ecosystem productivity, such as fisheries [*Diaz and Rosenberg*, 2011], can be negatively impacted. In addition, the occurrence of hypoxia can compromise drinking water supplies, cause beach closures, and necessitate restrictions on swimming, boating, and tourism [*NSTC*, 2000].

Summer hypoxia is a natural phenomenon in the central basin of Lake Erie, probably dating back thousands of years [*Delorme*, 1982]. However, evidence suggests that oxygen depletion rates increased in the 1950s due to anthropogenic factors [*Committee on Environment and Natural Resources*, 2010]. During summer, the strong vertical thermal gradient (i.e., stratification) reduces mixing and hence the flux of oxygen into the hypolimnion, where low

light restricts photosynthesis. The warm, light water near the surface acts as a barrier that prevents the colder, heavier bottom water from being replenished with oxygen. The lack of sufficient warmth and light in the bottom layer severely restricts the photosynthesis that produces oxygen. Decomposing phytoplankton and other settled organic matter consume oxygen, often reducing it to below the 2 mg/L hypoxic threshold in the hypolimnion, after they sink to the bottom from the surface water.

Hypoxic areal extent is an important indicator for tracking historical changes in water quality, and is often used as a response metric for water quality management [*Rabalais et al.*, 2002]. However, because of sparse sampling, the spatial and temporal dynamics of hypoxia in Lake Erie are poorly understood.

This work proposes the use of the Bayesian Information Criterion (BIC; [*Schwarz*, 1978]) for identifying a set of ancillary variables that best inform the distribution of DO in Lake Erie, followed by the application of Universal Kriging (UK, a.k.a. kriging with an external drift) and conditional realizations [*Chiles and Delfiner*, 1999], geostatistical interpolation methods that can incorporate these additional data sources to augment the sparse sampling of DO in Lake Erie. Together, these methods are used to characterize the spatial distribution of hypoxia in Lake Erie for August and September of 1987 to 2007.

To compensate for the sparseness of *in situ* DO measurements in Lake Erie, the data on ancillary variables with full spatial coverage were used to better represent the DO spatial variability. These ancillary variables could include any parameter that is correlated with DO, and therefore provides information about its distribution, such as water temperature [*Buzzelli et al.*, 2002] and salinity [*Zhang et al.*, 2006]. Remote sensing observations of parameters such as sea

surface temperature (SST) and surface chlorophyll α concentration offer other potentially useful sources of information because they are dense in spatial information that can be used to inform the spatial variability of DO. Specific data types used in this study are described in Section 5.2.

The objectives of this part of work are: (1) to produce accurate estimates of the spatial variability of DO, and of the interannual variability of the extent of the hypoxic zone, in the central basin of Lake Erie using BIC and UK as data fusion tools to combine remote sensing data with limited *in situ* measurements, (2) to understand and evaluate the utility of factors, such as bathymetry, SST and chlorophyll, in explaining the extent of hypoxia in Lake Erie, (3) to explore the impact of stratification dynamics on intra-seasonal changes in hypoxic extent, and (4) to develop a simple model of hypoxic extent based on the average measured bottom water DO concentration.

This work presents estimated DO concentrations only for the central basin of Lake Erie, because this is the basin where hypoxia has been found to occur most. The western basin is sufficiently shallow that the water column mixes and thereby diminishes hypoxia. The amount of phosphorous flowing into the eastern basin, on the other hand, is lower than that flowing into the west and central basins, and its water volume is sufficiently large that the DO remains relatively unaffected. Moreover, measurements of DO for these other two basins are not available for most years.

This chapter is organized as follows: Section 5.2 presents the data used in this study. Section 5.3 presents a brief overview of the inputs into geostatistical Universal Kriging, model selection method and conditional realizations. Section 5.4 presents the results of the analysis, including the

variation of hypoxic extent from August to September from 1987-2007. Section 5.5 summarizes the main conclusions of the study.

5.2 Data description

Dissolved Oxygen - The *in situ* DO data used in this study were collected by the U.S. EPA Great Lakes National Program Office [*Great Lake National Program Office (GLNPO)*, 2008], the National Water Research Institute of Environment Canada (NWRI, [*Rucinski et al.*, 2010]), and the National Oceanic and Atmospheric Administration (NOAA) Great Lakes Environmental Research Laboratory (GLERL). GLNPO and NWRI have been collecting DO data for April, August, September and October since the 1980s at ten (sometimes nine) fixed buoy stations in the central basin of Lake Erie (Figure 5-1, [*Esterby and Bertram*, 1993]). In addition, GLERL collected DO measurements at approximately 60 locations in the central basin of Lake Erie from May through October in 2005 and approximately 40 locations in the central basin in September 2007 for the International Field Years on Lake Erie (IFYLE) program [*Hawley et al.*, 2006]. Overall, the analysis presented here was based on 75 sampling cruises for August and September between 1987 and 2007 (Table 5-1), 61 of which detected hypoxia. Note that within this study period, no DO data are available for August 1992, 1994, and 1995 or for September 1991, 2000, and 2006. Note that, within the period covered here, no DO data were available in August 1992, 1994, and 1995 or in September 1991, 2000, and 2006.

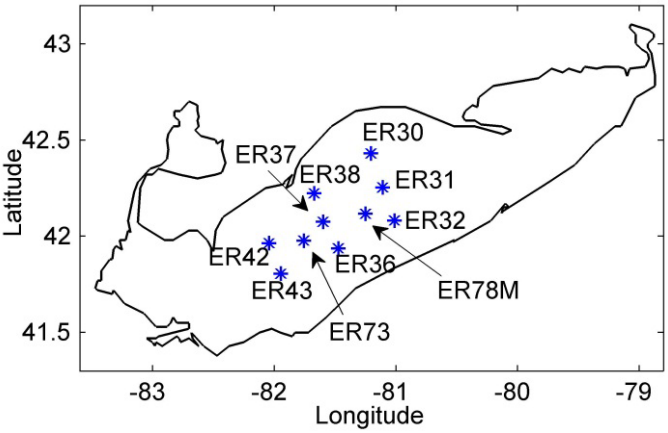

Figure 5-1 Regular DO measurement locations.

At each sampling location, DO concentrations were measured at about 3m, 1m, and <1m vertical intervals throughout the water column for the GLNPO, NWRI, and GLERL cruises, respectively. For the analysis presented here, the DO observations 1m or 2m above the lake bottom at each sampling location were used (depending on the deepest available observations), which are normally representative of the DO concentration in the hypolimnion.

The focus of our study was on August and September, the months when the hypoxic extent is typically at its maximum [*Bertram*, 1993]. The analysis was restricted to the central basin of Lake Erie, the basin most susceptible to hypoxia due to its depth and nutrient loading [*Rucinski et al.*, 2010].

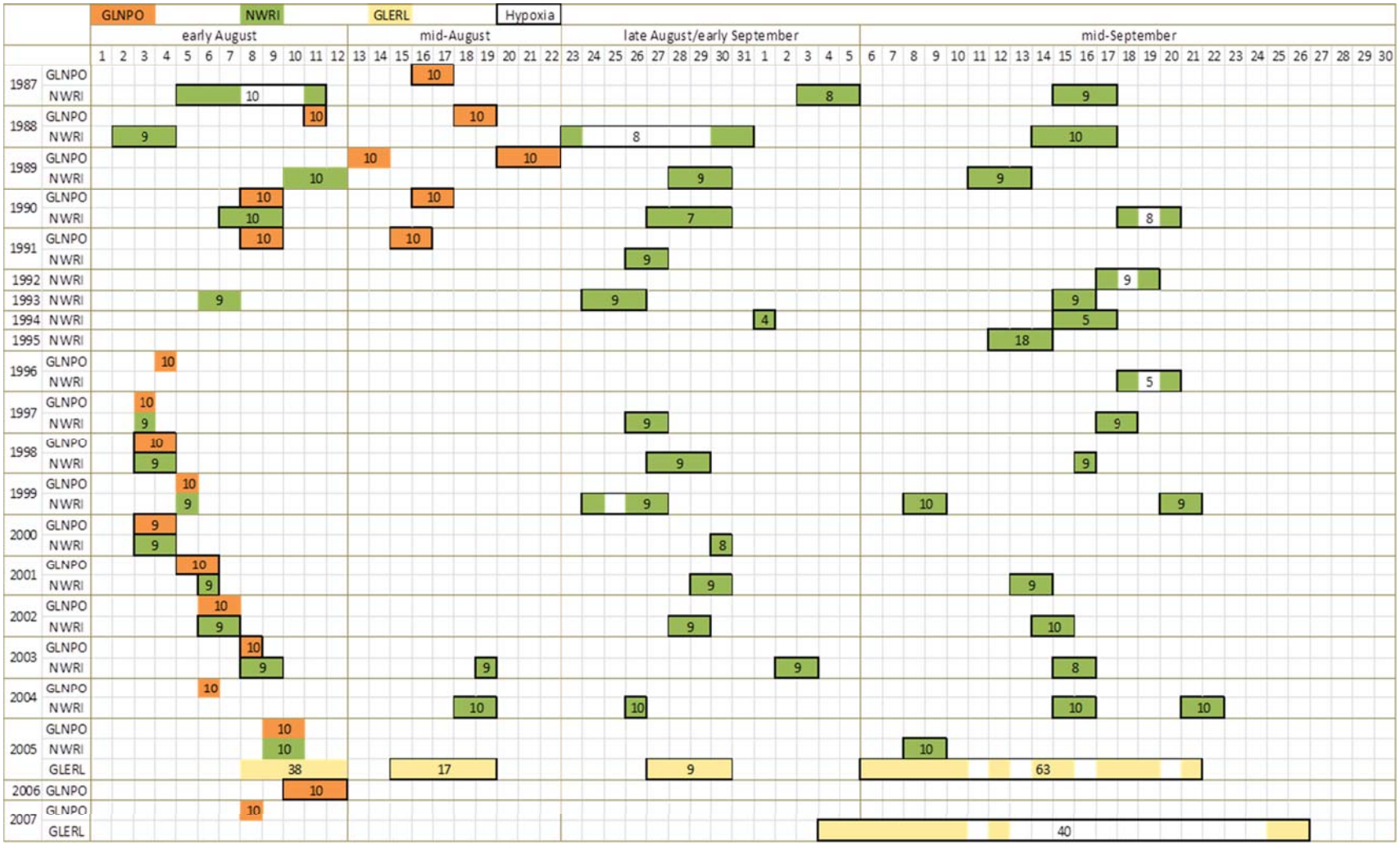

Table 5-1 Dates and total number of central basin DO measurements for all available GLNPO, NWRI, and GLERL cruises. The shading indicates the data source and the date(s) when sampling occurred in the central basin. The number for each cruise represents the number of observations in the central basin. Cruises that observed hypoxia (i.e. that had at least one observation below 2 mg/L) are outlined in black.

Auxiliary Data - To augment the sparse *in situ* DO measurements, auxiliary variables with full spatial coverage are included in the analysis. These variables, selected based on availability and known association with DO, include latitude, longitude, bathymetry, and satellite-derived monthly-average sea surface temperature (SST) and surface chlorophyll concentration from April to September.

The bathymetry data are a subset of the New Bathymetry of Lake Erie and Lake Saint Clair from the NOAA National Geophysical Data Center [*National Geophysical Data Center*, 2008]. The chlorophyll data are from the Sea-viewing Wide Field-of-view Sensor (SeaWiFS), and are available from the National Aeronautics and Space Administration (NASA) Ocean Color Web facility [*NASA Ocean Color Web facility*] at a resolution of 9 km × 9 km for 1998 onwards. SST data at 2.5 km × 2.5 km are available starting in 1992 from Great Lakes Surface Environmental Analysis [*Great Lakes Surface Environmental Analysis (GLSEA)*, 2008], a digital map of the Great Lakes surface temperature and ice cover produced daily from Advanced Very High Resolution Radiometer data by NOAA GLERL. These GLSEA data are produced specifically for the Great Lakes using a smoothing algorithm [*Schwab et al.*, 1999]. All data were regridded to a 2.5 km × 2.5 km resolution using nearest neighbor interpolation for the geostatistical analysis.

Analyzed Cases – Because the remote sensing data sets are not available for the full examined period, two case studies were designed and compared. Case 1 covered 1987 to 2007, and used only bathymetry, latitude, and longitude as auxiliary variables. Case 2 included these variables, plus the GLSEA SST and SeaWiFS chlorophyll data for April through September, and covered 1998 to 2007. The target analysis is annual DO distribution in the bottom water layer

during August and September, when hypoxia is typically at a maximum for Lake Erie [*Bertram*, 1993; *Hawley et al.*, 2006].

5.3 Methodology

This section describes the geostatistical framework for estimating yearly hypoxic extent for the central basin of Lake Erie.

5.3.1 Universal Kriging

UK uses auxiliary variables, in this work latitude, longitude, bathymetry, SST and chlorophyll, in addition to the limited *in situ* DO measurements, to obtain a detailed picture of the spatial distribution of DO in Lake Erie. UK has been widely used in environmental sciences, for applications such as estimating snow accumulation, temperature and precipitation in meteorology [*Arthern et al.*, 2006; *Erickson et al.*, 2005; *Haylock et al.*, 2008] characterizing contaminant distributions [*Jerrett et al.*, 2001], and identifying the sources and sinks of carbon dioxide in atmospheric science [*Mueller et al.*, 2010]. Within the context of estimating hypoxia, *Murphy et al.* [2010] recently used UK, among other approaches, to interpolate the spatial distribution of water quality parameters such as salinity, water temperature and dissolved oxygen in the Chesapeake Bay using water quality modeling outputs as ancillary sources of information. *Obenour et al.* [2012] used a UK-type model to separate the impacts of stratification and nutrient loading on DO in the Gulf of Mexico.

The role of auxiliary variables in UK is analogous to the role of independent variables (covariates) in multiple linear regression. Contrary to linear regression, however, UK also (i) accounts for the spatial correlation (i.e. smoothness) of the DO distribution, and (ii) is an exact interpolator, such that it reproduces all the available DO observations to within their measurement error. UK using data from all the cruises with hypoxic measurements were

performed simultaneously, such that the relationship between auxiliary variables and DO remained constant from cruise to cruise. The estimates of DO themselves, however, were cruise-specific, and no correlation was assumed among regression residuals from different cruises.

For each of the two examined cases, the $n \times 1$ observation vector $\mathbf{z}$ of DO is defined as:

$$\mathbf{z} = \begin{bmatrix} \mathbf{z}_1 \\ \mathbf{z}_2 \\ \vdots \\ \mathbf{z}_y \end{bmatrix} \tag{5.1}$$

where $\mathbf{z}_i$ (i = 1, 2, ... , y) are $n_i \times 1$ vectors of DO measurements, n_i is the number of DO measurements for the i^{th} cruise (i.e., $n = \sum_{i=1}^{y} n_i$), and y is the total number of cruises for which DO data are used in each case (y=61 for Case 1; y=32 for Case 2).

Within the UK framework, the DO distribution is modeled as the sum of a deterministic term (trend) and a zero-mean stochastic term (spatially-correlated residuals). The deterministic term represents the portion of the DO distribution that can be explained by the available auxiliary variables and spatially-constant cruise offsets (corresponding to each cruise), and the stochastic term represents the remaining portion of the observed variability:

$$\mathbf{z} = \mathbf{X_z}\boldsymbol{\beta} + \mathbf{z}_{res} \tag{5.2}$$

where $\mathbf{X_z}$ is a known $n \times (y+p)$ matrix of y categorical variables (corresponding to unique offsets [i.e. intercepts] for each cruise) and p auxiliary variables, $\boldsymbol{\beta}$ is a $(y+p) \times 1$ vector of unknown drift coefficients on these variables, and $\mathbf{z}_{res}$ is an $n \times 1$ vector of residuals. The approach used for selecting a subset of auxiliary variables from among those listed in Section 5.2 is presented in Section 5.3. Overall, the model of the trend, $\mathbf{X_z}$, is expressed as:

$$\mathbf{X_z} = \begin{bmatrix} \mathbf{1}_1 & \cdots & \emptyset & \mathbf{X}_1 \\ \vdots & \ddots & \vdots & \vdots \\ \emptyset & \cdots & \mathbf{1}_y & \mathbf{X}_y \end{bmatrix} \tag{5.3}$$

where $\mathbf{1}_i$ (i = 1, 2, ... , y) are $n_i \times 1$ vectors of ones, and $\mathbf{X}_i$ is an $n_i \times p$ matrix of auxiliary variables representing the trend for each cruise. Because the mean DO concentration is expected to change from cruise to cruise, $\mathbf{X_z}$ includes y columns of categorical variables (ones and zeros). The components in $\boldsymbol{\beta}$ that multiply the first y columns of $\mathbf{X_z}$ represent a constant offset in DO concentrations for each cruise.

The stochastic term $\mathbf{z}_{res}$ is modeled as spatially-correlated residuals. A spatial covariance function [*Chiles and Delfiner*, 1999], which quantifies the degree to which the spatial correlation between a pair of locations decays as a function of their separation distance h was defined as:

$$Q_{zz}(h) = \begin{cases} \sigma^2 + \sigma_{\mathbf{Q}}^2, & h = 0 \\ \sigma^2 \exp\left(-\frac{h}{l}\right), & h > 0 \end{cases} \qquad 5.4$$

where σ^2 is the variance of the portion of the residual DO variability that is spatially correlated, 3l is the practical correlation range, and $\sigma_{\mathbf{Q}}^2$ is the variance of the portion of the variability that is not spatially correlated (e.g. measurement error). These three model parameters were optimized by fitting the theoretical model (equation 5.4) to the empirical covariance of the residuals using nonlinear least-squares [*Bogaert and Russo*, 1999]. No covariance was assumed among cruises, and the overall $n \times n$ covariance matrix of the DO observations is therefore defined as:

$$\mathbf{Q_{zz}(h)} = \begin{bmatrix} \mathbf{Q_1} & \cdots & \emptyset \\ \vdots & \ddots & \vdots \\ \emptyset & \cdots & \mathbf{Q}_y \end{bmatrix} \qquad 5.5$$

where $\mathbf{Q}_i$ (i = 1, 2, ... , y) is an $n_i \times n_i$ covariance matrix for the residuals in each cruise, and all the $\mathbf{Q}_i$'s use the same covariance parameters (σ^2, $\sigma_{\mathbf{Q}}^2$, and l).

The covariance matrix is used in the UK system of linear equations:

$$\begin{bmatrix} \mathbf{Q_{zz}} & \mathbf{X_z} \\ \mathbf{X_z}^T & \mathbf{0} \end{bmatrix} \begin{bmatrix} \boldsymbol{\Lambda}^T \\ \mathbf{M} \end{bmatrix} = \begin{bmatrix} \mathbf{Q_{zs}} \\ \mathbf{X_s}^T \end{bmatrix} \qquad 5.6$$

where $\mathbf{Q_{zs}}$ is an $n \times m$ covariance matrix between the measurement and estimation locations obtained analogously to equation 5.5, the matrix $\mathbf{X_s}$ contains the same cruise-specific offsets and auxiliary variables as in $\mathbf{X_z}$ but defined at the estimation locations, and T denotes a matrix transposition. The system of equation 5.6 was solved for $\mathbf{\Lambda}$, which is an $m \times n$ matrix of weights assigned to each observation for each estimation location, and for $\mathbf{M}$, which is a $(y+p) \times m$ matrix of Lagrange multipliers. Finally, $\mathbf{\Lambda}$ and $\mathbf{M}$ were used to obtain estimates of the DO distribution, and their associated uncertainties, throughout the central basin:

$$\hat{\mathbf{s}} = \mathbf{\Lambda z} \tag{5.7}$$

$$\mathbf{V_{\hat{s}}} = -\mathbf{X_s M} + \mathbf{Q_{ss}} - \mathbf{Q_{zs}^T \Lambda^T} \tag{5.8}$$

where $\hat{\mathbf{s}}$ is an $m \times 1$ vector of DO estimates, $\mathbf{V_{\hat{s}}}$ is the covariance matrix representing the covariances associated with these estimates, and $\mathbf{Q_{ss}}$ is an $m \times m$ covariance matrix between the residuals of estimates, obtained in the same way and using the same parameters as in equation 5.4. The square roots of the diagonal elements of $\mathbf{V_{\hat{s}}}$ are the standard deviations (i.e., estimation uncertainties) of the DO estimates. Because the thermocline always appears at depths of 15m to 20m during summer [*Rao et al.*, 2008], and because shallower areas are usually oxygenated and rarely sampled, the DO estimates for areas with depths of less than 15m were constrained to be above the hypoxic threshold.

The best estimate of drift coefficients ($\widehat{\boldsymbol{\beta}}$) of the auxiliary variables were obtained as [*Chiles and Delfiner*, 1999]:

$$\widehat{\boldsymbol{\beta}} = (\mathbf{X_z^T Q_{zz}^{-1} X_z})^{-1} \mathbf{X_z^T Q_{zz}^{-1} z} \tag{5.9}$$

and their associated covariances were:

$$\mathbf{V_{\hat{\beta}}} = (\mathbf{X_z^T Q_{zz}^{-1} X_z})^{-1} \tag{5.10}$$

where the square roots of the diagonal elements of $\mathbf{V}_{\hat{\boldsymbol{\beta}}}$ are the estimation uncertainties of the individual parameters, and the off-diagonal terms represent their estimated covariances.

Ordinary Kriging (OK), one of the most common geostatistical approaches, is used in this work for comparison to the UK estimates. OK also uses a covariance function to represent the spatial correlation, but it cannot use information about other auxiliary variables. Therefore, the spatial structure representing the DO distribution is obtained only from the *in situ* measurements, the matrix $\mathbf{X_z}$ therefore becomes:

$$\mathbf{X_z} = \begin{bmatrix} \mathbf{1}_1 & \cdots & \emptyset \\ \vdots & \ddots & \vdots \\ \emptyset & \cdots & \mathbf{1}_y \end{bmatrix} \tag{5.11}$$

such that the (spatial) mean DO concentration is constant for each cruise but can differ from cruise to cruise. Otherwise, equations 5.6 - 5.10 remain unchanged for OK.

5.3.2 Geostatistical auxiliary variable selection

The Bayesian Information Criterion [*Anderson et al.*, 1998] is used as a statistical variable selection method to identify the subset of the auxiliary variables described in Section 5.2 to be used in UK. The purpose of auxiliary variable selection is to choose a subset of variables that can reliably represent the spatial distribution of DO [*Faraway*, 2004]. Adding more variables to the trend of a model will always reproduce more of the variability observed in the DO measurements, but will also make the model more complex. Moreover, some of the variables may not provide significant insights into the DO variability, and may only serve to represent a spurious correlation that is not reliable for estimating DO at unsampled locations. Therefore, it is necessary to identify a model that balances the degree to which available DO observations are reproduced (i.e., minimizes the sum of the squared residuals) with the complexity of the model (i.e. the number of selected variables).

Several statistical tests or criteria are available for model selection. BIC, based on Bayesian factor or posterior probability [*Schwarz*, 1978], is used in this study because it is able to evaluate both nested and non-nested competing models. Instead of comparing the statistical significance of the difference between two models as in traditional hypothesis-testing-based approaches, BIC ranks how well a model with a specific set of variables explains the observations. Mathematically, the BIC criterion is defined as [*Schwarz*, 1978]:

$$\mathrm{BIC} = -2\ln(L) + p\ln(n) \tag{5.12}$$

where L is likelihood of the observations.

If the residuals are normally distributed, the negative natural log of the likelihood becomes:

$$-ln(L) = \frac{n}{2} ln(2\pi) + \frac{1}{2} ln|\mathbf{Q_{zz}}| + \frac{1}{2}(\mathbf{z} - \mathbf{X_z}\boldsymbol{\beta})^T \mathbf{Q_{zz}^{-1}}(\mathbf{z} - \mathbf{X_z}\boldsymbol{\beta}) \tag{5.13}$$

After minimizing the likelihood with respect to the unknown drift coefficients $\boldsymbol{\beta}$ [*Hoeting et al.*, 2006], and ignoring the constant term, the log-likelihood function becomes:

$$-ln(L) = \frac{1}{2} ln|\mathbf{Q_{zz}}| + \frac{1}{2}\mathbf{z}^T(\mathbf{Q_{zz}^{-1}} - \mathbf{Q_{zz}^{-1}X_z}(\mathbf{X_z^T Q_{zz}^{-1} X_z})^{-1}\mathbf{X_z^T Q_{zz}^{-1}})\mathbf{z} \tag{5.14}$$

BIC is evaluated for each possible combination of auxiliary variables, and the set of variables with the lowest BIC is identified as the best model. The auxiliary variables of the best model are the then used in UK.

5.3.3 Conditional realizations

UK yields estimates of DO concentrations across space, but cannot be used directly to estimate the hypoxic extent (i.e. the area for which DO concentration is below a given threshold) and its associated uncertainty. Conditional realizations (a.k.a. simulations, [*Gutjahr et al.*, 1994;

Kitanidis, 1995; *Michalak et al.*, 2004]) of the DO distribution do this by providing equally-likely alternative DO distribution scenarios. The detailed procedure for generating conditional realizations is discussed by *Gutjahr et al.* [1994], *Kitanidis* [1995], and *Michalak et al.* [2004]. Each realization is an equally likely realization that follows the correlation structure **Q**, and reproduces the observations **z** within the measurement error $\mathbf{z}_{res}$ [*Michalak et al.*, 2004]. These realizations follow the spatial covariance $\mathbf{Q}_{ss}$ and are consistent with all available observations.

