Satellite prediction of coastal hypoxia in the northern Gulf of Mexico

Yingjie Li [a], Samuel V.J. Robinson [b], Lan H. Nguyen [b,*], Jianguo Liu [a,*]

[a] Center for Systems Integration and Sustainability, Department of Fisheries and Wildlife, Environmental Science and Policy Program, Michigan State University, East Lansing, MI 48823, USA
[b] Department of Biological Sciences, University of Calgary, Calgary, AB T2N 1N4, Canada

ARTICLE INFO

ABSTRACT

The growing number, size, and frequency of coastal hypoxia increasingly threaten marine ecosystem health and essential ecosystem services for human well-being. It is therefore urgent to use continuous and consistent observation and develop advanced tools to characterize and track the spatial and temporal change of coastal hypoxia. Satellite imagery with fine spatiotemporal resolution and global coverage has shown great potential for monitoring environmental changes, yet has rarely been applied to hypoxia mapping. To advance the understanding, we synthesized satellite-derived ocean color variables and dissolved oxygen measurements collected during 2014, and used random forest regression, lagged linear regression, and functional data analysis to estimate the spatiotemporal change of the hypoxia zone in the Gulf of Mexico. The three models achieved similar predictive accuracy ($\pm$1.2–1.4 mg/L dissolved oxygen), but the random forest regression performed the best in estimating the bottom dissolved oxygen from satellite-derived variables. Our models also revealed time lags of roughly 0–5 and 16–19 days between the surface water process (e.g., algae bloom and ocean warming) and bottom water hypoxia, which was rarely considered in previous hypoxia studies using satellite data. Finally, our models showed that the area of Gulf hypoxia increased gradually from May and reached a peak during mid-July and mid-August in 2014, and the hypoxia zone occurred in the estuary of the Mississippi River and Suwannee River during roughly 25% of summer days. In addition to predicting the size of hypoxic zones, our study provides additional information on where, when, and how long hypoxic zones persist with greater spatial details and enables modeling hypoxic zones at near-real-time (e.g., days) temporal scales. More importantly, we demonstrate the great potential of applying satellite remote sensing for spatially explicit hypoxia mapping, which could promote more cost-effective coastal hypoxia monitoring and assessment practices.

1. Introduction

Spreading coastal hypoxia zones (or "dead zones" – areas with a dissolved oxygen level below 2 mg/L) increasingly threatens marine ecosystem health and essential ecosystem services (e.g., fisheries and recreation) (Breitburg et al., 2018). This directly affects 10–12% of the global population who depend on coastal systems for their livelihoods (FAO, 2020). The past few decades have seen a growing number of dead zones due to excess agricultural inputs (e.g., fertilizers) leaking into river systems as well as sea warming (IPBES, 2019). Since 1995, at least 500 coastal dead zones have been reported near coasts (Breitburg et al., 2018), covering a combined area larger than the United Kingdom (Díaz et al., 2019). More alarmingly, the number and extent are also expected to increase under climate change scenarios (Zhang et al., 2013). Despite growing attention to the issue, continuous monitoring efforts have been focused on a few scattered coastal regions because traditional cruise-based observations are labor-intensive and costly (Matli et al., 2020).

Modeling of hypoxia offers a potential solution to this problem. So far, two major approaches have been widely used: process-based and statistical models. Process-based (mechanistic) models are useful to understand the hypoxia processes by accounting for biophysical and biogeochemical processes such as nutrient transport, primary production, and water stratification in order to predict dissolved oxygen (see a brief summary of the process in Fig. 1) (Wang and Justić, 2009; Scavia et al., 2013; Obenour et al., 2015; Fennel et al., 2016; Del Giudice et al., 2020). Yet, these models are often complicated and require tremendous

multidisciplinary knowledge and parameter data to build and calibrate. Statistical (empirical) models offer an alternative and more feasible way to approach hypoxia prediction. Models of this type aim to characterize the empirical statistical relationship between the size of summertime hypoxia zones and nutrient loads along with river discharge (Justić et al., 2007; Greene et al., 2009; Forrest et al., 2011; Turner et al., 2012). Existing studies using statistical models usually only predicted the hypoxia size but did not reveal the geospatial distribution of the hypoxic zones. In summary, two types of models implemented in existing studies have contributed significantly to understanding coastal hypoxia issues. However, few of them have produced spatially explicit information on hypoxia, partly because of the lack of sufficient water measurements (Matli et al., 2018, 2020). Existing cruise-based water measurements are costly and often sparsely distributed in space and time. Even in highly-studied areas such as the Gulf of Mexico, water measurements are still sparse, limiting the accuracy of hypoxia mapping across these large regions (Chen et al., 2014).

Remotely sensed satellite imagery has great potential for monitoring environmental changes at fine spatiotemporal scales, yet has rarely been applied to hypoxia mapping (Chen et al., 2014; Matli et al., 2020). A few studies have indicated that remotely-sensed Chlorophyll a (Chlor-a) from satellite imagery, such as MODIS (Moderate Resolution Imaging Spectroradiometer) and SeaWiFS (Sea-viewing Wide Field-of-view Sensor), can serve as a surrogate for mapping algae bloom and ocean hypoxia by combining in-situ sampling data (Klemas, 2011; Chen et al., 2014; Le et al., 2016). A recent study attempted to use a linear regression approach to characterize the relationship between surface dissolved oxygen (DO) and satellite-derived variables, such as Chlor-a concentration and water temperature (Kim et al., 2020). They found that surface DO and water temperature shows a strong inverse relationship, demonstrating the potential of remote sensing data to model DO. However, very few studies have applied time-series of satellite imagery to predict bottom DO, which is more concerning as hypoxia zones usually occur in the bottom water and cause severe impacts on coastal marine ecosystems (Diaz and Rosenberg, 2008; Rabotyagov et al., 2014; Breitburg et al., 2018; Tomasetti and Gobler, 2020). In addition, there is

usually a time lag between the surface process and the occurrence of bottom hypoxia (Fig. 1), but the relevant time scale is still unclear (Justić et al., 1993; Chen et al., 2014; Zhou et al., 2020). Thus, there is a great need to further explore the relationship between bottom DO and satellite-derived variables, and to develop robust models to estimate coastal hypoxia at fine spatiotemporal resolution.

In this study, we examined three different statistical models to analyze the relationship between multispectral satellite imagery and bottom DO: Random Forest Regression (RFR), Lagged Linear Regression (LLR), and Functional Data Analysis (FDA). Each model has its strengths (see details in Section 2.3). For instance, LLR has a simple structure that can provide fast estimates, while RFR is widely known for its efficiency in capturing hidden patterns in large and complex data, enabling high prediction accuracy (Cutler et al., 2007; Belgiu and Drăguţ, 2016; Chen et al., 2019). This is especially helpful as the formation of dead zones involves many complex biophysical and biogeochemical factors that are not completely understood. Similarly, FDA can incorporate complex time dependency into the model for prediction (Ramsay and Silverman, 2005; Yen et al., 2015). Here, we used the Gulf of Mexico as our area of study and compiled DO measurements at 332 locations across the Gulf area. 14 satellite-derived ocean color and ocean biology variables from the MODIS imagery collection were compiled as model inputs. With these three models and the compiled dataset, this study aims to answer two main research questions: (1) How well can satellite imagery be used to predict bottom dissolved oxygen across different models? (2) What was the spatiotemporal progression of hypoxia in the Gulf?

