Hydrodynamic and Biochemical Impacts on the Development of Hypoxia in the Louisiana–Texas Shelf Part II: Statistical Modeling and Hypoxia Prediction

Yanda Ou¹, Bin Li², Z. George Xue¹,³,⁴

¹Department of Oceanography and Coastal Sciences, Louisiana State University, Baton Rouge, LA, 70803, USA.
²Department of Statistical Sciences, Louisiana State University, Baton Rouge, LA, 70803, USA.
³Center for Computation and Technology, Louisiana State University, Baton Rouge, LA, 70803, USA.
⁴Coastal Studies Institute, Louisiana State University, Baton Rouge, LA, 70803, USA.

Correspondence to: Z. George Xue (zxue@lsu.edu)

Abstract. In this study, a novel ensemble regression model was developed for hypoxic area (HA) forecast in the Louisiana–Texas (LaTex) Shelf. The ensemble model combines a zero-inflated Poisson generalized linear model (GLM) and a quasi-Poisson generalized additive model (GAM) and considers predictors with hydrodynamic and biochemical features. Both models were trained and calibrated using the daily hindcast (2007–2020) by a three-dimensional coupled hydrodynamic–biogeochemical model embedded in the Regional Ocean Modeling System (ROMS). A promising HA forecast is provided by the ensemble model with a low RMSE (3,204 km²), a high R² (0.8005), and a precise performance in capturing hypoxic area peaks in the summers. To test its robustness, the model was further applied to a global forecast model and produces HA prediction from 2019 to 2020 with the adjusted predictors from the HYbrid Coordinate Ocean Model (HYCOM). Predicted HA shows a high agreement with the ROMS hindcast time series (RMSE=4,571 km², R²=0.8178). Our model can also predict the magnitude and onsets of summer HA peaks in both 2019 and 2020 with high accuracy. To the best of our knowledge, this ensemble model is by far the first one providing fast and accurate daily HA predictions for the LaTex Shelf while considering both hydrodynamic and biochemical effects. This study demonstrates that it is feasible to perform regional ocean HA prediction using global ocean forecast.

1 Introduction

The Louisiana–Texas (LaTex) Shelf has become a center of hypoxia (bottom dissolved oxygen, DO<2 mg L⁻¹) study since the 1980s (Rabalais et al., 2002; Rabalais et al., 2007a; Justić and Wang, 2014). Regular mid-summer Shelfwide cruises documented that the area and volume of hypoxic bottom water could reach up to 23,000 km² and 140 km³, respectively (Rabalais and Turner, 2019; Rabalais and Baustian, 2020). The aquatic environments, fisheries, and coastal economies are under threat of recurring hypoxia in summer (Chesney and Baltz, 2001; Craig and Bosman, 2013; De Mutsert et al., 2016; LaBone et al., 2020; Rabalais and Turner, 2019; Rabotyagov et al., 2014; Smith et al., 2014). Water column stratification and sediment oxygen consumption (SOC) are two main factors regulating the formation, evolution, and deconstruction of bottom
hypoxia from mid-May through mid-September (Bianchi et al., 2010; Conley et al., 2009; Fennel et al., 2011, 2013, 2016; Feng et al., 2014; Hetland and DiMarco, 2008; Justić and Wang, 2014; Laurent et al., 2018; McCarthy et al., 2013; Murrell and Lehrter, 2011; Rabalais et al., 2007b; Wang and Justić, 2009; Yu et al., 2015). However, prevailing prediction models for the hypoxic area (HA) rely most on nutrient-induced mechanisms rather than the hydrodynamic features. Turner et al. (2006) built a multiple linear regression model for summer HA prediction using the annual and May nitrogen flux (nitrate+nitrite) of the Mississippi River as the predictors. The model provides a robust annual prediction when no strong wind was present but underestimates the HA in windy years. Obenour et al. (2015) modeled HA using the empirical relationship between HA and bottom DO concentration derived by a Bayesian biophysical model. Their model accounted for primary biophysical processes solved for steady-state conditions, water transport, May total nitrogen loads by rivers, and parameterized water reaeration. Katin et al. (2021) further adjusted the Bayesian model by taking into account river flows, riverine bioavailable nitrogen loadings, and wind velocity in both summer (June–September) and non-summer (November–May) months. Summer riverine inputs are projected using non-summer riverine variables, river basin precipitation, and river basin temperature, while, however, summer wind velocity is resampled from historical records from 1985 to 2016. Therefore, the model is known as a pseudo-forecast model since predictors in future stages only include riverine inputs. The pseudo-forecast model explains 71 % and 41 %–48 % of the variability in hindcast (Del Giudice et al., 2020) and geostatistically estimated HA (Matli et al., 2018), respectively. Another Bayesian model was proposed for summer bottom DO concentration prediction taking account of May total nitrogen loads, distance from the Mississippi River mouth, and downstream velocity (Scavia et al., 2013). The summer HA is determined by hypoxic length (HA=57.8 hypoxic length) derived from summer bottom DO concentration. The model explains 69 % of the variability in observed HA by the mid-summer Shelfwide cruises. Different from linear regression and Bayesian analysis, Laurent and Fennel (2019) developed a weighted mean forecast method calibrated on the May nitrate loads and three-dimensional hindcast simulations (1985–2018). Once calibrated, the model requires the May nitrate loads for the forecast year as the only input to produce the seasonal forecast. The model can explain up to 76 % of the year-to-year variability of the HA observation. However, the model is not favorable for years with strong wind events during summer.

