Ensemble modeling informs hypoxia management in the northern Gulf of Mexico

Donald Scavia[a,b,1], Isabella Bertani[a], Daniel R. Obenour[c], R. Eugene Turner[d], David R. Forrest[e], and Alexey Katin[c]

[a]Graham Sustainability Institute, University of Michigan, Ann Arbor, MI 48104; [b]School for Environment and Sustainability, University of Michigan, Ann Arbor, MI 48104; [c]Department of Civil, Construction, and Environmental Engineering, North Carolina State University, Raleigh, NC 27695; [d]Department of Oceanography and Coastal Sciences, College of the Coast and Environment, Louisiana State University, Baton Rouge, LA 70803; and [e]Virginia Institute of Marine Science, College of William and Mary, Gloucester Point, VA 23062

Edited by Jonathan J. Cole, Cary Institute of Ecosystem Studies, Avon, NC, and approved July 7, 2017 (received for review March 31, 2017)

A large region of low-dissolved-oxygen bottom waters (hypoxia) forms nearly every summer in the northern Gulf of Mexico because of nutrient inputs from the Mississippi River Basin and water column stratification. Policymakers developed goals to reduce the area of hypoxic extent because of its ecological, economic, and commercial fisheries impacts. However, the goals remain elusive after 30 y of research and monitoring and 15 y of goal-setting and assessment because there has been little change in river nitrogen concentrations. An intergovernmental Task Force recently extended to 2035 the deadline for achieving the goal of a 5,000-km^2 5-y average hypoxic zone and set an interim load target of a 20% reduction of the spring nitrogen loading from the Mississippi River by 2025 as part of their adaptive management process. The Task Force has asked modelers to reassess the loading reduction required to achieve the 2035 goal and to determine the effect of the 20% interim load reduction. Here, we address both questions using a probabilistic ensemble of four substantially different hypoxia models. Our results indicate that, under typical weather conditions, a 59% reduction in Mississippi River nitrogen load is required to reduce hypoxic area to 5,000 km^2. The interim goal of a 20% load reduction is expected to produce an 18% reduction in hypoxic area over the long term. However, due to substantial interannual variability, a 25% load reduction is required before there is 95% certainty of observing any hypoxic area reduction between consecutive 5-y assessment periods.

ensemble modeling | hypoxia | Gulf of Mexico | nitrogen-loading targets

A large region of low-dissolved-oxygen (DO) bottom waters (hypoxia; DO < 2 mg·L^{-1}) has formed nearly every summer in the Gulf of Mexico for at least the last three decades (1–3). Hypoxia forms because of respiration of organic matter in bottom waters and a vertically stratified water column restricting reaeration (4–6). Both factors are related to the outflow of freshwater and nutrients from the Mississippi River Basin, which typically peaks in March through May. Nutrients stimulate phytoplankton production, much of which settles in early summer, and bottom water DO (BWDO) is consumed during its decomposition. River outflows create a fresher, warmer surface layer above a colder, saltier bottom layer, limiting the vertical diffusion of DO (2).

Policymakers developed hypoxia reduction goals because of ecological, economic, and commercial fisheries impacts (7–10). However, the goals remain elusive after >30 y of research and monitoring (3, 11) and >15 y of assessment and goal-setting (5, 12–16). A Gulf Task Force recently agreed to retain the 5-y moving average goal of a 5,000-km^2 hypoxic zone, but extended the deadline from 2015 to 2035 (17). The timeframe was reset because the 2015 hypoxic zone was 16,760 km^2 and the most recent 5-y average was 14,024 km^2—greater than the long-term average of 13,751 km^2 (1). Missing the goal was not unexpected because the 5-y average load of late spring nitrogen (N) from the Mississippi River remains similar to the 1980–1996 baseline (Fig. 1) and far from the 45% reduction the Task Force recommended (17). To support its adaptive management process and to ensure progress, the Task Force established a milestone of a 20% reduction in N load by 2025 (17). Two questions emerged during their deliberations: (*i*) Has the required load reduction changed? and (*ii*) what hypoxic area would result from meeting the 20% load reduction milestone?

We explore both questions with an ensemble of four different models developed independently to produce annual forecasts (Fig. 2) and to evaluate load reduction scenarios. The University of Michigan (U-M) model is a one-dimensional Bayesian adaptation of a river model that predicts DO downstream from sources of nutrient-stimulated organic matter, accounting for its decomposition and DO reaeration. The North Carolina State University (NCSU) Bayesian biophysical model is a mass-balance model of the eastern and western Gulf shelf that predicts DO in terms of nutrient-stimulated primary production, wind-driven transport, sedimentation, decomposition, sediment DO demand, and reaeration. The Louisiana State University/Louisiana University Marine Consortium (LSU/LUMCON) and the Virginia Institute of Marine Sciences (VIMS) models are linear regressions based on different assumptions and statistically identified relationships for the response of hypoxia to nutrients and stratification. Further details are in *Methods*.

The benefits of using multiple models are being increasingly recognized (18–23). Because each model is a different representation of the real world, the advantages of multiple models include viewing problems from different conceptual perspectives, testing the effects of different assumptions, and reducing decision risk (22). The integration of the models provides increased confidence when they agree and helps identify knowledge gaps when they do not. However, multimodel efforts addressing policy-relevant natural resource issues often lack rigorous frameworks to quantify how model uncertainties propagate into formal ensembles. As for climate and weather forecasting (24), there is a need to develop

Significance

The number of coastal hypoxia areas is spreading worldwide, with severe environmental and societal impacts. The second-largest hypoxic zone occurs in the northern Gulf of Mexico, where anthropogenic nutrient load is a key driving factor, as in many coastal waters. We address policy-relevant questions raised by Gulf stakeholders and decision-makers using an ensemble approach that integrates results from multiple models. Through development of a rigorous framework to propagate intramodel and intermodel uncertainty into the ensemble, we provide policymakers with the response of hypoxic area to a range of different nitrogen load reduction scenarios, with corresponding probabilistic statements that allow for quantitative risk assessment of alternative policy strategies.

Author contributions: D.S., D.R.O., and R.E.T. designed research; D.S., I.B., D.R.O., R.E.T., D.R.F., and A.K. performed research; I.B., D.R.O., R.E.T., D.R.F., and A.K. analyzed data; and D.S., I.B., D.R.O., and R.E.T. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Freely available online through the PNAS open access option.

[1]To whom correspondence should be addressed. Email: scavia@umich.edu.