2. Methods

2.1. Study area

We chose the Gulf of Mexico (Fig. 2) as a case study because it is the largest hypoxic zone in the United States and the second-largest worldwide. Hypoxia in the Gulf is largely caused by excess nutrients from the adjacent human-dominated Mississippi River basin (Diaz and Rosenberg, 2008; Pitcher et al., 2021). The Mississippi River basin
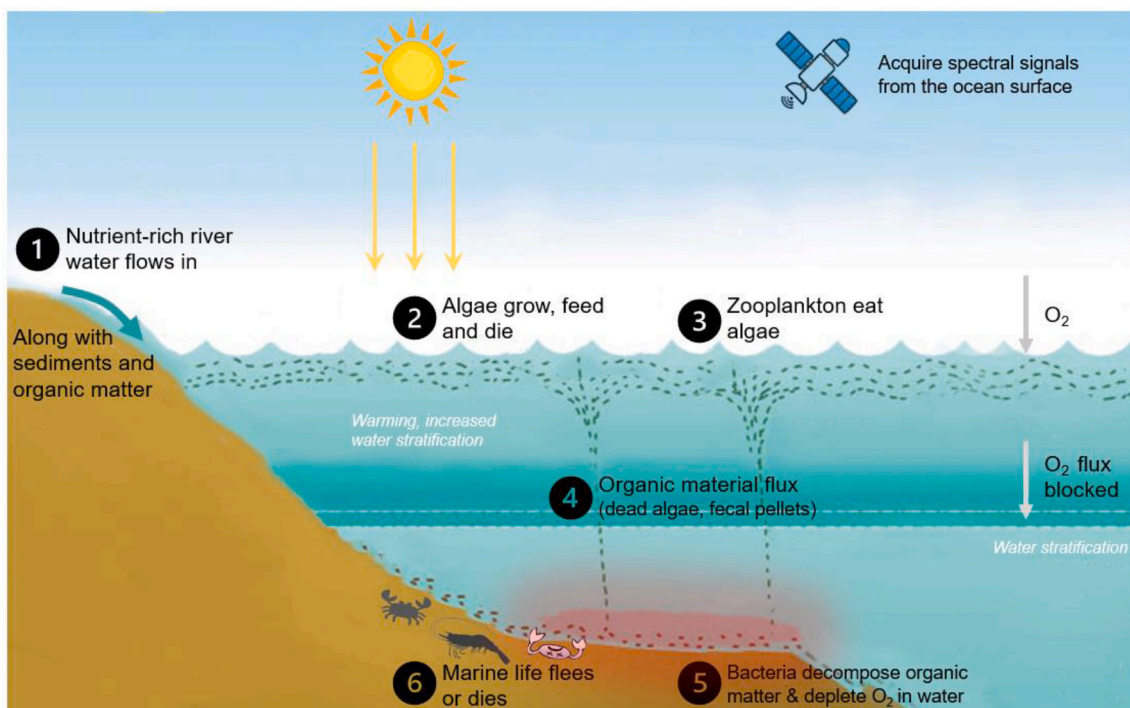

Fig. 1. Nutrient-based hypoxia formation process (based on Dagg and Breed, 2003; Bianchi et al., 2010; https://www.epa.gov/ms-htf/hypoxia-101; https://gulfhypoxia.net/).

Fig. 2. The geographic location of the Gulf of Mexico and the adjacent Mississippi River basin (light yellow region). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

drains approximately 41% of the land area of the conterminous United States, and the combined flows of the Mississippi and Atchafalaya Rivers account for 96% of the annual freshwater discharge and 98.5% of the total annual nitrogen load to the Gulf (Pitcher et al., 2021). It is reported that starting from the last half of the 20th century, there has been a marked increase in the concentration of nitrogen and phosphorus in the Lower Mississippi River, which has been attributed to agricultural activities, especially increased use of fertilizers, in the basin (Mitsch et al., 2001). The maximum spatial extent of the hypoxic zone was measured at 22,720 km^2 during the summer of 2017, which is approximately the same size as the state of New Jersey (NOAA, 2017). We chose 2014 as a focal year for a case study because there are available and abundant DO observation data in the year.

The key mechanism and process of nutrient-induced hypoxia formation are depicted in Fig. 1. Briefly, nutrient-rich water flows from the Mississippi River basin enrich coastal systems and lead to significant increases in algae blooms on the water's surface. Algae have a relatively short life span and sink down to the bottom waters after dying, where they are decomposed by bacteria, causing depletion of oxygen in the bottom water. In addition, during the summer, the global warming effect and freshwater-induced thermal stratification also restrict the vertical mixing of atmospheric oxygen into deeper waters (Dagg and Breed, 2003). As a consequence, oxygen consumption rates at the ocean bottom easily exceed those of resupply and result in hypoxia (Bianchi et al., 2010; Breitburg et al., 2018). Mobile organisms (e.g., fish and crustaceans) may escape the hypoxia area, but less mobile fauna experience stress or die from low oxygen levels, meaning that hypoxia can damage important commercial fisheries in the Gulf of Mexico (Smith et al., 2014).

2.2. Data

We used a coastal water quality dataset from 2014 that included water samples collected at 332 locations in the Gulf area (Fig. S1 – S4), along with a satellite imagery dataset from MODIS. The water quality dataset was collected by the Louisiana Universities Marine Consortium (LUMCON) and the Southeast Area Monitoring and Assessment Program (SEAMAP). The observation dataset includes measurements of dissolved oxygen (DO), temperature, and depth of sampling (Matli et al., 2018, 2020). We selected water samples collected at depths between 3 and 80

m mainly because sampling outside of this depth range is sparse and hypoxia is very rare beyond 80 m (Matli et al., 2018). DO observations collected from May 1 to September 31 were used in model development, as hypoxia occurs predominantly in summer (Matli et al., 2018, 2020; Rabalais et al., 1994; Pitcher et al., 2021). Satellite imagery data for our models were derived from the daily MODIS Aqua and Terra Ocean Color Data (NASA Goddard Space Flight Center, 2018). This level-3 product includes global daily ocean color and satellite ocean biology data at 4616 m resolution that were produced by NASA's Earth Observing System Data and Information System (EOSDIS). We used all available 14 satellite-derived variables, including ten ocean color bands and four ocean biology variables as inputs (see Table S1).

To better predict bottom DO with the corresponding satellite-derived variables, the observation dates of the two datasets must be matched with each other. Since the bottom DO measurements across the Gulf of Mexico were collected on different days (ranging between May 26 – August 2, 2014), we extracted MODIS data on the same day that the DO measurement was taken. In addition, we also extracted MODIS data on the day from 1 day to 60 days before each DO measurement was taken, in order to use a total 61-day satellite time series to identify the time lags between what was observed from space and that on the water bottom. We chose this time scale based on earlier studies that suggest the time lag ranges from days to up to 2 months (Justić et al., 1993; Chen et al., 2014; Zhou et al., 2020).

2.3. Models

We used three types of models — Random Forest Regression (RFR), Lagged Linear Regression (LLR), and Functional Data Analysis (FDA) — to model the relationship between bottom dissolved oxygen concentrations and remotely sensed multispectral information.

