

This work was written as part of one of the author's official duties as an Employee of the United States Government and is therefore a work of the United States Government. In accordance with 17 U.S.C. 105, no copyright protection is available for such works under U.S. Law.

Public Domain Mark 1.0

<https://creativecommons.org/publicdomain/mark/1.0/>

Access to this work was provided by the University of Maryland, Baltimore County (UMBC) ScholarWorks@UMBC digital repository on the Maryland Shared Open Access (MD-SOAR) platform.

Please provide feedback

Please support the ScholarWorks@UMBC repository by emailing scholarworks-group@umbc.edu and telling us what having access to this work means to you and why it's important to you. Thank you.



Hypoxia forecasting for Chesapeake Bay using artificial intelligence

Guangming Zheng^{a,c}, Stephanie Schollaert Uz^b, Pierre St-Laurent^d, Marjorie A. M. Friedrichs^d,
Amita Mehta^{b,e}, and Paul M. DiGiacomo^a

^a *NOAA/NESDIS/Center for Satellite Applications and Research, College Park, MD 20740*

^b *Earth Science Division, NASA Goddard Space Flight Center, Greenbelt, MD 20771*

^c *Cooperative Institute for Satellite Earth System Studies, Earth System Science Interdisciplinary
Center, University of Maryland College Park, College Park, MD 20740*

^d *Virginia Institute of Marine Science, William & Mary, Gloucester Point, VA 23062*

^e *NASA Goddard Earth Sciences Technology and Research, University of Maryland Baltimore
County, Baltimore, MD 21250*

Corresponding author: Guangming Zheng, guangming.zheng@noaa.gov

Early Online Release: This preliminary version has been accepted for publication in *Artificial Intelligence for the Earth Systems*, may be fully cited, and has been assigned DOI 10.1175/AIES-D-23-0054.1. The final typeset copyedited article will replace the EOR at the above DOI when it is published.

ABSTRACT: Seasonal hypoxia is a recurring threat to ecosystems and fisheries in the Chesapeake Bay. Hypoxia forecasting based on coupled hydrodynamic and biogeochemical models has proven useful for many stakeholders, as these models excel in accounting for the effects of physical forcing on oxygen supply, but may fall short in replicating the more complex biogeochemical processes that govern oxygen consumption. Satellite-derived reflectances could be used to indicate the presence of surface organic matter over the Bay. However, teasing apart the contribution of atmospheric and aquatic constituents from the signal received by the satellite is not straightforward. As a result, it is difficult to derive surface concentrations of organic matter from satellite data in a robust fashion. A potential solution to this complexity is to use deep learning to build end-to-end applications that do not require precise accounting of the satellite signal from atmosphere or water, phytoplankton blooms, or sediment plumes. By training a deep neural network with data from a vast suite of variables that could potentially affect oxygen in the water column, improvement of short-term (daily) hypoxia forecast may be possible. Here we predict oxygen concentrations using inputs that account for both physical and biogeochemical factors. The physical inputs include wind velocity reanalysis information, together with 3D outputs from an estuarine hydrodynamic model, including current velocity, water temperature, and salinity. Satellite-derived spectral reflectance data are used as a surrogate for the biogeochemical factors. These input fields are time series of weekly statistics calculated from daily information, starting 8 weeks before each oxygen observation was collected. To accommodate this input data structure, we adopted a model architecture of long short-term memory networks with 8 time steps. At each time step, a set of convolutional neural networks are used to extract information from the inputs. Ablation and cross validation tests suggest that among all input features, the strongest predictor is the 3D temperature field, with which the new model can outperform the state-of-the-art by ~20% in terms of median absolute error. Our approach represents a novel application of deep learning to address a complex water management challenge.

SIGNIFICANCE STATEMENT: This study presents a novel approach that combines deep learning and hydrodynamic model outputs to improve the accuracy of hypoxia forecasts in the Chesapeake Bay. By training a deep neural network with both physical and biogeochemical information as input features, the model accurately predicts oxygen concentration at any depth in the water column one day in advance. This approach has the potential to benefit stakeholders and inform adaptation measures during the recurring threat of hypoxia in the Chesapeake Bay. The success of this study suggests the potential for similar applications of deep learning to address complex water management challenges. Further research could investigate the application of this approach to different forecast lead time and other regions and ecosystem types.

1. Introduction

Hypoxia, typically defined as oxygen levels $<2 \text{ mg L}^{-1}$, has been identified as a recurring threat to the Chesapeake Bay ecosystem. Nutrient reductions have helped to mitigate the problem (Frankel et al. 2022), but are not sufficient to fully address it owing to the setback associated with climate change (Du et al. 2018; Irby et al. 2018). In fact, statistics from recent years have shown that hypoxia continues to have a significant impact on the bay's fisheries and aquaculture (see reports by VIMS (2022) and MDDNR (2022)). This problem has garnered significant attention from the scientific community, as well as policymakers and stakeholders who are concerned about the long-term health and sustainability of the Chesapeake Bay ecosystem. In preparation for upcoming space-borne spectroscopy missions, The National Aeronautics and Space Administration (NASA) recently began engaging potential users of its future satellite data prior to mission launch for the ocean color and aquatic community (Schollaert Uz et al. 2019; Culver et al. 2020; Scott and Urquhart 2020; Culver et al. 2022; Lee et al. 2022). Chesapeake Bay resource managers who monitor waters are keenly interested in improving the ability of remote sensing and other new technology to help detect harmful algal blooms and other threats to the growing aquaculture industry (Wolny et al. 2020).

Hypoxia forecasts help resource managers and policymakers to make informed decisions about managing coastal ecosystems and fisheries, and allow stakeholders to take proactive measures to minimize the impact on fisheries and aquaculture operations. Current approaches to forecasting hypoxia in the Chesapeake Bay involve the use of numerical models such as the Chesapeake Bay

Environmental Forecasting System (CBEFS) (St-Laurent et al. 2020; Bever et al. 2021; St-Laurent and Friedrichs 2024). This 3D mechanistic model simulates the physical and biogeochemical processes that lead to hypoxia and is forced with a wide range of real-time inputs including both terrestrial information (river discharge, nutrient loadings) and atmospheric data (air temperature, winds, humidity, precipitation, etc.). Model outputs include water temperature, salinity, currents, as well as biogeochemical variables including nutrient concentrations, phytoplankton and zooplankton biomass, dissolved organic matter, inorganic sediment, alkalinity, and dissolved oxygen concentrations.

In many estuarine ecosystems like the Chesapeake Bay, phytoplankton are the primary producer of organic matter and thus one of the key drivers of hypoxia (Su et al. 2020). Therefore, accurate and low-latency measurement and prediction of algal blooms beyond discrete field sampling is crucial for effective hypoxia forecasts. Satellite remote sensing has the potential to provide valuable data on algal biomass over large spatial scales and in a timely fashion (e.g., Aurin et al. 2013; Mouw et al. 2015). Although dissolved oxygen (DO) is not directly detectable from satellites, it is related to physical and biological processes that can be remotely sensed (Zheng and DiGiacomo 2020). However, assimilating satellite remote sensing data into current modeling approaches like CBEFS can be challenging due to issues such as data latency, quality, incomplete swaths or missing data due to cloud cover, and data processing requirements.

Because of the challenges associated with the assimilation of remote-sensing data into mechanistic models, there is growing interest in the potential of deep-learning driven artificial intelligence (AI) to improve our understanding and management of these complex systems. Deep-learning driven AI has already shown promise in a range of environmental applications, including prediction of weather patterns (Chantry et al. 2021), air pollution forecasting Masood and Ahmad 2021, tracking wildlife populations (Isabelle and Westerlund 2022), and managing water resources (Ghobadi and Kang 2023). In the context of hypoxia management, AI has the potential to provide new insights into the complex relationships between physical, biological, and environmental variables that drive hypoxia events (Yu et al. 2020; Valera et al. 2020). By analyzing large datasets with multiple variables, AI models can identify patterns and correlations that may not be readily apparent using traditional modeling approaches. In addition, the computational cost of AI models,

once training is completed, is low compared to the burden of running a 3D hydrodynamic model, making AI models very promising for operational applications.

The potential benefits of AI-driven approaches to environmental monitoring and management, create a strong motivation for developing AI algorithms for hypoxia forecasts (Schollaert Uz et al. 2020). By leveraging the power of AI to better understand and manage hypoxia and other environmental challenges, we can move towards more effective and sustainable management strategies that promote the long-term health and resilience of our coastal ecosystems. The ultimate goal of this research is to help provide resource managers and policymakers with the forecasts they need to make informed decisions about managing coastal ecosystems and fisheries, while also contributing to our broader understanding of the complex relationships between physical, biological, and environmental variables that drive hypoxia events.

2. Data

In order to predict daily mean DO concentrations within the water column of the Chesapeake Bay, it is necessary to obtain data that characterizes the supply, consumption, and advection of DO. The primary control on oxygen supply across the air-water interface is water column stratification and wind-driven vertical mixing, which are determined by factors such as temperature, salinity, wind speed, and direction. Oxygen consumption is primarily influenced by the availability of organic matter in the water column and sediments, as well as the rate of organic decomposition, which is largely controlled by temperature and the corresponding rate of microbial metabolism. Additionally, pH and redox potential also play a role in affecting oxygen consumption. It should be noted that temperature also directly affects the solubility of oxygen in water.

To address these considerations, we utilized temperature and salinity data to characterize water column stratification, wind data to characterize vertical mixing, satellite optical data to characterize the distribution of organic matter at the surface, and 3D current velocity data to characterize advection. Temperature also serves as an important indicator of organic matter decay rate. This study does not include any data to characterize distributions of organic matter in subsurface layers and sediments, as an effective approach to their characterization was not available at the time of the study. Satellite ocean color data were used as indicators of surface organic carbon content such

as dissolved organic matter, suspended particulate matter, and phytoplankton which is the major source of autochthonous organic carbon.

With respect to the prediction target, daily mean DO concentrations, it is characterized with field-measured DO samples. While these DO data were derived from instantaneous measurements, our primary objective is to accurately predict daily mean DO concentrations. The approach is intended to ensure a statistically robust agreement between model-predicted and measured DO on an average daily basis across an extensive period of two decades.

a. Data Sources

Temperature, salinity, and current velocity information are obtained from a 3D hydrodynamic model implemented for the Chesapeake Bay, i.e., the Chesapeake Bay Environmental Forecast System (CBEFS). CBEFS is an implementation of the Regional Ocean Modeling System (Shchepetkin and McWilliams 2005) originally designed for process-based studies of the Chesapeake Bay. It is a fully mechanistic model based on well-known biogeochemical processes and strictly enforces conservation of nitrogen, carbon, and oxygen (see St-Laurent et al. 2020; Bever et al. 2021; Frankel et al. 2022; St-Laurent and Friedrichs 2024, for detailed descriptions of the model). CBEFS is operated in a freely running mode, meaning that its skill at reproducing observations is entirely determined by its prognostic equations and parameters. There are no mechanisms in CBEFS to artificially bring the model closer to reality (e.g., data assimilation) even when observations are available or when systematic biases are present. As opposed to using the publicly available nowcast/forecast (www.vims.edu/cbefs), we used a CBEFS hindcast output, similar to that described by Frankel et al. (2022) except with a finer horizontal resolution of 600 m instead of 1.8 km. This hindcast uses the same configuration, including bathymetry and biogeochemical parameters, as the forecast; however, the forecast uses a different atmospheric forcing product (ECMWF ERA5 for hindcast, North American Mesoscale Forecast System for nowcast/forecast). Different terrestrial inputs of freshwater and biogeochemical constituents were also used. For the period 1985-2014, the hindcast configuration use the EPA's Phase6 Watershed Model (Hood et al. 2021), whereas CBEFS uses scaled real-time USGS freshwater discharge and a seasonal climatology of biogeochemical concentrations. After 2014, both the hindcast and nowcast/forecast use the same riverine forcing because Phase6 was not available beyond 2014 at the time of this work. The same wind data used

to force the CBEFS were used as one of the input features of our machine learning model, which is the 3-hourly wind data generated by the ECMWF ERA5.