These above-mentioned models share some similar shortages: (1) The effects of water column stratification are not included or only partially considered even though stratification is documented as a crucial factor in regulating HA variability. (2) The information of future conditions is limited although some models are built upon multiple predictors, thus these forecast models are indeed “pseudo-forecast” ones. (3) Most models only capture year-to-year HA variability and fail whenever winds are strong in summers. According to the hindcast results by our three-dimensional coupled hydrodynamic–biogeochemical model described in the accompanying paper (Part I), monthly and daily variabilities of HA cannot be neglected before and after strong wind events. In this study, we aimed to provide a new technique in HA prediction considering both stratification and biochemical effects and executing daily forecasts based on the forecasts of selected predictors. An important assumption is that the future conditions of predictors are accessible. Indeed, it can be fulfilled by using global forecast products such as the HYbrid Coordinate Ocean Model (HYCOM), which provides operational hydrodynamics forecasts for up to one week (eight
days). In the accompanying paper (Part I), we demonstrated that the hypoxic volume and the bottom HA over the LaTex Shelf are highly correlated. The former can be reproduced by the latter alone with a quadratic relationship. Thus, in this study, we focused on bottom HA predictions. The rest of the paper is organized as follows: a detailed description of methods and data is given in section 2. The employment of generalized linear models (GLMs) and generalized additive models (GAMs) is given in section 3. The ensemble HA prediction and its application using the global HYCOM is discussed in section 4.

2 Methods

2.1 Data descriptions

We adapted a three-dimensional coupled hydrodynamic–biogeochemical model embedded in the framework of the Regional Ocean Modeling System (ROMS) on the platform of Coupled Ocean–Atmosphere–Wave–Sediment Transport modeling system (COAWST, Warner et al., 2010) to the GoM (Gulf-COAWST, for detailed descriptions, validations, and results of the numerical model see Part I). Numerical hindcasts (hereafter denoted as ROMS hindcasts or ROMS simulations) are output daily from 1 January 2007 to 26 August 2020 and spatially averaged over the LaTex Shelf. In this study, we aim to produce a fast and accurate daily forecast of the shelf HA using models trained from the ROMS outputs.

2.1.1 Hydrodynamic-related predictors

Both water stratification and bottom biochemical processes modulate the variability of bottom DO concentration in the LaTex Shelf. Potential energy anomaly (PEA, in J m⁻³) is introduced as an estimate of water column stratification according to:

\[ PEA = \frac{1}{H} \int_{-H}^{0} (\bar{\rho} - \rho) gz dz, \tag{1} \]

where \( \rho \) is water density profile (estimated by water temperature and salinity profiles) over water column of depth \( H = h + \eta \), \( h \) is the location of the bed, \( \eta \) is water surface elevation, \( g \) is the gravitational acceleration (9.8 m s⁻²), \( z \) is the vertical axis, \( \bar{\rho} \) is the depth-integrated water density given by \( \bar{\rho} = \frac{1}{H} \int_{-H}^{0} \rho dz \) (Simpson and Hunter, 1974; Simpson et al., 1978; Simpson, 1981; Simpson and Bowers, 1981). The PEA represents the amount of energy per volume to homogenize the entire water column (Simpson and Hunter, 1974). Thus, a greater PEA value represents a more stratified water column. As a river-dominated area, water stratification in the LaTex Shelf is highly affected by freshwater-induced buoyancy from the Mississippi and Atchafalaya Rivers. Sea surface salinity (SSS) is a good proxy in representing the distribution and variability of river freshwater across the shelf. Indeed, the correlation of regionally averaged PEA and SSS is significantly high up to -0.88 (\( p<0.001 \); Figure 1a) which emphasizes the importance of freshwater-induced stratification. Therefore, we considered SSS as another candidate predictor besides PEA.
In the meantime, surface heating and wind mixing are other two factors influencing water stratification (Simpson and Hunter, 1974; Simpson et al., 1978) and can be quantified as follows:

\[
\frac{d(\text{PEA})}{dt} = \frac{agh}{2c} Q - \delta k_a \rho_a W^3, \tag{2}
\]

where \( Q \) is the rate of surface heat input, \( \alpha \) is the volume expansion coefficient, \( c \) is water specific heat capacity, \( \delta \) is coefficient of wind mixing, \( k_a \) is drag coefficient, \( \rho_a \) is humid air density near the sea surface, and \( W \) is the wind speed near sea surface. The first term on the right-hand side of Eq. (2) represents the rate of change of water stratification due to surface heating, while the second term is the rate of working by wind stress contributing negatively to water stratification. Therefore, the heat-induced change of PEA is proportional to the product of heat input and water depth, which is,

\[
\text{d(PEA)}_{\text{heat}} \propto Qh, \tag{3}
\]

The total net heat flux, a sum of net shortwave and net longwave radiation flux, is derived from the National Centers for Environmental Prediction Climate Forecast System (CFSR) 6-hourly products (Saha et al., 2010; 2011) in this study. The term \( Q_h \) is added to the candidate list of predictors and is denoted as \( \text{PEA}_{\text{heat}} \) (heat-induced PEA changes) for simplification.