SUSTAINABILITY SCIENCE

ENVIRONMENTAL SCIENCES

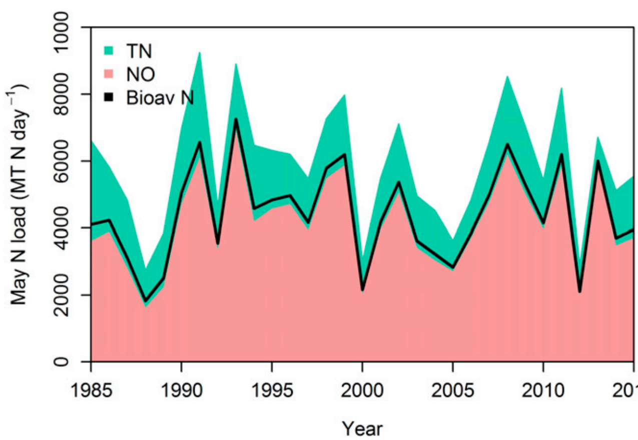

Fig. 1. May nitrogen loads from the Mississippi and Atchafalaya rivers (MT N·day^{-1}) for TN, nitrate plus nitrite nitrogen (NO), and bioavailable nitrogen (NO + ammonia + 12% of organic N).

probabilistic ensembles to help bridge the gap between research and policymaking (25–29). Our work advances that goal by formally propagating model uncertainty into our ensemble predictions to allow for quantitative risk assessment and policies that are robust within realistic uncertainty ranges.

Results

The U-M model (30) explained 69% of the variability in hypoxic area extent over 1985–2011. This model (and all others) was subjected to a leave-one-year-out cross-validation, in which it explained 45% of hypoxic area variability. This model's annual hypoxia forecasts have compared well with measurements, especially for years without storms or high winds (R^2 = 70%; Fig. 2). Its load–response curve indicates that a reduction in hypoxia to 5,000 km^2 requires a 58% [95% credible interval (CI): 49–70%] decrease in May total nitrogen (TN) load relative to the 1980–1996 average (Table 1). A 20% reduction would produce a hypoxic extent of 15,000 km^2 (95% CI: 13,500–16,500 km^2). The NCSU model, calibrated for 1985–2011, explains ∼75% of the variability in BWDO and 70% of the variability in hypoxic area. The cross-validation performed similarly to the full calibration (R^2 = 72% for BWDO) (31). This model was first used for hypoxia forecasts in 2015 (Fig. 2) and estimates that reaching the hypoxia goal requires a 56% (95% CI: 50–62%) decrease in spring bioavailable N load (Table 1). A 20% reduction would produce a hypoxic extent of 12,400 km^2 (95% CI: 10,800–14,000 km^2). The LSU/LUMCON model explains 92% of the variation in hypoxic extent since 2000, after removing 5 y with high wind/storms. The cross-validation on the same calibration dataset showed similar model performance (R^2 = 87%). Its annual forecasts have been fairly accurate since 2002 (Fig. 2). Without the five high-wind/storm years, its forecasts accounted for 56% of the observed variability, and the model indicates that a 56% (95% CI: 50–64%) reduction in the May nitrite+nitrate (NO) load is needed to meet the hypoxia goal (Table 1). A 20% reduction would produce a hypoxic extent of 15,600 km^2 (95% CI: 14,400–16,800 km^2). The VIMS model explained 64% of the variability in hypoxic area between 1985 and 2015, and the cross-validation explained 52% of the interannual variation. A simplified version of the model has been used to produce blind forecasts since 2014 (Fig. 2). This model indicates that an 80% (95% CI: 50%) reduction in the May NO load is needed to meet the hypoxia goal (Table 1), while a 20% load reduction would result in a hypoxic area of 12,900 km^2 (95% CI: 10,700–15,000 km^2).

The ensemble modeling results indicate that, under typical weather conditions, a 59% reduction in N load from the 1980–1996 baseline would be required to reach the goal. The 20% reduction interim goal would result in 13,900 km^2 (95% CI: 11,100–16,400 km^2), corresponding to an 18% reduction in hypoxic extent over the long-term, and a 1% reduction compared with the most recent 5-y period (Fig. 3 and Table 1).

Discussion

While multiple models have been used to inform hypoxia management in the past, in this work, multiple models are synthesized within a probabilistic framework to develop a "consensus" estimate of how the system will respond and to quantify the uncertainty in that estimate. By developing the ensemble, we explore and quantify uncertainty due to differences in model structure and inputs (32), which is not captured by the individual models themselves. Our four models differ substantially in mechanistic form, type of N load driver, internal sources of DO demand, and criteria for selecting calibration datasets. For example, the VIMS and LSU/LUMCON models are linear regressions with different assumptions on the form of the relationship between nutrients, stratification, and hypoxia, whereas the U-M and NCSU models are based on different mechanistically derived relationships. In addition, while the VIMS and LSU/LUMCON models use NO as the primary nutrient driver, the U-M model uses TN, and the NCSU model uses an estimation of bioavailable N (Fig. 1). The U-M and LSU/LUMCON models use nutrient load as the primary driver, and, as a result, their response curves are steeper than the VIMS curve, which attributes more variation in hypoxia to freshwater discharge and wind. The NCSU model provides a compromise in that load, freshwater discharge, and wind are incorporated into its mechanistic framework. The U-M, NCSU, and VIMS models use the period of record from 1985 for calibration, whereas the LSU/LUMCON model is calibrated from 2000 on, under the hypothesis that the nutrient–hypoxia relationship has changed over time. Finally, the U-M, NCSU, and VIMS models define outlier years according to quantitative meteorological criteria, whereas the LSU/LUMCON model defines outliers based on sea conditions during the midsummer sampling cruise (*Methods*).

