Sensitivity and uncertainty analysis of model hypoxia estimates for the Texas-Louisiana shelf

Jann Paul Mattern,[1,2] Katja Fennel,[2] and Michael Dowd[1]

[1]   Numerical ocean models are becoming increasingly important tools for marine research and for management of marine resources. It is therefore crucial that uncertainty in model predictions and model sensitivity to errors in the model inputs be quantified. We performed a combined sensitivity and uncertainty analysis for a realistic physical-biological model of the Texas-Louisiana shelf in the northern Gulf of Mexico. The model simulates the major physical and biological processes involved in the formation of the hypoxic zone that develops on the shelf every summer. With the help of a statistical emulator technique, we introduced uncertainty in selected model inputs and assessed the effects of these uncertainties on the predicted development and spatial distribution of bottom hypoxia. The uncertain inputs we examined belong to two categories: (i) biological inputs including river nutrient concentration, phytoplankton growth rate and initial and boundary conditions of biological variables, and (ii) physical inputs including freshwater river runoff, wind forcing, and mixing coefficients. We show that uncertainty in different inputs has distinct effects on model output which vary in magnitude, time, and space. Uncertainty in physical inputs was found to have a strong impact on estimates of hypoxia, e.g., hypoxic area estimates vary by more than 40%, due to a 20% variation in the freshwater river runoff.

1. Introduction

[2]   Quantification of uncertainty is an important part of numerical modeling. Knowledge of model uncertainty allows for an assessment of the reliability and precision of the model and therefore its general usefulness as a tool for prediction and analysis [*Karniadakis and Glimm*, 2006]. Model uncertainty is the model error due to incomplete knowledge of the simulated system, e.g., unknown boundary or initial conditions due to errors in the model formulation and equations or due to lack of (computational) resources to simulate the degree of complexity of the system, e.g., a low model resolution. Inherently connected to the concept of model uncertainty is that of model sensitivity which characterizes the response of model output to changes in its input. A model is said to be sensitive to a particular input if a slight change in the input triggers a large change in the output. Sensitivity thus contributes to uncertainty in outputs. Typically, many model inputs are not well known, which, combined

with model sensitivity, can lead to large uncertainty in model outputs.

[3]   In this study, we performed a combined sensitivity and uncertainty analysis for a three-dimensional (three spatial dimensions in addition to the time dimension) physical-biological model set on the Texas-Louisiana shelf. The model includes a nitrogen-based biological module coupled with dissolved oxygen dynamics and is described in *Fennel et al.* [2011, 2013]. In *Fennel et al.* [2013], it is shown that the model reproduces observed hypoxic extent well for certain configurations. The goal of this study is to assess the sensitivity of these predicted hypoxia estimates to model inputs of two categories: (i) biological inputs, including one biological parameter (the maximum growth rate of phytoplankton), biological boundary and initial conditions, and river nutrient concentration and (ii) physical inputs, including two parameters controlling subgrid scale horizontal mixing, wind forcing, which can have a significant impact on the extent of hypoxia [*Forrest et al.*, 2011; *Feng et al.*, 2012], and the amount of freshwater discharge.

[4]   Previous studies of model uncertainty and sensitivity have focused on physical ocean models [*Lermusiaux*, 2006; *Kim et al.*, 2010; *Thacker et al.*, 2012]. Sensitivity analyses are also commonly found in the field of ecosystem modelling [*Clancy et al.*, 2010; *Makler-Pick et al.*, 2011; *Gibson and Spitz*, 2011; *Melbourne-Thomas et al.*, 2011] but not typically for coupled three-dimensional biological-physical models. Only a few studies have investigated uncertainty propagation in physical-biological ocean models, e.g.,

---

[1]Department of Mathematics and Statistics, Dalhousie University, Halifax, Nova Scotia, Canada.
[2]Department of Oceanography, Dalhousie University, Halifax, Nova Scotia, Canada.

Corresponding author: J. P. Mattern, Department of Mathematics and Statistics, Dalhousie University, PO Box 15000, Halifax, NS B3H 4R2, Canada. (paul.mattern@dal.ca)

Béal et al. [2010] assessed the effect of mixing errors on biological properties in a physical-biological model, and *Cossarini et al.* [2009] simulate the error dynamics of a model of the Lagoon of Venice ecosystem using a data assimilative approach.

[5] Unless uncertainty is directly integrated into the model [see, e.g., *Lermusiaux*, 2006], an uncertainty or sensitivity analysis entails many model simulations and is thus computationally demanding, in particular for three-dimensional models. Typically, uncertainty analyses represent input uncertainty via random samples using Monte Carlo techniques [*Clancy et al.*, 2010; *Kim et al.*, 2010; *Melbourne-Thomas et al.*, 2011], a computationally inefficient approach that is not practical for three-dimensional models. Emulator-based approaches, such as the polynomial chaos expansion [introduced by *Wiener*, 1938] which we use here, offer a computationally more efficient alternative for propagating uncertainty in model inputs to their outputs [*Xiu and Karniadakis*, 2003; *Shen et al.*, 2010; *Thacker et al.*, 2012]. These techniques sample the input probability distribution in a nonrandom fashion and interpolate model output in between samples.

[6] Our results highlight the relative importance of the biological and physical model inputs on hypoxia predictions. The inputs cause distinct temporal and spatial distributions in model uncertainty, where high uncertainties are typically found on the inner shelf region in summer. We show that model predictions of oxygen concentration, hypoxia, and surface chlorophyll are sensitive to uncertainty in various model inputs, especially the physical inputs which perturb the model's stochastic flow field. Uncertainties in the inputs show strong local effects, such as the oxygen concentration in a specific grid cell, as well as larger scale features, such as the size of the predicted hypoxic area.

2. Methods

2.1. The Emulator: Polynomial Chaos Expansion

[7] Emulators provide a statistical approximation of unknown model output based on existing output (here, a *statistical* approximation means an approximation that provides error estimates, see *O'Hagan* [2006]). Based on the model output approximation, emulators can be used to propagate uncertainty in model inputs (model parameters, initial or boundary conditions, physical forcing, etc.) to its outputs [*Xiu and Karniadakis*, 2003; *Le Maître et al.*, 2004]. For this purpose, emulators use output from existing model simulations that are obtained by varying uncertain inputs. By interpolating between the existing output values, emulators can estimate model output for new, unknown values of the inputs and even approximate the probability distribution of the model output. Polynomial chaos expansion, the technique we use in this study, is an emulator based on polynomial interpolation (approximation error estimates are not provided directly but can be derived from convergence properties, see *Thacker et al.* [2012]); it has been used previously with ocean models by *Thacker et al.* [2012] and *Mattern et al.* [2012]. In *Mattern et al.* [2012], the approximation feature of polynomial chaos is used to estimate parameter values for a biological model. In this study, we take advantage of the fact that polynomial chaos allows one to make

approximations of model output probability distributions and use it to perform an uncertainty analysis.

[8] An introduction to the polynomial chaos technique is given in *Thacker et al.* [2012] and *Mattern et al.* [2012]. Here, we briefly review the main aspects. In the following, we consider the stochastic model input θ, which can be, for example, a model parameter or even a scaling factor for the model's boundary condition. Because we only consider variations of single model inputs in this study, we assume in the following that θ is one-dimensional. Although we do not consider the multidimensional case here, the theory translates in a straightforward manner into two or more dimensions for multiple inputs [*Xiu and Karniadakis*, 2002].

[9] Here, the model output of interest is represented by the function $f(\mathbf{x}, t, \theta)$. The output may be dependent on space $\mathbf{x}$, time t, and the uncertain input θ. With the polynomial chaos technique, f is approximated by a basis function expansion as follows:

$$f(\mathbf{x}, t, \theta) = \sum_{k=0}^{k_{\max}} a_k(\mathbf{x}, t)\, \phi_k(\theta) \; + \epsilon_{\mathrm{trunc}}(\theta), \tag{1}$$

where $a_k(\mathbf{x}, t)$ are the expansion coefficients, independent of the uncertain input θ, and the kth basis function $\phi_k(\theta)$ is a polynomial of order k in the parameter space defined by θ. The parameter $k_{\max}$ is the maximum order of polynomials used and determines the accuracy of the approximation, and $\epsilon_{\mathrm{trunc}}$ is the truncation error. Without cutoff, i.e., for $k_{\max} = \infty$, the approximation is exact and $\epsilon_{\mathrm{trunc}}(\theta) = 0$. However, the number of required model runs grows with $k_{\max}$, so that computational constraints force us to use relatively small values in applications with computationally expensive models.

[10] The choice of polynomials in equation (1) is dependent on the probability density function (pdf) of the parameter θ which we denote $p(\theta)$. The polynomials are chosen to be orthogonal with respect to p, and all common distributions have specific sets of polynomial basis functions [*Xiu and Karniadakis*, 2002]. For the beta distribution, which is used here for its finite support, the corresponding set of orthogonal polynomials are the Jacobi polynomials, and ϕ_k is the kth Jacobi polynomial.

[11] In order to approximate f in equation (1), the coefficients a_k need to be computed. For this purpose, equation (1) is rearranged to express the coefficients a_k in terms of $f(\mathbf{x}, t, \theta^{(i)})$, the model output at each so-called quadrature point $\theta^{(i)}$ for $i = 0, 1, \ldots, k_{\max}$ (see *Xiu and Karniadakis* [2002] for details). This is where existing model output ($k_{\max} + 1$ model simulations) are required for the emulator-based approximation. Once the coefficients have been computed, two important properties of the output distribution, the expected value and variance of f, are straightforward to calculate. They are given by

$$\mathbb{E}(f(\mathbf{x}, t, \theta)) = a_0(\mathbf{x}, t) \qquad \text{and} \tag{2}$$

$$\mathrm{var}\,(f(\mathbf{x}, t, \theta)) = \sum_{k=1}^{n} a_k^2(\mathbf{x}, t)\, N_k. \tag{3}$$

Here, $N_k = \int_S \phi_k(\theta)^2\, p(\theta)\, \mathrm{d}\theta$ is a normalization factor specific to the kth polynomial and independent of θ; S is the support of p (the region where $p(\theta) > 0$). The properties in equations (2) and (3) are the mean and variance of

the probability distribution of *f* based on the distribution of θ. Different inputs will produce different distributions in *f*; therefore, mean and variance are dependent on θ. Note that other properties of the probability distribution of *f*, including its pdf can also be approximated with the polynomial chaos expansion (section 2.6).

[12] When the polynomial chaos-based approximation is expanded beyond a single (one-dimensional) input, the computational cost increases exponentially with the number of stochastic inputs. As a result, a joint analysis of nine inputs with polynomials of order k_{max} would require $(k_{max} + 1)^9$ model runs, while only $(k_{max} + 1) \times 9$ would be required for an individual analysis of each input. Although methods have been proposed to reduce the number of required simulations by performing sparse sampling [see, e.g., *Blatman and Sundret*, 2010], we are specifically interested in the effect of individual inputs on the model under similar conditions. For this reason, we sample each input distribution in the same manner but restrict the uncertainty analysis in this study to one parameter at a time, i.e., we estimate the univariate distributions of the inputs and not their joint, multivariate distribution.

