Retrospective Analysis of Midsummer Hypoxic Area and Volume in the Northern Gulf of Mexico, 1985−2011

Daniel R. Obenour, Donald Scavia, Nancy N. Rabalais, R. Eugene Turner, and Anna M. Michalak

SEDAR74-RD84

September 2021

Environmental Science & Technology

Article

pubs.acs.org/est

Retrospective Analysis of Midsummer Hypoxic Area and Volume in the Northern Gulf of Mexico, 1985–2011

Daniel R. Obenour,[†,‡,*] Donald Scavia,[†,‡] Nancy N. Rabalais,[§] R. Eugene Turner,[∥] and Anna M. Michalak[⊥]

[†]School of Natural Resources & Environment, University of Michigan, Ann Arbor, Michigan 48109-1041, United States

[‡]Department of Civil & Environmental Engineering, University of Michigan, Ann Arbor, Michigan 48109-2125, United States ,

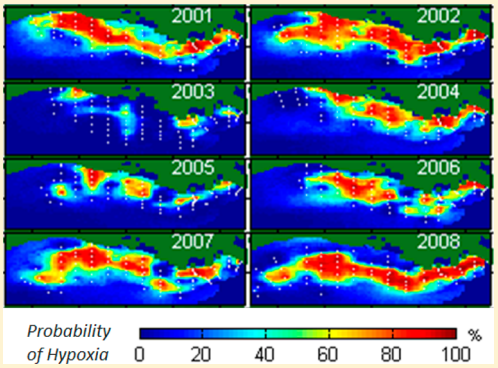
[§]Louisiana Universities Marine Consortium, 8124 Highway 56, Chauvin, Louisiana 70344, United States

[∥]Department of Oceanography and Coastal Sciences, Louisiana State University, Baton Rouge, Louisiana 70803, United States

[⊥]Department of Global Ecology, Carnegie Institution for Science, Stanford, California 94305, United States

Supporting Information

ABSTRACT: Robust estimates of hypoxic extent (both area and volume) are important for assessing the impacts of low dissolved oxygen on aquatic ecosystems at large spatial scales. Such estimates are also important for calibrating models linking hypoxia to causal factors, such as nutrient loading and stratification, and for informing management decisions. In this study, we develop a rigorous geostatistical modeling framework to estimate the hypoxic extent in the northern Gulf of Mexico from data collected during midsummer, quasi-synoptic monitoring cruises (1985–2011). Instead of a traditional interpolation-based approach, we use a simulation-based approach that yields more robust extent estimates and quantified uncertainty. The modeling framework also makes use of covariate information (i.e., trend variables such as depth and spatial position), to reduce estimation uncertainty. Furthermore, adjustments are made to account for observational bias resulting from the use of different sampling instruments in different years. Our results suggest an increasing trend in hypoxic layer thickness (p = 0.05) from 1985 to 2011, but less than significant increases in volume (p = 0.12) and area (p = 0.42). The uncertainties in the extent estimates vary with sampling network coverage and instrument type, and generally decrease over the study period.



1. INTRODUCTION

A large hypoxic zone has formed nearly every year in the northern Gulf of Mexico over at least the last few decades.[1−3] It is the largest human-caused coastal hypoxic zone in the western hemisphere and one of the largest worldwide.[1,4] Because of its negative ecological impacts and potential fisheries impacts,[5] the hypoxic zone has received attention from various stakeholders and policy makers who have developed plans to reduce its average size.[6,7]

The hypoxic zone is generally largest in summer due to the seasonal cycles of phytoplankton production and water column stratification, which are the two primary factors leading to dissolved oxygen (DO) depletion.[8−10] Both factors are related to the outflow from the Mississippi River Basin, which typically peaks in March–May. Nutrient loads from the Basin stimulate phytoplankton production in offshore waters, and much of the resulting organic matter eventually settles along the Louisiana–Texas shelf. Aerobic bacteria decompose the organic matter, consuming large quantities of DO. At the same time, freshwater outflows from the Basin promote stratification through a fresher surface layer, which warms in spring and summer, and overlays a colder, saltier water layer, inhibiting reoxygenation of bottom waters.[1]

The hypoxic extent, operationally defined as the region (area or volume) where DO concentrations are below 2 mg L^{-1}, is often assessed using data from "shelfwide" sampling cruises. The cruises are considered "quasi-synoptic" because changes in weather conditions (e.g., tropical storms) preceding or during the cruises can disrupt typical seasonal patterns in hypoxia. Cruises to document the occurrence of hypoxia, as well as related physical and biological parameters, have been performed by the Louisiana Universities Marine Consortium (LUMCON) beginning in 1985.[1,2] The estimated bottom-water hypoxic area[2] has been determined by interpolating between sampling locations and hand-contouring parallel to isobaths over a calibrated (planimeter) grid. These estimates have been used in multiple modeling studies linking hypoxia to nutrient loads from the Mississippi River basin and other environmental factors[11−14] and for setting policy goals to reduce the severity of hypoxia.[6,7] However, these estimates are generally conservative in that the most inshore, most offshore,

Received: March 4, 2013
Revised: July 8, 2013
Accepted: July 29, 2013
Published: July 29, 2013

   dx.doi.org/10.1021/es400983g | Environ. Sci. Technol. 2013, 47, 9808–9815

and most western extents of hypoxia are not always captured because of logistical constraints.[15,16] Also, these estimates do not quantify the uncertainty inherent in making estimates from limited observations.[17,18] Quantified uncertainties are useful for assessing the adequacy of existing sampling programs and for improving models that link hypoxic extent to environmental causes and effects.

Hypoxic area estimates only partially characterize the hypoxic extent. The thickness and volume of hypoxia are also important, as they relate to how hypoxia affects pelagic organisms.[19,20] In addition, hypoxic volume, rather than area, should be more closely related to the total oxygen deficit of the system, potentially making it a more useful metric for biogeochemical modeling studies.[4] While hypoxic volume measurements are available for other systems, such as Chesapeake Bay,[21,22] estimates for the Gulf have only been available for a subset of years, using an unpublished methodology.[4]

The primary goal of this study is to improve our knowledge of the Gulf's hypoxic extent over the 27-year study period (1985−2011) by systematically estimating midsummer hypoxic area and volume. These are the first Gulf hypoxic extent estimates to include "instrument bias adjustments" that account for the use of different sampling instruments, capable of being lowered to within different proximities of the sea floor, in different years.[23] Our approach also uses trends between DO, hypoxic fraction, depth, and spatial position to account for consistent, large-scale spatial patterns in DO, thereby improving the explanatory power of the model.[18] Our extent estimates are developed using a Monte Carlo-type simulation approach, rather than traditional interpolation, allowing for uncertainty quantification (i.e., confidence intervals).[17,18] These confidence intervals reflect the spatial stochasticity of the system, uncertainties in trends among variables, and uncertainties in instrument bias adjustments. Hypoxic volume is estimated using a novel two-step approach where DO is simulated first, and hypoxic fraction (i.e., the fraction of the water column that is hypoxic) is simulated second using DO as a trend variable.

2. METHODS

2.1. Data and Study Boundaries. We use DO data from LUMCON midsummer sampling cruises conducted between 1985 and 2011. Data for 1998−2008 are retrieved from the National Ocean Data Center, while data for other years are obtained directly from LUMCON.[24] Sampling locations are geo-referenced using the Universal Transverse Mercator (UTM) Zone 15 projection, and bathymetry is determined from a 3-arc-second digital elevation model obtained from the National Oceanic and Atmospheric Administration (NOAA).[25]

During cruises, DO is sampled using one or two types of instruments (Figure 1): a rosette-mounted DO probe and a hand-held DO probe, with the latter capable of being lowered closer to the sea floor.[9,23] (Hereafter, we refer to these instruments as the "rosette sampler" and "handheld sampler".) While instrument technology has changed over time, all instruments were calibrated against Winkler titrations, and postcruise corrections were made as necessary. For sampling events where both instruments were used, data are combined into synthesized profiles.[9] (Here, a "sampling event" refers to the depth profile of data collected at a specific latitude/longitude and time.)

We extract bottom water dissolved oxygen (BWDO) and minimum dissolved oxygen (MinDO) concentrations from the DO profiles. BWDO and MinDO concentrations greater than 7 mg L^{-1} (less than 1% of total samples), are treated as 7 mg L^{-1}, because higher concentrations are outliers representing supersaturation conditions that are not of interest in this study. For sampling events where hypoxia was present, we also extract the fraction of the water column that is within the hypoxic bottom layer (hereafter, bottom water hypoxic fraction, or BWHF) and, where other hypoxic layers are present, the total hypoxic fraction (THF) of all hypoxic layers. Bottom and upper layers of hypoxia are observed at 49% and 12% of sampling events, respectively. Here, we focus on the models for BWDO and BWHF because bottom layer results are most comparable to previous studies.[2,16] The models for MinDO and THF are formulated in the same way and yield comparable results, but for brevity are discussed only in Supporting Information (SI) Section S7.

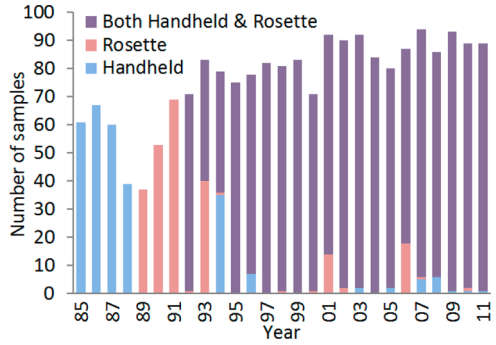


Figure 1. Number of locations sampled during the annual midsummer shelfwide cruises using hand-held and rosette instruments.

In this study, adjustments are made to address biases that arise from the use of different sampling instruments. For cruises when only the rosette sampler was used (Figure 1), one would expect the hypoxic extent to be underestimated because the rosette sampler does not reach as close to the sea floor as the hand-held sampler. We quantify this bias using data from sampling events where both instruments were used together. For these cases, observed BWDO and BWHF are calculated for both the synthesized profile and rosette-only profile. Probabilistic relationships are then developed between the synthesized results and the rosette-only results (Section S1 of the SI). When performing simulations (described below), we adjust the rosette-only observations by sampling from these relationships. A larger bias adjustment is required for the first 38 sampling events in 1991 because the ship's fathometer was not properly calibrated, causing the rosette sampler to be lowered 1.5 m less than it would have been otherwise (N.N. Rabalais, cruise records). For these observations, the probabilistic relationships between synthesized and rosette-only results are again determined based on sampling events where both instruments were functioning properly (as described above) but with the bottom 1.5 m of the rosette-only profiles removed.

Geostatistical modeling is performed on a 5 × 5 km^2 grid of estimation points (Figure 2), covering 342.5−837.5 km UTM easting and 3122.5−3292.5 km UTM northing, reflecting the general extent of sampling. The estimation grid is limited to depths between 3 and 80 m, which are typical of the Louisiana−Texas shelf region where hypoxia occurs.