There are several moderate resolution satellite sensors available and in this study we chose MODIS Aqua. MODIS Aqua has the longest time span which ensures that we obtain the largest training dataset possible. The commonly used "Level-2" ocean color data which are the product of atmospheric correction often have large gaps owing to unfavorable conditions such as turbid water, glint, clouds, and aerosols. To maximize the availability of ocean color data and leave extraction of predictive features to machine learning algorithms, we used top-of-atmosphere reflectance data and applied Rayleigh correction. MODIS Level-1A data were obtained from NASA's Ocean Biology Processing Group (OBPG). Using the OBPG software SeaDAS-OCSSW we obtained Rayleigh-correct spectral reflectance, ρ_s , at 12 visible, near-infrared, and shortwave infrared bands including 412, 443, 469, 488, 531, 551, 555, 645, 667, 678, 748, and 869 nm. Among these bands, spatial resolution for 645 nm is 250 m; for 469 and 555 nm it is 500 m, The rest of the spectral bands have spatial resolution of 1 km. However, the obtained Rayleigh-correct spectral reflectance we generate at all bands are at 250 m, where all other band reflectance are downsampled at 645 nm spatial resolution. we used "Level-1b" top-of-atmosphere reflectance data, applied Rayleigh correction, and obtained Rayleigh-correct spectral reflectance, ρ_s , at 12 visible, near-infrared, and shortwave infrared bands including 412, 443, 469, 488, 531, 551, 555, 645, 667, 678, 748, and 869 nm. MODIS data were obtained from NASA's Ocean Biology Processing Group (OBPG).

DO field data were obtained from the Chesapeake Bay Program data hub (<https://datahub.chesapeakebay.net/WaterQuality>). Sampling locations are illustrated in Figure 1. Monitoring in the mainstem is routinely conducted monthly, except during warmer months when it is done bi-weekly. There are tidal tributary stations which are generally sampled once per month. The mainstem and tidal tributary samples account for ~75% and 7%, respectively, of DO data used in this study. We also used a significant amount of data sampled semi-continuously at various depths at some shallow water sites along the shoreline of the mainstem Bay and tributaries (~6%), as well as at the surface (0.5 m) along some segments monitored monthly using a flow-through sampling system (~11%). All DOs were measured in situ typically using either YSI or Hydrolab sondes.

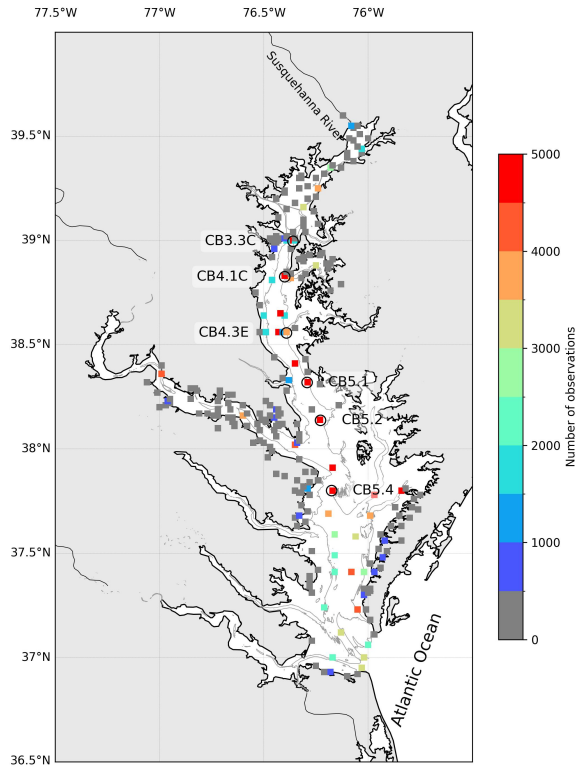


FIG. 1. Locations of CBP field sampling of dissolved oxygen concentration. Symbol color represents total number of observations in our datasets from 2002 to 2020. Open black circles mark the location of selected stations labeled with station names.

b. Data Preprocessing

The original data from multiple sources such as CBEFS outputs, MODIS, and ECMWF outputs were not presented in a spatial-temporally consistent format, and need to be preprocessed before being accommodated into the training datasets. The CBEFS outputs we used (Table 1), i.e., currents, temperature, and salinity, are daily averages. We binned the CBEFS outputs on a pixel-by-pixel basis into weekly means to reduce the volume of input data for our model. Satellite reflectance data were also binned in the same fashion, except that the 1st percentile as opposed to weekly mean values were used to reduce the influence of bright pixels that are likely associated with atmospheric sources. With respect to ECMWF-derived wind data, we calculated the 2D

probability distribution, P_{wind} , with respect to 10 bins of windspeed (0-10 m/s) and 12 bins of wind direction within weekly time windows for data corresponding to the grid point closest to a DO field sampling location. This is done to characterize the mixing effect of the winds, since calculating simple means did not result in a good representation of the wind forcing data.

In addition to temporal binning, the CBEFS outputs and MODIS reflectance data were also spatially reprojected to a evenly spaced rectangular grid, which is bounded by 36.6 – 39.6°N and 75.6 – 77.0°W with a resolution of 0.01 °. The MODIS data were originally in irregularly oriented swaths and were reprojected to this grid. The original CBEFS outputs are defined on a Cartesian horizontal grid and on 20 topography-following vertical levels. They were reprojected to our rectangular grid and also vertically every 1 m from 0 to 15 m. This is sufficient for most part of the Chesapeake Bay except in the regularly dredged main shipping channel that extends from the Atlantic Ocean to the Baltimore harbor. Finally, the CBEFS outputs were vertically averaged every 5 m to have three vertical layers in the final reprocessed data. Note that although the input features obtained from CBEFS outputs cover only three layers up to 15 m deep, our models are trained to predict DO at any depths in the Chesapeake Bay.

c. Construction of Training Datasets

Using the data described above we compiled a dataset comprising 162,274 pairs of training examples covering the time period from 2002 to 2020. Each example contains multiple scalar, vector, or tensor arrays as feature variables and one scalar target variable (Table 1), where the features are used to predict the target. For each individual example, the target is DO measured at a given date, latitude, longitude, and depth. The features include multi-dimensional CBEFS-derived current velocity, temperature, and salinity, MODIS-derived ρ_s , and P_{wind} data from 8 weeks to 1 day before the date of DO sampling, as well as three contextual scalar variables including day of year (DoY), DO sampling depth, and bottom depth. Thus, the model developed using this dataset can be used to predict DO with 1 day lead time. Each of the CBEFS- and MODIS-derived variables is a 8-week time series of weekly statistical arrays (typically mean, except for ρ_s which is 1st percentile to avoid cloud contamination as much as possible) sampled on a grid-by-grid basis in a rectangular window surrounding the target location. The size of the rectangular sampling window was made proportional to the time difference in the number of weeks, d , between the

sampling dates of DO and the feature variable, assuming that as we go back further in time, what happens at locations farther away could potentially impact the target. The initial sampling window is selected to be 10 km (along the same longitude) by 6 km (along the same latitude) for the first week, and enlarges d times for other weeks so that the sampling window is $10d$ km by $6d$ km. Longer size along the same longitude were selected because the general pattern of currents in the Chesapeake Bay is dominated by longitudinal flows, thus higher mobility in terms of water parcels. Although the sampling region varies with time difference d , the size of the horizontal array were kept identical by subsampling the data d -folds; data within every $d \times d$ grid points were averaged. The wind probability distribution P_{wind} was sampled at the same times but at a fixed single point for each given example. Therefore, the P_{wind} feature is a 8-week time series of probability distribution across various wind speeds and directions. Other feature variables are simply scalars associated with the field sampling day of year, sampling depth, and bottom depth. Any missing values in the input features are replaced with zeros, which is a common practice in machine learning.

TABLE 1. Data structure of each individual training example. Symbols: t , time, in number of weeks before the DO observation, size = 8; lat , latitude, size = 10; lon , longitude, size = 6; λ , light wavelength, size = 12; z , water depth, size = 3; w_{spd} , wind speed, size = 11; w_{dir} , wind direction, size = 8; u, v, w , current velocity data in longitudinal, latitudinal, and vertical directions, respectively; T , water temperature; S , salinity; P_{wind} , 2D probability distribution of wind with respect to its speed and direction; z_{sam} , DO sampling depth; z_{bot} , bottom depth; ECMWF, European Centre for Medium-Range Weather Forecasts. ERA5 is the fifth generation ECMWF atmospheric reanalysis of the global climate covering the period from January 1950 to present. MODIS-Aqua, Moderate Resolution Imaging Spectroradiometer onboard Aqua satellite. CBEFS, Chesapeake Bay Environmental Forecasting System. CBP, Chesapeake Bay Program.

Variables	Number of Dimensions	Coordinates	Source
ρ_s	4	t, lat, lon, λ	MODIS-Aqua
u, v, w, T, S	4	t, lat, lon, z	CBEFS Output
P_{wind}	3	t, w_{spd}, w_{dir}	ECMWF-ERA5 Output
Day of Year (DoY)	0	N/A	N/A
z_{sam}	0	N/A	CBP
z_{bot}	0	N/A	CBP
DO (target)	0	N/A	CBP

The entire dataset was split into training, validation, and test sets chronologically as opposed to randomly to avoid overfitting (Table 2). The test set used data from the years 2019 and 2020; these data were not included in the training. The year 2018 was used as validation set except during the cross validation studies when various other previous years were used alternatively. The number of examples varies from year to year but are generally consistent in size, accounting for 5–7% of the entire dataset except for 2002 and 2020 (Table 2). The much smaller size were attributable to the late start of the availability of MODIS Aqua data in 2002, and the impact of COVID-19 pandemic on field sampling activities in 2020.

TABLE 2. Data distribution by year and the split of training, validation, and test datasets. There are 162,274 examples in total.

Splits	Year	Size	Percentage
Training and Validation	2002	1578	1.0
	2003	8657	5.3
	2004	10016	6.2
	2005	8415	5.2
	2006	10393	6.4
	2007	10573	6.5
	2008	10047	6.2
	2009	9110	5.6
	2010	8125	5.0
	2011	8834	5.4
	2012	8844	5.5
	2013	9303	5.7
	2014	8949	5.5
	2015	9307	5.7
	2016	9300	5.7
	2017	9185	5.7
	2018	10020	6.2
Test	2019	8839	5.4
	2020	2779	1.7

3. Model Architecture

The machine learning model utilizes a Long-Short Term Memory (LSTM) framework, which is based on physical intuition to incorporate a diverse set of input data (Fig. 2). LSTM is chosen to learn the time series data because it is particularly well-suited for capturing temporal dependencies

and patterns in sequences, thanks to its ability to maintain long-term memory and selectively forget irrelevant information, which is crucial in predicting future values of DO. The prediction target, DO, was conceptualized as a result of a series of historical environmental conditions. At each time step, in this case, weekly, the environmental data corresponding to that specific time step were integrated into an LSTM cell.

