Title: Discerning effects of warming, sea level rise and nutrient management on long-term hypoxia trend in Chesapeake Bay

Authors: Wenfei Ni[1], Ming Li[1*], Jeremy Testa[2]

Affiliations:

[1] Horn Point Laboratory, University of Maryland Center for Environmental Science, Cambridge, MD 21613, United States

[2] Chesapeake Biological Laboratory, University of Maryland Center for Environmental Science, Solomon Island, MD 20688, United States

Submitted to *Science of The Total Environment*

*Correspondence to: Ming Li, Horn Point Laboratory., University of Maryland Center for Environmental Science, 2020 Horns Point Road, Cambridge, MD 21613. Email: mingli@umces.edu.

Abstract

Analyses of dissolved oxygen concentration in Chesapeake Bay over the past three decades suggested seasonally-dependent changes in hypoxic volume and an earlier end of hypoxic conditions. While these studies hypothesized and evaluated multiple potential driving mechanisms, quantitative evidence for the relative effects of various drivers has yet to be presented. In this study, a coupled physical-biogeochemical model was used to conduct hindcast simulations between 1985 and 2016. Additional numerical experiments, in which the long-term trends in external drivers were removed, were analyzed to discern the separate effects of temperature increase, sea level rise and nutrient reduction. After the removal of seasonal and interannual variations, dissolved oxygen concentration in all regions of the estuary showed a statistically significant declining trend: ~0.1 mg/L per decade. Most of this decline occurred during winter and spring while May-August hypoxic volumes showed no changes and September hypoxic volume showed a slight decrease (~0.9 km^3). Our simulations show that warming was the dominant driver of the long-term oxygen decline, overwhelming the effects of sea level rise and modest oxygen increases associated with nutrient reduction. There was no statistically significant trend in the initiation of hypoxia in spring, where the potential delay associated with nutrient reduction was offset by warming-induced oxygen declines, and both nutrient reduction and warming contributed to an earlier disintegration of hypoxia in the fall. These results suggest that recent warming has prevented oxygen improvements in Chesapeake Bay expected from nutrient input reductions and support the expectation that continued warming will serve to counter future nutrient management actions.

Keywords: hypoxia, nutrient management, climate change, long-term trend

1. Introduction

Anthropogenic nutrient enrichment of estuaries has contributed to the degradation of water quality by fueling phytoplankton production and associated depletion of oxygen (hypoxia) from bottom waters in coastal systems worldwide (Diaz and Rosenberg, 2008; Kemp et al., 2009; Zhang et al., 2009; Breitburg et al. 2018). Despite the fact that many coastal regions have made major public commitments to reduce nutrient loading and reverse this trend of declining water quality and habitat conditions, estuaries and coastal waters around the world continue to experience hypoxia and deteriorating water quality (Conley et al., 2009a; Duarte et al., 2009; Scavia et al., 2017; Fennel and Testa, 2019). A major impediment for achieving restoration successes is the complicating effect of climate variability and climate change. Large interannual fluctuations in river flows result in highly variable nutrient loading and strong interannual variability in hypoxia (Justic et al., 2003; Hagy et al., 2004; Bever et al., 2013; Li et al., 2016). Longer-term climate changes, such as warming and sea level rise, exert more subtle controls on biogeochemical processes, and their effects on hypoxia in coastal systems are not well understood.

Various nutriment management strategies have been implemented in coastal regions around the world. For example, Europe developed the Water Framework Directive (WFD) and Marine Strategy Framework Directive (MSFD) to coordinate and integrate catchment-to-coast management for the environmental assessment among European countries (Borja et al., 2010). In the Baltic Sea, continuous efforts have been made to reduce anthropogenic nutrient load since the late 1980s, first enacted through the Helsinki Commission (HELCOM) and later through an international agreement on the Baltic Sea Action Plan (BSAP) (HELCOM, 2007). With modest

73 reductions in nitrogen and phosphorus loads, there was only water quality recovery at specific
74 sites (Conley et al., 2009b). In the northern Gulf of Mexico, the nutrient management action plan
75 was released by Mississippi River/Gulf of Mexico Watershed Nutrient Task Force in 2001, but
76 there have been little signs of summer hypoxic area contraction (Rabalais et al., 2010; Van Meter
77 et al., 2018). On the other hand, major improvements in the dissolved oxygen concentration (O_2
78 hereafter) have been observed in Tampa Bay, the northwest Black Sea and the northern Adriatic
79 Sea, where anthropogenic nutrient inputs decreased substantially (Greening and Janicki, 2006;
80 Mee et al., 2006; Giani et al., 2012). Other coastal regions, such as the Bohai Sea and the Pearl
81 River Estuary, experienced expansion of hypoxia in recent years due to continuing increases in
82 nutrient loading (Xin et al., 2019; Qian et al., 2018; Zhai et al., 2019). These divergent
83 trajectories of water quality trends suggest that water quality recovery depends critically on the
84 magnitude of nutrient reduction. A modest nutrient reduction may not lead to the desired
85 outcomes in hypoxia reduction, particularly since other factors such as climate change also leads
86 to ocean deoxygenation (Breitburg et al., 2018).
87
88 Climate change causes sea level rise, temperature increases, and altered pattern of
89 precipitation and wind, which can produce varied effects on coastal hypoxia (Altieri and Gedan,
90 2015). In the Baltic Sea, the North Sea and Chesapeake Bay, sea level rise may lead to stronger
91 vertical stratification with consequent O_2 decline in the bottom water (Meier et al., 2017; Meire
92 et al., 2013; Ni et al., 2019). Warming may increase the duration and spatial extent of hypoxia
93 via reduced oxygen solubility, faster oxygen consumption and intensified internal nutrient
94 cycling (Meier et al., 2011; Lake and Brush, 2015; Irby et al., 2018). Over 30% of the observed
95 O_2 decrease in the Long Island Sound in the past 20 years has been attributed to warming