2.2. Model Formulation. Geostatistical methods provide an effective means to model data that exhibit spatial correlation.[17,26] The efficacy of these methods has been demonstrated in previous environmental analyses of rainfall,

dx.doi.org/10.1021/es400983g | Environ. Sci. Technol. 2013, 47, 9808−9815

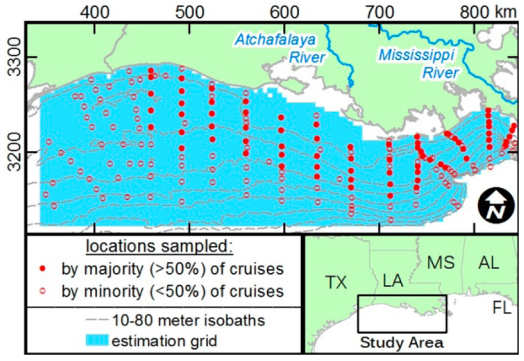


Figure 2. Study area bathymetry, sampling, and estimation locations.

snow depth, soil phosphorus, and water quality indicators.[18,21,27−29] Many of these studies have focused on mapping through spatial interpolation, and comparison studies have demonstrated that geostatistical interpolation outperforms simpler interpolation methods, such as inverse distance weighting.[21,28]

In a geostatistical model, the response variable, $\mathbf{z}$, is represented as a combination of deterministic and stochastic components:[26]

$$\mathbf{z} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\eta} + \boldsymbol{\varepsilon} \quad (1)$$

This formulation is similar to linear regression, but includes an additional term, $\boldsymbol{\eta}$, representing spatially correlated stochasticity, in addition to the more commonly modeled uncorrelated stochasticity, $\boldsymbol{\varepsilon}$. As in linear regression, $\mathbf{X}\boldsymbol{\beta}$ is the portion of $\mathbf{z}$ that can be expressed as a deterministic function of categorical and/or trend variables ($\mathbf{X}$) and their corresponding regression coefficients ($\boldsymbol{\beta}$). If the $\mathbf{X}$ term includes only cruise-specific categorical variables (vectors of zeros and ones that bin the data by cruise), then the model can essentially be used to perform ordinary kriging (OK). If $\mathbf{X}$ also includes trend variables, then it can be used to perform universal kriging (UK).[17]

In this study, the stochastic components ($\boldsymbol{\eta}$ and $\boldsymbol{\varepsilon}$) are found to be well represented by the commonly used exponential covariance function with a nugget effect,[26] defined as follows:

$$Q(z_i, z_j) = Q(h_{i,j}) = \begin{cases} \sigma_\eta^2 + \sigma_\varepsilon^2, & h_{i,j} = 0 \\ \sigma_\eta^2 \exp\left(-\dfrac{h_{i,j}}{r}\right), & h_{i,j} > 0 \end{cases} \quad (2)$$

where Q is the covariance between two observations (z_i and z_j) separated by distance ($h_{i,j}$); ${\sigma_\varepsilon}^2$ and ${\sigma_\eta}^2$ are parameters representing the variances of the stochasticity that is spatially correlated and is not spatially correlated, respectively; and r is a range parameter ($3r$ is approximately the distance at which observations are no longer spatially correlated). We allow for anisotropy in the covariance model by scaling $h_{i,j}$ using parameter α, which represents the ratio of east−west to north−south correlation ranges. Covariance function parameters are estimated using restricted maximum likelihood[30,31] and deterministic component parameters ($\boldsymbol{\beta}$) are estimated using generalized least-squares, as outlined in a previous study by Obenour et al.[9] that used geostatistical methods to explore biophysical drivers of DO depletion.

Geostatistical models are developed for both BWDO and BWHF to determine bottom layer hypoxic area and volume, respectively. The model for BWDO uses all observations, while the model for BWHF uses only observations from locations where BWDO is below the hypoxic threshold. Both models use the UK formulation to take advantage of potential relationships between the response variables and available trend variables.[18]

The deterministic components of the models include two types of trends. First, "constant trends" represent relationships between response and trend variables using regression coefficients that are the same for all cruises. For the BWDO model, potential constant trends include linear and quadratic trends with depth (Depth and Depth^2), easting (Easting and Easting^2), and northing (Northing and Northing^2). On the basis of an examination of model residuals, it was noted that the overall trend between depth and BWDO does not continue for depths greater than 40 m, and so depths of greater than 40 m are treated as 40 m in the final BWDO model development. This approach is justified because model residuals are evenly distributed across all depths when depths greater than 40 m are modified as described here (see Figure S3 of the SI). For BWHF, potential constant trends include linear and quadratic trends with easting, northing, and BWDO.

Second, "cruise-specific trends" (with cruise-specific regression coefficients for Easting) represent relationships that are specific to individual cruises. Allowing these trends was motivated by previous studies[16,23] that indicate the east−west distribution of hypoxia is influenced by alongshore current velocity, which can vary interannually in response to prevailing winds. When selected, these cruise-specific trends modify the constant trend.

To prevent overparameterization of the model, we use only trend variables selected through a geostatistical adaptation of the Bayesian information criterion (BIC),[9,32,33] wherein models with different subsets of trend variables are compared based on their BIC score, and the model with the lowest BIC score is optimal in terms of its explanatory power relative to its complexity. Because of the large number of variables considered in this study, an initial optimal model is first selected among those formed from constant trend variables only. Cruise-specific trend variables are then individually tested relative to that model, and only cruise-specific trends that improve the BIC score are included in the final model.

The deterministic portion of the model also includes categorical variables that bin data by cruise, reflecting the fact that the mean BWDO varies from cruise to cruise. In UK, the categorical variables essentially allow for different "intercepts" (as in linear regression), that shift the trend up or down to best fit the observations from each cruise. In OK, these categorical variables simply allow for a different mean for each cruise.

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

$$\begin{bmatrix} \mathbf{Q}_{oo} & \mathbf{X}_o \\ \mathbf{X}_o^T & \mathbf{0} \end{bmatrix} \begin{bmatrix} \boldsymbol{\Lambda}_y \\ -\mathbf{G}_y \end{bmatrix} = \begin{bmatrix} \mathbf{Q}_{oe,y} \\ \mathbf{X}_e^T \end{bmatrix} \quad (3)$$

where $\mathbf{Q}_{oo}$ is an $n \times n$ covariance matrix for the n observation locations (all cruises), with elements determined from eq 2. Because we did not assume correlation among stochasticity from different cruises, intercruise covariances are assigned a

dx.doi.org/10.1021/es400983g | Environ. Sci. Technol. 2013, 47, 9808−9815

value of zero. Similarly, $\mathbf{Q}_{oe,y}$ is an $n \times m$ covariance matrix of n observation locations and m estimation locations, and the rows of $\mathbf{Q}_{oe,y}$ that correspond to observations from cruises other than cruise y are assigned a value of zero. The matrix $\mathbf{X}_o$ is $n \times p$ and includes the p deterministic variables (trend and categorical) for the observation locations, and the matrix $\mathbf{X}_e$ ($m \times p$) includes the same variables for the estimation locations. The trend variables are normalized to a mean of zero and variance of one. Note that terms with a "y" subscript are cruise-specific.

In eq 3, $\mathbf{\Lambda}_y$ is an $n \times m$ matrix of cruise-specific weights that reflect both the spatial correlation structure and the deterministic trends. Also, $\mathbf{G}_y$ is a $p \times m$ matrix of Lagrange multipliers that can be used with $\mathbf{\Lambda}_y$ to determine location-specific estimation uncertainties. Using the weights, along with the observations $\mathbf{z}_o$ (an $n \times 1$ vector), one can develop estimates of the response variable across the estimation grid:

$$\mathbf{z}^0_{e,y} = \mathbf{\Lambda}_y^T \mathbf{z}_o \tag{4}$$

where $\mathbf{z}^0_{e,y}$ is an $m \times 1$ vector of interpolated BWDO or BWHF values for cruise y.

Although we use a UK model formulation, hypoxic area and volume are not determined by kriging (i.e., spatial interpolation), but instead by developing "*spatially consistent* Monte Carlo simulations"[17] which are often referred to as conditional realizations (CRs). While both kriging and CRs provide equivalent information for individual estimation locations, CRs are necessary to estimate spatially aggregated quantities (e.g., area and volume) probabilistically. A CR is performed by first creating an "unconditional realization" and then "conditioning" it to the observed data and deterministic trend.[17] Unconditional realizations (eq 5) include simulated values at the estimation locations ($\mathbf{z}^u_{e,y}$) as well as at the observation locations ($\mathbf{z}^u_{o,y}$):

$$\begin{bmatrix} \mathbf{z}^u_{e,y} \\ \mathbf{z}^u_{o,y} \end{bmatrix} = \mathbf{C}\left(\begin{bmatrix} \mathbf{Q}_{ee} & \mathbf{Q}^T_{oe,y} \\ \mathbf{Q}_{oe,y} & \mathbf{Q}_{oo} \end{bmatrix}\right)^T \mathbf{u} \tag{5}$$

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

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

$$\mathbf{z}^c_{e,y} = \mathbf{\Lambda}_y^T(\mathbf{z}_o - \mathbf{z}^u_{o,y}) + \mathbf{z}^u_{e,y} \tag{6}$$

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

The CRs are performed in two steps. First, BWDO concentration is simulated across the entire estimation grid. Then, BWHF is simulated over those locations where the simulated BWDO is below the hypoxic threshold. At hypoxic estimation locations, the simulated BWDO values are used as a trend variable in the BWHF model. Simulated values are limited to within realistic ranges (determined to be 0−7 mg L^{-1} for BWDO and 0.001−0.8 for BWHF, based on a review of the observed data). Without these constraints, the estimated hypoxic volume would be 6% smaller, on average, as a result of (unrealistic) negative hypoxic thicknesses.

The two-step CR process is repeated 1000 times for each cruise, resulting in 1000 realizations of both BWDO and BWHF. For each realization of BWDO, the hypoxic area is calculated as the number of estimation locations simulated to be hypoxic multiplied by the grid cell area (25 km^2); and for each realization of BWHF, the hypoxic volume is calculated as the vector of simulated BWHF values multiplied by the corresponding vector of water column depths, all multiplied by the grid cell area. From this ensemble of results, we determine the mean and 95% confidence intervals for the hypoxic area and volume for each cruise.

3. RESULTS

The BWDO and BWHF models include several parameters that characterize the deterministic and stochastic model components (eq 1). Regression coefficients for the BIC-selected trend variables, explaining a portion of the spatial variability in observed BWDO and BWHF, are provided in Table 1. The standard errors of these coefficients are low (i.e., p < 0.05), suggesting these trends are statistically significant. The coefficients for the cruise-specific categorical variables, accounting for year-to-year variability in the responses, are included in Section S3 of the SI. Overall, the deterministic model components explain 28% and 32% of the total (spatial plus interannual) variability in BWDO and BWHF, respectively, while the stochastic components of the models explain the remainder of the spatial variability. If trend variables are omitted (i.e., the OK formulation), then the deterministic components only account for interannual variability, and only 12% and 11% of the total variability in BWDO and BWHF is explained, respectively. A check of the linearity assumption implicit in these deterministic relationships is included in Section S2 of the SI.

Table 1. Regression Coefficients ($\hat{\beta}$) with Standard Errors ($\sigma_{\hat{\beta}}$) for Normalized, BIC-Selected Trend Variables in BWDO and BWHF Models[a,b]

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

[a]Parameters optimized by generalized least squares. [b]c.s.E = cruise specific trend for Easting, n.s.=not selected, n.a.=not available.

The BWDO model includes trends with easting, northing, and depth. The trend between BWDO and easting is quadratic with a minimum at 794 km easting (UTM coordinate, Figure 2), between the Mississippi and Atchafalaya river outfalls, which is reasonable given that these rivers provide the freshwater flows and nutrients that are important for hypoxia formation. In 1998, the only year for which a cruise-specific east−west trend with BWDO is selected, the minimum is shifted to 1111 km easting, outside of the study area, indicating that BWDO concentrations decrease monotonically with easting within the study area. The unique spatial distribution in 1998 has been

dx.doi.org/10.1021/es400983g | Environ. Sci. Technol. 2013, 47, 9808−9815

noted previously, and is generally attributed to unusually persistent eastward currents.[16] A similar BWDO pattern has been noted for 2009,[23] but it was not sufficiently strong to result in a cruise-specific trend in this analysis. The trend between BWDO and northing is linear, and suggests that BWDO concentrations are higher to the south. The trend between BWDO and bathymetry (depth) is quadratic with a minimum at 22 m. The overall deterministic trends in BWDO (Figure S7 of the SI) compare well with spatial patterns in hypoxic frequency, as mapped by Walker and Rabalais.[34]

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

The models also include covariance parameters that characterize their stochastic components. For BWDO, σ_ε^2 and σ_η^2, commonly referred to as the nugget and partial sill, were 0.39 and 2.33 (mg^2 L^{-2}), respectively. Since σ_ε^2 is smaller than σ_η^2, the majority of the stochasticity in the model is spatially correlated. The approximate range of spatial correlation ($3r$, per eq 2) is 94 km in the east−west direction and 53 km in the north−south direction. The greater correlation distance in the east−west direction was expected due to the dominant east−west current pattern.[35,36] For BWHF, the spatial correlation of the stochasticity is somewhat weaker with σ_ε^2 and σ_η^2 similar in magnitude, having values of 0.009 and 0.011 (unit-less), respectively. Also, the correlation range was 63 km in all directions (anisotropy was negligible). Overall, spatially correlated stochasticity accounts for the greatest portion of the variability in both BWDO and BWHF. As discussed in Obenour et al.,[9] the spatial correlation of the stochasticity is consistent with the effects of varying coastal current patterns, influencing the distribution of hypoxia over the spatial scales described above. In general, the correlation ranges are considerably longer than the typical distances between sampling locations (Figure 2), especially on the eastern shelf, suggesting that the sampling network is adequate for resolving the spatially correlated stochasticity of the system.