Each realization ($\mathbf{s}_{ci}$, $m \times 1$) is defined as [*Kitanidis*, 1996]:

$$\mathbf{s}_{ci} = \mathbf{\Lambda}(\mathbf{z} - \mathbf{z}_{ui}) + \mathbf{s}_{ui} \tag{5.15}$$

where $\mathbf{\Lambda}$ is the $m \times n$ matrix of weights defined in equation 5.6, and $\mathbf{z}_{ui}$ ($n \times 1$) and $\mathbf{s}_{ui}$ ($m \times 1$) are unconditional realizations at measurement locations and estimation locations, respectively, obtained from:

$$\begin{bmatrix} \mathbf{z}_{ui} \\ \mathbf{s}_{ui} \end{bmatrix} = \mathbf{C}^T \mathbf{u} \tag{5.16}$$

where **u** is an $(n+m) \times 1$ vector of normally distributed random values with zero mean and unit variance (note that a new vector **u** is generated for each realization), and **C** is the $(n+m) \times (n+m)$ matrix resulting from the Cholesky decomposition of the covariance matrix below:

$$\begin{bmatrix} \mathbf{Q}_{zz} & \mathbf{Q}_{zs} \\ \mathbf{Q}_{zs}^T & \mathbf{Q}_{ss} \end{bmatrix} = \mathbf{C}\mathbf{C}^T \tag{5.17}$$

Conditional realizations are generated for regions of the central basin with a depth greater than 15m, and the hypoxic area is calculated for each realization by summing the areas where the DO concentration is below 2 mg/L. A thousand realizations are generated for each cruise, and the results are used to develop a probabilistic estimate of the hypoxic extent.

5.4 Results and discussion

5.4.1 Variables explaining the spatial variability of DO in Lake Erie

The selected auxiliary variables, together with the cruise-specific offsets, explained 53% of the DO variability for the 1987-2007 Case 1 data set. Consistent variable selections between Cases 1 and 2 indicate that the difference in the timespans does not affect the significance of the auxiliary variables. The estimated drift coefficients, $\hat{\boldsymbol{\beta}}$, for the selected auxiliary variables (Table 1) explain a portion of the within-cruise spatial variability, while the cruise-specific offsets (not shown) account for temporal variability in DO due to other cofactors, such as nutrient loading or circulation [*Rao et al.*, 2008; *Rucinski et al.*, 2010]. The consistency of the $\hat{\boldsymbol{\beta}}$ values between cases further confirms that the relationships between DO and the auxiliary variables are consistent for different time periods.

Longitude, depth, and a quadratic depth term (i.e. depth squared) were selected through the BIC analysis as being significant for both examined cases (Table 5-2). Longitude was found to be positively correlated with DO, potentially acting as a proxy for phosphorus availability, which primarily enters the central basin from the west. Bathymetry was also correlated with DO, consistent with the fact that stratification is related to bathymetry (i.e., both the thermal structure and thickness of the bottom layer) in Lake Erie [*Loewen et al.*, 2007; *Schertzer et al.*, 1987]. Based on the regression coefficients for depth and depth-squared, the bottom water DO concentration is expected to be lowest at a station depth of around 23m (close to the 24m maximum depth of the central basin), all other factors being equal.

Case	Longitude (mg/L)/degree	Depth (mg/L)/m	Depth^2 (mg/L)/m^2
1	1.24±0.26	-1.40±0.22	0.03±0.006
2	1.17±0.33	-1.36±0.27	0.03±0.007

Table 5-2 Drift coefficients ($\widehat{\boldsymbol{\beta}}$) and associated uncertainties ($\boldsymbol{\sigma}_{\widehat{\boldsymbol{\beta}}}$) for the selected auxiliary variables for the two cases.

Neither of the remote sensing data sets (only available in Case 2) was found to significantly improve the model. In general, this suggests a more complex relationship between lake surface properties (SST and chlorophyll) and bottom DO. Surface conditions are likely decoupled from bottom conditions due to stratification and varying circulation patterns in the epilimnion and hypoliminion. For example, *Walker and Rabalais* [2006] suggested that a relationship between chlorophyll and hypoxia was not observed in the Gulf of Mexico due to the various physical and biological processes that confound a direct spatial correlation. In addition, the significance of chlorophyll may have been further diminished due to the quality of the satellite data product, which is known to have considerable uncertainty in the central basin [*Witter et al.*, 2009].

5.4.2 Spatial extent of summer hypoxia each year

Because the two examined cases yield consistent estimates of the hypoxic extent and use the same auxiliary variables, the discussion to the extents and uncertainties determined were restricted in Case 1. The extents were derived from the conditional realizations of the DO distribution (Figure 5-2), and the maximum estimated hypoxic extent for August and September of each year and associated uncertainty (Figure 5-3) show that the maximum extent occurs most often between late August and mid-September (Figure 5-2).

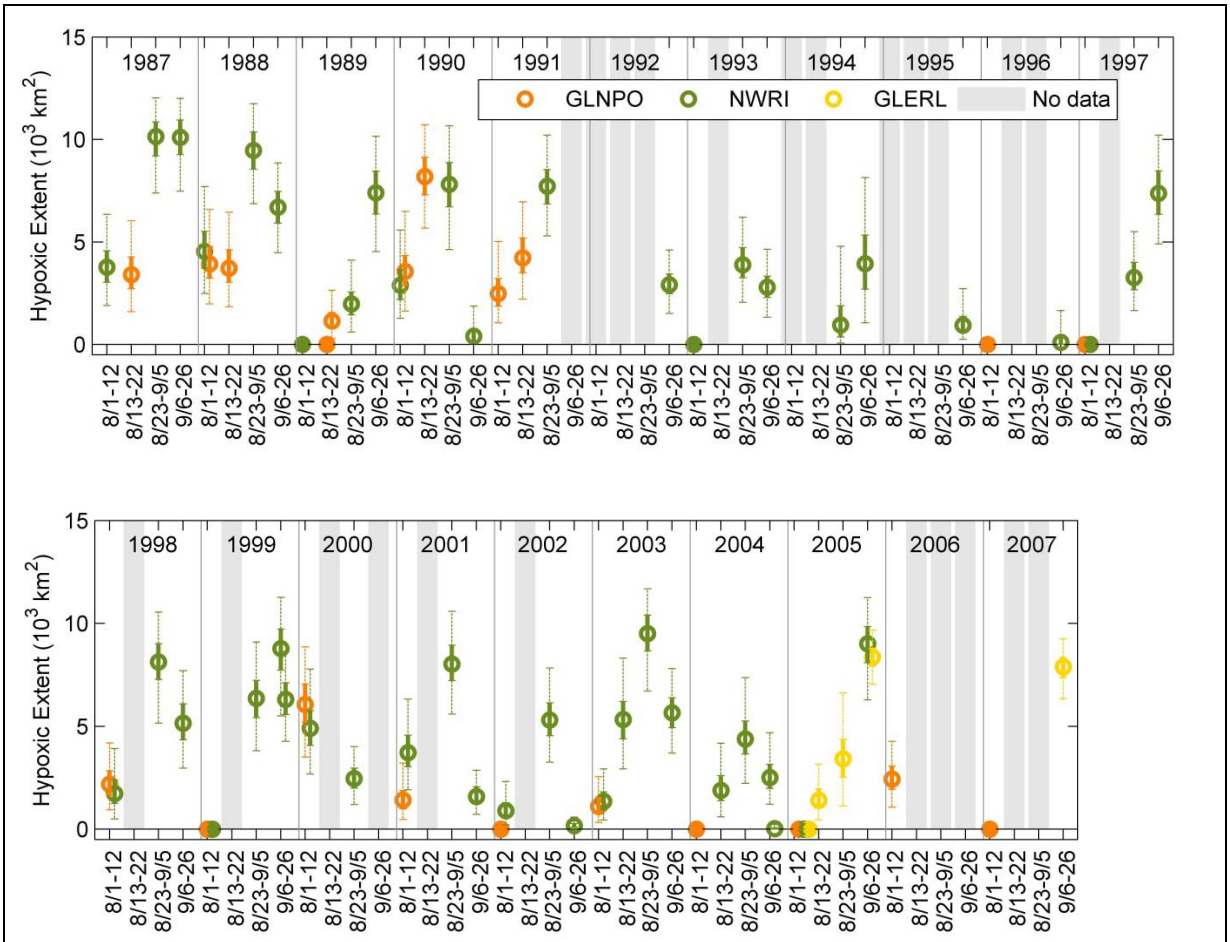

Figure 5-2 Plot of the median (circle), interquartile range (solid line), and 95% confidence intervals (dashed line) of the estimated hypoxic extent in early August (8/1-12), mid-August (8/13-22), late August and early September (8/23-9/5), and mid-September (9/6-26) from all available cruises with observed hypoxia from GLNPO, NWRI and GLERL for 1987-2007. The hypoxic extents of zero shown in solid circle represent cruises for which no hypoxia was observed, and periods shaded in grey represent times during which no data were available.

Results are qualitatively consistent with those of *Makarewicz and Bertram* [1991], as well as *Hawley et al.* [2006] (Figure 5-3). *Makarewicz and Bertram* [1991] reported that hypoxic extent decreased from the late 1980s to early 1990s as a result of the phosphorus load abatement programs, a part of the United States/Canada Great Lakes Water Quality Agreement of 1972 [*Dolan*, 1993]. *Hawley et al.* [2006] subsequently reported that hypoxic extent increased and

remained relatively high in the late 1990s and early 2000s, likely due to an increase in nonpoint source phosphorus loading or nutrient recycling by dreissenids (i.e. zebra and quagga mussels) that appeared in the system in the late 1980s [*Vanderploeg et al.*, 2009].

The smallest yearly hypoxic zones were observed in 1995 and 1996 (Figure 5-2, Figure 5-3), but these years had only one and two cruises, respectively (Table 5-3). In 1995, the only available data were from mid-September. In 1996, August sampling data were available for August 2-4, followed by sampling in mid-September, and at only five locations. It is therefore possible that periods with larger hypoxic zones were missed in those years. In 2002 and 2004, for example, the hypoxic extent was also small in early August and mid-September, but was larger during the interim period (Figure 5-3). On the other end of the spectrum, the largest estimated hypoxic extents exceeded 9000 km^2 (nearly two thirds of the surface area of the central basin deeper than 15m) in early and mid-September 1987, late August 1988, early September 2003, and mid-September 2005.

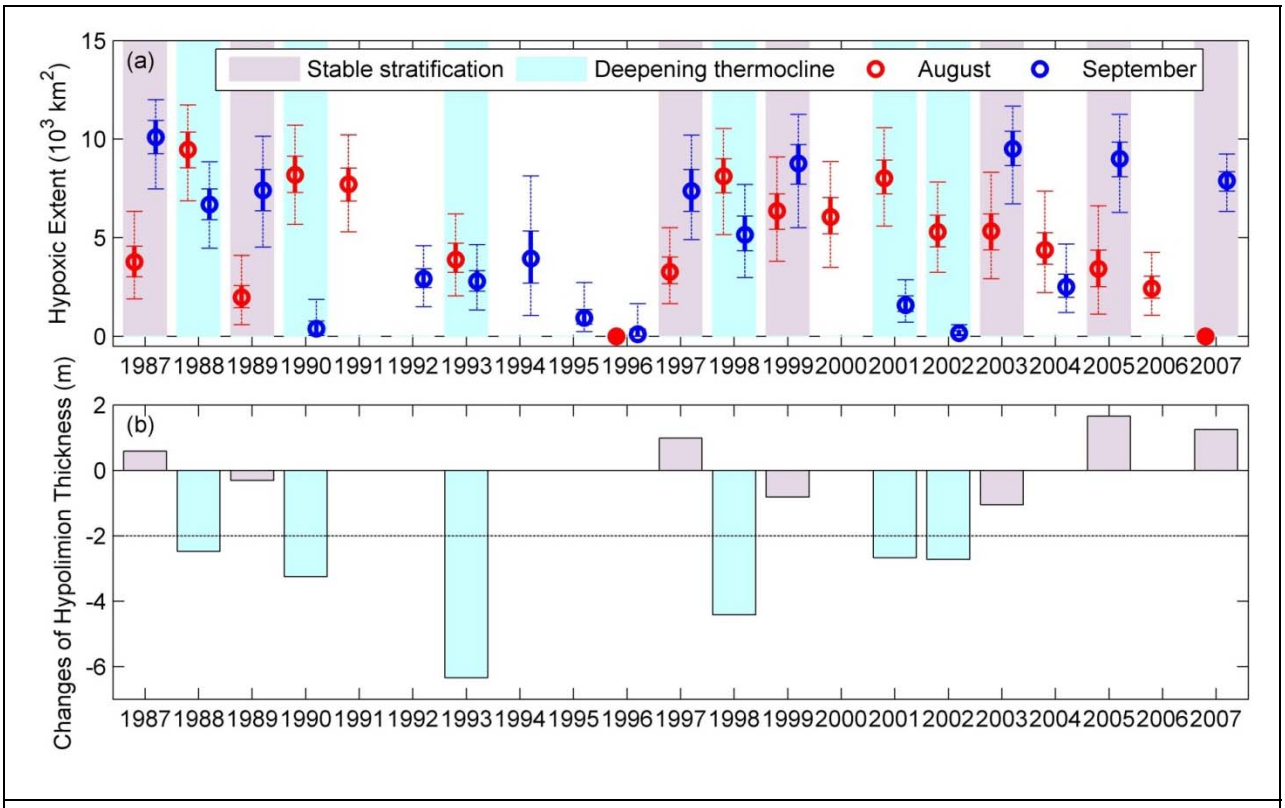

Figure 5-3 History of (a) the observed maximum hypoxic extent in August and September in the central basin of Lake Erie for 1987 to 2007, and (b) August to September change in hypolimnion thickness. In (a), solid circles represent months where cruises indicated no hypoxia; for months when hypoxia was observed, open circles represent the median, solid lines represent the interquartile range, and dashed lines represent the 95% confidence intervals based on the conditional realizations. Years when the August to September decrease in the hypolimnion thickness was less than 2m (i.e. years with stable stratification) are indicated in purple in panels (a) and (b); conversely, years with a decrease of more than 2m, corresponding to a deepening thermocline and early re-oxygenation, are indicated in light blue. Data of changes in hypolimnion thickness are not available for 1996 and 2004.

For a given number of measurements, the uncertainties associated with the estimated extents are generally higher for months with larger hypoxic areas. These uncertainties (representing as 95% confidence intervals) range from nearly zero for mid-September 2002 when hypoxia was negligible, to almost 6,000 km^2 for September 1999. Months with large hypoxic extents have considerable areas with estimated DO concentrations close to the 2mg/L hypoxic threshold, leading to the large uncertainty on the exact area that is hypoxic. Uncertainties are, as expected,

higher for cruises with fewer measurements, such as for mid-September 1994, when only 5 measurements were available and the 95% confidence intervals span 7,000 km^2 (Table 5-1, Table 5-3). Compared to other years with similar hypoxic extents, uncertainties for September 2005 and 2007 are smaller due to the more extensive DO measurements available for those two years through the IFYLE program by GLERL (Figure 5-3).

Source	Year	Month	Date	Hypoxic extent (10^3 km^2)				
				2.5%	25%	50%	75%	97.5%
NWRI	1987	Aug	5-7,11	1.90	3.02	3.77	4.56	6.34
GLNPO	1987	Aug	16-17	1.59	2.71	3.41	4.26	6.02
NWRI	1987	Sep	3-5	7.38	9.19	10.13	10.85	12.02
NWRI	1987	Sep	15-17	7.47	9.25	10.09	10.94	11.99
NWRI	1988	Aug	2-4	2.48	3.72	4.53	5.51	7.70
GLNPO	1988	Aug	11	1.97	3.23	3.93	4.77	6.57
GLNPO	1988	Aug	18-19	1.83	3.02	3.72	4.61	6.45
NWRI	1988	Aug	23,30-31	6.86	8.53	9.46	10.35	11.73
NWRI	1988	Sep	14-17	4.47	5.91	6.69	7.46	8.84
NWRI	1989	Aug	10-12	No observed hypoxia				
GLNPO	1989	Aug	13-14	No observed hypoxia				
GLNPO	1989	Aug	20-22	0.28	0.79	1.15	1.55	2.64
NWRI	1989	Aug	28-30	0.59	1.45	1.98	2.56	4.10
NWRI	1989	Sep	11-13	4.52	6.35	7.39	8.45	10.14
NWRI	1990	Aug	7-9	1.27	2.19	2.89	3.66	5.58
GLNPO	1990	Aug	8-9	1.63	2.86	3.56	4.34	6.48
GLNPO	1990	Aug	16-17	5.66	7.29	8.19	9.13	10.70
NWRI	1990	Aug	27-30	4.63	6.70	7.81	8.86	10.65
NWRI	1990	Sep	18,20	0.04	0.17	0.40	0.77	1.87
GLNPO	1991	Aug	8-9	1.05	1.86	2.48	3.20	5.02
GLNPO	1991	Aug	15-16	2.20	3.49	4.23	5.18	6.95
NWRI	1991	Aug	26-27	5.29	6.85	7.71	8.52	10.20
NWRI	1992	Sep	17,19	1.51	2.47	2.91	3.43	4.59
NWRI	1993	Aug	6-7	No observed hypoxia				
NWRI	1993	Aug	24-26	2.05	3.24	3.88	4.72	6.20
NWRI	1993	Sep	15-16	1.33	2.28	2.79	3.33	4.64
NWRI	1994	Sep	1	0.06	0.37	0.95	1.88	4.78
NWRI	1994	Sep	15-17	1.05	2.69	3.94	5.33	8.13
NWRI	1995	Sep	12-15	0.24	0.62	0.94	1.36	2.72

Source	Year	Month	Date	Hypoxic extent (10^3 km^2)				
				2.5%	25%	50%	75%	97.5%
GLNPO	1996	Aug	4	No observed hypoxia				
NWRI	1996	Sep	18,20	0.00	0.02	0.11	0.36	1.65
GLNPO	1997	Aug	3	No observed hypoxia				
NWRI	1997	Aug	3	No observed hypoxia				
NWRI	1997	Aug	26-27	1.66	2.66	3.27	4.01	5.50
NWRI	1997	Sep	17-18	4.90	6.34	7.37	8.45	10.20
GLNPO	1998	Aug	3-4	0.93	1.65	2.18	2.83	4.18
NWRI	1998	Aug	3-4	0.49	1.24	1.73	2.38	3.91
NWRI	1998	Aug	27-29	5.15	7.27	8.12	9.00	10.53
NWRI	1998	Sep	16	2.97	4.34	5.15	6.09	7.70
GLNPO	1999	Aug	5	No observed hypoxia				
NWRI	1999	Aug	5	No observed hypoxia				
NWRI	1999	Aug	24, 26-27	3.81	5.42	6.35	7.22	9.09
NWRI	1999	Sep	8-9	5.50	7.71	8.77	9.72	11.25
NWRI	1999	Sep	20-21	4.27	5.56	6.30	7.11	8.63
GLNPO	2000	Aug	3-4	3.49	5.19	6.06	7.04	8.85
NWRI	2000	Aug	3-4	2.68	4.07	4.90	5.76	7.77
NWRI	2000	Aug	30	1.20	1.98	2.46	2.96	4.01
GLNPO	2001	Aug	5-6	0.47	0.99	1.41	1.88	3.21
NWRI	2001	Aug	6	1.91	3.02	3.72	4.56	6.31
NWRI	2001	Aug	29-30	5.59	7.21	8.02	8.92	10.57
NWRI	2001	Sep	13-14	0.71	1.26	1.59	2.05	2.86
GLNPO	2002	Aug	6-7	No observed hypoxia				
NWRI	2002	Aug	6-7	0.20	0.57	0.89	1.29	2.31
NWRI	2002	Aug	28-29	3.24	4.53	5.30	6.13	7.82
NWRI	2002	Sep	14-15	0.02	0.09	0.16	0.27	0.59
GLNPO	2003	Aug	8	0.32	0.77	1.10	1.54	2.55
NWRI	2003	Aug	8-9	0.45	0.94	1.35	1.81	2.92
NWRI	2003	Aug	19	2.92	4.38	5.33	6.20	8.31
NWRI	2003	Sep	2-3	6.70	8.65	9.50	10.39	11.67
NWRI	2003	Sep	15-16	3.69	4.92	5.65	6.38	7.79
GLNPO	2004	Aug	6	No observed hypoxia				
NWRI	2004	Aug	18-19	0.59	1.38	1.89	2.59	4.16
NWRI	2004	Aug	26	2.22	3.65	4.38	5.25	7.35
NWRI	2004	Sep	15-16	1.21	1.98	2.51	3.14	4.67
NWRI	2004	Sep	21-22	0.00	0.01	0.03	0.07	0.22
GLNPO	2005	Aug	9-10	No observed hypoxia				
NWRI	2005	Aug	9-10	No observed hypoxia				

Source	Year	Month	Date	Hypoxic extent (10^3 km^2)				
				2.5%	25%	50%	75%	97.5%
GLERL	2005	Aug	8-12	No observed hypoxia				
GLERL	2005	Aug	15-19	0.45	1.02	1.41	1.95	3.15
GLERL	2005	Aug	27-30	1.12	2.52	3.42	4.36	6.61
NWRI	2005	Sep	8-9	6.27	8.08	9.00	9.84	11.24
GLERL	2005	Sep	6-21	7.02	7.91	8.36	8.78	9.66
GLNPO	2006	Aug	10-12	1.07	1.94	2.43	3.05	4.26
GLNPO	2007	Aug	8	No observed hypoxia				
GLERL	2007	Sep	4-26	6.33	7.35	7.89	8.34	9.24

Table 5-3 Median (50th percentile), interquartile range (25th and 75th percentiles), and 95% confidence intervals (2.5th and 97.5th percentiles) of estimated hypoxic extent from all available GLNPO, NWRI and GLERL cruises for 1987-2007. This table presents the same information as that in Figure 5-2, but in tabular form.

The maximum hypoxic extent and its location in August and September (Figure 5-4, Figure 5-5) vary from year to year. It is most common in the western and middle northern portion of the central basin in August, spreading east in September, and thus there is a greater probability of larger hypoxic areas in September relative to August (Figure 5-6).

Reconstructions of the historical extent of August (Figure 5-4) and September (Figure 5-5) hypoxia for Cases 1 and 2 show that the spatial distribution of DO is consistent between the cases, despite the differences in auxiliary variables. This again supports the robustness of the presented approach.

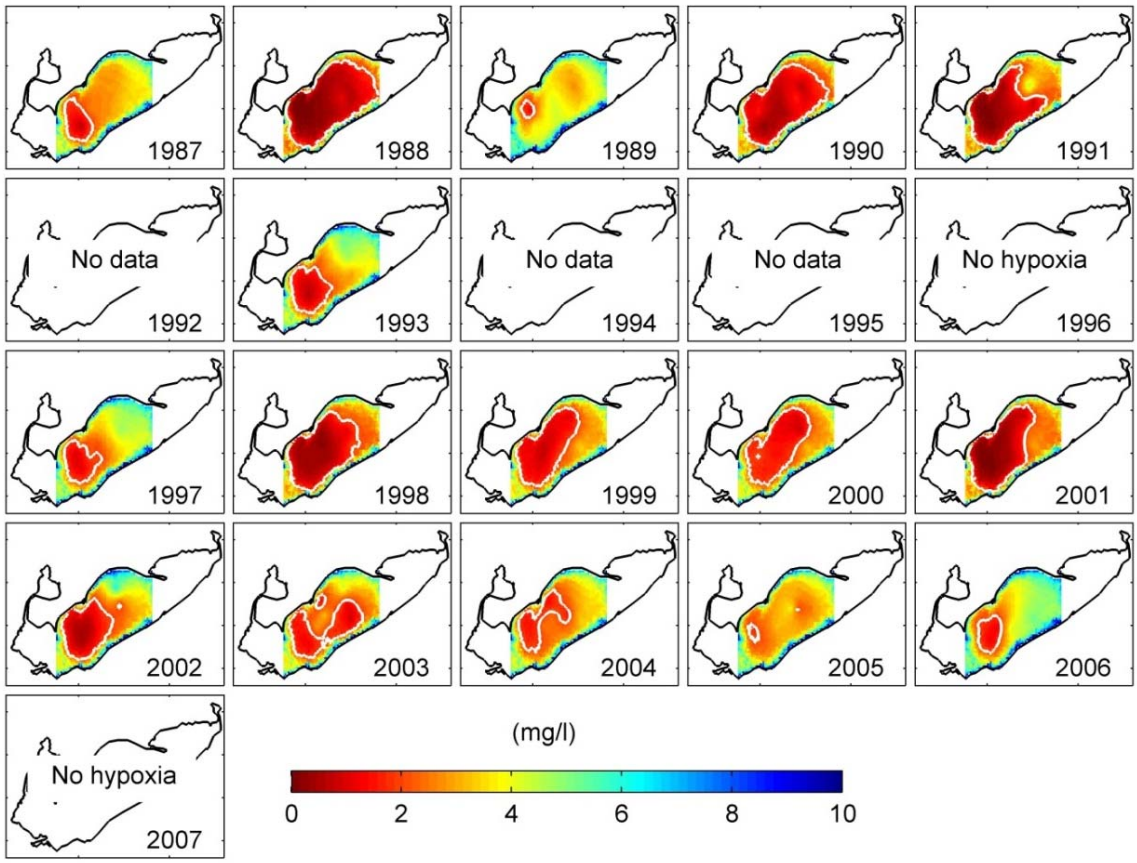

Figure 5-4 Estimated August DO concentrations using UK from the cruise with the observed maximum extent in each year. The hypoxic zone (DO < 2mg/L) is outlined in white.