2.2. Hypoxia on the Texas-Louisiana Shelf

[13] The Mississippi-Atchafalaya river system is fed by a vast $3,220,000$ km^2 drainage basin and supplies a high load of nutrients [recent estimates are at around 1.25 Tg nitrogen per year, *Aulenbach et al.*, 2007] to the Texas-Louisiana shelf. The nutrients fuel phytoplankton growth on the shelf and contribute to the formation of zones of low-dissolved oxygen in bottom waters [*Rabalais et al.*, 2002; *Dagg and Breed*, 2003]. Oxygen is depleted when sinking organic material is remineralized by bacteria-consuming oxygen in the process. In summer, when the water is highly stratified limiting oxygen resupply to the bottom waters, zones of low oxygen form. These zones are considered hypoxic when oxygen concentrations fall below a critical threshold of 63 mmol O$_2$ m^{-3} (the equivalent of 2 mg L^{-1}) that is considered harmful to many marine organisms. In fall, when water column stratification is eroded by cooling and increasing winds, the hypoxic area begins to shrink and disappears before reforming the following year. Hypoxia on the Texas-Louisiana shelf has been subject to many studies; a recent overview is given by *Bianchi et al.* [2010].

2.3. Model Description

[14] We are using the physical-biological model described in *Fennel et al.* [2011, 2013], its domain is shown in Figure 1. The model is based on the Regional Ocean Modeling System [*Haidvogel et al.* 2008] configured for the Texas-Louisiana shelf as described in *Hetland and DiMarco* [2008, 2012] and uses the biological model component of *Fennel et al.* [2006] with an additional variable for dissolved oxygen [*Fennel et al.*, 2013]. The main objective of the model is to simulate the biological nitrogen cycle and the distribution of dissolved oxygen over the Texas-Louisiana shelf which is heavily influenced by freshwater discharge and nutrient input from the Mississippi and Atchafalaya rivers [*Bianchi et al.*, 2010; *Laurent et al.*, 2012]. In addition to dissolved oxygen, the biological model component contains two variables for dissolved inorganic nitrogen (nitrate and ammonium), one variable to represent phytoplankton, one to represent zooplankton, and two detritus variables (small, suspended, and large, fast-sinking detritus). The model also contains a chlorophyll variable, allowing for the simulation of photoacclimation, adjustments to the phytoplankton's chlorophyll content.

[15] In the following, we provide a short overview with a focus on two model configurations for sediment oxygen consumption and the model inputs important for this study. For more detailed information of the model, see *Fennel et al.* [2011].

2.3.1. Instant Remineralization (IR) and Sediment Oxygen Consumption (SOC) Model Configurations

[16] As shown in *Fennel et al.* [2013], oxygen consumption by the sediments is a significant oxygen sink contributing to the development of hypoxia in our study region. Since the choice of parametrization for this process can have a large impact on hypoxia predictions, we use two different parametrizations here: instantaneous remineralization (IR), and the parametrization of sediment oxygen consumption (SOC) by *Hetland and DiMarco* [2008]. In the IR configuration, as the name suggests, sinking organic material is remineralized instantaneously when it reaches the sediments, leading to a consumption of oxygen in the bottom waters above. This approach has the advantage of making oxygen consumption dependent on the amount of organic matter that reaches the sediment, yet has the disadvantage of speeding up the process of remineralization which in reality
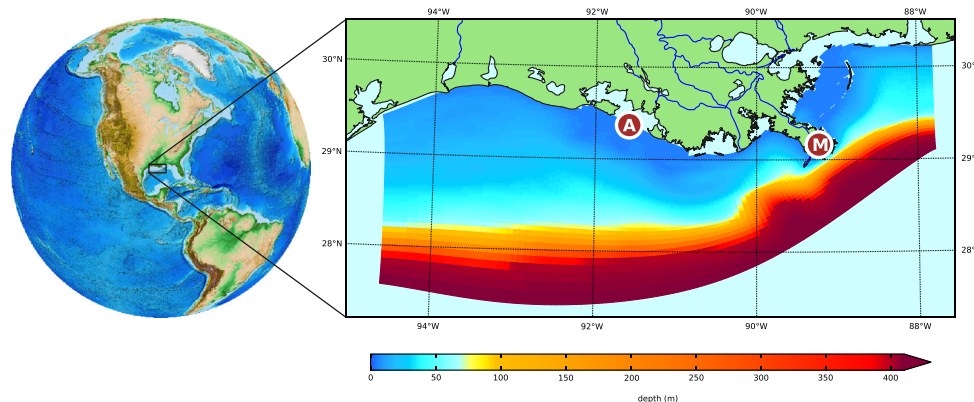

Figure 1. The model domain and its bathymetry. The Mississippi delta and mouth of Atchafalaya Bay are marked by brown circles with "M" and "A", respectively.

leads to a slower but longer lasting consumption of oxygen. In the SOC configuration, sediment oxygen consumption and nutrient production are parametrized based on observations and dependent on temperature and water oxygen concentration only. Consumption of oxygen in the sediments is assumed to occur only in water depths shallower than 50 m in the SOC configuration. This approach is able to simulate a more steady, slow consumption of oxygen; yet the oxygen flux is independent of the flux of organic matter to the sediments and thus decoupled from the supply of nutrients. In *Fennel et al.* [2013], it is shown that both treatments produce reasonable estimates of the hypoxic zone; thus, we examine both the IR and the SOC model configurations and test their response to uncertainty in the inputs.

2.4. Model Inputs of Interest

[17] The model inputs to which we introduce uncertainty are listed in Table 1; all variations were introduced as simple scalings. For example, uncertainty is introduced into the spatially and temporally varying wind field by multiplying it with a scaling factor, a scalar random variable, described in section 2.4.1 below. The univariate distribution of the scaling factor can be approximated at much lower computational cost than a multidimensional distribution which would be required to represent temporal or spatial variations in the input uncertainty. An alternative approach that would allow for the variation of multidimensional inputs would be a dimension reduction, e.g., via principal component analysis and subsequent variation of the lower dimensional representation [*Thacker et al.*, 2012].

[18] We selected these specific model inputs because they represent some of the least known inputs that we suspect would have a noticeable effect on the biological model output. We decided to include both physical and biological model inputs to compare their relative effects. The phytoplankton maximum growth rate was selected as the only representative of the biological parameters because we wanted to include a wider variety of inputs.

2.4.1. Input Distributions

[19] Since the aim of this study is to quantify and compare the effect of uncertainty in various model inputs, we applied the same relative amount of variation to each of the inputs. That is, we used a distribution with a constant ratio of standard deviation/mean for all inputs. While these input uncertainties do not reflect our best knowledge of the real uncertainty of the inputs (although accurate uncertainty estimates would be hard to obtain for most of them), this approach allows for straightforward, objective comparison of the effect on the output uncertainty and is analogous in its approach to sensitivity analyses.

[20] As mentioned above, all variations in the input are based on the variation of a scaling factor that is multiplied with the input quantity, e.g., the amount of river discharge. For a scaling factor that is equal to one, we obtain the baseline model simulation with standard parameters, which represents the scenario we consider most likely. Departure of the scaling factor from one creates perturbations in the model, relative to the baseline results.

[21] For our experiments, we assumed a beta distribution with a scaled support that allows for values of the scaling

Table 1. The Model Inputs That Are Varied in Our Experiments

Name	Abbreviation[a]	Description
River N load and discharge	*river*	Variation in freshwater discharge including all constituents such as nutrients (NO_3 and NH_4) and detritus. The variations are introduced to both the Mississippi and the Atchafalaya rivers.
Freshwater discharge	*disch*	Variation in the freshwater discharge excluding the chemical and biological constituents which are held at the same loads. Variations are introduced in the discharge of the Mississippi and the Atchafalaya rivers.
River N load	*river nut*	Variation in the river nutrient (NO_3 and NH_4) loads while the freshwater discharge is unchanged.
Phytoplankton maximum growth rate	μ_0	Variation in the maximum growth rate of phytoplankton, one of the parameters governing the nitrogen cycle in the biological model.
Nutrient initial conditions	*nut ini*	Variation in the initial conditions for the two biological nutrients NO_3 and NH_4.
Nutrient boundary conditions	*nut bry*	Variation in the boundary conditions for the two biological nutrients NO_3 and NH_4.
Horizontal diffusivity	*hdiff*	Variation in the horizontal mixing coefficient for tracer variables.[b]
Horizontal viscosity	*hvisc*	Variation in the horizontal mixing coefficient for momentum.[b]
Wind	*wind*	Variation in wind speed over the entire model domain.

[a]Used in figures.
[b]The horizontal mixing coefficients provide a parametrization of subgrid scale horizontal mixing.

factor between zero and two. The scaled beta distribution has the probability density function

$$p(x) = \frac{\frac{1}{2}\left(\frac{1}{2}x\right)^{\alpha-1}\left(1-\frac{1}{2}x\right)^{\beta-1}}{B(\alpha,\beta)} \text{ for } x \in \,]0,2[. \quad (4)$$

Here, α and β are the positive parameters that determine the shape of p, and $B(\alpha,\beta) = \int_0^1 x^{\alpha-1}(1-x)^{\beta-1}dx$ is the beta function which acts as a normalizing constant. We only consider beta distributions with parameters $\alpha = \beta > 1$ which results in symmetric distributions with mean values of one and with variances that can be adjusted via α and β. For this study, we select a standard deviation of 0.2; the resulting distribution is similar to a normal distribution with equal mean and standard deviation (Figure 2). One advantage of the beta distribution is that it is truncated and thus does not allow for scaling factors less than zero. The beta distribution also comes with its associated set of orthogonal polynomials, the Jacobi polynomials [*Xiu and Karniadakis*, 2003].

2.5. Shannon Entropy as a Measure of Uncertainty for Model Hypoxia

[22] A parcel of water is defined as hypoxic if its dissolved oxygen concentration is below the critical threshold of 63 mmol O_2 m^{-3}. In the model, the hypoxic area comprises all grid cells with an oxygen concentration below the threshold. If we consider a single grid cell at a fixed point in time, then it can be in one of two states: hypoxic or not hypoxic. When uncertainty is introduced into the model, each grid cell has a certain probability of being hypoxic. The aim is to identify a suitable measure to quantify the uncertainty of the state of the grid cell. If the probability of being hypoxic is zero, then it is certain that the grid cell is not hypoxic, and the measure of uncertainty should be zero to indicate that there is no uncertainty. For the same reason, the measure of uncertainty should be zero if it is certain that hypoxia is occurring and the probability of being hypoxic is one. In all other cases, our uncertainty measure should be greater than zero and should reach its maximum when both hypoxia and no hypoxia have equal probability, i.e., when the probability of hypoxia is $\frac{1}{2}$.

[23] The Shannon entropy introduced in *Shannon* [2001] (originally published in 1948) is a measure of uncertainty that fulfils the above criteria and is defined for a discrete random variable X that has n possible values $x_1, x_2, \ldots, x_n$ as

$$H(X) = -\sum_{i=1}^{n} p_i \, \log_b(p_i). \quad (5)$$

Here, $p_i = \text{Prob}(X = x_i)$ is the probability of X taking the value x_i, and b is the base of the logarithm and can be chosen based on the application. If any of the $p_i = 0$, then the product $p_i \, \log_b(p_i) = 0$ as well.