Using CR, we determined the mean and 95% confidence intervals for hypoxic area and volume of each cruise (Figure 3). The largest estimated hypoxic area was for 1996, but it was not significantly different from 1991, 1993, 1995, 1997, 1999, 2001, 2002, 2007, or 2008 ($p > 0.05$), given the uncertainty in the area estimates. The largest hypoxic volume estimate was for 2008, but it was not significantly different from 1993, 1996, 1999, 2004, or 2007 ($p > 0.05$). Both hypoxic area and volume were lowest in 1988, a drought year.[16]

Two sets of hypoxic area estimates, determined by LUMCON, are also included in Figure 3 for comparison. The first set are the original LUMCON estimates, determined by hand contouring, as described above.[2,19] The second set are revised estimates developed as part of this study; also by hand contouring, but using the updated BWDO values for this study (which are superior because they reflect postcruise DO calibrations). In most cases, the original LUMCON estimates (Figure 3, open squares) overlap the revised estimates (solid squares), and the majority of the LUMCON estimates fall within the 95% confidence intervals determined by this study.

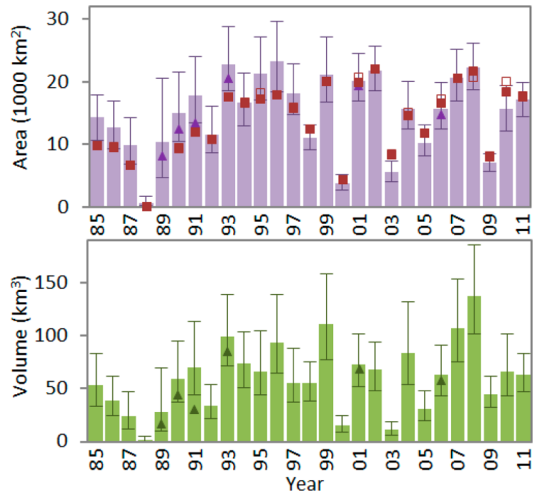


Figure 3. Bottom layer hypoxic extent estimates with 95% confidence intervals by year; estimates prior to making adjustments for instrument bias as triangles; previous LUMCON area estimates as open squares; revised LUMCON area estimates as solid squares.

However, for the first third of the study period (1985−1993), the new estimates are consistently higher than the LUMCON estimates.

LUMCON did not make an estimate for 1989, because only a portion of the shelf was sampled that year. Our method does allow for extent estimates for 1989, and the wide confidence intervals (Figure 3) reflect the relatively large uncertainties. Even though data for 1989 were limited, the overall spatial trends (based on the data from all years) help to constrain the variability in the CRs to within realistic ranges, across the study area. We note that the 1989 estimate presented here is consistent with previous estimates developed from nutrient loading models.[37,38]

There was considerable interannual variability in the spatial distribution of hypoxia and in the thickness of the hypoxic bottom layer (Figure 4). Years with similar hypoxic areas may have very different average hypoxic thicknesses and thus volumes. For example, 2002 and 2008 have similar hypoxic areas, but they are very different in terms of hypoxic thickness,

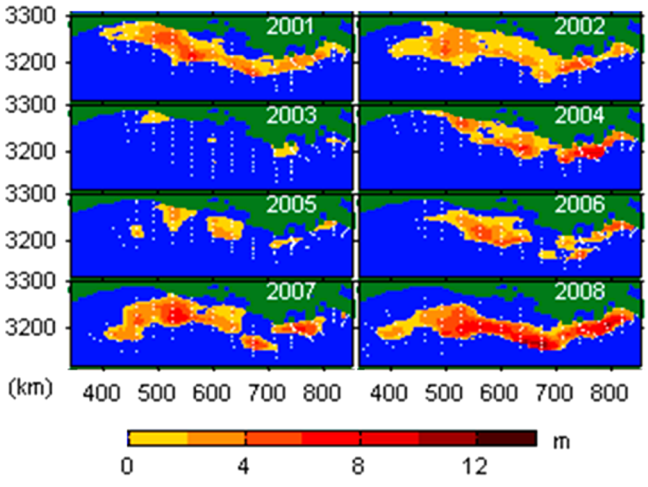


Figure 4. Example maps of estimated bottom layer hypoxic thickness (median values from CRs), 2001−2008; observation locations shown as white dots.

dx.doi.org/10.1021/es400983g | Environ. Sci. Technol. 2013, 47, 9808−9815

such that 2008 has approximately twice the hypoxic volume of 2002. The results indicate that the average bottom layer hypoxic thickness for the 27-year study period was 3.9 m, with the thickest average layers of approximately 6.2 and 6.3 m occurring in 2008 and 2009, respectively. Over the 27-year study period, hypoxic volume is correlated with area ($r^2 = 0.77$). Interestingly, hypoxic area and thickness are also somewhat positively correlated over most years ($r^2 = 0.37$, not including 1998 and 2009, which were subject to unusually strong eastward currents[16,23]), such that volume increases exponentially relative to area. Tabulated results and additional maps showing estimated BWDO, estimated hypoxic thickness, probability of hypoxia, and example CRs for the entire study period are included in Sections S4 and S5 of the SI.

4. DISCUSSION

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

The new hypoxic area estimates can be compared to the previous hypoxic area estimates developed by LUMCON.[2,16] For the 27-year study period, the previous area estimates increased at a highly significant rate of 2.6% per year (360 km^2 yr^{-1}, $p = 0.01$), substantially greater than the 0.9% rate for the new estimates. This dissimilarity is due primarily to differences in estimates for the earlier years of the study period (Figure 3). For 1985−1987, the new estimates are consistently higher than the previous estimates primarily because the geostatistical methodology accounts for the possibility of hypoxia occurring outside the envelopes of the cruises, which were relatively small in these years. For 1990, 1991, and 1993, the new estimates are also higher because of the instrument bias adjustments developed in this study. For 1985−1993 (except 1989, for which there is no previous estimate), the mean geostatistically determined hypoxic area is 39% (3650 km^2) greater than that of the previous estimates.

For 1994−2011, the new area estimates are in general agreement with the previous area estimates. Over this time period, the two sets of estimates are highly correlated ($r^2 = 0.88$), the means of the two data sets are of negligible difference, and only three of the previous estimates (1996, 2003, and 2010) fall outside the 95% confidence intervals of the new estimates (Figure 3). This suggests that when the biasing issues noted above are avoided, there is approximate agreement between the geostatistical and hand-contouring estimates.

The new extent estimates have implications for our understanding of how hypoxia is changing over long temporal scales. Multiple studies using nutrient loads to predict the previous hypoxic area estimates have suggested the Gulf is becoming increasingly susceptible to hypoxia, based on increasing hypoxic area relative to nutrient loading.[11,12,39,40] (Nutrient loading increased greatly in the 1970s but remained relatively stable throughout the study period.)[37,41] Increases in hypoxic susceptibility have also been suggested in other studies of the Gulf and other coastal systems.[42−44] The new hypoxic area estimates, however, exhibit relatively little increase over time, especially when compared to the previous estimates, potentially suggesting less system change during the study period than previously thought. But, the new hypoxic thickness and volume estimates do increase to a greater degree, potentially suggesting a more vertically oriented increase in hypoxic extent. Future studies could focus on recalibrating existing nutrient-loading models to these new hypoxic extent estimates to develop a refined understanding of how Gulf hypoxia may be changing over time.

Uncertainties in the new hypoxic extent estimates, represented by the 95% confidence intervals (Figure 3), reflect the limited spatial scope and resolution of shelfwide cruise sampling. Uncertainties are generally greatest in the earlier years when cruises were smaller and did not always use instruments that reached the sea floor. From 1985 to 1993, the mean relative standard error for hypoxic area was 23%, but it decreased to 11% for 1994−2011. Uncertainties also appear to be larger in years with relatively severe hypoxia (i.e., lower average BWDO and larger hypoxic area), likely because more of the estimation grid is subject to the possibility of hypoxia (in years with higher average BWDO, most of the estimation grid is determined to be well above the hypoxic threshold, such that simulated values rarely fall below the hypoxic threshold). In the future, the modeling framework presented here could be used to evaluate different sampling designs based on how well they constrain estimate uncertainty. The modeling approach could also be used to compare hypoxic extent results and associated uncertainties using hypoxic threshold choices other than 2 mg L^{-1}.

Regardless of the precision of the estimates, it is important to remember that they represent conditions at only certain points in time (i.e., during the shelfwide cruises), and that hypoxic extent can vary substantially throughout the summer due to changes in organic matter production, wind-driven mixing events, and fluctuating current patterns.[10,45,46] As a result, these estimates do not necessarily reflect hypoxic conditions over the entire summer (although back-to-back cruises have demonstrated fairly consistent hypoxic areas under stable weather conditions).[15] More detailed monitoring and biophysical modeling are needed to better understand the short- and long-term dynamics of hypoxia formation, and to mechanistically interpret the temporal variability in the extent estimates presented here.

The primary feature of the geostatistical approach, when compared to more traditional, interpolation-based approaches, is the use of simulations (i.e., CRs).[18] The most obvious benefit is the ability to quantify the uncertainty in extent estimates. A second benefit is that CRs provide more realistic extent estimates than can be derived from kriging (or largely equivalent methods such as Gauss-Markov smoothing[47]) alone. In this study, hypoxic area estimates derived from kriged maps (UK interpolation, Figure S14 of the SI) were

substantially lower than both our CR estimates and the LUMCON estimates. This is because the kriged maps (Figure S9 of the SI) tend to characterize large portions of the estimation grid as slightly above the hypoxic threshold. However, due to the stochasticity of the system, these locations still have some probability of being hypoxic. The CR approach, which samples from the uncertainty in the system, accounts for this possibility, and thus results in larger hypoxic area estimates. This is consistent with Chiles and Delfiner,[17] who argue that CR, rather than kriging, is the more appropriate approach for determining spatially aggregated quantities. A third advantage of the CR approach is that it provides a framework for performing instrument bias adjustments probabilistically, such that adjustment uncertainty is propagated to the hypoxic extent estimates. Finally, CR was fundamental to probabilistically determining hypoxic volume (in addition to area). The two-step CR approach, developed here, can be compared to other volume estimation methods, such as multilayer kriging[21] and three-dimensional CR,[22] which have been applied in Chesapeake Bay. A benefit of our approach is that it allows uncertainty quantification of hypoxic volume within a relatively simple (two-dimensional) geostatistical framework.

This study also demonstrates the benefits of including trend variables within the geostatistical model (i.e., the UK formulation) because the deterministic trends help reduce model uncertainty and result in more realistic extent estimates (Section S6 of the SI). The advantages of UK have been demonstrated previously for Lake Erie and Chesapeake Bay hypoxia,[18,21] but are perhaps even more salient for an open system such as the Gulf shelf. When trends in DO are not modeled, it has been necessary to limit the estimation grid around the bounds of the sampling cruise,[15,48] such that the size of the cruise can potentially bias the inferred hypoxic area. By including trend variables that explain the large-scale spatial patterns in BWDO and BWHF, it is possible to develop realistic CRs of DO across the entire study area, so the same estimation grid can be used for all cruises.

Finally, the results confirm that the hypoxic area on the Louisiana-Texas shelf greatly exceeds the Hypoxia Task Force goal of 5000 km^2 as a five-year running average.[6,7] The most recent five-year period from our study, 2007−2011, has a mean hypoxic area of 16 600 km^2 with a 95% confidence range of 15 100−18 000 km^2. Clearly, additional management measures are required if the hypoxic extent is to be reduced to comply with the Task Force goal.

