Science of the Total Environment 814 (2022) 152722

Contents lists available at ScienceDirect

Science of the Total Environment

journal homepage: www.elsevier.com/locate/scitotenv

Nitrogen reductions have decreased hypoxia in the Chesapeake Bay: Evidence from empirical and numerical modeling

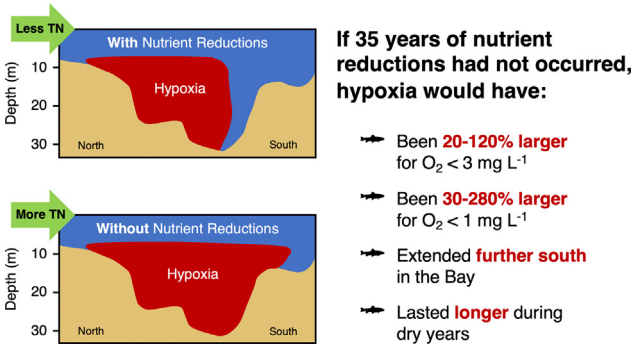
Luke T. Frankel [a,*], Marjorie A.M. Friedrichs [a], Pierre St-Laurent [a], Aaron J. Bever [b], Romuald N. Lipcius [a], Gopal Bhatt [c,d], Gary W. Shenk [c,e]

[a] Virginia Institute of Marine Science, William & Mary, 1370 Greate Road, Gloucester Point, VA, USA
[b] Anchor QEA LLC, 1201 3rd Avenue, Suite 2600, Seattle, WA, USA
[c] Chesapeake Bay Program Office, 1750 Forest Drive, Suite 130, Annapolis, MD, USA
[d] Department of Civil & Environmental Engineering, The Pennsylvania State University, 212 Sackett Building, University Park, PA, USA
[e] U.S. Geological Survey, Virginia and West Virginia Water Science Center, 1730 East Parham Road, Richmond, VA, USA

HIGHLIGHTS

- During 1985–2019, the Chesapeake Bay has undergone nitrogen reductions and warming.
- Data analysis and models were used to quantify the effects of nitrogen reductions.
- Reductions have decreased the duration and southern extent of hypoxia in the Bay.
- From 1985 to 2019, warming has offset 6–34% of these hypoxia improvements.
- Nutrient reductions have made Bay oxygen levels more resilient to climate change.

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:
Received 19 September 2021
Received in revised form 22 December 2021
Accepted 23 December 2021
Available online 31 December 2021

Editor: Ashantha Goonetilleke

Keywords:
Chesapeake Bay
Hypoxia
Oxygen
Nutrient reductions
Climate change
Water quality management

ABSTRACT

Seasonal hypoxia is a characteristic feature of the Chesapeake Bay due to anthropogenic nutrient input from agriculture and urbanization throughout the watershed. Although coordinated management efforts since 1985 have reduced nutrient inputs to the Bay, oxygen concentrations at depth in the summer still frequently fail to meet water quality standards that have been set to protect critical estuarine living resources. To quantify the impact of watershed nitrogen reductions on Bay hypoxia during a recent period including both average discharge and extremely wet years (2016–2019), this study employed both statistical and three-dimensional (3-D) numerical modeling analyses. Numerical model results suggest that if the nitrogen reductions since 1985 had not occurred, annual hypoxic volumes (O_2 < 3 mg L^{-1}) would have been ~50–120% greater during the average discharge years of 2016–2017 and ~20–50% greater during the wet years of 2018–2019. The effect was even greater for O_2 < 1 mg L^{-1}, where annual volumes would have been ~80–280% greater in 2016–2017 and ~30–100% greater in 2018–2019. These results were supported by statistical analysis of empirical data, though the magnitude of improvement due to nitrogen reductions was greater in the numerical modeling results than in the statistical analysis. This discrepancy is largely accounted for by warming in the Bay that has exacerbated hypoxia and offset roughly 6–34% of the improvement from nitrogen

Abbreviations: 3-D, Three-Dimensional; USEPA, U.S. Environmental Protection Agency; TMDL, Total Maximum Daily Load; GLM, Generalized Linear Model; CBP, Chesapeake Bay Program; WQMP, Water Quality Monitoring Program; USGS, U.S. Geological Survey; AIC, Akaike's Information Criterion; ROMS, Regional Ocean Modeling System; ISS, Inorganic Suspended Solids; DIN, Dissolved Inorganic Nitrogen; ON, Organic Nitrogen; PON, Particulate Organic Nitrogen; DON, Dissolved Organic Nitrogen; TN, Total Nitrogen; DOC, Dissolved Organic Carbon; DLEM, Dynamic Land Ecosystem Model; NAM, North American Mesoscale Forecast System; ADCIRC, Advanced Circulation model; RMSD, Root-Mean Squared Difference; RIM, River Input Monitoring; WRTDS, Weighted Regression on Time Discharge and Season; CAST, Chesapeake Assessment and Scenario Tool; LOESS, Locally Estimated Scatterplot Smoothing.

* Corresponding author at: FB Environmental Associates, Cocheco Mills #3, 383 Central Ave, Suite 267, Dover, NH 03820, USA.
E-mail address: frankelluke@gmail.com (L.T. Frankel).

http://dx.doi.org/10.1016/j.scitotenv.2021.152722

reductions. Although these results may reassure policymakers and stakeholders that their efforts to reduce hypoxia have improved ecosystem health in the Bay, they also indicate that greater reductions are needed to counteract the ever-increasing impacts of climate change.

1. Introduction

Hypoxia resulting from anthropogenic eutrophication has become one of the greatest threats to the health of estuarine and coastal ecosystems worldwide due to its ability to degrade habitat, decrease biodiversity, and alter food-web interactions (Diaz and Rosenberg, 2008). Despite ongoing management efforts, the severity of hypoxia globally is projected to increase in the future as an ever-increasing human population places more anthropogenic stress on coastal environments through continued land-use change, urbanization, and climate change (Altieri and Gedan, 2015). Although at a fundamental level the development of coastal hypoxia is similar across all systems, where oxygen use below the pycnocline draws down oxygen concentrations to hypoxic levels, the characteristics of hypoxia can vary substantially between systems due to differences in both natural and anthropogenic processes (Carstensen and Conley, 2019; Fennel and Testa, 2019; Howarth, 2008; Meier et al., 2019; Rabalais and Turner, 2019; Wang et al., 2018). Because of these differences, hypoxia needs to be studied individually for each system to develop management goals that take into account the combination of local and global drivers so that appropriate ecosystem restoration actions can be taken.

The Chesapeake Bay is one such system, for which hypoxia continues to negatively impact ecosystem health. The Bay is naturally a highly productive estuary that receives a large flux of nutrients from terrestrial sources due to its extensive watershed that spans six states and the District of Columbia, encompassing 164,200 km^2 (Kemp et al., 2005). The magnitude of this nutrient flux varies in response to changes in land-use within the watershed, the most notable of which occurred during the European colonization of the area in the 17th century, and the post-World War II increase in fertilizer usage throughout the watershed. Through analysis of both observational data (Hagy et al., 2004; Kemp et al., 2005; Officer et al., 1984) and proxies within sediment cores (Cooper and Brush, 1991, 1993), current low oxygen conditions within the Chesapeake Bay have been definitively attributed to anthropogenic eutrophication.

Management efforts to restore ecosystem health within the Chesapeake Bay began in 1976, in response to a visible decline in natural resources (USEPA, 1982). A Total Maximum Daily Load (TMDL) for the Chesapeake Bay was put in place in 2010, establishing pollutant load allocations for the various subbasins in the watershed necessary to achieve water quality standards that protect living resources. These allocated loads provide managers with a framework to develop Watershed Implementation Plans aimed at reducing nutrient pollution from point and nonpoint sources throughout the watershed (USEPA, 2010).

Despite the progress that has been made over the past 35 years in reducing nutrient concentrations entering the Bay (Harding et al., 2016a; Hirsch et al., 2010; Lefcheck et al., 2018; Testa et al., 2008), hypoxia remains a significant water quality issue for the Chesapeake Bay region. Seasonal hypoxia continues to persist in large portions of the Bay today (Bever et al., 2021), and a long-term decline is not directly visible from time series of hypoxic volume computed via interpolation of dissolved oxygen observations (Fig. S1). In fact, several recent high precipitation years (i.e., 2018 and 2019) have been associated with unusually extensive volumes of hypoxia (1056–1408 km^3 days compared to 882–904 km^3 days in 2016 and 2017 for $O_2 < 3$ mg L^{-1}). Wet years are typically associated with lower water quality due to higher nutrient loading from land, and the period from June 2018 to May 2019 was extremely wet in the Chesapeake Bay region. During that time, four out of six states in the watershed (Maryland, Virginia, Pennsylvania, and West Virginia) experienced the wettest 12-months on record dating back to 1895 (National Centers for Environmental Information, 2019).

In addition to the wet conditions in recent years, there are multiple other hypotheses for why the water quality improvements observed in the watershed are not being reflected in the estuary. One potential explanation for this lack of estuarine response is that the Chesapeake Bay has undergone a regime shift (Duarte et al., 2009), where positive feedbacks such as enhanced recycling of ammonium and phosphate in sediments and the decline in oyster populations under low oxygen conditions have transformed the system to further support hypoxia (Cerco and Noel, 2007; Kemp et al., 2005; Newell, 1988; Testa and Kemp, 2012). Changes to the physical dynamics within the Bay in response to external climate factors, such as the Bermuda highs influence on wind direction (Du et al., 2018; Scully, 2010a), have also been highlighted as a potential explanation. In addition to regional changes in wind patterns, large-scale increases in temperature and precipitation due to climate change (Ni et al., 2019; Hinson et al., 2021) may have already masked some of the expected improvements in bottom oxygen. Since there is already evidence that some of the positive feedbacks in biogeochemistry that further support hypoxia are weakening (Schulte et al., 2009; Testa et al., 2018), the impact of climate change is likely to become the most important factor counteracting the impact of management efforts within the Chesapeake Bay in the future (Irby etal., 2018; Ni et al., 2020).

It has been suggested that nutrient reductions over the 35-year period from 1985 to 2019 have made the Chesapeake Bay more resilient to years with high riverine discharge by decreasing hypoxia (Maryland Department of Natural Resources, 2019), however this idea has been challenged as a result of record-breaking precipitation throughout the watershed in 2018 and 2019 that fueled particularly large volumes of hypoxic water. In this study, the impact of decades of nutrient management efforts is examined by estimating how much worse hypoxia would have been in recent years, if no such nutrient reductions had taken place. Here the difference between the hypoxia that occurred and the hypoxia that would have occurred if no nutrient reductions had taken place is referred to as the increased "resilience" of the Bay for this time period. By generating a greater understanding of this increased resilience due to nutrient reductions, this study assists managers and policymakers in better evaluating the impact that management actions have already had, so they can continue to support and (or) modify these actions in the future to achieve the overall goal of improved water quality throughout the Bay.

2. Methods

To estimate the impact of watershed nitrogen reductions on hypoxia, this study employed empirical statistical methods in concert with three-dimensional (3-D) numerical modeling. The statistical method used generalized linear models (GLMs) to quantify changes in the relationship between riverine discharge and data-based estimates of hypoxic volume over the 35-year period from 1985 to 2019. The numerical modeling approach consisted of realistic and sensitivity simulations run over the four-year period from 2016 to 2019 to allow for the evaluation of two average discharge years (2016 and 2017) and two high discharge years (2018 and 2019) that occurred recently. These two approaches provide independent methods of evaluating any enhanced resilience of the Chesapeake Bay to adverse environmental conditions over the 1985 to 2019 time period.

2.1. Observational data

Physical and biogeochemical data have been collected within the Chesapeake Bay since 1984 as part of the Chesapeake Bay Program's (CBP's) water quality monitoring program (WQMP). This program spans the full extent of the estuary through the routine sampling of 49 fixed stations in the mainstem of the Bay and 105 fixed stations within the Bay's tidal tributaries (CBP, 2012). These stations are typically sampled once

per month, with mainstem and larger tributary stations sampled twice per month during the warmer parts of the year (various intervals from March–October before 2010, June–August after 2010). At each station, hydrographic vertical profiles measuring temperature, salinity, and dissolved oxygen (O_2) are made at 1–2 m intervals throughout the water column. Water samples are also taken from the surface and bottom of the water column and at locations above and below the pycnocline. Laboratory analysis is performed on these water samples for dissolved and particulate constituents (e.g., nutrients, pigments, and suspended solids; CBP, 2012).

Data-based estimates of hypoxic volume for specific thresholds, i.e., $O_2 < 3$ mg L^{-1} and $O_2 < 1$ mg L^{-1}, were calculated monthly to twice-monthly by spatially interpolating available data from CBP's WQMP during each sampling period. Thresholds of 3 and 1 mg L^{-1} were the focus of this analysis as they correspond to the O_2 requirements of ecologically important organisms found at depth in the Chesapeake Bay, such as benthic infauna and the Bay anchovy (USEPA, 2003). The interpolation to obtain hypoxic volume was performed using CBP's volumetric inverse distance squared interpolator program (USEPA, 2003). This was done in the same manner as previous studies (Bever et al., 2013, 2018), except that data from 56 stations were used instead of 13 (Fig. 1). Using 56 stations improves the estimates by filling spatial gaps in data coverage when key stations are missing, without substantially increasing the temporal uncertainty associated with a longer sampling period. The specific date assigned to each hypoxic volume estimate was the average of the collection dates of all the stations used in the interpolation. Following Bever et al. (2021), all data collected within ±7 days of a data collection event at station CB3.2 was used in an individual interpolation. CB3.2 was chosen as it has been sampled more often than any other station in the interpolation.

Annual hypoxic volume, otherwise known as cumulative hypoxic volume (Bever et al., 2013) or total annual hypoxic volume (Bever et al., 2021), was calculated in order to quantify the total amount of hypoxia for a given year. This was accomplished by first linearly interpolating the hypoxic volume estimates from each sampling period to obtain a value for each day, and then calculating the sum of these daily values over a calendar year. The resulting time series consists of annual estimates of volumes (km^3 days) of water with oxygen less than a specific threshold, spanning the 35-year period from 1985 to 2019.

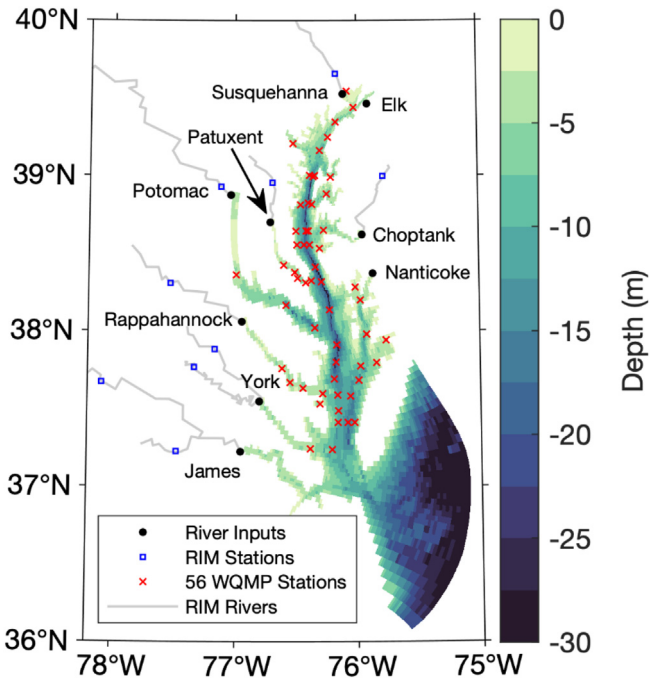


Fig. 1. ChesROMS-ECB numerical model bathymetry. Black circles represent the 9 river locations where terrestrial inputs enter the estuarine numerical model. Blue squares indicate the locations of the River Input Monitoring (RIM) stations. Red x's indicate the Water Quality Monitoring Program station locations used to estimate hypoxic volume and assess numerical model skill.