96 (Staniec and Vlahos, 2017). Winds and river flows can affect estuarine and coastal hypoxia by
97 regulating the vertical supply of O_2 to the bottom water (O'Donnell et al., 2008; Scully, 2013;
98 Hetland and DiMarco, 2008). Changes in wind speed, direction and freshwater discharge were
99 shown to regulate the interannual and decadal variations of hypoxia in coastal waters (Wilson et
100 al., 2014; Feng et al., 2012; Scully, 2010, 2016; Zillén et al., 2008). Thus, climatic forcing may
101 generate comparable impacts on coastal hypoxia as changes in nutrient loading (Meier et al.,
102 2011; Saravia et al., 2019), but the relative roles of climate change and nutrient loading in
103 driving long-term hypoxia trends in estuaries and coastal oceans are not quantified yet.
104
105 Chesapeake Bay is the largest estuary in the U.S. and has suffered from seasonal hypoxia
106 since 1950s (Figure 1a). Retrospective analyses of routine water quality monitoring data since
107 1985 have found no long-term trends in the magnitude of summer hypoxia but an apparent early
108 shift of the seasonal hypoxia cycle. Murphy et al. (2011) and Testa et al. (2018) found significant
109 increases in early summer hypoxia but a slight decrease in late summer hypoxia since 1985.
110 Zhou et al. (2014) reached a similar conclusion that the timing of the maximum hypoxic volume
111 shifted from late to early July but no long term trend was detected in the seasonal-maximum of
112 hypoxic volume. Both Zhou et al. (2014) and Testa et al. (2018) found that hypoxia terminated
113 earlier in the fall but did not detect significant shifts in the timing of hypoxia onset. Testa et al.
114 (2018) hypothesized that the earlier shift of summer hypoxia cycle was linked to the altered
115 external forcing such as nutrient loading and water temperature. In addition to the above
116 observational studies, modeling studies by Scully (2016) and Du et al. (2018) highlighted the
117 importance of physical processes in controlling hypoxia in Chesapeake Bay and suggested that a
118 worsened physical condition may have contributed to the hypoxia increase over the past few

decades. However, these models used simple parameterization of oxygen consumption and did not provide a full account of nutrient loading and associated biogeochemical processes important for hypoxia consumption. Therefore, it remains unclear what drove the long term changes in O_2 in Chesapeake Bay over the past 30 years.

The effects of nutrient management on hypoxia also need to be better understood. Coordinated plans to control point and nonpoint nutrient sources among state and federal stakeholders in the Chesapeake Bay watershed started in 1983. In 2010 the Total Maximum Daily Load (TMDL) was developed to set limits to meet water quality in the Chesapeake Bay and its tidal tributaries (Shenk et al., 2013). As a result, the dissolved nitrogen and phosphorus concentration in the Susquehanna River, the largest tributary of Chesapeake Bay, has reduced by ~15% and ~28% respectively over the past three decades (Zhang et al., 2015). Meanwhile, Chesapeake Bay has experienced a rapid rate of climate change, such as faster warming (~1.5 ^{o}C temperature increase) and higher relative sea level rise (twice the global mean rise rate) (Ding and Elmore, 2015; Boon and Mitchell, 2015). A major outstanding question is how the modest nutrient reduction and rapid climate change interplayed in driving the long-term hypoxia trend in this eutrophic estuary. Furthermore, data analysis is limited by coarse temporal (biweekly to monthly) and spatial (41 stations located in the main stem of Chesapeake Bay) resolutions in monitoring cruises. A numerical model provides a complementary approach. Once validated against observation data, it can generate high resolution and long term time series of physical and biogeochemical variables such that robust statistical analysis can be conducted to detect long term trends. The numerical model can also be used to test different driving mechanisms through scenario runs.

The plan for this paper is as follows. In section 2 we describe the coupled physical-biogeochemical model as well as the statistical model used to analyze the long-term time series of model output. In section 3 we present scenario model runs to discern the effects of warming, sea level rise and nutrient reduction on the long-term O_2 trend. Section 4 discusses the results in the broad context of climate change and nutrient reduction in the water quality management.

2. Methods

2.1 Coupled hydrodynamic-biogeochemical models (ROMS-RCA)

The hydrodynamic model is based on Regional Ocean Modeling System (ROMS) model (Shchepetkin and McWilliams, 2005; Haidvogel et al., 2008), which has been configured for Chesapeake Bay (Li et al. 2005) and validated against a wide variety of observational data (Li et al. 2005, 2006; Zhong and Li, 2006; Xie and Li, 2018). The model has 120×80 horizontal grids (~1-2 km resolution) and 20 sigma-coordinate layers in the vertical direction (Figure 1b). The ROMS model is driven by the atmospheric, riverine and oceanic forcing at the boundaries. Across the sea surface, the air-sea fluxes of momentum and heat are calculated by applying the standard bulk formula (Fairall et al., 2003) to the atmospheric products from the North American Regional Reanalysis (NARR) (Mesinger et al., 2006). At the offshore boundary, the model is forced by open-ocean sea level, temperature and salinity. The sea level includes both tidal and non-tidal components. The former is calculated from 10 harmonic constituents interpolated from the Oregon State University global inverse tidal model TPXO7 (Egbert and Erofeeva, 2002). The latter is obtained from de-tided water level at Duck, NC (NOAA station ID: 8651370). Temperature and salinity at the offshore boundary are interpolated from the World Ocean Atlas

(WOA) climatological averages (http://www.nodc.noaa.gov/OC5/woa13/). At the upstream boundaries of eight major tributaries, the freshwater flows are obtained from daily measurements at USGS gauging stations (http://nwis.waterdata.usgs.gov/nwis) (Figures 1a, 1b). Daily temperature at the tributaries is interpolated from Chesapeake Bay Program monitoring stations (https://www.chesapeakebay.net/).