Despite these differences, the individual model results show similar responses to load reductions (Fig. 3), and the mean ensemble results indicate that a 59% reduction from the 1980–1996 average N load is needed to meet the Task Force goal (Table 1). The worst-case scenario in our ensemble results (upper bound of predictive intervals) indicates that even an 80% load reduction (which is likely infeasible) may not meet the goal

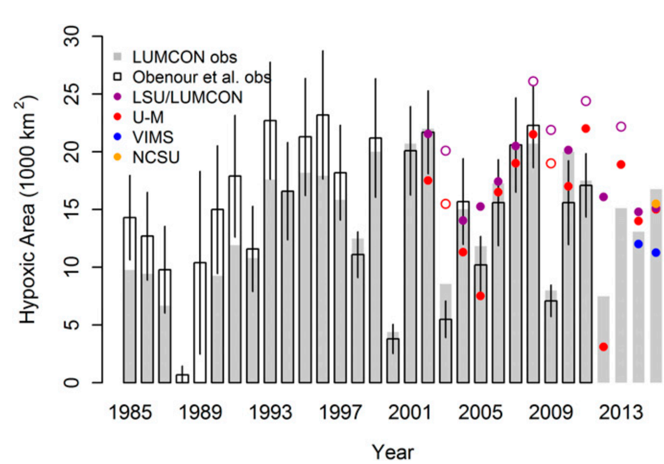

Fig. 2. Observed hypoxic area (1985–2015) and forecast track record (2002–2015) for the four models. Gray bars are estimates of hypoxia area from LUMCON (1), and open bars are geostatistical estimates (with 95% confidence intervals) (70). The annual forecasts made assuming normal weather conditions are shown as circles (red, U-M; purple, LSU/LUMCON; blue, VIMS; and orange, NCSU). Filled and open circles identify normal weather and "wind/storm" years, respectively (as defined by each model).

Table 1. Model estimates of the load reduction required to meet the 5,000-km² Task Force goal and estimates of the hypoxic area expected in response to a 20% load reduction

Model	Load reduction needed for 5000 km² hypoxia, % (95% CI)	Hypoxia area expected for a 20% load reduction, 1,000 km² (95% CI)
U-M	58 (49–70)	15.0 (13.5–16.5)
NCSU	56 (50–62)	12.4 (10.8–14.0)
LSU/LUMCON	56 (50–64)	15.6 (14.4–16.8)
VIMS	80 (50–)	12.9 (10.7–15.0)
Ensemble	59 (50–)	13.9 (11.1–16.4)

(Table 1 and Fig. 3). However, this scenario represents the tail of the probability distribution and is unlikely. We estimate that the probability that hypoxia would be <5,000 km² under an 80% load reduction is 87%. The 59% reduction is larger than the 45% reduction called for in the most recent Action Plan (17), but within the range of previous individual models (5, 30, 31, 33–36). This percentage is also higher than recommendations made for other eutrophic systems. For example, a multimodel effort (20) recommended a 40% phosphorus (P) load reduction for Lake Erie under the Great Lakes Water Quality Agreement (37); the Chesapeake Bay agreement calls for a 25% reduction in N load and a 24% reduction in P (38); and the Neuse River Strategy calls for a 30% reduction in N load (23).

The Task Force also set an interim N load reduction target of 20% by 2025 (17). Our analysis suggests that a 20% reduction may not have a measureable effect within the next 5–10 y. According to the ensemble modeling results, a 20% reduction will result in a 13,900-km² mean hypoxic region (95% CI: 11,100–16,400 km²), which is not significantly different from the current 5-y average of 14,024 km² (Table 1). However, over the long term (as opposed to a 5-y average, which is subject to annual anomalies), a 20% reduction would produce an 18% reduction in hypoxic area, relative to model-estimated current conditions (Fig. 3).

An important contribution of this study, relative to previous Gulf hypoxia modeling efforts, is the characterization of model structure and input uncertainty, in addition to model parameter and predictive error uncertainty (32). We developed two types of predictive intervals that distinguish different sources of uncertainty and which are relevant to different types of management questions (Fig. 3). The credible interval for the mean of the four models quantifies the deterministic uncertainty associated with the response curve (Fig. 3, shaded area), which can be partitioned between within-model uncertainty (i.e., parameter uncertainty) and across-model uncertainty (i.e., input and structural uncertainty). If hypoxia observations are taken over a large number of years (i.e., $n > 30$), we expect the observed mean will fall within this credible interval. For load reductions of 0–50%, the across-model uncertainty accounts for 60% of the total uncertainty on average. For an 80% reduction, the across-model uncertainty accounts for 70%, reflecting the increasing influence of structural uncertainty. At large load reductions, the VIMS predictions are substantially larger than the other models (Fig. 3) because the model attributes more variation in hypoxia to river discharge and wind forcing, and less to N load. Greater ensemble uncertainty is expected at large load reductions where observational data are scarce and the response of the system is less certain. This is especially true for empirical models, where extrapolating predictions near or outside of the calibration range requires caution. Across-model structural uncertainties underscore the need for continued research to address mechanistic uncertainties (discussed below). In general, our results show that across-model uncertainty represents a considerable portion of the overall deterministic uncertainty, and accounting for this uncertainty will lead to more dependable, albeit wider, credible intervals.

The 5-y predictive interval (Fig. 3, gray lines) addresses the additional uncertainty that arises when estimating a mean value from a limited sample size (e.g., the 5-y period used to assess the Hypoxia Task Force goal) (17). The portion of overall uncertainty due to model prediction error can be used to estimate the minimum load reduction required to achieve a statistically significant decrease in hypoxia across 5-y periods. Our results suggest that a 25% load reduction would be required to be 95% certain to observe a reduction in hypoxic area when comparing any two 5-y periods. This supports our finding that the 20% interim load reduction may not produce a measurable effect over 5- to 10-y time scales.

All models are simplified representations of the real world, and the models used here are no exception. Although N has historically been considered the main nutrient driver of hypoxia in the Gulf, the relative roles of N vs. P limitation of primary production, and how that ratio may change seasonally and spatially, remains unclear (39). Based on studies suggesting P limitation on the Gulf shelf at critical time periods, the Task Force adopted a dual nutrient strategy, with the same percent reduction goals for N and P (5, 17). While our models are not designed to answer this question, other modeling supports the dual strategy, with percentage reductions for N and P loads in line with our N load recommendations (40, 41). Expanding the models to incorporate both N and P loads, coupled with better understanding of the stoichiometry regulating primary productivity, has the potential to significantly improve our ability to assess different N and P load-reduction scenarios. Until then, we believe that current evidence points toward a dual-nutrient strategy as the most prudent management approach (39).