A time series of the average annual freshwater discharge entering the Chesapeake Bay over a water year (1 October–30 September) was obtained for the same period (1985–2019) from the U.S. Geological Survey (USGS; USGS, 2021). These values of total streamflow entering the Bay are calculated by the USGS using discharge data from gauges on the three largest rivers entering the Bay: the Susquehanna (01576000), Potomac (01646500), and James (02037500) rivers. Together, these rivers account for ~70% of the total freshwater flow entering the Bay. When examined in the context of hypoxia, averaging freshwater discharge over a water year is preferred to an average over a calendar year, since discharges that occur late in the calendar year (e.g., November) do not influence summer hypoxic conditions associated with that calendar year, but may affect hypoxia the following summer.

2.2. Statistical analysis

To investigate how the relationship between river discharge and hypoxic volume has changed over time, two 6-year time periods in the beginning and end of the interval from 1985 to 2019 were selected for analysis (i.e., 1985–1990 and 2014–2019). A duration of 6 years was chosen as a balance between sufficient separation between the two time periods and adequate sample size for each time period. For these two periods, five GLMs of varying complexity were analyzed, with annual hypoxic volume serving as the response variable (y) and different combinations of both a continuous independent variable of freshwater discharge (x_1) and a categorical factor of time period (x_2) serving as predictors. An information-theoretic approach was used to select the best GLM from this set of five using corrected AIC (Akaike Information Criterion) statistics and Likelihood Ratio Chi-Square tests (Anderson, 2008; Burnham and Anderson, 2010; Seitz et al., 2009; Table S1). The additive GLM predicting hypoxic volume from discharge and time period without interaction emerged as the best GLM (Eq. (1)):

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 \quad (1)$$

where β_0 = intercept and β_i = parameters for the corresponding variables. This statistical analysis was repeated for the two different classifications of hypoxia (i.e., $O_2 < 3$ mg L^{-1} and $O_2 < 1$ mg L^{-1}). The probabilities that the additive GLM was the best among the set were 0.93 and 0.85 for $O_2 < 3$ mg L^{-1} and $O_2 < 1$ mg L^{-1}, respectively (Table S1). The percent residual deviances explained by the additive GLM were low, accounting for 21.5% and 19.1% of the null deviance for $O_2 < 3$ mg L^{-1} and $O_2 < 1$ mg L^{-1}, respectively (Table S1). This statistical analysis was done using R (R Core Team, 2020).

2.3. Numerical model description

2.3.1. Estuarine model

Numerical model simulations were generated using a fully coupled hydrodynamic-biogeochemical estuarine model developed specifically for the Chesapeake Bay (ChesROMS-ECB: Da et al., 2018; Feng et al., 2015; Irby et al., 2018; St-Laurent et al., 2020, Bever et al., 2021). The hydrodynamic model is an application of the Regional Ocean Modeling System (ROMS; Shchepetkin and McWilliams, 2005) to the Bay (Fig. 1), with a curvilinear grid that has an average cell resolution of 1.7 km within the Bay (Xu et al., 2012) and 20 terrain-following vertical levels that have higher resolutions near the surface and bottom of the water column. The biogeochemical model represents full carbon and nitrogen cycles through multiple state variables including nitrate (NO_3^-), ammonium (NH_4^+), O_2, chlorophyll-a, phytoplankton, zooplankton, inorganic suspended solids (ISS), small and large detrital nitrogen and carbon, and semi-labile and refractory dissolved organic nitrogen and carbon (Feng et al., 2015), as well

as dissolved inorganic carbon and total alkalinity (St-Laurent et al., 2020). For simplicity, these variables were classified into specific groups that are referred to later in the text: dissolved inorganic nitrogen (DIN) includes NO_3^- and NH_4^+; particulate organic nitrogen (PON) includes phytoplankton, zooplankton, and small and large detrital nitrogen; dissolved organic nitrogen (DON) includes semi-labile and refractory pools; organic nitrogen (ON) represents the sum of PON and DON; and total nitrogen (TN) includes all nitrogen state variables (i.e., the sum of ON and DIN).

The numerical model equations are identical to those listed in St-Laurent et al. (2020), including those pertaining to the interaction between the sediment bed and the water column. Specifically, upon reaching the sea floor, the organic matter flux sinking from the water column is split into three fractions: (1) a fraction is resuspended and respired in the water column with this fraction being determined by the local hydrodynamical stress, (2) a fraction is permanently buried in the sediments, and (3) the remainder is assumed to be instantly respired by the sediments. This assumption of instantaneous respiration is an important limitation of the model as it does not allow for delayed respiration associated with the accumulation of organic matter in the sediments from previous phytoplankton blooms. Parameter values are also the same as in St-Laurent et al. (2020), with the exception of two minor modifications: (1) the remineralization rates of DON and dissolved organic carbon (DOC) at 0 °C were decreased from 0.00765 d^{-1} and 0.012 d^{-1}, respectively, to 0.0028 d^{-1} in order to improve modeled DON skill, and (2) the Jerlov water type parameter in the numerical model was increased from type 3 to 5, as in Hinson et al. (2021), to improve the accuracy of modeled temperatures in the mainstem Bay. This Jerlov water type parameter impacts how light is attenuated with depth, with higher numbers indicating greater attenuation.

2.3.2. Terrestrial inputs

All simulations were run with terrestrial inputs (i.e., freshwater, nutrients, and ISS) derived from a combination of USGS gauge data from all nine rivers represented in the estuarine model and two separate watershed models: the CBP's Phase 6 watershed model (CBP, 2017) and the Dynamic Land Ecosystem Model (DLEM; Pan et al., 2021; Tian et al., 2015; Yang et al., 2015a, 2015b; Yao et al., 2021). The use of two independent watershed models helps to quantify the uncertainty associated with estimating nutrient loading to the Bay by providing a range of possible values instead of a single value from one watershed model. Because the outputs from Phase 6 and DLEM do not yet extend past 2015, the freshwater discharge for 2015–2019 was derived from USGS gauge data (USGS, 2020) and upscaled for each river to include inputs from the portions of the watershed downstream of the gauges. As in Bever et al. (2021), daily gauge flow was regressed against daily DLEM flow that included watershed areas below the gauge for the period where the two time series overlap in 2014 and 2015. The resulting linear relationship was employed to convert 2016–2019 gauge data to estimated total watershed inputs for all simulations.

Concentrations of terrestrial nutrients and ISS in 2015–2019 were estimated using relationships between the different biogeochemical constituents and discharge ($m^3\ s^{-1}$) in the watershed models from 2010 to 2014 (Fig. S2). As a result, this method assumes that the biogeochemical conditions in the watershed from 2010 to 2014 were similar to those in 2015–2019. To develop these relationships between concentration and discharge, 2010–2014 concentrations were log-transformed for each river and binned by their corresponding log-transformed discharges every 0.25 log (discharge) value from 0 to 4. Within each of these bins, median log-transformed concentrations were calculated and used as the concentration estimates for the central discharge values in the bins. To make these concentration estimates continuous across all discharges, the values between these central discharges were obtained through interpolation using a piecewise cubic Hermite interpolating polynomial. For values outside of the central discharges for the highest and lowest bins with data, the median concentrations of these bins were used to estimate concentrations rather than extrapolating this relationship outside of the range of data. The resulting relationships were very similar to those obtained when a Locally Estimated Scatterplot Smoothing (LOESS) curve was fit to the log-transformed values. When multiplied by discharge, these concentrations for Phase 6 and DLEM yielded average annual loadings (mean ± standard deviation) of 145 ± 61 Gg yr^{-1} and 104 ± 36 Gg yr^{-1}, respectively, for inorganic nitrogen, and 40 ± 30 Gg yr^{-1} and 59 ± 29 Gg yr^{-1} for organic nitrogen over the 2016–2019 period.

To fully link the terrestrial and estuarine systems, variables from the watershed models must be partitioned into specific variables represented in the estuarine model. For Phase 6, the partitioning was done in the same manner as in Irby and Friedrichs (2019), where semi-labile DON concentrations in ChesROMS-ECB riverine inputs were estimated as 100% of the biological oxygen demand of organic nitrogen and 80% of phytoplankton nitrogen; refractory DON concentrations were estimated as 20% of the total watershed refractory DON; and small detrital nitrogen concentrations were estimated as 20% of phytoplankton nitrogen and 80% of the total watershed refractory DON. Since DLEM only provides estimates of total DON and ChesROMS-ECB requires estimates of the semi-labile and refractory components of DON, the relationship between discharge and the ratio of refractory DON to total DON from the Phase 6 model was used to partition total DON in DLEM into refractory and semi-labile pools. The resultant riverine inputs were provided to ChesROMS-ECB at nine river locations (Figs. 1 & S3).

2.3.3. Model forcing

ChesROMS-ECB was forced at the surface using estimates of air temperature, downwelling longwave radiation, net shortwave radiation, air pressure, precipitation, and relative humidity from a 0.25° resolution atmospheric reanalysis (ERA5) produced by the European Centre for Medium-Range Weather Forecasts (C3S, 2017; Fig. S3). Rather than using the wind from ERA5, a higher resolution wind product from the North American Mesoscale Forecast System (NAM) was used to force the numerical model (National Centers for Environmental Information, 2020), as ChesROMS-ECB has greater skill with respect to hypoxia using NAM winds compared to ERA5 (authors unpublished data). The NAM wind product (12 km resolution prior to 2018 and 3 km resolution thereafter) provides estimates of wind speed and direction that more effectively capture the position of gradients in wind speed over the Bay and along the coastline and has been successfully used for several years in the operational forecast version of ChesROMS-ECB (Bever et al., 2021).

The open ocean boundary was forced with monthly climatological averages of temperature and salinity computed from in-situ data in the World Ocean Database from 2008 to 2018 (Boyer et al., 2018; Fig. S3). These monthly averages were assumed to be representative of the central year over this interval (2013), and information for subsequent years (2014–2019) was computed by projecting this climatology forward using a long-term trend calculated from the same dataset over a longer period of time (Da et al., 2021). The long-term trends were computed individually for each month using a linear regression of depth-averaged values (0–35 m) along the boundary spanning the years 1985–2018 and the latitudinal interval from 36° to 37.8°N (Da et al., 2021; Hinson et al., 2021). Sea surface height forcing was derived using tidal harmonics from the Advanced Circulation (ADCIRC) model (Luettich et al., 1992) and hourly nontidal water levels from observation stations in Duck, NC and Lewes, DE (Da et al., 2018).

2.4. Numerical model skill assessment

A realistic numerical model hindcast (termed the "*Realistic*" scenario) was performed using the models and forcing described above for 2016–2019 (Table 1). Numerical model skill was assessed quantitatively by comparing available data from CBP's WQMP (Section 2.1) with hourly output from the model that was closest in space and time to each observation. This comparison was performed by computing the means and standard deviations of observations and model output individually, as well as the bias, root-mean squared difference (RMSD), and correlation between them. Specifically, observations of temperature, salinity, O_2, NO_3^-, NH_4^+, and DON, grouped into above and below 10 m depths at 56 stations

Table 1
Numerical model experiments with changes from the *Realistic* simulation highlighted in bold.

Simulation	Watershed DIN input	Watershed ON input	Temp.	Average difference from *Realistic* (km^3 days)[a]	
				(O_2 < 3 mg L^{-1})	(O_2 < 1 mg L^{-1})
Realistic	2016–2019	2016–2019	2016–2019	0	0
Expt. 1: ***1985 TN***	**1985**	**1985**	2016–2019	+575 (±169)	+360 (±95)
Expt. 2: ***1985 TN (High)***	**1985 + 0.3(%Δ)**	**1985 + 0.3(%Δ)**	2016–2019	+720 (±207)	+462 (±121)
Expt. 3: ***1985 TN (Low)***	**1985–0.3(%Δ)**	**1985–0.3(%Δ)**	2016–2019	+417 (±125)	+256 (±68)
Expt. 4: ***1985 DIN***	**1985**	2016–2019	2016–2019	+327 (±72)	+201 (±27)
Expt. 5: ***1985 ON***	2016–2019	**1985**	2016–2019	+277 (±183)	+169 (±112)
Expt. 6: ***1985 Temp***	2016–2019	2016–2019	**1985**	−101 (±16)	−45 (±16)

[a] Average difference between the sensitivity experiments and the "Realistic" scenario averaged across all four years and between simulations forced with the two different terrestrial inputs. Standard deviations are shown in parentheses.

(Fig. 1) from 2016 to 2019 were used in the comparison. These stations were chosen as they are the same stations used to derive interpolated estimates of hypoxic volume from the water quality monitoring cruises. In addition to being summarized in tables (Tables S2–S3), these metrics are displayed graphically through Target diagrams which display skill by representing standard-deviation normalized RMSD as the distance from the origin on a plot where normalized bias and normalized unbiased RMSD are on the x- and y-axes, respectively (Hofmann et al., 2008; Jolliff et al., 2009). The resulting figures from this analysis are described in Section 3.2.

Model-data comparison plots were also employed as a qualitative skill assessment to provide a more intuitive and descriptive representation of numerical model accuracy and to highlight differences in performance both spatially and temporally. Spatial performance was investigated by comparing seasonal averages of model output with observations from monitoring stations throughout the Bay. Temporal skill was examined through time series of hourly model output and available observations at key stations spanning the extent of hypoxia in the Bay. Additionally, skill with respect to hypoxic volume was investigated by comparing the time series of daily hypoxic volume from the numerical model with estimates of hypoxic volume based on the interpolation of observed O_2 from CBP's WQMP. Through this combination of both quantitative and qualitative skill assessments, the strengths and limitations of the model were identified, and a comprehensive assessment of model performance was achieved.

2.5. Sensitivity experiments

In addition to the *Realistic* scenario described above, six sensitivity experiments were conducted over the same period (2016–2019) to determine the impact that changing conditions had on the volume of hypoxic water (Table 1). All simulations included a one-year spin-up with a realistic initial condition beginning in January 2015 to ensure that the results from the first year of simulation did not include transient effects. To investigate the effect that nitrogen reductions had on hypoxia, the concentrations of specific nitrogen constituents were altered in the nutrient sensitivity experiments to reflect 1985 values while keeping everything else constant. In addition, the impact of temperature change on hypoxia was investigated by altering atmospheric inputs from the *Realistic* scenario to reflect 1985 conditions. The impact of each of these individual drivers was examined by comparing numerical model output from the sensitivity experiments with results from the *Realistic* scenario.

2.5.1. Experiment 1: 1985 TN

In the *1985 TN* experiment (Table 1) riverine nutrient concentrations in 2016–2019 were set to concentrations representative of 1985. Specifically, a percentage change in nitrogen loading from 1985 to 2019 was calculated for each of the nine river basins entering ChesROMS-ECB (Fig. 1). The load within each of these basins was subdivided into two separate components: the portion of the load from the watershed above the River Input Monitoring (RIM) stations (above RIM) and the portion of the load from the watershed below the RIM stations (below RIM). This was done because the above RIM loads could be estimated using the available data at the RIM stations, whereas the below RIM loads could only be estimated using watershed models. For the river basins that do not contain a RIM station (Nanticoke and Elk), the total loading was considered to be below RIM. For the river basins that contain multiple RIM stations in different tributaries (James and York; Fig. 1), the above RIM loading was represented by the sum of these stations (the Appomattox and James RIM stations for the James basin and the Pamunkey and Mattaponi RIM stations for the York basin).

For the above RIM portion of the watershed, TN and NO_3^- results from a "stationary" implementation of the Weighted Regression on Time, Discharge, and Season (WRTDS) were used to estimate the change in loading over the 35-year period. The stationary WRTDS model is similar to the standard WRTDS model (Hirsch, 2014; Hirsch et al., 2010), except that it allows a temporally invariant regression surface to be applied for concentration predictions over the entire time period (Zhang et al., 2016). For this specific case, the temporally invariant regression surfaces were reflective of specific years. To examine the change in loading from 1985 to 2019, WRTDS was run twice over this 35-year period: once with stationary 1985 nutrient concentrations and once with stationary 2019 nutrient concentrations (Q. Zhang, CBP, written comm., January 19, 2021). The resulting difference in loading between these two stationary models reflects the change in loads over the 35-year period under the discharge and seasonal conditions of a given year (Zhang et al., 2016). Since the numerical model simulations span the four-year period from 2016 to 2019, each with different discharges, a unique change in loading was obtained for each of the four years in that interval using this approach.