Daily variability of term \( (\delta k_a \rho_a W^3) \) is dominated by that of \( W^3 \), since the \( \rho_a \) fluctuates much less than the \( W^3 \) in a daily scale (Figure A1). We obtained the \( \rho_a \) according to (Picard et al., 2008):

\[
\rho_a = \frac{p M_d}{ZRT} \left[ 1 - x_v \left( 1 - \frac{M_v}{M_d} \right) \right], \tag{4}
\]

where \( p \) represents the absolute air pressure, \( M_d (=28.96546 \text{ g mol}^{-1}) \) is the molar mass of dry air, \( M_v (=18.01528 \text{ g mol}^{-1}) \) is the molar mass of water vapor, \( Z \) indicates compressibility, \( R (=8.314472 \text{ J mol}^{-1} \text{ K}^{-1}) \) is the molar gas constant, \( T \) is thermodynamic temperature, \( x_v \) is the mole fraction of water vapor. We assumed that air parcels at the sea surface are ideal gases \((Z = 1)\) and are always saturated with water vapor. Thus, \( x_v \) is a function of absolute air pressure \((p)\) and saturation vapor pressure of water \((p_{\text{sat}})\) and can be calculated as follows:

\[
x_v = \frac{p_{\text{sat}}}{p}, \tag{5}
\]

According to the Tetens equation (Monteith and Unsworth, 2014), \( p_{\text{sat}} \) (in Pa) can be estimated for the following:
\[ \rho_{\text{sat}} = 610.78e^{\frac{17.27(T-253.3)}{T}}, \quad (6) \]

Substitute Eqs. (5)–(6) to Eq. (4) with the assumption of \( Z = 1 \), we obtained air density as a function of both air pressure and air temperature in the following:

\[ \rho_a = \rho_a(T, p) = \frac{pM_d}{ZRT} \left[ 1 - \frac{1}{p} \left(1 - \frac{M_g}{M_d}\right) e^{\frac{17.27(T-253.3)}{T}} \right], \quad (7) \]

\[ \rho_a \text{ is then estimated using sea surface air pressure and air temperature 2 meters above the sea surface provided by NCEP CFSR 6-hourly products. Correlation of daily } \rho_a W^3 \text{ and } W^3 \text{ (provided by NCEP CFSR 6-hourly products) is significantly high as 0.9989 } (p<0.001, \text{ Figure A1}) \text{ emphasizing the importance of term } W^3 \text{ in controlling the daily variability of wind-induced PEA changes over the shelf. We, thus, approximated the relationship as:} \]

\[ \text{d}(\text{PEA})_{\text{wind}} \propto W^3, \quad (8) \]

The term \( W^3 \) is introduced as another candidate predictor and is denoted as \( \text{PEA}_{\text{wind}} \) (wind-induced PEA changes) for simplification.

2.1.2 Biochemical-related predictors

Sedimentary biochemical processes directly influence the bottom DO consumption rate. However, by far, global forecast model systems like HYCOM does not include biochemical fields. Therefore, the biochemical-related term SOC needs to be replaced by an alternative term (denoted as SOCalt) that does not rely on biochemical simulations. According to the SOC scheme stated in Eq. (8) and Eq. (10) in Part I, the biochemical features are attributed to the sedimentary particulate organic nitrogen (PONsed, derived from ROMS hindcasts) concentration. The total nitrate and nitrite load by the Mississippi River are used to represent the PONsed variability, because inorganic nitrogen is the primary nutrient resource for plankton bloom. Daily updates of measured riverine nitrate+nitrite loads are accessible from U.S. Geological Survey (USGS) National Water Information System (NWIS). Due to lateral transports and vertical settling of particulate organic matter, a leading period should be introduced to the time series of riverine nutrient loads. The optimal length of leading days is obtained by examining the highest linear correlation of regionally averaged ROMS-hindcast SOC and SOCalt following Eq. (9) and is calculated as 19 days (Figure A2a). The exponential term in Eq. (9) estimates the temperature-dependent decomposition rate of organic matter. A significant correlation coefficient between daily SOCalt and ROMS-hindcast SOC is found as 0.8157 \( (p<0.001, \text{ Figure A2}). \)

\[ \text{SOCalt} = \text{Mississippi River inorganic nitrogen loads (led by 19 days)} \cdot e^{0.0693T}, \quad (9) \]
where $T_b$ indicates bottom water temperature (in °C). Along with SOCalt, the temperature-dependent decomposition rate $e^{0.0693 T_b}$ is also considered as a candidate predictor in statistical models and is denoted as $DCP_{\text{Temp}}$ for simplification.

### 2.1.3 HA estimation

As listed in Table 1, there are six candidate predictors considered in the statistical models including four stratification-related variables (PEA, SSS, (Qh), and $W^3$) and two bottom biochemical variables (SOCalt and $e^{0.0693 T_b}$). For simplification, we denoted this variable as (Qh), $W^3$, and $e^{0.0693 T_b}$ as PEA$_{\text{heat}}$, PEA$_{\text{wind}}$, and $DCP_{\text{Temp}}$, respectively. Correlation coefficients matrix (Figure 1a) indicates that multicollinearity may become a problem in regression models since linear correlations among some predictors are significantly high, e.g., 0.76 ($p<0.001$) between PEA and SOCalt, and -0.88 ($p<0.001$) between PEA and SSS. The frequency distribution of HA (Figure 1b) illustrates that the response variable is highly right-skewed with ~51% of samples (2,506 out of 4,968) being exactly zero. The HA is estimated by the number of hypoxia cells times a constant value (area of the computational cell). Thus, the HA can be estimated by the number of grid cells when the Poisson and negative binomial regression models are applied. However, the great portion of zero samples leads to overdispersion (magnitude of variance $\gg$ magnitude of mean, i.e., 52,161,613 $\gg$ 4,378) and zero-inflated problems (Lambert, 1992). The overdispersion issue violates the mean-variance equality assumption employed in regular Poisson regression models, while zero-inflated problems can weaken the model performances.