2.3.1. Random Forest Regression (RFR)

Random Forest algorithms can handle data with high dimensionality and often achieve good prediction accuracy in comparison with other approaches such as maximum likelihood and single-layer neural networks (Belgiu and Drăguţ, 2016; Teluguntla et al., 2018). To estimate the appropriate time lag for optimal prediction, we successively fed daily satellite-derived variables (as predictors) from day 0 to day 60 to fit bottom DO measurements (as the response variable) using the

random forest algorithm (i.e., we trained different RFR classifiers for each potential time lag). Then we compared the model performance at each time lag and chose the one with the highest performance as our final prediction model. We used the built-in RF algorithm in the GEE cloud computing platform, as it allowed us to develop a complete workflow on a single platform and facilitate the application at regional scales.

2.3.2. Lagged Linear Regression (LLR)

Despite its predictive power, RFR does not offer descriptive information about the effects of individual predictors, aside from variable importance scores (Biau and Scornet, 2016). Linear regression can complement RFR and provide more information about how individual predictors affect DO. We first decomposed all the 14 satellite-derived variables into their principal components (PC) and used the first 5 components, which together represented 97% of the total variance in the satellite data. To reduce satellite data gaps (missing pixels or spectral variables caused by, for example, clouds, sun glint, inter-orbit gaps, ice, low light, etc.), we imputed PC values for pixel locations with fewer than 4 of the 14 variables missing using multivariate data imputation (*missMDA* library, Josse and Husson, 2016) in R. For location/time combinations with more than 4 missing satellite variables, we further considered the spatiotemporal dimensions and imputed the missing values using soap film smoothers from the *mgcv* R package (Wood et al., 2008; Wood, 2017). Smoothers (additive models) are similar to other spatial smoothing procedures (e.g., Kriging) but use basis functions to avoid time-consuming matrix inversion procedures, and soap-film smoothers model processes that wrap around geographic features (coastlines and river deltas, in our case). We used spatiotemporal smoothers (tensor product smooths) that considered both space and time as different types of "distances"; this allowed improved imputation, as missing data are often located close to the observed data in time, but not space (or vice-versa). To further improve computation time, we divided the entire time-series dataset into five time segments and fit smoothers within each segment for each PC were fitted using the 13,577 (30.9%) non-missing satellite data. All models converged properly, met assumptions of residual normality and equal variance, and returned reasonably accurate predictions (mean absolute error of predictions for each PC: 0.528, 0.728, 0.145, 0.270, 0.283).

To estimate the appropriate time lag for DO prediction, we fit linear models at various time lags (from 0 to 60 days, as done for RFR). For lags of $\tau = 0$–60 days, we then fit the following lagged linear model of DO on day *t*, where:

$$\hat{y} = b_0 + b_1 \times PC1_{t-\tau} + b_2 \times PC2_{t-\tau} + \ldots + b_5 \times PC5_{t-\tau}$$

$$DO_t \sim \text{Normal}(\hat{y}, \sigma) \tag{1}$$

2.3.3. Functional Data Analysis (FDA)

To examine the continuous effect of multiple time lags, we used function-on-scalar functional data analysis (FDA; Ramsay and Silverman, 2005). While RFR and LLR examine the effect of a single time lag on bottom DO measurements, FDA integrates the signal from multiple lagged predictors. In standard linear regression, we seek the coefficients β that are mapped onto a matrix of predictors X, where each column is an independent predictor of the data y:

$$\hat{y} = b_0 + X\beta$$

$$y \sim \text{Normal}(\hat{y}, \sigma) \tag{2}$$

In function-on-scalar FDA, the form is slightly different:

$$\hat{y} = b_0 + X\beta$$

$$\beta \sim f(z)$$

$$y \sim \text{Normal}(\hat{y}, \sigma) \tag{3}$$

The equations are similar, except that the rows of X now represent vectors of related data that are continuous predictors of *y* along the continuous domain of *z* (for example, sequences of measurements taken at different depths or time lags). β is now a smoothed function of *z*, meaning that the slope coefficients can change as a function of *z*, and Xβ represents the integrated effect of X on *y*. This is useful if predictions are to be made using functional predictors of either a scalar or a functional response (Yen et al., 2015). In our case, a vector of satellite-derived measurements at lags $\tau = 0$–60 days was used to predict bottom DO measurements.

We fit the following model:

$$\hat{y} = b_0 + X_{PC1}\beta_{PC1} + X_{PC2}\beta_{PC2} + \ldots + X_{PC5}\beta_{PC5}$$

$$\beta_{PCi} \sim f(\tau)_i$$

$$DO_t \sim \text{Normal}(\hat{y}, \sigma) \tag{4}$$

Each matrix X_{PCi} contains i-th principal components, where each row is a vector of the lagged satellite-derived data from $\tau = 0$–60 days for a given dissolved oxygen data point, DOt. β_{PCi} is a vector of coefficients that map onto the matrix, and *f* is a smoothed function (in this case, a penalized cubic spline) of time lag τ. This allows the effect of the spectral PC to change across days in a smooth manner, meaning that $X_{PC1}\beta_{PC1}$ represents the integrated effect of spectral PC1 on days $\tau = 0$–60 on the bottom DO. We used 10 basis dimensions for each smoother, meaning that the minimum length of data used for prediction was a 10-day lag (the highest effective degrees of freedom for the smoothers was 8.8, indicating that 10 basis dimensions were adequate). All lagged linear and FDA models were fit in R 4.1.1 (R Core Team, 2021), and the FDA models were fit using the *mgcv* library (Wood, 2017). FDA requires that all data in the matrix of predictors be non-missing, so satellite data with missing spectral PCs were imputed using the same approach as described for LLR.

2.4. Model evaluation

We calculated root mean square error (RMSE), mean absolute error (MAE), and R^2 to evaluate each model's predictive accuracy at each time lag. To test model performance, we used a random 70% of the samples as the training dataset and the remaining 30% as the testing dataset and repeated this procedure 1000 times to obtain the mean prediction error.

3. Results

3.1. Model performance comparison

Both the functional data analysis (FDA) and lagged linear regression (LLR) models achieved similar hypoxia prediction accuracy, but random forest regression (RFR) ultimately performed the best. The RFR model predicted bottom DO with an RMSE of 1.12 mg/L on the testing dataset (Table 1), and outperformed the other two models (LLR: 1.43, FDA: 1.15). Similarly, the RFR model explained about 66% of the variance in bottom DO, while LLR and FDA explained about 44% and 63%, respectively. RFR performed much better on the training data than on testing data (compare solid black lines to red lines in Fig. 3), possibly reflecting overfitting in the training dataset. Interestingly, the variance in predictions was also different between models, with FDA having a lower variance in test predictions than either LLR or RFR (compare the width of the grey confidence intervals in Fig. 3).

According to the FDA models that used the time series of lagged variables, the model prediction accuracy plateaus at roughly 30 days, indicating that the characteristic scale of the hypoxia process is most likely less than one month (Fig. 3). Further considering LLR and RFR using single-time lags for prediction, we found that, within the one-month range, LLR can achieve relatively high performance with satellite variables taken 16–19 days before. Similarly, RFR models also show

Table 1
Model performance of RFR, LLR, and FDA models.