These changes in loading were applied to the sensitivity experiments by first distributing them among the various ChesROMS-ECB nitrogen state variables (NO_3^-, NH_4^+, semi-labile DON, refractory DON, and small detrital nitrogen), as described in Appendix B. For NO_3^-, the change was calculated and applied directly since there is a NO_3^- state variable in the numerical model. For other nitrogen constituents, the values for 1985 and 2019 had to be estimated from TN loading by first calculating the load of non-nitrate nitrogen as the remaining pool of TN once NO_3^- is subtracted out. This methodology is consistent with previous approaches to estimate other nitrogen constituents from WRTDS output (Zhang et al., 2013; Zhang et al., 2015). This non-nitrate nitrogen loading was then subdivided into the individual constituents using the relative proportions of nonpoint source loading from the Phase 6 watershed model and subsequently used to calculate the change over the 35-year period.

The change in below RIM nitrogen loading from 1985 to 2019 was estimated using the Chesapeake Assessment and Scenario Tool (CAST), a web-based application of the CBP's Phase 6 watershed model (CBP, 2021). CAST provides annual loads of nonpoint source TN and point source NO_3^-, NH_4^+,

and organic nitrogen for the below RIM portions of the watershed in 1985 and 2019 based on average hydrologic conditions. As with the above RIM loading, these values needed to be distributed among the ChesROMS-ECB state variables (Appendix C). For point source DIN, changes in below RIM NO_3^- and NH_4^+ loading were applied directly to their respective nitrogen state variables. Point source organic nitrogen loading, in contrast, had to be subdivided into individual constituents using the relative proportions of point source loading from the Phase 6 dynamic watershed model before it could be applied to the appropriate nitrogen state variables. Nonpoint source TN loading also had to be subdivided into individual constituents using relative proportions of nonpoint source loading before it could be applied to the variables in ChesROMS-ECB.

The total change in loading between 1985 and 2019 ($\Delta(\text{Total})_{y,\,b,\,s}$, typically a negative number) was calculated from the above and below RIM changes in loading estimated for each of the nitrogen constituents (Eq. (2); Fig. 2). The subscript *y* represents each of the individual years from 2016 to 2019 (since there is a unique 35-year change in above RIM loading under the different hydrologic conditions for these years), the subscript *b* represents each of the nine river basins, and the subscript *s* represents the different nitrogen species. For PON, the total change in loading is calculated from small detrital nitrogen only as it is assumed that there are no long-term changes in phytoplankton, zooplankton, and large detrital nitrogen in the river inputs to the numerical model.

$$\Delta(\text{Total})_{y,b,s} = \Delta(\text{Above RIM})_{y,b,s} + \Delta(\text{Below RIM})_{b,s} \tag{2}$$

From this total change, a percentage change in loading relative to 2019 was estimated over the 35-year period (Eq. (3)) and used to create

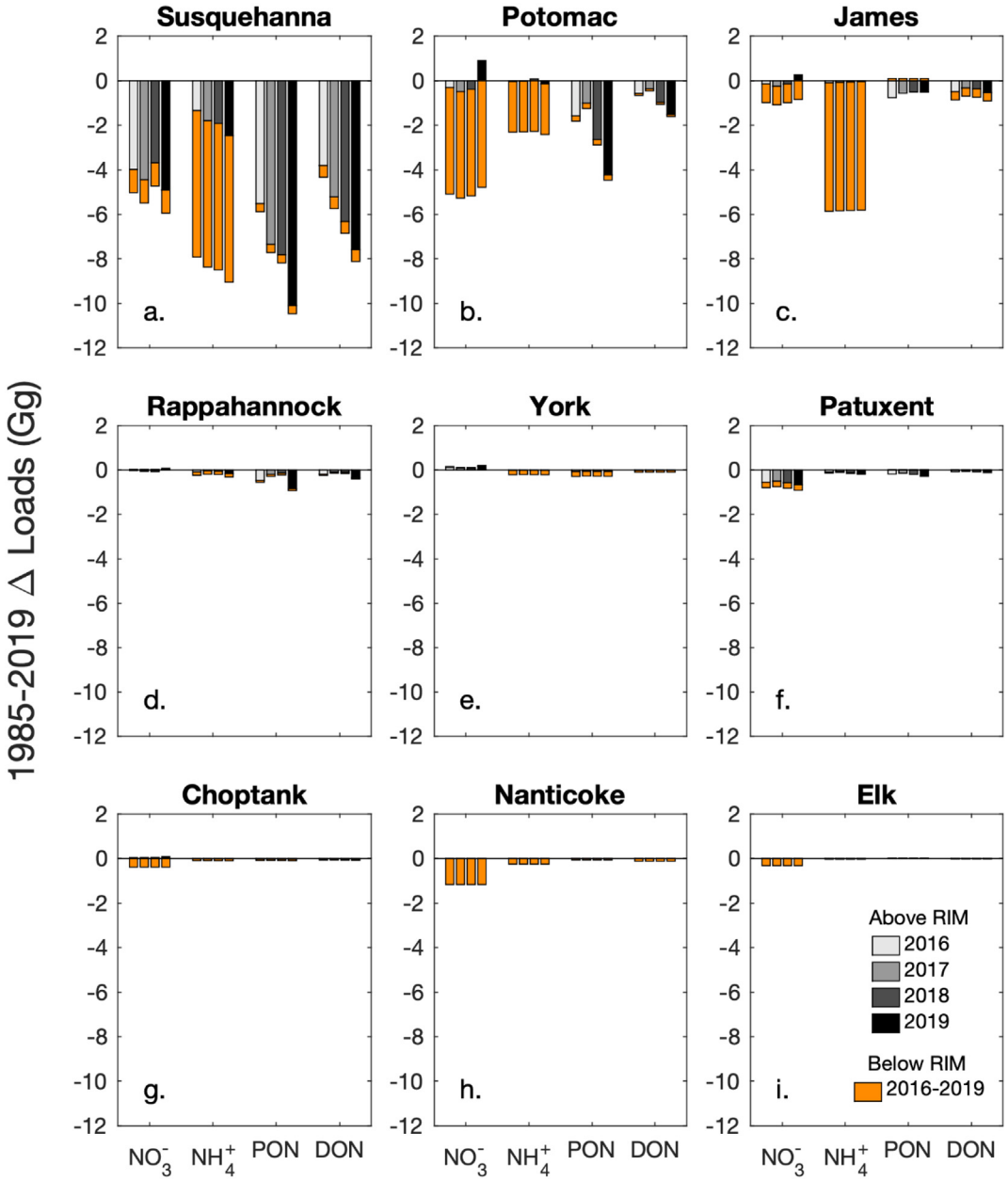


Fig. 2. Changes in loading from 1985 to 2019 for the different riverine nitrogen constituents. The grayscale bars represent the above RIM changes in loading, with unique values for 2016–2019 based on the different hydrologic conditions during those years. Orange bars represent the below RIM changes in loading (point source and nonpoint source), with constant values for 2016–2019 for each river-constituent combination.

multipliers (Eq. (4); Table S4) to adjust the riverine nutrient concentrations in the *Realistic* scenario to reflect concentrations in 1985 (Eq. (5)).

$$\%\Delta(\text{Total})_{y,b,s} = \frac{-\Delta(\text{Total})_{y,b,s}}{(\text{Above RIM})_{y,b,s}^{2019} + (\text{Below RIM})_{b,s}^{2019}} \times 100 \quad (3)$$

$$(\text{Multiplier})_{y,b,s} = \left(1 + \frac{\%\Delta(\text{Total})_{y,b,s}}{100}\right) \quad (4)$$

$$[1985\ \text{Levels}]_{y,b,s} = [\text{Realistic}]_{y,b,s} \times (\text{Multiplier})_{y,b,s} \quad (5)$$

Recreating concentrations using this delta approach is preferable to using realistic concentrations from the 1980s, as those past concentrations are strongly influenced by the meteorological conditions of that specific period and would produce inaccurate results if paired with the conditions over the simulation period. When multiplied by the 2016–2019 discharge values, these 1985 concentrations for Phase 6 (DLEM) yield average annual loadings (mean ± standard deviation) of 195 ± 73 Gg yr^{-1} (139 ± 46 Gg yr^{-1}) for inorganic nitrogen and 56 ± 36 Gg yr^{-1} (108 ± 38 Gg yr^{-1}) for organic nitrogen. In addition to the common metrics of daily hypoxic volume and annual hypoxic volume, a measure of "excess hypoxia" will also be investigated using this sensitivity experiment by showing the number of days at a particular location where hypoxia exists in the *1985 TN* simulation but not in the *Realistic* scenario.

2.5.2. Experiments 2 and 3: 1985 TN (High) and 1985 TN (Low)

To determine the sensitivity of the numerical model to the applied changes in nitrogen concentration, additional experiments were performed with larger and smaller percent changes in nutrient loading (Table 1). For *1985 TN* (*High*) and *1985 TN* (*Low*), the percent changes in loading from the *1985 TN* experiment were increased by 30% and decreased by 30%, respectively. For example, if the percent change in loading for a specific nitrogen constituent in a river was 50%, the percentage applied to *1985 TN* (*High*) would be 65% and the percentage applied to *1985 TN* (*Low*) would be 35%. This threshold of 30% was selected as it is similar in magnitude to the interannual variability of the multipliers used to obtain 1985-era loading for the different nitrogen species in the Susquehanna River (Table S4), which contributes the most to total nutrient loading to the Bay. For consistency, this same 30% adjustment was applied to all rivers and nitrogen species, such that the largest absolute changes in the sensitivity experiments were applied to the rivers and nitrogen species with the greatest percent changes in loading.

2.5.3. Experiments 4 and 5: 1985 DIN and 1985 ON

The changes to the riverine inputs for the *1985 DIN* and *1985 ON* sensitivity experiments were calculated using the same approach as for the *1985 TN* experiment (Section 2.5.1) except the concentrations of DIN and ON were altered individually for each experiment (Table 1). For *1985 DIN*, the percent changes in loading for NO_3^- and NH_4^+ were calculated by first estimating the above RIM and the below RIM changes in loading for the two constituents. These values were then used to calculate a total change (Eq. (2)) and a percentage change (Eq. (3)) in loading to adjust the concentrations of NO_3^- and NH_4^+ to reflect 1985 levels (Eqs. (4)–(5)). For *1985 ON*, the percent changes in loading for semi-labile DON, refractory DON, and small detrital nitrogen were calculated in the same manner by first estimating the above RIM and the below RIM changes in loading. As with the other sensitivity experiments, these above and below RIM changes were combined to estimate a total change (Eq. (2)) and a percentage change (Eq. (3)) in loading from which the concentrations of semi-labile DON, refractory DON, and small detrital nitrogen were altered to 1985 values (Eqs. (4)–(5)).

2.5.4. Experiment 6: 1985 Temp

A final sensitivity experiment was conducted (*1985 Temp*; Table 1) in which temperatures were altered to reflect 1985 conditions when the Bay was about 1.4 °C cooler in the summer (Hinson et al., 2021). As in Hinson et al. (2021), long term trends in air temperature and downwelling longwave radiation from 1985 to 2019 were calculated and used to adjust the 2016–2019 atmospheric conditions in the *Realistic* scenario. These trends were calculated individually for each month at all grid cells in the same ERA5 atmospheric reanalysis used to force the numerical model (Section 2.3.3). These monthly trends were then interpolated using a spline function in order to produce a smooth continuous time series that could be applied to the atmospheric forcing at each 3-h timestep. This analysis ignored other mechanisms that are contributing to long-term changes in Bay temperatures (e.g., changing ocean temperatures, changing river temperatures, sea level rise) as air temperature and downwelling longwave radiation are the dominant drivers of water temperature change in the upper- and mid-portion of the Bay where hypoxia is most severe (Hinson et al., 2021).

3. Results

3.1. Empirical data analysis

Predicted hypoxic volume decreased substantially from the first time period (1985–1990) to the second (2014–2019) both for $O_2 < 3$ mg L^{-1}

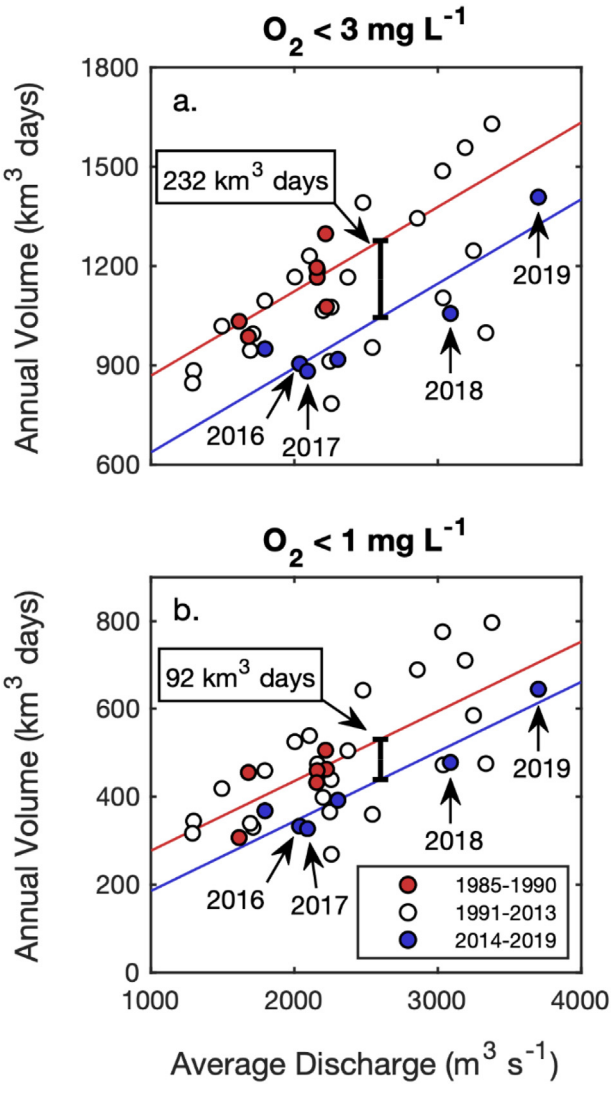


Fig. 3. GLM results for (a) $O_2 < 3$ mg L^{-1} and (b) $O_2 < 1$ mg L^{-1}. Red lines represent 1985–1990; blue lines represent 2014–2019. The vertical difference between these two lines represents the average 30-year decrease in hypoxic volume.

and $O_2 < 1$ mg L^{-1} (Fig. 3). The decrease was greater for the $O_2 < 3$ mg L^{-1} threshold (232 km^3 days) than for the $O_2 < 1$ mg L^{-1} (92 km^3 days). The magnitude of this difference was the same across all discharge conditions for each threshold, as indicated by the lack of an interaction effect between discharge and time period. However, the relative difference in predicted hypoxic volume decreased as average discharge increased, since the baseline level of hypoxia became larger but the difference between the two time periods was constant. Without the positive effect of time period, annual hypoxic volume for $O_2 < 3$ mg L^{-1} would have been 26% larger in 2016, 25% larger in 2017, 20% in 2018, and 18% in 2019. For $O_2 < 1$ mg L^{-1}, annual hypoxic volume would have been 26% larger in 2016 and 2017, 18% in 2018, and 15% in 2019 (Fig. 3).

To determine sensitivity of the results to the number of years included in each time period, the analysis was run for time period durations of 5–10 years. At 4 years and less, the low sample size per time period precluded sufficient statistical power to detect effects. For durations of 5–10 years, the additive GLM always emerged as the best among the set, similar to the results for the 6-year time period. The decrease in annual hypoxic volume was also similar for all durations, ranging from 135 to 232 km^3 days for $O_2 < 3$ mg L^{-1} and 69–101 km^3 days for $O_2 < 1$ mg L^{-1}. Thus, the decline in hypoxic volume related to time period was robust irrespective of time period duration.