**Table 1. Description of daily response variable and candidate predictors. The data cover a time range from 1 January 2007 to 26 August 2020. Prescribed min and max are used for min–max normalization.**

<table>
<thead>
<tr>
<th>Variables [units]</th>
<th>Description</th>
<th>Min</th>
<th>Median</th>
<th>Mean</th>
<th>Max</th>
<th>Prescribed (Min:Max)</th>
</tr>
</thead>
<tbody>
<tr>
<td>HA [km²]</td>
<td>Area of extremely low dissolved oxygen concentration (&lt; 2 mg L⁻¹)</td>
<td>0</td>
<td>0</td>
<td>4,378</td>
<td>40,561</td>
<td>Non-normalized</td>
</tr>
<tr>
<td>PEA [J m⁻³]</td>
<td>Potential energy anomaly measuring the water stratification</td>
<td>3.1</td>
<td>36.9</td>
<td>49.2</td>
<td>190.4</td>
<td>(0:200)</td>
</tr>
<tr>
<td>SSS [non-dim]</td>
<td>Sea surface salinity</td>
<td>20.7</td>
<td>31.8</td>
<td>31.4</td>
<td>34.4</td>
<td>(0:40)</td>
</tr>
<tr>
<td>PEA$_{\text{heat}}$ [W m⁻³]</td>
<td>=Qh, an approximation of surface heat-induced water stratification</td>
<td>-1,472.9</td>
<td>3,986.3</td>
<td>3,717.2</td>
<td>6,829.7</td>
<td>(-2,000:7,000)</td>
</tr>
</tbody>
</table>
PEA\textsubscript{wind} \textsuperscript{3}{m^{-3} s^{-3}} = W^3, an approximation of water stratification changes due to wind mixing

SOCal \textsuperscript{3}{mmol m^{-3} s^{-1}} An alternative term for sediment oxygen consumption.

DCP\textsubscript{Temp}[\text{non-dim}] = e^{0.0693T_b}, temperature-dependent decomposition rate of organic matter

<p>| | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.8</td>
<td>175.1</td>
<td>305.4</td>
<td>6,415.8</td>
</tr>
<tr>
<td></td>
<td>874,870</td>
<td>10,103,864</td>
<td>12,604,970</td>
<td>41,530,153</td>
</tr>
<tr>
<td></td>
<td>2.6</td>
<td>5.1</td>
<td>5.2</td>
<td>8.0</td>
</tr>
</tbody>
</table>

2.2 Data pre-processes

We applied the spatially averaged daily ROMS-derived predictors over the LaTex Shelf, then applied the min–max normalization (Eq. (10)) to the one-dimensional time series. Predictive models can be beneficial from the min–max normalization when applying to a new dataset since the method guarantees that the normalized predictors from different datasets range from 0 to 1 as the minimum and maximum values are prescribed. Note that the response is not normalized.

\[
X_{nor} = \frac{X_{org} - \text{Min}_{prescribed}}{\text{Max}_{prescribed} - \text{Min}_{prescribed}},
\]

where \(X_{nor}, X_{org}, \text{Min}_{prescribed}, \) and \(\text{Max}_{prescribed}\) represent normalized value, original value, prescribed minimum, and prescribed maximum, respectively. The daily samples are then split into a training set (for model construction) accounting for 80 \% of the total samples and a test set (for assessment of model performances) accounting for the rest 20 \%. To maintain the HA distribution in both sets, a random resampling method is applied in different HA intervals individually. For example, 80 \% of samples with HA=0 is chosen randomly for the training set out of all daily samples with HA=0, while the rest of samples with HA=0 is grouped into the test set. The HA=0 is the first interval to which the resampling process is applied, while the rest of samples are split every 5,000 km\textsuperscript{2}. However, the distribution of HA from each year is similar with a right-skewed structure and numerous zero values. Thus, even though through random processes, both the training and test sets contain samples from each year including samples with non-peak and peak HA. Samples shown in Figure 4 are listed sequentially in the time dimension from 2007 to 2020 but are not equally distributed along time, which means that the listed samples should not be regarded as time series. This splitting method increases the model applicability and provides a comprehensive assessment of prediction performances on both non-peak and peak HA.
3 Model construction