There is uncertainty in how internal nutrient loads and sediment oxygen demand (SOD) will modulate the system's response to external load reductions, which presents a challenge for long-term

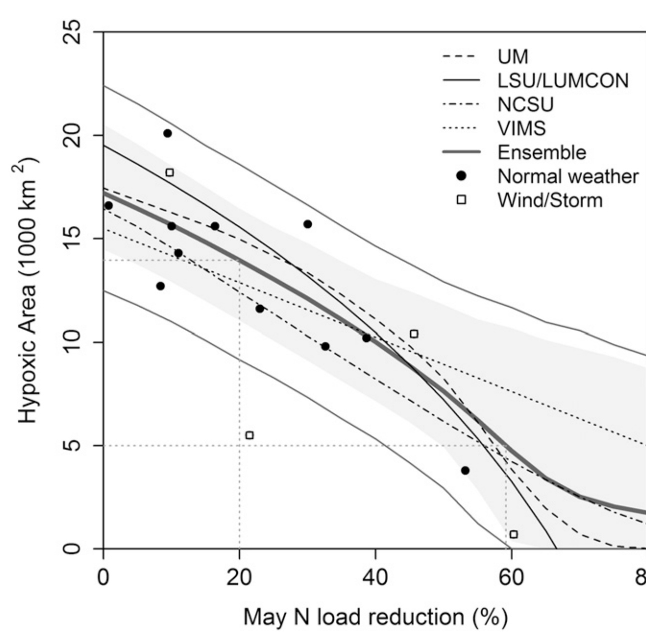

Fig. 3. Load–response curves for the individual models, with the ensemble mean curve, 95% CIs accounting for deterministic uncertainty (shaded area), and 95% predictive intervals accounting for prediction error (averaged across a 5-y period) and deterministic uncertainty (gray solid lines). Gray dotted vertical and horizontal lines indicate the mean hypoxic area expected in response to a 20% load reduction from the 1980–1996 long-term average, and the mean load reduction required to achieve the 5,000-km² Task Force goal. For comparison, observations (70) from 1985 to 2011 for years when the loads were below the 1980–1996 baseline are superimposed on the graph for normal years (filled circles) and wind/storm years (open boxes).

SUSTAINABILITY SCIENCE

ENVIRONMENTAL SCIENCES

forecasting. Currently, only the NCSU model accounts for this process by predicting changes in SOD as a function of long-term average nutrient loads. There is also uncertainty in predicting the impacts of climate change on meteorological and hydrological patterns, and how those impacts may affect the system's susceptibility to hypoxia. For example, the frequency, intensity, and timing of droughts and storms are predicted to shift as a result of global climate change (42), with potential interacting effects on the timing and amount of nutrient delivery, the intensity and duration of stratification, the solubility of oxygen, and biogeochemical cycling (43, 44). More research is needed to enhance our predictive understanding of how these and other processes may cause shifts in internal feedbacks and complex nonlinearities in response to management actions (45–47). However, until these structural uncertainties are decreased, we think that it is critical to develop probabilistic models and probabilistic ensembles of models whose results are communicated to decision-makers to support adaptive strategies that consider uncertainties in system response.

Despite these uncertainties, our results show that the hypoxia response to N load reductions is robust across substantially different and independent models, providing increased confidence that the load reduction proposed will achieve management goals. The strong relationship between nutrient loading and hypoxia illustrated here is also consistent with the results from sediment cores studies that indicate little to no hypoxia on the Gulf shelf before the escalation of nitrogen fertilizer use in the mid-1900s (48). However, it matters little whether the load reduction target is 30%, 45%, or 59% if insufficient resources are in place to make even modest reductions. Recent analyses by the Department of Agriculture (49, 50) comparing modeled nutrient losses between the present state and a hypothetical past without conservation practices indicate that there is some level of conservation effectiveness. However, while there are undoubtedly significant lag times between action on the land and changes in loads (51, 52), river nitrate concentrations have not declined since the 1980s (53, 54), and the current 5-y running average nitrate load to the Gulf is not significantly different from the 1980–1996 baseline (ref. 17; https://toxics.usgs.gov/hypoxia/mississippi/oct_jun/index.html) after US Farm Bill conservation programs have spent more than $28 billion in the 20 Mississippi Basin states since 1995 (55).

Most large-scale environmental restoration efforts are structured within an adaptive management framework (56)—one that sets goals, takes action, measures progress, and adjusts actions if needed. With little documented progress in loads or hypoxic extent, clearly something more or something different is needed (57). Several analyses have demonstrated a range of approaches and potential pathways toward the desired load reduction, including altering fertilizer application rates (58), using cover crops (59), nutrient management (60, 61), alternatives to corn-based biofuels (62, 63), and combinations of the above (12, 34, 64–68). Most of these studies emphasize the value of targeting funding to locations and practices that make the most difference. It is time to ask what is preventing more extensive implementation of some or all of these strategies.

Methods

Nutrient Loads. All models use N load estimates developed by the US Geological Survey (USGS) (https://toxics.usgs.gov/hypoxia/mississippi/nutrient_flux_yield_est.html). The U-M model uses the May TN load; the LSU/LUMCON and VIMS models use the May NO load; and the NCSU model uses a "bioavailable" N load calculated for the 30- to 90-d period preceding each cruise and estimated as inorganic nitrogen (NO and ammonia-N) plus 12% of the organic N (31). Previous studies suggested no apparent correlation between hypoxic area (or predictive errors) in 1 y and either N load or hypoxic extent in the previous year. Thus, including predictors representing carryover effects from the previous year did not substantially improve model fit (31, 35, 69). The loads used by each model show similar interannual patterns (Fig. 1). Reductions discussed herein are expressed as a percent reduction from the 1980–1996 average—the Task Force baseline (13, 14, 17).

Hypoxic Extent. The models are calibrated to hypoxic conditions derived from BWDO measurements taken during midsummer shelf-wide cruises (1). The LSU/LUMCON and VIMS models are calibrated to manually interpolated estimates (1), whereas the NCSU and U-M models are calibrated to geostatistically based estimates (70). While there are some differences between these two approaches, they track each other well over the last two decades (Fig. 2).

Impact of Weather Extremes. Because some models do not account for wind–hypoxia interactions, and because major storms are relatively rare, the response curves are generated for "normal weather" years. The results are conservative because storm events reduce hypoxia. For example, strong westerly winds can force the hypoxic region to "pile up" (shrinking the hypoxic area while maintaining volume), and storms can increase mixing, disrupting stratification and reaerating the water column. These disruptions can reduce the size of the hypoxic area at the time of the survey cruises, but are transitory. Monitoring data indicate that the DO levels return to prestorm conditions within 2 wk (71, 72).