The biogeochemical model is based on Row-Column AESOP (RCA) which includes a water column component (Isleib et al., 2007) and a sediment component (Di Toro, 2001; Testa et al., 2013; Brady et al., 2013). The water-column model includes state variables representing dissolved inorganic nitrogen, phosphorus, and silica, particulate and dissolved organic nitrogen and phosphorus, two phytoplankton groups and O_2. The sediment model has one aerobic layer and one anaerobic layer, and simulates the cycling of carbon, O_2, nitrogen, phosphorus and sulfur. RCA is driven by loads of dissolved and particulate nutrients from eight rivers and the adjacent Atlantic Ocean. Riverine concentration of phytoplankton, particulate and dissolved organic carbon, organic and inorganic nutrients are obtained from Chesapeake Bay Program (CBP) bi-weekly monitoring data at the eight tributaries. Nutrient concentrations at the offshore ocean boundary are acquired from the WOA 2013 climatology and Filippino et al. (2011). The ROMS-RCA model has been validated and used in several modeling studies (Testa et al., 2014; Li et al., 2016; Testa et al., 2017; Shen et al., 2019a,b, 2020). In particular, Li et al. (2016) demonstrated that ROMS-RCA reproduced interannual variations of hypoxic volume in Chesapeake Bay consistent with interpolated volumes based on observations.

In this paper ROMS-RCA were used to conduct hindcast simulations over three decades: 1985-2016. CBP have carried out regular monitoring cruises since 1985 and collected measurements of biogeochemical variables, such as O_2, at a number of monitoring stations, providing a rich data set to compare with the model results. ROMS was initialized on 1 January 1984 and ran continuously until 31 December 2016. Results from the spin-up period of 1984 were not analyzed. RCA was initialized on 1 January every year using the water-quality data collected in the preceding month (see Testa et al., 2014).

2.2 Statistical analysis approaches

Biogeochemical and physical variables in Chesapeake Bay display large seasonal and interannual variations. To detect long-term trends in O_2, salinity, and water temperature, we applied Generalized Additive Models (GAM) (Hastie and Tibshirani 1986, 1990; Wood, 2006). GAM was previously used to analyze long term trends in chlorophyll-a (*Chl-a*), nutrient concentrations and O_2 in Chesapeake Bay using monitoring data collected by CBP and others (Harding et al., 2015; Testa et al., 2018; Murphy et al., 2019).

Time series extracted from the model outputs were analyzed using GAM as below:

$$y_t \sim y_{t-1} + s(dnum) + s(doy) + s(sal) + ti(dnum, doy) + ti(dnum, sal) + ti(dnum, doy, sal) \quad (1)$$

where y_t represents the response variables such as O_2, y_{t-1} represents the same variables at the preceding time step to account for the autocorrelation, *dnum* is the number of months relative to the reference time (e.g. 1 for Jan. 1985), *doy* is the number of months in a year (e.g. 1 for January), and *sal* is the monthly averaged salinity representing the influence of river flow.

Among the functions used in GAM, *s*() is a smoothing function with thin plate regression splines, *ti*() represents the tensor product of two smooth functions to account for the interaction between these two variates. In Equation (1), *s*(*dnum*) represents the long-term residual, *s*(*doy*) represents the seasonal cycle, and *s*(*sal*) is meant to capture the effects of salinity on the interannual variations. The high-order term *ti*(*dnum, doy*) allows the seasonal cycle to change over time. In the GAM model for salinity, river discharge was used in Equation (1) instead of salinity (Beck and Murphy, 2017). Only the term *s*(*dnum*), *s*(*doy*) and *ti*(*dnum, doy*) were considered in the GAM model for temperature, since Chesapeake Bay is a relatively shallow system dominated by air-sea heat exchange and is relatively insensitive to temperature in riverine inflows.

To investigate estuary-wide responses in model simulations, we calculated the spatial averages over different regions of the Bay. The main stem of Chesapeake Bay was divided into four sub-regions: upper bay (oligohaline), upper mid-bay and lower mid-bay (mesohaline), and lower bay (polyhaline), following Irby et al. (2018) (Figure 1b). The mid-bay was split into two subregions to better account for the bathymetric and hydrodynamic controls on hypoxia in this region. In each region, temperature, salinity and O_2 in the surface and bottom layer were calculated from ROMS-RCA output and averaged at monthly intervals.

The non-parametric Mann-Kendall (MK thereafter) trend test was applied to the time series to statistically assess if there is a monotonic upward or downward trend of a variable over time (Mann, 1945; Kendall, 1975; Gilbert, 1987). A key assumption in the MK trend test is that the data points are not serially correlated over time. This would require that the time between two samples be sufficiently large so that there is no correlation between data collected at

different times. The MK test was applied to both the external forcing such as air temperature and offshore sea level as well as model state variables such as sea level, temperature, salinity, O_2 and hypoxic volume. Since these variables exhibit large seasonal and interannual variations, care is needed when applying the MK test to detect long-term trends. We either calculated monthly averages or used the GAM model to remove the short term signals. To calculate the linear trends of the time series, we used the non-parametric Theil-Sen estimator or Sen's slope estimator (Theil, 1950; Sen, 1968; Wilcox, 2001). The significance level α was set as 0.05 for both tests.