3.1 Model built-up process

Several regression models are explored using the statistical programming language R. To find the “best” model balancing both model interpretability and prediction performance, a procedure is conducted for model selection (Figure 2) and is summarized below. (1) Choose a regression model. (2) Apply an exhaustive best-subset searching approach to the chosen model. Models with possible combinations of candidate predictors from the ROMS training set are built. A 10-fold cross-validation (CV) method is applied to each model yielding 10 root-mean-square errors (RMSEs) and 1 corresponding mean. The candidate predictors of PEA and SOCalt are forced into each subset. Thus, the number of fitted models with a subset size of $k$ is $C(6 - 2, k - 2) = \frac{4!}{(6-k)!((k-2)!}, 2 \leq k \leq 6$ (the total number of candidate predictors is 6). The optimal subset of this size is found as the one with the lowest mean CV RMSE among these models. The best subset is then obtained by comparing mean CV RMSEs of the optimal subsets of different sizes. (3) Steps (1)–(2) are repeated for the selected M candidate regression models. (4) Prediction performances of different models with the corresponding best subsets are assessed by the 10-fold CV RMSEs and Bootstrap (1,000 iterations) aggregating (i.e., Bagging) ensemble algorithms. The Bagging method builds the given model $N (=1,000)$ times during each of which the given model is trained using different samples chosen randomly and repeatedly from the ROMS training set and is executed for HA prediction using samples in the ROMS test set. The ensemble means and ensemble 95 % prediction intervals (PIs) of forecast HA are given according to the prediction results in the 1,000...
The best model (Model X in Figure 2) is chosen according to the comparisons of the 10-fold CV RMSEs and the Bagging results.

Figure 2. A flow chart of building up regression models.

3.2 Generalized linear models (GLMs)

3.2.1 Regular GLMs and zero-inflated GLMs

The response variable can be treated as count data. Regular Poisson (function glm in R package “stats” version 3.6.2), quasi-Poisson (function glm in R package “stats” version 3.6.2), and negative binomial (function glm.nb in R package “MASS” version 7.3-54; Venables and Ripley, 2002) GLMs are explored in this section. The latter two GLMs are known for solving overdispersion problems by relaxing the mean-variance equality assumption. These GLMs make use of a natural log link function. Thus, a natural logarithm of the area of a single ROMS cell (~ 25.56 km$^2$) is added to the models as an offset term (an additional intercept term).

In addition, the overdispersion issue can result from the great percentage (~51%) of zero values in the response variable (Figure 1b). Zero-inflated GLMs (using function zeroinfl in R package “pscl” version 1.5.5; Jackman, 2020; Zeileis et al., 2008) are developed for dealing with response variables of this kind. Rather than resetting dispersion parameters, a zero-inflated count model is a two-component mixture model blending a count model and a zero-excess model. The count model is usually a Poisson or negative binomial GLM (with log link), while the zero-excess model is a binomial GLM (with logit link in this study) estimating the probability of zero inflation. An offset term of log (25.56) is also introduced into the count model. Instead of applying the best-subset searching to the count and zero-excess models simultaneously, in this study, the searching...
is conducted respectively for these two models to reduce demands of computational resources. The best subset of the zero-excess model (binomial GLM) is given first. The best subset of the count model (Poisson or negative binomial GLMs) is then provided blending the zero-excess model with the corresponding selected best subset fixed.

However, it is hard to determine whether a given zero value of HA is excessive, instead, it is relatively easy to model hypoxia occurrence assuming that all the zero values are excessive. A new binary response, hypoxia, stated in Eq. (11) is introduced for modeled hypoxia occurrence using regular binomial GLMs (function glm in R package “stats” version 3.6.2). The hypoxia is equal to 0 when HA is 0 (no hypoxia), otherwise, is equal to 1. The optimal model selected three predictors: PEA, SOCalt, and DCP_{temp} (Figure 3b).

\[
\text{hypoxia} = \begin{cases} 
0, & \text{no hypoxia} \\
1, & \text{hypoxia occurs}
\end{cases}
\] (11)

### 3.2.2 Performance of GLMs

The zero-inflated Poisson GLM serves as the best GLM in terms of prediction performances since it has the lowest mean CV RMSE (Figure 3a) among the five candidate GLMs. The relaxation of the mean-variance equality assumption by the negative binomial GLM and the quasi-Poisson GLM does not guarantee salient improvement of performances when comparing their CV RMSEs to those of regular Poisson GLM. The zero-inflated negative binomial GLM yields poorest performance with the largest mean CV RMSE. The mean CV RMSEs of zero-inflated Poisson GLM hit the trough (3,621 km²) at the size of five. However, the greatest drop of RMSEs (3,671 km²) occurs at the size of three beyond which the RMSEs remain stable. It is worth considering a model with fewer predictors satisfying model interpretability. Thus, the best zero-inflated Poisson GLM accounts for three predictors (PEA, SOCalt, and DCP_{temp}) in the count model and three predictors (PEA, SOCalt, and DCP_{temp}) in the zero-excess model. As indicated in the correlation matrix (Figure 1a), the robustness of a model can be impaired by multicollinearity which can be estimated by variance inflation factors (VIFs). VIFs among the selected predictors are 2.60, 2.43, and 1.23 for PEA, SOCalt, and DCP_{temp}, respectively. The VIFs are all less than 5 suggesting that predictors subsets involved in both the count and the zero-excess models are merely violated by multicollinearity. For simplicity, the best zero-inflated Poisson GLM is symbolized as GLMzip3.