■ ASSOCIATED CONTENT

Supporting Information

A description of the instrument bias adjustments, a graphical test of linearity for the deterministic components of the models, the intercept parameters for the categorical variables, a set of results maps, tabulated bottom layer hypoxic extent results, and a summary of the models for MinDO and THF. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*Phone: 734-647-2464; fax: 734-647-6730; e-mail: obenour@umich.edu.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We are grateful to the LUMCON personnel who collected and provided quality control for the data used in this study. The work was supported in part by the US EPA STAR Fellowship program, NOAA's Center for Sponsored Coastal Ocean Research grants NA09NOS4780204, NA06NPS4780197, NA09NOS4780204, and NA16OP2670, and the University of Michigan's Graham Sustainability Institute. This is NGOMEX Contribution 176.

■ REFERENCES

(1) Rabalais, N. N.; Turner, R. E.; Wiseman, W. J. Gulf of Mexico hypoxia, aka "The dead zone". *Annu. Rev. Ecol. Syst.* **2002**, *33*, 235−263.

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

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

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

(5) Rabalais, N. N.; Turner, R. E. *Coastal Hypoxia: Consequences for Living Resources and Ecosystems. Coastal and Estuarine Studies 58*; American Geophysical Union: Washington, D.C., 2001.

(6) *Mississippi River/Gulf of Mexico Watershed Nutrient Task Force Action Plan for Reducing, Mitigating, And Controlling Hypoxia in the Northern Gulf of Mexico*; Office of Wetlands, Oceans, and Watersheds, U.S. Environmental Protection Agency: Washington, DC, 2001, http://water.epa.gov/type/watersheds/named/msbasin/history.cfm.

(7) *Mississippi River/Gulf of Mexico Watershed Nutrient Task Force Gulf Hypoxia Action Plan 2008 for Reducing Mitigating, And Controlling Hypoxia in the Northern Gulf of Mexico and Improving Water Quality in the Mississippi River Basin*; U.S. Environmental Protection Agency, Office of Wetlands, Oceans, and Watersheds: Washington, D.C., 2008, http://water.epa.gov/type/watersheds/named/msbasin/actionplan.cfm.

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

(9) Obenour, D. R.; Michalak, A. M.; Zhou, Y.; Scavia, D. Quantifying the impacts of stratification and nutrient loading on hypoxia in the northern Gulf of Mexico. *Environ. Sci. Technol.* **2012**, *46* (10), 5489−5496.

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

(11) Forrest, D. R.; Hetland, R. D.; DiMarco, S. F. Multivariable statistical regression models of the areal extent of hypoxia over the Texas−Louisiana continental shelf. *Environ. Res. Lett.* **2011**, *6* (4), 045002.

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

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

(14) Turner, R. E.; Rabalais, N. N.; Justic, D. Predicting summer hypoxia in the northern Gulf of Mexico: Riverine N, P, and Si loading. *Mar. Pollut. Bull.* **2006**, *52* (2), 139−148.

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

(16) Rabalais, N. N.; Turner, R. E.; Sen Gupta, B. K.; Boesch, D. F.; Chapman, P.; Murrell, M. C. Hypoxia in the northern Gulf of Mexico: Does the science support the plan to reduce, mitigate, and control hypoxia? *Estuaries Coasts* **2007**, *30* (5), 753−772.
(17) Chiles, J.-P.; Delfiner, P. *Geostatistics: Modeling Spatial Uncertainty*; Wiley: New York, 1999.
(18) Zhou, Y.; Obenour, D. R.; Scavia, D.; Johengen, T. H.; Michalak, A. M. Spatial and temporal trends in Lake Erie hypoxia, 1987−2007. *Environ. Sci. Technol.* **2013**, *47* (2), 899−905.
(19) Zhang, H. Y.; Ludsin, S. A.; Mason, D. M.; Adamack, A. T.; Brandt, S. B.; Zhang, X. S.; Kimmel, D. G.; Roman, M. R.; Boicourt, W. C. Hypoxia-driven changes in the behavior and spatial distribution of pelagic fish and mesozooplankton in the northern Gulf of Mexico. *J. Exp. Mar. Biol. Ecol.* **2009**, *381*, S80−S91.
(20) Kimmel, D. G.; Boicourt, W. C.; Pierson, J. J.; Roman, M. R.; Zhang, X. S. A comparison of the mesozooplankton response to hypoxia in Chesapeake Bay and the northern Gulf of Mexico using the biomass size spectrum. *J. Exp. Mar. Biol. Ecol.* **2009**, *381*, S65−S73.
(21) Murphy, R. R.; Curriero, F. C.; Ball, W. P. Comparison of spatial interpolation methods for water quality evaluation in the Chesapeake Bay. *J. Environ. Eng.−ASCE* **2010**, *136* (2), 160−171.
(22) Zhou, Y.; Scavia, D.; Michalak, A. M. Nutrient loading and meteorological conditions explain interannual variability of hypoxia in the Chesapeake Bay, (**submitted manuscript**).
(23) Turner, R. E.; Rabalais, N. N.; Justic, D. Predicting summer hypoxia in the northern Gulf of Mexico: Redux. *Mar. Pollut. Bull.* **2012**, *64* (2), 319−324.
(24) Rabalais, N. N. *Louisiana Hypoxia Surveys*, available from NOAA National Oceanographic Data Center (NODC); http://www.nodc.noaa.gov/General/getdata.html. (Accessed January, 2011.)
(25) NOAA National Geophysical Data Center (NGDC) Coastal Relief Model; http://www.ngdc.noaa.gov/mgg/coastal/startcrm.htm. (Accessed June, 2010.)
(26) Zimmerman, D. L.; Stein, M. Classical Geostatistical Methods. In *Handbook of Spatial Statistics*; Gelfand, A. E., Diggle, P. J., Fuentes, M., Guttorp, P., Eds.; CRC Press: Boca Raton, FL, 2010.
(27) Erickson, T. A.; Williams, M. W.; Winstral, A. Persistence of topographic controls on the spatial distribution of snow in rugged mountain terrain, Colorado, United States. *Water Resour. Res.* **2005**, *41* (4).
(28) Goovaerts, P. Geostatistical approaches for incorporating elevation into the spatial interpolation of rainfall. *J. Hydrol.* **2000**, *228* (1−2), 113−129.
(29) Qian, S. S. Estimating the area affected by phosphorus runoff in an Everglades wetland: A comparison of universal kriging and Bayesian kriging. *Environ. Ecol. Stat.* **1997**, *4* (1), 1−26.
(30) Patterson, H. D.; Thompson, R. Recovery of inter-block information when block sizes are unequal. *Biometrika* **1971**, *58* (3), 545−554.
(31) Zimmerman, D. L. Likelihood-Based Methods. In *Handbook of Spatial Statistics*; Gelfand, A. E., Diggle, P. J., Fuentes, M., Guttorp, P., Eds.; CRC Press: Boca Raton, FL, 2010.
(32) Mueller, K. L.; Yadav, V.; Curtis, P. S.; Vogel, C.; Michalak, A. M. Attributing the variability of eddy-covariance CO_2 flux measurements across temporal scales using geostatistical regression for a mixed northern hardwood forest. *Global Biogeochem. Cycles* **2010**, *24*.
(33) Schwarz, G. Estimating the dimension of a model. *Ann. Stat.* **1978**, *6* (2), 461−464.
(34) Walker, N. D.; Rabalais, N. N. Relationships among satellite chlorophyll a, river inputs, and hypoxia on the Louisiana continental shelf, gulf of Mexico. *Estuaries Coasts* **2006**, *29* (6B), 1081−1093.
(35) Wiseman, W. J.; Kelly, F. J. Salinity variability within the Louisiana coastal current during the 1982 flood season. *Estuaries* **1994**, *17* (4), 732−739.
(36) Zavala-Hidalgo, J.; Morey, S. L.; O'Brien, J. J. Seasonal circulation on the western shelf of the Gulf of Mexico using a high-resolution numerical model. *J. Geophys. Res., [Oceans]* **2003**, *108* (C12), C123389.
(37) Scavia, D.; Donnelly, K. A. Reassessing hypoxia forecasts for the Gulf of Mexico. *Environ. Sci. Technol.* **2007**, *41* (23), 8111−8117.
(38) Turner, R. E.; Rabalais, N. N.; Swenson, E. M.; Kasprzak, M.; Romaire, T. Summer hypoxia in the northern Gulf of Mexico and its prediction from 1978 to 1995. *Mar. Environ. Res.* **2005**, *59* (1), 65−77.
(39) Liu, Y.; Evans, M. A.; Scavia, D. Gulf of Mexico hypoxia: Exploring increasing sensitivity to nitrogen loads. *Environ. Sci. Technol.* **2010**, *44* (15), 5836−5841.
(40) Turner, R. E.; Rabalais, N. N.; Justic, D. Gulf of Mexico hypoxia: Alternate states and a legacy. *Environ. Sci. Technol.* **2008**, *42* (7), 2323−2327.
(41) Goolsby, D. A.; Battaglin, W. A. Long-term changes in concentrations and flux of nitrogen in the Mississippi River Basin, USA. *Hydrol. Processes* **2001**, *15* (7), 1209−1226.
(42) Stow, C. A.; Qian, S. S.; Craig, J. K. Declining threshold for hypoxia in the Gulf of Mexico. *Environ. Sci. Technol.* **2005**, *39* (3), 716−723.
(43) Kemp, W. M.; Testa, J. M.; Conley, D. J.; Gilbert, D.; Hagy, J. D. Temporal responses of coastal hypoxia to nutrient loading and physical controls. *Biogeosciences* **2009**, *6* (12), 2985−3008.
(44) Conley, D. J.; Carstensen, J.; Vaquer-Sunyer, R.; Duarte, C. M. Ecosystem thresholds with hypoxia. *Hydrobiologia* **2009**, *629* (1), 21−29.
(45) Hetland, R. D.; DiMarco, S. F. How does the character of oxygen demand control the structure of hypoxia on the Texas−Louisiana continental shelf? *J. Mar. Syst.* **2008**, *70* (1−2), 49−62.
(46) Wang, L. X.; Justic, D. A modeling study of the physical processes affecting the development of seasonal hypoxia over the inner Louisiana−Texas shelf: Circulation and stratification. *Cont. Shelf Res.* **2009**, *29* (11−12), 1464−1476.
(47) Emery, W. J.; Thomson, R. E. *Data Analysis Methods in Physical Oceanography*, 2nd and rev. ed.; Elsevier: Amsterdam/New York, 2001.
(48) NOAA Gulf of Mexico Hypoxia Watch: Hypoxia Contour Process; http://www.ncddc.noaa.gov/hypoxia/. (Accessed October, 2012.)

Supporting Information

for

"A retrospective analysis of mid-summer hypoxic area and volume in the northern Gulf of Mexico, 1985-2011"

by

Daniel R. Obenour*, Donald Scavia, Nancy N. Rabalais, R. Eugene Turner, and Anna M. Michalak

consisting of

7 sections with 16 figures and 4 tables in 17 pages.

*corresponding author, obenour@umich.edu

S1. Instrument adjustment

Using data from sampling events where both the rosette and handheld samplers were deployed, we developed relationships between the hypoxic conditions (BWDO and hypoxic thickness) derived from the synthesized data (both instruments) and the hypoxic conditions derived from rosette-only data. Figure S1 presents BWDO values derived from the synthesized data (S_{BO}) versus BWDO values from the rosette-only data (R_{BO}). We divided the data into two different categories (blue and red, Figure S1), where the blue data meet the following criterion:

$$S_{BO} - R_{BO} > 2\sigma_{\varepsilon} \qquad \text{eq S1}$$

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

$$\hat{S}_{BO,blue} = 0.973R_{BO,blue} + \epsilon \qquad \text{eq S2}$$

The remaining (red) data can be modeled as a uniform distribution between zero and the threshold criterion used in eq S1:

$$\hat{S}_{BO,red} \sim \mathrm{U}(0, [R_{BO,blue} - 2\sigma_{\varepsilon}]) \qquad \text{eq S3}$$

Using these relationships, we can simulate values of S_{BO} for rosette-only sampling events. (Each conditional realization is assigned a unique set of simulated values.) For R_{BO} less than $2\sigma_{\varepsilon}$ mg L^{-1}, eq S2 always applies. For R_{BO} greater than $2\sigma_{\varepsilon}$ mg L^{-1}, eq S2 is applied at an 88.9% probability and eq S3 at an 11.1% probability. These percentages reflect the actual partitioning of the data as presented in Figure S1. From a physical perspective, application of eq S3 represents situations where there is a thin, high-density, bottom layer that is not reached by the rosette. Conversely, the application of eq S2 implies that the rosette did reach the bottom-most layer of water. When performing simulations, we do not sample from the error term (ϵ) in eq S2, as this variability is expected to be primarily reflective of the microvariability already accounted for within the covariance model.