[24] In this application, X is the state of a grid cell which can take one of $n = 2$ values: $x_1 = $ "hypoxia" and $x_2 = $ "no hypoxia". We let $p_{\text{hypox}} = p_1$ be the probability of X being hypoxic and choose $b = 2$. Then, the Shannon entropy simplifies to

$$H(X) = -p_{\text{hypox}} \, \log_2(p_{\text{hypox}}) - (1-p_{\text{hypox}}) \, \log_2(1-p_{\text{hypox}}). \quad (6)$$
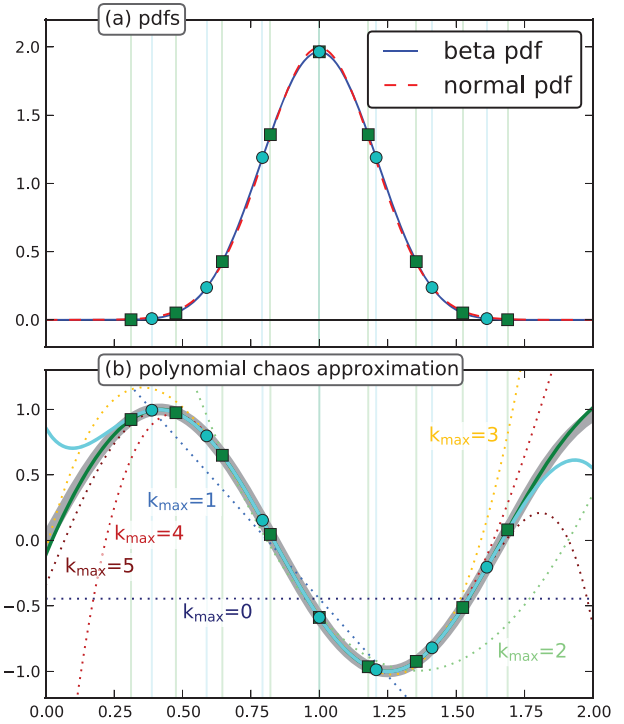

Figure 2. (a) The pdf of the scaled beta distribution in comparison to a normal distribution and (b) an example of polynomial chaos-based approximation on the same $[0,2]$ interval. Both pdfs in Figure 2a have equal mean (1.0) and standard deviation (0.2); the beta pdf is zero beyond the $[0,2]$ interval, whereas the support of the normal distribution is not limited. The thick, grey sine wave in Figure 2b represents an example function f to be approximated by polynomial chaos (equation (1)). Cyan circles and green squares mark the seven and nine polynomial chaos quadrature points, respectively, at which the pdf was sampled in the polynomial chaos expansion (in Figure 2a) and at which exact estimates of f are obtained (in Figure 2b). The solid cyan and green lines in Figure 2b show the polynomial chaos-based approximations of f for seven and nine quadrature points, respectively. The effect of using fewer basis functions by lowering k_{max} on the seven quadrature point approximation of f is illustrated by the dotted lines in Figure 2b. Note that (in Figure 2b) approximation errors can be very large outside the range of quadrature points but that (in Figure 2a) these regions have very small probability values.

The above expression is shown in Figure 3a as a function of p_{hypox} and behaves as desired: If the probability of hypoxia is zero ($p_{\text{hypox}} = 0$) or one ($p_{\text{hypox}} = 1$), then the Shannon entropy is zero. In between, the Shannon entropy is greater than zero and at $p_{\text{hypox}} = \frac{1}{2}$ reaches its maximum value of one. It has been shown generally that the Shannon entropy fulfils the criteria expected from a measure of uncertainty for a discrete random variable [*Shannon*, 2001; *Jaynes*, 1957]. In this study, we will use its binary form, equation (6), and refer to it simply as entropy. We will also frequently use the term uncertainty instead of entropy when referring to the property of the model hypoxia estimates.
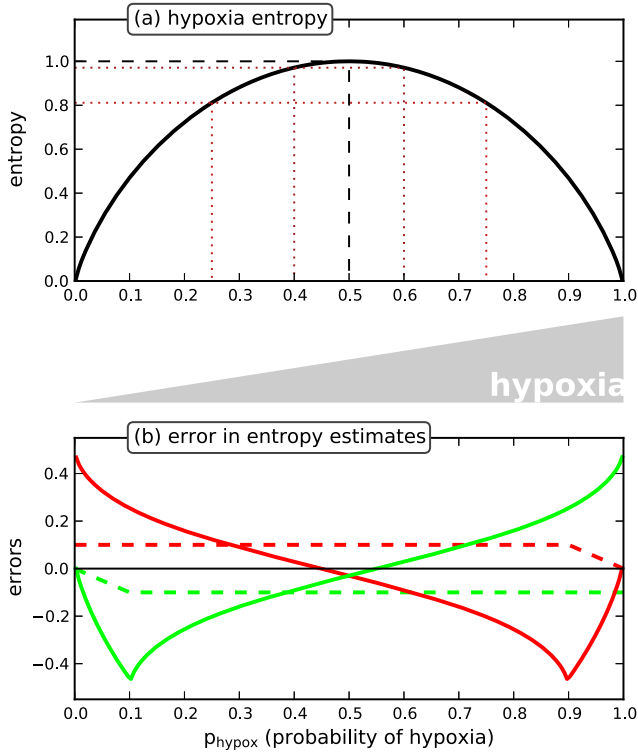
Figure 3. (a) Hypoxia entropy (black line) and (b) error in hypoxia entropy (solid lines) as a function of p_{hypox}, the probability of hypoxia. The dotted lines in Figure 3a mark the values of p_{hypox} that are used in Figures 10 and 11. The dashed lines in Figure 3b depict an error of 0.1 in p_{hypox} in form of an underestimation (green) and overestimation (red) due, e.g., to truncating the polynomial chaos expansion. (Due to restricting p_{hypox} to values between zero and one, underestimation and overestimation errors are below 0.1 for $p_{hypox} < 0.1$ and $p_{hypox} > 0.9$, respectively.) The solid lines show the resulting error in the entropy propagated via the functional relationship in Figure 3a (equation (6)). While the error in p_{hypox} is constant, the error in hypoxia entropy is reduced for p_{hypox} close to 0.5 and amplified as p_{hypox} approaches zero or one.

2.6. Estimating Properties of the Output Distribution

[25] After performing the required model simulations (performed for the input values at the quadrature points; see section 2.1, Figure 2a), we obtained the model outputs at the quadrature points. These outputs represent samples from the probability distribution of the model output (for a graphical example, see Figure 2). Our goal is to estimate properties of the output distribution from these samples. The polynomial chaos expansion offers a straightforward and numerically efficient way to estimate the mean and standard deviation of the output distribution (equations (2) and (3)). However, other properties we are interested in, such as the median, other quantiles, and the entropy, cannot be obtained in such an efficient way. Instead, we have to resort to a procedure of interpolation and binning for their approximation.

[26] In the following, we consider a scalar model output $f(\theta)$ that is dependent on the univariate model input θ; θ could, for example, be a model parameter. The procedure

below describes the approximation of $p_f(f(\theta))$, the pdf of the model output, based on $p(\theta)$, the pdf of θ. For higher dimensional model output, the procedure needs to be repeated for every entry in the output vector or matrix.

[27] (i) Create a fine equidistant grid in parameter space, covering values of θ where $p(\theta)$ is greater than zero (in practice, greater than a small positive number). Let θ^{grid}_i for $i = 1, 2, \ldots, n$ denote the resulting n grid points. For this application, we select $n = 1000$.

[28] (ii) Interpolate the model output in parameter space, i.e., estimate $\hat{f}(\theta) \approx f(\theta)$ based on known output at the quadrature points.

[29] (iii) Form pairs of $\hat{f}$ and p for each grid value, i.e., $\left(\hat{f}\left(\theta^{grid}_i\right), p\left(\theta^{grid}_i\right)\right)$ for $i = 1, 2, \ldots, n$.

[30] (iv) Bin the values of $\hat{f}\left(\theta^{grid}_i\right)$ for $i = 1, 2, \ldots, n$ into m bins (nonoverlapping intervals). Let x_j for $j = 1, 2, \ldots, m+1$ be the limits of the bins. Based on the number of grid points n, we select $m = 100$ (see notes below).

[31] (v) Sum the probability values associated with each bin, i.e., compute

$$\hat{s}_j = \sum_{i : x_j \leq \hat{f}\left(\theta^{grid}_i\right) < x_{j+1}} p\left(\theta^{grid}_i\right) \qquad \text{for } j = 1, 2, \ldots, m.$$

[32] (vi) Finally, normalize the sums to obtain $s_j = \frac{\hat{s}_j}{\sum_{k=1}^{m} \hat{s}_k}$, which is an approximation of the probability of the model output being in the jth bin, i.e., $s_j \approx \int_{x_j}^{x_{j+1}} p_f(x) dx$. At this

point, we have associated model output values with probability values. After the normalization in step (vi), we have obtained an estimated pdf for the output. This pdf can then be used to obtain the desired distribution properties. In practice, we used $m = 100$ bins of equal size to cover the interval between minimum and maximum value of the interpolated model output (step (iv)) in conjunction with $n = 1000$ grid points (step (i)). An increase in the number of bins typically requires a finer grid with more points.

[33] For the purpose of estimating the entropy of a binary event (such as model oxygen below the hypoxic threshold; see section 2.5), it is sufficient to use two bins in step (iv) so that all output values below the threshold are collected in the first bin. After the normalization step (vi), the sum of input pdf values in the first bin is an estimate of the probability of the event having the first of two possible outcomes (in our example, an oxygen concentration below the threshold).

[34] We tested the above procedure by estimating the mean and standard deviation of multiple distributions of different model outputs and comparing the values with the estimates obtained directly from the polynomial chaos expansion. These estimates differ by at most 0.005% in our tests and thus agree to the precision that is relevant to this study. Note that the above procedure can be replaced by a Monte Carlo procedure based on random samples obtained from the emulator approximation. However, we found the Monte Carlo procedure to be less efficient as it required a larger sample size and thus more computing time. No matter which of the aforementioned procedures is used to turn the model output into estimates of the model distribution, the limiting factor for their accuracy is typically the
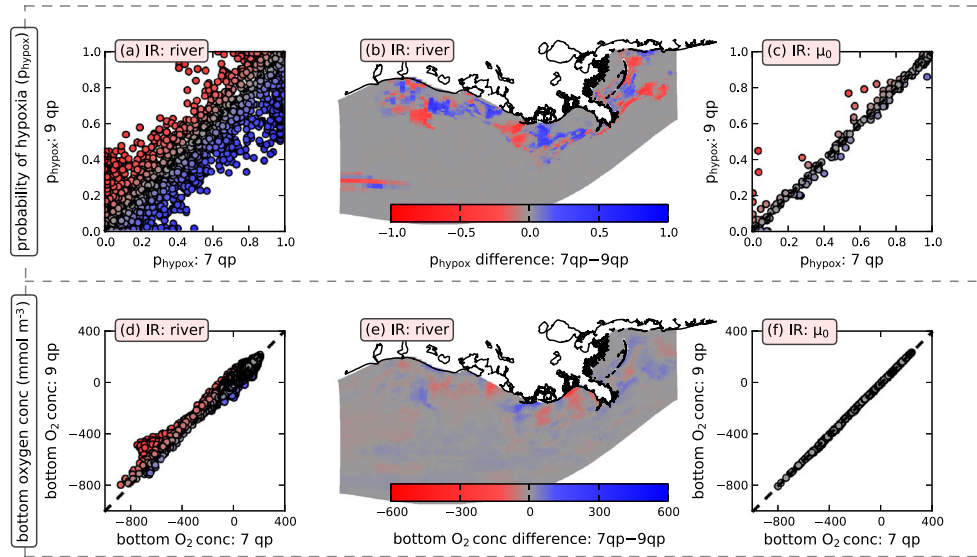
Figure 4. Comparison of emulator-based estimates of the probability of (a–c) bottom hypoxia and (d–f) the bottom oxygen concentration for seven and nine quadrature points (qp) at a day in mid-July for instant remineralization (IR) configuration. The scatter plots depict uncertainty in the model output due to uncertainty in *river*, one of the physical inputs, and μ_0, a biological input. In the scatter plots, distance from the diagonal is emphasized by color intensity. For the physical input, which displays a higher emulator error, the maps (b and e) visualize the spatial location of the values with high variation (the colors correspond to those in the scatter plot Figures 4a and 4d, respectively).

number of quadrature points. More terms in the polynomial chaos expansion in combination with more quadrature points lead to a smaller error in the interpolation and more accurate estimates, but they also require more model simulations (section 2.1, *Xiu and Karniadakis* [2002]).