The Bagging ensemble method is implemented to estimate the prediction performance of GLMzip3 (Figure 4a). The Bagging means of predicted HA provides an RMSE of 3,610 km² and an R² of 0.7511 against the ROMS hindcasts. The Bagging 95 % PIs of the predicted values are restricted within a narrow range with a slight increase at the predicted peaks. We want to address that the comparisons are not between the time series. The training set and test set are resampled according to different HA intervals, while the distributions of HA in each year are similar. Thus, HA in both the training set and test set contains
observations of peak and non-peak values in each year. The results suggest that GLMzip3 is capable of providing not only accurate but also stable HA forecasts. Nevertheless, we noted salient overestimations (e.g., around the 30th and 920th samples) and underestimations (e.g., around the 540th and 830th samples) at some peaks. Instead of the prediction performance at non-peak HA, here we are more focused on forecasts at HA peaks which impose more threats to the shelf ecosystem. In section 3.3, GAMs are investigated with an expectation of further improvements in peak predictions by considering non-parametric or non-linear effects of the predictors.

![Graph](image-url)

**Figure 3.** Comparisons of mean 10-fold CV RMSEs among different regression models with various sizes of predictors subsets. The response variable in (b) binomial GLM and (a) other models is hypoxia occurrence (hypoxia) and hypoxic area (HA), respectively. Note that the CV RMSE of negative binomial GAM with the size of six is out of the range shown. CV RMSE curves of the Poisson GLM, negative binomial GLM, and quasi-Poisson GLM overlap, while those of Poisson GAM and quasi-Poisson GAM overlap. The minimum size of predictor subsets is two since PEA and SOCalt are forced into every subset.
Figure 4. Comparisons of model predicted HA and ROMS-hindcast HA in the test set. RMSEs and R²s are derived between model Bagging mean and ROMS-hindcast HA.

3.2.3 Model interpretation for GLMzip3

We applied the complete ROMS training set to the model construction of GLMzip3 and found the coefficients for PEA, SOCalt, and DCP\textsubscript{Temp} (Table 2) are all significantly positive ($p<0.001$) in the count model, while coefficients for these predictors are significantly negative ($p<0.001$) in the zero-excess model. The count model simulates the HA while the zero-excess model estimates the probability of HA being zero. Higher PEA is consistent with stronger water stratification, while higher SOCalt and DCP\textsubscript{Temp} are both corresponding to higher sediment oxygen consumption. Therefore, there is no surprise that higher PEA, SOCalt, and DCP\textsubscript{Temp} are related to greater HA and higher hypoxia occurrence or lower probability of HA being zero. Results indicate that the GLMzip3 essentially builds up reasonable relationships between the response and predictors variables with a high agreement with physical and biochemical mechanisms.

Table 2. Regression coefficients of GLMzip3.

<table>
<thead>
<tr>
<th></th>
<th>Count model coefficients (Poisson with log link)</th>
<th>Zero-excess model coefficients (binomial with logit link)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Estimate  Std. Error  z value  Pr ($&gt;</td>
<td>z</td>
</tr>
<tr>
<td>Intercept</td>
<td>1.9897  0.0021  948.2  &lt;2E-16***</td>
<td>9.1993  0.3181  28.9  &lt;2E-16***</td>
</tr>
<tr>
<td>PEA</td>
<td>2.6763  0.0016  1681.4  &lt;2E-16***</td>
<td>-10.0945  0.5986  -16.9  &lt;2E-16***</td>
</tr>
<tr>
<td>SOCalt</td>
<td>0.9228  0.0014  663.6  &lt;2E-16***</td>
<td>-8.7784  0.5508  -15.9  &lt;2E-16***</td>
</tr>
<tr>
<td>DCP\textsubscript{Temp}</td>
<td>3.5940  0.0031  1168.2  &lt;2E-16***</td>
<td>-9.4939  0.4346  -21.9  &lt;2E-16***</td>
</tr>
</tbody>
</table>
Significance codes: 0 (***), 0.001 (**), 0.01 (*)

Log-likelihood: -2.4E6 on 8 degrees of freedom

3.3 Generalized additive models (GAMs)

GAMs are explored with an expectation of improving prediction performance in HA peaks by introducing non-parametric effects of predictors. Using function "gam" in R package “mgcv” (version 1.8-36; Wood, 2011) with smooth functions as pure thin plate regression splines (degree of freedom=9; Wood, 2003), three GAMs are studied and compared, i.e., Poisson GAM, quasi-Poisson GAM, and negative binomial GAM. Following the same procedure in GLM exploration, the best subset searching approach is applied to the GAMs first. The mean 10-fold CV RMSEs for the Poisson and quasi-Poisson GAMs (Figure 3a) exhibit insignificant differences and are the lowest among those for all GLMs and GAMs studied. Although the mean CV RMSEs for these two types of GAMs both reach the lowest at the size of five, the best size is considered as three (considering PEA, SOCalt, and DCP\_temp) at which CV RMSEs exhibit most saline decline, and beyond which mean CV RMSEs stabilize around 3,200 km². The negative binomial GAM has the greatest mean CV RMSEs among the GAMs studied and has an extremely high mean CV RMSE at the size of six. It is, therefore, dropped out of the list of candidate models. The quasi-Poisson GAM with three predictors involved (symbolized as GAMqsp3) is chosen as the best GAM since it relaxes the mean-variance equality assumption which should not be applied to the HA dataset due to the overdispersion issue.