The U-M, NCSU, and VIMS models identified 6 of the 31 y since 1985 with unusual weather based on quantitative criteria (30). "High wind years" (1998 and 2009) were years with unusually strong westerly winds in the 2-mo period preceding the shelf-wide cruise. The westerly wind velocities for these years were 0.96 and 1.11 $m \cdot s^{-1}$, respectively, compared with a mean of -0.44 $m \cdot s^{-1}$. "Storm years" (1988, 1989, 1997, and 2003) met two criteria: (*i*) National Oceanic and Atmospheric Administration (NOAA) storm track data showed tropical storms or hurricanes in the vicinity of the study area within 2 wk of the shelf-wide cruise; and (*ii*) wind stress (wind speed squared) at coastal stations was unusually high in the 2 wk leading up to the shelf-wide cruises (>35 $m^2 \cdot s^{-2}$). The LSU/LUMCON model eliminated 5 of the 16 y since 2000 (2003, 2008, 2009, 2011, and 2013) based on sea conditions within 2 wk before and during the cruise, later confirmed by review of National Weather Service records. The identification of different years by different models reflects an intrinsic level of subjectivity in classifying a year as having "unusual" weather. Including all models in the ensemble allows us to better capture that uncertainty.

U-M Model. This is an adaptation of a river model (73), predicting DO concentrations downstream from point sources of organic matter loads. It uses mass balance equations to estimate the DO consumed during organic matter decomposition and to predict the DO deficit (74) in longitudinal profiles of subpycnocline DO concentration downstream from the outflows of the Mississippi and Atchafalaya Rivers. Organic matter loads are calculated from the May TN loads by converting TN to algal carbon and the associated DO demand based on Redfield and respiratory ratios (75). The length of the hypoxic zone is the sum of all locations along the longitudinal profile where DO is below the hypoxic threshold. Hypoxic length is converted to area through empirical relationship (30). The model is calibrated through Bayesian inference over the period 1985–2011 (30). The model has been compared with others (18) and was used to explore N vs. P control (76), provide guidance for the Gulf Action Plans (13, 14), and explore impacts of climate change (77). The load–response curve was developed with parameter estimates for normal weather years.

NCSU Model. This Bayesian biophysical model predicts BWDO concentrations in western and eastern segments of the Louisiana–Texas shelf that are separated by the Atchafalaya River outfall (31). BWDO predictions are converted to hypoxic area by using empirical relationships between mean BWDO and hypoxic area, both of which are determined from a geostatistical model (70). The model is a steady-state solution to mechanistic, mass-balance equations (31), and is calibrated within a Bayesian framework that accounts for prior information on model parameters. It systematically characterizes parameter and prediction uncertainty (78). Important prior information includes SOD (79) and vertical carbon flux (80) measurements. The east and west shelf sections are segmented into upper and lower layers, where upper layers accommodate transport of freshwater and nutrients across the shelf. The transport of flow and loads along the shelf are regulated by long-term, along-shore wind velocities obtained from NOAA weather stations. Monthly water flow and load estimates are linearly interpolated to determine bioavailable N loads for consecutive 30-d averaging periods leading up to the beginning of the annual shelf-wide cruises. Surface-layer nutrients are subject to an effective settling rate that accounts for organic matter production and sinking. The DO in the bottom layer is controlled by water column DO demand, SOD, and reaeration which is influenced by the rate of freshwater flow and short-term wind stresses.

In the original model application (31), SOD is treated as a "long-term" demand that does not change from year to year. However, SOD is expected to respond to sustained nutrient loading reductions that reduce organic matter fluxes to the sediment over the long-term (74, 81, 82). To quantify this change in SOD, we use a relationship developed for coastal estuaries (83):

$$SOD = a\left(\frac{L_c}{1 + kL_c h}\right)^b$$

where SOD is sediment oxygen demand (mol O_2 $m^{-2} \cdot y^{-1}$), L_c is organic carbon deposition rate (mol C $m^{-2} \cdot y^{-1}$), h is the thickness of the lower layer (m; ~20 m for both shelves), and a, b, and k are parameters with mean values of 0.76, 0.79, and 0.00079, respectively. The average vertical carbon flux in the original model (31) was 5.48 mol C $m^{-2} \cdot y^{-1}$ and SOD was 3.76 mol C $m^{-2} \cdot y^{-1}$. Based on Eq. **1**, the modeled carbon flux produces an SOD, which is 29.8% lower than the SOD determined by the model. Therefore, Eq. **1** is multiplied by 1.425 for our load-reduction scenarios. Some adjustment is to be expected, given the uncertainty in Eq. **1** and the modeled benthic fluxes, and the differences between the Gulf shelf and the estuaries used to establish Eq. **1** (31, 82). To create the load–response curve, the model was run for 21 normal weather years with actual wind and flow conditions, and a range of N load reductions. Thus, the response curves are based on average hydrological and meteorological conditions under various load reductions.

LSU/LUMCON Model. This model is a regression of summer hypoxic zone size as a function of May nitrate load, similar to what has been done for lakes (84, 85). The model assumes that the hypoxic zone is driven mostly by the spring N load and that other influences cause variations around a relatively stable baseline. Previous studies (36, 86) suggested that the relationship between load and hypoxic extent changed over time due to the system becoming more sensitive to N loading as the result of incremental changes in organic matter accumulated in the sediments (87), increases in the nitrate fraction of the total N pool, and long-term climate change. As a result, the model coefficients varied over time, but stabilized after 15 y (36). Therefore, the model is calibrated for years 2000–2015. The N loading data are log10-transformed to linearize the curvilinear relationship observed between load and the estimates of hypoxic extent.

VIMS Model. This regression model uses river NO concentration, river discharge, and wind to estimate hypoxic area (35). To create load–response curves for the present analysis, the river NO concentration was varied to achieve load reductions, and river discharge and wind were held constant at the 2000–2015 averages.

Creating an Ensemble. The ensemble load–response curve uses individual models' curves and their parameter uncertainty. For each load reduction, we determined each model's predictive distribution based on the model's mean and 95% CIs. The ensemble predictive distribution for each load was based on Monte Carlo (MC) sampling from the predictive distributions of the individual models (10,000 samples from each model), thereby obtaining an estimate of the deterministic uncertainty (i.e., including input, structural, and parameter uncertainty) associated with mean ensemble predictions. Negative predictions obtained through MC sampling were set to zero. We compared ensemble deterministic uncertainty with that obtained by accounting for both deterministic uncertainty and the uncertainty associated with model prediction error. We determined each model's prediction error variance from the last 10 y (2005–2015; excluding 2009 as unusual weather) of blind-forecast errors, using a maximum-likelihood estimation. Whenever blind forecasts were not available, then the forecast error was set to the model's cross-validation error, or to the mean absolute error of the U-M and LSU/LUMCON blind forecasts for that year, whichever was greater. Because the Task Force goal is based on a 5-y running average, we divided the models' prediction error variance by 5 to estimate the uncertainty associated with a 5-y mean prediction. We then added each model's prediction error variance to the variance associated with deterministic uncertainty to get the overall predictive variance. The ensemble distribution was generated through MC sampling from the overall predictive distributions of the individual models, as described above.