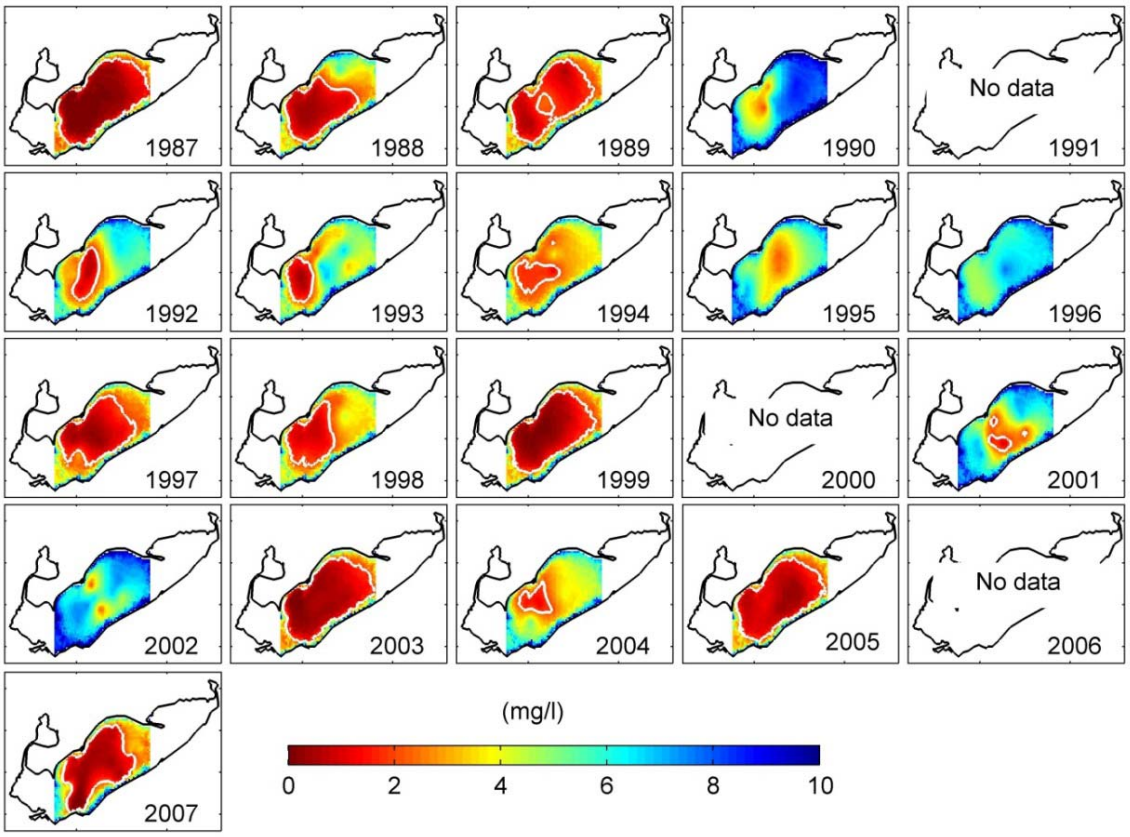

Figure 5-5 Estimated September DO concentrations using UK from the cruise with the observed maximum extent in each year. The hypoxic zone (DO < 2mg/L) is outlined in white.

Results show that hypoxia is most common in the western and middle northern portion of the central basin in August, and spreads to the central portion in September (Figure 5-6). Consistent with Figure 5-3a, Figure 5-6 also shows a larger area with a greater probability of hypoxia in September relative to August.

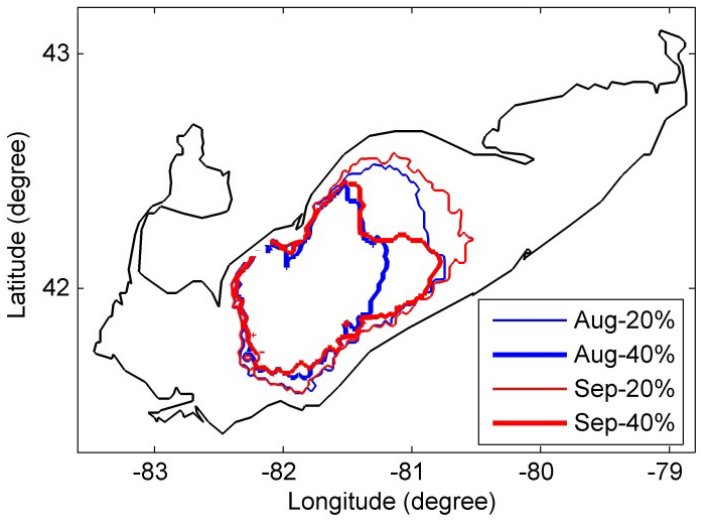

Figure 5-6 Areas with 20% and 40% probability of hypoxic conditions in August and September from 1987 to 2007, based on Case 1 estimated hypoxic areas (Figure 5-4 and Figure 5-5).

Annual estimates of maximum hypoxic extent are generally consistent with *Burns et al.* [2005], who provided the previously most complete analysis of Lake Erie hypoxia (Figure 5-7). Their estimates are based on a location-by-location selection of observation with the lowest DO throughout the year. One would expect that using the lowest measured concentration (especially from different times and different locations) would tend to overestimate the maximum extent of hypoxia. In addition, the lack of information from auxiliary variables makes it difficult to represent DO distributions given the very limited number of *in situ* DO observations, as seen by a comparison of UK and OK estimates (Figure 5-8, Figure 5-9). A quantitative or probabilistic comparison was not possible because *Burns et al.* [2005] did not include the actual values of hypoxic area or its uncertainty. Nonetheless, the estimates presented here are qualitatively consistent with *Burns et al.* [2005], supporting the notion that central basin hypoxia was more extensive in the late 1980s and late 1990s, compared to the early 1990s.

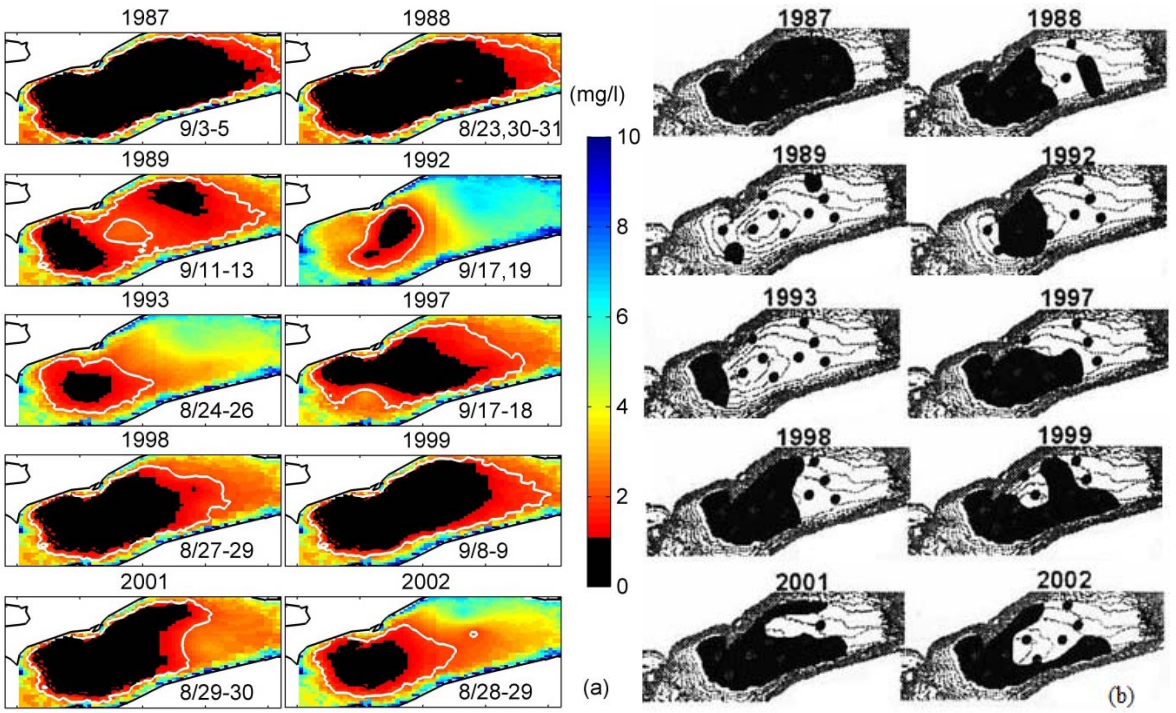

Figure 5-7 Comparison of (a) estimates from this study to (b) estimates from *Burns et al.* [2005]. In panel (a), white contours represent the 2 mg/L boundary, while areas with estimated concentrations below 1 mg/L are presented in black for consistency with *Burns et al.* [2005]. Subplots in (a) are based on the largest observed hypoxic zone for a given year (Figure 5-4 and Figure 5-5).

5.4.3 Validation of Universal Kriging and Ordinary Kriging

Figure 5-8 illustrates the benefit of including the auxiliary variables in UK relative to relying solely on the DO observations as is the case for OK. These sample maps, presented for September 1998, show the difference between OK results and UK results. Because areas close to eastern basin are generally not hypoxic, the hypoxic extent represented in UK (Figure 5-8b) is likely more representative of the actual DO distribution compared to that from OK (Figure 5-8a). Both methods have the lowest uncertainties near the measurement locations, but the uncertainties for OK are generally higher than those for UK away from the observed locations (Figure 5-8c and Figure 5-8d).

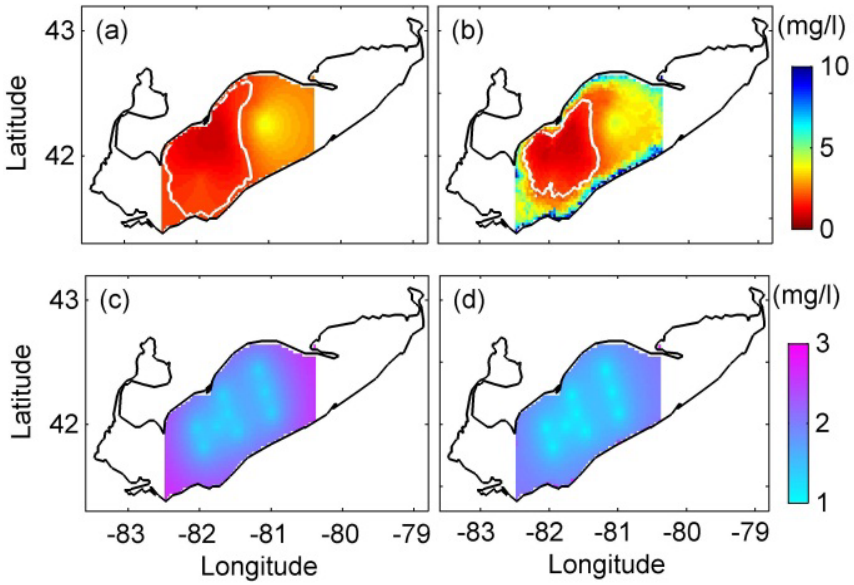

Figure 5-8 Estimated DO concentration for September 1998 using (a) OK and (b) UK; Kriging uncertainty (expressed as one standard deviation) using (c) OK and (d) UK. The white line defines the boundary of the hypoxic zone (DO<2mg/L).

The superior performance of UK was further confirmed through cross validation using data from 2005, when substantially more DO observations are available (Figure 5-9). To compare the methods, the DO measurements at the ten locations (Figure 5-9a) that are sampled in most years were used to predict DO throughout the central basin, and then compared the estimated DO distributions (Figure 5-9c for UK; Figure 5-9b for OK) with estimates obtained using the expanded data set of 63 observations (Figure 5-9d and Figure 5-9e). The shape of the hypoxic area derived from the limited set of ten observations using UK is relatively consistent with the more extensive observations (Figure 5-9d) and associated estimates (Figure 5-9e), whereas the hypoxic area predicted by OK using only the limited data set (Figure 5-9b) deviates from the GLERL observations and associated estimates. Overall, UK provides more reliable estimates, and with lower uncertainties, relative to OK.

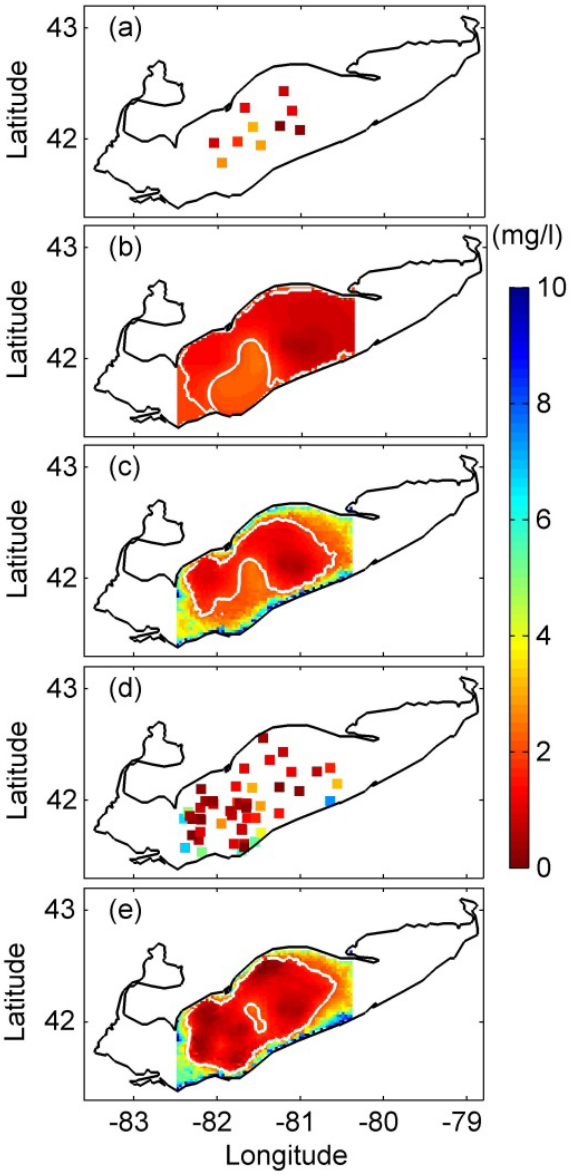

Figure 5-9 Validation results for September 2005. Best estimate of DO concentration obtained from OK (panel b) and UK (panel c), using only the 10 measurement locations available in most years (panel a). All available observations for September 2005 are presented in panel (d). The estimated hypoxic extent using all the available observations is presented in panel (e).

5.4.4 Relationship of areal extent to average bottom DO concentration and thermocline depth

The estimates of hypoxic extent presented above were based on a sequential application of BIC, UK, and conditional realizations. Based on these results, a simple exponential relationship was developed for predicting hypoxic extent using the square of the average measured DO from the ten regular sampling locations (Figure 5-1, Figure 5-10). This model explains 97% of the variability in these estimated hypoxic extents:

$$E = 9.30\exp(-DO_m^2/7.09) \qquad 5.18$$

where E is hypoxic extent (10^3 km^2) and DO_m is the mean of the DO concentration (mg/L) across the ten sites with regular observations (Figure 5-1). The two model parameters were estimated through a least-squares fit to the estimated hypoxic extent from all conditional realizations, and therefore account for the varying uncertainty of the estimates across individual cruises. The ten regular sampling locations could be used as index stations for obtaining estimates of hypoxic extent for times when detailed analyses such as the one presented are not done.

During the sampling cruises, vertical temperature profile data were also collected, allowing us to analyze our results relative to the thermal structure of the lake. The change in hypolimnion thickness (a correlate for thermocline depth) from August to September (Figure 5-3b) is an important predictor of seasonal change in hypoxic extent. For each month, the hypolimnion thickness was determined by averaging the measured hypolimnion thickness across monitoring stations for the cruise with the maximum observed hypoxic extent. Substantial deepening of the thermocline between August and September, as indicated by a decrease in hypolimnion thickness of more than two meters, is associated with early re-oxygenation of the basin, and a corresponding decrease in the hypoxic extent. This decrease in hypoxic extent is statistically

significant for four of the six such years for which DO data were available in both August and September (p~0 for 1990, 2001, 2002; p=0.05 for 1988; p=0.06 for 1998; p=0.18 for 1993). Conversely, a smaller change in the hypolimnion thickness is consistent with more stable stratification and an expansion of hypoxic extent between August and September. This expansion is significant for six of the seven such years (p~0 for 1987, 1989, 2007; p=0.01 for 1997, 2005; p=0.02 for 2003; p=0.13 for 1999). This finding illustrates the importance of timing and thermal structure on the size of the hypoxic zone; and these factors should be considered (along with biological drivers, e.g. nutrient stimulated productivity) when exploring the interannual variability of hypoxia in Lake Erie.

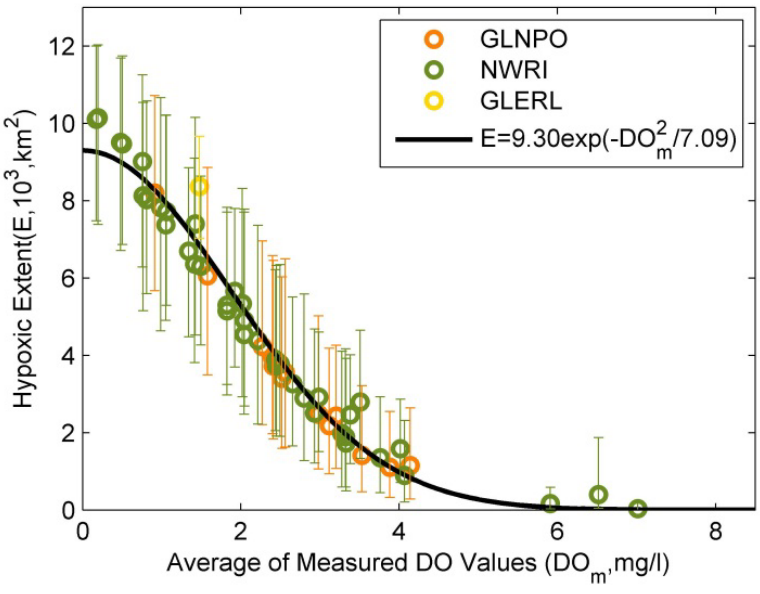

Figure 5-10 Predicted hypoxic extent based on average DO concentration from the ten index sampling locations and equation 5.18 (R^2=0.97). GLNPO, NWRI, and GLERL data are those from Figure 5-2, together with 95% confidence intervals.

5.5 Conclusions

Hypoxia occurs in the central basin of Lake Erie during most summers, but its spatial extent has been poorly understood due to sparse sampling. Universal Kriging and conditional realizations were used here to provide quantitative estimates of the extent of hypoxia in the central basin of Lake Erie for August and September of 1987 to 2007, and their associated uncertainties.

Bayesian Information Criterion was used to test the significance of auxiliary variables to the DO spatial distribution in the bottom of Lake Erie. Longitude was found to be positively correlated with DO, potentially acting as a proxy for phosphorus availability, because phosphorus primarily enters the central basin from the west. Bathymetry and its squared term were also shown to be correlated with DO distribution. The satellite observations of sea surface temperature and satellite chlorophyll were not found to significantly correlated with the spatial distribution of DO.

Consistent with previous findings (e.g., [*Hawley et al.*, 2006]), the hypoxic extent was generally found to have been lowest in the mid-1990s, with the late 1980s (1987, 1988) and the 2000s (2003, 2005) experiencing the largest hypoxic zones. The maximum hypoxic extent and its location in August and September vary from year to year. Hypoxia is most common in the western and middle northern portion of the central basin in August, spreading East in September, and thus there is a greater probability of larger hypoxic areas in September relative to August.

A simple exponential relationship based on the squared average measured bottom DO explains 97% of the estimated variability in the hypoxic extent. The ten regular sampling locations could therefore potentially be used as index stations for obtaining estimates of hypoxic extent for times when detailed analyses such as the one presented are not done.

The change in the observed maximum extent between August and September is found to be sensitive to the corresponding variability in the hypolimnion thickness. Substantial deepening of the thermocline between August and September, as indicated by a decrease in hypolimnion thickness of more than two meters, is associated with early re-oxygenation of the basin, and a corresponding decrease in the hypoxic extent.

CHAPTER 6 Temporal Variability of Hypoxic Volume in Chesapeake Bay

Hypoxia (low dissolved oxygen) was first reported in Chesapeake Bay in the 1930s, primarily due to nutrient loads and summertime stratification of the water column. The understanding of the seasonal and interannual variability of hypoxia is of great importance for proper water quality management and forecasting. The purpose of this chapter is to estimate the variability of hypoxic volume in Chesapeake Bay from 1985 to 2010 at a half-monthly temporal resolution, and to examine the effects of nutrient loads and weather patterns on this variability.

6.1 Introduction

Chesapeake Bay, the largest and the most productive estuary on the East Coast of the United States, is plagued with bottom-water hypoxia (dissolved oxygen, DO < 2mg/L) due to agricultural and industrial development and population growth along its shores and headwaters [*Cerco and Cole*, 1993]. While first reported in the 1930s [*Newcombe and Horne*, 1938; *Officer et al.*, 1984], hypoxia in Chesapeake Bay became more common and widespread in the late 1950s and early 1960s due to increased anthropogenic nutrient influx [*Cronin and Vann*, 2003].

It is widely understood that nitrogen loads and summertime stratification are two primary factors leading to hypoxia in coastal systems [*NSTC*, 2000]. Based on the fact that increases in nitrogen loading cause expansions in the hypoxic volume in Chesapeake Bay [*Flemer et al.*, 1983; *Hagy et al.*, 2004; *Liu and Scavia*, 2010], management efforts to improve water quality have focused on reducing nitrogen loading [*U. S. EPA*, 2002b]. Many hypoxia prediction models have thus been developed to guide management and policy in regulating nitrogen loading into the Bay (e.g., [*Cerco and Cole*, 1993; *Evans and Scavia*, 2011; *Hagy et al.*, 2004; *Kemp et al.*, 2005; *Murphy et al.*, 2011; *Scavia et al.*, 2006]). In addition to the nitrogen loading, vertical stratification is another primary factor leading to hypoxia formation in coastal systems, by

inhibiting the reaeration of bottom waters [*NSTC*, 2003]. However, the linear correlation between spring tributary nitrogen loading and summertime hypoxia is relatively low [*Hagy et al.*, 2004], and spring total nitrogen loading from major tributaries can only explain a small portion of the variability of summertime hypoxic volume [*Scully*, 2010a]. As a result, it remains uncertain how much reduction in hypoxia could be achieved from only restricting the total nitrogen loading from tributaries into Chesapeake Bay. A major source of this uncertainty may come from an inadequate understanding of the relationship between hypoxia and nitrogen loading. Thus, a better understanding of the controls of uncertainty in hypoxic volume estimates is needed.

Weather patterns (e.g., precipitation, wind) also play an important role in affecting both nitrogen loading and vertical stratification, which in turn impact hypoxic volume. For example, wind was shown to determine the concentration of DO in the subpycnocline layer (i.e., the bottom water layer) of Chesapeake Bay [*Malone et al.*, 1986; *Sanford et al.*, 1990]. Recent analyses suggest the duration of wind along particular directions, rather than the magnitude of wind speed, has a stronger impact on hypoxia [*Feng et al.*, 2012]. Previous studies suggested that the duration of westerly wind over Chesapeake Bay was correlated well with the summer hypoxic volume between 1950 and 2007 [*Scully*, 2010a; b]. However, *Murphy et al.* [2011]did not detect significant contributions from wind between 1985 and 2009. This discrepancy further adds to the need for a detailed investigation of how wind influences hypoxia.

Detailed analysis of the variability of hypoxic volume in Chesapeake Bay is thus needed to explore the effects of nitrogen loading and weather patterns. To map the detailed hypoxic conditions throughout the Bay, spatial interpolation methods such as simple linear interpolation, inverse distance weighting, and geostatistical methods have often been used [*Bahner*, 2006;

Murphy et al., 2011; *Rossi and Posa*, 1990]. However, because Chesapeake Bay is extremely narrow and shallow with variable depth, it creates some unique challenges for spatial interpolation.