3.2. Skill of the realistic hindcast

Numerical model hindcasts using terrestrial information from both DLEM and Phase 6 captured the spatiotemporal distribution of multiple variables within the Chesapeake Bay. For NO_3^-, the numerical model matched observations well for interannual variability in the along estuary concentration gradient, with higher concentrations extending farther down the Bay during wet years (Fig. 4c–d, k–l) compared to average discharge years (Fig. 4a–b, i–j). Modeled concentrations were generally lower than observed values from April–October in the mesohaline portion of the Bay (150–250 km from the mouth); this discrepancy was greater in

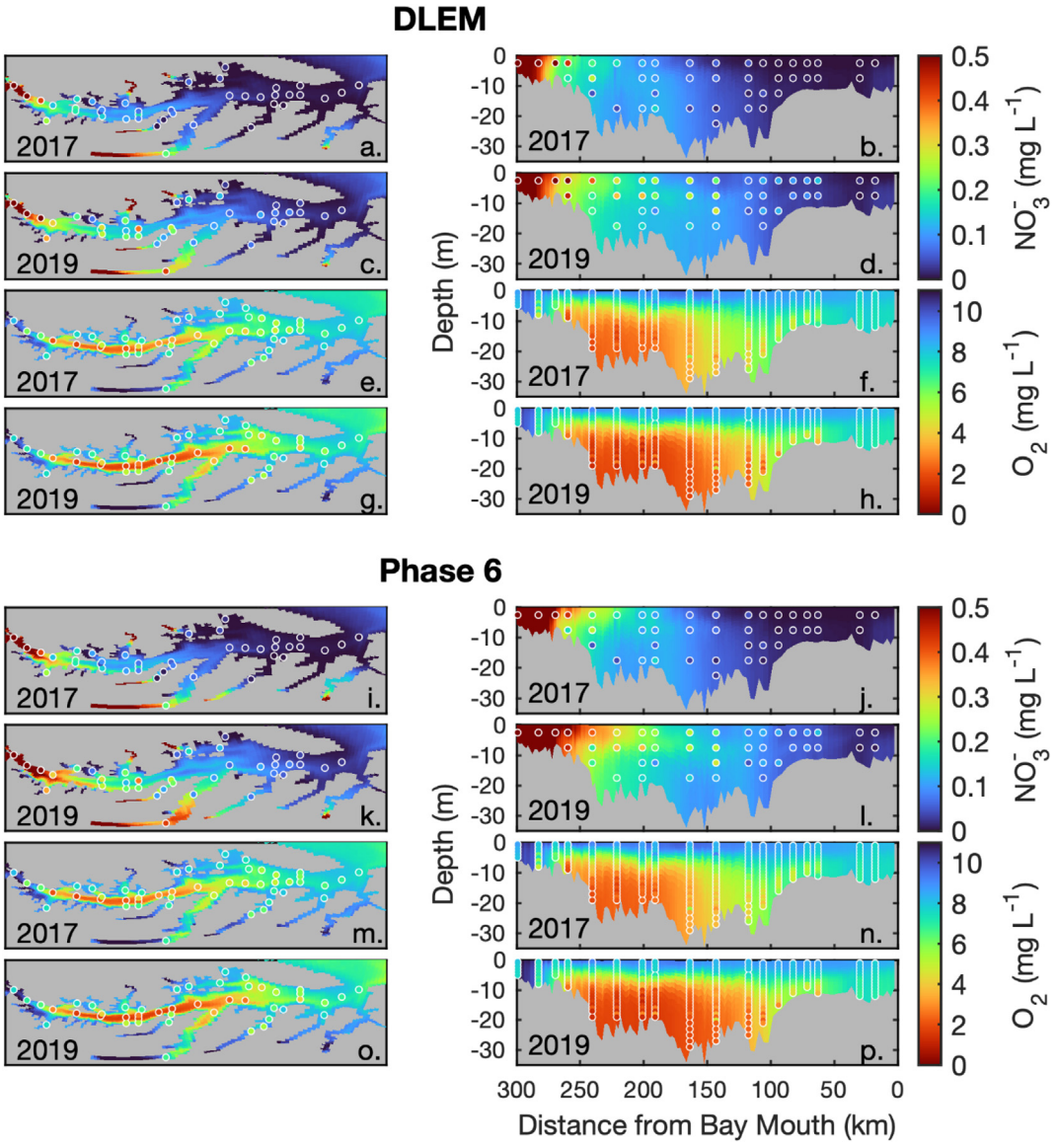


Fig. 4. Model-data comparison plots for NO_3^- (a–d, i–l) and O_2 (e–h, m–p). Values shown are April–October averages for the corresponding years, with numerical model output averaged only on days with observational data. Panels on the left (a, c, e, g, i, k, m, o) show maps of bottom concentrations while panels on the right (b, d, f, h, j, l, n, p) show depth transects throughout the mainstem Bay. DLEM (a–h) and Phase 6 (i–p) simulation results are shown. In all panels, the head of the Bay (north) is to the left and the mouth of the Bay (south) is to the right. NO_3^- observations are binned every 5 m due to limited data. Averages containing less than three observations are omitted as they do not represent conditions throughout the entire 7-month time period. Results are shown for an average discharge year (2017) and a wet year (2019).

DLEM simulations (Fig. 4a–d) compared to Phase 6 simulations (Fig. 4i–l). For O_2, the numerical model results matched the horizontal and vertical gradients in concentration and captured the greater O_2 drawdown in wet years (2019) compared to average discharge years (2017; Fig. 4e–h, m–p). The greatest discrepancy between modeled and observed O_2 was an overprediction of bottom concentrations in the southern mesohaline portion of the Bay around 60–120 km from the mouth (Fig. 4e–h, m–p). These higher modeled concentrations were apparent in time series of bottom O_2 at station CB6.3, though most stations in the mainstem of the Bay captured the seasonal cycle of oxygen well (Fig. S4). An explanation for this discrepancy is currently under investigation; however, it is possible that the model's lack of phosphate limitation causes too much production and nitrogen uptake in the upper Bay, resulting in too little nitrogen advecting down to the lower Bay where the resulting underestimate of organic matter causes too little bottom oxygen utilization.

The model demonstrated similar skill for temperature, salinity, and oxygen compared to the suite of Chesapeake Bay numerical models evaluated by Irby et al. (2016), as shown by the small normalized RMSD values relative to the standard deviation of the observations (Fig. S5). These variables had low biases when compared to observations, with modeled temperatures generally 0.4 °C warmer throughout the water column, modeled salinities similar at the surface and 0.6 higher in waters below 10 m, and modeled O_2 concentrations 0.2 mg L^{-1} higher in the surface and 0.3–0.4 mg L^{-1} higher at depths below 10 m (Tables S2–S3). Unlike oxygen, there was no spatial pattern in skill for salinity and temperature, with both along estuary and vertical gradients effectively captured in the numerical model.

Numerical model skill was lower and varied more between the DLEM and Phase 6 simulations for nitrogen constituents (e.g., NO_3^-, NH_4^+, and DON). For NO_3^-, the numerical model bias ranged from −0.03–0.05 mg L^{-1} (Tables S2–S3), with greater skill observed in the upper 10 m compared to depths below 10 m (Fig. S5). For NH_4^+, modeled concentrations were generally 0.05–0.11 mg L^{-1} larger than observed values (Tables S2–S3). Numerical model skill was better for DON, with biases ranging from −0.02–0.05 mg L^{-1} and smaller RMSD values compared to NH_4^+ (Fig. S5; Tables S2–S3). In the upper Bay, DON concentrations were generally more accurate for DLEM compared to Phase 6 as indicated by smaller biases at stations with latitudes greater than 39°N (Figs. S6–S8).

Numerical model estimates of hypoxic volume generally matched the seasonal and interannual changes in the data-based estimates of hypoxic volume obtained by spatially interpolating O_2 concentrations from the CBP's WQMP. For both $O_2 < 3$ mg L^{-1} and $O_2 < 1$ mg L^{-1}, the development, continuation, and breakup of hypoxia from April–October was accurately represented in the numerical model for each of the four years studied (Fig. 5a–d, g–j). Volumes were larger for the *Realistic* scenario using Phase 6 compared to that using DLEM, with DLEM generally showing better agreement with the interpolated estimates in the beginning of the hypoxia season (April–June) and Phase 6 showing better agreement at the end of the season (July–October). This match between interpolated estimates and hypoxic volumes from the numerical model was generally better for the wet years (2018 and 2019) than the average discharge years (2016 and 2017) evaluated.

3.3. 1985 TN sensitivity experiment results

The influence of nitrogen reductions on hypoxia was investigated by comparing hypoxic volume results from the *1985 TN* sensitivity experiment with those from the *Realistic* scenario (Fig. 5). Despite the interannual variability in hypoxic volume between wet and dry years, the average daily increase in hypoxic volume in the *1985 TN* simulation compared to the *Realistic* scenario during the hypoxia season (i.e., defined as being when hypoxia in *1985 TN* was greater than 0 km^3) was similar across all four years. This average increase ranged from 3.5–4.6 km^3 for DLEM and 1.7–3.1 km^3 for Phase 6 during the four years when considering a threshold of $O_2 < 3$ mg L^{-1} (Fig. 5a–d), and 2.7–2.8 km^3 (DLEM) and 1.4–2.0 km^3 (Phase 6) for $O_2 < 1$ mg L^{-1} (Fig. 5g–j). In terms of temporal differences between the two simulations, the hypoxia season was longer in *1985 TN* during the average discharge years of 2016 and 2017, beginning earlier in the year for both O_2 thresholds (Fig. 5a–b, g–h) and ending later in the year for $O_2 < 1$ mg L^{-1} (Fig. 5g–h). For the wet years of 2018 and 2019, the length of the hypoxia season was similar in the *Realistic* and *1985 TN* simulations (Fig. 5a–d, g–j).

The metric of annual hypoxic volume can be used to summarize the overall magnitude of hypoxia for a given year (Fig. 5e–f, k–l). The increase in annual hypoxic volume in *1985 TN* compared to *Realistic* is greater for DLEM simulations, for which annual volumes with $O_2 < 3$ mg L^{-1} increased by 120% in 2016, 97% in 2017, 54% in 2018, and 46% in 2019 (Fig. 5e) and annual volumes with $O_2 < 1$ mg L^{-1} increased by 288% in 2016, 179% in 2017, 108% in 2018, and 80% in 2019 (Fig. 5k; Table S5). These increases are smaller in the simulations using Phase 6, where annual volumes increased by 20–75% for $O_2 < 3$ mg L^{-1} (Fig. 5f) and by 33–165% for $O_2 < 1$ mg L^{-1} (Fig. 5l; Table S5). The error bars on annual hypoxic volumes for *1985 TN* represent results from the *TN* (*High*) and the *TN* (*Low*) simulations respectively. Even considering a range of ± 30%, the *1985 TN* annual hypoxic volumes for both DLEM and Phase 6 were higher than the *Realistic* (2016–2019) results.

Although the percent differences in annual hypoxic volume between the *1985 TN* and *Realistic* simulations varied from year to year, the absolute differences were more similar. For $O_2 < 3$ mg L^{-1}, the absolute differences ranged from 611 to 789 km^3 days for DLEM (Fig. 6a) and from 315 to 552 km^3 days for Phase 6 (Fig. 6b; Table S6). These differences between years were even more similar for $O_2 < 1$ mg L^{-1}, where they ranged from 433 to 465 km^3 days for DLEM (Fig. 6c) and 233–313 km^3 days for Phase 6 (Fig. 6d; Table S6). The magnitude of these differences in the DLEM simulations was nearly double of those in the Phase 6 simulations for both O_2 criteria. If the overall impact of nitrogen reductions is estimated as the average of all differences from both DLEM and Phase 6 (mean ± standard deviation), the effect is 575 ± 169 km^3 days (360 ± 95 km^3 days) for $O_2 < 3$ mg L^{-1} ($O_2 < 1$ mg L^{-1}; Table 1).

Despite the fact that the impact of nutrient reductions on hypoxia was evident to some extent throughout the entire portion of the Bay that regularly experiences hypoxia, the effect was concentrated in a few geographic areas. These areas are highlighted by the "excess hypoxia" metric that shows where hypoxia existed in the *1985 TN* simulation but not in the *Realistic* scenario (Fig. 7). The area that experienced the most excess hypoxia from 2016 to 2019 was the mesohaline portion of the Bay between the Patuxent and Rappahannock Rivers (~100–200 km from the Bay mouth), where there were generally 50–90 additional days of hypoxia for both $O_2 < 3$ mg L^{-1} and $O_2 < 1$ mg L^{-1}. The exact position of this region of excess hypoxia in the mesohaline portion of the Chesapeake Bay varied depending on the year and the O_2 threshold examined. For $O_2 < 3$ mg L^{-1}, this region was closer to the mouth of the Bay (Fig. 7a–h), with the wet years of 2018 and 2019 showing a further along-estuary extent compared to the average discharge years of 2016 and 2017. For $O_2 < 1$ mg L^{-1}, the region of excess hypoxia was further from the mouth of the Bay than $O_2 < 3$ mg L^{-1} but with a similar interannual pattern (Fig. 7i–p). An additional distinct region of excess hypoxia was present near the oligohaline portion of the Bay (~225–275 km from the mouth) at depths shallower than 10 m for $O_2 < 3$ mg L^{-1} (Fig. 7b, d, f, h). This region was most apparent in 2016 (Fig. 7a–b), where hypoxia would have existed for more than 80 additional days if nutrient reductions had not occurred. The number of excess hypoxia days in this region was progressively smaller in the three subsequent years from 2017 to 2019 (Fig. 7c–h). A similar upper Bay region was observed around 240 km from the Bay mouth for $O_2 < 1$ mg L^{-1} in 2016 and 2017 (Fig. 7i–l), however this area of excess hypoxia was observed at depths greater than 10 m.

In the deep channel where hypoxia typically persists for most of the summer (depths > 10 m, 150–250 km from the Bay mouth), there were a greater number of additional days with hypoxic conditions during the average discharge years of 2016 and 2017 compared to the wet years of 2018 and 2019. This distinction is most apparent for $O_2 < 3$ mg L^{-1}, where in

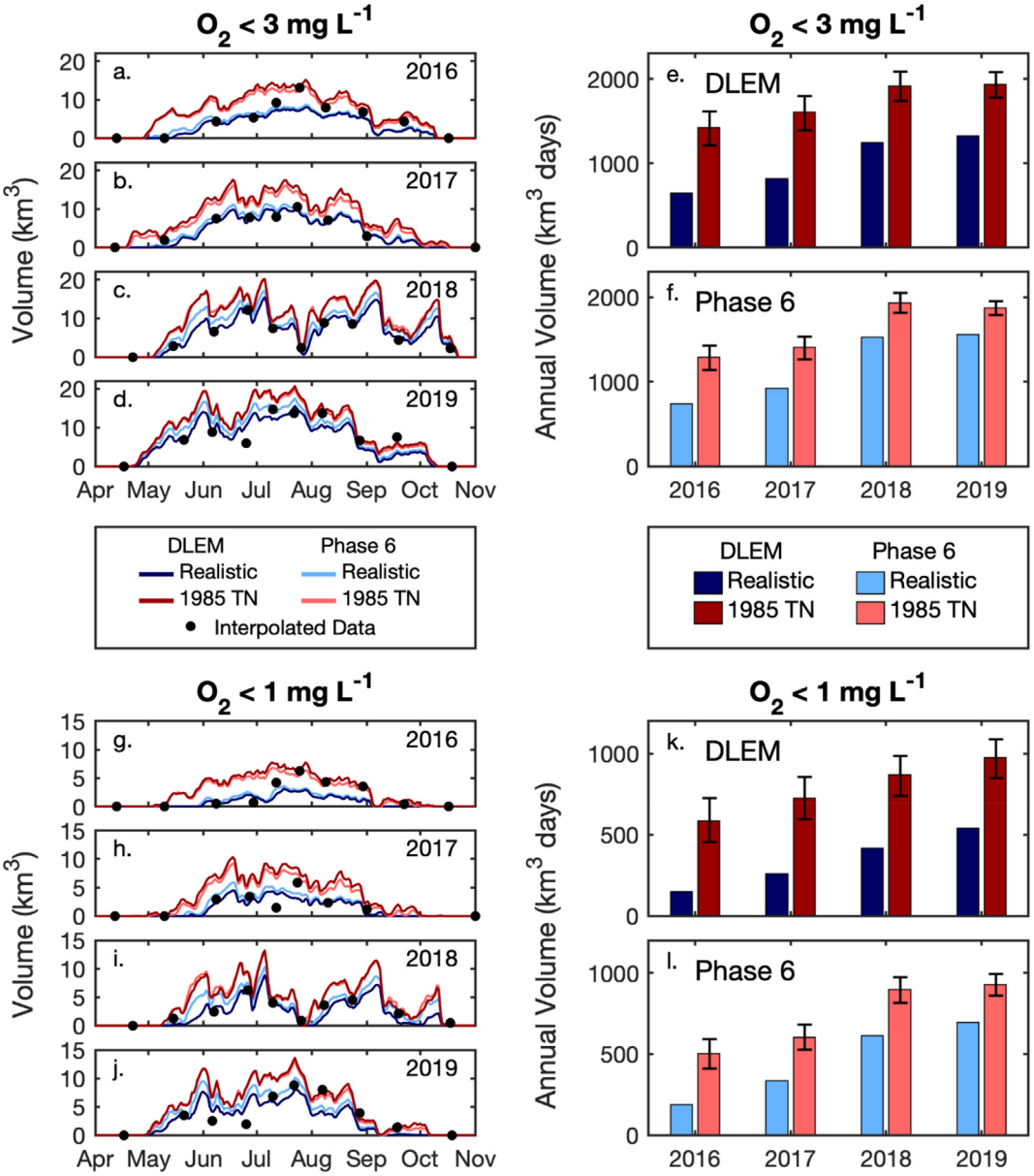


Fig. 5. Daily hypoxic volume results for the *Realistic* scenario and the *1985 TN* sensitivity experiment for $O_2 < 3$ mg L^{-1} (a–d) and $O_2 < 1$ mg L^{-1} (g–j), for each of the four years studied, with black circles representing interpolated hypoxic volume estimates. Total annual hypoxic volumes are displayed for both $O_2 < 3$ mg L^{-1} (e–f) and $O_2 < 1$ mg L^{-1} (k–l), with darker colors representing DLEM (e & k) and lighter colors representing Phase 6 (f & l). Error bars represent annual hypoxic volumes from the *1985 TN* (*High*) experiment (upper limit) and the *1985 TN* (*Low*) experiment (lower limit). Note that 2016–2017 are average discharge years and 2018–2019 are wet years.

some locations in the deep channel the number of days of additional hypoxia was less than 10 in 2018–2019 (Fig. 7e–h) but greater than 30 in 2016–2017 (Fig. 7a–d). The difference in the number of additional hypoxic days between average discharge years and wet years was still present for $O_2 < 1$ mg L^{-1}, however the difference in additional days was around 10 days compared to the greater than 20-day difference observed for $O_2 < 3$ mg L^{-1}. The same spatial patterns were observed in the simulation using Phase 6 (Fig. S9); however, the magnitude was lower with values ranging up to 70 days instead of the 90-day range observed in DLEM.