Component plots of model GAMqsp3 (Figure 5) imply that HA generally increases as the chosen predictors increase. The smooth functions of PEA and DCP\_temp are considerably greater than the smooth function of SOCalt indicating that the contributions of the former two predictors are greater than the effect of SOCalt on the daily variability of HA. Note that the fitted HA equals the summation of all smooth function terms. Such results agree with those found by model GLMzip3. However, the component plots provide more detailed information about the rate of changes of HA. The effective degrees of freedom range from 8 to 8.54 indicating strong non-linear effects of the predictors on the changes of HA. The HA is more sensitive to the predictors in the low-value ranges but becomes nearly stable in the medium- and high-value ranges of predictors. It implies that bottom hypoxia develops rapidly in early summer when water stratification and sediment oxygen demand start to increase. The bottom hypoxic water further extends with a much lower expansion speed as the stratification and SOC further intensify. Nevertheless, the smooth function of PEA is slightly greater also with a more acute slope than those found for SOCalt and DCP\_temp in the medium- and high-value regimes of the predictors. It indicates that the HA variability is more related to the hydrodynamic changes in the shelf than the biochemical effects. The result is consistent with the findings by previous studies of the shelf hypoxia (Yu et al., 2015; Mattern et al., 2013) emphasizing that the physical impacts are stronger than the biological impacts on HA estimates. A short conclusion is made that the GAMqsp3 model provides reasonable interpretations on the hypoxic area mechanisms.
Figure 5. Component plots of model GAMqsp3. Solid black lines represent the mean of the smooth function, while the pink area denotes the range of mean ± 1SE. Numbers in brackets represent effective degrees of freedom for the corresponding smooth terms. Black bars at the x axis indicate the density of corresponding predictors. Dashed black lines are straight lines of zero along the predictor domains. Note that the predictors shown have been normalized.

The prediction performance of GAMqsp3 is estimated using the Bagging ensemble method (Figure 4b). The RMSE and $R^2$ between the Bagging mean and ROMS-hindcast HA is 3,134 km$^2$ and 0.8093, respectively. They are 13 % lower and 8 % higher than the corresponding statistics found for the GLMzip3, respectively, suggesting that GAMqsp3 outcompetes GLMzip3 in terms of overall performance. However, GAMqsp3 tends to produce underestimated predictions at HA peaks (like peaks around the 310th and 920th samples) some of which are overestimated by the GLMzip3. Therefore, instead of determining the best model out of the two, ensemble HA predictions blending efforts of both GLMzip3 and GAMqsp3 are carried out with an expectation to improve model performance in the peak forecast. We assumed that the contributions of GLMzip3 and GAMqsp3 are equally weighted since there is no clue showing the apparent superiority of either model in HA peak predictions. We thus averaged the predicted HA by GLMzip3 and GAMqsp3 and calculated the 95 % PIs given the Bagging results of these models (Figure 4c). As expected, the overall performance of the ensemble forecast is somewhere between the performance of GLMzip3 and GAMqsp3 with an RMSE of 3,204 km$^2$ and an $R^2$ of 0.8005. However, some HA peak events (like peaks around the 310th and 920th samples) which are overestimated by GLMzip3 but are underestimated by GAMqsp3 are accurately predicted by the ensemble approach. The ensemble model provides higher accuracy in peak forecast given minor sacrifices in overall performance.
4 Discussion

A promising HA forecast is provided by the ensemble model with a low RMSE (3,204 km$^2$), a high $R^2$ (0.8005), and a precise performance in capturing hypoxic area peaks in the summers. The power of the prediction model relies on the availability of the forecast of predictors. In this section, we discuss the model's transferability using an independent global ocean product.

The Global Ocean Forecasting System (GOFS) 3.1 provides global daily analysis products and an eight-day forecast in a daily interval with a horizontal resolution of 1/12 °. The products (hereafter referred to HYCOM-derived products) are derived by a 41-layer HYCOM global model with data assimilated via the Navy Coupled Ocean Data Assimilation (NCODA) system (Cummings, 2005; Cummings and Smedstad, 2013). Daily data from 1 January 2007 to 26 August 2020 are retrieved and studied. Predictors of PEA, SOCalt, and DCP$\text{temp}$ are reconstructed using HYCOM-derived variables and Mississippi River daily total nitrate and nitrite loadings downloaded from the USGS NWIS. Relationships of ROMS-derived and HYCOM-derived predictors are examined in Figure 6. The magnitudes of HYCOM-derived SOCalt and DCP$\text{temp}$ match up with the corresponding ROMS-derived predictors, respectively, although HYCOM-derived predictors are found slightly greater. Simple linear regression for these predictors illustrates that the linear relationships between the ROMS and HYCOM products are significant with the $R^2$ ranging from 0.93 to 0.95. The intercept terms are at least one-order smaller than the magnitudes of corresponding predictors. Therefore, the HYCOM global products are deemed to agree with the ROMS hindcasts for SOCalt and DCP$\text{temp}$. Nevertheless, the magnitude of HYCOM-derived PEA is found much lower than the ROMS-derived PEA (Figure 6a). Simple linear regression indicates a significant linear relationship between the natural log transformation of PEA from the two datasets ($R^2=0.69$).

At land-sea interfaces, the HYCOM global model is forced by monthly riverine discharges, which weakens the model performance in coastal regions. The hydrodynamics in the LaTex Shelf is highly affected by the freshwater and momentum from the Mississippi and the Atchafalaya Rivers. Monthly river forcings in HYCOM are essentially weaker than daily forcings used in our ROMS set up and can result in a less stratified water column (i.e., lower PEA). Therefore, it is necessary to scale the magnitude of HYCOM-derived PEA to that of the ROMS hindcast. It can be achieved by using the natural log transformation and simple linear regression as discussed. We then adjusted HYCOM-derived PEA but kept the HYCOM-derived SOCalt and DCP$\text{temp}$ unchanged before the application of the ensemble model.