Model	Image Data Type	Training error			Testing error		
		RMSE	MAE	R^2	RMSE	MAE	R^2
FDA	Gap-filled, daily	1.067	0.809	0.634	1.150 (±0.08)	0.882 (±0.063)	0.630 (±0.027)
LLR	Gap-filled, daily	1.391	1.101	0.459	1.425 (±0.091)	1.128 (±0.076)	0.444 (±0.067)
RFR	Raw, daily	0.508	0.365	0.943	1.121 (±0.257)	0.822 (±0.174)	0.662 (±0.160)

Shown are RMSE (mg/L) and MAE (mg/L), as well as the R^2 between estimated and observed DO. The training error refers to errors (or residuals) from the model using the training dataset (70% of the whole dataset), while testing error refers to errors from the model fit on the training dataset and tested on the withheld 30%. Means ± SDs are shown for the testing error, taken from a set of 1000 independent simulations. For details on daily gap-filling, see the Methods section.
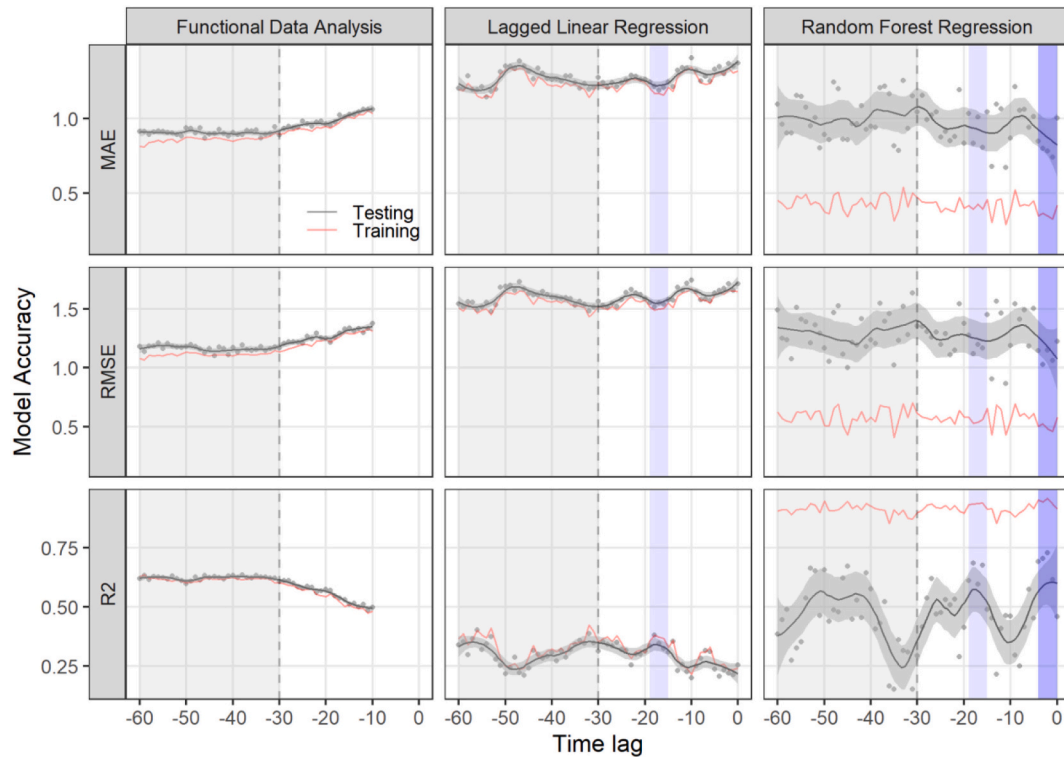

Fig. 3. Training error (red lines) and testing error (black lines/grey shade) for the three modeling approaches. A minimum time lag of 10 days was used for FDA, as 10 basis functions were used for each smoother. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)
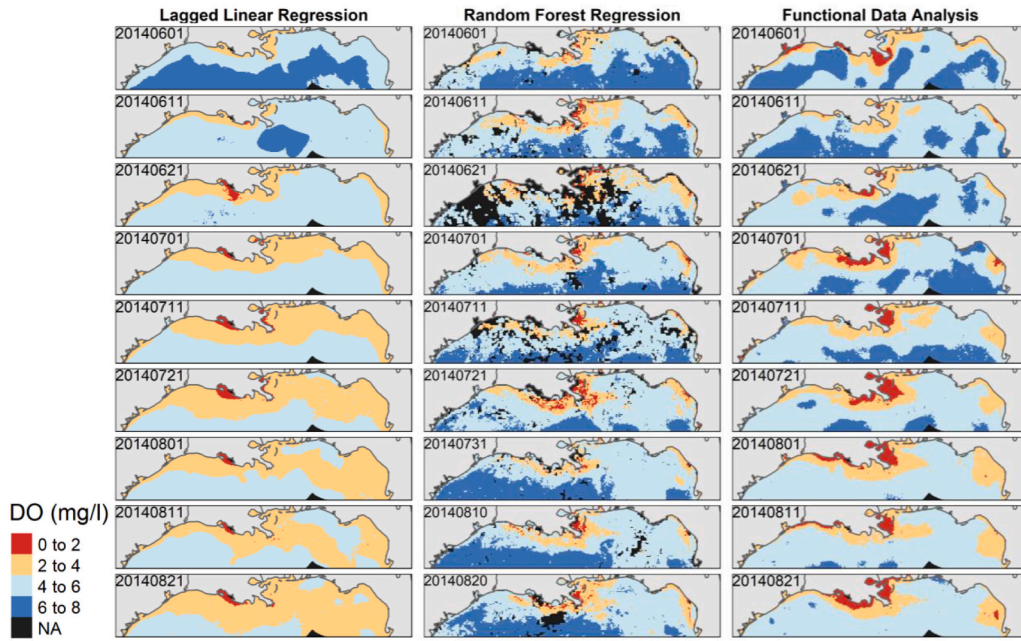
Model accuracy is evaluated by RMSE (mg/L), MAE (mg/L), as well as the R^2 between estimated and observed DO. The x-axis for the FDA models shows how model accuracy changes when more lagged days are added to the vector of predictors, starting at −10 days, and extending to −60 days. The x-axis for the LLR and RFR models shows the single time lags used to fit each model (e.g., −10 indicates that satellite-derived data from 10 days prior was used). Training data outputs are shown in red lines, while testing data are shown in black lines and grey shades. Lines represent mean values, and shaded areas around the lines represent 95% confidence intervals. The blue shaded area indicates the dates of imagery that can predict DO with higher accuracy.

high prediction power at this time window. Additionally, RFR can achieve the highest prediction accuracy with satellite variables from 0 to 4 days prior (Fig. 3). Although lags beyond one month (e.g., LLR with a time lag of 55–60 days and RRF with a time lag of 45–50 days) also performed well, we chose the time ranges that can achieve the highest accuracy for prediction and are also consistent across models (e.g., the one-month range indicated by FDA). These findings suggest that models using satellite data can reveal similar time lags between surface process and bottom hypoxia as indicated in the existing literature (Justić et al., 1993; Chen et al., 2014; Zhou et al., 2013; Zhou et al., 2020). This would be critically useful for choosing satellite imagery at the proper time frame as inputs for more accurately predicting bottom DO and mapping the extent of dead zones.
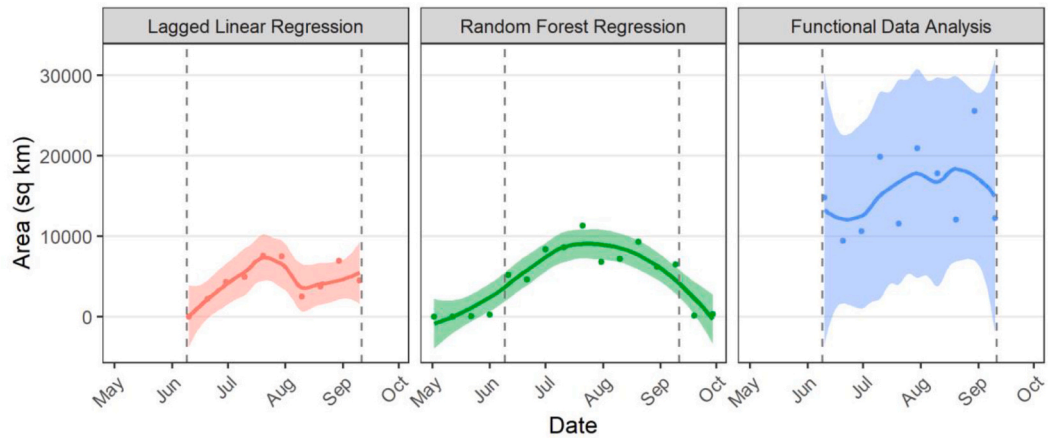
3.2. Spatiotemporal progression of hypoxia in Summer 2014