For hypoxic thickness, the relationship between the synthesized and rosette-only data is somewhat simpler (Figure S2). Here, the relationship between synthesized thickness (S_{Th}) and rosette thickness (R_{Th}) can be approximately modeled using a simple linear regression with normally distributed residuals (the units of the equation are meters):

$$\hat{S}_{Th} = R_{Th} + 0.82 + \epsilon \qquad \text{where ... } \epsilon \sim \mathrm{N}(0,0.36) \qquad \text{eq S4}$$

When performing simulations, we do sample from the error term (ϵ) of eq S4, as this error is expected to be primarily reflective of the variability in the maximum rosette sampling depth, rather than the natural variability in the thickness of the hypoxic layer. For observations that are not hypoxic based on the rosette measured BWDO, but become hypoxic when performing the instrument adjustment for BWDO (R_{BO} > 2 mg L^{-1} and $\hat{S}_{BO}$ < 2 mg L^{-1}), we multiply $\hat{S}_{Th}$ (which

is 0.82 + ϵ in this case) by a sample from a standard uniform distribution because it is unclear what portion of the offset is hypoxic. Although more realistic, this additional step has a negligible impact on results.

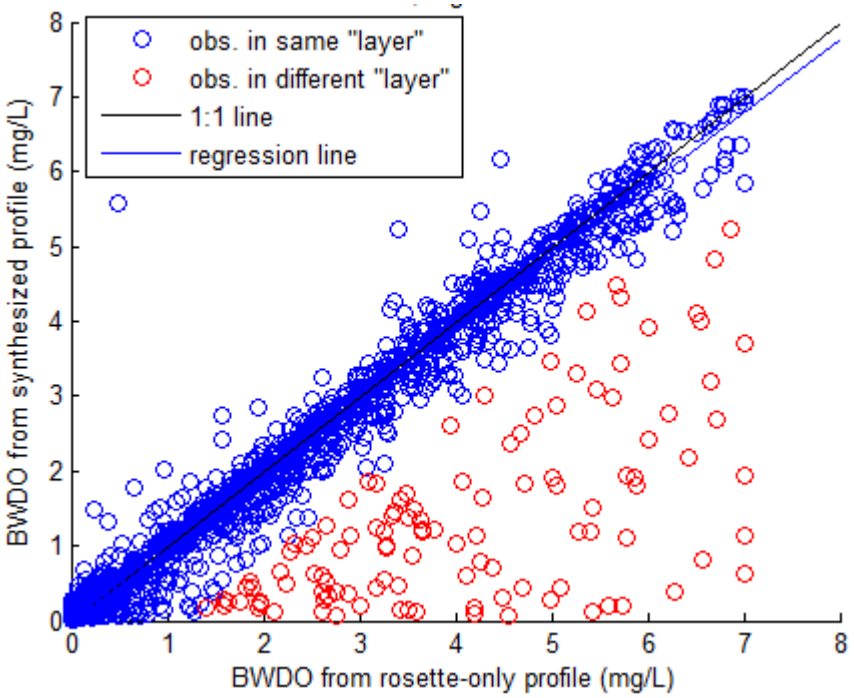


Figure S1: BWDO from synthesized data (S_{BO}) vs. BWDO from rosette instrument only (R_{BO})

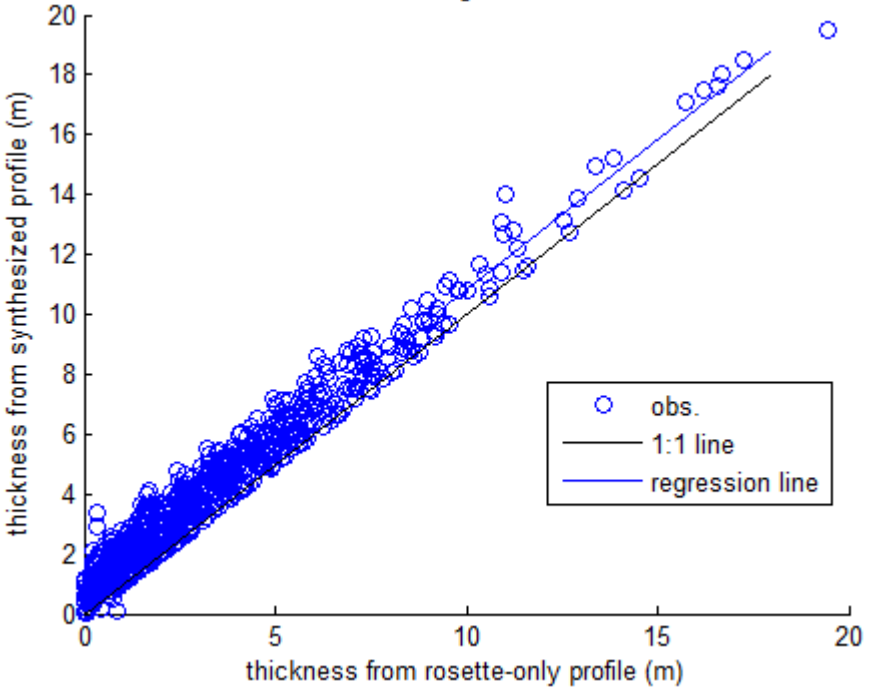


Figure S2: Hypoxic thickness from synthesized data (S_{Th}) vs. hypoxic thickness from rosette instrument (R_{Th})

As described in the primary text, in 1991 a larger bias adjustment was required for the first 38 sampling events because the ship's fathometer was not functioning correctly, causing the rosette sampler to be lowered 1.5 meters less than it would have been otherwise. The same type of adjustment was performed for these events, using eq's S5, S6 and S7, which are analogous to eq's S2, S3, and S4, respectively. For $R_{1.5,BO}$ greater than $2\sigma_\varepsilon$ mg L^{-1}, eq S2 is applied at an 67.5% probability and eq S3 at an 32.5% probability.

$$\hat{S}_{BO,blue} = -0.163 + 0.967R_{1.5,BO,blue} + \epsilon \qquad \text{eq S5}$$

$$\hat{S}_{BO,red} \sim \mathrm{U}(0, (R_{1.5,BO,blue} - 2\sigma_\varepsilon)) \qquad \text{eq S6}$$

$$\hat{S}_{Th} = R_{1.5,Th} + 2.3 + \epsilon \qquad \text{where ... } \epsilon \sim \mathrm{N}(0,0.39) \qquad \text{eq S7}$$

S2. Test of linearity assumption for deterministic trends

In Figures S3 and S4, we plot the residuals (stochastic portion) of the UK models for BWDO and hypoxic fraction, respectively, versus each of the trend variables used in these models. Because the residuals are generally evenly distributed around zero throughout the ranges of the trend variables, the linear model formulation appears reasonable. Note that the models do include nonlinear transformations of the trend variables (i.e. depth-squared) but they are incorporated within a linear modeling framework.