2.7. Uncertainty in Emulator Estimates

[35] The emulator approximation of model output contains the truncation or interpolation error $\epsilon_{\text{trunc}}(\theta)$ (equation (1)). To estimate the magnitude of this error for the biological output and the derived uncertainty estimates (section 2.6), we compared the emulator estimates for two different numbers of quadrature points, i.e., for two different values of $k_{\max}$. Based on the experiments in *Mattern et al.* [2012] with seven quadrature points, we selected $k_{\max} = 6$ and 8 which corresponds to emulator approximations based on seven and nine quadrature points, respectively. For both values of $k_{\max}$, we obtained the emulator approximations for most model inputs (section 2.4). (We did not obtain approximations for the nutrient initial conditions and nutrient boundary conditions inputs based on nine quadrature points. These inputs have such a low effect on the output (see, e.g., results in Figure 6) that we feel confident that seven quadrature points are sufficient for the emulator approximation.)

[36] If the model is well-approximated by the emulator, then the change from seven to nine quadrature points should be accompanied by a small reduction in the truncation error and little change in the output estimates. Output estimates that vary considerably indicate that the output is not approximated well by the emulator with seven quadrature points (a large change in $\epsilon_{\text{trunc}}(\theta)$ would indicate that the sum in equation (1) is not close to converging and thus even nine quadrature points are not sufficient).

[37] By comparing selected uncertainty estimates that are based on the emulator approximations for seven and nine quadrature points, we found that the truncation error varied considerably among different inputs. Based on the magnitude of the error, we can divide the inputs into two distinct classes. The first class consists of inputs that induce changes into the biological model directly and do not affect the physical model (e.g., biological parameters, biological initial, and boundary conditions). The effect of these inputs is well-approximated by the emulator, and the seven and nine point-based uncertainty estimates are very similar (Figure 4). The second class of inputs consists of those that change the physics of the model and thereby indirectly alter the biology (e.g., both river runoff scenarios, diffusivity, and wind). The biological model output appears to be very sensitive to changes in these parameters, and the emulator-based approximation is not precise at different locations and varies greatly between seven and nine quadrature points (Figure 4). High uncertainties in the emulator estimates are constrained to regions along the coast and are especially evident close to the location of the Mississippi Delta and Atchafalaya Bay. Here, the increase in the number of quadrature points from seven to nine can lead to a notable increase in the estimated probability of hypoxia in one grid cell, accompanied by a notable decrease in a neighboring cell. Our results in section 3 show that these regions of elevated emulator uncertainty are highly dynamic and also exhibit the highest model uncertainty, a finding that is discussed in section 4.

2.7.1. Propagation of Uncertainty Into Entropy Estimates

[38] In light of the high uncertainty for some of the emulator-based estimates of the probability of hypoxia (Figure 4), it is important to note how the emulator's truncation error propagates from probability estimates to

entropy estimates (section 2.5). Due to the log transformation in equation (6), the error in entropy estimates is increased for probability estimates close to zero or close to one (corresponding to low entropy values; see Figure 3), whereas entropy errors are diminished at probability values close to 0.5 (corresponding to high entropy values). Low entropy values therefore have a higher uncertainty than high entropy values.

3. Results

[39] Uncertainty propagates through the model equations and leads to distinct expressions of uncertainty in the different model outputs. Because they are of high importance, we focus our analysis on bottom oxygen and bottom hypoxia. In order to also gain an understanding of the expression of uncertainty at the surface, we close this section with an examination of uncertainty in the model's surface chlorophyll predictions.

3.1. Baseline Oxygen Dynamics Without Uncertainty

[40] First, we briefly describe the simulated oxygen dynamics without introduced uncertainty for the SOC and the IR configurations. We focus the description on the temporal development of dissolved oxygen at a location on the shelf between the Mississippi Delta and the mouth of Atchafalaya Bay (from here on simply referred to as the *station*, see star in Figure 5a) and a spatial snapshot of bottom oxygen across the model domain on 17 July 2004, a time when the simulated hypoxic zone typically reaches one of its largest extents (Figure 5).

[41] Bottom water oxygen concentration at the station shows a decline throughout the spring of 2004, beginning in

April. The decline is briefly interrupted by several mixing events that deliver oxygen-rich surface water to the bottom. Both SOC and IR models show hypoxic conditions at the sea floor in late May and strongly hypoxic conditions throughout most of July and the first half of August. In mid-August, strong mixing increases bottom oxygen and leads to nonhypoxic conditions. At the station location, the IR configuration exhibits more persistent hypoxic conditions with lower oxygen levels in comparison to the SOC model which switches more frequently between hypoxic and nonhypoxic conditions.

[42] Hypoxic areas extend along the coast downstream of the Mississippi Delta and Atchafalaya Bay in both model configurations (Figures 5a and 5b). In the SOC configuration, the hypoxic area appears less homogeneous and more fractured in comparison to the IR configuration, and the SOC bottom oxygen concentration field shows more small scale variability. For a comparison with observations, see *Fennel et al.* [2013].

3.2. Uncertainty in Bottom Oxygen

[43] We now describe the effect of uncertainty in model inputs on model outputs (section 2.4), focusing first on bottom oxygen concentrations at the station. The probability distribution of bottom oxygen and its seasonal development is characterized by quantiles which we averaged over periods of 3 months (Figure 6). (Note that, in comparison to quantiles of the time-averaged distribution, i.e., calculating the time-average first, then the quantiles, these time-averaged quantiles are influenced less by days of extremely high or low oxygen concentrations. They therefore provide a better representation of an average day within each 3 month
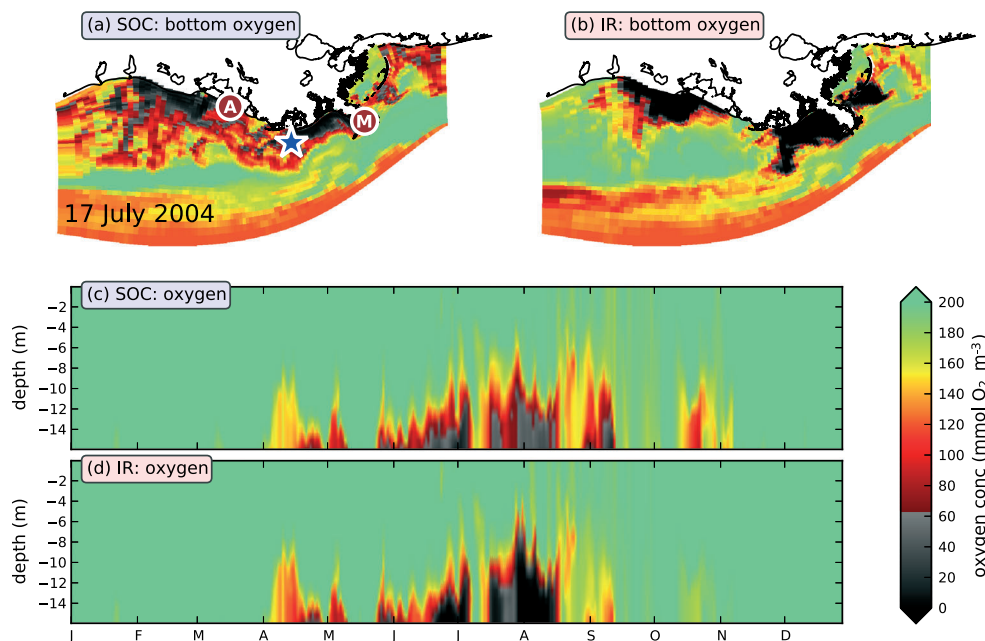

Figure 5. Snapshots of the simulated bottom oxygen concentration on 17 July 2004 for the (a) SOC and (b) IR configurations and (c and d) corresponding time-depth plots of the oxygen concentration at the station marked by the blue star in between the Mississippi Delta (brown circle with "M") and mouth of Atchafalaya Bay (brown circle with "A"). Gray and black colors mark hypoxic areas with oxygen concentrations below 63 mmol O_2 m^{-3}.
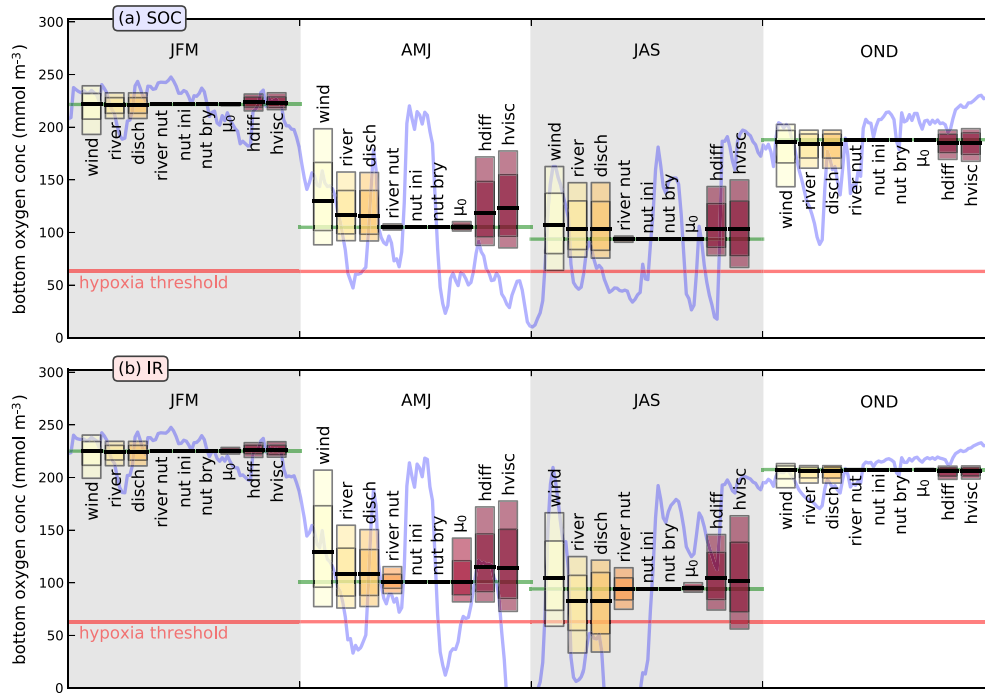
Figure 6. Three-monthly (time-averaged) quantiles of the bottom oxygen concentrations at the station marked in Figure 5 for 2004 (3 month periods are separated by background color; JFM = "January, February, March", etc.). Each bar marks the regions between 0.1 and 0.9 quantiles (outer bar), between the 0.25 and 0.75 quantiles (inner bar), and the median (solid black line) for one uncertain input. The blue solid line in the background marks the development of bottom oxygen in the baseline simulation without uncertainty; its 3 month averages are displayed as horizontal green lines.

period.) Based on the interquartile range (the difference between the 0.75 and 0.25 quantiles), which characterizes the dispersion of the output distribution, we found that the effect of the different inputs varies considerably, and their effect shows a strong seasonal dependence.