This work applied geostatistical Universal Kriging and conditional realizations to estimate hypoxic volume and its associated uncertainty by using the auxiliary variables which are correlated with the DO distribution. The purposes of this work are (i) to estimate early-April to late-October hypoxic volumes and associated uncertainties for 1985 to 2010; (ii) to show the frequency and duration of hypoxia over these 26 years; and (iii) to explore the degree to which nitrogen loads and weather patterns explain the variability of hypoxia.

This chapter is organized as follows: Section 6.2 presents the data used in this study. Section 6.4 provides a brief overview of the inputs used in the geostatistical methods including the model parameter optimization. Section 6.5 presents the results of the analysis, including the seasonal and yearly variability of estimated hypoxic volumes, factors affecting variation in hypoxic volume from April to October for 1985 through 2010, and the comparison of the estimated hypoxic volume from this work with the results from two other models. The last section (Section 6.6) summarizes the main conclusions of the study.

6.2 Data description

6.2.1 Data for estimating hypoxic volume

Chesapeake Bay includes many sub-estuaries and tributaries (Figure 6-1), but the analysis presented here was restricted to the mainstem of the Bay. The *in situ* DO data were collected by the Chesapeake Bay Program and their collaborators, and were obtained from the Chesapeake Bay Program Water Quality Database [*Chesapeake Bay Program*, 2011]. This database included

measurements starting in July 1984, with stations sampled once each month during late fall and winter months and twice each month during the warmer months. The sampling cruises generally took less than a week for each. There are about forty fixed monitoring stations in the mainstem at which water quality parameters are measured at intervals of 1m to 2m through the water column (Figure 6-1). The DO data collected twice a month from the early April to late October were used to calculate hypoxic volumes in this work.

In addition to the *in situ* DO measurements, auxiliary variables with full spatial coverage were also included in the analysis. These variables include east-west location, north-south location, bathymetry (i.e., the total depth of each station), and measurement depth (i.e., the depth at which each measurement was conducted), which describe the physical characteristics of each estimation location in the Bay. The bathymetry data, part of the National Oceanic and Atmospheric Administration (NOAA) Coastal Geospatial Data Project, were averaged to 1 km × 1 km resolution from their initial 1 m × 1 m resolution. All horizontal locations in this work were georeferenced using the coordinate system North American Datum of 1983 (NAD83), Universal Transverse Mercator (UTM) Zone 18 North. Because stratification greatly affects hypoxia, as described in Section 6.1, these physical characteristics reflecting stratification are of great importance for estimating the DO spatial distribution. Chesapeake Bay stratification driven primarily by salinity differences varies spatially and decreases with distance toward the Bay mouth (close to Atlantic Ocean).

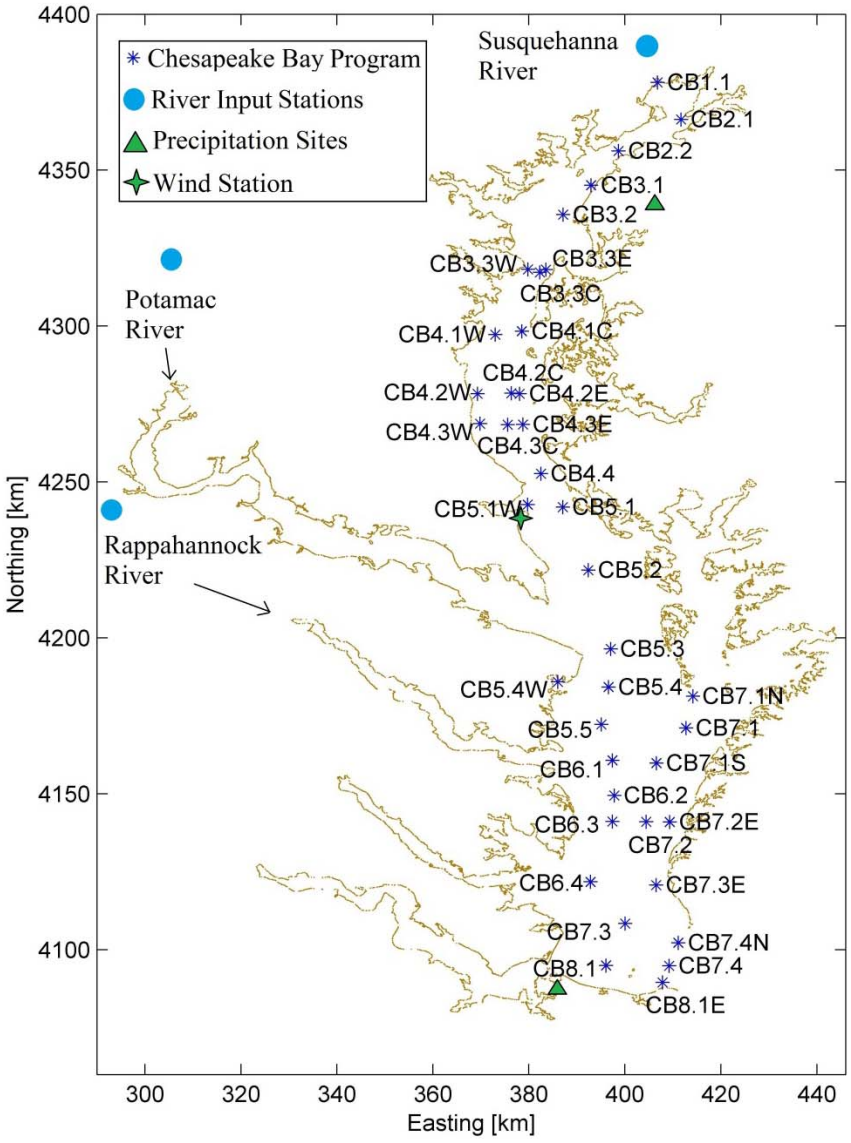

Figure 6-1 Locations of the Chesapeake Bay Program monitoring stations (blue stars) in the mainstem of Chesapeake Bay, river input monitoring stations (blue circles), precipitation monitoring sites (green triangles), and the wind monitoring station (green star).

6.2.2 Data for analyzing the estimated hypoxic volume

Nitrogen loads, precipitation, and wind were used to analyze the interannual variability in the hypoxic volumes. Monthly average total nitrogen (TN) loads data have been collected by the U.S. Geological Survey Chesapeake Bay River Input Monitoring Program [*USGS*, 2010] since

1981 for the main tributaries. Among these tributaries, Susquehanna, Potomac, James, and Rappahannock Rivers provide the largest nutrient loads to Chesapeake Bay. However, James River is located at the mouth of the Bay, the load of which is thus not considered in this work. Susquehanna, Potomac and Rappahannock all flow directly into the Bay, which together account for more than 90% of the TN of all tributary inputs (Table 6-1 and Figure 6-1). Therefore, the total spring nitrogen (TN) loads from these three tributaries from January to May were used in this work. Precipitation and wind data were obtained from the NOAA's National Climatic Data Center [*NCDC*, Accessed: August, 2012] and the National Weather Service [*NWS*, Accessed: August 2012]. The spring precipitation data from April to May were taken as the average of the measurements from Chestertown Station, Maryland and Norfolk Station, Virginia, which both provide long-term monitoring (Figure 6-1). Wind data are measured at Patuxent River Station, located close to the middle of the Bay (Figure 6-1). The wind data include average wind speed, and wind duration from different directions. The wind duration from a certain direction represents the length of time that wind blow over the Bay from this direction. The wind duration considered in this analysis was that from April to August, because the dominant wind direction changes after August according to the available wind data from 1985 to 2010.

Tributary	Station Name
Susquehanna	Susquehanna River near Conowingo, MD
Potomac	Potomac River at Chain Bridge, Washington DC
Rappahannock	Rappahannock River near Fredericksburg, VA

Table 6-1 Three main tributaries of Chesapeake Bay and their land use and wastewater discharge.

6.3 Models comparison for estimated hypoxic volume in Chesapeake Bay

The most expansive set of hypoxic volumes in Chesapeake Bay were estimated by the Chesapeake Bay Interpolation (CBI) Tool [*Bahner*, 2006] and Main-channel Ordinary Kriging (OK) [*Murphy et al.*, 2011], which both provide long term hypoxic volume estimation of the Bay. Including the model presented in this work, all three models have a similar timespan and the same temporal resolution (Table 6-2).

6.3.1 Chesapeake Bay Interpolator Tool (CBI)

Chesapeake Bay Interpolator Tool [*Bahner*, 2006] applied a three-dimensional inverse distance weighting (IDW) method to estimate the DO concentration for both the mainstem and tributaries in Chesapeake Bay. Only the hypoxic volume of the segments located in the mainstem was extracted for the comparison. IDW is commonly used and is easy to automate, but it only uses distance during interpolation without considering either spatial structure or estimation uncertainties. EPA recognized that geostatistical kriging could be a more accurate option and should be tested more for use in the Bay [*U. S. EPA*, 2007].

6.3.2 Main-channel Ordinary Kriging (Main-channel OK)

Murphy et al. [2011] kriged DO concentration only in the main channel (vertical and north-south direction) of the Bay, and then assumed the concentrations were constant laterally across the mainstem of the Bay to save extensively computational cost. By making this assumption, the near-shore zone is also hypoxic if the mainstem is hypoxic; however, it is unlikely to have hypoxia in the near-shore zone in reality. Therefore, the final results normally overestimate the actual hypoxic volume, especially in the summer. Unlike UK, OK does not use auxiliary variables. The spatial distribution of DO `in this work` was obtained only from the *in situ* measurements.

Model	Method	Spatial range	Spatial resolution	Time	Temporal resolution	Reference
CBI	Three-dimensional inverse distance weighting	Mainstem and tributaries	1km (east-west) × 1km (north-south) × 1m (vertical)	June-September 1985-2009	Half-month	*Bahner* [2006]
Main-channel OK	Two-dimensional OK	Mainstem	2km (north-south) × 1m (vertical)	May-September 1984-2009	Half-month	*Murphy et al.* [2011]
3D conditional realizations	Three-dimensional UK and conditional realizations	Mainstem	1km (east-west) × 1km (north-south) × 1m (vertical)	April-October 1985-2010	Half-month	This work

Table 6-2 Models of estimating spatial distribution of DO concentrations for comparison.

6.4 Methodology

The approach used to quantify hypoxia in this chapter is similar to the methods described in Chapter 5, but the geostatistical analysis presented here was expanded from two dimensions to three. The estimation timespan ranges from early April to late October (i.e., fourteen half-monthly intervals in total), covering the entire seasonal cycle of hypoxia for 1985 to 2010. The data for each half-month period were analyzed individually without assuming any conceptual temporal covariance among different periods or different years to avoid slightly unrealistic smoothing of extreme hypoxic volumes. The spatial resolution of DO concentrations in this work is 1 km (north-south direction) × 1 km (west-east direction) × 1 m (vertical direction).

A model selection method (i.e., Bayesian Information Criterion; see Section 6.4.4) was applied to identify which of the four auxiliary variables (i.e., east-west location [UTM X], north-

south location [UTM Y], bathymetry, and measurement depth) were significant variables in explaining the DO distribution in each half-month period (Table 6-3). UTM X and measurement depth were selected as significant variables for all time periods. UTM Y was also selected as a significant variable for most periods except early April, late September and October. Bathymetry was selected as a significant variable for eight time periods over the total fourteen periods. As the UTM Y and bathymetry are constant variables and the analyzed system is the same Bay, all four auxiliary variables were used for all time periods to keep the methods and analysis consistent.

	Early Apr	Late Apr	Early May	Late May	Early Jun	Late Jun	Early Jul	Late Jul	Early Aug	Late Aug	Early Sep	Late Sep	Early Oct	Late Oct
UTM X	*	*	*	*	*	*	*	*	*	*	*	*	*	*
UTM Y		*	*	*	*	*	*	*	*	*	*			
bathymetry		*	*		*	*		*	*			*	*	
measurement depth	*	*	*	*	*	*	*	*	*	*	*	*	*	*

Table 6-3 Model selection results for the DO distribution in each half-month period. Variables indicated with asterisks are the significant ones based on the model selection results.

The following Section 6.4.1 to Section 6.4.4 describe the methods used to estimate the spatial distribution of DO and the hypoxic volume for a given half-month period, using early July from 1985-2010 as an example. A detailed description of these methods is available in Chapter 5, while only the key steps are presented here.

6.4.1 Estimating DO spatial distribution using Universal Kriging

The preliminary steps of the UK setup are shown below. All 26-year DO data ($\mathbf{z}$, $n \times 1$) are organized as follows:

$$\mathbf{z} = \begin{bmatrix} \mathbf{z}_{1985} \\ \mathbf{z}_{1986} \\ \vdots \\ \mathbf{z}_{2010} \end{bmatrix} \quad 6.1$$

where $\mathbf{z}_i$ (i = 1985, ... , 2010) is an $n_i \times 1$ vector, and n_i is the number of measurements within a given half-month period (e.g., early July) during year i (i.e., $n = \sum_{i=1985}^{2010} n_i$).

Within the UK framework, the DO distribution was modeled as the sum of a deterministic term and a zero-mean spatially correlated residual term. The deterministic term represents the portion of the DO distribution that can be explained by the categorical variables (ones and zeros) and auxiliary available, and the residual term represents the remaining portion of the observed variability. The categorical variables represent the spatially-constant yearly offsets (i.e., intercepts) corresponding to each year. Therefore, the measurement data ($\mathbf{z}$) could also be written as:

$$\mathbf{z} = \mathbf{X_z}\boldsymbol{\beta} + \mathbf{z}_{res} \tag{6.2}$$

where $\mathbf{X_z}$ is a known $n \times (26+4)$ matrix of categorical variables and auxiliary variables that explain a portion of the DO variability in space, $\boldsymbol{\beta}$ is a $(26+4) \times 1$ vector of unknown regression coefficients corresponding to these variables, 26 is the number of categorical variables (i.e., the number of examined years), 4 is the number of auxiliary variables (i.e., latitude, longitude, measurement depth, bathymetry), and $\mathbf{z}_{res}$ is an $n \times 1$ vector of residuals. Note that the $\boldsymbol{\beta}$ values are different for each half-month period. Overall, the deterministic term ($\mathbf{X_z}$, $n \times (26+4)$) is expressed as:

$$\mathbf{X_z} = \begin{bmatrix} \mathbf{1}_{1985} & \cdots & \emptyset & \mathbf{X}_{1985} \\ \vdots & \ddots & \vdots & \vdots \\ \emptyset & \cdots & \mathbf{1}_{2010} & \mathbf{X}_{2010} \end{bmatrix} \tag{6.3}$$

where $\mathbf{1}_i$ (i = 1985, ... , 2010) is an $n_i \times 1$ vector of ones composing categorical variables, and $\mathbf{X}_i$ (i = 1985, ... , 2010) is an $n_i \times 4$ matrix of auxiliary variables.

For the residual term ($\mathbf{z}_{res}$), an $n \times n$ generalized covariance matrix ($\mathbf{Q_{zz}}$) is used to represent its spatial structure:

$$\mathbf{Q_{zz}} = \begin{bmatrix} \mathbf{Q}_{1985} & \cdots & \emptyset \\ \vdots & \ddots & \vdots \\ \emptyset & \cdots & \mathbf{Q}_{2010} \end{bmatrix} \tag{6.4}$$

where $\mathbf{Q}_i$ (i = 1985, ... , 2010) is a covariance matrix for the residuals within a given half-month period (e.g., early July) of each year. All $\mathbf{Q}_i$'s use the same covariance parameters (σ^2, $\sigma^2_{\mathbf{Q}}$, and l):

$$Q_i(h) = \begin{cases} \sigma^2 + \sigma^2_{\mathbf{Q}}; & h = 0 \\ \sigma^2 \exp\left(-\frac{h}{l}\right); & otherwise \end{cases} \tag{6.5}$$

where h is the separation distance between measurement locations, σ^2 is the variance of the portion of the residual DO variability that is spatially correlated, 3l is the practical correlation range, and $\sigma^2_{\mathbf{Q}}$ is the measurement error.

The covariance matrix is used in the UK system of linear equations:

$$\begin{bmatrix} \mathbf{Q_{zz}} & \mathbf{X_z} \\ \mathbf{X_z^T} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{\Lambda}^T \\ \mathbf{M} \end{bmatrix} = \begin{bmatrix} \mathbf{Q_{zs}} \\ \mathbf{X_s^T} \end{bmatrix} \tag{6.6}$$

where $\mathbf{Q_{zs}}$ is an $n \times m$ covariance matrix between the measurement and estimation locations obtained analogously to equation 6.4, m is the number of estimation locations, the $m \times (26+4)$ matrix $\mathbf{X_s}$ contains the same four auxiliary variables as in $\mathbf{X_z}$ but defined at the estimation locations, and T denotes a matrix transposition. The system of equation 6.6 is solved for $\mathbf{\Lambda}$, which is an $m \times n$ matrix of weights assigned to each observation for each estimation location, and for $\mathbf{M}$, which is a $(26+4) \times m$ matrix of Lagrange multipliers. Finally, $\mathbf{\Lambda}$ and $\mathbf{M}$ are used to obtain estimates of the DO distribution, and their associated uncertainties:

$$\hat{\mathbf{s}} = \mathbf{\Lambda z} \tag{6.7}$$

where $\hat{\mathbf{s}}$ is an $m \times 1$ vector of the final DO estimates.

6.4.2 Obtain spatial structure of DO using Restricted Maximum Likelihood

The Restricted Maximum Likelihood (REML) approach maximizes the likelihood of available observations after marginalizing with respect to the unknown regression coefficients (**β**) [*Snodgrass and Kitanidis*, 1997; *Zhou and Michalak*, 2009]. This is equivalent to minimizing the following cost function to obtain the structural parameters in $\mathbf{Q_{zz}}$:

$$L = \frac{1}{2} ln|\mathbf{Q_{zz}}| + \frac{1}{2} ln|\mathbf{X_z^T Q_{zz}^{-1} X_z}| + \frac{1}{2}\mathbf{z}^T \mathbf{\Xi z} \tag{6.8}$$

$$\mathbf{\Xi} = \mathbf{Q_{zz}^{-1}} - \mathbf{Q_{zz}^{-1} X_z (X_z^T Q_{zz}^{-1} X_z)^{-1} X_z^T Q_{zz}^{-1}} \tag{6.9}$$

where | | denotes matrix determinant.

Because of the stratification between surface water and bottom water in Chesapeake Bay from late spring to late summer, the detrended DO (i.e., $\mathbf{z}_{res}$) along the vertical direction still shows inconsistent spatial distribution between the surface and sub-pycnocline layers. Due to the discontinuous spatial patterns of DO along vertical directions, REML is not able to accurately estimate the covariance parameters (i.e., REML has difficulty finding the local minimum of the likelihood values). To solve this problem, covariance parameters in the horizontal and vertical directions were obtained separately, and were then combined. These covariance parameters are different for each half-month period (Table 6-4). Taking early July for example, equation 6.10 shows its covariance matrix ($Q_i(h)$, unit: [mg/L]^2):

$$Q_i(h) = \begin{cases} 3.4 + 0.6; & h_{ew} = 0, h_{ns} = 0, h_v = 0 \\ 3.4\exp\left(-\sqrt{\left(\frac{h_{ew}}{26.0}\right)^2 + \left(\frac{h_{ns}}{98.4}\right)^2 + \left(\frac{h_v}{13.3}\right)^2}\right); & otherwise \end{cases} \tag{6.10}$$

where h_{ew}, h_{ns} and h_v are the separation distances along east-west, north-south and vertical directions, respectively. This equation suggests that the DO data are spatially correlated within

3*26.0 km ($3l_{ew}$) in east-west direction, within 3*98.4 km ($3l_{ns}$) in north-south direction, and within 3*13.3 m ($3l_v$) in vertical direction. As expected, the variability of DO concentration is smoothest (i.e., the correlation length is longest) along the north-south direction because the surface fresh water flows from north to south, and the bottom sea water flows from south to north.

Period	Sill (σ^2, $[mg/L]^2$)	Range parameter along E-W direction (l_{ew}, km)	Range parameter along N-S direction (l_{ns}, km)	Range parameter along vertical direction (l_v, m)	Nugget ($\boldsymbol{\sigma_Q^2}$, $[mg/L]^2$)
Early April	1.8	36.9	87.8	21.2	0.4
Late April	2.5	25.4	79.4	15.7	0.2
Early May	2.7	19.9	82.9	13.0	0.3
Late May	3.3	20.7	79.5	13.2	0.3
Early June	2.8	28.7	95.6	12.1	0.6
Late June	2.8	17.5	61.4	10.6	0.3
Early July	3.4	26.0	100.0	13.3	0.6
Late July	3.0	25.0	73.7	12.4	0.7
Early August	2.9	21.0	69.8	12.6	0.5
Late August	2.4	19.3	77.3	12.6	0.6
Early September	2.0	15.8	79.2	13.5	0.5
Late September	1.6	27.5	58.4	18.2	0.5
Early October	1.3	18.2	51.9	19.5	0.2
Late October	1.1	25.4	59.0	18.2	0.2

Table 6-4 Covariance (**Q**) parameters of detrended DO for each half-month period (e.g., equation 6.10 for early July).

6.4.3 Estimating hypoxic volumes and associated uncertainties using conditional realizations

UK provides a direct, quantitative assessment of the DO concentration; including an assessment of the uncertainty associated with the estimated DO concentrations. However, UK cannot be used directly to estimate the uncertainty associated with the hypoxic volume (i.e. the total volume for which DO concentration is below the 2 mg/L threshold). To provide a probabilistic estimate of the hypoxic volume uncertainty for each year, conditional realizations of the DO distribution were generated [*Gutjahr et al.*, 1994; *Kitanidis*, 1995; *Zhou and Michalak*, 2009]. Each conditional realization is defined as:

$$\mathbf{s}_{ci} = \mathbf{\Lambda}(\mathbf{z} - \mathbf{z}_{ui}) + \mathbf{s}_{ui} \tag{6.11}$$

where $\mathbf{\Lambda}$ is the $m \times n$ matrix of weights defined in equation 6.6, and $\mathbf{z}_{ui}$ and $\mathbf{s}_{ui}$ are unconditional realizations at measurement locations and estimation locations, respectively, obtained from:

$$\begin{bmatrix}\mathbf{z}_{ui}\\ \mathbf{s}_{ui}\end{bmatrix} = \mathbf{C}^T\mathbf{u} \tag{6.12}$$

where $\mathbf{u}$ is an $(n+m) \times 1$ vector of normally distributed random values with zero mean and unit variance (note that a new vector $\mathbf{u}$ is regenerated for each realization), and $\mathbf{C}$ is the $(n+m) \times (n+m)$ matrix resulting from the Cholesky decomposition of the covariance matrix below:

$$\begin{bmatrix}\mathbf{Q_{zz}} & \mathbf{Q_{zs}}\\ \mathbf{Q_{zs}^T} & \mathbf{Q_{ss}}\end{bmatrix} = \mathbf{C}\mathbf{C}^T \tag{6.13}$$

Conditional realizations were only generated for locations that are deeper than 2m because (i) scientifically, the pycnocline depth is normally around 10m [*Olson and Shenk*, 2003], and (ii) computationally, it saves extensive computational cost. The hypoxic volume was calculated for each conditional realization by summing the volumes where the predicted DO concentration is below 2 mg/L. A thousand conditional realizations were generated for each half-month for which

hypoxia was estimated, and the results were used to develop a probabilistic estimate of the hypoxic volume.

6.4.4 Auxiliary variable selection

Bayesian Information Criterion (BIC) is based on the Bayesian factor or the posterior probability of a model, and considers both the goodness of fit and the number of variables in the model [*Anderson et al.*, 1998; *Schwarz*, 1978]. The implementation of BIC was discussed in the methodology section of Chapter 5. BIC was evaluated for each possible subset of auxiliary variables, and the set of variables with the lowest BIC was identified as the best model. BIC was used several times in this work. As described in beginning of Section 6.4, BIC was used to identify the significance of latitude, longitude, bathymetry, and measured depth to the DO spatial distribution in the mainstem of the Bay. In addition, BIC was also used multiple times to select a subset of variables that can reliably represent the temporal variability of the hypoxic volume from 1985 to 2010.

6.5 Results and discussions

This section first shows the variability of hypoxia at half-monthly intervals and a summary of the frequency of occurrence of hypoxia by location in the Bay between 1985 and 2010, followed by an examination of the impact of nutrient loading and weather patterns on the hypoxic volume. The final part of this section compares estimated hypoxic volumes from conditional realizations with those from other two models.