3.4. 1985 DIN, ON, and Temperature sensitivity experiment results

The relative contributions that DIN and ON reductions had on the total difference in hypoxia between the *1985 TN* and *Realistic* simulations differed depending on the terrestrial inputs that were used for the experiments. For the DLEM simulations, reductions in ON accounted for, on average, ~60% of the total change in hypoxia compared to ~40% for DIN reductions (Fig. 6a, c). The results were the opposite for Phase 6, where reductions in DIN accounted for ~70–80% of the total change in hypoxia and ON reductions accounted for the remaining ~20–30% (Fig. 6b, d). The influence of changing DIN and ON concentrations was almost additive, with the sum of the additional hypoxia resulting from increasing the two nutrient types individually in *1985 DIN* and *1985 ON* being almost equal to the additional hypoxia resulting from increasing them simultaneously (*1985 TN*).

As expected, altering 2016–2019 air temperatures and downwelling longwave radiation to reflect cooler conditions characteristic of 1985 decreased hypoxia (Fig. 6). The magnitude of this decrease was almost identical for the DLEM and Phase 6 simulations, however it varied interannually. The greatest impact occurred during the years with more hypoxia, with

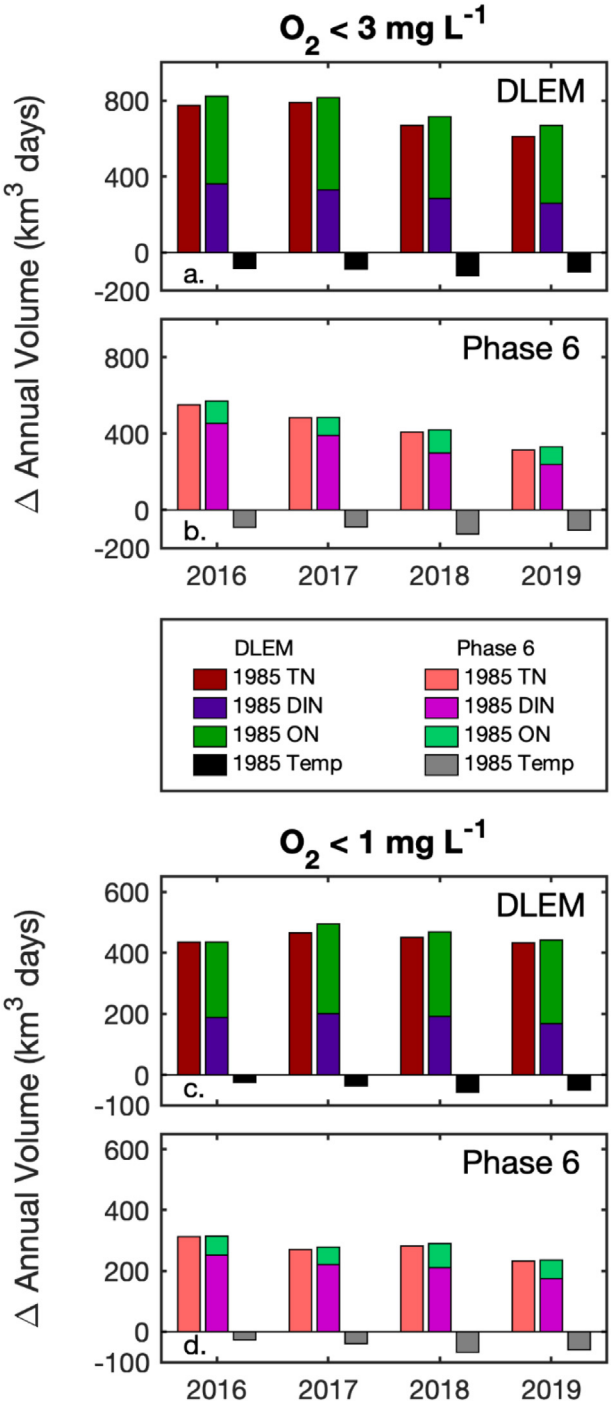


Fig. 6. Differences in annual hypoxic volume between the *Realistic* scenario and sensitivity experiments: *1985 TN*, *1985 DIN*, *1985 ON* and *1985 Temp*. Values are displayed for DLEM (a, c) and Phase 6 (b, d) for both $O_2 < 3$ mg L^{-1} (a, b) and $O_2 < 1$ mg L^{-1} (c, d).

1985 Temp having 122–126 km^3 days and 101–106 km^3 days less hypoxia in 2018 and 2019, respectively, compared to the *Realistic* scenario for $O_2 < 3$ mg L^{-1} (Fig. 6a–b; Table S6). For $O_2 < 1$ mg L^{-1}, hypoxia decreased by 57–68 km^3 days in 2018 and 50–60 km^3 days in 2019 in the *1985 Temp* simulation (Fig. 6c–d; Table S6). In contrast, during the average discharge years of 2016 and 2017, this decrease was substantially smaller: 85–92 km^3 days for $O_2 < 3$ mg L^{-1} and 24–39 km^3 days for $O_2 < 1$ mg L^{-1}.

When the magnitude of this temperature effect (*1985 Temp*) was compared to that of the TN effect (*1985 TN*), the relative impact was different for the simulations forced with the two types of terrestrial inputs. For Phase 6, the relative impact of temperature was greater, counteracting 17–34% of the TN effect when considering a threshold of $O_2 < 3$ mg L^{-1} (Fig. 6b) compared to only 11–18% for DLEM (Fig. 6a; Table S6). The same is true for $O_2 < 1$ mg L^{-1}, where temperature counteracts 9–26% of the TN effect for Phase 6 (Fig. 6d) and only 6–13% for DLEM (Fig. 6c; Table S6).

4. Discussion

4.1. Spatiotemporal impacts of nitrogen reductions vary by hydrological conditions

One advantage of using a 3-D numerical model to investigate the enhanced resilience of the Chesapeake Bay is that it can be used to identify what areas have benefitted the most from nutrient reductions. Hypoxia in the Chesapeake Bay is largely constrained by the depth of pycnocline and the bathymetry of the deep mainstem channel (Bever et al., 2018; Irby et al., 2016). Although the influence of these constraints varies depending on which O_2 threshold is examined, these vertical and horizontal boundaries generally cause the primary difference in annual hypoxic volume between years to be related to changes in the along estuary extent of low oxygen water. Results from sensitivity experiments showed that the greatest impact of nutrient reductions generally occurs at the southern end of where hypoxia develops in the Bay, somewhere between the Patuxent and Rappahannock Rivers (~100–200 km from the Bay mouth), irrespective of year and oxygen threshold (Fig. 7). This result agrees well with other studies that have found that hypoxia extends further south (closer to the Rappahannock) in the Bay during years with greater nitrogen loading (Murphy et al., 2011; Testa et al., 2018). Since hypoxia develops as a result of oxygen drawdown from the benthic decay of previous phytoplankton blooms (Officer et al., 1984), it can only form in regions where deep waters have a sufficient supply of sinking organic matter from the surface. As a result, hypoxic conditions are primarily found in the mesohaline Bay where surface phytoplankton concentrations are greatest (Zheng and DiGiacomo, 2020). Hypoxia generally does not form in the upper portions of the oligohaline Bay where the water column is well mixed and there is low phytoplankton biomass due to light limitation near the estuarine turbidity maximum (Harding et al., 1986). Therefore, when the total nutrient input to the Bay decreases, the distribution of phytoplankton in surface waters responds by changing how far it extends down-estuary (Harding et al., 2016b). The extent of hypoxia responds accordingly, and hence the greatest and most consistent impact of nutrient reductions is seen near the southernly extent of hypoxia in the Bay.

Another region that experienced a decrease in the number of days with hypoxic conditions as a result of nutrient reductions was located near the top of the pycnocline in the oligohaline portion of the Bay around 5–10 m, particularly during dry years (Fig. 7a–d, i–l). The curtailment of hypoxia at this location suggests that with less nutrient input, oxygen penetrated deeper in the water column as a result of less phytoplankton growth, especially during drier years when light limitation is less dominant in the upper Bay (surface ISS concentrations can be up to 80% lower for such years in the model). The effect of nutrient reductions became less prominent in this region during the higher discharge years of 2018 and 2019, when it is likely that light limitation rather than nutrients plays a larger role in controlling phytoplankton growth (Fisher et al., 1999; Fig. 7e–h, m–p). Future monitoring efforts aimed at capturing interannual changes in the extent of hypoxia due to nutrient reductions should focus on collecting higher resolution data near the northern and southern extents of hypoxia where O_2 concentrations are most sensitive to additional nutrient inputs.

Differences between wet and average discharge years were also apparent when the duration of hypoxia is examined in the context of nutrient reductions. Simulation results conducted here highlight that when terrestrial nutrient inputs were higher in the 1980s, the hypoxia season was generally longer during average discharge years with 10–20 additional days of

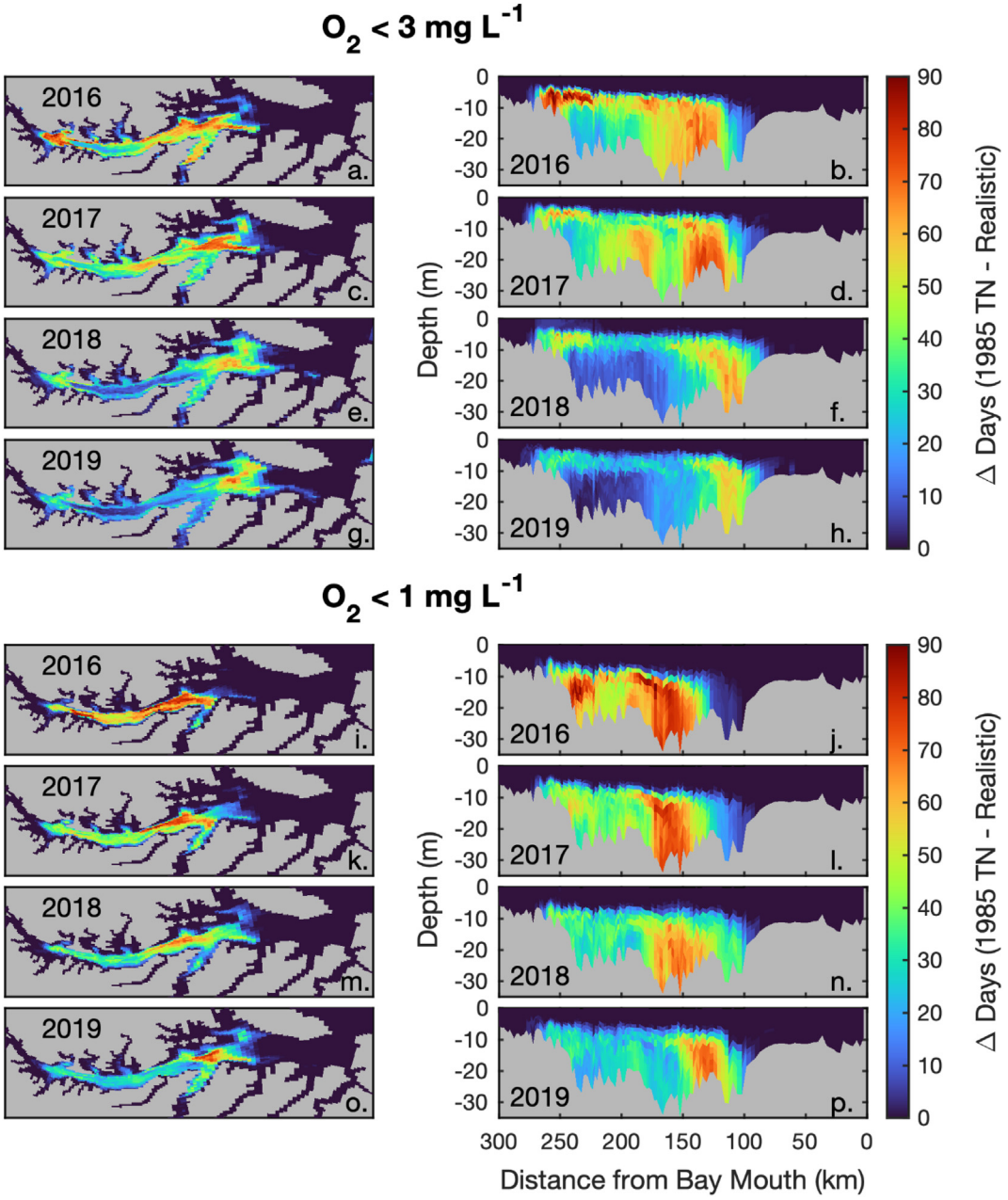


Fig. 7. The number of days of excess hypoxia (defined as where hypoxia exists in *1985 TN* but not in the *Realistic* scenario) for bottom grid cells (a, c, e, g, i, k, m, o) and cells along a mainstem transect (b, d, f, h, j, l, n, p). Results are displayed by year for both $O_2 < 3$ mg L^{-1} (a–h) and $O_2 < 1$ mg L^{-1} (i–p). In all panels, the head of the Bay (north) is to the left and the mouth of the Bay (south) is to the right. The values shown here are only for DLEM simulations; analogous results for Phase 6 are shown in Fig. S9.