The environmental data were utilized for various purposes within the model. Satellite reflectance was utilized as a marker for surface organic matter. Temperature and salinity (density) were employed to determine the degree of vertical stratification and rate of organic matter decomposition, respectively. Currents were used to identify the movement of water parcels. Wind was evaluated as a factor impacting oxygen supply through wind-induced mixing at the air-water interface. Consequently, the reflectance, temperature and salinity, currents, and wind data were processed independently before being combined into the machine learning model. To obtain organic matter information from reflectance data on a grid-specific basis, the model was trained to learn the spectral features in the reflectance data through the implementation of 3D convolutions using kernels of size $1 \times 1 \times n$, where n represents the number of channels in the input reflectance data. Analogous operations were applied to the temperature and salinity data to obtain localized information on stratification and organic matter decomposition, as well as to the combined data of reflectance, density, and currents. Subsequently, dense layers were applied to the combined organic flux and decomposition information to estimate oxygen demand. oxygen supply is estimated by processing the wind data using a 2D CNN to obtain a vector characterizing the overall degree of vertical mixing.

After assimilating the environmental data at all time steps, the output of the LSTM are combined with other contextual information including DoY and sampling and bottom depths corresponding to the target. The prediction of DO is made using several more densely connected neural network layers.

To optimize the hyperparameters (external configurations that is set arbitrarily and whose value cannot be estimated from data) of the model that ingests all input features discussed above, we utilized the Ray Tune library (<https://docs.ray.io>) to search in a hyperparameter space. The hyperparameter space is defined by the following dimensions: batch size, dropout rate, number of convolution filters of various Conv2D and Conv3D layers, kernel size, pooling type, learning rate,

number of dense layers and number of neurons in them, and etc. Ray Tune uses various search algorithms such as Grid Search, Random Search, and Bayesian Optimization to efficiently navigate through this hyperparameter space to find the optimal set of parameters for our model. The best set of hyperparameters is selected based on the validation error.

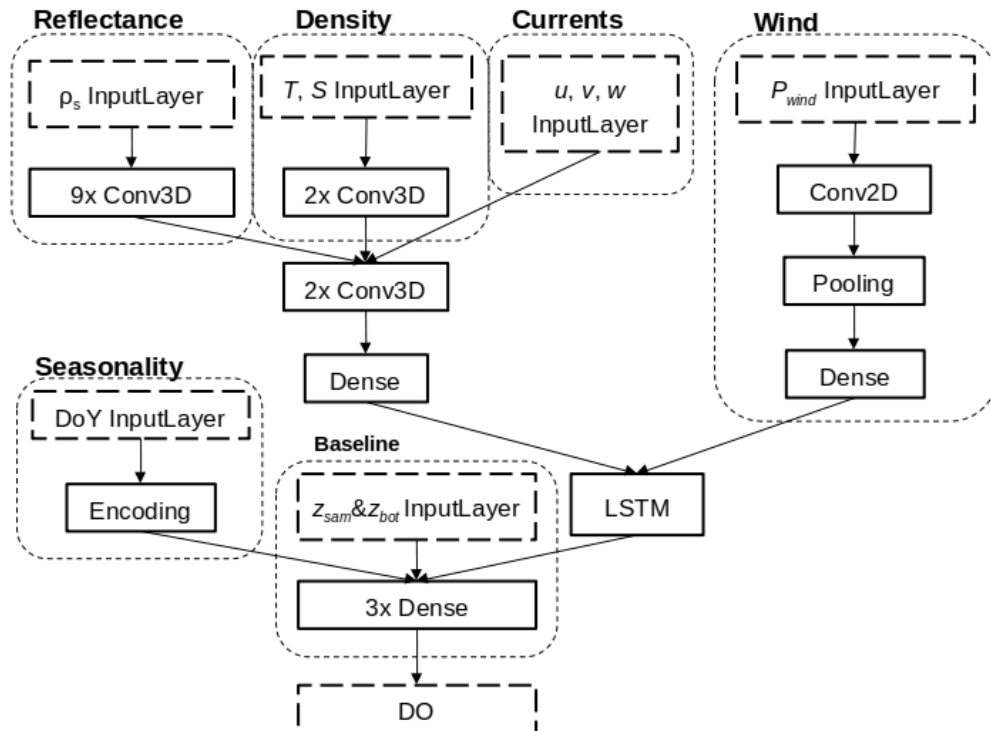


FIG. 2. Model architecture of using multiple data streams to predict the concentration of dissolved oxygen in the Chesapeake Bay. InputLayer, a layer to receive input data. Conv2D and Conv3D, 2D and 3D convolutional neural networks, respectively. LSTM, Long Short-Term Memory. Dashed rectangles represent different components that learn from different data streams. See Table 1 for details of various input variables.

4. Ablation Study

We conducted an ablation study on the model to understand the contribution of each component to the overall performance of the model, and to identify any redundant or unnecessary features that may be removed to improve the efficiency and accuracy of the model without sacrificing its performance. Various individual components and their combinations were tested. The results are shown in Table 3 in incremental order of validation mean absolute error (MAE). Each configuration is named after the physical meaning of the additional input features used. For convenience, an abbreviated name is also given to each configuration. The simplest Baseline model (B) has only two input features, i.e., sampling and bottom depths. All other configurations of model setups involve additional input features.

From the standpoint of single-feature prediction power, temperature is the single most important predictor, outperforming reflectance, currents, wind, salinity, and day of year by a significant margin of at least 0.3 mg L^{-1} in terms of validation MAE. However, if one wishes to push the limit of model performance, the validation MAE can be further lowered by up to 0.15 mg L^{-1} , at much higher computing cost because of the inclusion of more input features.

5. Cross Validation

The above results were obtained using 2018 as the validation year, which is known to be a wet year (see <https://www.usgs.gov/centers/chesapeake-bay-activities/science/freshwater-flow-chesapeake-bay>) and its use as the validation set could potentially skew the model training result. To ensure that the model is unbiased by the selection of validation year, we conducted a cross validation by using different years from 2002 to 2017 as validation year, while holding 2019 and 2020 as the years included in the test set. Cross validation experiments were performed for the four best performing configurations, i.e., DCW, DCWS, RDCW, RDCWS (see Table 3 for details regarding these configurations).

The cross validation results (Fig. 3; Table 4) show that the MAE on the test set is similar, $\sim 0.9 \text{ mg L}^{-1}$, even when using different years for validation. This result was consistent across the evaluation of the four different model configurations considered (DCW, RDCW, DCWS, and RDCWS). No correlation between wet/dry years versus test-set MAE is observed, although the validation MAE appears to be somewhat associated with the annual mean streamflow into the Chesapeake Bay, with

TABLE 3. Different configurations of the model with various input features. The configurations were sorted based on the validation error (MAE) using 2018 as the validation year. The "Baseline" model uses only z_{sam} and z_{bot} as input. Density represent both temperature and salinity.

Model Configuration		Additional Input Variables	Validation MAE
Abbre.	Full Name		
B	Baseline		2.457
S	Seasonality	DoY	2.264
D _S	Salinity	S	2.153
W	Wind	P_{wind}	1.403
C	Currents	u, v, w	1.208
R	Reflectance	ρ_s	1.196
WS	Wind-Seasonality	P_{wind}, DoY	1.174
RD _T	Reflectance-Temperature	ρ_s, T	0.903
D _T	Temperature	T	0.902
D _T WS	Temperature-Wind-Seasonality	T, P_{wind}, DoY	0.896
RDC	Reflectance-Density-Currents	ρ_s, T, S, u, v, w	0.889
D	Density	T, S	0.888
DC	Density-Currents	T, S, u, v, w	0.885
DCW	Density-Currents-Wind	T, S, u, v, w, P_{wind}	0.869
RDCW	Reflectance-Density-Currents-Wind	$\rho_s, T, S, u, v, w, P_{wind}$	0.861
RDCWS	Reflectance-Density-Currents-Wind-Seasonality	$\rho_s, T, S, u, v, w, P_{wind}, DoY$	0.857
DCWS	Density-Currents-Wind-Seasonality	$T, S, u, v, w, P_{wind}, DoY$	0.851

high errors for the wet years 2003, 2004, and 2011. However, even this correlation is inconsistent because the validation MAE is also high for two "normal" years of 2005 and 2006, but low for the wet year of 2018. After careful analysis and comparison of the performance of these models, it was determined that the choice of which year was used as the validation set has a small influence on the performance of the trained model on the test set. The consistent results across the various validation years demonstrate the robustness and reliability of the methodology.

The four model configurations also have similar validation and test MAE. This indicates that all four model configurations have similar validation and test MAE (Fig. 3; Table 4), indicating their ability to accurately predict DO with only minor variations in performance. To make the most efficient and cost-effective model, the DCW configuration was selected as the final model due to its simplicity with the least number of input features, which results in lower computing cost and complexity compared to the other configurations. We hereafter refer to this model configuration as the Hypoxia forecast based on AI, HypoxAI.

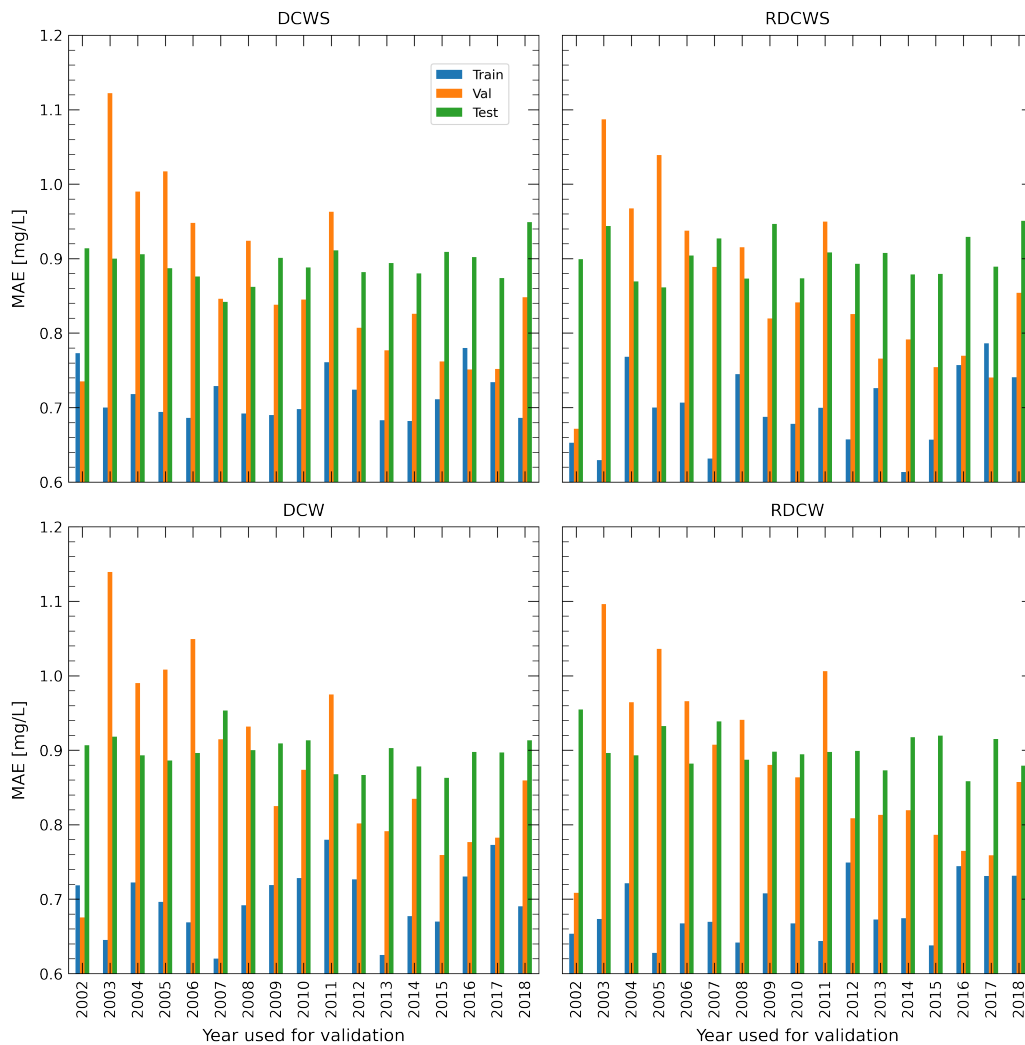


FIG. 3. Mean absolute error calculated for DO-derived from four versions of the model: DCW, DCWS, RDCW, RDCWS, which were cross validated using different years as the source of validation data. Results are calculated from the test set data, i.e., years 2019 and 2020.