2.3 Model scenarios

Warming, sea level rise and nutrient management are the three major factors driving the long-term changes in Chesapeake Bay. To tease out how each of these factors affected O_2 and hypoxic volume in the estuary, we conducted three scenario model runs in which the long-term trend in temperature (DtrTEMP run) or sea level (DtrSLR run) or nutrient loading (DtrNut run) was removed in the model forcing. For comparison, the hindcast run is hereafter denoted as the Base run.

Both the air temperature over Chesapeake Bay and the riverine temperature in the tributaries have increased over the past several decades (Ding and Elmore, 2015; Rice and Jastram, 2015). The riverine temperature influences the estuarine temperature through river inflows while the air temperature affects it through the air-sea heat fluxes. To remove the long-term trends in temperature forcing in the DtrTEMP run, MK test was applied to the temperature time series at the upstream boundary of the eight major tributaries as well as the heat flux-related variables (air temperature, humidity, solar radiations) at NARR grids over Chesapeake Bay:

$$T_{detrend} = T_{original} - Slope_T \times (Time - 1/1/1985) \quad (2)$$

where $T_{original}$ represents the original time series of each variable (i.e. river temperature, surface air temperature, humidity, downwelling long-wave radiation and net short wave radiation), $T_{detrend}$ represents the detrended variables. $Slope_T$ is the Theil-Sen slope calculated from the original data. Temperature in the Mid-Atlantic Bight also increased over the past few decades (Mountain, 2003), but no long-term temperature series were available near the mouth of Chesapeake Bay.

To remove the sea level rise in the DtrSLR run, we removed the linear trend of the observed de-tided sea level time series at Duck, North Carolina. The detrended non-tidal sea level component was then added to the tidal sea level to produce the sea level time series at the offshore boundary:

$$SL_{detrend} = H_{Duck} - Slope_H \times (Time - 1/1/1985) + SL_{tide} \quad (3)$$

where H_{Duck} represents the de-tided sea level time series at Duck, N.C., $Slope_H$ is the Theil-Sen slope obtained from the de-tided time series at Duck, SL_{tide} is the tidal sea level and $SL_{detrend}$ represents the detrended sea level.

Riverine nitrate and phosphate concentrations have decreased moderately due to nutrient management practices (Langland et al., 2007; Zhang et al., 2015) (Figure 2). At the same rate of discharge from the Susquehanna River, the nitrate loading was appreciably lower in 2001-2016 than in 1985-2000 (Figure 2a), as reported in Testa et al. (2018). To examine how the nutrient management affected the long-term trend in hypoxia, we conducted the DtrNut run in which the trends in the riverine nutrient concentration were removed (Figures 2b-c). Since hypoxia in

Chesapeake Bay is mostly caused by autochthonous primary production fueled by inorganic nutrients, only the time series of nitrate+nitrite and phosphate concentration from the Susquehanna River were detrended. Because the nutrient loads were highly influenced by streamflow and season, we applied the additive approach based on the flow-normalized nutrient concentration from USGS Chesapeake Bay Nontidal Monitoring Program (https://cbrim.er.usgs.gov). The monthly averaged maximum concentrations during 1985-1989 ("peak period") were estimated, and the differences in the nutrient concentration between this peak period and each year in 1985 to 2016 were calculated. These differences were then added to the original riverine nutrient concentration:

$$C_{detrend} = C_{original} + (CF_{peak\ period} - CF) \quad (4)$$

where $C_{detrend}$ and $C_{orginal}$ are the detrended and original riverine nutrient concentration respectively, CF is the flow-normalized nutrient concentration during 1985-2016, and $CF_{peak\ period}$ is the monthly averaged nutrient concentration during 1985-1989. Although the river flows displayed strong interannual variations, they showed no long term trend between 1985 and 2016 (Figure 2d). Therefore, we did not conduct a scenario run on the river flows.

3. Model-simulated long term changes and comparison with observations

Results from the hindcast model simulation (the Base run) of 1985-2016 are presented, with a focus on identifying long term trends in dissolved oxygen and hypoxia. To support the use of model simulations to quantify trends, we first evaluated the model-predicted sea level, temperature and O_2 in Chesapeake Bay against observations.

3.1 Sea level rise and warming

303 Model-predicted monthly averaged water levels at selected NOAA tidal gauging stations
304 (locations marked in Figure 1a) were compared with monthly averaged observations in Figure 3.
305 The model captured the seasonal and interannual sea level variations in the estuary. The Taylor
306 diagram provides a quantitative evaluation of the model's skill (Taylor, 2001). The correlation
307 coefficient between the predicted and observed monthly sea levels is around 0.95 and the
308 standard deviations of the predicted and observed sea level are close to each other (Figure 4a).
309 Mann-Kendall trend test indicates that both the modelled and observed sea levels show a
310 statistically significant upward trend, with small *p*-values (Table 1). The Theil-Sen estimator was
311 used to calculate these linear trends. The modelled sea levels at the tidal gauge stations rose at a
312 rate of 4.7-4.9 mm/year while the observed sea level rise rate varied from 4.3 to 6.6 mm/year
313 between 1985 and 2016. The ROMS model was forced by the offshore sea level at Duck, North
314 Carolina, which rose at a rate of 4.5 mm/year. Land subsidence associated with glacial isostatic
315 adjustment (Engelhart et al., 2009; Miller et al. 2013) and ground water extraction produced non-
316 uniform relative sea level rise rates in Chesapeake Bay (Boesch et al., 2018). This was not
317 considered in our model runs. Nevertheless, the difference between the predicted and observed
318 long-term trend over the past three decades was relatively small and should not significantly
319 affect hypoxia simulations. The root-mean-square-error was 3-4 cm (Table 1).
320
321 Next we compared the predicted and observed surface and bottom water temperature at a
322 few stations arrayed along the center deep channel of the Bay (Figure 1a, Figure 5). The CBP
323 monitoring cruises collected temperature measurements at bi-weekly or monthly intervals. These
324 data are directly compared with hourly model output, revealing that the model captured the
325 observed temperature time series in both surface and bottom waters (Figures 4b and 4c). Both the

long-term annual averages from observations and model simulation show clear increasing trends from 1985 to 2016 (Figure 5). Over the 32-year time period, the observed and modelled annual-mean surface water temperature increased by 0.9-1.95 °C, with the p-value generally less than 0.05 (Table 2). The bottom temperature increase was slightly smaller but was statistically significant at most stations.