6.5.1 Hypoxic frequency from 1985 to 2010

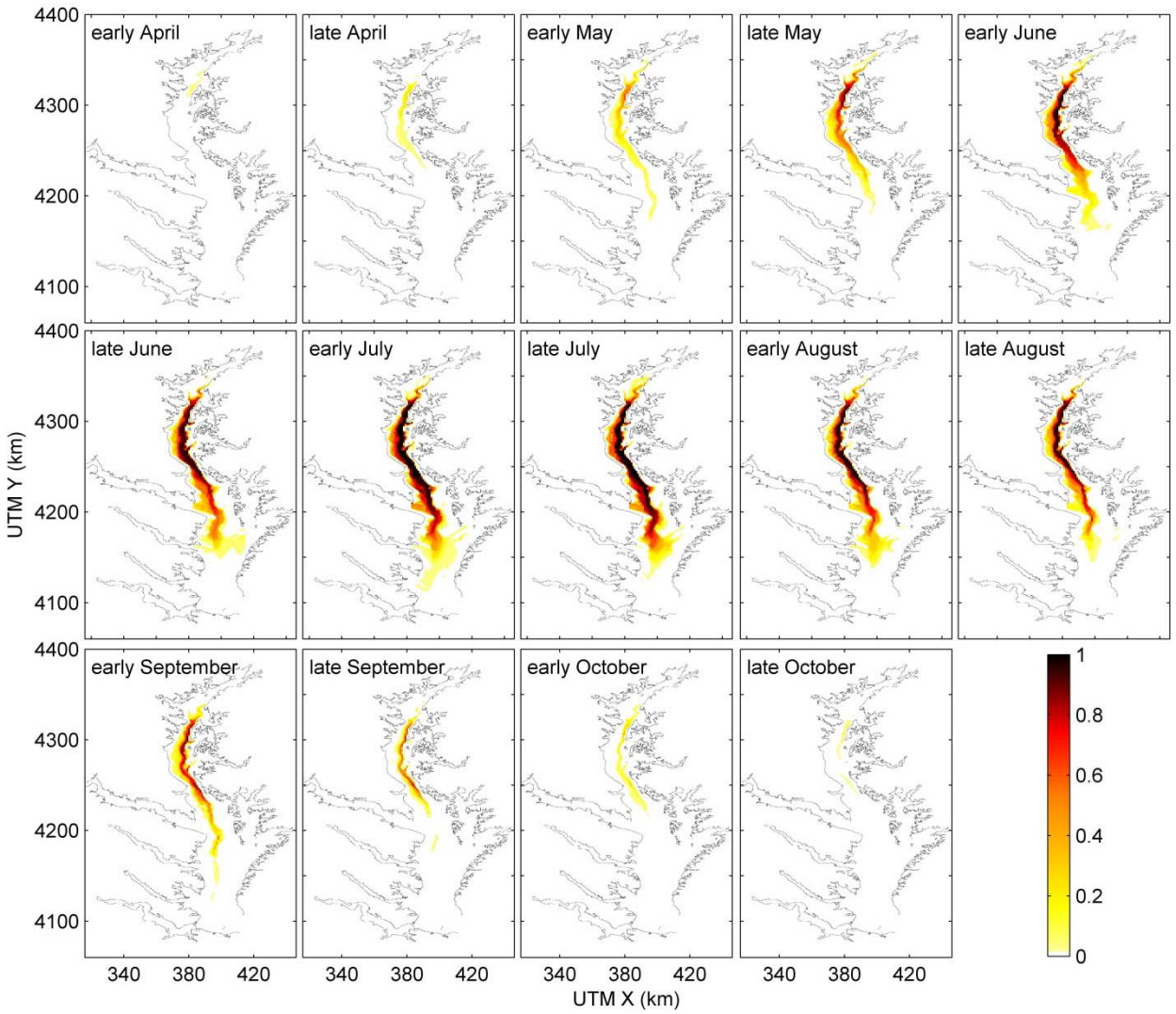

Figure 6-2 Time frequency of hypoxia (from UK) in horizontal direction of Chesapeake Bay for every half-month from 1985 to 2010.

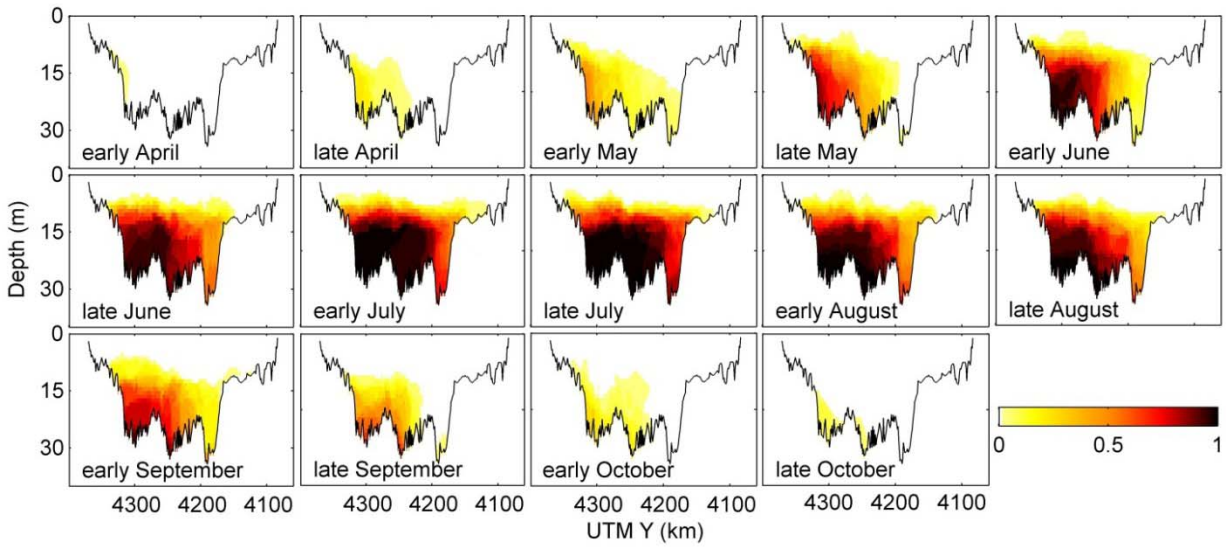

Figure 6-3 Time frequency of hypoxia (from UK) in vertical direction of Chesapeake Bay for every half-month from 1985 to 2010.

The fraction of years for which hypoxic conditions occur at a given location for each half-month period was estimated using the best estimates of DO from 1985 to 2010 (equation 6.7). Hypoxia was most frequent at depths greater than 10m to 15m of the mainstem (i.e., along the main/central channel) during the 26-year period under consideration (Figure 6-2 and Figure 6-3). This is expected because the central channel in the upper Bay and middle Bay is the deepest channel with strongest stratification, and its observed pycnocline depth is usually located around 10m [*Olson and Shenk*, 2003]. The Bay is narrow along the east-west direction, thus the volume of water below the pycnocline is small enough that oxygen is depleted quickly. The near-shore zone and lower Bay zone suggest lower probabilities of hypoxia because they are not sufficiently deep to develop stratification.

Differences in hypoxic frequency in Chesapeake Bay are suggested even at a half-monthly resolution (Figure 6-2 and Figure 6-3). Hypoxia appears earliest in the upper Bay in April,

although the low probabilities imply that hypoxia may only appear in this month in a minority of given years. Hypoxia appears first in the upper Bay because the primary input of nitrogen stems from the Susquehanna River, located at the head of the Bay [*Evans and Scavia*, 2011], which accounts for about 64% of the total tributary TN loads in spring [*Murphy et al.*, 2011]. In addition, during this time of year stratification starts to form in the Bay. The melting snow and precipitation increase the freshwater flow in the surface of Chesapeake Bay, which results in a density gradient along the vertical direction within the Bay. In June, hypoxia begins to spread south, to the middle and partial lower Bay zones. By July, hypoxic probabilities and extent have peaked; starting in late August hypoxia starts to disappear, first from the lower, then the middle, and finally the upper Bay. This decline is due to cooling surface waters that eventually sink disturbing the stable stratification and re-oxygenating the bottom water. By October, hypoxia has disappeared entirely from the Bay, with only extremely low possibilities of hypoxia in the upper Bay remaining.

6.5.2 Half-monthly estimated hypoxic volumes from 1985 to 2010

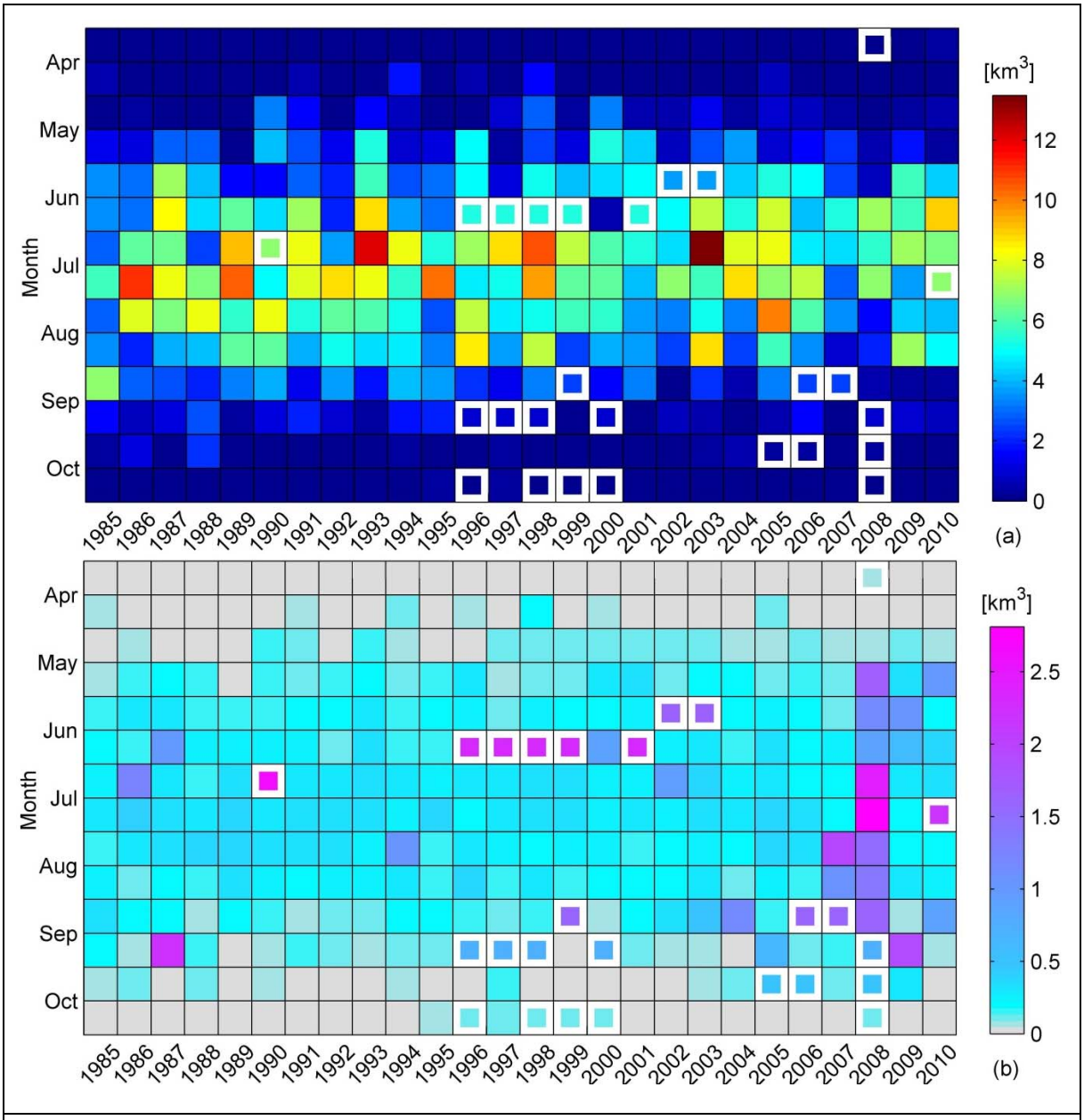

Figure 6-4 Hypoxic volumes (a) and associated standard errors (b) of Chesapeake Bay from 3D conditional realizations from April to October 1985-2010. The cells with white borders represent the time periods for which DO measurements were not available, and these values represent the average volumes and standard error for each half-month.

The estimated half-monthly hypoxic volumes and their associated uncertainties, derived from conditional realizations, are shown in the solid cells of Figure 6-4. Some hypoxic volumes and uncertainties represented in the cells using white borders were not estimated directly from

observations, because the DO measurements are not available for those periods. The hypoxic volume denoted by the cells with white borders was obtained by averaging the hypoxic volumes shown in solid cells during the same half-month when DO measurements are available. The uncertainty denoted in the cells with white borders of each half-month was calculated as the standard deviation of the hypoxic volumes from conditional realizations during the same half-month.

Hypoxic volumes exhibit a strong seasonal cycle, with the maximum volumes consistently found in July or August from 1985 to 2010 (Figure 6-4a). The summer months (i.e., June to August) with large hypoxic volumes have considerable volumes with estimated DO concentrations close to the 2mg/L hypoxic threshold, leading to a slightly larger uncertainty on the exact volume that is hypoxic (Figure 6-4b). However, the spring and fall months with small hypoxic volumes have considerable volumes with DO concentrations mostly above this hypoxic threshold, which results in smaller uncertainty of hypoxic volume. During the 26-year period, the largest volume of oxygen-depleted water appeared in early July 2003 due to heavy rain that spring, which brought intensive nutrient loading into the Bay within a short time [*Lewis et al.*, 2007]. However, hypoxia was greatly alleviated by late July due to storm-induced mixing in the water [*Lewis et al.*, 2007]. This implies that the hypoxic volumes in a half-monthly resolution can be greatly affected by short-term extreme events.

6.5.3 Selection of the auxiliary variables to hypoxic volume variability

Several variables were included in the analysis to examine their significance in explaining the interannual variability of hypoxic volume from 1985 to 2010. These variables include total spring nitrogen (TN) loads, mean spring precipitation, average wind speed, and directional wind duration.

The TN loads used in this work were those from the Susquehanna, Potomac and Rappahannock Rivers. Loading from the Susquehanna River is of particular interest in most research because this river contributes a large fraction of nitrogen loading to the system and is among the major tributaries in flowing directly into the mainstem of Chesapeake Bay [*Hagy et al.*, 2004]. However, *Murphy et al.* [2011] used TN loading from both the Susquehanna and Potomac Rivers to explain the variability of hypoxic volume. Although TN loading from the Rappahannock was ignored in accounting for the total loads into the Chesapeake Bay in previous research (e.g., [*Murphy et al.*, 2011]), this work shows that including the loading from Rappahannock yields an improved explanation of the variability of hypoxia (to be discussed in Section 6.5.1) on the order of 1% (i.e., R^2 decreased 0.01). This slight improvement makes sense because the loading from the Rappahannock, though not negligible, is much smaller than the other two tributaries.

BIC was used to test the significance of wind speed in this work. *Hunter et al.* [2008] suggested that only wind above a certain speed threshold could affect the stratification or the bloom through the water column. Taking Lake Taihu (in China) with an area of 2,250 km² and an average depth of 2 m for example, 4 m/s was indicated as the wind speed threshold capable of inducing turbulent mixing [*Cao et al.*, 2006]. For Lake Erie, 7.7 m/g was used as the effective wind speed threshold for affecting the stratification [*Wynne et al.*, 2010]. As the dimensions of Chesapeake Bay are intermediate to those of Lake Taihu and Lake Erie, 6 m/s was chosen for this work as the criterion for affecting the wind-induced energy propagating into the bottom layer of the Bay. Therefore, two kinds of monthly wind data were tested in this work, which were calculated (i) from hourly wind data with speeds greater than 6 m/s and (ii) from all the hourly wind data, respectively. However, the model selection results shows that all wind data (BIC = -

73.0, $R^2 = 0.85$) better explain the hypoxic variability than the wind data above 6m/s (BIC = -66.2, $R^2 = 0.83$). Consequently, only the results using all wind data are shown in the following analysis.

BIC was also used to test the significance of wind duration along different directions in explaining the variability of hypoxia from 1985 to 2010, together with TN loading data and precipitation data. Wind duration from the four cardinal directions (N, E, S, W) was used first, and followed by that from the four intercardinal directions (SW, NE, SE, NW). Preliminary BIC results suggest that the SW wind duration and NE wind duration were shown as the most and the second-most significant wind factors, respectively. NE wind duration and SW wind duration were positively and negatively correlated with hypoxia (refer to Section 6.5.4 for details). To further explore whether a single parameter representing both durations could be a better substitute, the ratio between them (i.e., [SW wind duration] / [NE wind duration]) was selected for the best model according to BIC. Therefore, a more intuitive term (i.e., dominant wind effect), calculated as [SW wind duration] / [NE wind duration] minus 1, was used in this work, which measures the relative effect of the wind duration from these two directions. If the SW wind is dominant, the dominant wind effect should be negatively correlated with hypoxia. However, if the NE wind is dominant, the dominant wind effect should be positively correlated with hypoxia.

6.5.4 Relationship of auxiliary variables to the variability of hypoxic volume

Hypoxia is a complicated process involving many physical and biochemical reactions, changing every half-month (Figure 6-4a). Weather events can affect nutrient loading and mixing in the short term (e.g., early July 2003). To avoid the short-term effect caused by extreme events, this work considers hypoxic conditions throughout the entire year in the analysis for Chesapeake

Bay. For example, the summer hypoxia of 1996, a year with one of the highest nitrogen loads, was less extensive than that of 1995 [*Hagy et al.*, 2004]. The research presented here reveals that although the July or early July hypoxic volume in 1995 is much larger than that in 1996, the mean hypoxic volume from April to October for 1995 representing the overall yearly hypoxic condition is actually smaller than 1996 (Figure 6-4a).

To account for the overall hypoxic condition of each year, this work uses mean hypoxic volume and hypoxic volume intensity of each year. Mean hypoxic volume represents the average half-monthly hypoxic volume from April to October. Hypoxic volume intensity is calculated as the sum of each half-monthly hypoxic volume from April to October multiplied by fifteen days (i.e., a half-month). These two indices are proportional to each other, though their values are not identical. Since there is no hypoxia before April or after October each year, the mean hypoxic volume and hypoxic volume intensity in this work can be taken to reflect the total hypoxic condition for each year.

From 1985 to 2010, spring mean precipitation (**P** [26×1], cm), spring TN loads (**N** [26×1], 10^6 kg/day), and a variable representing the dominant wind effect (**W** [26×1], i.e., [SW wind duration] / [NE wind duration] minus 1) were selected by BIC and explained 85% of the variability in the mean hypoxic volume from April to October. This best model selected is:

$$\mathbf{V} = 1.56 + 3.16 \cdot \mathbf{N} + 0.06 \cdot \mathbf{P} - 0.38 \cdot \mathbf{W} \qquad 6.14$$

where **V** [km^3] is a 26×1 vector representing the mean hypoxic volumes from 1985 to 2010. The correlation coefficients of the variables were shown in Table 6-5. Among the three variables, the importance of each variable can be isolated, because any pairs of these variables were not well correlated.

	TN	Precipitation	Dominant wind effect
TN	1	0.36	0.20
Precipitation		1	0.03
Dominant wind effect			1

Table 6-5 Correlation coefficient for each pair of the three coefficients.

The contribution of these three variables and the constant is shown in a stacked bar plot (Figure 6-5). From 1985 to 2010, TN contributes 25.7-61.3% of mean hypoxic volume with a mean of 37%. Precipitation contributes10.1-24.2% of the mean hypoxic volume with a mean of 17.3%. Dominant wind effect contributes 0.1-27.3% of the mean hypoxic volume with a mean of 8.8%. The 95% confidence intervals associated with these mean hypoxic volume estimates are the approximate uncertainties, representing both the spatial uncertainties of each time period and temporal interpolation uncertainties of the periods without available DO measurements (Figure 6-4b). The yearly uncertainty of mean hypoxic volume was the square root of the average variance (i.e., square of the uncertainty in Figure 6-4b) of all the half-monthly hypoxic volumes from April to October during the same year. Note that 2007 and 2008 are the years with the fewest DO measurements in space during the summer months. Their extremely sparse sampling results in large uncertainties associated with the estimated hypoxic volumes (Figure 6-4), and therefore the estimated volumes of these two years were not included in this analysis.

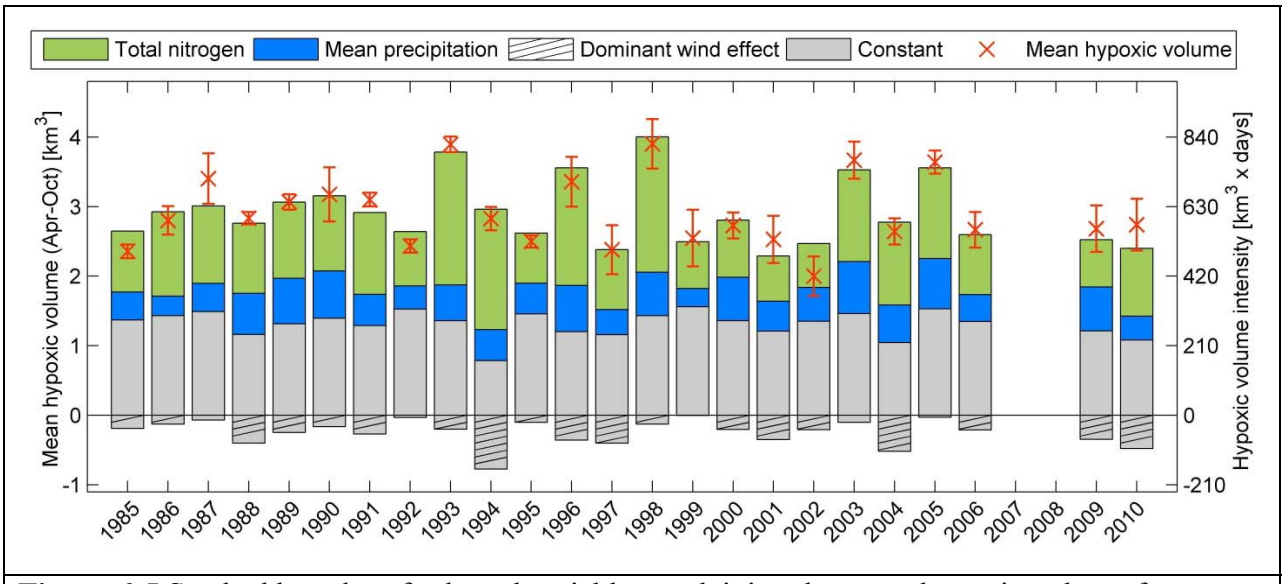

Figure 6-5 Stacked bar plot of selected variables explaining the mean hypoxic volume from April to October (or hypoxic volume intensity) during hypoxic season. Dominant wind effect is calculated as the [SW wind duration] / [NE wind duration] minus 1. Red cross and red line represent the mean and 95% confidence interval of hypoxic volume obtained from Figure 6-4.

	R^2 of each variable (R^2_{ind})	R^2 for all variables in the best model (R^2_{all})	R^2 when each variable is removed from the best model (R^2_{rem})
TN	0.64	0.85	0.36
Precipitation	0.31		0.78
Dominant wind effect	0.05		0.73

Table 6-6 Coefficient of determination (R^2) of each individual variable and combination of these variables to the variability of mean hypoxic volumes.

The positive correlation between TN loads and hypoxic volume makes intuitive sense, because greater nitrogen loads will result in more consumption of DO in the bottom of the Bay. TN loads from the Susquehanna, Potomac and Rappahannock were shown to explain more variability of the mean hypoxic volume than the other two selected variables for all the examined

years (Table 6-6). TN alone explains 64% variability of mean hypoxic volume, and the prediction capability of the model decreased by 49% (i.e., R^2 decreased from 0.85 to 0.36) without taking TN into account. This confirms that TN loading is the primary contributing factor to the variability of hypoxia in Chesapeake Bay from 1985 to 2010.

This work shows for the first time that spring precipitation is positively correlated with the mean hypoxic volume from April to October, and the impact of precipitation was not examined in previous studies. This positive correlation is expected because precipitation is generally assumed to influence nitrogen loads by increasing river flow or soil erosion. Precipitation alone explains 31% variability of mean hypoxic volume, and the prediction capability of the model decreases 7% without precipitation (Table 6-6). This result confirms the importance of precipitation in rendering an accurate prediction of hypoxic volume.

The precipitation appears to represent nonpoint sources of nitrogen downstream from the tributary monitoring sites. There are four main sources of nitrogen input into the Bay: (i) nitrogen from tributary monitoring (i.e., the TN loading used in this work), (ii) nonpoint sources downstream from the tributary monitoring sites, (iii) nitrogen from wastewater treatment plants downstream from the tributary monitoring sites, and (iv) atmospheric deposition of nitrogen to tidal water (Table 6-7, [*USGS*, 2011]). Due to data availability and measurement restrictions, only the nitrogen loads from the tributary monitoring sites were included in this work. As the nitrogen loads from tributary monitoring sites were measured in the upstream of the tributaries (Figure 6-1), the nonpoint sources downstream from the tributary monitoring sites that could be affected by precipitation were not included in the TN loading used in this work. The TN loading from tributary monitoring sites represent nearly twice the loading from nonpoint sources downstream from these monitoring sites (Table 6-7), which is analogous to the relationship

between the contribution of monitored TN (37%) and the contribution of mean precipitation (17.3%) to the variability of mean hypoxic volume. This indicates that even though the nonpoint sources of nitrogen downstream from the tributary monitoring sites were not available to this work, its contribution is probably captured by that attributed to precipitation.