Using the three developed models, we created hypoxia time-series maps for the Gulf of Mexico (Fig. 4a). The hypoxic area increased gradually starting in May, reached a peak during mid-July and mid-August in 2014, then started to shrink after the end of August (Fig. 4b). The three models estimated the area of the hypoxia zone at about 10,342 km² (standard error, se = 1939) during the peak period in 2014, which is close to the reported hypoxic area (13,080 km²) during the same period by the Hypoxia Research Team at LUMCON (https://gulfhypoxia.net/research/shelfwide-cruises/). FDA (mean = 15,585 km², se = 2281) appears to overestimate the hypoxia area, while LLR (mean = 5828 km², se = 1687) and RFR (mean = 8647 km², se = 1037) appear to underestimate the value likely because of missing pixels
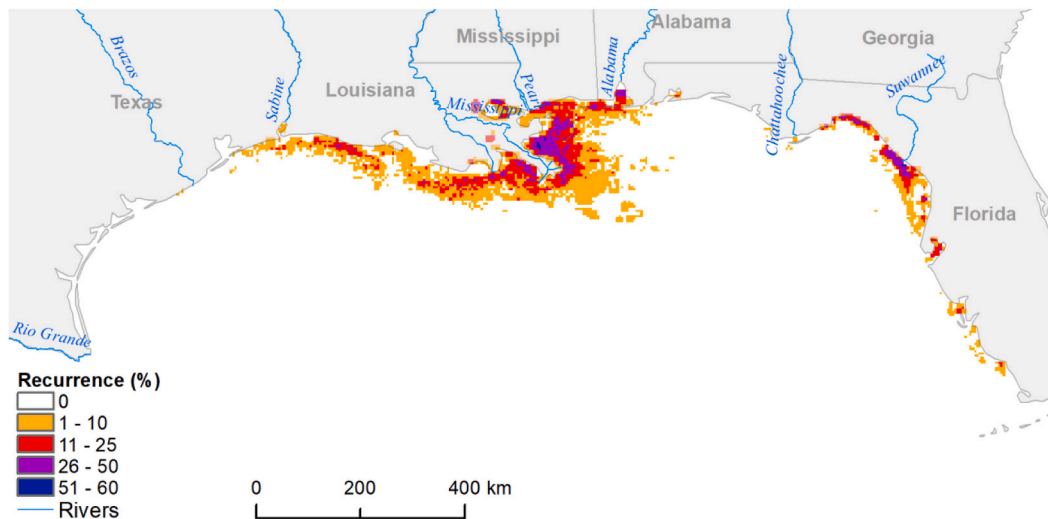
(a)

(b)

(c)

(caption on next page)

Fig. 4. Spatial and temporal change of hypoxia area (a and b), and hypoxia recurrence map (c) in 2014. In (a), The NA values (pixels in black) on the DO maps by the RFR model are due to the use of non-imputed satellite imagery as inputs. For RFR, to close the satellite imagery gap and consider the time lag, we aggregated the daily satellite data by taking the average pixel value across a five-day imagery stack (A 5-day imagery stack can achieve good spatial coverage as well as maintain temporal resolution, see Fig. S5). The individual data points in (b) represent the predicted hypoxia area from June to early September at 10-day intervals, the solid lines represent the smoothed conditional means by the local polynomial regression fitting method, and the shaded areas around the lines represent 95% confidence intervals. The hypoxia recurrence map (c) is based on the RFR model only, which has the highest prediction accuracy. Hypoxia recurrence here was measured by the percentage of time with hypoxia detected during the summertime (from June to September).

in its input imagery (Fig. 4a).

Most of the hypoxia zones were identified in the nearshore regions, and occurred more frequently in the estuary of the Mississippi and Suwannee Rivers (Fig. 4c). We found over one-quarter of the summertime hypoxia occurred in these two estuary areas, while other regions had less persistent hypoxia zones. Although the intensity and extent of the hypoxia zones differ between the three models, we found the general geospatial patterns identified were similar, indicating that all three models were able to capture the underlying signals (Fig. 4a) but with some uncertainties (e.g., errors from image imputations and difference in model predictive power).

4. Discussion

Our analysis shows that satellite imagery can be used to predict coastal hypoxia zones with a high degree of accuracy. The three types of models — random forest regression, lagged linear regression, and functional data analysis — achieved similar prediction accuracy, but the random forest regression performed the best in estimating the bottom DO from satellite-derived information. Our work shows that commonly-used models can be used in conjunction with remotely sensed data to predict and map hypoxic zones. The spatially explicit hypoxia maps can not only reveal the geospatial characteristic of hypoxia (at ~4 km resolution) but also the change over time at finer temporal scales (e.g., across days and weeks).

4.1. Strengths of and uncertainties in our models

Our models provide insight on selecting the appropriate time window of Ocean Color satellite imagery to map hypoxia. Both the LLR and RFR models achieved high prediction accuracy using satellite imagery at 16–19 days before water sampling dates, and RFR further revealed that imagery from 0 to 4 days before water sampling can achieve the highest accuracy, suggesting that the time lag between the water surface processes and the emergence of bottom hypoxic zones might operate on the scale of days to weeks. This aligns with previous work on hypoxia time lags, as Chen et al. (2014) found a similar short lag time (3–4 days) between the peak of Chlor-a concentrations and bottom DO decline on the coast of the East China Sea while Zhou et al. (2020) revealed a lag of 1–8 weeks and Justić et al. (1993) implied a lag of up to 8 weeks. Although our findings are supported by earlier studies, there is still a lack of sufficient and direct evidence to determine the exact time lag between surface processes and bottom hypoxia. One explanation could be that the time lag may vary across years due to changing climate and ocean/river interactions. For example, earlier studies related to time lag were conducted in different years and even across decades (Justić et al., 1993; Justić et al., 2003; Chen et al., 2014; Zhou et al., 2013; Zhou et al., 2020). Since our research focuses on hypoxia prediction using remote sensing data, rather than the complex biological mechanism, we hope future studies can use more explicit process-based "mechanical-statistical" models to address this research gap when detailed and continuous observation data become available. Nevertheless, our results suggest that choosing an appropriate time window of satellite imagery, instead of using monthly aggregates, can help achieve better hypoxia prediction and mapping accuracy.