Further analysis of regionally averaged temperature from the model shows that temperature increased in every sub-region of Chesapeake Bay and this upward trend was statistically significant, with the *p*-value less than 0.01 (Table 3). Temperature in the deep mid-bay rose most: ~1.8 °C at the surface and ~1.6 °C in the bottom. In comparison, temperature in the surface and bottom waters of shallow and upper bay increased by ~1.2 °C. Surface and bottom salinity increased ~(0.2-0.3) psu between 1985 and 2016, consistent with the retrospective data analysis by Hilton et al. (2008) (Table 3). The upward trend is statistically significant, with small *p*-values.

3.2. Long-term changes in O_2 and hypoxic volume

Figure 6 shows the time series of the modelled and observed O_2 concentration in the surface and bottom waters at the CBP monitoring stations between 1985 and 2016. The model outputs were saved at 4-hourly intervals while the sampling data were collected at bi-weekly or weekly intervals. The model captured the seasonal cycle of dissolved oxygen as well as the interannual variations (Figures 4d and 4e). For the surface O_2, the correlation coefficient ranges from 0.70 to 0.95 and the normalized standard deviations is around 1 except that it falls to 0.85 at station CB 4.1C. For the bottom O_2, the correlation coefficient ranges from 0.87 to 0.95 and the

normalized standard deviations is around 1 at stations CB 3.1 and CB 4.1C but is 1.1/1.3 at stations CB 5.2/CB 6.2 where the model-predicted variance is larger than the observed. At station CB 6.2, the model did not always capture the observed seasonal minimum, which occasionally dropped below 2 mg/L.

To identify the long term O_2 trends, we averaged O_2 concentration over the four subregions in the Bay and obtained the regionally averaged O_2 concentrations for both the surface and bottom waters. The GAM model was used to fit these time series, as illustrated in Figures 7a and 7b. GAM fits the data with high skill, with the adjusted regression coefficient R^2 ranging from 0.91 to 0.96. After the seasonal cycle was removed, the residuals displayed large interannual variations, as reported in previous studies (e.g. Li et al., 2016). However, O_2 declined in all subregions (Table 4). The O_2 in the two mid-bay subregions declined by ~0.3 mg/L on average over the past three decades. The reduction was smaller (0.1 – 0.2 mg/L) in the upper and lower bay. For comparison, the O_2 saturation concentration decreased by 0.20-0.33 mg/L over the same period. Mann-Kendall tests were conducted to determine the statistical significance of the long term trends in the time series of the regionally averaged O_2. All the O_2 decline trends were statistically significant, with the *p*-value much less 0.01 (Table 4).

Monthly averaged hypoxic volumes (with O_2 concentration less 2 mg/L) were calculated for May-September, and their variations in 1985-2016 are shown in Figure 8. The non-parametric MK test was applied to the time series of the hypoxic volume to identify possible monotonic long term trends. The Sen's slope in the September hypoxic volume was -0.028 km^3/yr, resulting in a total reduction of 0.90 km^3 over the 32-year period. In comparison, the

372 Sen's slope in the July hypoxic volume was 0.014 km^3/yr, amounting to an increase of 0.46 km^3.
373 The Sen's slope in May, June and August hypoxic volume was less than 0.01, indicating no
374 apparent long term trend. However, the MK test showed that none of these trends are statistically
375 significant, with the *p*-value exceeding 0.05. This suggests that the long-term trend in the
376 hypoxic volume was insignificant when compared with the large interannual variations.
377
378 The apparent contradiction between Figures 7 and 8 motivated us to examine the seasonal
379 averaged O_2, as shown for the bottom O_2 in the upper mid-bay where hypoxia generally occurred
380 (Figure 9). O_2 declined by 0.61 mg/L in winter and by 0.54 mg/L in spring between 1985 and
381 2016. In contrast, O_2 declined by 0.35 mg/L in summer but increased by 0.13 mg/L in fall over
382 the same period. In summary, O_2 declined by ~0.3 mg/L over the past 3 decades, but at much
383 faster rates during the winter and spring seasons. On the other hand, O_2 increased slightly in the
384 fall, resulting in a smaller hypoxic volume.
385

386 **4. Model-scenario analyses to discern driving mechanisms**

387
388 To discern the roles of warming, sea level rise and nutrient management in driving the
389 long term trends in hypoxia in Chesapeake Bay, we analyzed the three scenario model runs
390 DtrTEMP, DtrSLR and DtrNut in which the effects of warming, sea level rise and nutrient
391 reduction were removed.
392
393 The long-term term in the GAM fit to the regionally averaged surface and bottom O_2,
394 namely *s(dnum)*, is shown in Figure 10. First, in all the four subregions, there were no detectable
395 differences in *s(dnum)* between the Base run and the DtrSLR run in surface O_2 (except for small