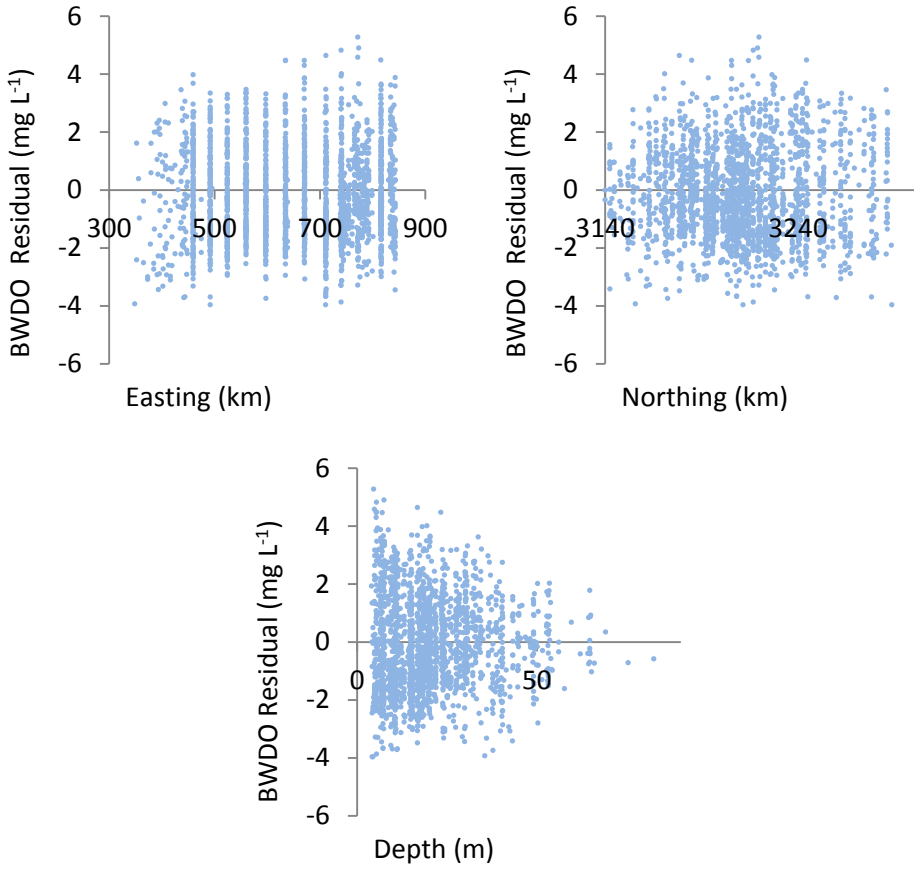


Figure S3: BWDO residuals (stochastic portion of UK model) vs. covariates

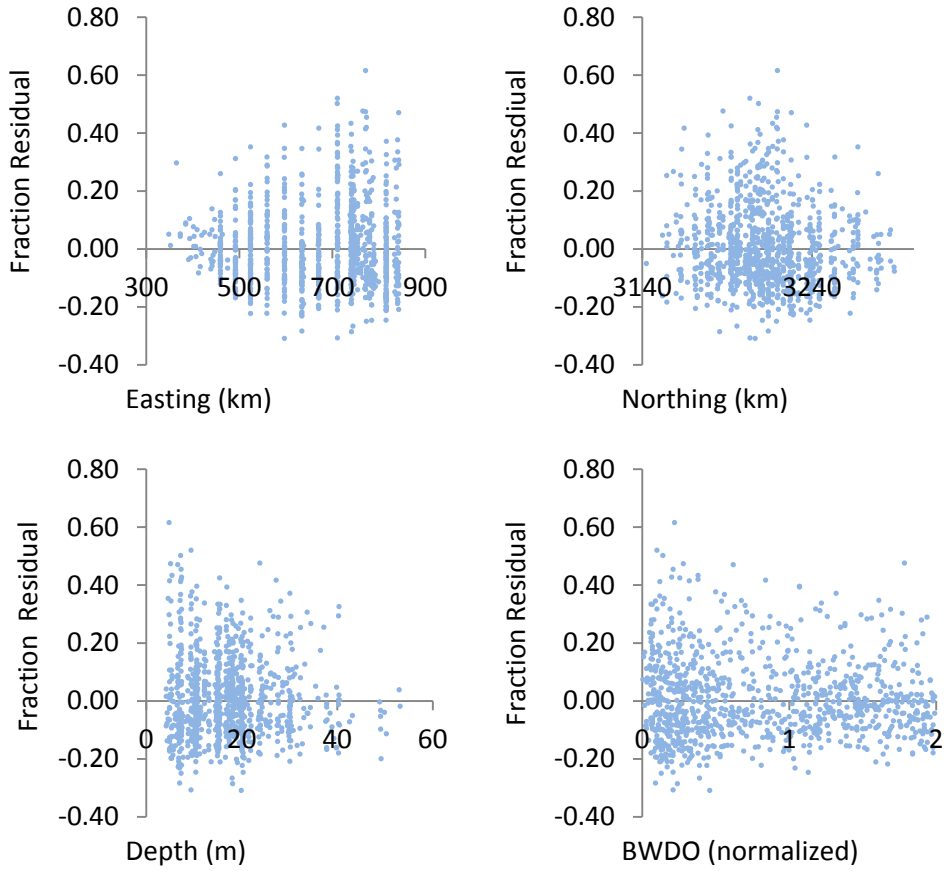


Figure S4: BWHF residuals (stochastic portion of UK model) vs. covariates

S3. Intercept values for BWDO and BWHF models

Table S1: Annual, cruise-specific intercept values for BWDO and BWHF models

year	BWDO (mg L^{-1})	BWHF (-)
1985	2.43	0.188
1986	2.71	0.179
1987	3.13	0.161
1988	5.40	0.096
1989	3.30	0.111
1990	2.84	0.164
1991	2.71	0.145
1992	2.76	0.141
1993	1.91	0.200
1994	2.30	0.213
1995	1.94	0.184
1996	1.65	0.209
1997	2.03	0.189
1998	2.71	0.180
1999	1.82	0.306
2000	4.19	0.240
2001	2.13	0.221
2002	1.97	0.188
2003	3.63	0.149
2004	2.22	0.322
2005	2.99	0.229
2006	2.51	0.217
2007	2.01	0.292
2008	1.84	0.262
2009	3.82	0.296
2010	2.10	0.269
2011	2.39	0.202

S4. Tabulated bottom layer hypoxic area and volume estimates

Table S2 tabulates the geostatistical extent estimates presented graphically in Figure 3 of the main text (based on CRs from UK model formulation)

Table S2: Bottom layer hypoxic areas and volume results

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

S5. Results maps (based on CRs from UK model formulation)

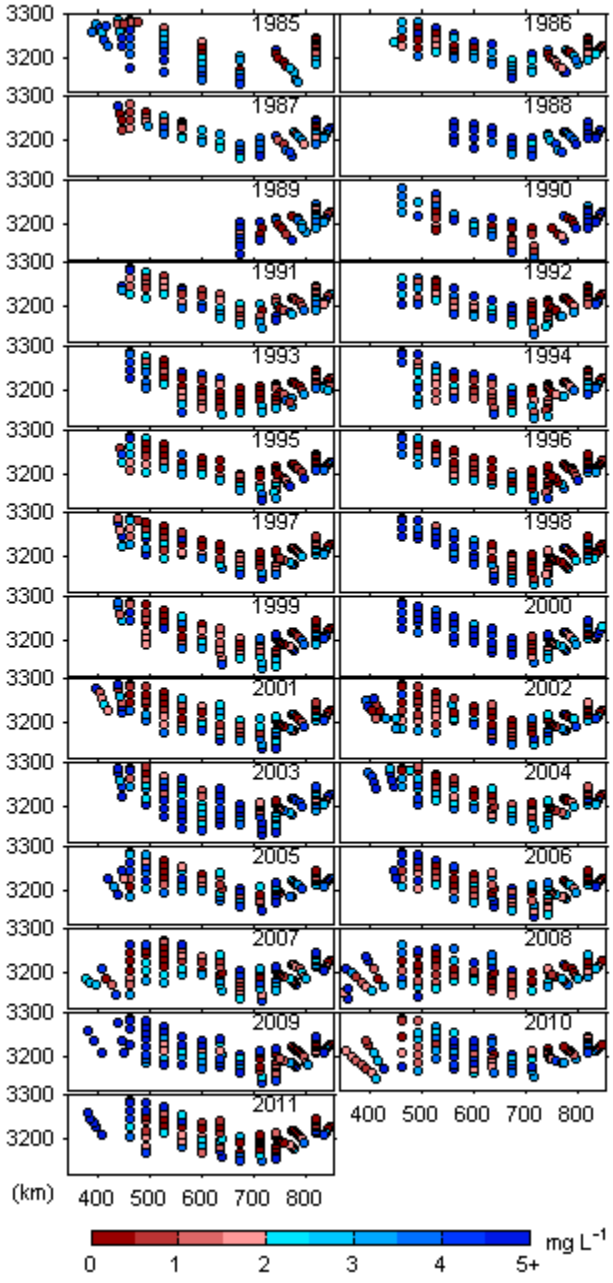


Figure S5: Observed BWDO concentration

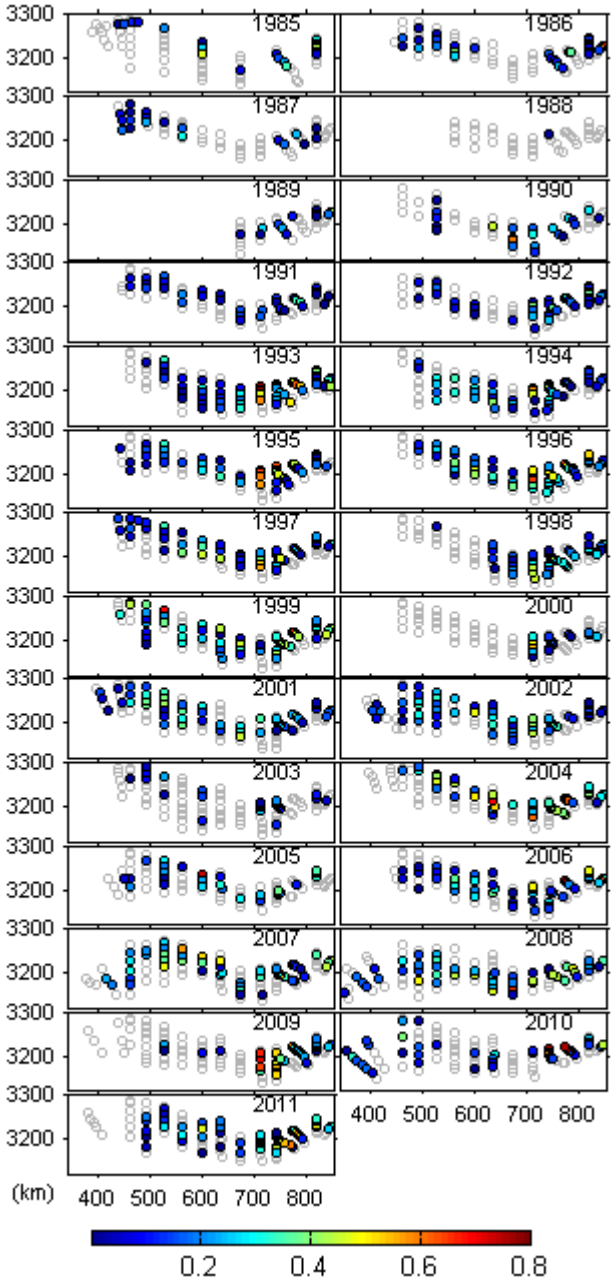


Figure S6: Observed BWHF

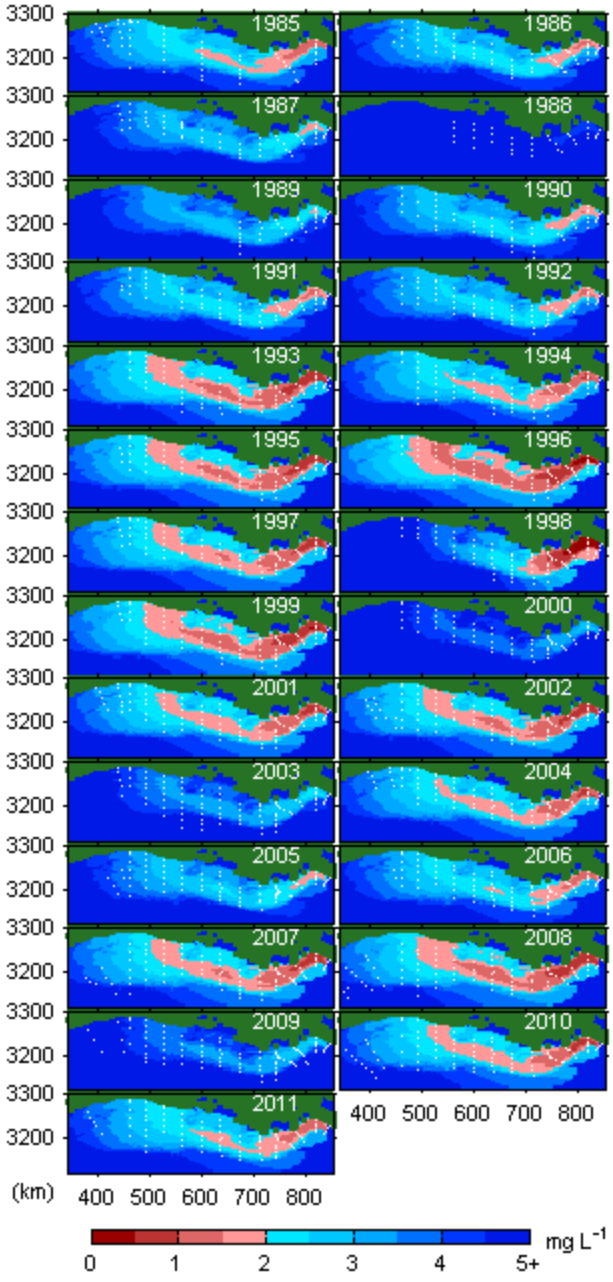


Figure S7: BWDO deterministic trend (note that the spatial pattern is the same for all years, except 1998)

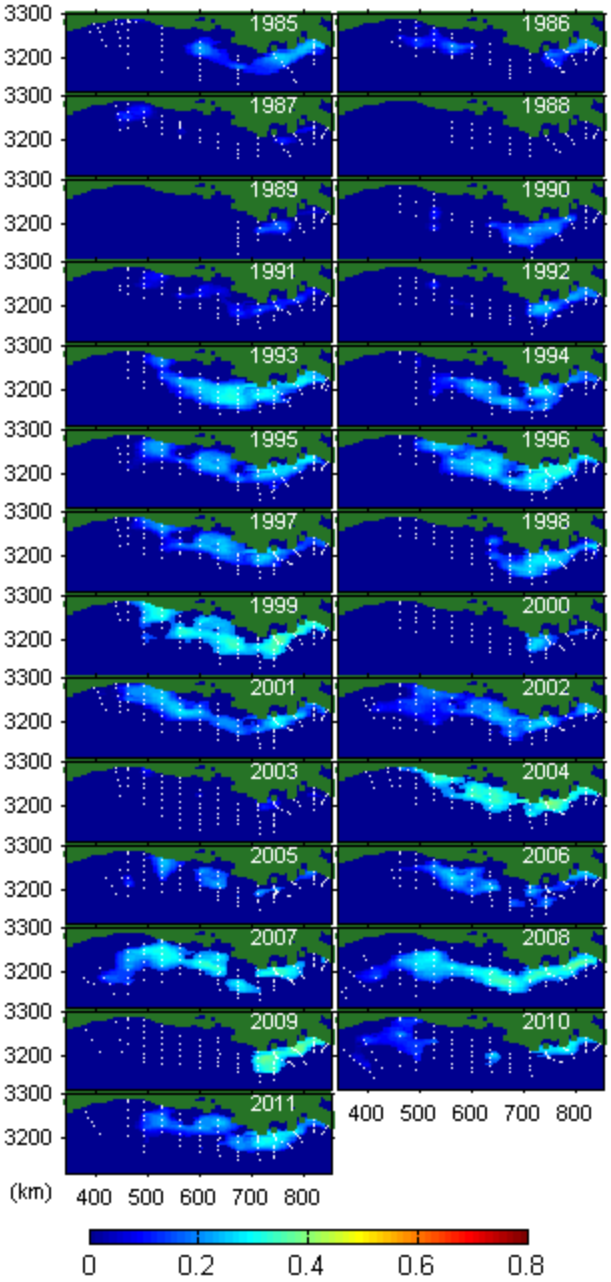


Figure S8: BWHF deterministic trend (using kriged BWDO as a trend variable)