Our hypoxia mapping encountered a trade-off between achieving high model accuracy and producing seamless DO prediction maps. Smooth DO maps by LLR and FDA predictions used a mixture of raw and imputed data (i.e., using multivariate imputation for pixels with partly-missing variables, and spatiotemporal smoothers for missing pixels). However, the DO maps by RFR are not smooth (i.e., prediction with missing values). This is because neither multivariate imputation nor spatiotemporal smoothers are currently available on the Google Earth Engine platform, meaning that producing smooth DO maps requires downloading and storing large satellite datasets, which can take hours to days to run (depending on the scale of analysis). Therefore, we suggest that LLR and FDA are currently better-suited for smaller-scale analyses, as imputation over large areas is very computationally-intensive and time-consuming. RFR had better prediction accuracy, but fell short of generating seamless maps due to data gaps in satellite imagery collection. Possible ways to achieve both accuracy and seamless mapping could be potentially done by adopting advanced gap-filling techniques for RFR in Google Earth Engine (Moreno-Martínez et al., 2020). This suggests that RFR is more appropriate for regional or global scales of study, where gaps in data are not as much of a problem. However, given that both linear models had similar model fits RFR, they could be used to supplement RFR by producing continuous hypoxia maps across space and over time, even with slightly lower predictive accuracy (Fig. 4a, Fig. S6).

In terms of the variable features derived from satellite imagery, we found that sea surface temperatures (SST), Chlor-a concentrations, and most of the blue bands played relatively important roles in the RFR (Fig. S7). This aligns with previous studies, which revealed that surface water temperature increase can facilitate algae bloom and is also one of the major causes of ocean water stratification (Dagg and Breed, 2003; Rabalais et al., 2009; Matli et al., 2020), as it can make dead zones worse by warming a layer of surface water that locks colder and denser water near to the seafloor (Fig. 1). In addition to surface temperatures, algal blooms (represented by Chlor-a and fluorescence measurements) have been widely recognized as the dominant driver of hypoxia both in freshwater and oceans (Behrenfeld et al., 2009; Scavia et al., 2017; Shen et al., 2019; Jane et al., 2021). Other factors, such as wind change, might affect the mixing of oxygen from the surface into deeper waters (Scavia et al., 2017; Matli et al., 2020), but we did not consider this variable in this study because the lack of spatial data of wind and we wanted to focus on illustrating the usefulness of empirical satellite-derived information in predicting coastal DO. Lastly, like many other statistical hypoxia models, our models cannot capture the movement of organic material flux in the coastal seawater. It is possible that the spatial extent of bottom-water hypoxia is larger than the surface signals (e.g., chlor-a and surface temperature anomaly) over the shelf because of the spread of organic matters. For instance, the estimated hypoxia area by our models ($10,342$ km^2) is close to but smaller than the measured hypoxic area ($13,080$ km^2) by the Hypoxia Research Team.

4.2. Implications and future directions

Missing data imputation presents both a challenge and an opportunity in remote sensing, especially when linking ground-based measurements to satellite-based observations. Imputation is not commonly used in remote sensing (satellite-based data are plentiful but prone to data gaps), but is much more common when using ground-based data (costly to obtain but have fewer data gaps). However, when satellite data are used to model ground-based data, imputation of missing satellite data may be a better approach than simply discarding missing observations. Finally, some kinds of smoothing (either on the input or

the output) are usually required if we want continuous predictions. Bayesian data imputation (Gelman et al., 2013; McElreath, 2020) could be used to fill data gaps, but this comes with the cost of additional model complexity and computation time. We suggest that spatiotemporal processes are a relatively straightforward alternative that could be used to supplement most modeling techniques. Here, our work used a soap-film smoothing approach, as it is appropriate for wrapping around coastal areas, but one could also use standard cubic splines or older techniques such as inverse-distance weighting. Unfortunately, Google Earth Engine currently does not provide any kind of smoothing procedure, so these procedures will have to be done client-side in platforms such as Python or R.

Spatially explicit mapping is critical for estimating the size of coastal hypoxia zones and assessing the socio-environmental impacts. Compared to previous studies that focused on predicting the size (Murphy et al., 2010; Turner et al., 2012; Scavia et al., 2017), this study provides fine-resolution information on where, when, and how long the hypoxia occurred and persisted using satellite data. Although recent studies using geostatistical models can also characterize the spatial and temporal variation in coastal hypoxia zones (Matli et al., 2018, 2020), these studies mainly rely on field measurement data and have not taken into account the abundant information that can be derived from satellite imagery (Matli et al., 2020). Those geostatistical models would be suitable for regions with rich field observation data but might not be able to monitor hypoxia occurrence in regions with few field measurements. Although the important role of on-site field measurements in modeling coastal hypoxia (Rabalais et al., 2007), satellite imagery with global coverage, long-term, and continuous observations, could supplement field measurements and provide fine-resolution and long-term temporal and spatial information for hypoxia modeling. Thus, this study can help further promote the integration of remote sensing techniques in the practical hypoxia monitoring and assessment, and inform stakeholders with low-cost but enriched information on hypoxia dynamics.

Although satellite imagery can help narrow the data gap, there is still an urgent need to initiate a global coast observatory network with a mission to synthesize and share data for better understanding, predicting, and communicating the changing coasts in the Anthropocene. This is because coastal DO observation data are critical for hypoxia modeling and mapping, but, such data are either not publicly available or distributed across multiple data sources. It is extremely challenging to get reliable and consistent observation data, even for one single coastal system, let alone global coasts. A data initiative through the network might help fill the gap. Role models such as the Global Lake Ecological Observatory Network (https://gleon.org/) has successfully linked and gathered lake observatories with research sites on more than 300 lakes across six continents, which greatly enabled large-scale investigation of hypoxia phenomena in the global lakes (Jane et al., 2021). With more field observation data compiled, future research can combine satellite imagery with other geospatial biological datasets for more explicit process-based (e.g., three dimensions, 3D) (Justić and Wang, 2014), long-term and global-scale hypoxia modeling and monitoring. Also, it would be helpful to use the integrated framework of telecoupling to identify knowledge gaps and link distant places that apply excessive fertilizers with coastal hypoxia (Hull and Liu, 2018; Liu and Yang, 2013; Liu et al., 2018). Building such linkages can help predict the timing and amount of the excessive fertilizers that reach global coasts based on the locations of fertilizer applications and leakages as well as the speed of river water flows. High spatiotemporal resolution mapping of coastal hypoxia using satellite imagery can greatly facilitate such efforts toward sustainable coasts.

Code availability

The R scripts and data used for analysis can be found at https://github.com/samuelVJrobinson/hypoxiaMapping

CRediT authorship contribution statement

Yingjie Li: Conceptualization, Methodology, Formal analysis, Visualization, Writing – original draft, Writing – review & editing. **Samuel V.J. Robinson:** Conceptualization, Methodology, Formal analysis, Writing – original draft, Writing – review & editing. **Lan H. Nguyen:** Conceptualization, Methodology, Writing – review & editing. **Jianguo Liu:** Conceptualization, Writing – review & editing, Supervision, Funding acquisition.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The authors do not have permission to share data.

Acknowledgments

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.rse.2022.113346.

References

Behrenfeld, M.J., Westberry, T.K., Boss, E.S., O'Malley, R.T., Siegel, D.A., Wiggert, J.D., Franz, B.A., McClain, C.R., Feldman, G.C., Doney, S.C., Moore, J.K., Dall'Olmo, G., Milligan, A.J., Lima, I., Mahowald, N., 2009. Satellite-detected fluorescence reveals global physiology of ocean phytoplankton. Biogeosciences 6, 779–794. https://doi.org/10.5194/bg-6-779-2009.