ACKNOWLEDGMENTS. This work was supported in part by NOAA Grants NA12OAR4320071 and NA09NOS780204 and the University of Michigan Graham Sustainability Institute.

1. LUMCON (2017) Hypoxia in the Northern Gulf of Mexico. Available at: www.gulfhypoxia.net/Research/. Accessed March 8, 2017.
2. Rabalais NN, Turner RE, Wiseman WJ (2002) Gulf of Mexico hypoxia, A.K.A. "The Dead Zone." *Annu Rev Ecol Syst* 33:235–263.
3. Rabalais NN, Turner RE, Scavia D (2002) Beyond science into policy: Gulf of Mexico hypoxia and the Mississippi River. *BioSci* 52:129–142.
4. Wiseman WJ, Rabalais NN, Turner RE, Dinnel SP, MacNaughton A (1997) Seasonal and interannual variability within the Louisiana coastal current: Stratification and hypoxia. *J Mar Syst* 12:237–248.
5. US Environmental Protection Agency (2007) Hypoxia in the Northern Gulf of Mexico: An update by the EPA Science Advisory Board (Environmental Protection Agency, Washington, DC), Technical Report EPA-SAB-08-003. Available at https://yosemite.epa.gov/sab/SABPRODUCT.NSF/C3D2F27094E03F90852573B800601D93/$File/EPA-SAB-08-003complete.unsigned.pdf. Accessed February 2, 2017.
6. Obenour DR, Michalak AM, Zhou Y, Scavia D (2012) Quantifying the impacts of stratification and nutrient loading on hypoxia in the northern Gulf of Mexico. *Environ Sci Technol* 46:5489–5496.
7. Rabalais NN, Turner RE (2001) *Coastal Hypoxia: Consequences for Living Resources and Ecosystems*, Coastal and Estuarine Studies (American Geophysical Union, Washington, DC), Vol 58.
8. Smith MD, et al. (2017) Seafood prices reveal impacts of a major ecological disturbance. *Proc Natl Acad Sci USA* 114:1512–1517.
9. Langseth BJ, et al. (2014) Effect of changes in dissolved oxygen concentrations on the spatial dynamics of the Gulf Menhaden fishery in the northern Gulf of Mexico. *Mar Coast Fish* 6:223–234.
10. O'Connor T, Whitall D (2007) Linking hypoxia to shrimp catch in the northern Gulf of Mexico. *Mar Pollut Bull* 54:460–463.
11. Rabalais NN, et al. (2010) Dynamics and distribution of natural and human-caused hypoxia. *Biogeosciences* 7:585–619.
12. National Science and Technology Council Committee on Environment and Natural Resources (2000) Integrated assessment of hypoxia in the Northern Gulf of Mexico (National Science and Technology Council Committee on Environment and Natural Resources, Washington, DC). Available at oceanservice.noaa.gov/products/hypox_final.pdf. Accessed March 1, 2017.
13. Mississippi River/Gulf of Mexico Watershed Nutrient Task Force (2001) Action plan for reducing, mitigating, and controlling hypoxia in the Northern Gulf of Mexico (Mississippi River/Gulf of Mexico Watershed Nutrient Task Force, Washington, DC). Available at https://www.epa.gov/sites/production/files/2015-03/documents/2001_04_04_msbasin_actionplan2001.pdf. Accessed March 1, 2017.
14. Mississippi River/Gulf of Mexico Watershed Nutrient Task Force (2008) 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 (Mississippi River/Gulf of Mexico Watershed Nutrient Task Force, Washington, DC). Available at https://www.epa.gov/sites/production/files/2015-03/documents/2008_8_28_msbasin_ghap2008_update082608.pdf. Accessed March 1, 2017.
15. Mississippi River/Gulf of Mexico Watershed Nutrient Task Force (2013) Reassessment 2013: Assessing progress made since 2008 (Mississippi River/Gulf of Mexico Watershed Nutrient Task Force, Washington, DC). Available at https://www.epa.gov/sites/production/files/2015-03/documents/hypoxia_reassessment_508.pdf. Accessed March 1, 2017.
16. Mississippi River/Gulf of Mexico Watershed Nutrient Task Force (2016) Looking forward: The strategy of the federal members of the Hypoxia Task Force (Mississippi River/Gulf of Mexico Watershed Nutrient Task Force, Washington, DC). Available at https://www.epa.gov/sites/production/files/2016-12/documents/federal_strategy_updates_12.2.16.pdf. Accessed March 1, 2017.
17. Mississippi River/Gulf of Mexico Watershed Nutrient Task Force (2015) 2015 Report to Congress. (Mississippi River/Gulf of Mexico Watershed Nutrient Task Force, Washington, DC). Available at https://www.epa.gov/sites/production/files/2015-10/documents/htf_report_to_congress_final_-_10.1.15.pdf. Accessed March 1, 2017.
18. Scavia D, Justić D, Bierman VJ (2004) Reducing hypoxia in the Gulf of Mexico: Advice from three models. *Estuaries* 27:419–425.
19. Bierman VJ (1980) A comparison of models developed for phosphorus management in the Great Lakes. *Phosphorus Management Strategies for Lakes*, eds Loehr R, Martin C, Rast W (Ann Arbor Science, Ann Arbor, MI), pp 235–255.
20. Scavia D, DePinto JV, Bertani I (2016) A multi-model approach to evaluating target phosphorus loads for Lake Erie. *J Great Lakes Res* 42:1139–1150.
21. Scavia D, et al. (2017) Multiple models guide strategies for agricultural nutrient reductions. *Front Ecol Environ* 15:126–132.
22. Weller DE, et al. (2013) Multiple models for management in the Chesapeake Bay (Scientific and Technical Advisory Committee, Edgewater, MD), STAC Publication 14-004.
23. Stow CA, Roessler C, Borsuk ME, Bowen JD, Reckhow KH (2003) Comparison of estuarine water quality models for total maximum daily load development in Neuse River Estuary. *J Water Resour Plan Manage* 129:307–314.
24. Gneiting T, Raftery AE (2005) Atmospheric science. Weather forecasting with ensemble methods. *Science* 310:248–249.
25. Bankes SC (2002) Tools and techniques for developing policies for complex and uncertain systems. *Proc Natl Acad Sci USA* 99:7263–7266.
26. Clark JS, et al. (2001) Ecological forecasts: An emerging imperative. *Science* 293:657–660.
27. Harwood J, Stokes K (2003) Coping with uncertainty in ecological advice: Lessons from fisheries. *Trends Ecol Evol* 18:617–622.
28. Müller C (2011) Agriculture: Harvesting from uncertainties. *Nat Clim Chang* 1:253–254.
29. Schindler DE, Hilborn R (2015) Sustainability. Prediction, precaution, and policy under global change. *Science* 347:953–954.
30. Scavia D, Evans MA, Obenour DR (2013) A scenario and forecast model for Gulf of Mexico hypoxic area and volume. *Environ Sci Technol* 47:10423–10428.
31. Obenour DR, Michalak AM, Scavia D (2015) Assessing biophysical controls on Gulf of Mexico hypoxia through probabilistic modeling. *Ecol Appl* 25:492–505.