$O_2 < 3$ mg L^{-1} and 41–63 additional days of $O_2 < 1$ mg L^{-1} in the Bay (Fig. 5a–b, g–h). During wet years, the duration of hypoxia was more similar between the 1980s and 2010s terrestrial nutrient inputs, with only 4–10 additional days of $O_2 < 3$ mg L^{-1} and 3–22 additional days of $O_2 < 1$ mg L^{-1} for simulations with 1980s inputs (Fig. 5c–d, i–j). Other studies have similarly found a decline in the duration of hypoxia in the Bay over the years from 1985 to 2010 and have primarily attributed this to an earlier termination in the fall rather than a later initiation in the spring (Testa et al., 2018; Zhou et al., 2014). In agreement with the results found here, this decline has been linked to a decrease in January–May nitrogen loading (Murphy et al., 2011), with lower nutrient input in the spring limiting organic matter production in summer and subsequently decreasing water column respiration and allowing oxygen concentrations to rise above hypoxic levels earlier in the fall (Testa et al., 2017). This mechanism explains the different responses in duration between average discharge years and wet years, with nutrient reductions having a smaller impact when the Bay is farther from this limitation threshold during wet years when there is ample nutrient input to support spring and summer phytoplankton growth. Results from the sensitivity experiments conducted here also suggest that this nutrient limitation effect may be occurring at the beginning of the hypoxia season as well, with nutrient reductions limiting phytoplankton growth and water column respiration early in the spring, allowing bottom oxygen levels to remain above hypoxic levels until later in the year as occurred in the summer of 2020 (Virginia Institute of Marine Science and Anchor QEA, LLC, 2020). Ni et al. (2020) found a similar effect by showing that nutrient reductions have delayed the onset of hypoxia by a few days, however, their study did not discuss how this effect varies between years with different riverine discharge.

4.2. Increased resilience estimated from statistical and numerical modeling analyses

Results from both statistical and numerical modeling analyses indicate that the Chesapeake Bay has become more resilient to adverse environmental conditions in recent years as a result of the nutrient reductions that have decreased hypoxia over the past 30–40 years. When the magnitude of the effect is compared between the two methods, the resilience estimates from the numerical modeling simulations are larger than those from the statistical analysis. This discrepancy, however, is to be expected as the two methods differ in a fundamental way. The statistical results do not directly link the observed interannual changes in hypoxic volume to changes in nutrient concentrations, since nutrient data are not included in the GLM. A link can only be presumed from the fact that since nutrient reductions have occurred in the watershed since the 1980s (Ator et al., 2020; Murphy et al., 2011; Testa et al., 2008), the same freshwater discharge in the 2010s is expected to have a lower nutrient load than in the 1980s. However, other factors also influence the observed change in resilience between the two time periods, including long-term changes in temperature (Hinson et al., 2021; Ni et al., 2020), stratification (Du et al., 2018; Murphy et al., 2011), and prevailing wind direction (Scully, 2010a, 2010b), which have all been shown to impact hypoxia on decadal timescales.

Unlike the statistical analysis which does not account for factors other than freshwater discharge, the numerical modeling approach directly links nutrients and resilience by isolating the impact of nutrient load reductions over the past 35 years. By solely changing nutrient concentrations, this mechanistic approach separates the impact of management efforts from long-term changes in environmental conditions. As a result, if a long-term change in environmental conditions exists over the 35-year time period, the magnitude of the resilience from the numerical modeling analysis would be expected to be different from the magnitude of the resilience from the GLM analysis. The role of environmental conditions is examined in more detail in the next section.

4.3. Warming temperatures partially counteract improvements from nitrogen reductions

Water temperatures in the Chesapeake Bay have increased by roughly 1.4 °C in the summer over the past 35 years, driven primarily by warming atmospheric temperatures and increased downwelling longwave radiation (Hinson et al., 2021). In terms of climate change impacts on hypoxia, warming estuarine temperatures are of principal concern due to their high probability of occurrence in the future (Muhling et al., 2018) and their ability to promote hypoxia through enhanced stratification, decreased oxygen solubility at the surface, and an increased rate of microbial respiration at depth (Altieri and Gedan, 2015; Irby et al., 2018). Since stratification in the Chesapeake Bay is driven primarily by vertical gradients in salinity rather than temperature (Murphy et al., 2011), and since surface and bottom waters in the Bay have warmed at similar rates in the past (Hinson et al., 2021), decreased solubility and increased respiration are the two primary mechanisms by which warming exacerbates hypoxia in the Bay. Not only are warmer temperatures in the mid-21st century expected to have a large impact on hypoxia in the Bay (Irby et al., 2018), but the results presented here also reveal that the impact of warming has already begun. The temperature sensitivity experiment shows that hypoxic volumes in 2016–2019 would have been 7–13% (9–16%) smaller for $O_2 < 3$ mg L^{-1} ($O_2 < 1$ mg L^{-1}) if temperatures had not increased from the levels that they were in 1985. Although these percentages may seem small, this enhancement due to warming has already offset roughly 11–34% (6–26%) of the improvements in hypoxia from nutrient reductions for $O_2 < 3$ mg L^{-1} ($O_2 < 1$ mg L^{-1}; Fig. 6).

Even though the improvements in hypoxic volume resulting from nutrient reductions would have been larger without atmospheric warming, nutrient reductions have still managed to outpace the negative effects of temperature and increase bottom oxygen concentrations in the Chesapeake Bay. The retrospective analysis in this study is consistent with projections of future change in the Bay, which show that hypoxic conditions will continue to improve despite the negative impacts of climate change, as long as the nutrient reduction targets set in the TMDL are met (Irby et al., 2018). The opposite however was found in a similar study of warming temperatures and nutrient reductions in the Chesapeake Bay (Ni et al., 2020), where rising temperatures between 1985 and 2016 cancelled out any potential benefits from nutrient reductions by decreasing O_2 concentrations in bottom waters by a larger amount than they have been increased by nutrient reductions. Although Ni et al. (2020) focused on average O_2 concentrations rather than hypoxic volume and used a continuous simulation from 1985 to 2016 rather than applying a delta approach as used here, it is unlikely that these differences alone explain why the results differ from those found in this study. An alternative explanation could be that the nutrient reductions calculated by Ni et al. (2020) underestimate the magnitude of reductions observed in the Bay as they only consider above RIM changes in NO_3^-, NO_2^-, and PO_4^- in the Susquehanna River for their analysis. Large reductions in other nitrogen constituents (e.g., NH_4^+, DON, PON) and below RIM NO_3^- loading have also occurred in the Susquehanna River, and the nitrogen reductions found in other rivers entering the Bay are non-trivial (Fig. 2). It is likely that the total impact of nitrogen reductions is larger in this study as a result of accounting for these additional nitrogen reductions in the Chesapeake Bay watershed.

The relative impacts that nutrient reductions and climate change have on hypoxia can vary dramatically between systems. In the Gulf of Mexico, a similar nutrient dominant effect was found where a proposed 30% reduction in NO_3^- loading from the Mississippi River decreased the occurrence of hypoxia by a larger magnitude than the increase in occurrence resulting from a 4 °C increase in water temperature (Justić et al., 2003). The opposite is generally found for studies of future change in the Baltic Sea (Meier et al., 2011, 2012), where the projected impacts from climate change tend to outweigh the anticipated impact of future nutrient reductions, owing primarily to enhanced precipitation in the watershed increasing river runoff rather than warming temperatures. As a result of these differences, studies examining the competing impacts of nutrient reductions and climate change on hypoxia need to be performed individually for each system in order to account for different responses to local and global drivers.

Finally, warming temperatures can partially account for the discrepancy in the magnitude of nutrient reduction impacts between the statistical analysis and the numerical modeling results, since increasing temperatures in the Bay have at least partially counteracted the impact of nutrient reductions. For $O_2 < 3$ mg L^{-1}, the estimated additional hypoxic volume that would have occurred without nutrient reductions is 83–557 km^3 days greater in the numerical modeling results than the statistical results depending on the year and the terrestrial inputs. For $O_2 < 1$ mg L^{-1}, this difference between the two methods ranges from 141 to 373 km^3 days. Since the temperature effect ranges from 85 to 126 km^3 days for $O_2 < 3$ mg L^{-1} and 24–68 km^3 days for $O_2 < 1$ mg L^{-1} (Fig. 6), it accounts for 16–72% of the discrepancy between the two methods for $O_2 < 3$ mg L^{-1} (128% for Phase 6 in 2019 since the temperature effect is greater than the difference between the two methods for that year) and 7–42% of the discrepancy for $O_2 < 1$ mg L^{-1}. The remaining difference between the two methods can be attributed to a variety of factors, including long term changes in other environmental factors not accounted for in the numerical modeling simulations, as well as inherent differences and uncertainties associated with the two methods.

4.4. Sources of uncertainty

One of the largest sources of uncertainty in this study comes from estimating nitrogen input into the Bay. Over a large geographic area such as the Chesapeake Bay watershed, it is infeasible to collect concentration data with high temporal resolution for the various nitrogen species entering the Bay from different riverine sources. Statistical methods such as WRTDS can be used to fill temporal gaps in data at specific station locations (Hirsch, 2014; Hirsch et al., 2010), however these methods are unable to provide information on the portions of the watershed that are unmonitored. As a

result, estuarine numerical models are reliant on terrestrial watershed models to provide the necessary spatial coverage. These terrestrial models are calibrated and validated using available data (Pan et al., 2021; Shenk et al., 2012; Yao et al., 2021), however uncertainty still exists, particularly for nitrogen constituents and locations with limited data. To quantify this uncertainty, all simulations were run in duplicate using terrestrial input relationships derived from two separate watershed models (DLEM and Phase 6), both of which have been demonstrated to successfully generate estimates of nitrogen loading to the Chesapeake Bay (CBP, 2017; Pan et al., 2021).

The two watershed models produce very similar estimates of freshwater discharge; however, because they represent nutrient cycling processes on land differently, the relationships between discharge and nutrient concentrations vary considerably. Using both sets of estimates in this analysis results in a range of values that is more likely to capture the actual nutrient loading than either of the watershed models individually and provides increased confidence in the results of this study. Other studies have found that multiple numerical model approaches like this are more robust than relying on a single numerical model, since model performance varies uniquely under different conditions for each model (Apostel et al., 2021; Irby et al., 2016; Singh et al., 2005).

In comparing the results from the DLEM and Phase 6 simulations, the largest discrepancy in annual hypoxic volume between the two sets of terrestrial inputs occurs in the organic nitrogen sensitivity experiments. Results from the inorganic nitrogen sensitivity experiments better agree between DLEM and Phase 6 (Fig. 6; Table S6) due to the fact that the magnitude of the increase in loading to recreate 1985 conditions is more similar in these simulations (roughly $+35$ Gg yr^{-1} for DLEM and roughly $+50$ Gg yr^{-1} for Phase 6) compared to the organic nitrogen simulations (roughly $+49$ Gg yr^{-1} for DLEM and roughly $+16$ Gg yr^{-1} for Phase 6). This larger difference in the organic nitrogen loading change applied to the sensitivity experiments causes the increase in annual hypoxic volume to be four times greater for DLEM compared to Phase 6 (Fig. 6; Table S6). Since the multipliers used to adjust nitrogen concentrations in the sensitivity experiments are the same for both terrestrial inputs (Eqs. (4)–(5)), the larger increase in organic nitrogen loading in DLEM is due to the fact that the organic nitrogen multipliers are large (Table S4) and DLEM has a higher baseline organic nitrogen loading of roughly 59 Gg yr^{-1} compared to 40 Gg yr^{-1} for Phase 6. Even though the multipliers for ammonium are large and Phase 6 does have a higher baseline inorganic nitrogen loading than DLEM (145 Gg yr^{-1} compared to 104 Gg yr^{-1}), the increase in inorganic nitrogen loading is more similar for the DLEM and Phase 6 because the higher baseline in Phase 6 is primarily due to higher NO_3^- loading and the NO_3^- multipliers are generally small (Table S4).

When examining organic nitrogen skill in the numerical model simulations, comparisons between numerical model results and observed values indicate that simulations with DLEM generally overestimate DON concentrations in the oligohaline and mesohaline Bay (Figs. S6 & S8) while simulations with Phase 6 generally underestimate DON concentrations in the same region (Figs. S7–S8). It is clear from these differences that there is considerable uncertainty in estimating organic nitrogen loading entering the Bay, owing primarily to the lack of organic nitrogen data in streams and rivers throughout the Bay's watershed. Since results from the sensitivity experiments shown here demonstrate that organic nitrogen loading can have a substantial impact on hypoxia, future monitoring efforts aimed at improving the understanding of how nutrients impact the Bay should focus on obtaining more information regarding the cycling of organic nitrogen in the watershed. This would not only improve the calibration and validation of the watershed models but would also allow statistical methods such as WRTDS to be employed directly to organic nitrogen data so that more accurate loading estimates can be obtained.

An additional source of uncertainty in this study comes from estimating the magnitude of nitrogen reductions over the time period studied (1985–2019). To limit this uncertainty, the data-based WRTDS approach was used to calculate nutrient reductions for the portion of the watershed located above the RIM stations (Fig. 2). For the small portion of the watershed not captured by these stations (below RIM), nutrient reduction estimates were calculated from watershed model output. Since it is difficult to quantify the uncertainty associated with this joint data-numerical modeling approach, the percent changes in loading obtained were both increased and decreased by 30% to determine the sensitivity of the method applied. The difference in hypoxic volume between these high and low loading scenarios is similar in magnitude to the difference in additional volume between the DLEM and Phase 6 simulations, suggesting that the uncertainty associated with the terrestrial inputs (DLEM versus Phase 6) is of a similar magnitude to $\pm 30\%$ of the change in nutrient loading over the 35-year period. Additionally, the annual hypoxic volumes from the low loading scenario are far from overlapping with the realistic volumes in 2016–2019 (Fig. 5e–f, k–l), indicating that the conclusion of increased resilience remains valid under a broad range of assumptions for the terrestrial inputs. Although uncertainty exists, the use of a data-based approach to estimate a substantial portion of the nutrient reductions in the watershed should provide an added level of confidence in the values obtained and the subsequent results from the sensitivity experiments.

Uncertainty in this study is also associated with multiple methodological assumptions that are made for both the statistical and numerical modeling analyses. For the statistical analysis, there is uncertainty in both the interpolated hypoxic volumes and the freshwater discharge estimates. Hypoxic volume cannot be measured directly and is therefore estimated by interpolating O_2 data collected over multiple days. As a result, there is a degree of temporal uncertainty associated with this non-synoptic sampling (potentially ~5 km^3 for $O_2 < 2$ mg L^{-1}) as well as uncertainty from the spatial interpolation (potentially ~2.4 km^3 for $O_2 < 2$ mg L^{-1}; Bever et al., 2013). In addition, the discharge values used in the statistical analysis are estimates of average annual total streamflow into the Bay calculated from discharge data in the Susquehanna, Potomac, and James rivers. Although the discharge in these three rivers is well constrained, uncertainty exists when scaling these values up to reflect the entire watershed. In the numerical modeling analysis, uncertainty is introduced at every step in the process from numerical model formulation and parameterization to the forcing that is used to generate the simulations. As was done in this study, this uncertainty can be minimized by using available data to evaluate numerical model processes and improve the accuracy of the results; however, a baseline level of uncertainty still exists. Despite the underlying uncertainties in both the statistical and numerical modeling approaches, the general agreement between these two independent methods of estimating resilience provides an added level of confidence that the overall findings from this study are robust.

4.5. Informing the management effort

The primary objective of this research is to quantify the degree to which nutrient reductions in the watershed have improved O_2 conditions in the Chesapeake Bay in recent years despite the adverse environmental conditions posed by warming temperatures and unusually wet conditions. This information is of interest to watershed managers, policymakers, and other stakeholders who have devoted a large amount of time and resources toward improving water quality conditions in the Bay (Boesch, 2019; USEPA, 2010). The increased resilience shown in this study is encouraging, demonstrating that despite the occurrence of large hypoxic volumes in recent years, efforts to reduce nutrient input to the Bay have improved water quality by effectively preventing more hypoxia from developing.