[44] Uncertainties in the wind forcing, the main driver of vertical mixing, have the strongest effect on bottom oxygen for both model configurations. Effects of similar magnitude are caused by the horizontal mixing coefficients (*hdiff*, *hvisc*; these abbreviations for the inputs are defined in Table 1) and the freshwater river runoff (included in both *river* and *disch*). In contrast, nutrient initial and boundary conditions have little to no impact on the bottom oxygen concentration at the station. The phytoplankton growth rate (μ_0) and the river nutrient input (*river nut*) are the only inputs that have a distinctly different effect on the two model configurations. As expected, their effect is relatively strong in the IR configuration because sediment oxygen consumption is determined by sinking organic material, in contrast to the SOC configuration where oxygen consumption is largely decoupled from the biological nutrient cycle. Some of the probability distributions in Figure 6 exhibit considerable skewness and a deviation of the median from the baseline (based on the model without uncertainty; dotted blue line in Figure 6), which are due to nonlinearities in the model.

[45] The time dependence of the effect of the uncertain inputs is consistent across inputs for both model configurations. From April through September, the effect of the inputs on bottom oxygen is considerably higher than during the rest of the year. The reason for the low effect early and late in

the year is that temperatures are lower and the ocean is well-mixed, a model behavior that cannot be disrupted by the uncertainty in the input we prescribed.

3.3. Entropy of Hypoxia

3.3.1. Temporal Development of Entropy

[46] The entropy of hypoxia quantifies the uncertainty in model hypoxia estimates and reaches its highest value if the probability of hypoxia is 0.5 (section 2.5). The entropy is low for most of the year and near the ocean surface (Figure 7). In these cases, hypoxia uncertainty is small due to a low probability of hypoxia; variation of the model inputs has a very low probability of causing hypoxia near the surface or when the water column is well mixed in fall and winter. High entropy values and therefore high uncertainty are only found close to the bottom and from mid-May to mid-September, in the times and places where hypoxia is likely to occur. Here, the SOC and IR model configurations show some differences.

[47] In the SOC configuration, the inputs that have a strong effect on bottom oxygen (see previous section 3.1) cause high uncertainties in bottom hypoxia from June through September. The maximum likelihood region of hypoxia, comprised of the model's bottom grid cell where the probability of hypoxia is greater than 0.5 (region outlined in green in Figure 7), varies strongly among different model inputs (compare Figures 7a–7d) and is characterized by high entropy values. Thus, estimates of the number and duration of hypoxic events that the station experienced in 2004 are highly uncertain. This, of course, is only true for the model
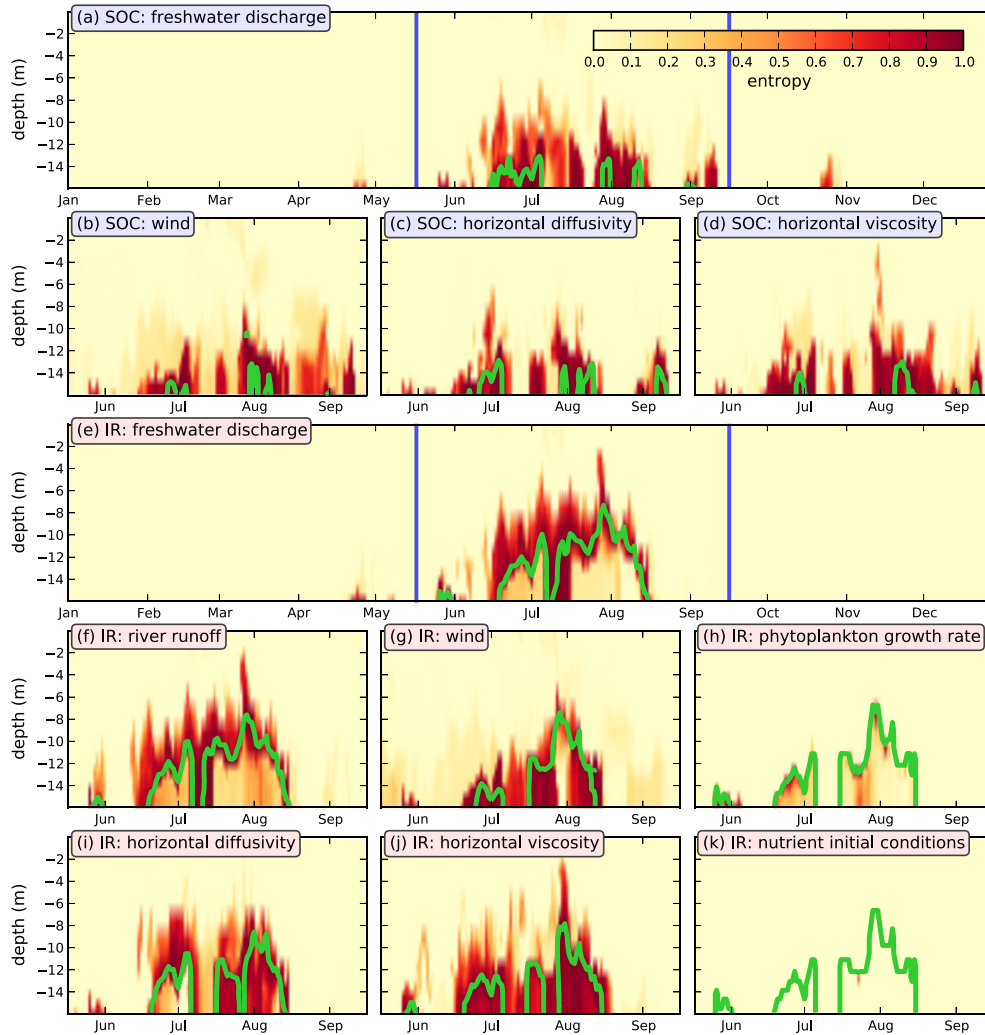
Figure 7. Time-depth plots of hypoxia entropy for selected inputs in the (a–d) SOC and (e–k) IR configurations for the station marked in Figure 5. (a and e) The large panels show the development of entropy for the entire year 2004, and smaller panels focus on the time span from mid-May to mid-September (marked by the blue vertical lines). Regions outlined in green have a probability > 0.5 of being hypoxic. The inputs that are not shown in this figure typically exhibit low uncertainty in oxygen estimates and have small entropy values close to zero. (k) As a representative of these inputs, we have included the *nutrient initial conditions* input.

inputs that have a strong effect on bottom oxygen, but not for those with low effects such as the nutrient initial conditions (the effect of variations in the initial nutrients for the IR configuration is included in Figure 7k; in the SOC configuration, entropy values are equally low).

[48] The IR model configuration results in stronger, more persistent hypoxic events at the station (Figure 5d) which correspond to larger maximum likelihood regions (Figure 7e–7k). Some inputs, such as the river runoff (Figure 7f), exhibit low entropy values within the maximum likelihood region; these correspond to points where the model is nearly certain that hypoxia will occur.

[49] To summarize the temporal development of entropy, we obtained 3 month averages of bottom hypoxia entropy at the station (Figure 8). In both model configurations, uncertainties due to variation of inputs are close to zero early in the year and become significantly larger in April. During the second half of the year, uncertainties are larger in

the SOC configuration compared to the IR, with the exception of those caused by the μ_0 and *river nut* inputs which are generally lower in the SOC configuration due to the weaker coupling between the nitrogen cycle and sediment remineralization (the biological model component affects the sediment oxygen consumption in the SOC configuration through the bottom oxygen concentration only). Generally, uncertainty values of bottom hypoxia correspond well to the variations caused in the bottom oxygen concentrations (Figure 7).

3.3.2. Spatial Distribution of Entropy

[50] The spatial distribution of hypoxia entropy in mid-July (Figure 9) corresponds well to our previous observations: Outer shelf regions exhibit low uncertainty and a low probability of hypoxia; high entropy values occur on the inner shelf and mostly along the fringe of the maximum likelihood hypoxic area. The maximum likelihood hypoxic area contains low entropy values; these correspond to high
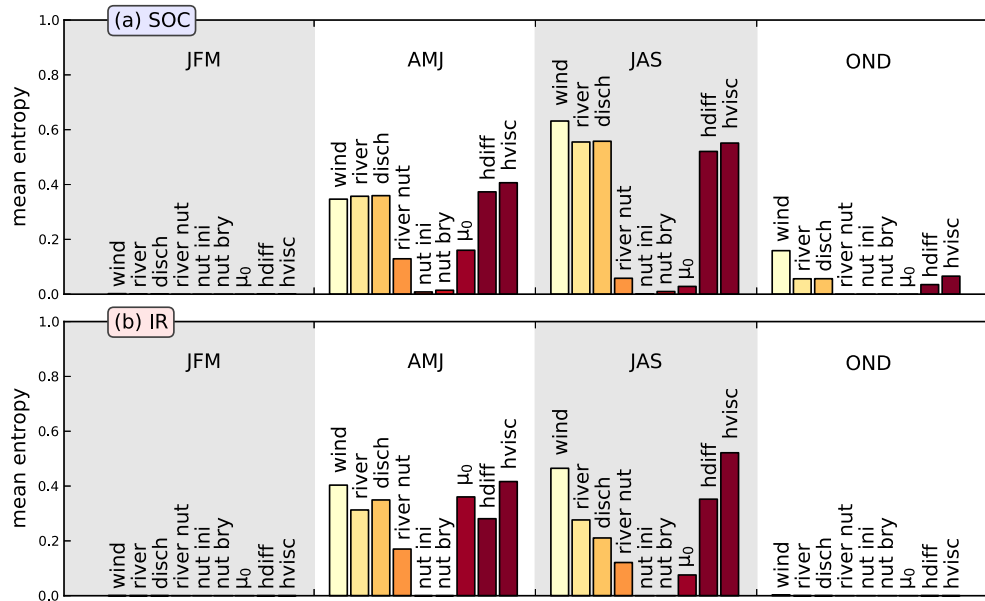
Figure 8. Three-monthly (time-averaged) bottom hypoxia entropy values for the (a) SOC and (b) IR configurations at the location of the station marked in Figure 5.

probabilities of hypoxia. The station (blue star in Figure 9) is in the fringe of the high entropy region which explains the high uncertainty values in the temporal development observed previously (Figure 7).

[51] Consistent with hypoxia predictions in the baseline simulations, the SOC model configuration shows a less homogeneous and more fractured region of high entropy compared to IR (compare SOC Figures 9a–9d to IR Figures 9e–9i). All model inputs create a region of low entropy within the maximum likelihood hypoxic area close to the Mississippi Delta and mouth of Atchafalaya Bay. These correspond to a high probability of hypoxia and low uncertainty.