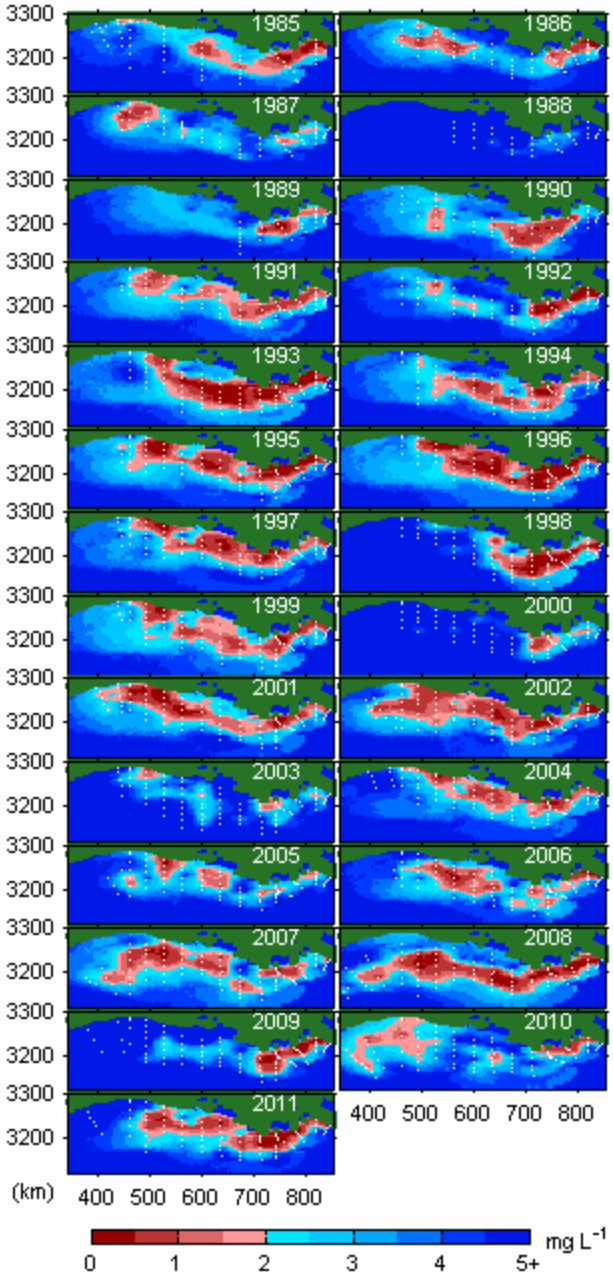


Figure S9: Median BWDO concentration from CR (note that these results are the same as kriged results, except they also include the instrument bias adjustment, which was implemented through the CR process)

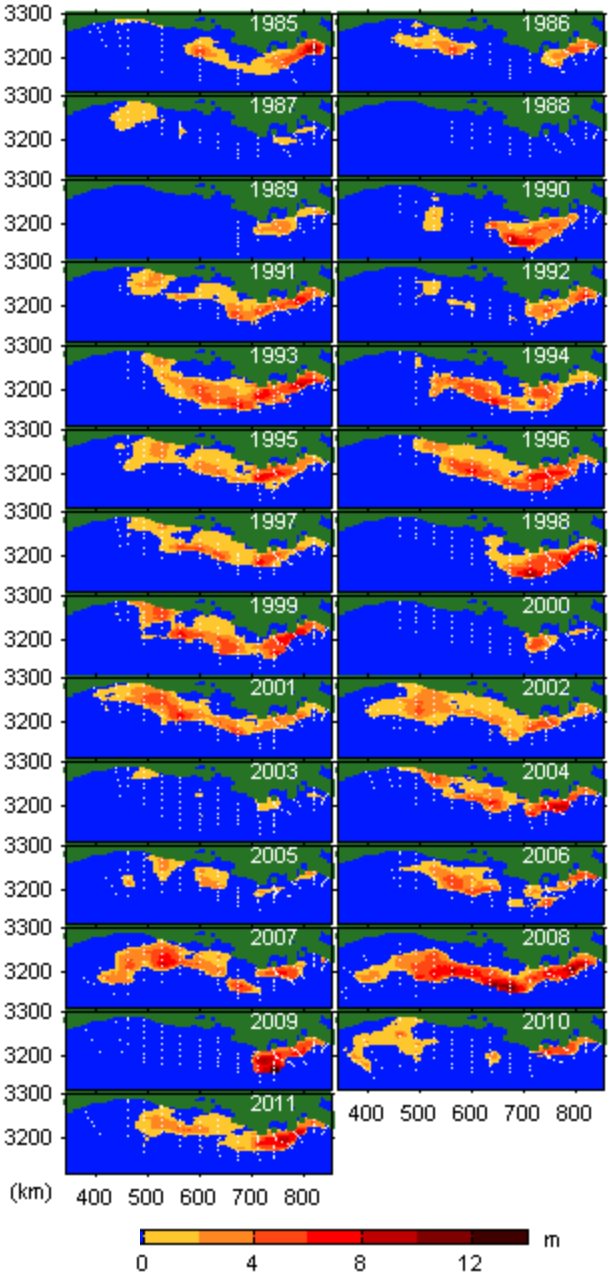


Figure S10: Median BW hypoxic thickness from CR

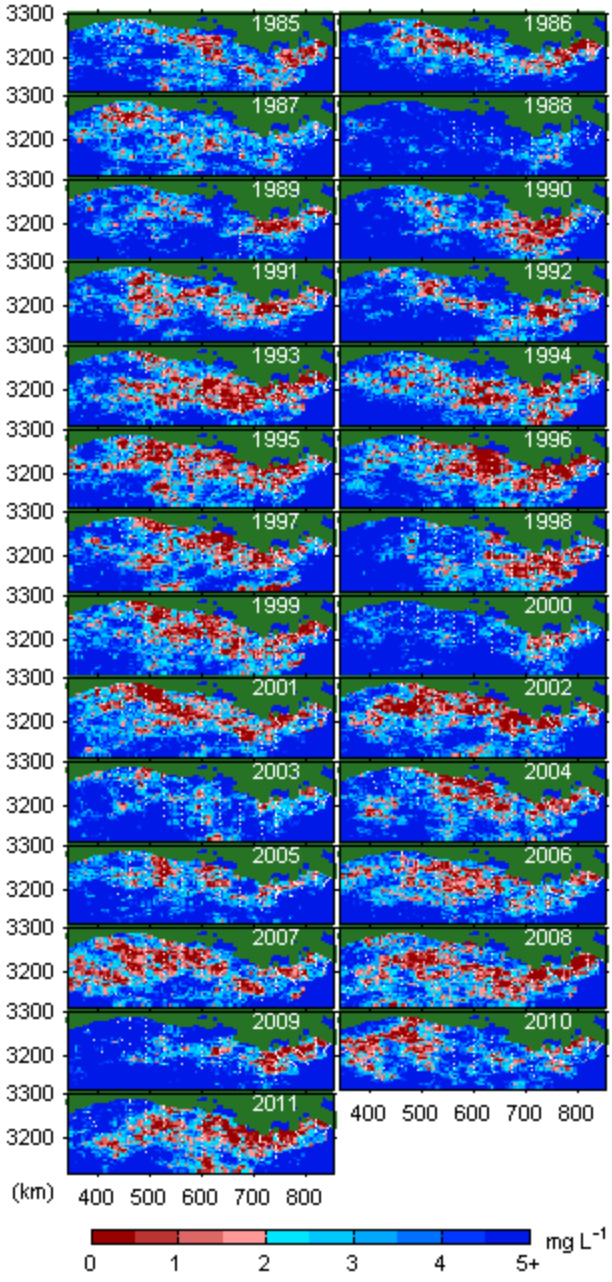


Figure S11: Example conditional realizations of BWDO concentration

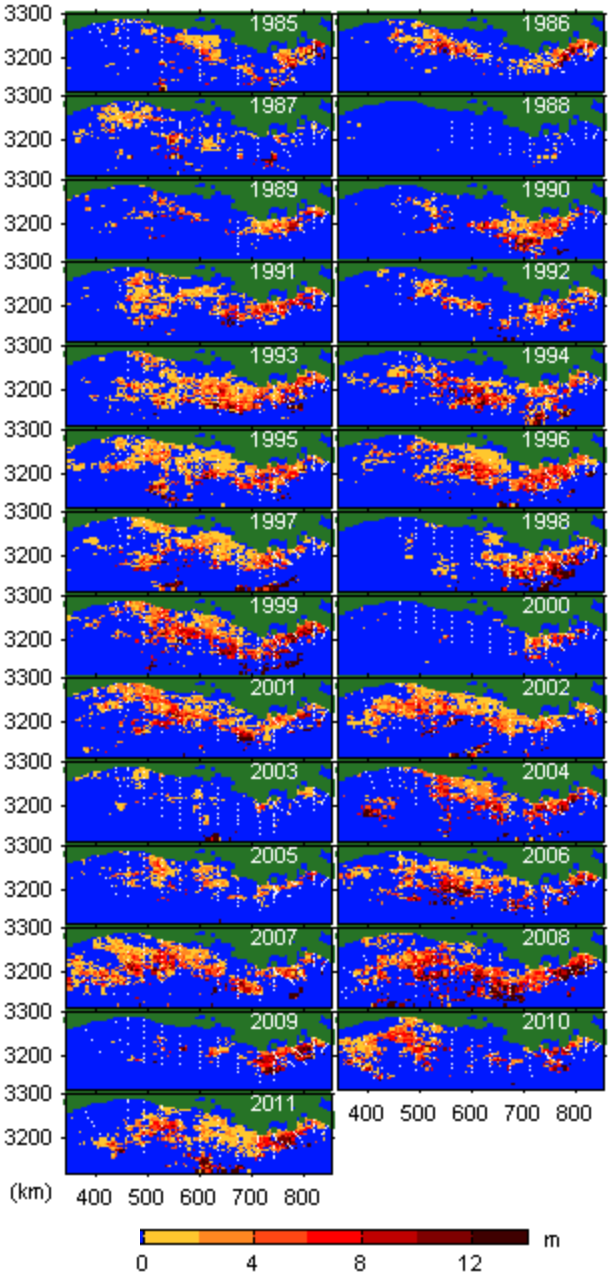


Figure S12: Example conditional realizations of BW hypoxic thickness

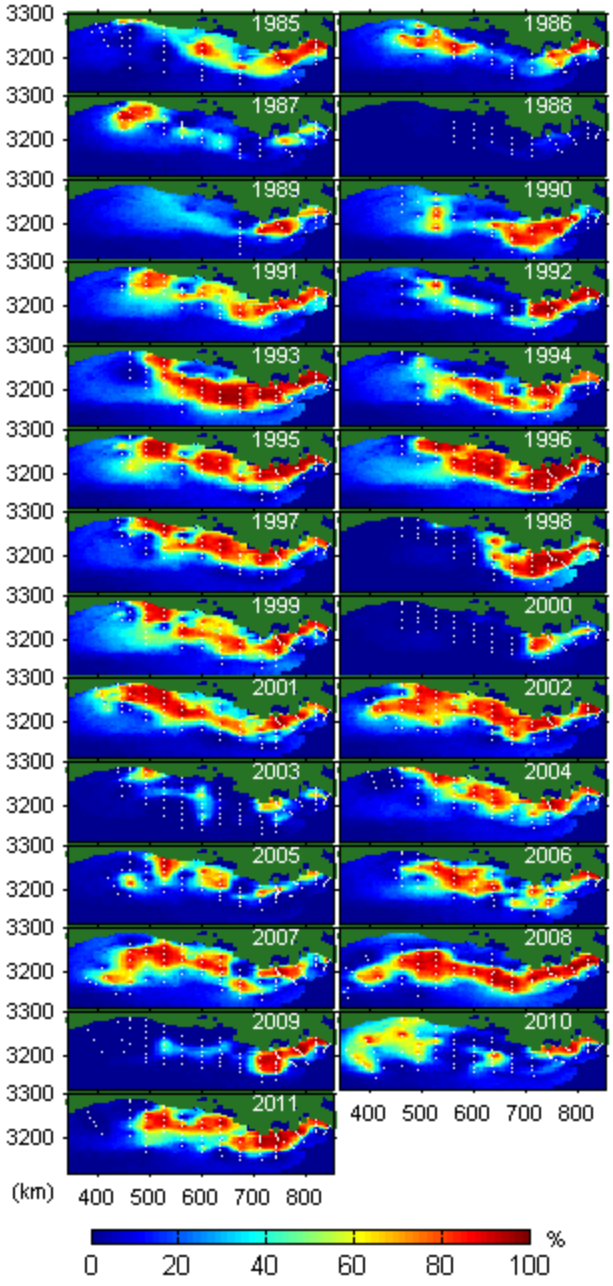


Figure S13: Probability of hypoxia

S6. Comparison of bottom layer extent results using different methods

The hypoxic extent can be determined using a variety of different interpolation and simulation-based methods. Figure S14 compares the preferred area estimates (CRs from UK, with instrument adjustment, Table S2) to areas inferred from less-optimal methods. This comparison includes interpolation ('kriged') estimates, which are consistently lower than the results determined from other methods, as described in the main text.