32. Walker WE, et al. (2003) Defining uncertainty: A conceptual basis for uncertainty management in model-based decision support. *Integrated Assess* 4:5–17.
33. Greene RM, Lehrter JC, Hagy JD, 3rd (2009) Multiple regression models for hindcasting and forecasting midsummer hypoxia in the Gulf of Mexico. *Ecol Appl* 19: 1161–1175.
34. Rabotyagov SS, et al. (2014) Cost-effective targeting of conservation investments to reduce the northern Gulf of Mexico hypoxic zone. *Proc Natl Acad Sci USA* 111: 18530–18535.
35. Forrest DR, Hetland RD, DiMarco SF (2011) Multivariable statistical regression models of the areal extent of hypoxia over the Texas–Louisiana continental shelf. *Environ Res Lett* 6:45002.
36. Turner RE, Rabalais NN, Justić D (2012) Predicting summer hypoxia in the northern Gulf of Mexico: Redux. *Mar Pollut Bull* 64:319–324.
37. Great Lakes Water Quality Agreement (2015) Recommended phosphorus loading targets for Lake Erie—Annex 4 Objectives and Targets Task Team final report to the Nutrients Annex Subcommittee (Great Lakes Water Quality Agreement) (US EPA, Chicago). Available at https://binational.net//wp-content/uploads/2015/06/nutrients-TT-report-en-sm.pdf. Accessed March 1, 2017.
38. US Environmental Protection Agency (2010) Chesapeake Bay total maximum daily load for nitrogen, phosphorus, and sediment (U.S. Environmental Protection Agency, Chesapeake Bay Program Office, Annapolis, MD). Available at https://www.epa.gov/chesapeake-bay-tmdl/chesapeake-bay-tmdl-document. Accessed March 1, 2017.
39. Dodds WK (2006) Nutrients and the "dead zone": The link between nutrient ratios and dissolved oxygen in the northern Gulf of Mexico. *Front Ecol Environ* 4:211–217.
40. Laurent A, Fennel K (2014) Simulated reduction of hypoxia in the northern Gulf of Mexico due to phosphorus limitation. *Elem Sci Anthr* 2:22.
41. Feist TJ, et al. (2016) Modeling the relative importance of nutrient and carbon loads, boundary fluxes, and sediment fluxes on Gulf of Mexico hypoxia. *Environ Sci Technol* 50:8713–8721.
42. Intergovernmental Panel on Climate Change (2012) Managing the risks of extreme events and disasters to advance climate change adaptation. A special report of Working Groups I and II of the Intergovernmental Panel on Climate Change, eds Field CB, et al. (Cambridge Univ Press, Cambridge, UK).
43. Meier HEM, et al. (2011) Hypoxia in future climates: A model ensemble study for the Baltic Sea. *Geophys Res Lett* 38:L24608.
44. Loecke TD, et al. (2017) Weather whiplash in agricultural regions drives deterioration of water quality. *Biogeochemistry* 133:7–15.
45. Carpenter SR (2005) Eutrophication of aquatic ecosystems: Bistability and soil phosphorus. *Proc Natl Acad Sci USA* 102:10002–10005.
46. Carstensen J, Sánchez-Camacho M, Duarte CM, Krause-Jensen D, Marbà N (2011) Connecting the dots: Responses of coastal ecosystems to changing nutrient concentrations. *Environ Sci Technol* 45:9122–9132.
47. Duarte CM, Conley DJ, Carstensen J, Sánchez-Camacho M (2009) Return to Neverland: Shifting baselines affect eutrophication restoration targets. *Estuaries Coasts* 32:29–36.
48. Rabalais NN, Turner RE, Sen Gupta BK, Platon E, Parsons ML (2007) Sediments tell the history of eutrophication and hypoxia in the northern Gulf of Mexico. *Ecol Appl* 17: S129–S143.
49. US Department of Agriculture, Natural Resources Conservation Service (2012) Assessment of the effects of conservation practices on cultivated cropland in the Upper Mississippi River Basin, Conservation Effects Assessment Project (CEAP) (US Department of Agriculture, Natural Resources Conservation Service, Washington, DC).
50. White MJ, et al. (2014) Nutrient delivery from the Mississippi River to the Gulf of Mexico and effects of cropland conservation. *J Soil Water Conserv* 69:26–40.
51. Van Meter KJ, Basu NB (2015) Catchment legacies and time lags: A parsimonious watershed model to predict the effects of legacy storage on nitrogen export. *PLoS One* 10:e0125971.
52. Van Meter KJ, Basu NB, Van Cappellen P (2017) Two centuries of nitrogen dynamics: Legacy sources and sinks in the Mississippi and Susquehanna River Basins. *Global Biogeochem Cycles* 31:2–23.
53. Sprague LA, Hirsch RM, Aulenbach BT (2011) Nitrate in the Mississippi River and its tributaries, 1980 to 2008: Are we making progress? *Environ Sci Technol* 45:7209–7216.
54. Murphy JC, Hirsch RM, Sprague LA (2013) Nitrate in the Mississippi River and its tributaries, 1980 - 2010: An update (US Geological Survey, Reston, VA), USGS Scientific Investigations Report 2013-5169.
55. EWG (2017) Environmental Working Group Conservation Database. Available at https://conservation.ewg.org/?_ga=1.152939057.968936435.1488972853. Accessed March 8, 2017.
56. Zedler JB (2017) What's new in adaptive management and restoration of coasts and estuaries? *Estuaries Coasts* 40:1–21.
57. Porter PA, Mitchell RB, Moore KJ (2015) Reducing hypoxia in the Gulf of Mexico: Reimagining a more resilient agricultural landscape in the Mississippi River watershed. *J Soil Water Conserv* 70:63A–68A.
58. McIsaac GF, David MB, Gertner GZ, Goolsby DA (2001) Nitrate flux in the Mississippi River. *Nature* 414:166–167.
59. Kling CL, et al. (2014) LUMINATE: Linking agricultural land use, local water quality and Gulf of Mexico hypoxia. *Eur Rev Agric Econ* 41:431–459.