[52] Based on the spatial maps of hypoxia probability, we computed bounds for the size of the hypoxic area on 17 July 2004 (Figure 10). For this purpose, we summed the areas of the model grid cells with a probability of hypoxia greater than a given threshold. For a threshold of 0.5, we obtained the maximum likelihood area of hypoxia. Other thresholds can define uncertainty estimates for the size of the hypoxic area. For example, the interval in between the thresholds of 0.6 and 0.4 corresponds to the size estimate of the area with entropy values $\geq$ 0.97 (Figure 3a). The wind input, both river inputs and the horizontal mixing coefficients induce wide ranges into the size estimates, especially in the SOC model configuration. For example, for the river inputs *river* and *disch*, the area between the 0.6 and 0.4 probability thresholds is greater than 5000 km^2, more than 40% of the maximum likelihood estimate (Figure 10a). In the SOC configuration, the sizes of the maximum likelihood areas for the high uncertainty inputs also deviate strongly from the estimates of the baseline simulation without uncertainty. Here, it is most apparent that the symmetric distribution of the uncertain input can lead to a highly skewed response in the model.

[53] In summary, some inputs result in large uncertainty in hypoxia predictions, while two inputs (nutrient initial and

boundary conditions) have nearly no effect on bottom oxygen. Uncertainty in hypoxia predictions is temporally and spatially constrained to when and where hypoxia occurs, e.g., variation of model inputs cannot induce hypoxic conditions when the water column is well mixed.

3.4. Comparison of Hypoxic Area Estimates With Observations

[54] For the freshwater discharge, one of the inputs the hypoxia predictions are most sensitive to, we derived a multiyear time series of hypoxic extent (Figure 11). The time series exhibits an annual cycle, with hypoxia typically disappearing from November to March and reaching its maximum extent in late July or August. The uncertainty in the simulated hypoxic extent displays a seasonal cycle, as well, with high uncertainty during the summer months and small uncertainty at times of small hypoxic area estimates. The deviation of the median size of the hypoxic area from the baseline simulation, which was already observed in Figure 10, persists throughout the four years of simulation.

[55] Using the time series, we perform a qualitative comparison with hypoxic area estimates obtained from observations from the LUMCON cruises [*Rabalais et al.*, 2007] for the years 2004–2007 (black bars in Figure 11). In the years 2004 and 2006, both the maximum likelihood estimate and the baseline simulation estimate agree with the observation-based estimate (Figure 11a and 11c). In 2005, the baseline simulation is close to the observations, while the simulation with uncertainty yields lower estimates of the hypoxic area so that the observations lie outside of the region with hypoxia probability of 0.25 or higher (Figure 11b). In 2007, the simulation with uncertainty again underestimates the size of the hypoxic area, yet the observations are within the region of hypoxia probability greater than 0.25 (Figure 11d). In other words, by extending the maximum likelihood hypoxic area to the area that has a probability of hypoxia of 0.25 or higher (based on the uncertainty in
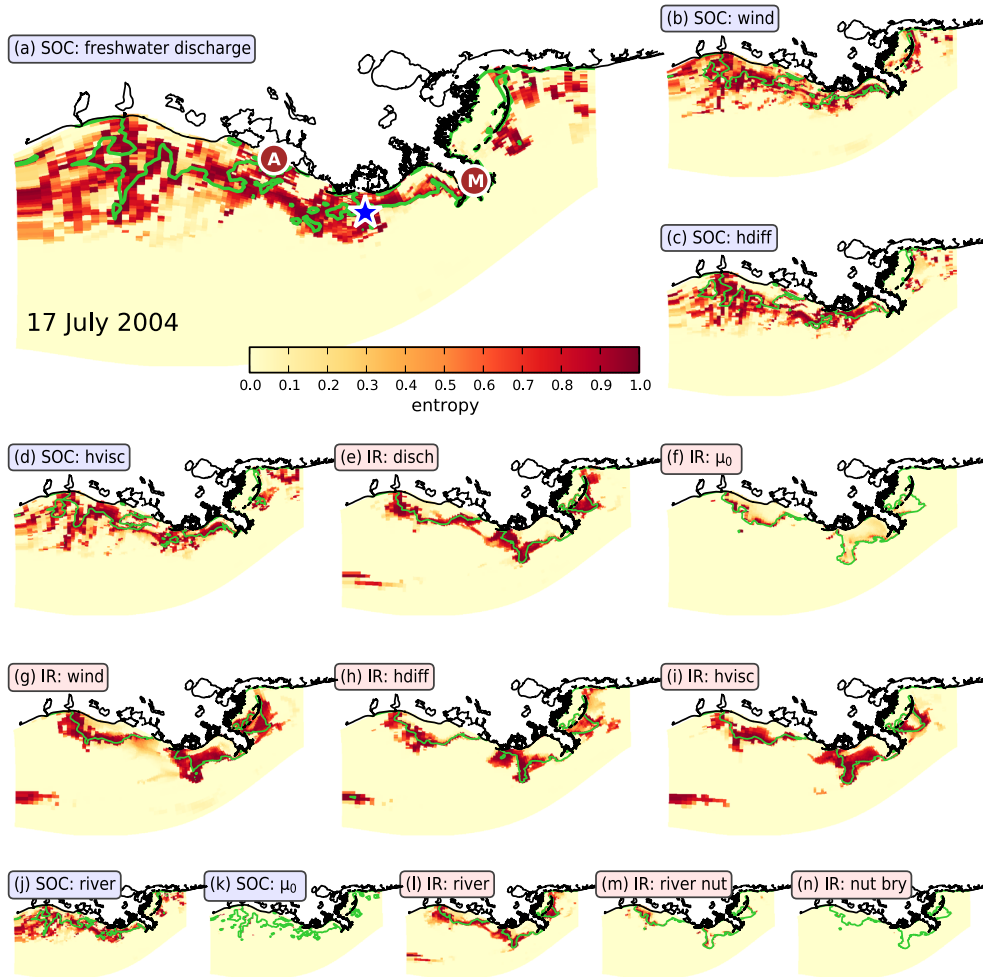
Figure 9. Spatial maps of bottom hypoxia uncertainty (entropy) for selected inputs to the SOC and IR configurations on 17 July 2004. Regions outlined in green have a probability > 0.5 of being hypoxic. The bottom row of panels displays inputs with low entropy values and (j and l) the SOC and IR *river* input maps which are nearly identical to (a and e) their respective *disch* input maps. (a) The blue star marks the station corresponding to the time-depth plots in Figure 7; the Mississippi Delta and mouth of Atchafalaya Bay are marked by brown circles with "M" and "A", respectively.

the freshwater discharge), the prediction corresponds to the observation-based estimate. Uncertainties in the freshwater discharge could help explain the model misfit here.

[56] It should be noted that the observational estimates contain an error as well due to nonsynopticity in the observations, sampling error in the spatially and temporally varying field, and the error associated with interpolation. Nevertheless, simple analyses like this can help to gain an understanding of model error sources and their effect on model output.

3.5. Uncertainty in Surface Chlorophyll Estimates

[57] In order to assess the expression of uncertainty in model variables not directly related to oxygen, we briefly focus on the surface chlorophyll estimates of the model. Surface chlorophyll is often used in data assimilation and model validation applications due to the wealth of available satellite observations [see, e.g., *Allen et al.*, 2007; *Mattern et al.*, 2010]. To obtain an overview of the uncertainty in surface chlorophyll, we computed its time-averaged 3 month quan-

tiles at the station (Figure 12). In comparison to the bottom oxygen quantiles (Figure 6), it is evident that the same inputs have a strong effect on both quantities, albeit in a different order. The inputs that include freshwater river runoff (*river*, *disch*) have a very pronounced impact on surface chlorophyll at the station, and the extra nutrient input of the *river* compared to the *disch* input manifests itself in a slightly more spread out distribution. As with bottom oxygen, seasonal changes in the effect of the inputs are apparent, and all of the inputs have a higher impact in the middle of the year. The choice of SOC or IR model configuration has no noticeable impact on the surface chlorophyll field at the station and leads to negligible differences in the time-averaged quantiles (compare Figure 12a and 12b).

[58] To get an idea of the spatial spread of uncertainty, we obtained the mean and standard deviation of the surface chlorophyll field on 17 July 2004 for two inputs: *disch*, which has a strong effect on surface chlorophyll at the station, and μ_0 with a weaker effect. Both inputs show similar patterns in the mean surface chlorophyll field (Figure 13).
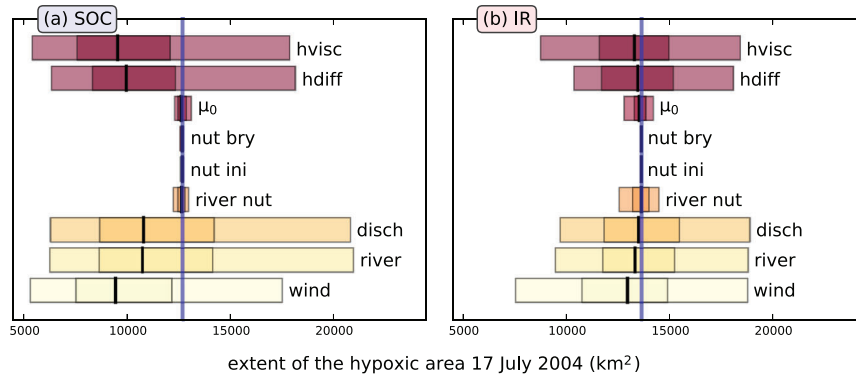
Figure 10. The distribution of the size of the hypoxic area on 17 July 2004 induced by uncertainty in the inputs. Each bar corresponds to one uncertain input to the (a) SOC and (b) IR model configuration. The outer (lighter) bars mark the size of the area that has a probability of hypoxia between at least 0.75 (lower limits of outer bars) and at least 0.25 (upper limits of outer bars). The inner (darker) bars mark the [0.4, 0.6] probability interval, and the black lines correspond to the area with a probability that is at least 0.5, the maximum likelihood area. The blue vertical lines mark the size of the hypoxic area in the SOC and IR baseline simulations without uncertainty. The entropy values corresponding to these probabilities are marked in Figure 3. The time development of these properties for the SOC *disch* input are shown in Figure 11.
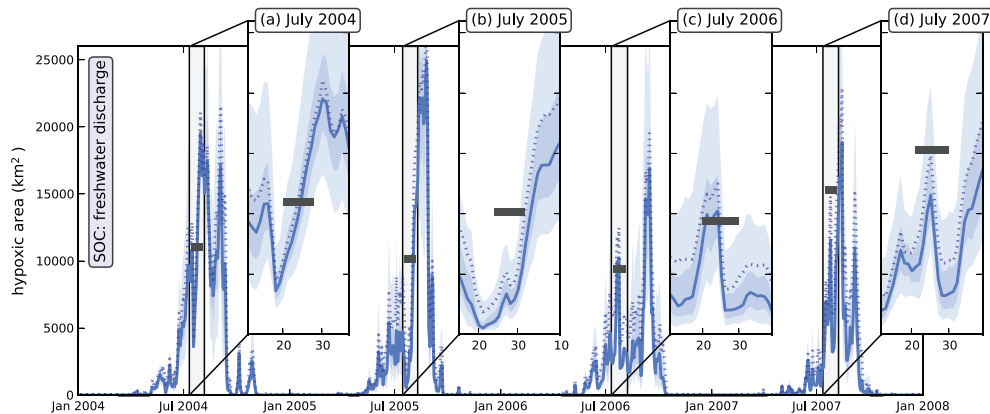

Figure 11. Temporal development of the distribution of the size of the hypoxic area for uncertain *disch* input for the SOC model configuration in comparison with estimates based on observations. The colored regions in the plot correspond to the freshwater discharge to the SOC model configuration. The outer region is the size of the area that has a probability of hypoxia between at least 0.75 (lower limit of outer region) and at least 0.25 (upper limit of outer region). The inner corresponds to probability values of 0.6 and 0.4. The thick blue line corresponds to the area with a probability that is at least 0.5, the maximum likelihood size of the hypoxic area. The dotted line shows the development of the size of the hypoxic area in the baseline simulation without uncertainty. The large black bars correspond to estimates of the hypoxic area based on observations from the Louisiana Universities Marine Consortium (LUMCON) cruises which are performed annually in late July; the width of each bar marks the date range of the corresponding cruise.