Belgiu, M., Drăguţ, L., 2016. Random forest in remote sensing: a review of applications and future directions. ISPRS J. Photogramm. Remote Sens. 114, 24–31. https://doi.org/10.1016/j.isprsjprs.2016.01.011.

Bianchi, T.S., DiMarco, S.F., Cowan, J.H., Hetland, R.D., Chapman, P., Day, J.W., Allison, M.A., 2010. The science of hypoxia in the northern Gulf of Mexico: a review. Sci. Total Environ. 408, 1471–1484. https://doi.org/10.1016/j.scitotenv.2009.11.047.

Biau, G., Scornet, E., 2016. In: A random forest guided tour. TEST 25, pp. 197–227. https://doi.org/10.1007/s11749-016-0481-7.

Breitburg, D., Levin, L.A., Oschlies, A., et al., 2018. Declining oxygen in the global ocean and coastal waters. Science 359, eaam7240. https://doi.org/10.1126/science.aam7240.

Chen, J., Ni, X., Liu, M., et al., 2014. Monitoring the occurrence of seasonal low-oxygen events off the Changjiang estuary through integration of remote sensing, buoy observations, and modeling. J. Geophys. Res.Oceans 119, 5311–5322. https://doi.org/10.1002/2014JC010333.

Chen, S., Hu, C., Barnes, B.B., et al., 2019. Improving ocean color data coverage through machine learning. Remote Sens. Environ. 222, 286–302. https://doi.org/10.1016/j.rse.2018.12.023.

Cutler, D.R., Edwards Jr., T.C., Beard, K.H., et al., 2007. Random forests for classification in ecology. Ecology 88, 2783–2792. https://doi.org/10.1890/07-0539.1.

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

Del Giudice, D., Matli, V.R.R., Obenour, D.R., 2020. Bayesian mechanistic modeling characterizes Gulf of Mexico hypoxia: 1968–2016 and future scenarios. Ecol. Appl. 30, e02032 https://doi.org/10.1002/eap.2032.

Diaz, R.J., Rosenberg, R., 2008. Spreading dead zones and consequences for marine ecosystems. Science 321, 926–929. https://doi.org/10.1126/science.1156401.

Díaz, S., Settele, J., Brondízio, E.S., et al., 2019. Summary for policymakers of the global assessment report on biodiversity and ecosystem services of the intergovernmental science-policy platform on biodiversity and ecosystem services. Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services.

FAO, 2020. The State of World Fisheries and Aquaculture 2020: Sustainability in Action. Food and Agriculture Organization of the United Nations, Rome.

Fennel, K., Laurent, A., Hetland, R., et al., 2016. Effects of model physics on hypoxia simulations for the northern Gulf of Mexico: a model intercomparison. J. Geophys. Res.Oceans 121, 5731–5750. https://doi.org/10.1002/2015JC011577.

Forrest, D.R., Hetland, R.D., DiMarco, S.F., 2011. Multivariable statistical regression models of the areal extent of hypoxia over the Texas-Louisiana continental shelf. Environ. Res. Lett. 6, 045002 https://doi.org/10.1088/1748-9326/6/4/045002.

Gelman, A., Carlin, J.B., Stern, H.S., et al., 2013. Bayesian Data Analysis, 3rd edition. Chapman and Hall/CRC, Boca Raton.

Greene, R.M., Lehrter, J.C., Iii, J.D.H., 2009. Multiple regression models for hindcasting and forecasting midsummer hypoxia in the Gulf of Mexico. Ecol. Appl. 19, 1161–1175. https://doi.org/10.1890/08-0035.1.

Hull, V., Liu, J., 2018. Telecoupling: a new frontier for global sustainability. Ecol. Soc. 23 https://doi.org/10.5751/ES-10494-230441.

IPBES, 2019. Global assessment report on biodiversity and ecosystem services of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services.

Justić, D., Rabalais, N.N., Eugene Turner, R., Wiseman, W.J., 1993. Seasonal coupling between riverborne nutrients, net productivity and hypoxia. Mar. Pollut. Bull. 26, 184–189. https://doi.org/10.1016/0025-326X(93)90620-Y.

Justić, D., Rabalais, N.N., Turner, R.E., 2003. Simulated responses of the Gulf of Mexico hypoxia to variations in climate and anthropogenic nutrient loading. J. Marine Syst. Nutr. Dynam. Coast.Ecosyst. Link.Phys. Biol. Process. 42, 115–126. https://doi.org/10.1016/S0924-7963(03)00070-8.

Josse, Julie, Husson, François, 2016. missMDA: A Package for Handling Missing Values in Multivariate Data Analysis. J. Stat. Softw. 70 (1), 1–31. https://doi.org/10.18637/jss.v070.i01.

Justić, D., Bierman, V.J., Scavia, D., Hetland, R.D., 2007. Forecasting Gulf's hypoxia: the next 50 years?Estuaries and Coasts. J. ERF 30, 791–801. https://doi.org/10.1007/BF02841334.

Justić, D., Wang, L., 2014. Assessing temporal and spatial variability of hypoxia over the inner Louisiana–upper Texas shelf: application of an unstructured-grid three-dimensional coupled hydrodynamic-water quality model. Cont. Shelf Res. 72, 163–179. https://doi.org/10.1016/j.csr.2013.08.006.

Jane, S.F., Hansen, G.J.A., Kraemer, B.M., et al., 2021. Widespread deoxygenation of temperate lakes. Nature 594, 66–70. https://doi.org/10.1038/s41586-021-03550-y.

Kim, Y.H., Son, S., Kim, H.-C., et al., 2020. Application of satellite remote sensing in monitoring dissolved oxygen variabilities: a case study for coastal waters in Korea. Environ. Int. 134, 105301 https://doi.org/10.1016/j.envint.2019.105301.

Klemas, V., 2011. Remote sensing of algal blooms: an overview with case studies. J. Coast. Res. 34–43 https://doi.org/10.2112/JCOASTRES-D-11-00051.1.

Le, C., Lehrter, J.C., Hu, C., Obenour, D.R., 2016. Satellite-based empirical models linking river plume dynamics with hypoxic area and volume. Geophys. Res. Lett. 43, 2693–2699. https://doi.org/10.1002/2015GL067521.

Liu, J., Yang, W., 2013. Integrated assessments of payments for ecosystem services programs. Proc. Nat. Acad. Sci. 110, 16297–16298. https://doi.org/10.1073/pnas.1316036110.

Liu, J., Dou, Y., Batistella, M., Challies, E., Connor, T., Friis, C., Millington, J.D., Parish, E., Romulo, C.L., Silva, R.F.B., Triezenberg, H., Yang, H., Zhao, Z., Zimmerer, K.S., Huettmann, F., Treglia, M.L., Basher, Z., Chung, M.G., Herzberger, A., Lenschow, A., Mechiche-Alami, A., Newig, J., Roche, J., Sun, J., 2018. Spillover systems in a telecoupled anthropocene: typology, methods, and governance for global sustainability. Curr.Opin.Environ.Sustain.Syst. Dynam. Sustain. 33, 58–69. https://doi.org/10.1016/j.cosust.2018.04.009.

Matli, V.R.R., Laurent, A., Fennel, K., et al., 2020. Fusion-based hypoxia estimates: combining geostatistical and mechanistic models of dissolved oxygen variability. Environ. Sci. Technol. https://doi.org/10.1021/acs.est.0c03655.

Matli, V.R.R., Fang, S., Guinness, J., et al., 2018. A space-time geostatistical assessment of hypoxia in the northern Gulf of Mexico. Environ. Sci. Technol. https://doi.org/10.1021/acs.est.8b03474.