60. Ribaudo M, Marshall E, Aillery M, Scott S (2016) Reducing the dead zone in the Gulf of Mexico: Assessing the costs to agriculture. Selected Paper Prepared for Presentation at the 2016 Agricultural & Applied Economics Association Annual Meeting, Boston, MA, July 31–August 2. Available at ageconsearch.umn.edu//handle/235197. Accessed February 2, 2017.
61. Ribaudo M, Key N, Sneeringer S (December 29, 2016) The potential role for a nitrogen compliance policy in mitigating Gulf hypoxia. *Appl Econ Perspect Policy*, 10.1093/aepp/ppw029.
62. Costello C, Griffin WM, Landis AE, Matthews HS (2009) Impact of biofuel crop production on the formation of hypoxia in the Gulf of Mexico. *Environ Sci Technol* 43: 7985–7991.
63. Donner SD, Kucharik CJ (2008) Corn-based ethanol production compromises goal of reducing nitrogen export by the Mississippi River. *Proc Natl Acad Sci USA* 105:4513–4518.
64. Doering OC, Kling CL, Nassauer JI, Scavia D (2007) Agricultural policy choices. *From the Corn Belt to the Gulf: Societal and Environmental Implications of Alternative Agricultural Futures*, eds Nassauer JI, Santelmann MV, Scavia D (Resources for the Future, Washington, DC), pp 185–200.
65. Scavia D, Nassauer JI (2007) Policy insights from integrated assessments and alternative futures. *From the Corn Belt to the Gulf: Societal and Environmental Implications of Alternative Agricultural Futures*, eds Nassauer JI, Santelmann M, Scavia D (Resources for the Future, Washington, DC), pp 1–9.
66. Santelmann M, et al. (2007) An integrated assessment of alternative futures for corn belt agriculture. *From the Corn Belt to the Gulf: Societal and Environmental Implications of Alternative Agricultural Futures*, eds Nassauer JI, Santelmann M, Scavia D (Resources for the Future, Washington, DC), pp 162–174.
67. Scavia D, Mitsch WJ, Doering OC (2007) Improving water quality from the corn belt to the Gulf. *From the Corn Belt to the Gulf: Societal and Environmental Implications of Alternative Agricultural Futures*, eds Nassauer J, Santelmann M, Scavia D (Resources for the Future, Washington, DC), pp 175–184.
68. McLellan E, et al. (2015) Reducing nitrogen export from the corn belt to the Gulf of Mexico: Agricultural strategies for remediating hypoxia. *J Am Water Resour Assoc* 51: 263–289.
69. Liu Y, Evans MA, Scavia D (2010) Gulf of Mexico hypoxia: Exploring increasing sensitivity to nitrogen loads. *Environ Sci Technol* 44:5836–5841.
70. Obenour DR, Scavia D, Rabalais NN, Turner RE, Michalak AM (2013) Retrospective analysis of midsummer hypoxic area and volume in the northern Gulf of Mexico, 1985-2011. *Environ Sci Technol* 47:9808–9815.
71. Walker ND, Rabalais NN (2006) Relationships among satellite chlorophylla, river inputs, and hypoxia on the Louisiana Continental shelf, Gulf of Mexico. *Estuaries Coasts* 29:1081–1093.
72. Rabalais NN, et al. (2007) Hypoxia in the northern Gulf of Mexico: Does the science support the plan to reduce, mitigate, and control hypoxia? *Estuaries Coasts* 30:753–772.
73. Streeter HW, Phelps EB (1925) A study of the pollution and natural purification of the Ohio River (US Public Health Service, Rockville, MD), Public Health Bulletin 146.
74. Chapra SC (2008) *Surface Water-Quality Modeling* (Waveland, Long Grove, IL).
75. Scavia D, Rabalais NN, Turner RE, Justić D, Wiseman WJ (2003) Predicting the response of Gulf of Mexico hypoxia to variations in Mississippi River nitrogen load. *Limnol Oceanogr* 48:951–956.
76. Scavia D, Donnelly KA (2007) Reassessing hypoxia forecasts for the Gulf of Mexico. *Environ Sci Technol* 41:8111–8117.
77. Donner SD, Scavia D (2007) How climate controls the flux of nitrogen by the Mississippi River and the development of hypoxia in the Gulf of Mexico. *Limnol Oceanogr* 52:856–861.
78. Lunn DJ, Thomas A, Best N, Spiegelhalter D (2000) WinBUGS–A Bayesian modelling framework: Concepts, structure, and extensibility. *Stat Comput* 10:325–337.
79. Lehrter JC, Beddick DL, Devereux R, Yates DF, Murrell MC (2012) Sediment-water fluxes of dissolved inorganic carbon, O2, nutrients, and N2 from the hypoxic region of the Louisiana continental shelf. *Biogeochemistry* 109:233–252.
80. Redalje DG, Lohrenz SE, Fahnenstiel GL (1994) The relationship between primary production and the vertical export of particulate organic matter in a river-impacted coastal ecosystem. *Estuaries* 17:829–838.
81. Di Toro DM, Paquin PR, Subburamu K, Gruber DA (1990) Sediment oxygen demand model: Methane and ammonia oxidation. *J Environ Eng* 116:945–986.
82. Rucinski DK, DePinto JV, Scavia D, Beletsky D (2014) Modeling Lake Erie's hypoxia response to nutrient loads and physical variability. *J Great Lakes Res* 40:151–161.
83. Borsuk ME, Higdon D, Stow C, Reckhow K (2001) A Bayesian hierarchical model to predict benthic oxygen demand from organic matter loading in estuaries and coastal zones. *Ecol Modell* 143:165–181.
84. Vollenweider RA (1968) Scientific fundamentals of the eutrophication of lakes and flowing waters, with particular reference to nitrogen and phosphorus as factors in eutrophication (Organisation for Economic Co-operation and Development, Paris), OECD Technical Report DAS/CSI/68.27.
85. Vollenweider RA (1975) Input-output models. *Schweizerische Zeitschrift für Hydrol* 37:53–84.
86. Turner RE, Rabalais NN, Justić D (2008) Gulf of Mexico hypoxia: Alternate states and a legacy. *Environ Sci Technol* 42:2323–2327.
87. Turner RE, Rabalais NN (1994) Coastal eutrophication near the Mississippi river delta. *Nature* 368:619–621.