Surface chlorophyll is highest close to the Mississippi Delta and the mouth of Atchafalaya Bay, in regions where the nutrient input into the model is high. As expected, uncertainty, described here by the standard deviation, is lower for the μ_0 input. For the *disch* input, high values of the standard deviation are found along the fringes of the high chlorophyll areas, similar to the spread of entropy along the boundary of the hypoxic area (Figure 9). Variations in the inputs appear to affect the surface patch of high chlorophyll in a similar way they affect the hypoxic zone at the bottom: Spatial offsets in large chlorophyll gradients have a similarly strong effect on model uncertainty as offsets of the boundary of the hypoxic area.

4. Discussion

[59] We performed a series of experiments to assess how uncertainty in selected inputs of a physical-biological model of the Texas-Louisiana shelf propagates into its output. This uncertainty analysis was based on an emulator approach, the
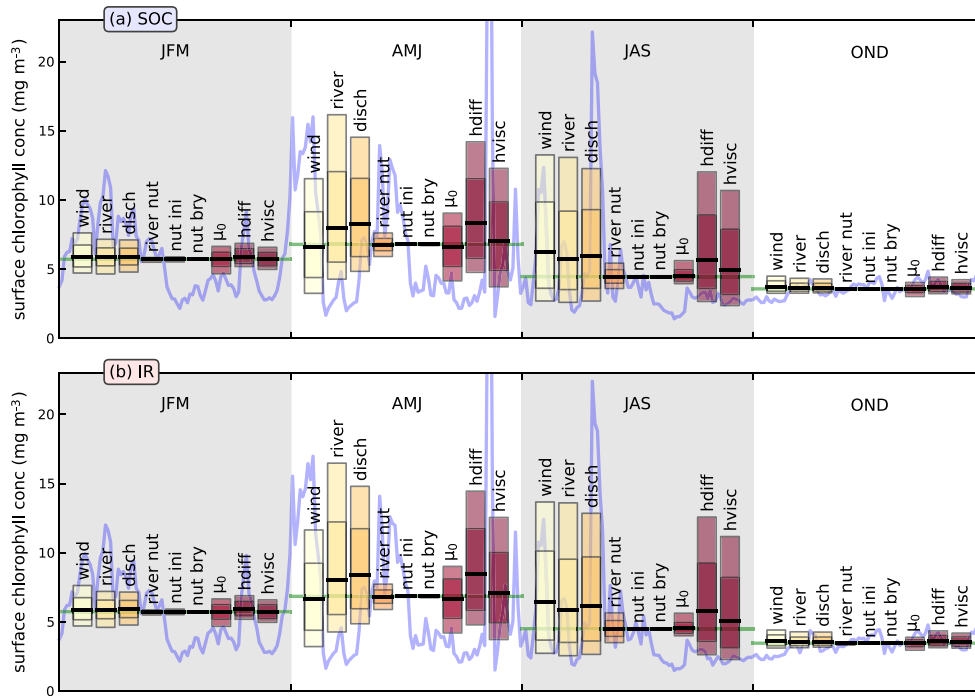
Figure 12. Three-monthly (time-averaged) quantiles of the surface chlorophyll distribution at the location of the station marked in Figure 5 for 2004 (3 month periods are separated by background color; JFM = "January, February, March", etc.). Each bar marks the region between the 0.1 and 0.9 quantiles (outer bar), the region between the 0.25 and 0.75 quantiles (inner bar), and the median (solid black line) for one uncertain input, in one season in 2004. The blue solid line in the background marks the development of surface chlorophyll in the baseline simulation without uncertainty; its seasonal averages are displayed as horizontal green lines. In contrast to bottom oxygen (Figure 6), differences between the SOC and IR configuration are not very pronounced.
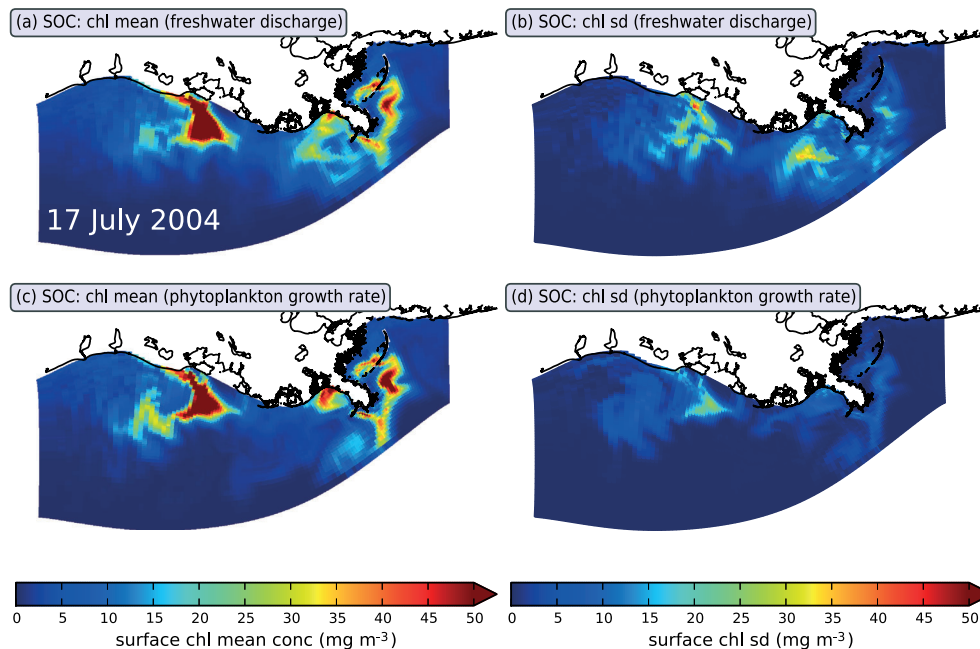

Figure 13. Mean and standard deviation (sd) of surface chlorophyll due to uncertainty the *disch* and μ_0 input for the SOC configuration.

polynomial chaos expansion, to approximate output distributions based on prescribed input distributions. We selected nine inputs from both the physical and biological components of the model and used the same relative amount of uncertainty for all inputs. We focused on uncertainty in estimates of oxygen and bottom water hypoxia expressed through the biological model component. The goals of this study were to examine the effect of different inputs in time and space and to compare their impacts. Two benthic parametrizations, IR and SOC, were considered in the examination. While our experiments were performed using the polynomial chaos expansion, we expect that similar results can be obtained using other emulator approaches or simpler random sampling-based Monte Carlo techniques, albeit the latter typically require a higher computational effort in the form of more model simulations to reach a similar level of accuracy.

[60] In order to quantify the uncertainty in model estimates of hypoxia, we used entropy, a simple function of the probability of hypoxia (Figure 3a). In cases where we are interested in finite numbers of events (here, there are two events: the oxygen at a grid cell concentration can either be below the hypoxic threshold or above), entropy is the suitable way to quantify uncertainty [*Shannon*, 2001]. High entropy values that we encounter on many occasions (Figures 7 or 9) indicate that model hypoxia estimates can be highly uncertain.

[61] As we have shown in section 2.7.1, the emulator-based uncertainty estimates themselves are not free of error. While uncertainties in the biological inputs (river nutrient load, phytoplankton maximum growth rate, nutrient initial, and boundary conditions) are propagated into the output reliably, there is a higher error when estimating the model response to variations in the physical inputs (wind, freshwater discharge, and vertical mixing coefficients; Figure 4). Lower emulation error can be achieved by adding more terms to the polynomial chaos expansion; yet this requires more quadrature points, i.e., more samples from the input distribution, and comes at a much higher computational cost. What is interesting in regard to the emulator error is that it spatially correlates with the region of high model uncertainty (compare Figure 4b and Figure 9). The cause of the error is that the emulator does not sample the output on a sufficiently dense grid to capture its variability. One reason for this model behavior is that, in contrast to the biological inputs, changes in the physical inputs can affect the stochastic flow field, e.g., instabilities and eddies in the model, leading to highly variable small scale changes. It is this high sensitivity of the model to the physical inputs that leads to the emulator error. For this reason, we hypothesize that the high model uncertainty estimates are not caused by a systematic overestimation by the emulator, but rather that both model uncertainty and emulator error are caused by the same phenomenon, the highly variable model output due to perturbations in the circulation. Thus, the emulator estimate is right at least qualitatively by assigning high uncertainty values to the regions in question.

[62] The results of this study offer insights into the sensitivity of the model and allow for an assessment of the impact of the uncertain inputs at each point in time and space of the model simulations. The most obvious result is that two inputs, the nutrient boundary and initial conditions, have very little influence on the simulated bottom oxygen concentrations and the surface chlorophyll content, while the model is considerably more sensitive to the other inputs we investigated. Uncertainty in physical model inputs had the strongest effect on both bottom oxygen and surface chlorophyll. This includes freshwater river runoff, consistent with other studies identifying rivers as large sources of uncertainty in coastal regions [*Blumberg and Georgas*, 2008; *Cossarini et al.*, 2009]. While changes in the nutrient boundary conditions had little effect, our model also has a high sensitivity to the physical boundary conditions, as shown in *Fennel et al.* [2013]. Variations in input to the biological model, including the river nutrient load, had generally a much smaller impact.

[63] Besides model sensitivity, this study provides some insights into model uncertainty, due to errors in the inputs, although we cannot provide a full assessment of model uncertainty here, since the model responses to inputs were observed individually and not all inputs were considered. Most importantly, we deliberately chose the same relative error for all inputs in order to better compare their effect, instead of selecting a specific distribution for each input which would reflect our actual knowledge of the input processes. However, we can draw some important conclusions with regard to model error from our results. Uncertainty in the inputs typically has little effect on oxygen or chlorophyll outside the biologically very productive zone close to the river mouths or outside the spring and summer months (April through September). Within the region and period of higher error, however, uncertainties due to just a single input can be considerable. Estimates of hypoxia are strongly influenced by all of the physical inputs we investigated (Figure 9), as are size estimates of the hypoxic region (Figure 10). Even if the input uncertainties are decreased (in this study, we used a distribution with an standard deviation of 20%), it is very likely that their combined impact on bottom oxygen estimates should lead to high errors in hypoxia estimates.