The Bagging approach is implemented again to assess the performances of the ensemble model. During each iteration (N=1,000), the GLMzip3 and GAMqsp3 are trained using the ROMS training set and then applied to the adjusted HYCOM-derived predictors for HA prediction from 1 January 2019 to 26 August 2020. The ensemble method provides averages and 95% PIs of predicted HA blending Bagging results by GLMzip3 and GAMqsp3. Compared to the ROMS-hindcast HA, the ensemble model performs an overall accurate HA forecast with an RMSE and an $R^2$ of 4,571 km$^2$ and 0.8178, respectively.
(Figure 7). The HA peaks in both 2019 and 2020 summers are well captured by the model with slight underestimates at the first peak and slight overestimates at the second. The width of 95% PI is larger during high HA periods suggesting less stability in the HA peak forecast. Discharges measurements for the Mississippi and the Atchafalaya Rivers are provided and updated daily by USGS NWIS, assuring a possible improvement of HYCOM model performance in the LaTex Shelf. Once the performance of hydrodynamics predictions in the LaTex Shelf is guaranteed, predictions performance of the ensemble model on HA would be further improved.

Figure 6. Scatter plots of (a) log(PEA) (unit: log (J m⁻³)), (b) SOCalt (unit: mmol m⁻³ s⁻¹), and (c) DCPtemp (unit: 1) between ROMS and HYCOM simulations. Note that the solid red lines represent linear regression lines, while the dashed grey lines are diagonals with a slope of 1 and an intercept of 0. Daily data compared are from 2007 to 2020.
Figure 7. Daily time series of predicted HA by ensemble model ((GLMzip3+GAMqsp3)/2) when applied to adjusted HYCOM products and ROMS-hindcast HA from 2019 to 2020.

5 Conclusion

In this study, an ensemble HA forecast model for the LaTex Shelf is developed using the state-of-the-art statistic programming language R. The model is trained using numeric simulations from 1 January 2007 to 26 August 2020 generated by a coupled hydrodynamic–biogeochemical model. Before splitting data into a training set and a test set, we applied regional average over the LaTex Shelf and min-max normalization to the hindcast data.

Multiple GLMs (regular Poisson GLMs, quasi-Poisson GLMs, negative binomial GLMs, zero-inflated Poisson GLMs, and zero-inflated negative binomial GLMs) and GAMs (regular Poisson GAMs, quasi-Poisson GAMs, and regular negative binomial GAMs) are assessed for HA prediction. Comparisons of model prediction performance illustrate that an ensemble model combing the prediction efforts of a zero-inflated Poisson GLM and a quasi-Poisson GAM provides the most accurate HA forecast with a variability explanation high up to 80 %, a low overall RMSE of 3,204 km², and a high precision in forecasting peak HA when compared to the hindcasts by the coupled model. Predictors PEA, SOCalt, and DCP\text{Temp} are involved in the GLM and GAM. Statistically significant coefficients for the predictors (for the GLMzip3) and component plots (for the GAMqsp3) agree well with the physical and biochemical mechanisms.

The ensemble model is then migrated to the GOFS 3.1 products based on HYCOM, which provides eight-day forecast of global hydrodynamics. The ensemble model is trained using the ROMS training set and then is used for the HA prediction covering the period from January 1st, 2019 to August 26th, 2020. The prediction is robust when compared to the ROMS simulations (2019–2020), with a low overall RMSE (4,571 km²) and a high $R^2$ (0.8178). The model can also accurately predict the magnitude and onset of summer HA peaks in 2019 and 2020, respectively. To our best knowledge, this ensemble model is the first model providing efficient yet accurate daily HA forecast for the LaTex Shelf while considering both hydrodynamic and biochemical effects. This model is also the first model successfully applying global hydrodynamic forecast in regional HA predictions.
Author contribution: Bin Li and Z. George Xue designed the experiments and Yanda Ou carried them out. Yanda Ou developed the model code and performed the simulations. Yanda Ou, Bin Li, and Z. George Xue prepared the manuscript.

Competing interests: The authors declare that they have no conflict of interest.

Acknowledgment: Research support was provided through the Bureau of Ocean Energy Management (M17AC00019, M20AC10001). We thank Dr. Jerome Fiechter at UC Santa Cruz for sharing his NEMURO model codes. Computational support was provided by the High-Performance Computing Facility (clusters SuperMIC and QueenBee3) at Louisiana State University.
Appendix A:

Figure A1. A scatter plot of $\delta a W^3$ and $W^3$ and their linear correlation.

Linear correlation = 0.9989 ($p < 0.0001$)
Figure A2. (a) Lead/lag correlation coefficients between ROMS hindcast daily SOC and SOCalt \( (\text{Mississippi River inorganic nitrogen loads } \cdot e^{0.06937t}) \) with the Mississippi nitrogen loads led by different days; (b) daily time series of ROMS hindcast SOC and SOCalt when the Mississippi nitrogen loads led by 19 days. Time series of compared is averaged over the LaTex Shelf and is normalized.

Reference