Source of loading	Percent of total
Tributary monitoring sites	58%
Nonpoint sources downstream from the tributary monitoring sites	23%
Wastewater treatment plants downstream from the tributary monitoring sites	13%
Atmospheric deposition of nitrogen into tidal water	6%

Table 6-7 Nitrogen sources in Chesapeake Bay, [*USGS*, 2011].

The dominant wind factor (i.e., [SW wind duration] / [NE wind duration] minus 1) alone explains 5% of the variability of mean hypoxic volume. In addition, the prediction capability of the model decreases 12% (i.e., R^2 reduced from 0.85 to 0.73) without this wind factor (Table 6-6). The significance of wind in this context is expected because wind affects hypoxia by influencing the stratification in the Bay.

NE wind duration and SW wind duration were positively and negatively correlated, respectively, with hypoxic volume. These results are consistent with the findings by *Valle-Levinson et al.* [1998] and *Guo and Valle-Levinson* [2008], who demonstrated that the northeasterly winds could enhance stratification in the lower part of Chesapeake Bay. Stronger stratification usually results in larger hypoxic volume, which confirms the positive correlation between NE wind duration and hypoxic volume. In addition, *Cho et al.* [2012] also confirmed that down-estuary (e.g., NE wind) local wind stress tends to enhance stratification under moderate wind speeds, and the up-estuary (e.g., SW wind) local wind stress tends to reduce

stratification by reversing gravitational circulation. However, the wind effect on hypoxia detected in this work conflicts with that found from 1950-2009 by *Scully* [2010a]. It is important to note, though, that *Scully* [2010a] pointed out a shift in SE and westerly wind frequencies that occurred in the early 1980s, which is beyond the timespan examined in this work. Thus, the different timespans considered may thus be the main reason for the contradictory results.

To further explore the impact of nutrient loading and weather patterns to hypoxia, other indexes of hypoxic conditions such as the half-monthly hypoxic volume in the summer and the maximum half-monthly hypoxic volume during each year, were examined in this analysis. The spring TN, precipitation, and dominant wind effect explain 46% variability of the maximum hypoxic volume observed in each year. In addition, these variables explained 43% of the variability in all half-monthly (including early July) hypoxic volumes across years. This further confirms that the nutrient loading during spring and the dominant wind effect during spring and summer affect the overall hypoxic condition for each year.

6.5.5 The duration and the final and maximum months of hypoxia from 1985 to 2010

The duration and seasonal timing of hypoxia have changed from 1985 to 2010 (Figure 6-6). A standard of 0.75km^3 (~2% of the water volume in the mainstem) was used as the threshold of the appearance of hypoxia, which is the average uncertainty (i.e., red bar in Figure 6-5) for the modeled hypoxic volume in this work. There is no significant change for the appearance of hypoxia from 1985 to 2010 (p=0.38). However, the final month of hypoxia moves from October to September (p=0.005), and the duration of seasonal hypoxia decreases from five to less than four months (p=0.02). Although the hypoxic duration decreased, the mean hypoxic volume does not suggest a significant decline. This is because the hypoxic volumes in September and October are much smaller than those in the summer months (i.e., June to August).

Additionally, the appearance of the maximum hypoxic volume has moved from late July to early July ($p = 0.07$). Previous research suggested that this seasonality shift in the maximum hypoxic volume may be due to sea level rises, because a rising sea level could result in enhanced stratification strength [*Murphy et al.*, 2011]. However, a significant correlation was not detected. It is possible that more complex models are needed to properly identify this relationship.

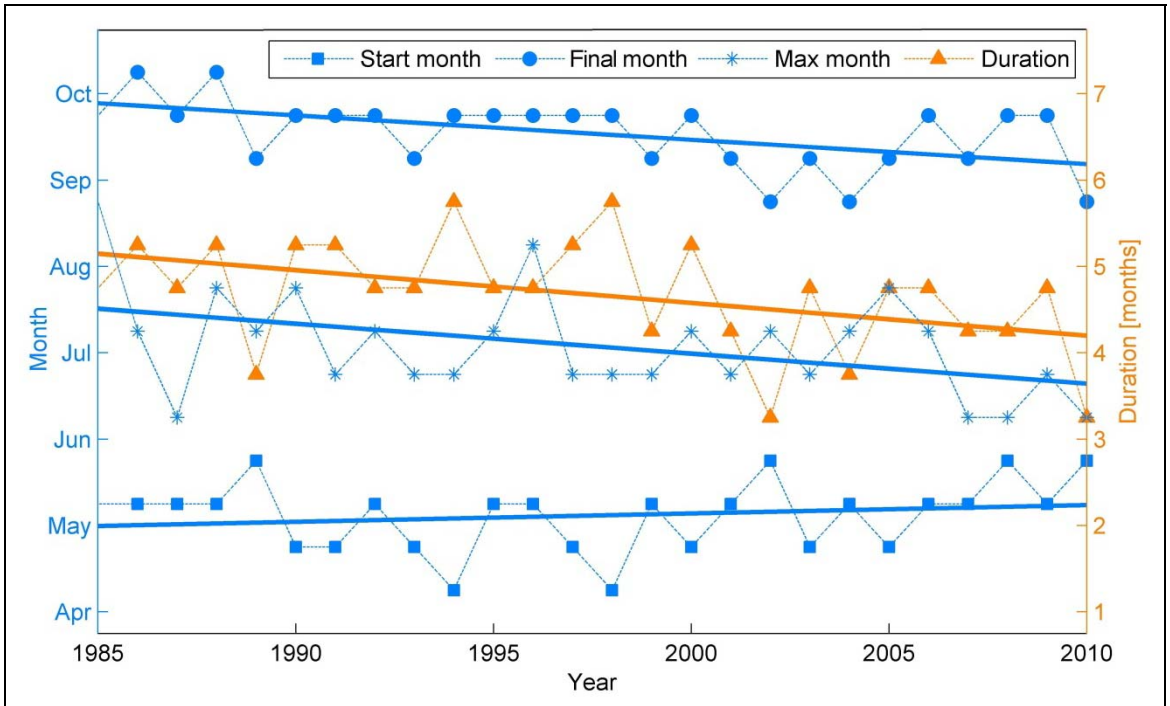

Figure 6-6 Start month (blue squares), final month (blue circles), maximum month (blue stars) and duration of hypoxic volume (orange triangles) in Chesapeake Bay from 1985 to 2010.

6.5.6 Model comparison results

The three-dimensional (3D) conditional realization results presented in this work are more comprehensive when compared to Main-channel OK and CBI in the sense that they provide estimation uncertainties associated with the best estimates. Main-channel OK and CBI can only provide the best estimates of hypoxic volumes, and are incapable of providing their associated estimated uncertainties. Because of the sparser sampling (i.e., some regular measurement stations

were skipped) in early July 1986 and 2002, and July to September 2008, the uncertainties (i.e., 95% confidence intervals) of the hypoxic volume during these time periods are much larger than other periods (Figure 6-7). This suggests the current DO network is reliable in estimating hypoxic volume using conditional realization, and reducing the number of measurements could result in larger uncertainties.

As expected, the estimated hypoxic volume figure from Main-channel OK are generally larger than 3D conditional realization results, especially in July and August (Figure 6-7), because the DO concentrations in the central channel of the Bay are lower than those in the near-shore zone (Figure 6-2). The assumption made by Main-channel OK of constant DO concentration along the lateral direction exaggerates the hypoxic volume.

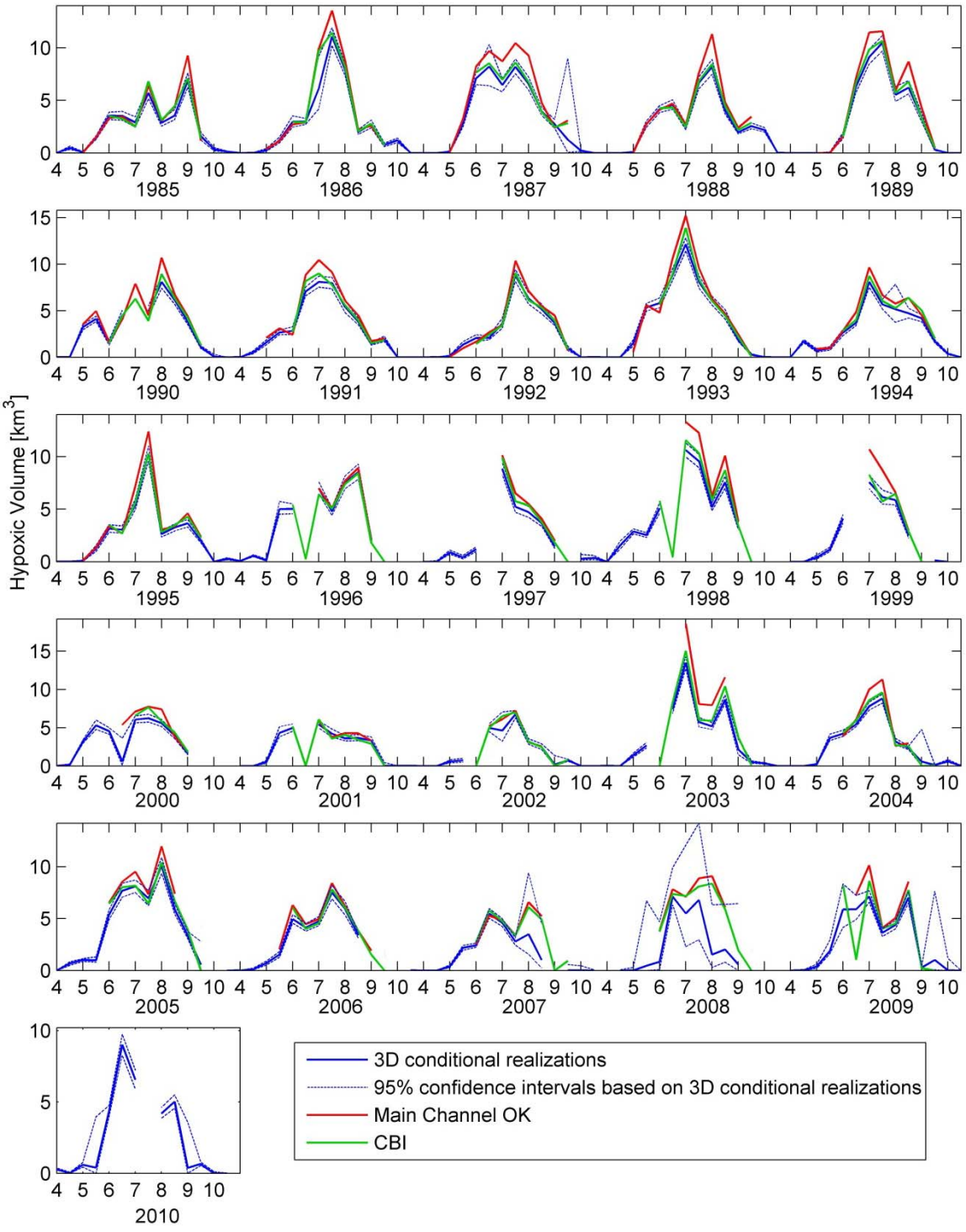

Figure 6-7 Half-monthly hypoxic volume estimates from 3D conditional realizations in blue lines, CBI in green lines and Main-channel OK in red lines. Note that the y axes are different for each subplot.

6.6 Conclusions

Geostatistical methods were implemented to estimate the hypoxic volume from April to October 1985-2010 and the associated uncertainties for Chesapeake Bay. This analysis shows that the mean hypoxic volume from April to October is better correlated with the examined factors than the maximum or any half-monthly hypoxic volume, suggesting that the mean hypoxic volume is better for elucidating mechanistic relationships between hypoxia and its contributing factors. The temporal trend in total hypoxic volume from 1985 to 2009 of this work is consistent with the results from *Bahner* [2006] and *Murphy et al.* [2011], although these three models were computed differently.

Whereas previous work had only considered TN loading from the Susquehanna [*Evans and Scavia*, 2011; *Hagy et al.*, 2004; *Scavia et al.*, 2006], or the Susquehanna and Potomac [*Murphy et al.*, 2011], including the loading from the Rappahannock was found to slightly improve the explanation of the interannual variability in mean hypoxic volume.

In addition to nitrogen loading, precipitation and wind duration are also shown to have significant contribution to the variability of hypoxic volumes in Chesapeake Bay. According to the analysis in this work, higher springtime precipitation leads to higher hypoxic volume, even after monitored tributary nitrogen loading are taken into account. Precipitation appears to influence the variability of hypoxia by affecting the nonpoint source of nitrogen downstream the river monitoring stations. The presented results confirm that wind duration is a better predictor of hypoxia than wind speed [*Scully et al.*, 2011], and that southwesterly winds are correlated with decreased hypoxic volume, while northeasterly winds lead to higher hypoxic volume.

This work also shows that the maximum hypoxic volume has been moving from late July to early July over the examined period, but there is no significant trend in the maximum hypoxic volume [*Murphy et al.*, 2011]. Furthermore, no significant trend was found in the timing of onset of hypoxia, but the end of the hypoxic period moved from October to September over the examined period.

CHAPTER 7 Conclusions and Future Directions

Each objective of this dissertation contributed in its own way to providing better scientific understanding of water quality in surface water or attribute (e.g., organic matter, contaminants) distributions in water sediments. This chapter summarizes the key findings and the overall contributions of this dissertation, and provides future directions.

7.1 Contributions of this dissertation

7.1.1 Objective 1: Geostatistical downscaling for data with a non-uniform resolution

The first objective of this dissertation involved the development of a geostatistical downscaling method for dealing with non-uniform resolution data sets, using pseudodata and the total organic carbon distribution in Passaic River sediment as examples. Both cases suggested that geostatistical downscaling provides more reliable estimates and associated uncertainties at a fine and uniform resolution when compared to the traditional point-to-point Ordinary Kriging (OK) that cannot account for the non-uniform resolution of the available data.

The methodology developed in this work can easily be adapted to other areas where the resolutions of the available samples is not uniform (e.g., remote sensing images with multi-resolution [*Zhan et al.*, 2012]), and therefore can improve estimation for engineers and scientists from a wide variety of backgrounds.

As part of this study, published code of geostatistical downscaling written in MATLAB, algorithms, and documentation have been made freely available to the scientific community (http://www.puorg.engin.umich.edu/).

7.1.2 Objective 2: Spatial and temporal trends in Lake Erie hypoxia from 1987 to 2007

The second objective of this dissertation was to provide a comprehensive analysis of the hypoxic extent in Lake Erie during August and September for the years 1987 to 2007 by using geostatistical methods. This work presented a geostatistical model selection approach that is able to determine the important factors for explaining the variability of bottom water dissolved oxygen (DO) concentrations. This work also provided the first quantitative history of hypoxic extent and its uncertainty in Lake Erie for each summer from 1987 to 2007.

The methods developed and the estimates of long-term hypoxic extent presented in this dissertation would benefit the research aimed at reducing hypoxic extent and improving water quality in this region. For example, this work revealed the impact of stratification on the summertime evolution of hypoxia. Therefore, in order to predict the hypoxia in Lake Erie, it is necessary to consider the stratification situation and the factors affecting it.

In addition, this work also provided a set of geostatistical methods that are easily transferable to other regions, with subsequent work applying these approaches in Chesapeake Bay (Chapter 6) and the Gulf of Mexico [*Obenour*, manuscript].

7.1.3 Objective 3: Impact of nutrient loading and weather patterns on seasonal hypoxia in Chesapeake Bay for 1985 to 2010

The third objective of this dissertation involved the estimation of the variability in hypoxic volume in half-monthly intervals for Chesapeake Bay over the period 1985 to 2010, and the examination of the effect of nutrient loading and weather patterns (e.g., wind, precipitation) on the estimated variability in hypoxic volume. Hypoxia varies annually, depending on the timing and magnitude of springtime and summertime stratification, nutrient loading, and weather patterns in the water body and watershed. Knowledge of the historical seasonality and inter-

annual variability of hypoxia provided in this work is of great importance for proper forecast of hypoxic volume in future.

Although management efforts directed toward reducing nitrogen loading have been implemented to improve the water quality in Chesapeake Bay, the observed correlation between nitrogen loads and summertime hypoxia is relatively low according to previous research. By estimating hypoxic volume within a larger timespan of each year, this work provided a thorough understanding of the variability of hypoxia. The research presented in this dissertation better explained the effect of nitrogen loading on hypoxia relative to the existing literature. Nitrogen loads together with weather patterns (e.g., precipitation, wind) explain more of the variability of the average annual hypoxic volume from April to October of the years 1985 to 2010 than that of the maximum hypoxic volume of each year or the hypoxic volume of any single half-month. This accurate quantification of the effect of nitrogen loads on hypoxia will benefit the water quality management strategies of reducing nutrient input and therefore improving the water quality in Chesapeake Bay.

7.1.4 Comparison between the research on the hypoxia in Lake Erie and that of Chesapeake Bay

The occurrence of hypoxia is increasing in surface waters worldwide, and represents a significant threat to the health and economy of both Lake Erie and Chesapeake Bay [*Committee on Environment and Natural Resources*, 2010]. Hypoxia in these two systems is caused by a combination of nutrient-driven production of phytoplankton organic matter, decomposition, and stratification of the water column. Although hypoxia can occur in natural and undisturbed systems, a dramatic increase in Lake Erie and Chesapeake Bay hypoxia has been linked to nutrient (i.e., nitrogen and phosphorus) and organic matter enrichment resulting from human

activities [*Committee on Environment and Natural Resources*, 2010]. According to the results from this dissertation, the relationship to nutrient loading is simpler and more clear in the Chesapeake Bay than in Lake Erie.

In addition to the hypoxia in Chesapeake Bay and Lake Erie, algal blooms are another main concern caused by the input of nutrients to these two systems [*Committee on Environment and Natural Resources*, 2010]. Lake Erie is divided into three basins (i.e., western, central and eastern) according to their bathymetry and locations. Algal blooms and hypoxia of Lake Erie mainly occur in western basin and central basin, respectively [*Becker et al.*, 2009; *Vincent et al.*, 2004]. However, these two phenomena of Chesapeake Bay occur both mainly in upper Bay and lower Bay [*Harding et al.*, 1994]. Consequently, it is possible that the response of algal blooms and hypoxia respond differently to nutrient loads for Lake Erie and Chesapeake Bay.

According to the variability of hypoxic extent from 1987 to 2007, hypoxic extent in Lake Erie decreased from the late 1980s to the early 1990s as a result of phosphorus load abatement programs, part of the United States/Canada Great Lakes Water Quality Agreement of 1972 [*Dolan*, 1993]. However, significant changes in the hypoxic volume in Chesapeake Bay were not observed from 1985 to 2010. In addition, the appearance of the maximum hypoxic volume was shown to have moved from late July to early July in Chesapeake Bay. However, similar change in the maximum hypoxic volume in Lake Erie was not detected.

Hypoxia in the two systems usually reaches maximum in July and September of each year, respectively. In Lake Erie, hypoxia appears in the western and middle part of central basin in August, and spreads to the eastern part of the central basin in September. This is because most nutrients enter the lake mainly from the west. Chesapeake Bay, hypoxia first appears in the upper

Bay, then spread to the middle and lower Bay. This is because most nutrients enter the Bay from the north of upper Bay. These results are consistent with the fact that the nutrient availability is the primary cause of the appearance of hypoxia in different systems.

Spatial trends play a significant role in estimating the bottom DO distributions in both Lake Erie and Chesapeake Bay. Weather patterns were shown to have a great impact on the inter-annual variability of hypoxic volume in Chesapeake Bay. However, such research was not conducted in Lake Erie. Therefore, there is a need to further explore the impact of weather patterns on the variability of hypoxic extent in Lake Erie.

7.1.5 Overall contributions

This dissertation focused on developing and applying geostatistical approaches for improving estimates of the spatial distribution of water quality parameters in surface water and organic matter in water sediment. In order to provide more accurate estimation tools, this dissertation developed several geostatistical data fusion models that make optimal use of data with different resolutions and from different sources. The infrastructure developed as part of this dissertation including the modeling framework, and code of the data fusion models developed fit into the community-wide effort to improve the quality of surface water.

Overall, knowledge of the spatial distribution of water quality parameters leads to a better understanding of physical, chemical, and ecological dynamics, making it easier to protect the world's precious water resources. This work improved the estimates of attributes (e.g., organic matter, contaminants) in sediment, which would benefit remedial actions. In addition, this work also improved the estimates of the variability of hypoxia in Lake Erie and Chesapeake Bay for the recent two-plus decades. In providing these more accurate estimates of hypoxic extent, this

dissertation further quantified the drivers of the long-term variability of water quality (e.g., hypoxia).

7.2 Future directions

This section describes specific avenues for future research for the three stated objective based on the findings of this dissertation. The section also provides some future directions for larger community-wide for water quality research.

7.2.1 Research on the spatial distribution of attributes in water sediment

This work demonstrated that geostatistical downscaling can provide more reliable estimates of total organic carbon at a fine and uniform resolution by using the *in situ* data with non-uniform resolutions in space. Future work could include:

- Expanding the method from two to three dimensions in order to provide more detailed information on the condition of the sediment contaminants. This dissertation only provided a two-dimensional estimate along fluvial and vertical directions.
- Qualifying the impact of river hydrology, erosion, and sedimentation on the distribution of organic carbon and contaminants in sediment and associated uncertainties. As the Passaic River is a tidal system, the distribution of organic carbon and contaminants in the inner bends and outer bends of the river may suggest different spatial structures. Knowing the correlation between the river dynamics and the distribution of organic carbon and contaminants could provide guidance for selecting sampling locations, which would further provide more information for interpolation contaminant distribution in space.

- Applying the developed method to other contaminants (e.g., PCB) distributions. Different contaminants are known to have different spatial distribution patterns. As discussed in Chapter 2, distribution of contaminants is of great importance for remedial actions. Therefore, it is necessary to provide more accurate estimations and associated uncertainty bounds of contaminant distribution using this method.
- Modifying the method developed in the first objective of this dissertation for non-normally distributed data. Current geostatistical methods are not suitable for non-normal data [*Kitanidis and Shen*, 1996]. In addition, covariance parameter optimization approaches (e.g., Restricted Maximum Likelihood) should also be correspondingly modified to obtain the representative spatial structure of such non-normal data.

7.2.2 Research on hypoxia in Lake Erie

While this work provided data on the variability of hypoxic extent in August and September for years 1987 to 2007 in Lake Erie, there are a few direct extensions of the current work that would help to make it more directly applicable to water quality studies:

- Examining the effect of nutrient loads and weather patterns on the variability of hypoxia in Lake Erie. Some extended work from this dissertation suggested that the linear correlation between phosphorous loading and hypoxic extent is weak (not discussed in the present work). It is therefore necessary to examine nonlinear or lagged relationships between nutrient loads and hypoxic extent in future research. In addition, it is highly likely that the combined contribution of phosphorous loading and weather patterns is greater than that of either factor alone. Chapter 6 confirmed that weather patterns (e.g., wind, precipitation) could present partial contributions to nutrient loading or stratification

situations, and therefore affects the variability in hypoxic volume in Chesapeake Bay. Similar contributions from these weather patterns might also be detected in Lake Erie.

- Forecasting hypoxic extent for the years when DO measurements are not available (both historically and in the future). Geostatistical methods can only estimate hypoxic extent when there are *in situ* DO measurements. Therefore, other methods are needed to estimate hypoxic extent for the years when such measurements do not exist or are otherwise unavailable. Similarly, it is also necessary to develop models to forecast hypoxia in future so as to inform water quality management and decision-making.

7.2.3 Research on hypoxia in Chesapeake Bay

This part of the work presented the variability of hypoxic volume from April to October for the years 1985 to 2010. Further, this work demonstrated the importance of nitrogen loading from tributaries and weather patterns in explaining the mean hypoxic volume from April to October (representing the yearly hypoxic condition). Future directions include:

- Evaluating all the sources of nutrient loads to Chesapeake Bay and examining their contribution to hypoxia. Currently, almost all studies included this dissertation focused solely on the nutrient loads at USGS monitoring stations from tributaries. There are other sources of nutrient loads such as those from wastewater treatment plants downstream from the monitoring stations, nonpoint sources downstream from the monitoring stations, and atmospheric deposition into tidal water [*USGS*, 2011]. However, direct measurements are not available for some of these sources of nutrients. Therefore, it is necessary to find factors that could represent these unmeasured nutrient loads. For example, this dissertation pointed out that spring precipitation appears to represent the effect of nonpoint sources of nutrients downstream from the river monitoring stations.

- Exploring the impact of extreme events on short-term variability of hypoxia. It remains unclear what effects the variability of hypoxic volume in for Chesapeake Bay in the short-term (e.g., half-month) [*Lewis et al.*, 2007]. For example, how extreme events (e.g., hurricanes, heavy and more frequent storms) affect hypoxia in the short-term (e.g., early July 2003) is another key factor of understanding the changes of water quality.