[64] In section 3.4, we contrasted hypoxic area estimates based on observations with uncertainty estimates due to river inflow. We have shown that the variations we introduced into the river inflow can, in three out of four cases, explain the discrepancy with the observations. Assuming that the error in the observations is relatively small, the predicted uncertainty cannot explain the discrepancy that occurs in 1 year. Here, an adjustment of the river inflow in the model would thus likely not lead to much improved fit to the observations, and one has to turn to other sources of model uncertainty for improvement.

5. Conclusions

[65] We analyzed the impact of uncertainty in nine physical and biological model inputs on hypoxia predictions from a physical-biological model of the Texas-Louisiana shelf in the northern Gulf of Mexico. For the purpose of propagating uncertainty from model inputs to the outputs, we used a model emulator, the polynomial chaos expansion, which offers a straightforward approach to propagating uncertainty. While the computational cost of an emulator-based uncertainty analysis is significant, it offers the possi-

bility of a thorough assessment of model sensitivity for a select number of model inputs.

[66] We found that the impact of the different inputs varied considerably: the physical inputs typically had a stronger effect than the biological inputs, with variations in the river freshwater discharge and the wind field having the strongest impact on both concentrations of bottom oxygen and surface chlorophyll. Two of the biological inputs we evaluated, nutrient initial and boundary conditions, had a negligible effect on the output we assessed. Strong effects were constrained to regions with high biological activity in both time and space. The uncertainty introduced into the system is considerable and especially affects estimates of hypoxia and the hypoxic area. For example, hypoxic area estimates of the model vary by more than 5000 km^2 or by around 40% (probability of hypoxia between 0.4 and 0.6) due to 20% variation in the river inflow.

References

Allen, J. I., J. T. Holt, J. Blackford, and R. Proctor (2007), Error quantification of a high-resolution coupled hydrodynamic-ecosystem coastal-ocean model: Part 2. Chlorophyll-a, nutrients and SPM, *J. Mar. Syst.*, *68*(34), 381–404, doi:10.1016/j.jmarsys.2007.01.005.

Aulenbach, B., H. Buxton, W. Battaglin, and R. Coupe, (2007), Streamflow and nutrient fluxes of the Mississippi-Atchafalaya River basin and subbasins for the period of record through 2005, *Technical report*, U. S. Geological Survey.

Béal, D., P. Brasseur, J.-M. Brankart, Y. Ourmières, and J. Verron (2010), Characterization of mixing errors in a coupled physical biogeochemical model of the North Atlantic: Implications for nonlinear estimation using Gaussian anamorphosis, *Ocean Sci.*, *6*(1), 247–262, doi:10.5194/os-6-247-2010.

Bianchi, T., S. DiMarco, J. H. Cowan Jr., R. Hetland, P. Chapman, J. Day, and M. Allison (2010), The science of hypoxia in the Northern Gulf of Mexico: A review, *Sci. Total. Environ.*, *408*(7), 1471–1484, doi:10.1016/j.scitotenv.2009.11.047.

Blatman, G., and B. Sudret (2010), An adaptive algorithm to build up sparse polynomial chaos expansions for stochastic finite element analysis, *Probabilist. Eng. Mech.*, *25*(2), 183–197, doi:10.1016/j.probengmech.2009.10.003.

Blumberg, A., and N. Georgas (2008), Quantifying uncertainty in estuarine and coastal ocean circulation modeling, *J. Hydraul. Eng.*, *134*(4), 403–415, doi:10.1061/(ASCE)0733-9429(2008)134:4(403).

Clancy, D., J. E. Tanner, S. McWilliam, and M. Spencer (2010), Quantifying parameter uncertainty in a coral reef model using Metropolis-Coupled Markov Chain Monte Carlo, *Ecol. Model.*, *221*(10), 1337–1347, doi:10.1016/j.ecolmodel.2010.02.001.

Cossarini, G., P. F. J. Lermusiaux, and C. Solidoro (2009), Lagoon of Venice ecosystem: Seasonal dynamics and environmental guidance with uncertainty analyses and error subspace data assimilation, *J. Geophys. Res.-Oceans*, *114*(C6), C06,026, doi:10.1029/2008JC005080.

Dagg, M. J., and G. A. Breed (2003), Biological effects of Mississippi River nitrogen on the northern gulf of Mexico–A review and synthesis, *J. Mar. Syst.*, *43*(34), 133–152, doi:10.1016/j.jmarsys.2003.09.002.

Feng, Y., S. F. DiMarco, and G. A. Jackson (2012), Relative role of wind forcing and riverine nutrient input on the extent of hypoxia in the northern Gulf of Mexico, *Geophys. Res. Lett.*, *39*(9), L09,601, doi:10.1029/2012GL051192.

Fennel, K., J. Wilkin, J. Levin, J. Moisan, J. O'Reilly, and D. Haidvogel (2006), Nitrogen cycling in the Middle Atlantic Bight: Results from a three-dimensional model and implications for the North Atlantic nitrogen budget, *Global Biogeochem. Cycles*, *20*(3), GB3007, doi:10.1029/2005GB002456.

Fennel, K., R. Hetland, Y. Feng, and S. DiMarco (2011), A coupled physical-biological model of the Northern Gulf of Mexico shelf: Model description, validation and analysis of phytoplankton variability, *Biogeosciences*, *8*(7), 1881–1899, doi:10.5194/bg-8-1881-2011.

Fennel, K., J. Hu, A. Laurent, M. Marta-Almeida, and R. Hetland (2013), Sensitivity of hypoxia predictions for the Northern Gulf of Mexico to sediment oxygen consumption and model nesting, *J. Geophys. Res. Oceans*, *118*, doi:10.1002/jgrc.20077.

Forrest, D. R., R. D. Hetland, and S. F. DiMarco (2011), Multi-variable statistical regression models of the areal extent of hypoxia over the Texas–Louisiana continental shelf, *Environ. Res. Lett.*, *6*(4), 045,002, doi:10.1088/1748-9326/6/4/045002.

Gibson, G., and Y. Spitz (2011), Impacts of biological parameterization, initial conditions, and environmental forcing on parameter sensitivity and uncertainty in a marine ecosystem model for the Bering Sea, *J. Mar. Syst.*, *88*(2), 214–231, doi:10.1016/j.jmarsys.2011.04.008.

Haidvogel, D., et al. (2008), Ocean forecasting in terrain-following coordinates: Formulation and skill assessment of the regional ocean modeling system, *J. Comput. Phys.*, *227*(7), 3595–3624, doi:10.1016/j.jcp.2007.06.016.

Hetland, R. D., and S. F. DiMarco (2008), How does the character of oxygen demand control the structure of hypoxia on the Texas–Louisiana continental shelf? *J. Mar. Syst.*, *70*(12), 49–62, doi:10.1016/j.jmarsys.2007.03.002.

Hetland, R. D., and S. F. DiMarco (2012), Skill assessment of a hydrodynamic model of circulation over the Texas–Louisiana continental shelf, *Ocean Model.*, *4344*(0), 64–76, doi:10.1016/j.ocemod.2011.11.009.

Jaynes, E. T. (1957), Information theory and statistical mechanics, *Phys. Rev.*, *106*, 620–630, doi:10.1103/PhysRev.106.620.

Karniadakis, G. E., and J. Glimm (2006), Uncertainty quantification in simulation science, *J. Comput. Phys.*, *217*(1), 1–4, doi:10.1016/j.jcp.2006.06.009.

Kim, S., R. M. Samelson, and C. Snyder (2010), Toward an uncertainty budget for a coastal ocean model, *Mon. Weather Rev.*, *139*(3), 866–884, doi:10.1175/2010MWR3352.1.

Laurent, A., K. Fennel, J. Hu, and R. Hetland (2012), Simulating the effects of phosphorus limitation in the Mississippi and Atchafalaya River plumes, *Biogeosciences*, *9*(11), 4707–4723, doi:10.5194/bg-9-4707-2012.

Lermusiaux, P. F. (2006), Uncertainty estimation and prediction for interdisciplinary ocean dynamics, *J. Comput. Phys.*, *217*(1), 176–199, doi:10.1016/j.jcp.2006.02.010.

Makler-Pick, V., G. Gal, M. Gorfine, M. R. Hipsey, and Y. Carmel (2011), Sensitivity analysis for complex ecological models. A new approach, *Environ. Modell. Softw.*, *26*(2), 124–134, doi:10.1016/j.envsoft.2010.06.010.

Mattern, J. P., K. Fennel, and M. Dowd (2010), Introduction and assessment of measures for quantitative model-data comparison using satellite images, *Remote Sens.*, *2*(3), 794–818, doi:10.3390/rs2030794.

Mattern, J. P., K. Fennel, and M. Dowd (2012), Estimating time-dependent parameters for a biological ocean model using an emulator approach, *J. Mar. Syst.*, *9697*(0), 32–47, doi:10.1016/j.jmarsys.2012.01.015.

Le Maître, O. L., O. Knio, H. Najm, and R. Ghanem (2004), Uncertainty propagation using Wiener-Haar expansions, *J. Comput. Phys.*, *197*(1), 28–57, doi:10.1016/j.jcp.2003.11.033.

Melbourne-Thomas, J., C. Johnson, and E. Fulton (2011), Characterizing sensitivity and uncertainty in a multiscale model of a complex coral reef system, *Ecol. Model.*, *222*(18), 3320–3334, doi:10.1016/j.ecolmodel.2011.07.014.

O'Hagan, A. (2006), Bayesian analysis of computer code outputs: A tutorial, *Reliab. Eng. Syst. Safe.*, *91*(10-11), 1290–1300, doi:10.1016/j.ress.2005.11.025.

Rabalais, N., R. Turner, B. Sen Gupta, D. Boesch, P. Chapman, and M. Murrell (2007), Characterization and long-term trends of hypoxia in the northern Gulf of Mexico: Does the science support the action plan? *Estuar. Coast.*, *30*(5), 753–772.

Rabalais, N. N., R. E. Turner, and D. Scavia (2002), Beyond science into policy: Gulf of Mexico hypoxia and the Mississippi River, *BioScience*, *52*(2), 129–142, doi:10.1641/0006-3568(2002)052[0129:BSIPGO]2.0.CO;2.

Shannon, C. E. (2001), A mathematical theory of communication, *ACM SIGMOBILE MC2R*, *5*(1), 3–55, doi:10.1145/584091.584093.

Shen, C. Y., T. E. Evans, and S. Finette (2010), Polynomial chaos quantification of the growth of uncertainty investigated with a Lorenz model, *J. Atmos. Oceanic Tech.*, *27*(6), 1059–1071, doi:10.1175/2009JTECHO727.1.

Thacker, W., A. Srinivasan, M. Iskandarani, O. Knio, and M. Le Hnaff (2012), Propagating boundary uncertainties using polynomial expansions, *Ocean Model.*, *4344*(0), 52–63, doi:10.1016/j.ocemod.2011.11.011.

Wiener, N. (1938), The homogeneous chaos, *Am. J. Math.*, *60*(4), 897–936, doi:10.2307/2371268.

Xiu, D., and G. Karniadakis (2002), The Wiener-Askey polynomial chaos for stochastic differential equations, *SIAM J. Sci. Comput.*, *24*(2), 619–644, doi:10.1137/S1064827501387826.

Xiu, D., and G. E. Karniadakis (2003), Modeling uncertainty in flow simulations via generalized polynomial chaos, *J. Comput. Phys.*, *187*(1), 137–167, doi:10.1016/S0021-9991(03)00092-5.