This comparison also includes estimates developed using CRs from an OK formulation (without trend variables). The "OK" estimates have average confidence intervals more than twice as wide as those from UK; and the OK hypoxic area and volume estimates are 53% and 121% greater than the UK estimates, respectively. OK tends to over-estimate the extent of hypoxia outside of the sampling cruise envelope, because unlike UK, OK does not use trend variables to represent large-scale spatial patterns in DO and hypoxic fraction. These trends (Figures S7 and S8) generally indicate that conditions become less hypoxic as one moves away from the most intensively sampled areas of the shelf.

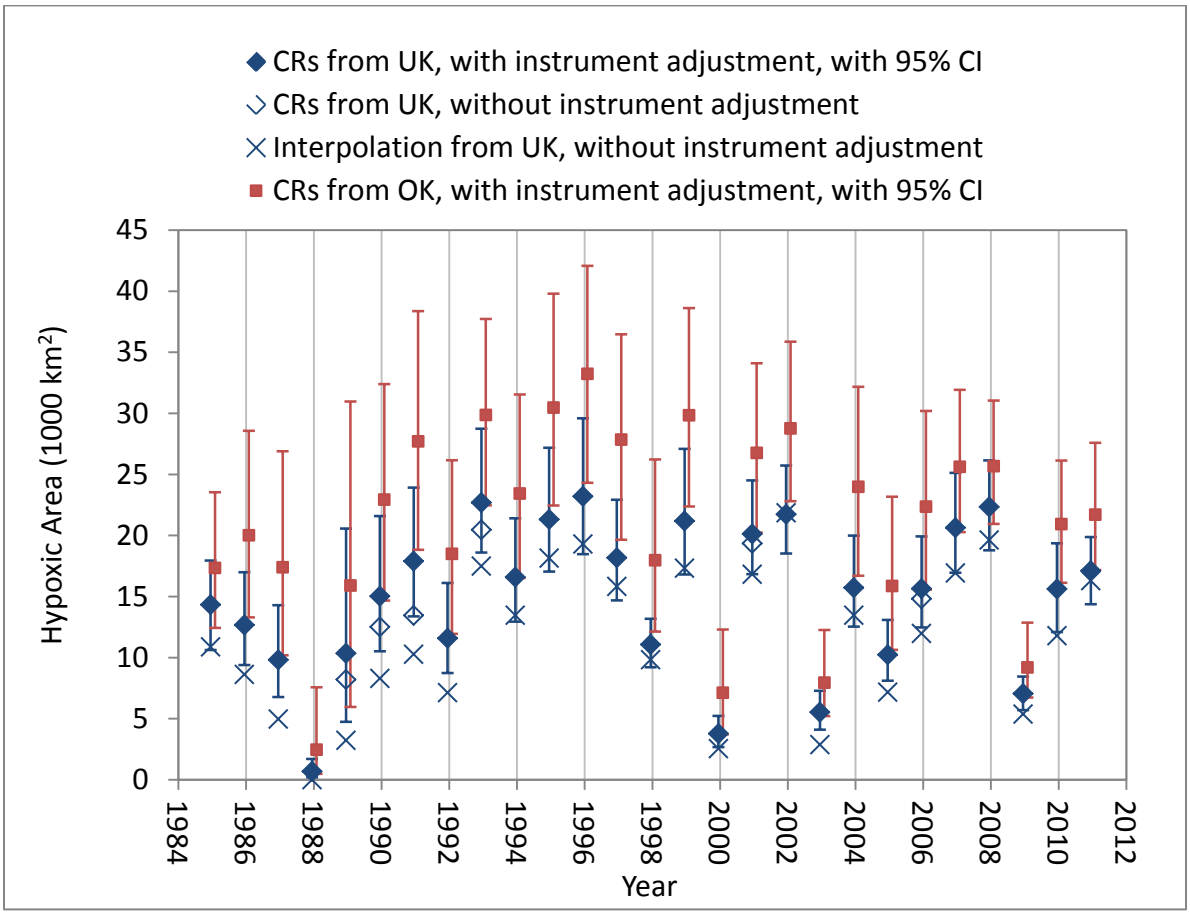


Figure S14: Hypoxic extent estimates developed using different methods

S7. Models for MinDO and THF

In the main text, we describe the models for BWDO and BWHF, which yield the area and volume of the hypoxic bottom layer. However, models can also be developed for the minimum dissolved oxygen (MinDO) and total hypoxic fraction (THF) which yield the total hypoxic area and volume. These models account for layers of hypoxia existing higher (i.e., suspended) in the water column. The parameterization and results of these models are provided below. Table S3 corresponds to Table 1 in the primary text. Table S4 presents the total hypoxic area and volume results (CR+UK methodology), corresponding to Table S2. Figures S15 and S16 present maps of expected MinDO and total hypoxic thickness, corresponding to Figure S9 and S10.

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

Table S3: MinDO and THF model parameters

Variable	MinDO (mg L^{-1})		THF	
	β	σ_β	β	σ_β
E	-0.74	0.09	n.s.	
E^2	0.32	0.07	n.s.	
N	-0.45	0.09	-0.0054	0.007
N^2			-0.0164	0.006
D	-2.45	0.18	n.a.	
D^2	2.39	0.16	n.a.	
O	n.a.		-0.080	0.004
O^2	n.a.		n.s.	
c.s. E 1998	-1.29	0.43	n.s.	
c.s. E 2010	1.04	0.33	n.s.	

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

Table S4: Total hypoxic area and hypoxic volume results*

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

*Note that these results are for total hypoxic area and volume. They are different from the bottom-layer results (Table S2) which were the focus of the primary text.

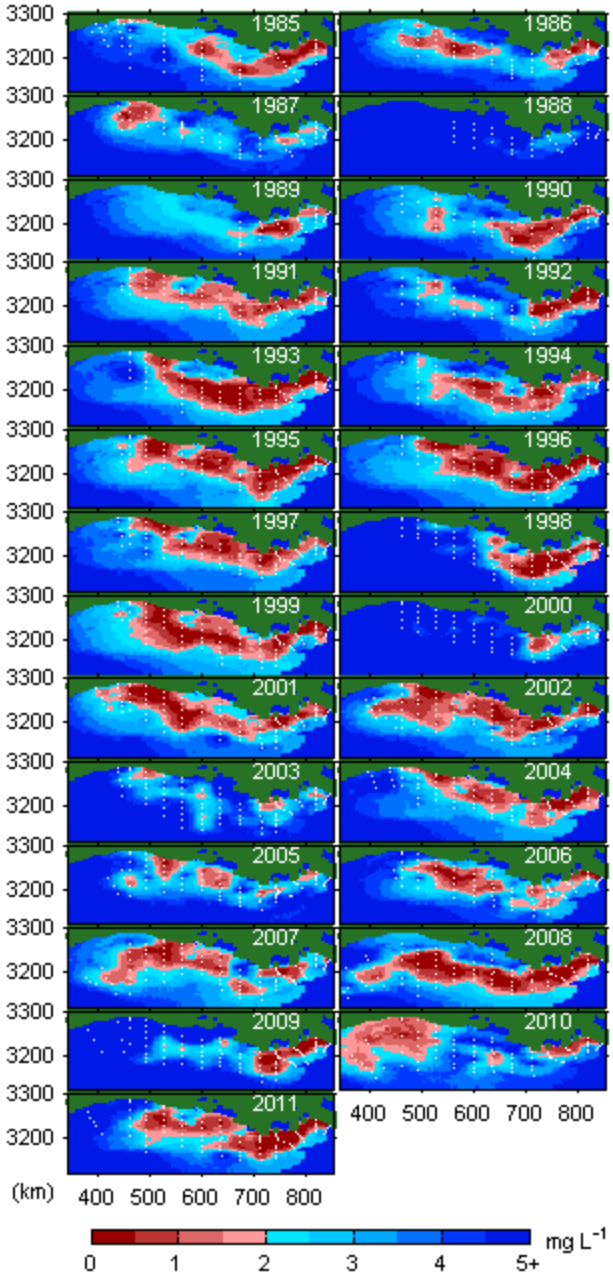


Figure S15: Median MinDO concentration

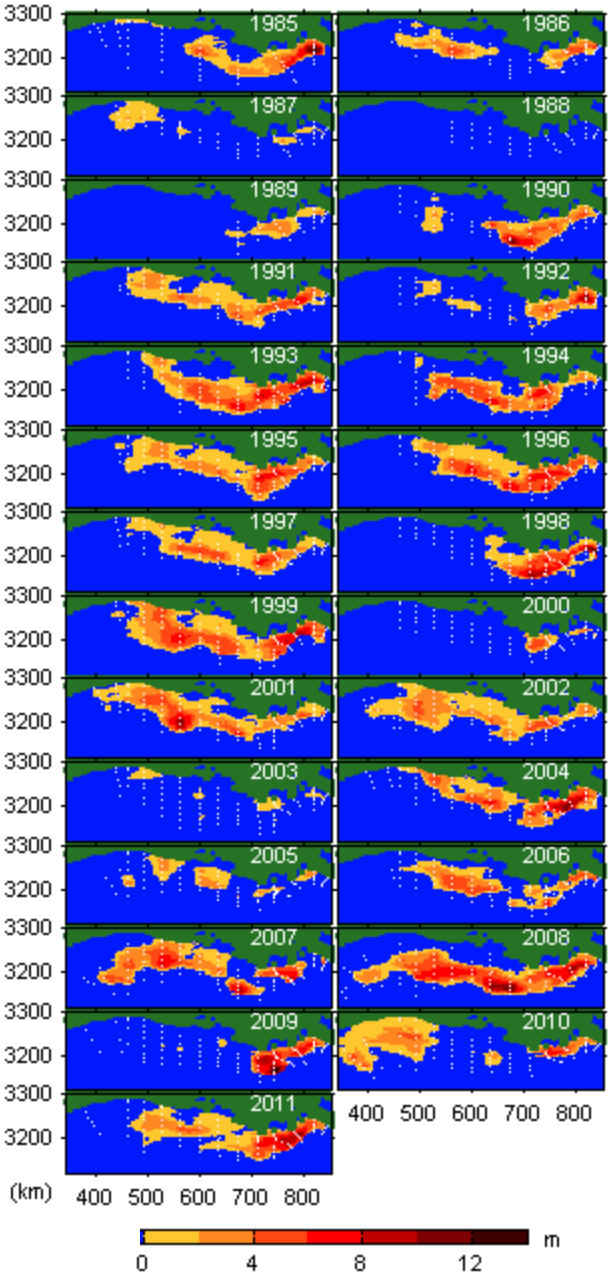


Figure S16: Median total hypoxic thickness