differences in the upper bay). This suggests that removing sea level rise had little effects on the long-term trend in the Bay. Second, the DtrTEMP run exhibited the largest difference from the Base run. For example, in the DtrTEMP run, surface O_2 showed no trends at all in the upper bay and upper mid-bay (Figures 10a, 10c). It increased slightly in the lower mid-bay and the lower bay between 1985 and 1995 and decreased slightly between 1995 and 2016, but the net reduction over the 32 years was considerably smaller than the Base run. Bottom O_2 in the DtrTEMP run showed little trend in the upper bay and in the lower bay (Figures 10b, 10h). In the bottom water of the two mid-bay subregions, O_2 in the DtrTEMP run trended upwards in 1990-2000, downwards in 2000-2010 and upwards again after 2010, but did not display a definitive direction for change (Figures 10d, 10f). In comparison, *s(dnum)* in the Base run displayed an overall declining trend over the 32 years, even though it showed low-frequency variations. Third, the DtrNut run showed much smaller differences from the Base run than from the DtrTEMP run. The departure of O_2 trend in DtrNut from Base run was in the same direction as DtrTEMP in surface water (Figures 10a, 10c), while they were generally opposite in the bottom water in most regions of the Bay (Figures 10d, 10f). It means that without nutrient reduction, O_2 experienced slightly smaller declines in surface water but slightly larger declines in bottom water.

Another way to tease out the individual effects of sea level rise, warming and nutrient reduction is to compare the regionally averaged O_2 in bottom water between the early and latter stages of the three decades. O_2 was averaged over two decadal periods: 1985-1994 (Period 1) and 2007-2016 (Period 2). The reason we chose 10-year average was to minimize the influence of interannual variability. Figures 11a-d show the difference $\Delta O_2 = O_2^{\text{Period2}} - O_2^{\text{Period1}}$ between Period 1 and Period 2 from the Base run and three scenario runs. In the Base run, O_2 decreased

419 during most of times in a year, except during late summer and early Fall (August-October) when
420 O_2 showed a slight increase (< 0.1 mg/L) in the middle and lower bay. This is consistent with the
421 declining September hypoxic volume shown in Figure 8. In contrast, ΔO_2 was negative in other
422 seasons, reaching (-0.2 to -0.4) mg/L in the lower and upper bay and (-0.5 to -0.7) mg/L in the
423 two mid-bay regions. It is interesting to note that the O_2 reduction was considerably larger during
424 winter and spring. This resolves the apparent contradiction between Figures 7 and 8 because the
425 largest O_2 reduction occurred during the non-hypoxic seasons.
426
427 We further compared ΔO_2 between Period 1 and Period 2 in the three scenario runs
428 (Figures 11a-d). Compared with the Base run, ΔO_2 in the upper bay from the DtrSLR run
429 exhibited overall upward shift, indicating a smaller decline in winter-summer and a larger gain in
430 fall. This suggests that sea level rise contributed to O_2 decline even though it effect was relatively
431 small. ΔO_2 in the mid and lower bay regions did not differ from the Base run. O_2 decline in the
432 upper bay from the DtrNut run was smaller in the winter but larger in late spring when compared
433 with the Base run. In the middle and lower bay, summer O_2 reduction was larger in the DtrNut
434 run than in the Base run, implying a larger O_2 decline without nutrient reduction. The most
435 dramatic difference with the Base run still came from the DtrTEMP run, where the O_2 decline
436 during winter to early spring was substantially reduced all over the bay. ΔO_2 even became
437 positive from late spring to fall, implying that the bottom O_2 concentration would have increased
438 without warming. Moreover, bottom O_2 in the upper bay decreased in late winter and summer
439 but increased in spring and fall.
440

441 Since the O_2 concentration was similar between the Base run and scenario runs in Period
442 1, a comparison of the O_2 concentration between the Base and scenario runs in Period 2 shed
443 further light on the effects of the individual forcing (Figures 11e-h). Clearly bottom O_2 in the
444 DtrTEMP run was much higher than the Base run throughout the year, with the largest difference
445 in winter-spring and the smallest difference in fall. The surplus reached 0.5-0.7 mg/L in the
446 upper bay and mid-bay and 0.3 mg/L in the lower bay. This clearly demonstrated that bottom O_2
447 would be considerably higher without warming. It was also verified by the considerably smaller
448 bottom O_2 trend in the DtrTEMP run (Table 4). Bottom O_2 in the DtrNut run was somewhat
449 lower than the Base run in the mid-bay, indicating that O_2 concentration would be ~0.1 mg/L
450 lower without nutrient reduction. Bottom O_2 was greater in the DtrSLR run than in the Base run,
451 particularly in the upper bay, indicating that O_2 would be higher without sea level rise. In
452 conclusion, the effect of warming was strongest during winter-spring that predominately drove
453 the decline of bottom O_2. It was one order of magnitude larger than the effects of sea level rise
454 and nutrient reduction. Both warming and sea level rise caused the decline of bottom O_2, while
455 nutrient loads reduction led to the recovery of bottom O_2 during summer in the middle bay.
456
457 In addition to O_2 concentration and hypoxic volume, we examined whether the onset and
458 termination of hypoxia shifted over the past three decades. Using the mid-bay station CB4.1C as
459 an example, we calculated the day of a year (hypoxia initiation day, T_{ini}) when O_2 first fell below
460 2 mg/L in spring and the day of a year (hypoxia termination day, T_{ter}) when O_2 rose above 2
461 mg/L in fall (Figure 12). There were large scatters in T_{ini} spanning from late April to early June,
462 although T_{ini} most often occurred in the month of May (Figure 12a). According to the MK trend
463 test, T_{ini} showed no long term trend in the Base run. No trend in T_{ini} was detected either in the