From an ecological perspective, this curtailment of hypoxia has important implications for ecosystem health in the Bay. Numerical model results indicate that if nutrient reductions did not occur, 50–90 days of additional hypoxia would have occurred at specific locations in the oligohaline and southern mesohaline portions of the Chesapeake Bay for the thresholds of $O_2 < 3$ mg L^{-1} and $O_2 < 1$ mg L^{-1} (Fig. 7). At these locations where oxygen remained above this 3 mg L^{-1} threshold for a large portion of the summer due to nutrient reductions, demersal fish had greater habitat availability since this community tends to avoid areas with $O_2 < 3$ mg L^{-1} (Breitburg, 2002; Buchheister et al., 2013). Less hypoxic water is also

beneficial for other economically important species in the Bay such as striped bass, which have been shown to have a reduced aerobic scope when subjected to water with $O_2 < 3$ mg L^{-1} (Lapointe et al., 2014). Although hypoxia may still be present, a decrease in duration of hypoxic conditions at some of these locations is critical for the recruitment of ecological important species like the Bay anchovy, whose eggs and larvae are threatened when exposed to water with $O_2 < 3$ mg L^{-1} for durations of ~30 days (USEPA, 2003). In addition to better recruitment and habitat availability, a shift from more severe and expansive hypoxic events to ones that are short-lived and localized may also better support higher trophic levels by allowing greater access to prey as benthic infauna move closer to the sediment-water interface in search of O_2 while still remaining accessible from higher O_2 waters (Long and Seitz, 2008; Nestlerode and Diaz, 1998). Within the benthic community itself, a decrease in the number of days with $O_2 < 1$ mg L^{-1} can be the difference between life and death, as the majority of macrobenthic species experience mortality when conditions are below this threshold for only a few days (Llansó, 1992; Seitz et al., 2009). Thus, any decrease in the presence of hypoxia at both thresholds is likely to result in greater species diversity and an improved overall health of the ecosystem. Because this study has shown that oxygen conditions at these critical ecological thresholds have improved, Bay managers and policymakers now have additional evidence to support the continuation and enhancement of current nutrient reduction strategies.

Although the results from this work demonstrate that nutrient management efforts in the Chesapeake Bay have been effective, the impact that increasing temperatures have already had on hypoxia presents a cautionary tale. An increase in summer water temperature of roughly 1.4 °C over the 35-year period from 1985 to 2019 has already offset between 6 and 34% of the improvements in hypoxia due to nutrient reductions, and temperatures in the Bay are expected to warm by an additional 2–5 °C over the course of the 21st century (Muhling et al., 2018). So far management efforts have been able to outpace the competing impact of temperature, and this is expected to continue through the middle of the 21st century if the nutrient reduction targets set in the 2010 Chesapeake Bay TMDL are met (Irby et al., 2018). This dynamic may shift toward the end of the century however, particularly if the high estimate of a 5 °C increase in water temperature is achieved.

Nutrient management efforts should not be viewed as futile even in the face of such a large degree of warming; on the contrary, they may be the key to minimizing the overall impact that increasing temperatures as a result of climate change have on hypoxia. The impact of warming on hypoxia varies depending on the size of the hypoxic region, with a greater impact during years with large hypoxic volumes (2018–2019) compared to years with smaller hypoxic volumes (2016–2017; Fig. 6). Since the impact of temperature on O_2 solubility is not a function of freshwater discharge, this greater temperature effect in wetter years is likely due to enhanced microbial respiration at depth as a result of more organic matter during years with higher nutrient loading. If nutrient reductions can successfully limit this buildup of organic matter in bottom waters in the future, the impact of temperature on hypoxia may be minimized. The findings of this study emphasize the importance of not only accounting for warming temperatures in determining the magnitude of nutrient reductions required to raise bottom O_2 concentrations in the Bay in the future, but also the timeline by which these reductions should take place in order to minimize future negative effects.

5. Summary and conclusions

This study has demonstrated that nutrient reductions from 1985 to 2019 have made the Chesapeake Bay more resilient to warming atmospheric temperatures and high discharge years by preventing additional hypoxia from developing. If nutrient reductions had not occurred, hypoxic volumes would have been ~50–120% (~80–280%) greater for $O_2 < 3$ mg L^{-1} ($O_2 < 1$ mg L^{-1}) during the average discharge years of 2016–2017 and ~20–50% (~30–100%) greater for $O_2 < 3$ mg L^{-1} ($O_2 < 1$ mg L^{-1}) during the wet years of 2018–2019. Despite these interannual differences in the relative magnitude of resilience, the absolute magnitude of resilience is similar from 2016 to 2019, suggesting that nutrient reductions are equally effective at curtailing hypoxia under a range of hydrologic conditions. This reduction in hypoxia is primarily attributed to (1) a decrease in the along estuary extent of hypoxia in the mainstem of the Bay and (2) a decrease in the length of the hypoxia season during average discharge years. As a result, future monitoring efforts would benefit from collecting high resolution O_2 data near the northernly and southernly extent of hypoxia. In addition, since hypoxia is strongly dependent on organic nitrogen loading, a focus on collecting more organic nitrogen data throughout the Chesapeake Bay watershed is recommended.

Although nutrient reductions have substantially decreased Chesapeake Bay hypoxia over the past 35 years, these decreases would have been larger if warming temperatures had not offset 6–34% of the improvement in hypoxia by decreasing oxygen solubility and increasing the rate of microbial respiration. Once this temperature effect is taken into consideration, the magnitude of increased resilience from the data-based statistical analysis and the numerical modeling analysis agree reasonably well, providing an added level of confidence in the results from this study. Even though nutrient reductions have been able to outpace the negative effects of increasing temperatures thus far, future warming may reverse this dynamic if adequate nutrient management strategies are not implemented. Overall, the positive impact that nutrient reductions have had on hypoxia in recent years demonstrate that if management actions are implemented, the health of the Chesapeake Bay ecosystem can be improved despite adverse environmental conditions caused by future climate change.

CRediT authorship contribution statement

Luke T. Frankel: Software, Formal analysis, Methodology, Validation, Visualization, Investigation, Writing – original draft. **Marjorie A.M. Friedrichs:** Conceptualization, Investigation, Methodology, Supervision, Resources, Writing – review & editing, Funding acquisition. **Pierre St-Laurent:** Investigation, Methodology, Software, Formal analysis, Data curation, Writing – review & editing. **Aaron J. Bever:** Software, Resources, Writing – review & editing. **Romuald N. Lipcius:** Methodology, Formal analysis, Writing – review & editing. **Gopal Bhatt:** Formal analysis, Resources. **Gary W. Shenk:** Formal analysis, Resources, Writing – review & editing.

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.

Acknowledgements

Many thanks to the Ecosystem Dynamics and Global Ecology Laboratory at Auburn University and Qian Zhang from the CBP for providing model output used to perform this analysis. Additionally, the authors would like to acknowledge the members of the Chesapeake Hypoxia Analysis and Modeling Program's Management Transition Advisory Group and the CBP's Modeling Workgroup for providing feedback that has greatly contributed to this research. This manuscript has greatly improved from the feedback provided by the editor, Neil Ganju of the USGS, and five anonymous reviewers. The research presented here was funded by the National Oceanic and Atmospheric Administration's National Centers for Coastal Ocean Science under award NA16NOS4780207 to the Virginia Institute of Marine Science. The authors also acknowledge William & Mary Research Computing for providing computational resources and technical support that have contributed to the results reported within this paper (https://www.wm.edu/it/rc). The model results and data used in this manuscript are permanently available and can be downloaded online from the William & Mary ScholarWorks data repository associated with this article (https://doi.org/10.25773/st3m-kb83). This is contribution number 4063 of the Virginia Institute of Marine Science, William & Mary. Any use of trade, firm, or product names

is for descriptive purposes only and does not imply endorsement by the U.S. Government.

Funding

This work was supported by the National Oceanic and Atmospheric Administration's National Centers for Coastal Ocean Science [NA16NOS4780207] and the Virginia Institute of Marine Science Office of Academic Studies.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.scitotenv.2021.152722.

References

Altieri, A.H., Gedan, K.B., 2015. Climate change and dead zones. Glob. Chang. Biol. 21 (4), 1395–1406. https://doi.org/10.1111/gcb.12754.

Anderson, D.R., 2008. Model Based Inference in the Life Sciences: A Primer on Evidence. Springer.

Apostel, A., Kalcic, M., Dagnew, A., Evenson, G., Kast, J., King, K., Martin, J., Muenich, R.L., Scavia, D., 2021. Simulating internal watershed processes using multiple SWAT models. Sci. Total Environ. 759, 143920. https://doi.org/10.1016/j.scitotenv.2020.143920.

Ator, S.W., Blomquist, J.D., Webber, J.S., Chanat, J.G., 2020. Factors driving nutrient trends in streams of the Chesapeake Bay watershed. J. Environ. Qual. 49 (4), 812–834. https://doi.org/10.1002/jeq2.20101.

Bever, A.J., Friedrichs, M.A.M., Friedrichs, C.T., Scully, M.E., 2018. Estimating hypoxic volume in the Chesapeake Bay using two continuously sampled oxygen profiles. J. Geophys. Res. Oceans 123 (9), 6392–6407. https://doi.org/10.1029/2018JC014129.

Bever, A.J., Friedrichs, M.A.M., Friedrichs, C.T., Scully, M.E., Lanerolle, L.W.J., 2013. Combining observations and numerical model results to improve estimates of hypoxic volume within the Chesapeake Bay, USA: improving hypoxic volume estimates. J. Geophys. Res. Oceans 118 (10), 4924–4944. https://doi.org/10.1002/jgrc.20331.

Bever, A.J., Friedrichs, M.A.M., St-Laurent, P., 2021. Real-time environmental forecasts of the Chesapeake Bay: model setup, improvements, and online visualization. Environ. Model Softw. 140, 105036. https://doi.org/10.1016/j.envsoft.2021.105036.

Boesch, D.F., 2019. Barriers and bridges in abating coastal eutrophication. Front. Mar. Sci. 6, 123. https://doi.org/10.3389/fmars.2019.00123.

Boyer, T.P., Baranova, O.K., Coleman, C., Garcia, H.E., Grodsky, A., Locarnini, R.A., Mishonov, A.V., Paver, C.R., Reagan, J.R., Seidov, D., Smolyar, I.V., Weathers, K.W., Zweng, M.M., 2018. In: Mishonov, A.V. (Ed.), World Ocean Database 2018. NOAA Atlas NESDIS 87.

Breitburg, D., 2002. Effects of hypoxia, and the balance between hypoxia and enrichment, on coastal fishes and fisheries. Estuaries 25 (4), 767–781. https://doi.org/10.1007/BF02804904.

Buchheister, A., Bonzek, C., Gartland, J., Latour, R., 2013. Patterns and drivers of the demersal fish community of Chesapeake Bay. Mar. Ecol. Prog. Ser. 481, 161–180. https://doi.org/10.3354/meps10253.

Burnham, K.P., Anderson, D.R., 2010. Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach. 2. ed. Springer.

Carstensen, J., Conley, D.J., 2019. Baltic Sea hypoxia takes many shapes and sizes. Limnol. Oceanogr. Bull. 28 (4), 125–129. https://doi.org/10.1002/lob.10350.

Cerco, C.F., Noel, M.R., 2007. Can oyster restoration reverse cultural eutrophication in Chesapeake Bay? Estuar. Coasts 30 (2), 331–343. https://doi.org/10.1007/BF02700175.

Chesapeake Bay Program, 2017. Chesapeake Bay Program Phase 6 Watershed Model. . ftp://ftp.chesapeakebay.net/modeling/Phase6/Draft_Phase_6/Documentation/.

Chesapeake Bay Program, 2021. Chesapeake Assessment and Scenario Tool (CAST) (Version 2019) [Computer Software]. Chesapeake Bay Program Office.

Chesapeake Bay Program (CBP), 2012. Guide to Using Chesapeake Bay Program Water Quality Monitoring Data. https://www.chesapeakebay.net/documents/3676/wq_data_userguide_10feb12_mod.pdf.

Cooper, S.R., Brush, G.S., 1991. Long-term history of Chesapeake Bay anoxia. Science 254 (5034), 992–996. https://doi.org/10.1126/science.254.5034.992.

Cooper, S.R., Brush, G.S., 1993. A 2,500-year history of anoxia and eutrophication in Chesapeake Bay. Estuaries 16 (3), 617. https://doi.org/10.2307/1352799.

Copernicus Climate Change Service (C3S), 2017. ERA5: Fifth generation of ECMWF atmospheric reanalyses of the global climate. Copernicus Climate Change Service Climate Data Store (CDS). https://cds.climate.copernicus.eu/cdsapp#!/home.

Da, F., Friedrichs, M.A.M., St-Laurent, P., 2018. Impacts of atmospheric nitrogen deposition and coastal nitrogen fluxes on oxygen concentrations in Chesapeake Bay. J. Geophys. Res. Oceans 123 (7), 5004–5025. https://doi.org/10.1029/2018JC014009.

Da, F., Friedrichs, M.A.M., St-Laurent, P., Shadwick, E.H., Najjar, R.G., Hinson, K., 2021. Mechanisms driving decadal changes in the carbonate system of a coastal plain estuary. J. Geophys. Res. Oceans 126 (6). https://doi.org/10.1029/2021JC017239.

Diaz, R.J., Rosenberg, R., 2008. Spreading dead zones and consequences for marine ecosystems. Science 321 (5891), 926–929. https://doi.org/10.1126/science.1156401.

Du, J., Shen, J., Park, K., Wang, Y.P., Yu, X., 2018. Worsened physical condition due to climate change contributes to the increasing hypoxia in Chesapeake Bay. Sci. Total Environ. 630, 707–717. https://doi.org/10.1016/j.scitotenv.2018.02.265.

Duarte, C.M., Conley, D.J., Carstensen, J., Sánchez-Camacho, M., 2009. Return to Neverland: shifting baselines affect eutrophication restoration targets. Estuar. Coasts 32 (1), 29–36. https://doi.org/10.1007/s12237-008-9111-2.

Feng, Y., Friedrichs, M.A.M., Wilkin, J., Tian, H., Yang, Q., Hofmann, E.E., Wiggert, J.D., Hood, R.R., 2015. Chesapeake Bay nitrogen fluxes derived from a land-estuarine ocean biogeochemical modeling system: model description, evaluation, and nitrogen budgets. J. Geophys. Res. Biogeosci. 120 (8), 1666–1695. https://doi.org/10.1002/2015JG002931.

Fennel, K., Testa, J.M., 2019. Biogeochemical controls on coastal hypoxia. Annu. Rev. Mar. Sci. 11 (1), 105–130. https://doi.org/10.1146/annurev-marine-010318-095138.

Fisher, T.R., Gustafson, A.B., Sellner, K., Lacouture, R., Haas, L.W., Wetzel, R.L., Magnien, R., Everitt, D., Michaels, B., Karrh, R., 1999. Spatial and temporal variation of resource limitation in Chesapeake Bay. Mar. Biol. 133 (4), 763–778. https://doi.org/10.1007/s002270050518.

Hagy, J.D., Boynton, W.R., Keefe, C.W., Wood, K.V., 2004. Hypoxia in Chesapeake Bay, 1950–2001: long-term change in relation to nutrient loading and river flow. Estuaries 27 (4), 634–658. https://doi.org/10.1007/BF02907650.

Harding, L.W., Gallegos, C.L., Perry, E.S., Miller, W.D., Adolf, J.E., Mallonee, M.E., Paerl, H.W., 2016. Long-term trends of nutrients and phytoplankton in Chesapeake Bay. Estuar. Coasts 39 (3), 664–681. https://doi.org/10.1007/s12237-015-0023-7.

Harding, L.W., Mallonee, M.E., Perry, E.S., Miller, W.D., Adolf, J.E., Gallegos, C.L., Paerl, H.W., 2016. Variable climatic conditions dominate recent phytoplankton dynamics in Chesapeake Bay. Sci. Rep. 6 (1), 23773. https://doi.org/10.1038/srep23773.

Harding, L.W., Meeson, B.W., Fisher, T.R., 1986. Phytoplankton production in two east coast estuaries: photosynthesis-light functions and patterns of carbon assimilation in Chesapeake and Delaware Bays. Estuar. Coast. Shelf Sci. 23 (6), 773–806. https://doi.org/10.1016/0272-7714(86)90074-0.

Hinson, K.E., Friedrichs, M.A.M., St-Laurent, P., Da, F., Najjar, R.G., 2021. Extent and causes of Chesapeake Bay warming. J. Am. Water Resour. Assoc. 1752–1688, 12916. https://doi.org/10.1111/1752-1688.12916.

Hirsch, R.M., 2014. Large biases in regression-based constituent flux estimates: causes and diagnostic tools. Journal of the American Water Resources Association 50 (6), 1401–1424. https://doi.org/10.1111/jawr.12195.

Hirsch, R.M., Moyer, D.L., Archfield, S.A., 2010. Weighted regressions on time, discharge, and season (WRTDS), with an application to Chesapeake Bay river inputs. Journal of the American Water Resources Association 46 (5), 857–880. https://doi.org/10.1111/j.1752-1688.2010.00482.x.