7.2.4 Community-wide future directions

The methods implemented for the sediment work (Chapter 4) in this dissertation focus only on snapshots in time accounting for spatial correlation in the observed data. Given the dynamic nature of river sediments, studies that take into account temporal (e.g., seasonal) variability should be considered for an ecological risk assessment [*Chapman and Wang*, 2001]. For the data that are temporally correlated, the space-time analysis was shown to provide more information for management regarding contaminant distribution by including more information both in space and time. Space-time models have been extensively used in several areas of environmental research, such as precipitation forecasts (e.g., [*Amani and Lebel*, 1997]), assessment of wind energy resources (e.g., [*Haslett and Raftery*, 1989]), and estimation of surface ozone levels (e.g., [*Guttorp et al.*, 1994; *Sampson et al.*, 1994]). Therefore, future studies in contaminated sediment area could include temporal correlation in addition to spatial correlation.

Additionally, sediments also affect the hypoxia in the water because they contain nutrients and organic matter (e.g. organic carbon) accumulated over the years [*Kennish*, 2002; *Rabalais et al.*, 2007b; *Turner et al.*, 2008]. Sediments thereby increase oxygen demand, and, in turn, expand the extent of hypoxia in bottom water [*Turner et al.*, 2008]. To decrease the effect of organic matter or nutrients in sediments on hypoxia, it is necessary to obtain their distribution in water

sediments. In addition to the organic matter and nutrients, the impact of sediment contaminants to aquatic environment is another problem posing threats to the future health and viability of ecosystems worldwide [*Kennish*, 2002]. *Testa and Kemp* [2012] suggested the positive feedback between hypoxia and sediment recycling of nutrients in Chesapeake Bay using quantitative analysis of long-term monitoring data. However, it remains unknown if their feedback is the same in the non-hypoxic locations, slight hypoxic locations and severe hypoxic locations. It is therefore necessary to improve understanding of how nutrients and organic matter accumulated in water sediments affect the bottom water hypoxia spatially, which will lead to better models for forecasting hypoxia.

As shown in this dissertation, weather patterns greatly affect water quality. Therefore, collaboration between researchers from these two research areas is needed to assess the impact of weather patterns on water quality changes and the response of water quality changes to variable weather patterns or climate. Many current nutrient control management designed to improve water quality in relatively stable climatic conditions may be ill prepared to adapt to future changes in climate, consumption, and population. Currently, research on the impact of weather patterns and climate changes on water quality usually takes place on the local level (e.g., [*Frick et al.*, 2008; *Lipp et al.*, 2001; *Quinn et al.*, 2001]). There is a need for more in-depth analysis involving both water quality forecasts and weather forecasts at a continental or even global scale. A global perspective on water quality changes is important for ensuring sustainable water use.

References

Adriaens, P., M. Y. Li, and A. M. Michalak (2006), Scaling Methods of Sediment Bioremediation Processes and Applications, in *International Symposium on Environmental Biotechnology*, edited, pp. 217-227, Leipzig, GERMANY, doi:10.1002/elsc.200520127.

Ahmed, S., and G. Demarsily (1987), Comparison of Geostatistical Methods for Estimating Transmissivity Using Data on Transmissivity and Specific Capacity, *Water Resources Research*, *23*(9), 1717-1737, doi:10.1029/WR023i009p01717.

Amani, A., and T. Lebel (1997), Lagrangian kriging for the estimation of Sahelian rainfall at small time steps, *Journal of Hydrology*, *192*(1-4), 125-157, doi:10.1016/s0022-1694(96)03104-6.

Anderson, D. R., K. P. Burnham, and G. C. White (1998), Comparison of Akaike Information Criterion and Consistent Akaike Information Criterion for Model Selection and Statistical Inference from Capture-Recapture Studies, *Journal of Applied Statistics*, *25*(2), 263-282.

Arthern, R. J., D. P. Winebrenner, and D. G. Vaughan (2006), Antarctic Snow Accumulation Mapped Using Polarization of 4.3-cm Wavelength Microwave Emission, *Journal of Geophysical Research, [Atmospheres]*, *111*(D6), doi:10.1029/2004jd005667.

Bahner, L. (2006), User Guide for the Chesapeake Bay and Tidal Tributary Interpolator*Rep.*, NOAA Chesapeake Bay Office, Annapolis, MD.

Barabas, N., P. Goovaerts, and P. Adriaens (2001), Geostatistical Assessment and Validation of Uncertainty for Three-Dimensional Dioxin Data from Sediments in an Estuarine River, *Environmental Science & Technology*, *35*(16), 3294-3301, doi:10.1021/es010568n.

Barabas, N., P. Goovaerts, and P. Adriaens (2004), Modified polytopic vector analysis to identify and quantify a dioxin dechlorination signature in sediments. 2. Application to the passaic river, *Environmental Science & Technology*, *38*(6), 1821-1827, doi:10.1021/es026229r.

Becker, R. H., M. I. Sultan, G. L. Boyer, M. R. Twiss, and E. Konopko (2009), Mapping cyanobacterial blooms in the Great Lakes using MODIS, *Journal of Great Lakes Research*, *35*(3), 447-453, doi:10.1016/j.jglr.2009.05.007.

Bertram, P. E. (1993), Total Phosphorus and Dissolved-Oxygen Trends in the Central Basin of Lake Erie, 1970-1991, *Journal of Great Lakes Research*, *19*(2), 224-236.

Bogaert, P., and D. Russo (1999), Optimal Spatial Sampling Design for the Estimation of the Variogram Based on a Least Squares Approach, *Water Resources Research*, *35*(4), 1275-1289, doi:10.1029/1998wr900078.

Bourennane, H., D. King, A. Couturier, B. Nicoullaud, B. Mary, and G. Richard (2007), Uncertainty Assessment of Soil Water Content Spatial Patterns Using Geostatistical Simulations: An Empirical Comparison of a Simulation Accounting for Single Attribute and a Simulation Accounting for Secondary Information, *Ecological Modelling*, *205*(3-4), 323-335, doi:10.1016/j.ecolmodel.2007.02.034.

Breitburg, D. L. (1990), Near-Shore Hypoxia in the Chesapeake Bay - Patterns and Relationships among Physical Factors, *Estuarine Coastal and Shelf Science*, *30*(6), 593-609, doi:10.1016/0272-7714(90)90095-9.

Burns, N. M., D. C. Rockwell, P. E. Bertram, D. M. Dolan, and J. J. H. Ciborowski (2005), Trends in Temperature, Secchi Depth, and Dissolved Oxygen Depletion Rates in the Central Basin of Lake Erie, 1983-2002, *Journal of Great Lakes Research*, *31*, 35-49.

Burrough, P. A., and R. A. McDonnell (1998), *Principles of Geographic Information Systems*, Oxford University Press.

Butcher, J. B. (1996), Co-kriging to Incorporate Screening Data: Hudson River Sediment PCBs, *Water Resources Bulletin*, *32*(2), 349-356.

Buzzelli, C. P., R. A. Luettich, S. P. Powers, C. H. Peterson, J. E. McNinch, J. L. Pinckney, and H. W. Paerl (2002), Estimating the Spatial Extent of Bottom-Water Hypoxia and Habitat Degradation in a Shallow Estuary, *Marine Ecology-Progress Series*, *230*, 103-112, doi:10.3354/meps230103.

Cao, H. S., F. X. Kong, L. C. Luo, X. L. Shi, Z. Yang, X. F. Zhang, and Y. Tao (2006), Effects of wind and wind-induced waves on vertical phytoplankton distribution and surface blooms of Microcystis aeruginosa in Lake Taihu, *Journal of Freshwater Ecology*, *21*(2), 231-238, doi:10.1080/02705060.2006.9664991.

Carrick, H. J., J. B. Moon, and B. F. Gaylord (2005), Phytoplankton Dynamics and Hypoxia in Lake Erie: A Hypothesis Concerning Benthic-Pelagic Coupling in the Central Basin, *Journal of Great Lakes Research*, *31*, 111-124.

Carroll, S. S., and N. Cressie (1996), A comparison of geostatistical methodologies used to estimate snow water equivalent, *Water Resources Bulletin*, *32*(2), 267-278.

Cerco, C. F., and T. Cole (1993), 3-Dimensional Eutrophication Model of Chesapeake Bay, *Journal of Environmental Engineering-ASCE*, *119*(6), 1006-1025, doi:10.1061/(ASCE)0733-9372(1993)119:6(1006).

Cerco, C. F., and M. R. Noel (2004), Process-Based Primary Production Modeling in Chesapeake Bay, *Marine Ecology-Progress Series*, *282*, 45-58, doi:10.3354/meps282045.

Cerco, C. F., M. R. Noel, and S. C. Kim (2006), Three-dimensional management model for Lake Washington, part II: Eutrophication modeling and skill assessment, *Lake and Reservoir Management*, *22*(2), 115-131.

Chapman, P. M., and F. Y. Wang (2001), Assessing sediment contamination in estuaries, *Environmental Toxicology and Chemistry*, *20*(1), 3-22, doi:10.1897/1551-5028(2001)020<0003:ascie>2.0.co;2.

Chapra, S. C. (1979), Applying Phosphorus Loading Models to Embayments, *Limnology and Oceanography*, *24*(1), 163-168.

Chehata, M., D. Jasinski, M. C. Monteith, and W. B. Samuels (2007), Mapping Three-Dimensional Water-Quality Data in the Chesapeake Bay Using Geostatistics, *Journal of the American Water Resources Association*, *43*(3), 813-828, doi:10.1111/j.1752-1688.2007.00065.x.

Chen, C.-C., G.-C. Gong, and F.-K. Shiah (2007), Hypoxia in the East China Sea: One of the largest coastal low-oxygen areas in the world, *Marine Environmental Research*, *64*(4), 399-408, doi:10.1016/j.marenvres.2007.01.007.

Chesapeake Bay Program (2011), Chesapeake Bay Program Water Quality Database (1984–present). edited, http://www.chesapeakebay.net/data_waterquality.aspx.

Chiles, J. P., and P. Delfiner (1999), *Geostatistics: Modeling Spatial Uncertainty*, Wiley, New York.

Cho, K.-H., H. V. Wang, J. Shen, A. Valle-Levinson, and Y.-c. Teng (2012), A modeling study on the response of Chesapeake Bay to hurricane events of Floyd and Isabel, *Ocean Modelling*, *49-50*, 22-46, doi:10.1016/j.ocemod.2012.02.005.

Clark, J. S., et al. (2001), Ecological Forecasts: An Emerging Imperative, *Science*, *293*(5530), 657-660, doi:10.1126/science.293.5530.657.

Cloern, J. E. (1999), The relative importance of light and nutrient limitation of phytoplankton growth: A simple index of coastal ecosystem sensitivity to nutrient enrichment, *Aquatic Ecology*, *33*(1), 3-16, doi:10.1023/a:1009952125558.

Committee on Environment and Natural Resources (2010), Scientific Assessment of Hypoxia in U.S. Coastal Waters*Rep.*, 154 pp, Interagency Working Group on Harmful Algal Blooms, Hypoxia, and Human Health of the Joint Subcommittee on Ocean Science and Technology. Washington, DC.

Coopersmith, E. J., B. Minsker, and P. Montagna (2011), Understanding and forecasting hypoxia using machine learning algorithms, *Journal of Hydroinformatics*, *13*(1), 64-80, doi:10.2166/hydro.2010.015.

Crawford, D. W., N. L. Bonnevie, and R. J. Wenning (1995), Sources of pollution and sediment contamination in Newark Bay, New-Jersey, *Ecotoxicology and Environmental Safety*, *30*(1), 85-100, doi:10.1006/eesa.1995.1010.

Cressie, N. A. C. (1993), *Statistics for Spatial Data*, 900 pp., John Wiley & Sons, Inc.

Cronin, T. M., and C. D. Vann (2003), The Sedimentary Record of Climatic and Anthropogenic Influence on the Patuxent Estuary and Chesapeake Bay Ecosystems, *Estuaries*, *26*(2A), 196-209, doi:10.1007/BF02695962.

Daskalov, G. M. (2003), Long-Term Changes in Fish Abundance and Environmental Indices in the Black Sea, *Marine Ecology-Progress Series*, *255*, 259-270, doi:10.3354/meps255259.

Dauer, D. M., A. J. Rodi, and J. A. Ranasinghe (1992), Effects of Low Dissolved-Oxygen Events on the Macrobenthos of the Lower Chesapeake Bay, *Estuaries*, *15*(3), 384-391, doi:10.2307/1352785.

Delhomme, J. P. (1979), Spatial Variability and Uncertainty in Groundwater-Flow Parameters - Geostatistical Approach, *Water Resources Research*, *15*(2), 269-280, doi:10.1029/WR015i002p00269.

Delorme, L. D. (1982), Lake Erie Oxygen - The Prehistoric Record, *Canadian Journal of Fisheries and Aquatic Sciences*, *39*(7), 1021-1029, doi:10.1139/f82-137.

Diaz, R. J., and R. Rosenberg (2008), Spreading Dead Zones and Consequences for Marine Ecosystems, *Science*, *321*(5891), 926-929, doi:10.1126/science.1156401.

Diaz, R. J., and R. Rosenberg (2011), Introduction to Environmental and Economic Consequences of Hypoxia, *International Journal of Water Resources Development*, *27*(1), 71-82, doi:10.1080/07900627.2010.531379.

DiToro, D. M., and J. P. Connolly (1980), Mathematical Models of Water Quality in Large Lakes, Part 2: Lake Erie.*Rep.*, Report No. EPA-600/3-80-065, report to Large Lakes Research Station, ERL-Duluth, Grosse Ile, MI 48138.

Dolan, D. M. (1993), Point-Source Loadings of Phosphorus to Lake Erie - 1986-1990, *Journal of Great Lakes Research*, *19*(2), 212-223.

Donovan, E. P., D. F. Staskal, K. M. Unice, J. D. Roberts, L. C. Haws, B. L. Finley, and M. A. Harris (2008), Risk of gastrointestinal disease associated with exposure to pathogens in the sediments of the Lower Passaic River, *Applied and Environmental Microbiology*, *74*(4), 1004-1018, doi:10.1128/aem.01203-07.

El Idrysy, E. H., and F. De Smedt (2007), A comparative study of hydraulic conductivity estimations using geostatistics, *Hydrogeology Journal*, *15*(3), 459-470, doi:10.1007/s10040-007-0166-0.

Erickson, and A. M. Michalak (2006), Merging of Variable-Resolution Imagery using Geostatistics and Sensor PSFs, in *American Society for Photogrammetry and Remote Sensing (ASPRS)* edited, Reno, Nevada.

Erickson, M. W. Williams, and A. Winstral (2005), Persistence of topographic controls on the spatial distribution of snow in rugged mountain terrain, Colorado, United States, *Water Resources Research*, *41*(4), doi:10.1029/2003wr002973.

Esterby, S. R., and P. E. Bertram (1993), Compatibility of Sampling and Laboratory Procedures Evaluated for the 1985 3-Ship Intercomparison Study on Lake Erie, *Journal of Great Lakes Research*, *19*(2), 400-417.

Evans, M. A., and D. Scavia (2011), Forecasting hypoxia in the Chesapeake Bay and Gulf of Mexico: model accuracy, precision, and sensitivity to ecosystem change, *Environmental Research Letters*, *6*(1), doi:10.1088/1748-9326/6/1/015001.

Faraway, J. J. (2004), *Linear models with R*, Chapman and Hall.

Fasbender, D., J. Radoux, and P. Bogaert (2008), Bayesian data fusion for adaptable image pansharpening, *Ieee Transactions on Geoscience and Remote Sensing*, *46*(6), 1847-1857, doi:10.1109/tgrs.2008.917131.

Feng, Y., S. F. DiMarco, and G. A. Jackson (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*, doi:10.1029/2012gl051192.

Fienen, M. N., P. K. Kitanidis, D. Watson, and P. Jardine (2004), An application of Bayesian inverse methods to vertical deconvolution of hydraulic conductivity in a heterogeneous aquifer at Oak Ridge National Laboratory, *Mathematical Geology*, *36*(1), 101-126, doi:10.1023/B:MATG.0000016232.71993.bd.

Fish, C. (1960), Limnological survey of eastern and central Lake Erie, 1928–29.US Fish and Wildlife Service Special ScienceReport, *Fisheries*, *334*.

Flemer, D. A., et al. (1983), Chesapeake Bay: A Profile of Environmental Change Appendices in Chesapeake Bay: A Profile of Environmental Change*Rep.*, 120 pp, U.S. Environmental Protection Agency Chesapeake Bay Program Office. Washington, D.C.

Foerstner, U., and S. E. Apitz (2007), Sediment remediation: US focus on capping and monitored natural recovery - Fourth international Battelle Conference on remediation of contaminated sediments, *Journal of Soils and Sediments*, *7*(6), 351-358, doi:10.1065/jss2007.10.256.

Folger, D. W. (1972), Characteristics of estuarine sediments of the United States*Rep.*, 94 pp.

French, J. R., T. Spencer, A. L. Murray, and N. S. Arnold (1995), Geostatistical analysis of sediment deposition in 2 small tidal wetlands, Norfolk, UK, *Journal of Coastal Research*, *11*(2), 308-321.

Frick, W. E., Z. Ge, and R. G. Zepp (2008), Nowcasting and forecasting concentrations of biological contaminants at beaches: A feasibility and case study, *Environmental Science & Technology*, *42*(13), 4818-4824, doi:10.1021/es703185p.

Georgakarakos, S., and D. Kitsiou (2008), Mapping abundance distribution of small pelagic species applying hydroacoustics and Co-Kriging techniques, *Hydrobiologia*, *612*, 155-169, doi:10.1007/s10750-008-9484-z.

Gohin, F., and G. Langlois (1993), Using geostatistics to merge insitu measurements and remotely-sensed observations of sea-surface temperature, *International Journal of Remote Sensing*, *14*(1), 9-19.

Gotway, C. A., and L. J. Young (2002), Combining incompatible spatial data, *Journal of the American Statistical Association*, *97*(458), 632-648.

Gourdji, S. M., K. L. Mueller, K. Schaefer, and A. M. Michalak (2008), Global monthly averaged CO2 fluxes recovered using a geostatistical inverse modeling approach: 2. Results including auxiliary environmental data, *Journal of Geophysical Research-Atmospheres*, *113*(D21), doi:10.1029/2007jd009733.

Great Lake National Program Office (GLNPO) (2008), edited.

Great Lakes Surface Environmental Analysis (GLSEA) (2008), edited.

Gunster, D. G., C. A. Gillis, N. L. Bonnevie, T. B. Abel, and R. J. Wenning (1993), Petroleum and hazardous chemical spills in Newark Bay, New-Jersey, USA from 1982 to 1991, *Environmental Pollution*, *82*(3), 245-253, doi:10.1016/0269-7491(93)90126-9.

Guo, X., and A. Valle-Levinson (2008), Wind effects on the lateral structure of density-driven circulation in Chesapeake Bay, *Continental Shelf Research*, *28*(17), 2450-2471, doi:10.1016/j.csr.2008.06.008.

Gutjahr, A., B. Bullard, S. Hatch, and L. Hughson (1994), Joint Conditional Simulations and the Spectral Approach for Flow Modeling, *Stochastic Hydrology and Hydraulics*, *8*(1), 79-108, doi:10.1007/bf01581391.

Guttorp, P., W. Meiring, and P. D. Sampson (1994), A space-time analysis of ground-level ozone data, *Environmetrics*, *5*(3), 241-254, doi:10.1002/env.3170050305.

Hagy, J. D., W. R. Boynton, C. W. Keefe, and K. V. Wood (2004), Hypoxia in Chesapeake Bay, 1950-2001: Long-term change in relation to nutrient loading and river flow, *Estuaries*, *27*(4), 634-658.

Hagy, J. D., and M. C. Murrell (2007), Susceptibility of a northern Gulf of Mexico estuary to hypoxia: An analysis using box models, *Estuarine Coastal and Shelf Science*, *74*(1-2), 239-253, doi:10.1016/j.ecss.2007.04.013.

Harding, L. W., E. C. Itsweire, and W. E. Esaias (1994), Estimates of phytoplankton biomass in the Chesapeake Bay from aircraft remote-sensing of chlorophyll concentrations, 1989-92, *Remote Sensing of Environment*, *49*(1), 41-56, doi:10.1016/0034-4257(94)90058-2.

Haslett, J., and A. E. Raftery (1989), Space-time modeling with long-memory dependence - assessing Irelands wind power resource, *Applied Statistics-Journal of the Royal Statistical Society Series C*, *38*(1), 1-50, doi:10.2307/2347679.

Hawley, N., T. H. Johengen, Y. R. Rao, S. A. Ruberg, D. Beletsky, S. A. Ludsin, B. J. Eadie, D. J. Schwab, T. E. Croley, and S. B. Brandt (2006), Lake Erie hypoxia prompts Canada-U.S. study, *Eos Trans. AGU*, *87*(32), 313, doi:10.1029/2006EO320001.

Haylock, M. R., N. Hofstra, A. M. G. K. Tank, E. J. Klok, P. D. Jones, and M. New (2008), A European daily high-resolution gridded data set of surface temperature and precipitation for 1950-2006, *Journal of Geophysical Research, [Atmospheres]*, *113*(D20), doi:10.1029/2008jd010201.

Hengl, T., G. B. M. Heuvelink, and D. G. Rossiter (2007), About regression-kriging: From equations to case studies, in *Geostats-UK 2005 Meeting*, edited, pp. 1301-1315, Belfast, NORTH IRELAND, doi:10.1016/j.cageo.2007.05.001.

Herdendorf, C. E. (1984), Lake Erie Water Quality 1970-1982: A Management Assessment*Rep.*, USEPA, GLNPO, Chicago IL.

Himmelheber, D. W., K. D. Pennell, and J. B. Hughes (2007), Natural attenuation processes during in situ capping, *Environmental Science & Technology*, *41*(15), 5306-5313, doi:10.1021/es0700909.

Hoeksema, R. J., R. B. Clapp, A. L. Thomas, A. E. Hunley, N. D. Farrow, and K. C. Dearstone (1989), Cokriging model for estimation of water-table elevation, *Water Resources Research*, *25*(3), 429-438.

Hoeting, J. A., R. A. Davis, A. A. Merton, and S. E. Thompson (2006), Model selection for geostatistical models, *Ecological Applications*, *16*(1), 87-98.

Hopkinson, C. S., and J. J. Vallino (1995), The relationships among mans activities in watersheds and estuaries - a model of runoff effects on patterns of estuarine community metabolism, *Estuaries*, *18*(4), 598-621, doi:10.2307/1352380.

Hunter, P. D., A. N. Tyler, N. J. Willby, and D. J. Gilvear (2008), The spatial dynamics of vertical migration by Microcystis aeruginosa in a eutrophic shallow lake: A case study using high spatial resolution time-series airborne remote sensing, *Limnology and Oceanography*, *53*(6), 2391-2406, doi:10.4319/lo.2008.53.6.2391.

Isaaks, E. H., and R. M. Srivastava (1989), *An introduction to applied geostatistics*, Oxford University Press.

Jerrett, M., R. T. Burnett, P. Kanaroglou, J. Eyles, N. Finkelstein, C. Giovis, and J. R. Brook (2001), A GIS - environmental justice analysis of particulate air pollution in Hamilton, Canada, *Environment and Planning A*, *33*(6), 955-973, doi:10.1068/a33137.

Kemp, W. M., et al. (2005), Eutrophication of Chesapeake Bay: historical trends and ecological interactions, *Marine Ecology-Progress Series*, *303*, 1-29.

Kennish, M. J. (2002), Environmental threats and environmental future of estuaries, *Environmental Conservation*, *29*(1), 78-107, doi:10.1017/s0376892902000061.

Kim, T., Y. P. Sheng, and K. Park (2010), Modeling water quality and hypoxia dynamics in Upper Charlotte Harbor, Florida, U.S.A. during 2000, *Estuarine Coastal and Shelf Science*, *90*(4), 250-263, doi:10.1016/j.ecss.2010.09.006.

Kitanidis, P. K. (1995), Quasi-Linear Geostatistical Theory for Inversing, *Water Resources Research*, *31*(10), 2411-2419, doi:10.1029/95wr01945.

Kitanidis, P. K. (1996), Analytical expressions of conditional mean, covariance, and sample functions in geostatistics, *Stochastic Hydrology and Hydraulics*, *10*(4), 279-294, doi:10.1007/bf01581870.

Kitanidis, P. K. (1997), *Introduction to Geostatistics: Applications in Hydrogeology*, Cambridge University Press.

Kitanidis, P. K., and R. W. Lane (1985), Maximum-likelihood parameter-estimation of hydrologic spatial processes by the Gauss-Newton method, *Journal of Hydrology*, *79*(1-2), 53-71.

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

Kitanidis, P. K., and E. G. Vomvoris (1983), A geostatistical approach to the inverse problem in groundwater modeling (steady-state) and one-dimensional simulations, *Water Resources Research*, *19*(3), 677-690.

Kravchenko, A., and D. G. Bullock (1999), A comparative study of interpolation methods for mapping soil properties, *Agronomy Journal*, *91*(3), 393-400.

Kurunc, A., K. Yurekli, and O. Cevik (2005), Performance of two stochastic approaches for forecasting water quality and streamflow data from Yesil iota rmak River, Turkey, *Environmental Modelling & Software*, *20*(9), 1195-1200, doi:10.1016/j.envsoft.2004.11.001.