6. Evaluation of HypoxAI in Comparison with CBEFS

The overall agreement between HypoxAI-predicted and observed DO in the test set in comparison with CBEFS-predicted DO are shown in Fig. 4 and Table 5. In this test set, DO observations were made at various sampling depths and covered a broad range of bathymetry including both deep

TABLE 4. Cross validation with different validation year and four different versions of model with and without reflectance and seasonality, in addition to the input of density, currents, and wind data. All MAEs represent the average values of all setups of using different years for validation.

Model versions	Training MAE [mg L ⁻¹]	Validation MAE [mg L ⁻¹]	Test MAE [mg L ⁻¹]
DCW	0.699	0.881	0.898
RDCW	0.683	0.881	0.902
DCWS	0.714	0.868	0.893
RDCWS	0.696	0.860	0.902

and shallow stations. HypoxAI surpasses CBEFS in terms of the number of locations where it can make predictions, with a capability to make ~10% more predictions, mostly in shallow areas of the tributaries which are out of the scope of predictions for the CBEFS we used. For the subset of samples where both HypoxAI and CBEFS provided predictions, HypoxAI demonstrated better performance, e.g., the test set mean absolute error (MAE) was lower by 0.18 mg L⁻¹, or 16%. For the other subset of samples where CBEFS did not make any prediction, HypoxAI exhibits consistently good accuracy compared with its performance on the previous subset, actually with somewhat better skill.

TABLE 5. Testset Error Statistics for Model Version DCW using 2018 as Validation Year.

Model	Samples	RMSE	MAE	R ²	Slope	Bias	N
	All	1.263	0.913	0.847	0.846	-0.003	11618
HypoxAI	No CBEFS predictions	1.197	0.857	0.876	0.898	0.126	1156
	CBEFS made valid predictions	1.270	0.920	0.838	0.837	-0.017	10462
CBEFS	CBEFS made valid predictions	1.538	1.098	0.763	0.892	0.155	10462

While we have evaluated the overall performance of the model using the cross-validation experiments described above, it is also important to examine whether the model makes realistic predictions with respect to vertical DO profiles. The evaluation of the model's performance in predicting vertical DO profiles will provide further insight into the accuracy and robustness of the model. Therefore, we compared observed DO profiles with those predicted by the CBEFS and HypoxAI for six selected stations along the Chesapeake Bay mainstem, which are relatively deep and known to experience the worst hypoxia issues. We analyzed both a cooler month (April, Fig. 5) and a warmer month (July, Fig. 6) and found that HypoxAI demonstrated reasonable vertical

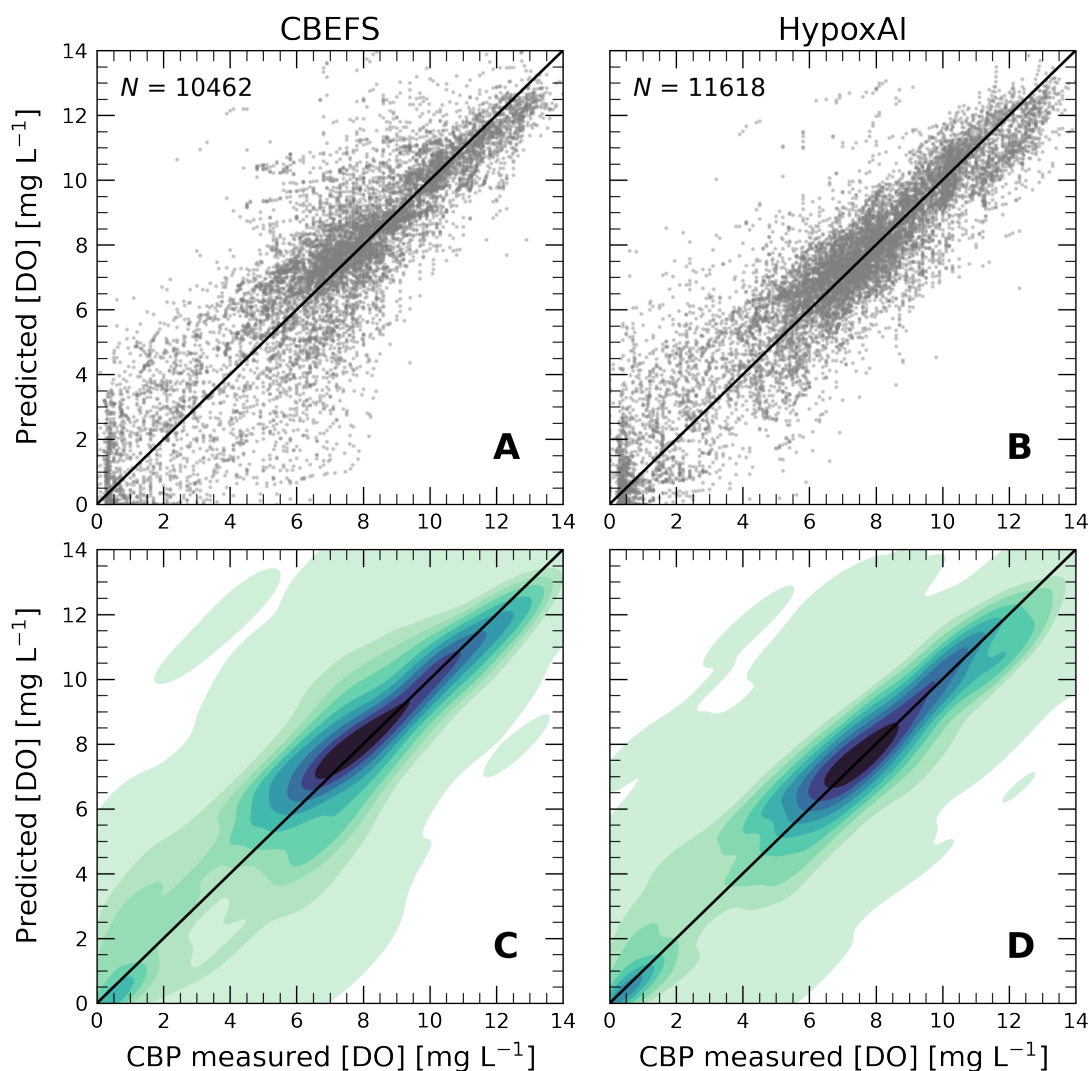


FIG. 4. Comparison between model-derived and measured DO for HypoxAI (A,C) and CBEFS (B,D) using an independent test dataset (2019-2020).

DO profiles for both. For the cooler month examples, the HypoxAI-predicted profiles exhibited variable vertical shape and were in better agreement with field measurements than those predicted by the CBEFS model. However, we also observed that the measurements showed abrupt changes in DO, which are not well-reproduced by the HypoxAI-predicted profiles, as they tend to be smooth and lack fine structures. Several CBEFS profiles in Fig. 5 show vertically well mixed DO profile, suggesting that the model has not yet simulated the onset of hypoxia on this particular date. There-

fore the disagreement between CBEFS and field measurements might be mainly a timing issue. For the warmer month example, we found that CBEFS and HypoxAI were able to reproduce some stratification in the vertical DO profiles, but still struggled to match the abrupt changes in DO seen in the field measurements, similar to some of the examples in the cooler month.

In addition to the visual comparison of the predicted and measured DO profiles, we also conducted a quantitative assessment of the model performance by calculating statistical differences between the model-derived and field-measured DO profiles for all stations deeper than 10 m. Our results indicate that the accuracies of both CBEFS and HypoxAI-derived DO values vary with month and depth (Fig. 7). They both agree reasonably well with field measurement except in April. This suggests that prediction of the exact timing of the onset of hypoxia is difficult. Overall the HypoxAI exhibits the best agreement with measurements in June and August. With respect to depth, we observed an increased error in the predicted profiles at depths between 5-10 m. This represents a challenge for both models to accurately predict sharp the DO changes across the oxycline which is a well-known problem (Irby et al. 2016).

These results suggest that, overall, the HypoxAI model demonstrated reasonable performance in predicting vertical DO profiles and may be a valuable tool for predicting hypoxia conditions in the Chesapeake Bay. However, there are still some limitations and areas for improvement, particularly in reproducing fine-scaled structures in the water column. Like any models, HypoxAI is limited by the accuracy of its inputs. If neither CBEFS hydrodynamics and mixing nor other input information captures the right temperature and salinity profiles, HypoxAI's capability to resolve vertical structures will be compromised. Therefore, the development of more precise tools to characterize the hydrodynamics continues to be crucial in future research.

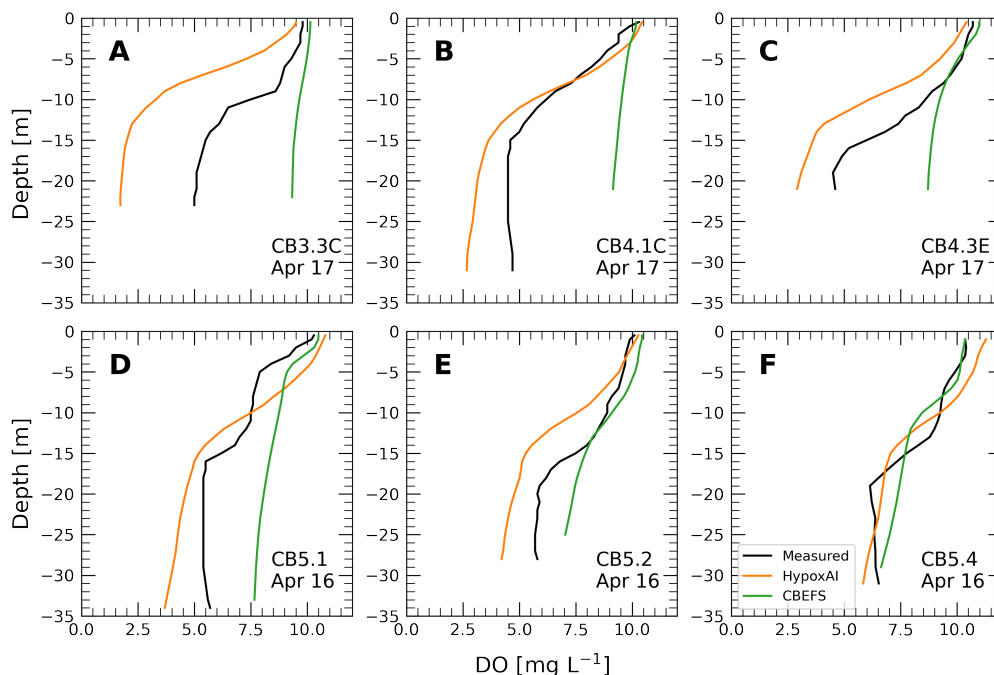


FIG. 5. Comparison of DO Profiles in April 2019 at six stations along the Chesapeake Bay mainstem representative of the growth stage of hypoxia. See Fig. 1 for station locations.