Hofmann, E.E., Droun, J.-N., Fennel, K., Friedrichs, M.A.M., Haidvogel, D., Lee, C., Mannino, A., McClain, C., Najjar, R.G., O'Reilly, J., Pollard, D., Previdi, M., Seitzinger, S., Siewert, J., Signorini, S., Wilkin, J., 2008. Eastern US continental shelf carbon budget integrating models, data assimilation, and analysis. Oceanography 21 (1), 86–104 JSTOR.

Howarth, R.W., 2008. Coastal nitrogen pollution: a review of sources and trends globally and regionally. Harmful Algae 8 (1), 14–20. https://doi.org/10.1016/j.hal.2008.08.015.

Irby, I.D., Friedrichs, M.A.M., 2019. Evaluating confidence in the impact of regulatory nutrient reduction on Chesapeake Bay water quality. Estuar. Coasts 42 (1), 16–32. https://doi.org/10.1007/s12237-018-0440-5.

Irby, I.D., Friedrichs, M.A.M., Da, F., Hinson, K.E., 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., Friedrichs, M.A.M., Friedrichs, C.T., Bever, A.J., Hood, R.R., Lanerolle, L.W.J., Li, M., Linker, L., Scully, M.E., Sellner, K., Shen, J., Testa, J., Wang, H., Wang, P., Xia, M., 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.

Jolliff, J.K., Kindle, J.C., Shulman, I., Penta, B., Friedrichs, M.A.M., Helber, R., Arnone, R.A., 2009. Summary diagrams for coupled hydrodynamic-ecosystem model skill assessment. J. Mar. Syst. 76 (1–2), 64–82. https://doi.org/10.1016/j.jmarsys.2008.05.014.

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. Mar. Syst. 42 (3–4), 115–126. https://doi.org/10.1016/S0924-7963(03)00070-8.

Kemp, W., Boynton, W., Adolf, J., Boesch, D., Boicourt, W., Brush, G., Cornwell, J., Fisher, T., Glibert, P., Hagy, J., Harding, L., Houde, E., Kimmel, D., Miller, W., Newell, R., Roman, M., Smith, E., Stevenson, J., 2005. Eutrophication of Chesapeake Bay: historical trends and ecological interactions. Mar. Ecol. Prog. Ser. 303, 1–29. https://doi.org/10.3354/meps303001.

Lapointe, D., Vogelbein, W., Fabrizio, M., Gauthier, D., Brill, R., 2014. Temperature, hypoxia, and mycobacteriosis: effects on adult striped bass Morone saxatilis metabolic performance. Dis. Aquat. Org. 108 (2), 113–127.

Lefcheck, J.S., Orth, R.J., Dennison, W.C., Wilcox, D.J., Murphy, R.R., Keisman, J., Gurbisz, C., Hannam, M., Landry, J.B., Moore, K.A., Patrick, C.J., Testa, J., Weller, D.E., Batiuk, R.A., 2018. Long-term nutrient reductions lead to the unprecedented recovery of a temperate coastal region. Proc. Natl. Acad. Sci. 115 (14), 3658–3662. https://doi.org/10.1073/pnas.1715798115.

Llansó, R.J., 1992. Effects of hypoxia on estuarine benthos: the lower Rappahannock River (Chesapeake Bay), a case study. Estuar. Coast. Shelf Sci. 35 (5), 491–515. https://doi.org/10.1016/S0272-7714(05)80027-7.

Long, W., Seitz, R., 2008. Trophic interactions under stress: hypoxia enhances foraging in an estuarine food web. Mar. Ecol. Prog. Ser. 362, 59–68. https://doi.org/10.3354/meps07395.

Luettich, R.A., Westerink, J.J., Scheffner, N.W., 1992. ADCIRC: An Advanced Three-dimensional Circulation Model for Shelves, Coasts, and Estuaries. Report 1, Theory and Methodology of ADCIRC-2DD1 and ADCIRC-3DL (Technical Report DRP-92-6). Coastal Engineering Research Center, Vicksburg, MS.

Maryland Department of Natural Resources, 2019. Chesapeake Bay Health Shows Resilience in 2018 Report Card. https://news.maryland.gov/dnr/2019/05/21/chesapeake-bay-health-shows-resilience-in-2018-report-card/.

Meier, H.E.M., Andersson, H.C., Eilola, K., Gustafsson, B.G., Kuznetsov, I., Müller-Karulis, B., Neumann, T., Savchuk, O.P., 2011. Hypoxia in future climates: a model ensemble study for the Baltic Sea. Geophys. Res. Lett. 38 (24). https://doi.org/10.1029/2011GL049929 n/a-n/a.

Meier, H.E.M., Eilola, K., Almroth-Rosell, E., Schimanke, S., Kniebusch, M., Höglund, A., Pemberton, P., Liu, Y., Väli, G., Saraiva, S., 2019. Disentangling the impact of nutrient load and climate changes on Baltic Sea hypoxia and eutrophication since 1850. Clim. Dyn. 53 (1–2), 1145–1166. https://doi.org/10.1007/s00382-018-4296-y.

Meier, H.E.M., Hordoir, R., Andersson, H.C., Dieterich, C., Eilola, K., Gustafsson, B.G., Höglund, A., Schimanke, S., 2012. Modeling the combined impact of changing climate and changing nutrient loads on the Baltic Sea environment in an ensemble of transient simulations for 1961–2099. Clim. Dyn. 39 (9–10), 2421–2441. https://doi.org/10.1007/s00382-012-1339-7.

Muhling, B.A., Gaitán, C.F., Stock, C.A., Saba, V.S., Tommasi, D., Dixon, K.W., 2018. Potential salinity and temperature futures for the Chesapeake Bay using a statistical downscaling spatial disaggregation framework. Estuar. Coasts 41 (2), 349–372. https://doi.org/10.1007/s12237-017-0280-8.

Murphy, R.R., Kemp, W.M., Ball, W.P., 2011. Long-term trends in Chesapeake Bay seasonal hypoxia, stratification, and nutrient loading. Estuar. Coasts 34 (6), 1293–1309. https://doi.org/10.1007/s12237-011-9413-7.

National Centers for Environmental Information, 2019. National Temperature and Precipitation Maps. NOAA National Centers for Environmental Information. https://www.ncdc.noaa.gov/temp-and-precip/us-maps/.

National Centers for Environmental Information, 2020. North American Mesoscale Forecast System (NAM). NOAA National Centers for Environmental Information. https://www.ncei.noaa.gov/products/weather-climate-models/north-american-mesoscale.

Nestlerode, J., Diaz, R., 1998. Effects of periodic environmental hypoxia on predation of a tethered polychaete, Glycera Americana: implications for trophic dynamics. Mar. Ecol. Prog. Ser. 172, 185–195. https://doi.org/10.3354/meps172185.

Newell, R.I.E., 1988. Ecological changes in Chesapeake Bay: are they the result of overharvesting the American oyster (Crassostrea virginica)? In: Lynch, M.P., Krome, E.C. (Eds.), Understanding the Estuary – Advances in Chesapeake Bay Research. Chesapeake Research Consortium, pp. 536–546 Publication 129

Ni, W., Li, M., Ross, A.C., Najjar, R.G., 2019. Large projected decline in dissolved oxygen in a eutrophic estuary due to climate change. J. Geophys. Res. Oceans 124 (11), 8271–8289. https://doi.org/10.1029/2019JC015274.

Ni, W., Li, M., Testa, J.M., 2020. Discerning effects of warming, sea level rise and nutrient management on long-term hypoxia trends in Chesapeake Bay. Sci. Total Environ. 737, 139717. https://doi.org/10.1016/j.scitotenv.2020.139717.

Officer, C.B., Biggs, R.B., Taft, J.L., Cronin, L.E., Tyler, M.A., Boynton, W.R., 1984. Chesapeake Bay anoxia: origin, development, and significance. Science 223 (4631), 22–27. https://doi.org/10.1126/science.223.4631.22.

Pan, S., Bian, Z., Tian, H., Yao, Y., Najjar, R.G., Friedrichs, M.A.M., Hofmann, E.E., Xu, R., Zhang, B., 2021. Impacts of multiple environmental changes on long-term nitrogen loading from the Chesapeake Bay watershed. Journal of geophysical researchBiogeosciences 126 (5). https://doi.org/10.1029/2020JG005826.

R Core Team, 2020. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. https://www.R-project.org/.

Rabalais, N.N., Turner, R.E., 2019. Gulf of Mexico hypoxia: past, present, and future. Limnol. Oceanogr. Bull. 28 (4), 117–124. https://doi.org/10.1002/lob.10351.

Schulte, D.M., Burke, R.P., Lipcius, R.N., 2009. Unprecedented restoration of a native oyster metapopulation. Science 325 (5944), 1124–1128. https://doi.org/10.1126/science.1176516.

Scully, M.E., 2010a. The importance of climate variability to wind-driven modulation of hypoxia in Chesapeake Bay. J. Phys. Oceanogr. 40 (6), 1435–1440. https://doi.org/10.1175/2010JPO4321.1.

Scully, M.E., 2010b. Wind modulation of dissolved oxygen in Chesapeake Bay. Estuar. Coasts 33 (5), 1164–1175. https://doi.org/10.1007/s12237-010-9319-9.

Seitz, R.D., Dauer, D.M., Llansó, R.J., Long, W.C., 2009. Broad-scale effects of hypoxia on benthic community structure in Chesapeake Bay, USA. J. Exp. Mar. Biol. Ecol. 381, S4–S12. https://doi.org/10.1016/j.jembe.2009.07.004.

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

Shenk, G.W., Wu, J., Linker, L.C., 2012. Enhanced HSPF model structure for Chesapeake Bay watershed simulation. J. Environ. Eng. 138 (9), 949–957. https://doi.org/10.1061/(ASCE)EE.1943-7870.0000555.

Singh, J., Knapp, H.Vernon, Arnold, J.G., Demissie, M., 2005. Hydrolocial modeling of the Iroquois River watershed using HSPF and SWAT. J. Am. Water Resour. Assoc. 41 (2), 343–360. https://doi.org/10.1111/j.1752-1688.2005.tb03740.x.

St-Laurent, P., Friedrichs, M.A.M., Najjar, R.G., Shadwick, E.H., Tian, H., Yao, Y., 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.

Testa, J.M., Kemp, W.M., 2012. Hypoxia-induced shifts in nitrogen and phosphorus cycling in Chesapeake Bay. Limnol. Oceanogr. 57 (3), 835–850. https://doi.org/10.4319/lo.2012.57.3.0835.

Testa, J.M., Kemp, W.M., Boynton, W.R., 2018. Season-specific trends and linkages of nitrogen and oxygen cycles in Chesapeake Bay. Limnol. Oceanogr. 63 (5), 2045–2064. https://doi.org/10.1002/lno.10823.

Testa, J.M., Kemp, W.M., Boynton, W.R., Hagy, J.D., 2008. Long-term changes in water quality and productivity in the Patuxent River Estuary: 1985 to 2003. Estuar. Coasts 31 (6), 1021–1037. https://doi.org/10.1007/s12237-008-9095-y.

Testa, J.M., Li, Y., Lee, Y.J., Li, M., Brady, D.C., Di Toro, D.M., Kemp, W.M., 2017. Modeling physical and biogeochemical controls on dissolved oxygen in Chesapeake Bay: lessons learned from simple and complex approaches. In: Justic, D., Rose, K.A., Hetland, R.D., Fennel, K. (Eds.), Modeling Coastal Hypoxia. Springer International Publishing, pp. 95–118 https://doi.org/10.1007/978-3-319-54571-4_5.

Tian, H., Yang, Q., Najjar, R.G., Ren, W., Friedrichs, M.A.M., Hopkinson, C.S., Pan, S., 2015. Anthropogenic and climatic influences on carbon fluxes from eastern North America to the Atlantic Ocean: a process-based modeling study. J. Geophys. Res. Biogeosci. 120 (4), 757–772. https://doi.org/10.1002/2014JG002760.

U.S. Geological Survery (USGS), 2020. National Water Information System. http://waterdata.usgs.gov/nwis.

U.S. Geological Survery (USGS), 2021. Chesapeake Bay Estimated Streamflow. https://www.usgs.gov/centers/cba/science/freshwater-flow-chesapeake-bay.

USEPA, 1982. Chesapeake Bay Program Technical Studies: A Synthesis. US Environmental Protection Agency.

USEPA, 2003. Ambient Water Quality Criteria for Dissolved Oxygen, Water Clarity and Chlorophyll a for the Chesapeake Bay and Its Tidal Tributaries. Chesapeake Bay Program Office, p. 343 EPA 903-R-03-002.

USEPA, 2010. Chesapeake Bay Total Maximum Daily Load for Nitrogen, Phosphorus and Sediment. US Environmental Protection Agency, US Environmental Protection Agency Chesapeake Bay Program Office, Annapolis, MD.

Virginia Institute of Marine Science, & Anchor QEA, LLC, 2020. 2020 Chesapeake Bay Dead Zone Report , pp. 1–2.. https://www.vims.edu/research/topics/dead_zones/_docs/2020_cb_hypoxia_reportcard.pdf.

Wang, B., Hu, J., Li, S., Yu, L., Huang, J., 2018. Impacts of anthropogenic inputs on hypoxia and oxygen dynamics in the Pearl River estuary. Biogeosciences 15 (20), 6105–6125. https://doi.org/10.5194/bg-15-6105-2018.

Xu, J., Long, W., Wiggert, J.D., Lanerolle, L.W.J., Brown, C.W., Murtugudde, R., Hood, R.R., 2012. Climate forcing and salinity variability in Chesapeake Bay, USA. Estuar. Coasts 35 (1), 237–261. https://doi.org/10.1007/s12237-011-9423-5.

Yang, Q., Tian, H., Friedrichs, M.A.M., Hopkinson, C.S., Lu, C., Najjar, R.G., 2015. Increased nitrogen export from eastern North America to the Atlantic Ocean due to climatic and anthropogenic changes during 1901–2008. J. Geophys. Res. Biogeosci. 120 (6), 1046–1068. https://doi.org/10.1002/2014JG002763.

Yang, Q., Tian, H., Friedrichs, M.A.M., Liu, M., Li, X., Yang, J., 2015. Hydrological responses to climate and land-use changes along the North American East Coast: a 110-year historical reconstruction. Journal of the American Water Resources Association 51 (1), 47–67. https://doi.org/10.1111/jawr.12232.

Yao, Y., Tian, H., Pan, S., Najjar, R.G., Friedrichs, M.A.M., Bian, Z., Li, H., Hofmann, E.E., 2021. Riverine carbon cycling over the past century in the Mid-Atlantic region of the United States. Journal of geophysical researchBiogeosciences 126 (5). https://doi.org/10.1029/2020JG005968.

Zhang, Q., Brady, D.C., Ball, W.P., 2013. Long-term seasonal trends of nitrogen, phosphorus, and suspended sediment load from the non-tidal Susquehanna River Basin to Chesapeake Bay. Sci. Total Environ. 452–453, 208–221. https://doi.org/10.1016/j.scitotenv.2013.02.012.

Zhang, Q., Brady, D.C., Boynton, W.R., Ball, W.P., 2015. Long-term trends of nutrients and sediment from the nontidal Chesapeake watershed: an assessment of Progress by river and season. Journal of the American Water Resources Association 51 (6), 1534–1555. https://doi.org/10.1111/1752-1688.12327.

Zhang, Q., Hirsch, R.M., Ball, W.P., 2016. Long-term changes in sediment and nutrient delivery from Conowingo Dam to Chesapeake Bay: effects of reservoir sedimentation. Environ. Sci. Technol. 50 (4), 1877–1886. https://doi.org/10.1021/acs.est.5b04073.

Zheng, G., DiGiacomo, P.M., 2020. Linkages between phytoplankton and bottom oxygen in the Chesapeake Bay. J. Geophys. Res. Oceans 125 (2). https://doi.org/10.1029/2019JC015650.

Zhou, Y., Scavia, D., Michalak, A.M., 2014. Nutrient loading and meteorological conditions explain interannual variability of hypoxia in Chesapeake Bay. Limnol. Oceanogr. 59 (2), 373–384. https://doi.org/10.4319/lo.2014.59.2.0373.