McElreath, R., 2020. Statistical Rethinking: A Bayesian Course with Examples in R and STAN, 2nd edition. Chapman and Hall/CRC, Boca Raton.

Mitsch, W.J., Day, J.W., Gilliam, J.W., Groffman, P.M., Hey, D.L., Randall, G.W., Wang, N., 2001. Reducing nitrogen loading to the Gulf of Mexico from the Mississippi River basin: strategies to counter a persistent ecological problem: Ecotechnology—the use of natural ecosystems to solve environmental problems—should be a part of efforts to shrink the zone of hypoxia in the Gulf of Mexico. Bioscience 51, 373–388. https://doi.org/10.1641/0006-3568(2001)051[0373:RNLTTG]2.0.CO;2.

Moreno-Martínez, Á., Izquierdo-Verdiguier, E., Maneta, M.P., et al., 2020. Multispectral high resolution sensor fusion for smoothing and gap-filling in the cloud. Remote Sens. Environ. 247, 111901 https://doi.org/10.1016/j.rse.2020.111901.

Murphy, R.R., Curriero, F.C., Ball, W.P., 2010. Comparison of spatial interpolation methods for water quality evaluation in the Chesapeake Bay. J. Environ. Eng. 136, 160–171. https://doi.org/10.1061/(ASCE)EE.1943-7870.0000121.

NASA Goddard Space Flight Center Ocean Biology Processing Group, Ocean Ecology Laboratory, 2018. Moderate-resolution Imaging Spectroradiometer (MODIS) Aqua Ocean Color Data. NASA OB.DAAC, Greenbelt, MD, USA. https://doi.org/10.5067/AQUA/MODIS/L3M/CHL/2018. (Accessed 27 July 2021).

NOAA, 2017. Gulf of Mexico 'dead zone' is the largest ever measured. https://www.noaa.gov/media-release/gulf-of-mexico-dead-zone-is-the-largest-ever-measured. Accessed 27 Jul 2021.

Obenour, D.R., Michalak, A.M., Scavia, D., 2015. Assessing biophysical controls on Gulf of Mexico hypoxia through probabilistic modeling. Ecol. Appl. 25, 492–505. https://doi.org/10.1890/13-2257.1.

Pitcher, G.C., Aguirre-Velarde, A., Breitburg, D., et al., 2021. System controls of coastal and open ocean oxygen depletion. Prog. Oceanogr. 102613 https://doi.org/10.1016/j.pocean.2021.102613.

R Core Team, 2021. R: A language and environment for statistical computing (R Version 4.0. 3, R Foundation for Statistical Computing, Vienna, Austria, 2020). Vienna, Austria.

Rabalais, N.N., Wiseman, W.J., Turner, R.E., 1994. Comparison of continuous records of near-bottom dissolved oxygen from the hypoxia zone along the Louisiana coast. Estuaries 17, 850. https://doi.org/10.2307/1352753.

Rabalais, N., Turner, R., Sen Gupta, B., Boesch, D., Chapman, P., Murrell, M., 2007. Characterization and longterm trends of hypoxia in the northern Gulf of Mexico: does the science support the action Plan? Estuar. Coasts 30, 753–772.

Rabalais, N.N., Turner, R.E., Díaz, R.J., Justić, D., 2009. Global change and eutrophication of coastal waters. ICES J. Mar. Sci. 66, 1528–1537. https://doi.org/10.1093/icesjms/fsp047.

Rabotyagov, S.S., Kling, C.L., Gassman, P.W., et al., 2014. The economics of dead zones: causes, impacts, policy challenges, and a model of the Gulf of Mexico hypoxic zone. Rev. Environ. Econ. Policy 8, 58–79. https://doi.org/10.1093/reep/ret024.

Ramsay, J., Silverman, B.W., 2005. Functional Data Analysis, 2nd edn. Springer-Verlag, New York.

Scavia, D., Evans, M.A., Obenour, D.R., 2013. A scenario and forecast model for Gulf of Mexico hypoxic area and volume. Environ. Sci. Technol. 47, 10423–10428. https://doi.org/10.1021/es4025035.

Scavia, D., Bertani, I., Obenour, D.R., et al., 2017. Ensemble modeling informs hypoxia management in the northern Gulf of Mexico. PNAS 114, 8823–8828. https://doi.org/10.1073/pnas.1705293114.

Shen, F., Tang, R., Sun, X., Liu, D., 2019. Simple methods for satellite identification of algal blooms and species using 10-year time series data from the East China Sea. Remote Sens. Environ. 235, 111484 https://doi.org/10.1016/j.rse.2019.111484.

Smith, M.D., Asche, F., Bennear, L.S., Oglend, A., 2014. Spatial-dynamics of hypoxia and fisheries: the case of Gulf of Mexico Brown shrimp. Mar. Resour. Econ. 29, 111–131. https://doi.org/10.1086/676826.

Teluguntla, P., Thenkabail, P.S., Oliphant, A., et al., 2018. A 30-m landsat-derived cropland extent product of Australia and China using random forest machine learning algorithm on Google Earth Engine cloud computing platform. ISPRS J. Photogramm. Remote Sens. 144, 325–340. https://doi.org/10.1016/j.isprsjprs.2018.07.017.

Tomasetti, S.J., Gobler, C.J., 2020. Dissolved oxygen and pH criteria leave fisheries at risk. Science 368, 372–373. https://doi.org/10.1126/science.aba4896.

Turner, R.E., Rabalais, N.N., Justić, D., 2012. Predicting summer hypoxia in the northern Gulf of Mexico: redux. Mar. Pollut. Bull. 64, 319–324. https://doi.org/10.1016/j.marpolbul.2011.11.008.

Wang, L., Justić, D., 2009. A modeling study of the physical processes affecting the development of seasonal hypoxia over the inner Louisiana-Texas shelf: circulation and stratification. Cont. Shelf Res. 29, 1464–1476. https://doi.org/10.1016/j.csr.2009.03.014.

Wood, S.N., Bravington, M.V., Hedley, S.L., 2008. Soap film smoothing. J. Roy.Stat.Soc. Ser. B (Stat. Methodol.) 70, 931–955. https://doi.org/10.1111/j.1467-9868.2008.00665.x.

Wood, S.N., 2017. Generalized Additive Models: An Introduction with R, 2nd edn. Chapman and Hall/CRC, Boca Raton.

Yen, J.D.L., Thomson, J.R., Paganin, D.M., et al., 2015. Function regression in ecology and evolution: FREE. Methods Ecol. Evol. 6, 17–26. https://doi.org/10.1111/2041-210X.12290.

Zhang, J., Cowie, G., Naqvi, S.W.A., 2013. Hypoxia in the changing marine environment. Environ. Res. Lett. 8, 015025 https://doi.org/10.1088/1748-9326/8/1/015025.

Zhou, F., Chai, F., Huang, D., Wells, M., Ma, X., Meng, Q., Li, H., 2020. Coupling and decoupling of high biomass phytoplankton production and hypoxia in a highly dynamic coastal system: the changjiang (Yangtze River) estuary. Front. Mar. Sci. 7, 259. https://doi.org/10.3389/fmars.2020.00259.

Zhou, Y., Obenour, D.R., Scavia, D., et al., 2013. Spatial and temporal trends in Lake Erie hypoxia, 1987–2007. Environ. Sci. Technol. 47, 899–905. https://doi.org/10.1021/es303401b.