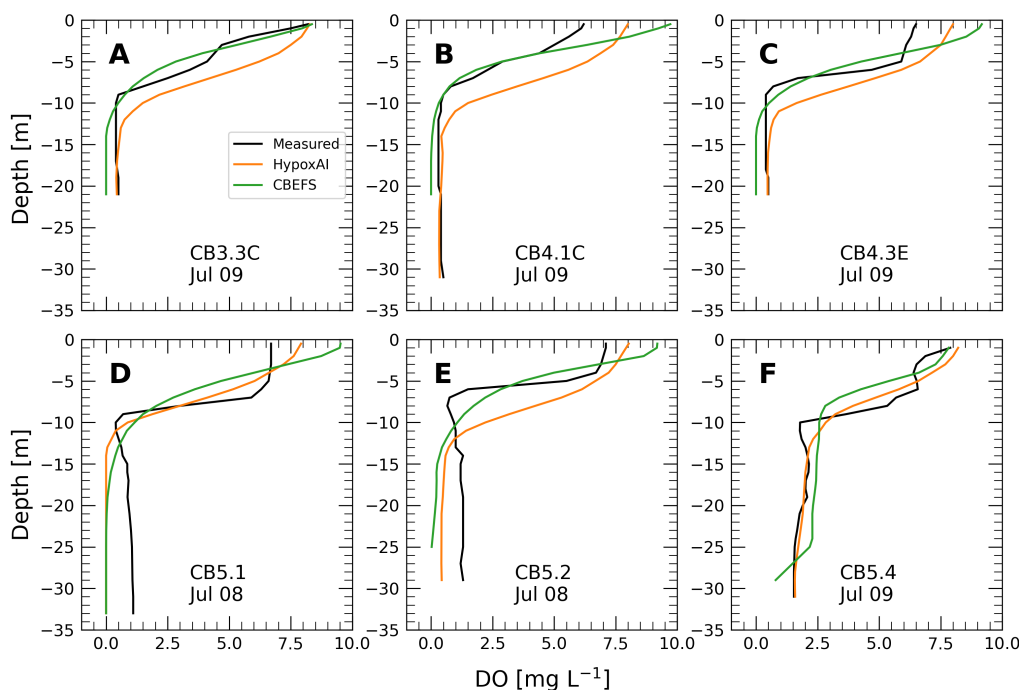


FIG. 6. Same as Fig. 5 except for July 2019 representative of the peak hypoxia period.

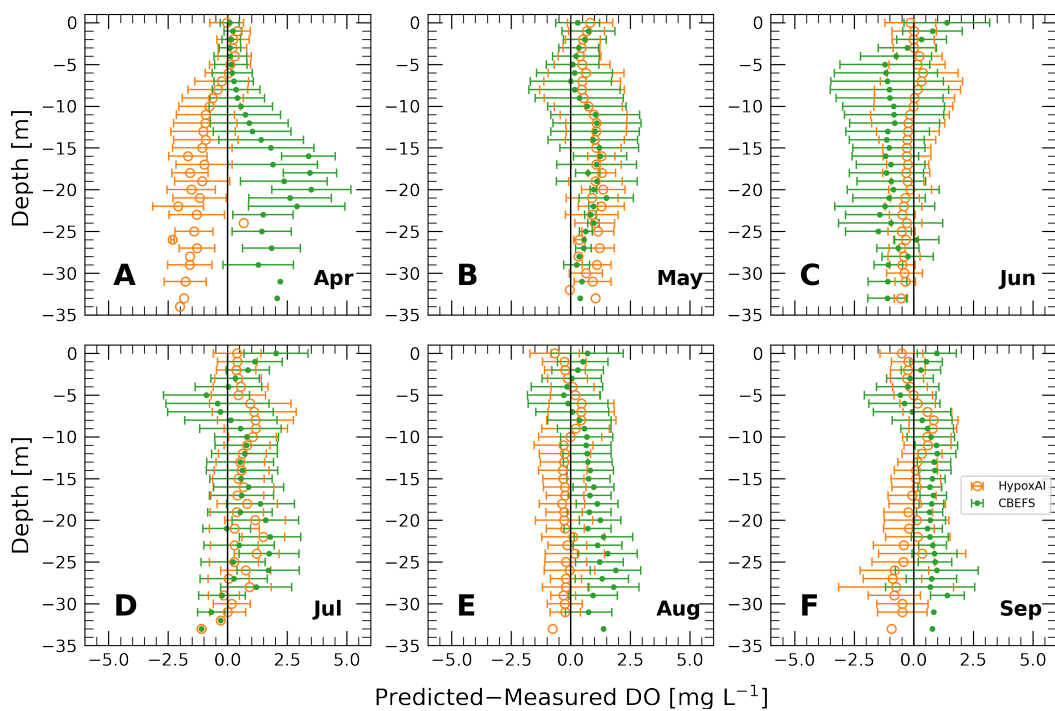


FIG. 7. Monthly difference between predicted and measured DO profiles during warm season.

7. Discussions

Hypoxia is the culmination of the complex interplay between oxygen supply and consumption processes that can occur over an extended period, ultimately resulting in oxygen depletion in a body of water when oxygen consumption out-competes supply. To predict this issue, our study utilized machine learning techniques to account for multiple key processes associated with oxygen supply and consumption, incorporating not only hydrodynamic information provided by CBEFS, but also atmospheric reanalysis products as well as satellite-derived remote-sensing data which more accurately account for the influence of algal biomass. One notable result of this endeavor is the cross-validation study (Fig. 3 and Table 4) which demonstrated stability in the test set outcomes, regardless of the specific year used for validation. This is interesting because environmental variables in the Chesapeake Bay have been experiencing significant changes (Ni et al. 2020; Turner et al. 2021; Hinson et al. 2022; Frankel et al. 2022) over the time period encompassed by our dataset. It suggests that despite the long-term environmental changes, the dynamics between short-term features and short-term DO concentrations remained sufficiently consistent to train a reliable model. This might also indicate the stability of such models in the face of future climate change. However, we advise against applying the current version of the model to future projections, because scenarios exhibiting sudden or extreme changes beyond the model's trained range would result in poor model performance. Updating and retraining the model by incorporating new data into the existing training dataset will be key to maintaining the model's relevance and accuracy in a rapidly changing climate scenario.

While our study demonstrated the efficacy of modern artificial intelligence in predicting this environmental phenomenon, it also opened up more questions for future research, particularly in understanding the intricate dynamics governing oxygen levels. One notable finding is that although satellite-derived Rayleigh-corrected reflectance is a significant predictor of dissolved oxygen in the Chesapeake Bay, its addition to input features does not improve model accuracy beyond the strongest predictor, which is 3D water temperature. This can be attributed to several factors. First, benthic respiration plays a significant role in oxygen consumption in the Chesapeake Bay (Officer et al. 1984; Kemp et al. 1992; Li et al. 2015). This process is not directly observable via remote sensing, but is strongly influenced by temperature (Roehm 2005; Murphy et al. 2011). Since water temperature significantly affects the metabolic rates of not only benthic but also pelagic organisms,

temperature directly impacts the rate of oxygen consumption across both benthic and pelagic zones. This could partly explain why the 3D water temperature emerges as an overwhelmingly strong predictor of DO. Second, remote sensing detects color of phytoplankton pigment at the surface, used as a proxy for the standing stock of phytoplankton, whereas it is the primary productivity that is more relevant to bottom DO which can vary temporally. This variability is highlighted in the study by Zheng and DiGiacomo (2020), which demonstrates that the correlation between remote-sensing reflectance and bottom DO levels in deeper parts of the Chesapeake Bay is more pronounced during warmer seasons with active algal growth, compared to periods of cooler seasons with less algal activity. Finally, the use of Rayleigh-corrected top-of-atmosphere reflectance in this study as opposed to atmospherically corrected remote-sensing reflectance might have introduced more "noise" than "signal". We chose to use Rayleigh-corrected Level-1 data to maximize data coverage, whereas Level-2 remote-sensing reflectance data suffer from high data-masking due to the presence of unfavorable conditions such as clouds, aerosols, glint, and high turbidity. The hope was to leave more feature engineering to the model, i.e., allow the model to find relevant features in the reflectance data. However, our results indicate that, for the specific task of this study, there might have been too much noise in the Rayleigh-corrected data and/or there are not enough training examples for the model to learn relevant features in these data. As such, further research is necessary to determine the extent to which satellite ocean color data should be processed to improve feature engineering relevant for hypoxia forecasting.

Although an improvement in model accuracy with the addition of Rayleigh-corrected MODIS reflectance data is absent, it is important to note that this finding was made when 3D water temperature data from the output of a hydrodynamic model is used. In situations where such 3D data are unavailable, e.g., when only satellite data are provided, the potential contribution of ocean color data to the forecasting of hypoxia in the Chesapeake Bay must be further evaluated. Further research is necessary to assess how well the model perform with only satellite ocean color and SST data.

With respect to the selection of input data sources, we primarily utilized data from the Chesapeake Bay Environmental Forecasting System (CBEFS), which incorporates various terrestrial inputs. However, it is important to acknowledge that this model includes some form of groundwater biogeochemistry, which plays a critical role in the bay's hydrological and ecological dynamics

(Brookfield et al. 2021), but not in an explicit way. Incorporating relevant datasets could account for chemical pathways from groundwater discharge in supporting or suppressing algal blooms and subsequent hypoxic events. In addition, the integration of microbial datasets could enhance the model's skill through better understanding of the succession of microbial and algal blooms (Sison-Mangus et al. 2016; Cheng et al. 2021). Despite their potential value, the incorporation of chemical and microbial datasets into existing predictive models is not without challenges. One of the primary limitations of these datasets is their restricted spatial and temporal coverage, which can introduce significant uncertainties into the model. These uncertainties can stem from both the inherent variability in microbial and chemical processes and the sporadic nature of data collection. Furthermore, from an operational perspective, which is the ultimate goal of this study, the utility of these datasets is constrained by data latency issues. Timely data acquisition is essential for operational forecasting, and the delayed availability of chemical and microbial data can limit their applicability in real-time environmental prediction systems. Therefore, while the integration of these datasets holds promise for advancing our predictive capabilities, careful consideration must be given to their limitations, uncertainties, and data latency.

Another important area of future research is the handling of missing values in input data. In our study, we addressed the challenge of missing data by adopting the convention of filling missing values with zeros, a common practice in computer science. While filling missing values with zeros is a straightforward and computationally efficient method, it may not accurately represent the underlying environmental processes; in addition, zero represents a physically realistic value for some variables such as temperature, salinity, and reflectance. Therefore, filling input variables with zero could potentially introduce biases into the model. In environmental science, more sophisticated imputation techniques have been developed, ranging from simple statistical methods to advanced machine learning approaches that consider spatial and temporal correlations to predict missing data points (Fernandes et al. 2015; Stock et al. 2020; Hamid Mohebzadeh and Biswas 2021). However, such data were unavailable for Rayleigh-corrected MODIS data, and developing such gap-filling algorithms for our datasets would be a major undertaking that warrants a separate study. Thus, the impact of replacing missing values with zeros on the statistical integrity of our environmental variables and, consequently, on the model's performance remains an area of uncertainty and requires further investigation.