Kyriakidis, P. C. (2004), A Geostatistical Framework for Area-to-Point Spatial Interpolation, *Geographical Analysis*, *36*(3), 259-289, doi:10.1111/j.1538-4632.2004.tb01135.x.

Kyriakidis, P. C., J. Kim, and N. L. Miller (2001), Geostatistical mapping of precipitation from rain gauge data using atmospheric and terrain characteristics, *Journal of Applied Meteorology*, *40*(11), 1855-1877.

Lake, J. L., N. I. Rubinstein, H. Lee, C. A. Lake, J. Heltshe, and S. Pavignano (1990), Equilibrium partitioning and bioaccumulation of sediment-associated contaminants by infaunal organisms, *Environmental Toxicology and Chemistry*, *9*(8), 1095-1106, doi:10.1897/1552-8618(1990)9[1095:epabos]2.0.co;2.

Leenaers, H., J. P. Okx, and P. A. Burrough (1990), Comparison of spatial prediction methods for mapping floodplain soil pollution, *Catena*, *17*(6), 535-550, doi:10.1016/0341-8162(90)90028-c.

Lewis, B. L., B. T. Glazer, P. J. Montbriand, G. W. Luther, III, D. B. Nuzzio, T. Deering, S. Ma, and S. Theberge (2007), Short-term and interannual variability of redox-sensitive chemical parameters in hypoxic/anoxic bottom waters of the Chesapeake Bay, *Marine Chemistry*, *105*(3-4), 296-308, doi:10.1016/j.marchem.2007.03.001.

Li, D. J., J. Zhang, D. J. Huang, Y. Wu, and J. Liang (2002), Oxygen depletion off the Changjiang (Yangtze River) Estuary, *Science in China Series D-Earth Sciences*, *45*(12), 1137-1146.

Lipp, E. K., N. Schmidt, M. E. Luther, and J. B. Rose (2001), Determining the effects of El Nino - Southern Oscillation events on coastal water quality, *Estuaries*, *24*(4), 491-497, doi:10.2307/1353251.

Little, L. S., D. Edwards, and D. E. Porter (1997), Kriging in estuaries: As the crow flies, or as the fish swims?, *Journal of Experimental Marine Biology and Ecology*, *213*(1), 1-11, doi:10.1016/s0022-0981(97)00006-3.

Liu, Y., and D. Scavia (2010), Analysis of the Chesapeake Bay Hypoxia Regime Shift: Insights from Two Simple Mechanistic Models, *Estuaries and Coasts*, *33*(3), 629-639, doi:10.1007/s12237-009-9251-z.

Loewen, M. R., J. D. Ackerman, and P. F. Hamblin (2007), Environmental implications of stratification and turbulent mixing in a shallow lake basin, *Canadian Journal of Fisheries and Aquatic Sciences*, *64*(1), 43-57, doi:10.1139/f06-165.

MacDonald, R., D. MacKay, and B. Hickie (2002), A contaminant amplification in the environment, *Environmental Science & Technology*, *36*(23), 456A-462A, doi:10.1021/es022470u.

Magar, V. S., and R. J. Wenning (2006), The role of monitored natural recovery in sediment remediation, *Integrated environmental assessment and management*, *2*(1), 66-74, doi:10.1897/1551-3793(2006)2[66:tromnr]2.0.co;2.

Makarewicz, J. C., and P. Bertram (1991), Evidence for the Restoration of the Lake Erie Ecosystem - Water-Quality, Oxygen Levels, and Pelagic Function Appear to Be Improving, *Bioscience*, *41*(4), 216-223, doi:10.2307/1311411.

Malone, T. C., W. M. Kemp, H. W. Ducklow, W. R. Boynton, J. H. Tuttle, and R. B. Jonas (1986), Lateral variation in the production and fate of phytoplankton in a partially stratified estuary, *Marine Ecology-Progress Series*, *32*(2-3), 149-160, doi:10.3354/meps032149.

Matheron, G. (1963), Principles of Geostatistics, *Economic Geology*, *58*.

Michalak, A. M. (2008), A Gibbs Sampler for Inequality-Constrained Geostatistical Interpolation and Inverse Modeling, *Water Resources Research*, *44*(9), doi:10.1029/2007wr006645.

Michalak, A. M., L. Bruhwiler, and P. P. Tans (2004), A geostatistical approach to surface flux estimation of atmospheric trace gases, *Journal of Geophysical Research-Atmospheres*, *109*(D14), doi:10.1029/2003jd004422.

Michalak, A. M., A. Hirsch, L. Bruhwiler, K. R. Gurney, W. Peters, and P. P. Tans (2005), Maximum Likelihood Estimation of Covariance Parameters for Bayesian Atmospheric Trace Gas Surface Flux Inversions, *Journal of Geophysical Research-Atmospheres*, *110*(D24), doi:10.1029/2005jd005970.

Michalak, A. M., and P. K. Kitanidis (2004), Estimation of historical groundwater contaminant distribution using the adjoint state method applied to geostatistical inverse modeling, *Water Resources Research*, *40*(8), doi:10.1029/2004wr003214.

Mouser, P. J., D. M. Rizzo, W. F. M. Roling, and B. M. Van Breukelen (2005), A multivariate statistical approach to spatial representation of groundwater contamination using hydrochemistry and microbial community profiles, *Environmental Science & Technology*, *39*(19), 7551-7559, doi:10.1021/es0502627.

Mueller, K. L., S. M. Gourdji, and A. M. Michalak (2008), Global monthly averaged CO2 fluxes recovered using a geostatistical inverse modeling approach: 1. Results using atmospheric measurements, *Journal of Geophysical Research-Atmospheres*, *113*(D21), doi:10.1029/2007jd009734.

Mueller, K. L., V. Yadav, P. S. Curtis, C. Vogel, and A. M. Michalak (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*, doi:10.1029/2009gb003642.

Murphy, R. R., F. C. Curriero, and W. P. Ball (2010), Comparison of Spatial Interpolation Methods for Water Quality Evaluation in the Chesapeake Bay, *Journal of Environmental Engineering-Asce*, *136*(2), 160-171, doi:10.1061/(asce)ee.1943-7870.0000121.

Murphy, R. R., W. M. Kemp, and W. P. Ball (2011), Long-Term Trends in Chesapeake Bay Seasonal Hypoxia, Stratification, and Nutrient Loading, *Estuaries and Coasts*, *34*(6), 1293-1309, doi:10.1007/s12237-011-9413-7.

NASA Ocean Color Web facility, edited.

National Geophysical Data Center (2008), edited, http://www.ngdc.noaa.gov/

National Research Council (2000), *Clean Coastal Waters*, National Academy Press, Washington D.C.

National Research Council (2007), Sediment Dredging at Superfund Megasites: Assessing the Effectiveness*Rep.*, 236 pp.

NCDC (Accessed: August, 2012), National Climatic Data Center, edited, http://www.ncdc.noaa.gov/cdo-web/.

Newcombe, C. L., and W. A. Horne (1938), Oxygen-poor waters of the Chesapeake Bay, *Science*, *88*.

Nixon, S. W., S. L. Granger, and B. L. Nowicki (1995), An assessment of the annual mass balance of carbon, nitrogen, and phosphorus in Narragansett Bay, *Biogeochemistry*, *31*(1), 15-61.

NOAA (2002), Watershed Database and Mapping Projects, edited, http://response.restoration.noaa.gov/cpr/watershed/watershedtools.html.

NSTC (2000), An integrated assessment of hypoxia in the northern Gulf of Mexico*Rep.*

NSTC (2003), An assessment of coastal hypoxia and eutrophication in U.S. waters*Rep.*

NWS (Accessed: August 2012), National Weather Service edited, http://www.weather.gov/.

Obenour, D. R. (manuscript), Reanalysis for mid-summer hypoxic extent in the northern Gulf of Mexico, 1985-2008.

Obenour, D. R., A. M. Michalak, Y. Zhou, and D. Scavia (2012), Quantifying the Impacts of Stratification and Nutrient Loading on Hypoxia in the Northern Gulf of Mexico, *Environmental Science & Technology*, *46*(10), 5489-5496, doi:10.1021/es204481a.

Officer, C. B., R. B. Biggs, J. L. Taft, L. E. Cronin, M. A. Tyler, and W. R. Boynton (1984), Chesapeake Bay anoxia - origin, development, and significance, *Science*, *223*(4631), 22-27.

Olson, M., and G. Shenk (2003), Calculation of the Pycnocline - Chesapeake Bay Program, edited, p. Powerpoint Slides.

Pannone, M., and P. K. Kitanidis (2001), Large-time spatial covariance of concentration of conservative solute and application to the Cape Cod tracer test, *Transport in Porous Media*, *42*(1-2), 109-132, doi:10.1023/a:1006704215335.

Pardo-Iguzquiza, E., M. Chica-Olmo, and P. M. Atkinson (2006), Downscaling cokriging for image sharpening, *Remote Sensing of Environment*, *102*(1-2), 86-98, doi:10.1016/j.rse.2006.02.014.

Pardo-Iguzquiza, E., P. A. Dowd, and D. I. F. Grimes (2005), An automatic moving window approach for mapping meteorological data, *International Journal of Climatology*, *25*(5), 665-678, doi:10.1002/joc.1128.

Parker, C. A., and J. E. Oreilly (1991), Oxygen depletion in Long-Island Sound - a historical-perspective, *Estuaries*, *14*(3), 248-264.

Pearson, T. H., and R. Rosenberg (1978), Macrobenthic succession in relation to organic enrichment and pollution of the marine environment, *Annual review of oceanography and marine biology*, *16*.

Perelo, L. W. (2010), Review: In situ and bioremediation of organic pollutants in aquatic sediments, *Journal of Hazardous Materials*, *177*(1-3), 81-89, doi:10.1016/j.jhazmat.2009.12.090.

Phillips, D. L., J. Dolph, and D. Marks (1992), A comparison of geostatistical procedures for spatial-analysis of precipitation in mountainous terrain, *Agricultural and Forest Meteorology*, *58*(1-2), 119-141.

Pihl, L., S. P. Baden, and R. J. Diaz (1991), Effects of periodic hypoxia on distribution of demersal fish and crustaceans, *Marine Biology*, *108*(3), 349-360.

Pokryfki, L., and R. E. Randall (1987), Nearshore Hypoxia in the Bottom Water of the Northwestern Gulf of Mexico from 1981 to 1984, *Marine Environmental Research*, *22*(1), 75-90, doi:10.1016/0141-1136(87)90081-x.

Quinn, N. W. T., N. L. Miller, J. A. Dracup, L. Brekke, and L. F. Grober (2001), An integrated modeling system for environmental impact analysis of climate variability and extreme

weather events in the San Joaquin Basin, California, *Advances in Environmental Research*, *5*(4), 309-317, doi:10.1016/s1093-0191(01)00081-8.

Rabalais, N. N. (2005), *Eutrophication*, The global coastal ocean. Multiscale interdisciplinary processes. The sea: ideas and observations on progress in the study of the seas.

Rabalais, N. N., R. E. Turner, and D. Scavia (2002), Beyond science into policy: Gulf of Mexico hypoxia and the Mississippi River, *Bioscience*, *52*(2), 129-142, doi:10.1641/0006-3568.

Rabalais, N. N., R. E. Turner, B. K. Sen Gupta, D. F. Boesch, P. Chapman, and M. C. Murrell (2007a), 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., R. E. Turner, B. K. Sen Gupta, E. Platon, and M. L. Parsons (2007b), Sediments tell the history of eutrophication and hypoxia in the Northern Gulf of Mexico, *Ecological Applications*, *17*(5), S129-S143, doi:10.1890/06-0644.1.

Rabalais, N. N., R. E. Turner, and W. J. Wiseman Jr (2001), Hypoxia in the Gulf of Mexico, *Journal of Environmental Quality*, *30*.

Rao, Y. R., N. Hawley, M. N. Charlton, and W. M. Schertzer (2008), Physical processes and hypoxia in the central basin of Lake Erie, *Limnology and Oceanography*, *53*(5), 2007-2020.

Rathbun, S. L. (1998), Spatial modelling in irregularly shaped regions: Kriging estuaries, *Environmetrics*, *9*(2), 109-129, doi:10.1002/(sici)1099-095x(199803/04)9:2<109::aid-env279>3.0.co;2-l.

Ridgway, J., and G. Shimmield (2002), Estuaries as repositories of historical contamination and their impact on shelf seas, *Estuarine Coastal and Shelf Science*, *55*(6), 903-928, doi:10.1006/ecss.2002.1035.

Riley, G. A. (1946), Factors controlling phytoplankton populations on Georges Bank, *Journal of Marine Research*, *6*(1), 54-73.

Robertson, G. P. (1987), Geostatistics in ecology - interpolating with known variance, *Ecology*, *68*(3), 744-748.

Rosenberg, R., and L. O. Loo (1988), Marine eutrophication induced oxygen deficiency - effects on soft bottom Fauna, western Sweden, *Ophelia*, *29*(3), 213-225.

Rossi, M., and D. Posa (1990), 3-D mapping of dissolved-oxygen in Mar-Piccolo - a case-study, *Environmental Geology and Water Sciences*, *16*(3), 209-219, doi:10.1007/bf01706045.

Rucinski, D. K., D. Beletsky, J. V. DePinto, D. J. Schwab, and D. Scavia (2010), A simple 1-dimensional, climate based dissolved oxygen model for the central basin of Lake Erie, *Journal of Great Lakes Research*, *36*(3), 465-476, doi:10.1016/j.jglr.2010.06.002.

Salehin, M., A. I. Packman, and M. Paradis (2004), Hyporheic exchange with heterogeneous streambeds: Laboratory experiments and modeling, *Water Resources Research*, *40*(11), doi:10.1029/2003wr002567.

Sampson, P. D., P. Guttorp, W. Meiring, S. S. Amer Stat Assoc, and Environm (1994), *Spatio-temporal analysis of regional ozone data for operational evaluation of an air quality model*, 46-55 pp.

Sandberg, E. (1994), Does short-term oxygen depletion affect predator-prey relationships in zoobenthos - experiments with the isopod saduria entomon, *Marine Ecology-Progress Series*, *103*(1-2), 73-80.

Sanford, L. P., K. G. Sellner, and D. L. Breitburg (1990), Covariability of dissolved-oxygen with physical processes in the summertime Chesapeake Bay, *Journal of Marine Research*, *48*(3), 567-590.

Scavia, D., E. L. A. Kelly, and J. D. Hagy, III (2006), A simple model for forecasting the effects of nitrogen loads on Chesapeake Bay hypoxia, *Estuaries and Coasts*, *29*(4), 674-684.

Schabenberger, O., and C. A. Gotway (2005), *Statistical methods for spatial data analysis*, Boca Raton : Chapman & Hall/CRC.

Schertzer, W. M., J. H. Saylor, F. M. Boyce, D. G. Robertson, and F. Rosa (1987), Seasonal Thermal Cycle of Lake Erie, *Journal of Great Lakes Research*, *13*(4), 468-486.

Schumacher, B. A. (2002), Methods for the Determination of Total Organic Carbon (TOC) in Soils and Sediments*Rep.*, Ecological Risk Assessment Support Center, EPA.

Schwab, D. J., G. A. Leshkevich, and G. C. Muhr (1999), Automated mapping of surface water temperature in the Great Lakes, *Journal of Great Lakes Research*, *25*(3), 468-481.

Schwarz, G. (1978), Estimating Dimension of a Model, *Annals of Statistics*, *6*(2), 461-464, doi:10.1214/aos/1176344136.

Scully, M. E. (2010a), The Importance of Climate Variability to Wind-Driven Modulation of Hypoxia in Chesapeake Bay, *Journal of Physical Oceanography*, *40*(6), 1435-1440, doi:10.1175/2010jpo4321.1.

Scully, M. E. (2010b), Wind Modulation of Dissolved Oxygen in Chesapeake Bay, *Estuaries and Coasts*, *33*(5), 1164-1175, doi:10.1007/s12237-010-9319-9.

Scully, M. E., W. R. Geyer, and J. H. Trowbridge (2011), The Influence of Stratification and Nonlocal Turbulent Production on Estuarine Turbulence: An Assessment of Turbulence Closure with Field Observations, *Journal of Physical Oceanography*, *41*(1), 166-185, doi:10.1175/2010jpo4470.1.

Sheng, Y. P. (1990), Evolution of a Three-Dimensional Cuivilinear-Grid Hydrodynamic Model for Estuaries, Lakes and Coastal Watches: CH3D, in *Estuarine and Coastal Modeling, ASCE*, edited by M. L. Spaulding, pp. 40-49.

Snodgrass, M. F., and P. K. Kitanidis (1997), A geostatistical approach to contaminant source identification, *Water Resources Research*, *33*(4), 537-546, doi:10.1029/96wr03753.

Stefanovic, D. L., and H. G. Stefan (2002), Two-dimensional temperature and dissolved oxygen dynamics in the littoral region of an ice-covered lake, *Cold Regions Science and Technology*, *34*(3), 159-178.

Stow, C. A., and D. Scavia (2009), Modeling hypoxia in the Chesapeake Bay: Ensemble estimation using a Bayesian hierarchical model, *Journal of Marine Systems*, *76*(1-2), 244-250, doi:10.1016/j.jmarsys.2008.05.008.

Sun, A. Y. (2007), A robust geostatistical approach to contaminant source identification, *Water Resources Research*, *43*(2), doi:10.1029/2006wr005106.

Testa, J. M., and W. M. Kemp (2012), Hypoxia-induced shifts in nitrogen and phosphorus cycling in Chesapeake Bay, *Limnology and Oceanography*, *57*(3), 835-850, doi:10.4319/lo.2012.57.3.0835.

Trangmar, B. B., R. S. Yost, and G. Uehara (1985), Application of geostatistics to spatial studies of soil properties, *Advances in Agronomy*, *38*, 45-94.

Turner, R. E., N. N. Rabalais, and D. Justic (2008), Gulf of Mexico hypoxia: Alternate states and a legacy, *Environmental Science & Technology*, *42*(7), 2323-2327, doi:10.1021/es071617k.

U. S. EPA (2002a), Mid-Atlantic Integrated Assessment (MAIA) Estuaries 1997-98: Summary Report*Rep.*, 115 pp.

U. S. EPA (2002b), The State of the Chesapeake Bay. A Report to the Citizens of the Bay Region.*Rep.*, Annapolis, Maryland.

U. S. EPA (2002c), Superfund Site Information Search, edited.

U. S. EPA (2005), Contaminated Sediment Remediation Guidance for Hazardous Waste Sites*Rep.*, United States Environmental Protection Agency.

U. S. EPA (2007), Ambient water quality criteria for dissolved oxygen, water clarity and chlorophyll a for the Chesapeake Bay and its tidal tributaries 2007 addendum.*Rep. EPA 903-R-07-003*, U.S. Environmental Protection Agency Region III, Chesapeake Bay Program Office, and Region III Water Protection Division, Annapolis, MD.

U.S. EPA (2008), Report on the Environment*Rep.*, EPA.

USGS (2010), Measuring Nutrient and Sediment Loads to Chesapeake Bay, edited, http://va.water.usgs.gov/chesbay/RIMP/.

USGS (2011), Chesapeake Bay river input monitoring program, , edited, http://cbrim.er.usgs.gov/featuremeasuringnutrient-sedimentloads.html.

Valettesilver, N. J. (1993), The use of sediment cores to reconstruct historical trends in contamination of estuarine and coastal sediments, *Estuaries*, *16*(3B), 577-588, doi:10.2307/1352796.

Valle-Levinson, A., J. L. Miller, and G. H. Wheless (1998), Enhanced stratification in the lower Chesapeake Bay following northeasterly winds, *Continental Shelf Research*, *18*(13), 1631-1647, doi:10.1016/s0278-4343(98)00067-3.

Vanderoost, R., H. Heida, and A. Opperhuizen (1988), Polychlorinated Biphenyl Congeners in Sediments, Plankton, Molluscs, Crustaceans, and Eel in a Freshwater Lake: Implications, of Using Reference Chemicals and Indicator Organisms in Bioaccumulation Studies, *Archives of Environmental Contamination and Toxicology*, *17*(6), 721-729.

Vanderploeg, H. A., T. H. Johengen, and J. R. Liebig (2009), Feedback between zebra mussel selective feeding and algal composition affects mussel condition: did the regime changer pay a price for its success?, *Freshwater Biology*, *54*(1), 47-63, doi:10.1111/j.1365-2427.2008.02091.x.

Vincent, R. K., X. M. Qin, R. M. L. McKay, J. Miner, K. Czajkowski, J. Savino, and T. Bridgeman (2004), Phycocyanin detection from LANDSAT TM data for mapping

cyanobacterial blooms in Lake Erie, *Remote Sensing of Environment*, *89*(3), 381-392, doi:10.1016/j.rse.2003.10.014.

Vollenweider, R. A. (1979), Concept of Nutrient Load as a Basis for the External Control of the Eutrophication Process in Lakes and Reservoirs, *Zeitschrift Fur Wasser Und Abwasser Forschung-Journal for Water and Wastewater Research*, *12*(2), 46-56.

Walker, R. P. McNutt, and C. K. Maslanka (1999), The potential contribution of urban runoff to surface sediments of the Passaic River: Sources and chemical characteristics, *Chemosphere*, *38*(2), 363-377, doi:10.1016/s0045-6535(98)00186-6.

Walker, and N. N. Rabalais (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, Q. R., D. Kim, D. D. Dionysiou, G. A. Sorial, and D. Timberlake (2004), Sources and remediation for mercury contamination in aquatic systems - a literature review, *Environmental Pollution*, *131*(2), 323-336, doi:10.1016/j.envpol.2004.01.010.

Wells, D., and J. Hill (2004), A synthesis of sediment chemical contaminant studies in the Maryland Coastal Bays in *Marylangd's Coastal Bays Ecosystem Health Assessment*, edited, Maryland Department of Natural Resources.

Welsh, B. L., and F. C. Eller (1991), Mechanisms controlling summertime oxygen depletion in western Long-Island Sound, *Estuaries*, *14*(3), 265-278.

Western, A. W., and G. Bloschl (1999), On the spatial scaling of soil moisture, *Journal of Hydrology*, *217*(3-4), 203-224, doi:10.1016/s0022-1694(98)00232-7.

Witter, D. L., J. D. Ortiz, S. Palm, R. T. Heath, and J. W. Budd (2009), Assessing the application of SeaWiFS ocean color algorithms to Lake Erie, *Journal of Great Lakes Research*, *35*(3), 361-370, doi:10.1016/j.jglr.2009.03.002.

Woodbury, A. D., and E. A. Sudicky (1991), The geostatistical characteristics of the borden aquifer, *Water Resources Research*, *27*(4), 533-546.

Wu, J., W. A. Norvell, and R. M. Welch (2006), Kriging on highly skewed data for DTPA-extractable soil Zn with auxiliary information for pH and organic carbon, *Geoderma*, *134*(1-2), 187-199, doi:10.1016/j.geoderma.2005.11.002.

Wynne, T. T., R. P. Stumpf, M. C. Tomlinson, and J. Dyble (2010), Characterizing a cyanobacterial bloom in western Lake Erie using satellite imagery and meteorological data, *Limnology and Oceanography*, *55*(5), 2025-2036, doi:10.4319/lo.2010.55.5.2025.

Yeh, T. C. J., and J. Q. Zhang (1996), A geostatistical inverse method for variably saturated flow in the vadose zone, *Water Resources Research*, *32*(9), 2757-2766, doi:10.1029/96wr01497.

Zeller, C., and B. Cushing (2006), Panel discussion: remedy effectiveness: what works, what doesn't?, *Integrated environmental assessment and management*, *2*(1), 75-79, doi:10.1897/1551-3793(2006)2[75:pdreww]2.0.co;2.

Zhan, W., Y. Chen, J. Wang, J. Zhou, J. Quan, W. Liu, and J. Li (2012), Downscaling land surface temperatures with multi-spectral and multi-resolution images, *International Journal of Applied Earth Observation and Geoinformation*, *18*, 23-36, doi:10.1016/j.jag.2012.01.003.

Zhang, X., M. Roman, D. Kimmel, C. McGilliard, and W. Boicourt (2006), Spatial variability in plankton biomass and hydrographic variables along an axial transect in Chesapeake Bay, *Journal of Geophysical Research-Oceans*, *111*(C5), doi:10.1029/2005jc003085.

Zhou, Y., and A. M. Michalak (2009), Characterizing Attribute Distributions in Water Sediments by Geostatistical Downscaling, *Environmental Science & Technology*, *43*(24), 9267-9273, doi:10.1021/es901431y.

Zimmerman, D. A., et al. (1998), A comparison of seven geostatistically based inverse approaches to estimate transmissivities for modeling advective transport by groundwater flow, *Water Resources Research*, *34*(6), 1373-1413, doi:10.1029/98wr00003.

Zimmerman, D. A., C. Pavlik, A. Ruggles, and M. P. Armstrong (1999), An experimental comparison of ordinary and universal kriging and inverse distance weighting, *Mathematical Geology*, *31*(4), 375-390.