464 DtrSLR and DtrNut runs. On the other hand, T_{ini} in the DtrTemp had an appreciable upward
465 trend, with the Sen's slope of 1.8 days/decade. This suggests that the onset of hypoxia would
466 have been delayed by ~6 days without the warming. This result can be seen more clearly when
467 we plot the difference in the hypoxia initiation day ΔT_{ini} between the three scenario runs and the
468 Base run (Figure 12b). ΔT_{ini} was nearly zero in the DtrSRL run, indicating that sea level rise had
469 no effects on hypoxia initiation. ΔT_{ini} was slightly negative (up to 5 days after 2005, with an
470 outlier of 10 days in 2012) in the DtrNut run, indicating that nutrient reduction delayed hypoxia
471 initiation in recent years. More strikingly, ΔT_{ini} in the DtrTemp run reached 5-10 days between
472 2000 and 2016, suggesting that warming caused hypoxia to develop 5-10 days earlier in spring.
473
474 There were also substantial scatters in the hypoxia termination day T_{ter} (Figure 12c). The
475 MK trend test showed that T_{ter} shifted earlier in the Base run as well as in the three scenario runs
476 if year 2003 and 2011 were removed due to the passage of Hurricane Isabel (2003) and Tropical
477 Storm Lee (2011) in fall. We also calculated the difference in the hypoxia termination day ΔT_{ter}
478 between the three scenario runs and the Base run (Figure 12d). Once again ΔT_{ter} was nearly zero
479 in the DtrSLR run, indicative of no influence from the sea level rise. ΔT_{ter} in both the DtrTemp
480 and DtrNut runs was positive but small. This shows that warming and nutrient reduction had
481 less influence on the hypoxia termination than the hypoxia initiation, although both drove an
482 earlier termination by several days.
483

484 **5. Discussion and conclusion**
485

To investigate long term changes in O_2 in Chesapeake Bay, we used a coupled hydrodynamic-biogeochemical (ROMS-RCA) model to conduct hindcast simulations between 1985 and 2016. The model provided high frequency (4-hourly) and fine resolution (1 km, 20 vertical layers) outputs of the three-dimensional fields of O_2 and other physical/biogeochemical state variables. ROMS-RCA accurately captured the observed O_2 time series at the CBP monitoring stations, with the correlation coefficient around 0.9 and the normalized standard deviation in the range of (0.9-1.1). This enabled a more robust statistical analysis (using model output) than what could have been achieved using sparse water quality data (biweekly or monthly intervals and 41 stations distributed in the main stem) collected at the CBP monitoring stations. The model reproduced the overall long-term O_2 trend in all Bay regions, and after the seasonal and interannual variations are removed, the dissolved oxygen concentration in all regions of the estuary showed a statistically significant downward trend: decreasing ~0.3 mg/L over the past three decades.

The majority of the modeled O_2 decline occurred during the winter and spring seasons, with a magnitude of (0.5-0.6) mg/L (and before hypoxic conditions develop). Despite these early season declines that could potentially increase the vulnerability to early season hypoxia, the hypoxic volume in May-August showed little change over three decades. However, September O_2 concentrations increased by 0.13 mg/L, the September hypoxic volume showed a slight increase (~0.9 km^3), and the timing of hypoxia breakup shifted earlier in the fall. This finding is consistent with previous findings, including a retrospective data analysis by Murphy et al. (2011) that found a moderate decrease in late summer hypoxia, and the data analysis by Zhou et al. (2014) and Testa et al. (2018) who concluded that the end of hypoxic period shifted earlier over

509 the past three decades. Both our model results and previous observational studies found no
510 significant trend in the timing of hypoxia onset in the Chesapeake Bay, although an increase of
511 early summer hypoxia volume was detected in retrospective data analysis (Zhou et al., 2014;
512 Murphy et al., 2011; Testa et al., 2018).
513
514 The numerical model also afforded an opportunity to investigate the underlying
515 mechanisms that contributed to long term changes in O_2. We conducted additional numerical
516 experiments in which we removed trends in each long term driver to discern the separate effects
517 of temperature increases, sea level rise, and nutrient reduction. Warming was found to be the
518 dominant driver of the long-term oxygen decline whereas sea level rise had a minor effect
519 (Figures 10-12). Without warming, O_2 concentrations in all regions of the estuary showed no or
520 weak long term trends over the past three decades (Figure 10 and Table 4). The magnitude of O_2
521 reduction in the Base run slightly exceeded the solubility effects in the surface water of mid-bay
522 regions and in the bottom water of lower mid-bay (Table 4). This suggested that in addition to
523 the change in oxygen saturation, warming also affected O_2 via biological processes in the water
524 column. Between January and June, warming resulted in ~4% larger water column respiration in
525 the Base run than in the DtrTemp run and hence resulted in lower O_2 (Figure 13). Between July
526 and September, warming generated lower rates of water column respiration because of the faster
527 depletion of organic matter in earlier summer (Basenback, 2019) and reduced primary
528 production. This metabolic response counteracted the solubility effect, leading to smaller O_2
529 reductions in summer than in winter and spring, and even an increase in late summer (Figures
530 11a-d) that is consistent with the September oxygen increases. A further comparison between the
531 Base run and DtrTemp showed that warming would lead to an earlier initialization of hypoxia in