To broaden the scope, the application of artificial intelligence in Earth sciences benefit from a thorough examination of the model's architecture and its correlation with the observed results. This would involve a detailed analysis of the optimized hyperparameters to understand their influence on model performance. For instance, examining how the optimized hyperparameters reflect the specific challenges of DO prediction, or comparing our model's architecture and performance with similar models in environmental science, could provide deeper insights into the capabilities and limitations of LSTM models in this field. Additionally, exploring potential modifications to the model architecture to enhance its predictive accuracy and adaptability to different environmental contexts would be a significant contribution to this area of research. Such comprehensive evaluations would guide future developments in this rapidly evolving domain.

8. Summary

This study developed a deep learning model, called HypoxAI, which is capable of integrating data from multiple sources with disparate data structures. The model was designed to predict DO concentrations in the Chesapeake Bay based on a variety of environmental variables, including temperature, salinity, satellite-derived Rayleigh-correction reflectance, currents, and wind. In evaluating the performance of HypoxAI, the results showed good overall agreement with measurements, and the model was able to learn multiple types of vertical DO profiles, which suggests its ability to capture complex patterns and variability in DO concentrations.

Our study also found that the 3D temperature field was the strongest predictor of DO levels, followed by satellite-derived Rayleigh-corrected reflectance, currents, wind, and salinity. Having 3D temperature field as an input feature is sufficient to optimize model performance; adding the satellite reflectance data as we did here, did not improve the AI model skill, likely due to atmospheric interference, as discussed.

However, it can be a significant undertaking to obtain 3D temperature field which requires running a hydrodynamic model like the CBEFS. Future studies could investigate how the model would perform using only more readily available data sources as predictors such as satellite-derived sea surface temperature and ocean color, and how to improve the feature engineering of ocean color data. This approach could potentially reduce operational costs while maintaining accurate DO predictions.

Acknowledgments. This work has been performed under funding from NASA Earth Science Technology Office (ESTO) Advanced Information Systems Technology (AIST) Program under Grant NNH18ZDA001N-AIST, and NOAA Ocean Remote Sensing program. Funding for St-Laurent and Friedrichs is from NOAA's Center for Coastal Ocean Science under award NA16NOS4780207 to the Virginia Institute of Marine Science.

Data availability statement. HypoxAI models and scripts to run them can be found at <https://github.com/coastwatch/HypoxAI-ChesapeakeBay>. Training datasets are available at <https://issues.pangaea.de/browse/PDI-37066>. Information on the availability of CBEFS outputs can be found at <https://doi.org/10.25773/q2kh-rd09>

APPENDIX A

Model Training Histories

a. Ablation study

Training histories of all model configurations evaluated in the ablation study are shown in Fig. A1, which were summarized in Table 3 in the main text.

b. Cross Validations

Shown here are training histories using different year as validation dataset for four cross validation models, DCW (Fig. A2), DCWS (Fig. A3), RDCW (Fig. A4), and RDCWS (Fig. A5), which were summarized in Table 4 and Fig. 3.

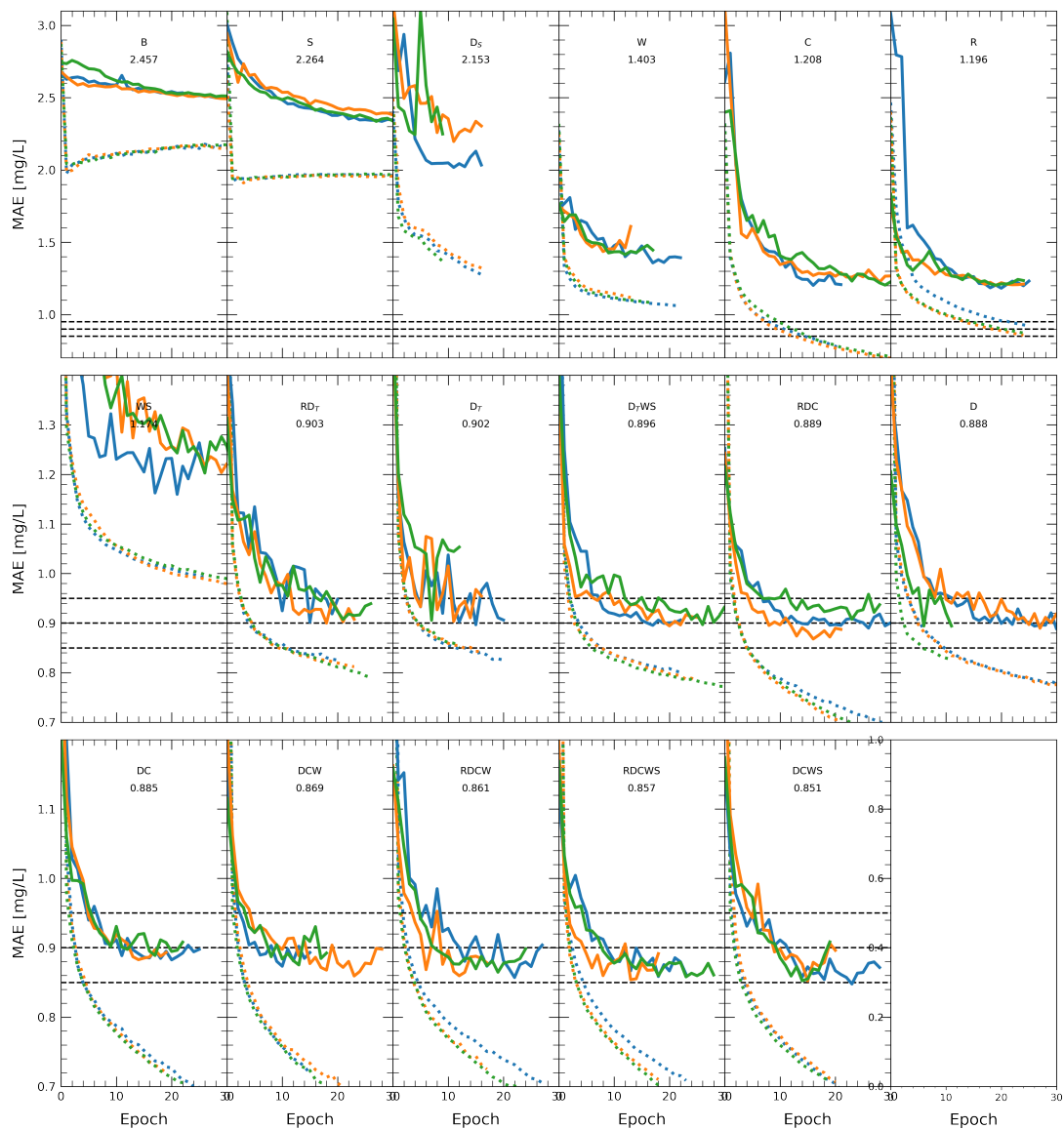


FIG. A1. Training histories of different model configurations with various combinations of input features.

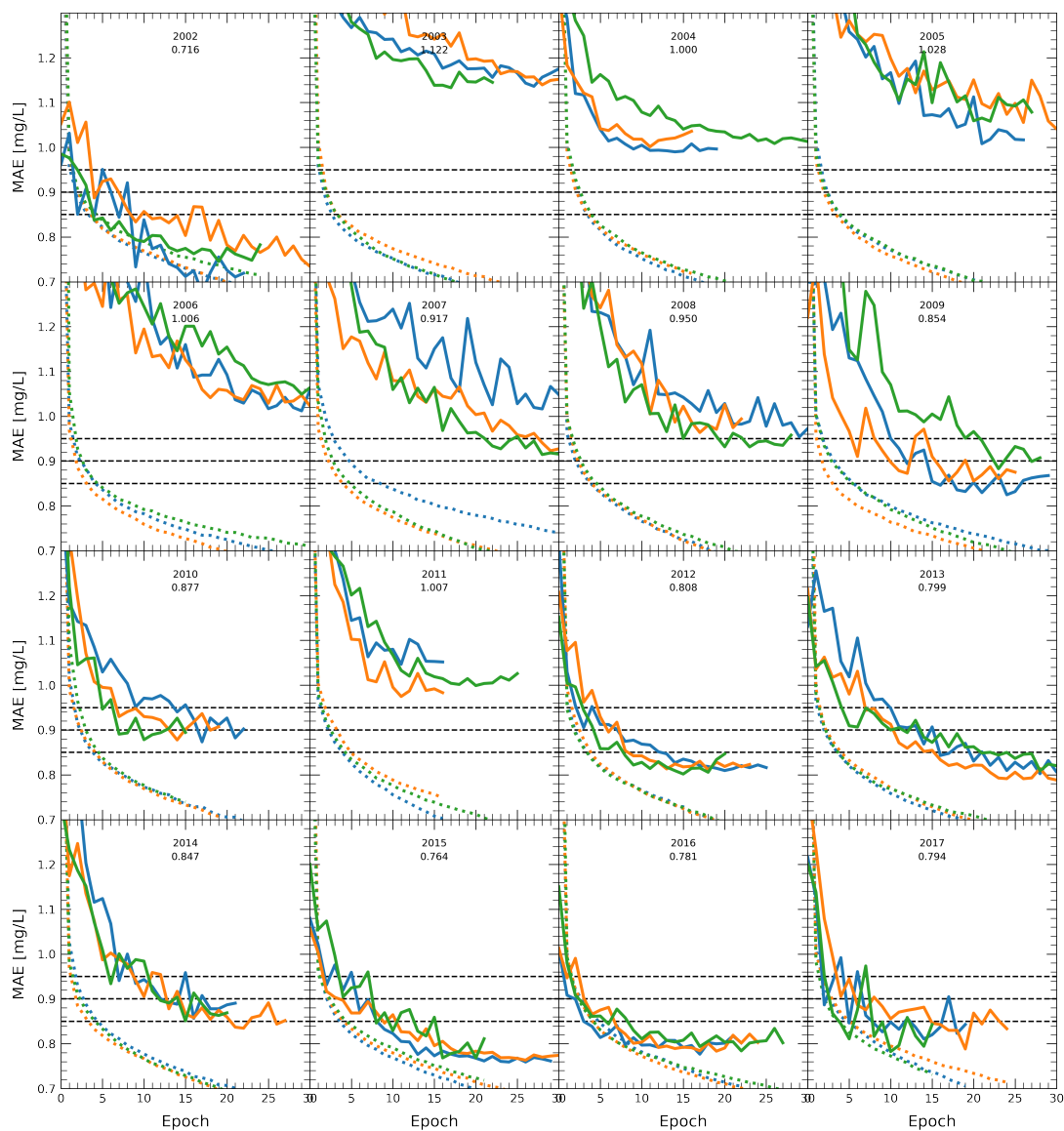


FIG. A2. Training histories for model DCW using different year as validation.

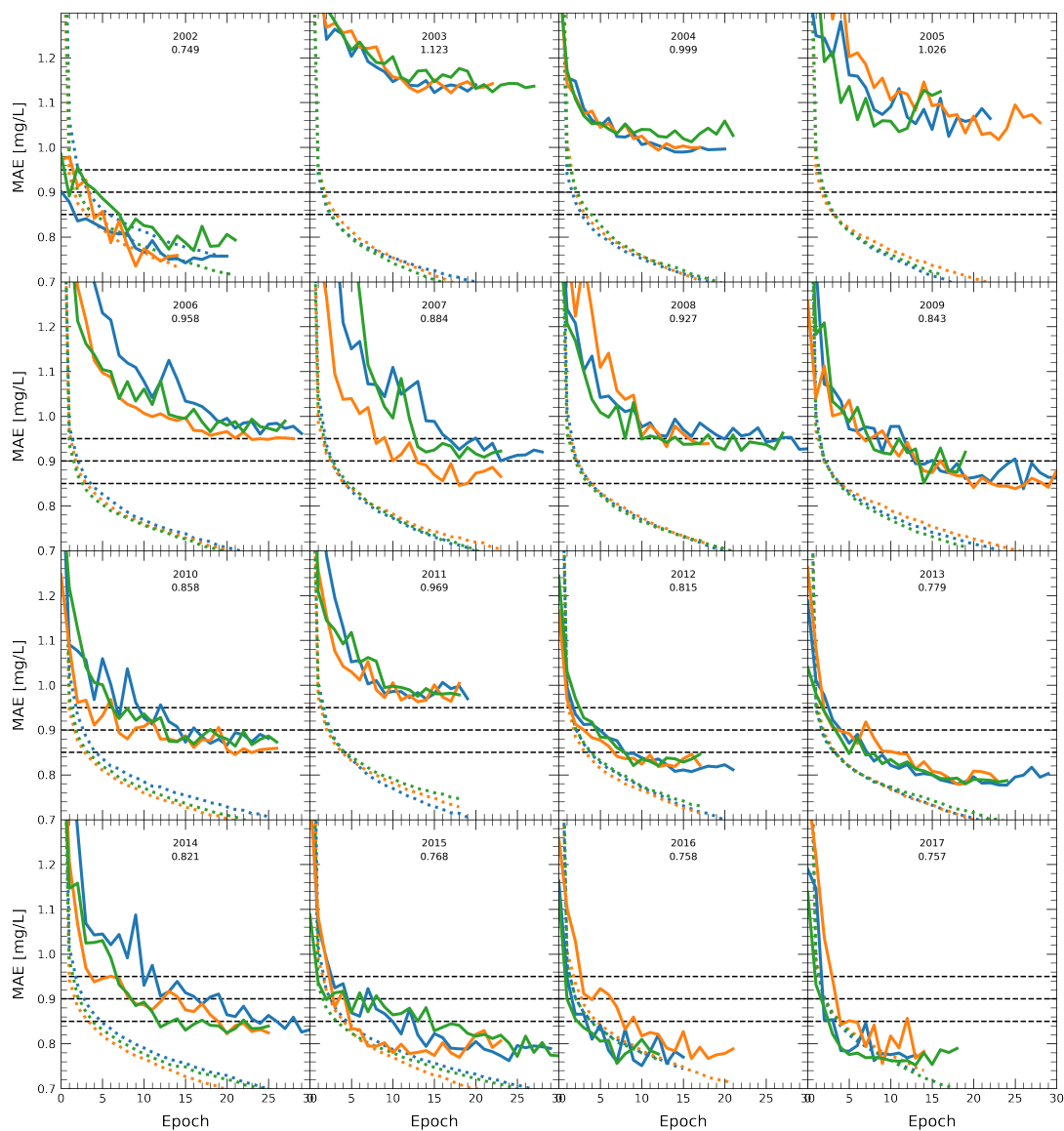


FIG. A3. Same as Fig A2 except for model DCWS.

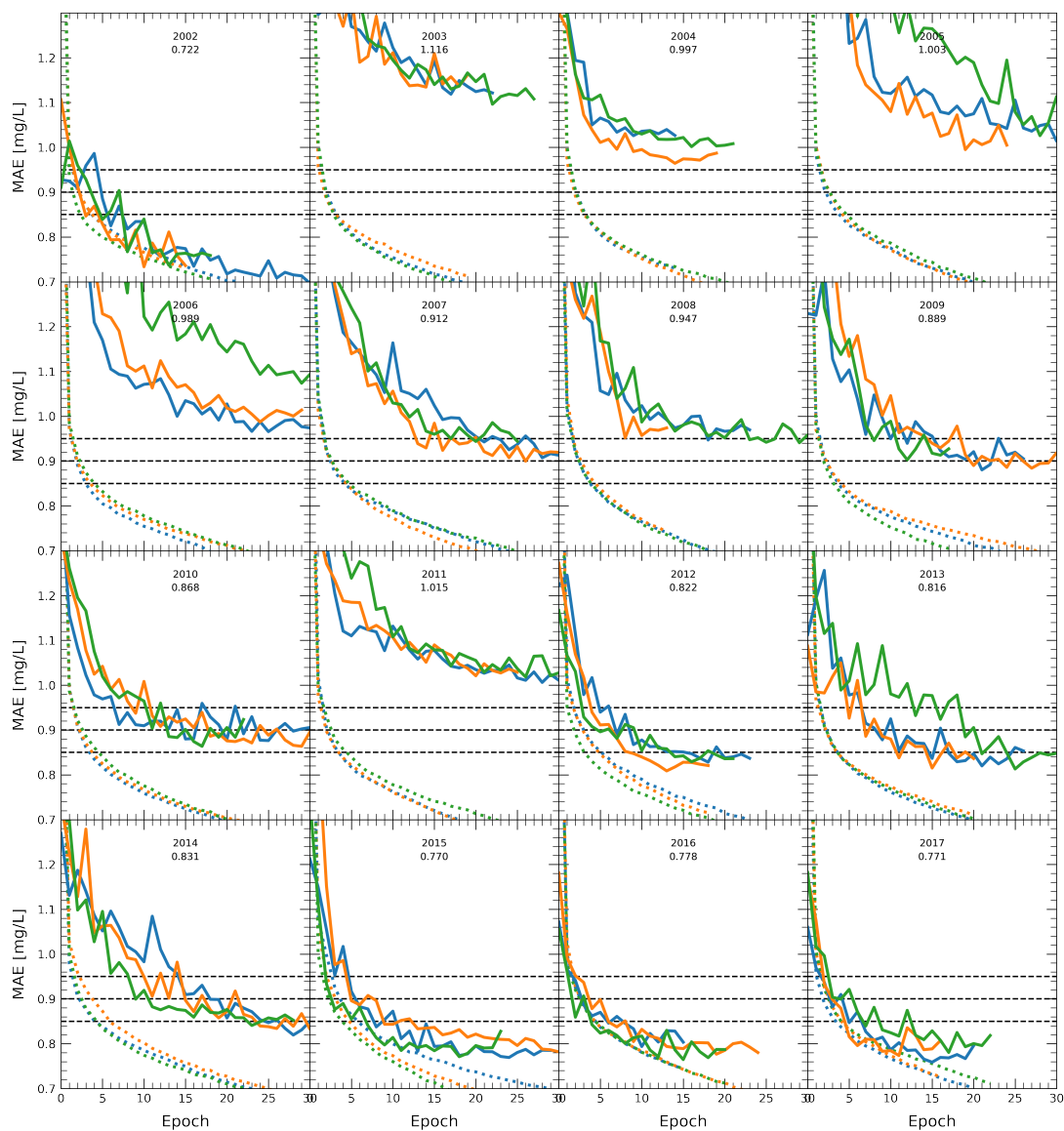


FIG. A4. Same as Fig A2 except for model RDCW.

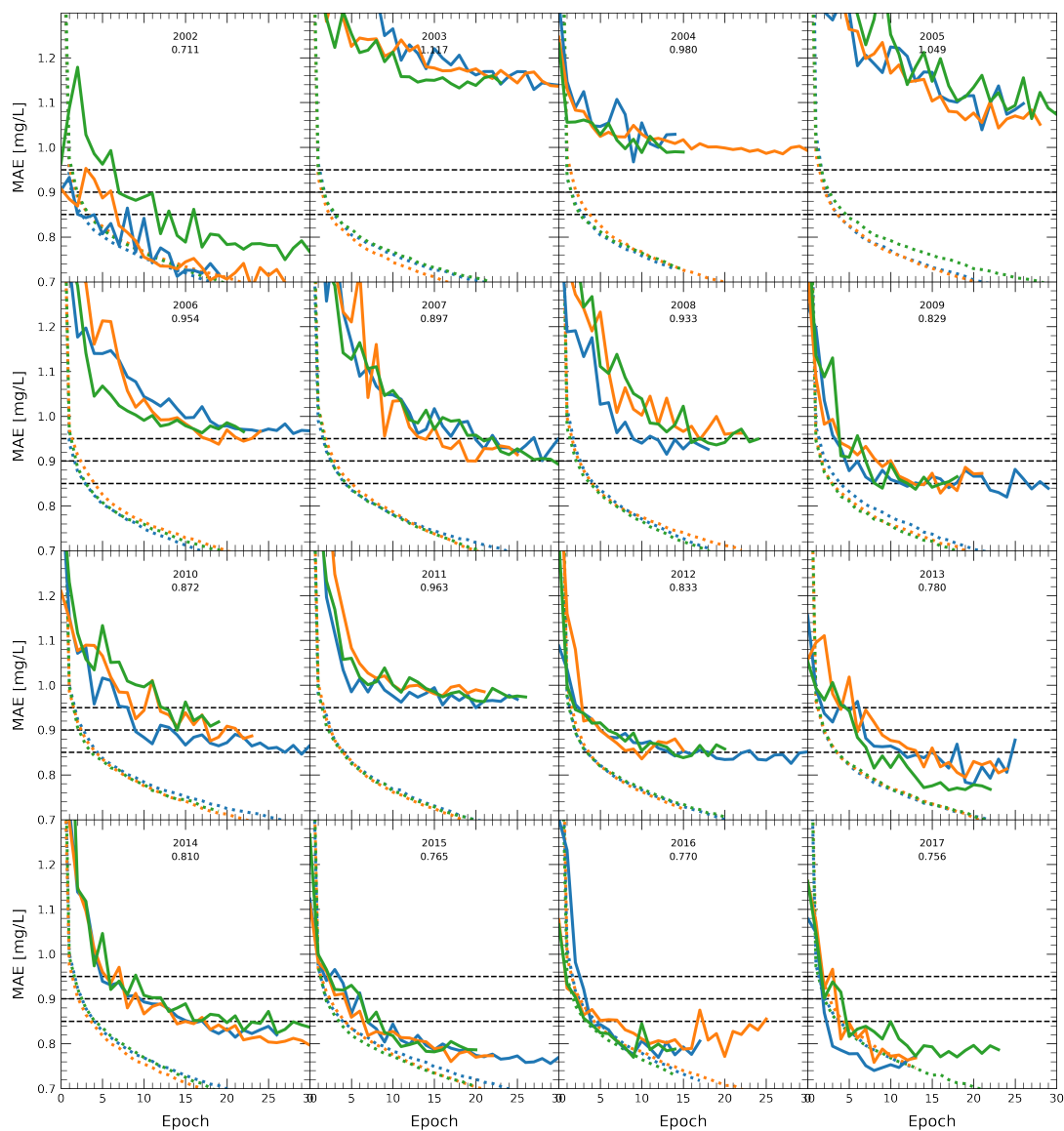


FIG. A5. Same as Fig A2 except for model RDCWS.

References

- Aurin, D., A. Mannino, and B. Franz, 2013: Spatially resolving ocean color and sediment dispersion in river plumes, coastal systems, and continental shelf waters. *Remote Sensing of Environment*, **137**, 212–225, <https://doi.org/10.1016/j.rse.2013.06.018>.
- Bever, A. J., M. A. Friedrichs, and P. St-Laurent, 2021: Real-time environmental forecasts of the chesapeake bay: Model setup, improvements, and online visualization. *Environmental Modelling and Software*, **140**, 105 036, <https://doi.org/10.1016/j.envsoft.2021.105036>.
- Brookfield, A. E., A. T. Hansen, P. L. Sullivan, J. A. Czuba, M. F. Kirk, L. Li, M. E. Newcomer, and G. Wilkinson, 2021: Predicting algal blooms: Are we overlooking groundwater? *Science of The Total Environment*, **769**, 144 442, <https://doi.org/10.1016/j.scitotenv.2020.144442>.
- Chantry, M., H. Christensen, P. Dueben, and T. Palmer, 2021: Opportunities and challenges for machine learning in weather and climate modelling: hard, medium and soft ai. *Philosophical Transactions of the Royal Society A*, **379 (2197)**, 20200 083, <https://doi.org/10.1098/rsta.2020.0083>.
- Cheng, Y., V. N. Bhoot, K. Kumbier, M. P. Sison-Mangus, J. B. Brown, R. Kudela, and M. E. Newcomer, 2021: A novel random forest approach to revealing interactions and controls on chlorophyll concentration and bacterial communities during coastal phytoplankton blooms. *Scientific reports*, **11 (1)**, 19 944.
- Culver, T., and Coauthors, 2020: SBG User Needs and Valuation Study Final Report September, 2020. Zenodo, <https://doi.org/10.5281/zenodo.6347764>.
- Culver, T., and Coauthors, 2022: SBG User Needs and Valuation Study Final Report, December 2021. Zenodo, <https://doi.org/10.5281/zenodo.6347789>.
- Du, J., J. Shen, K. Park, Y. P. Wang, and X. Yu, 2018: Worsened physical condition due to climate change contributes to the increasing hypoxia in chesapeake bay. *Science of The Total Environment*, **630**, 707–717, <https://doi.org/10.1016/j.scitotenv.2018.02.265>.
- Fernandes, J. A., X. Irigoien, J. A. Lozano, I. Inza, N. Goikoetxea, and A. Pérez, 2015: Evaluating machine-learning techniques for recruitment forecasting of seven north east atlantic fish species. *Ecological Informatics*, **25**, 35–42.

- Frankel, L. T., M. A. Friedrichs, P. St-Laurent, A. J. Bever, R. N. Lipcius, G. Bhatt, and G. W. Shenk, 2022: Nitrogen reductions have decreased hypoxia in the chesapeake bay: Evidence from empirical and numerical modeling. *Science of The Total Environment*, **814**, 152722, <https://doi.org/10.1016/j.scitotenv.2021.152722>.
- Ghobadi, F., and D. Kang, 2023: Application of machine learning in water resources management: A systematic literature review. *Water*, **15** (4), <https://doi.org/10.3390/w15040620>.
- Hamid Mohebzadeh, P. D., Esmail Mokari, and A. Biswas, 2021: A machine learning approach for spatiotemporal imputation of modis chlorophyll-a. *International Journal of Remote Sensing*, **42** (19), 7381–7404, <https://doi.org/10.1080/01431161.2021.1957513>.
- Hinson, K. E., M. A. Friedrichs, P. St-Laurent, F. Da, and R. G. Najjar, 2022: Extent and causes of chesapeake bay warming. *JAWRA Journal of the American Water Resources Association*, **58** (6), 805–825, <https://doi.org/10.1111/1752-1688.12916>.
- Hood, R. R., and Coauthors, 2021: The chesapeake bay program modeling system: Overview and recommendations for future development. *Ecological Modelling*, **456**, 109635, <https://doi.org/10.1016/j.ecolmodel.2021.109635>.
- Irby, I. D., M. A. M. Friedrichs, F. Da, and K. E. Hinson, 2018: The competing impacts of climate change and nutrient reductions on dissolved oxygen in chesapeake bay. *Biogeosciences*, **15** (9), 2649–2668, <https://doi.org/10.5194/bg-15-2649-2018>.
- Irby, I. D., and Coauthors, 2016: Challenges associated with modeling low-oxygen waters in chesapeake bay: a multiple model comparison. *Biogeosciences*, **13** (7), 2011–2028, <https://doi.org/10.5194/bg-13-2011-2016>.
- Isabelle, D. A., and M. Westerlund, 2022: A review and categorization of artificial intelligence-based opportunities in wildlife, ocean and land conservation. *Sustainability*, **14** (4), <https://doi.org/10.3390/su14041979>.
- Kemp, W., P. Sampou, J. Garber, J. Turtle, and W. Boynton, 1992: Seasonal depletion of oxygen from bottom waters of chesapeake bay: roles of benthic and planktonic respiration and physical exchange processes. *Marine Ecology Progress Series*, **85**, 137–152, <https://doi.org/10.3354/meps085137>.

- Lee, C. M., N. F. Glenn, E. N. Stavros, J. Luvall, K. Yuen, C. Hain, and S. Schollaert Uz, 2022: Systematic integration of applications into the surface biology and geology (sbg) earth mission architecture study. *Journal of Geophysical Research: Biogeosciences*, **127** (4), e2021JG006720, <https://doi.org/10.1029/2021JG006720>.
- Li, Y., M. Li, and W. M. Kemp, 2015: A budget analysis of bottom-water dissolved oxygen in chesapeake bay. *Estuaries and Coasts*, **38** (6), 2132–2148.
- Masood, A., and K. Ahmad, 2021: A review on emerging artificial intelligence (ai) techniques for air pollution forecasting: Fundamentals, application and performance. *Journal of Cleaner Production*, **322**, 129072, <https://doi.org/10.1016/j.jclepro.2021.129072>.
- MDDNR, 2022: Chesapeake bay hypoxia reports. URL <https://dnr.maryland.gov/waters/bay/pages/hypoxia-reports.aspx>, accessed 29 March 2023, Online.
- Mouw, C. B., and Coauthors, 2015: Aquatic color radiometry remote sensing of coastal and inland waters: Challenges and recommendations for future satellite missions. *Remote Sensing of Environment*, **160**, 15–30, <https://doi.org/10.1016/j.rse.2015.02.001>.
- Murphy, R. R., W. M. Kemp, and W. P. Ball, 2011: Long-term trends in chesapeake bay seasonal hypoxia, stratification, and nutrient loading. *Estuaries and Coasts*, **34**, 1293–1309, <https://doi.org/10.1007/s12237-011-9413-7>.
- Ni, W., M. Li, and J. M. Testa, 2020: Discerning effects of warming, sea level rise and nutrient management on long-term hypoxia trends in chesapeake bay. *Science of The Total Environment*, **737**, 139717, <https://doi.org/10.1016/j.scitotenv.2020.139717>.
- Officer, C. B., R. B. Biggs, J. L. Taft, L. E. Cronin, M. A. Tyler, and W. R. Boynton, 1984: Chesapeake bay anoxia: Origin, development, and significance. *Science*, **223** (4631), 22–27, <https://doi.org/10.1126/science.223.4631.22>.
- Roehm, C., 2005: *Respiration in wetland ecosystems*, 83–102. <https://doi.org/10.1093/acprof:oso/9780198527084.003.0006>.
- Schollaert Uz, S., T. J. Ames, N. Memarsadeghi, S. M. McDonnell, N. V. Blough, A. V. Mehta, and J. R. McKay, 2020: Supporting aquaculture in the chesapeake bay using artificial intelligence to detect poor water quality with remote sensing. *IGARSS 2020 - 2020*

IEEE International Geoscience and Remote Sensing Symposium, 3629–3632, <https://doi.org/10.1109/IGARSS39084.2020.9323465>.

Schollaert Uz, S., G. E. Kim, A. Mannino, P. J. Werdell, and M. Tzortziou, 2019: Developing a community of practice for applied uses of future pace data to address marine food security challenges. *Frontiers in Earth Science*, **7**, <https://doi.org/10.3389/feart.2019.00283>.

Scott, J. P., and E. Urquhart, 2020: Leveraging design principles to inform the next generation of nasa earth satellites. *Oceanography*, **33**, 128–129, <https://doi.org/10.5670/oceanog.2020.416>.

Shchepetkin, A. F., and J. C. McWilliams, 2005: The regional oceanic modeling system (roms): a split-explicit, free-surface, topography-following-coordinate oceanic model. *Ocean Modelling*, **9** (4), 347–404, <https://doi.org/10.1016/j.ocemod.2004.08.002>.

Sison-Mangus, M. P., S. Jiang, R. M. Kudela, and S. Mehic, 2016: Phytoplankton-associated bacterial community composition and succession during toxic diatom bloom and non-bloom events. *Frontiers in Microbiology*, **7**, 1433.

St-Laurent, P., and M. A. M. Friedrichs, 2024: On the sensitivity of coastal hypoxia to its external physical forcings. *Journal of Advances in Modeling Earth Systems*, **16** (1), e2023MS003 845, <https://doi.org/10.1029/2023MS003845>.

St-Laurent, P., M. A. M. Friedrichs, R. G. Najjar, E. H. Shadwick, H. Tian, and Y. Yao, 2020: Relative impacts of global changes and regional watershed changes on the inorganic carbon balance of the chesapeake bay. *Biogeosciences*, **17** (14), 3779–3796, <https://doi.org/10.5194/bg-17-3779-2020>.

Stock, A., A. Subramaniam, G. L. Van Dijken, L. M. Wedding, K. R. Arrigo, M. M. Mills, M. A. Cameron, and F. Micheli, 2020: Comparison of cloud-filling algorithms for marine satellite data. *Remote Sensing*, **12** (20), <https://doi.org/10.3390/rs12203313>.

Su, J., and Coauthors, 2020: Source partitioning of oxygen-consuming organic matter in the hypoxic zone of the chesapeake bay. *Limnology and Oceanography*, **65** (8), 1801–1817, <https://doi.org/10.1002/lno.11419>.

- Turner, J. S., C. T. Friedrichs, and M. A. M. Friedrichs, 2021: Long-term trends in chesapeake bay remote sensing reflectance: Implications for water clarity. *Journal of Geophysical Research: Oceans*, **126** (12), e2021JC017959, <https://doi.org/10.1029/2021JC017959>.
- Valera, M., R. K. Walter, B. A. Bailey, and J. E. Castillo, 2020: Machine learning based predictions of dissolved oxygen in a small coastal embayment. *Journal of Marine Science and Engineering*, **8** (12), <https://doi.org/10.3390/jmse8121007>.
- VIMS, 2022: Dead-zone report card: Compare the annual severity of chesapeake bay hypoxia. URL https://www.vims.edu/research/topics/dead_zones/forecasts/report_card/index.php, accessed 29 March 2023, Online.
- Wolny, J. L., and Coauthors, 2020: Current and future remote sensing of harmful algal blooms in the chesapeake bay to support the shellfish industry. *Frontiers in Marine Science*, **7**, <https://doi.org/10.3389/fmars.2020.00337>.
- Yu, X., J. Shen, and J. Du, 2020: A machine-learning-based model for water quality in coastal waters, taking dissolved oxygen and hypoxia in chesapeake bay as an example. *Water Resources Research*, **56** (9), e2020WR027227, <https://doi.org/https://doi.org/10.1029/2020WR027227>.
- Zheng, G., and P. M. DiGiacomo, 2020: Linkages between phytoplankton and bottom oxygen in the chesapeake bay. *Journal of Geophysical Research: Oceans*, **125** (2), e2019JC015650, <https://doi.org/10.1029/2019JC015650>.