532 the spring and an earlier breakup of hypoxia in the fall (Figure 12). The warming induced change
533 in water column respiration (Figure 13) is consistent with the hypothesis in Testa et al. (2018)
534 that warming caused a shift in phenology and resulted in the "speeding-up" of organic matter
535 consumption and biogeochemical cycle. It is also consistent with empirical studies that found
536 mid-Bay sediment respiration to be vulnerable to late summer organic matter limitation (Boynton
537 and Kemp 2008), which warming would exacerbate.
538
539 Since warming played a central role in driving long term O_2 trends, we further examined
540 how the model-predicted warming trends compared with the retrospective analyses of long term
541 temperature measurements in Chesapeake Bay. In a comparison on modeled and observation-
542 based temperature increase rates at four monitoring stations (Table 2), the direction of warming
543 trends were similar across the estimates, but there are quantitative differences in the magnitude
544 of the trend slope. For example, the warming trend in bottom temperature measured at CB4.1C
545 had a *p*-value of 0.1 and was not statistically significant. At CB5.2, the Sen's slope for the
546 observed surface/bottom temperature was 0.88/0.75 ^{o}C/32yr whereas the Sens's slope for the
547 modelled temperature time series was 1.95/1.43 ^{o}C/32yr. The differences at other stations were
548 smaller but still substantial. A couple of factors may have contributed to these differences.
549 Temperature data at the monitoring stations were collected at bi-weekly or monthly intervals but
550 not at fixed times every year, making it hard to discern long term trends from large seasonal and
551 diurnal variations. For example, the average diurnal variations of observed high-frequency
552 surface temperature measurements at Gooses Reef Buoy in the mid-bay
553 (https://buoybay.noaa.gov) during 2011-2016 ranged from 0.65 ^{o}C in the winter to 1.87 ^{o}C in the
554 early summer. The ROMS model was forced by the NARR reanalysis products at the sea

surface (e.g., air temperature, wind speed, barometric pressure, total shortwave radiation), which have coarse resolution and may not resolve the air-sea heat fluxes across the surface of Chesapeake Bay with high precision. However, the model-predicted temperature increase of ~1.5°C over the past three decades was well within the range of the reported temperature increases in the Bay (Table S1). Retrospective analyses of long-term temperature records at a number of locations in the estuary, including the main stem, tributaries and rivers, showed that the warming rate fell onto (0.2 – 0.8) °C/decade, which were equivalent to 0.6-2.7°C/32 years (Table S1). However, these warming rates varied temporally at different time periods and spatially from the freshwater regions to the ocean. In the study focusing on main stem of Chesapeake Bay over recent decades, Ding and Elmore (2015) analyzed the remote sensing observations with finer temporal resolution than CBP measurements and found that the surface water temperature increased by ~0.5-1.0°C/decade (1.6-3.2°C/32 years) from 1984 to 2010. The estimation from our model results mostly fell into this range (Table 3). However, the other in-situ measurement based studies suggested relatively smaller warming rate (<1.0 °C/32 years) than Ding and Elmore (2015).

To address the uncertainty in predicting the long-term temperature trends and its effects on hypoxia simulation, we conducted another model run in which the warming rate was set 0.2°C/decade lower than the Base run, i.e., the water temperature only increased by ~1.0°C between 1985 and 2016. The average O_2 reduction was smaller by < 0.2 mg/L during 2007-2016 in this test run compared to Base run (Figure S1). Nevertheless, it is much smaller than the difference between the Base run and DtrTemp run (up to 0.5 mg/L) (Figures 11e-h). In other words, the O_2 decline caused by the hypothetically smaller temperature increase would still

outweigh the O_2 changes induced by nutrient reduction and SLR. Therefore, the conclusion that warming effects exceeded nutrient reduction effects and the overall model predictions of long-term O_2 trends were not overly sensitive to uncertainty in the temperature prediction. The conclusions of our study are focused on a well-validated model's response to data-constrained external perturbations. Despite the known limitations of such modeling approaches, their predictions represent a consistent and mechanistic response to long-term changes and an alternative to long-term observations that have their own limitations. Because neither approach is a perfect representation of the estuarine environment, utilizing the two methods in concert helps build confidence in our understanding of controls on hypoxia.

Chesapeake Bay has been subject to extensive efforts to reduce watershed nutrient inputs since the 1980s. In 1987, a commitment was made to reduce controllable sources of both nitrogen (N) and phosphorous (P) by 40% by the year 2000. Although the actual implementation fell short of the goals, there were appreciable declines of riverine dissolved nitrogen load, as shown in Figure 2 and reported earlier in Zhang et al. (2015), Testa et al. (2018), and Harding et al. (2015). However, the model results presented here suggested that this nutrient reduction only played a minor role in driving the long-term O_2 trend over the past three decades. The bottom O_2 increased by <0.1 mg/L in the middle parts of the Bay, which was about one third of the O_2 increase due to warming (Figure 11). Both our model results and previous observational studies (Testa et al., 2018) showed no significant trend in the timing of hypoxia onset. The model scenario analyses suggested that the effect of nutrient reduction might offset the potential earlier hypoxia caused by warming (Figure 11c). Interestingly, nutrient reduction worked in concert with warming to cause reduced hypoxic volume and earlier termination of hypoxia in the fall, as

601 shown in Figure 11d and Figure 8e. To our knowledge, this seasonally-dependent response of
602 hypoxia to changes in warming and nutrient load has not been reported for other coastal
603 ecosystem experiencing hypoxia.
604
605 A major finding of this study is that warming (~0.05 °C yr^{-1}) due to climate change was a
606 more important driver of the long-term O_2 trend (~0.01 mg L^{-1} yr^{-1}) than nutrient management
607 over the past 3 decades. Similar findings have been reported in other estuaries and seas. In the
608 Baltic Sea, O_2 continues to decline over the past two decades even though riverine nitrogen and
609 phosphorus loads have reduced (Carstensen et al., 2014; Andersen et al., 2017). At the Boknis
610 Eck station in the Baltic Sea, Lennartz et al. (2014) observed an oxygen decline rate of ~0.015
611 mg L^{-1} yr^{-1} and attributed it to decreased oxygen solubility and increased respiration associated
612 with warming. In the northern Gulf of Mexico, the total hypoxic area has not shown any
613 significant long-term trend since 1985 despite the implementation of the nutrient management
614 strategy (Kemp et al., 2009; Conley et al., 2009c; Obenour et al., 2013). In Long Island Sound,
615 the implementation of TMDL management has resulted in ~40% reduction of the hypoxia area
616 since 2000 (Klee, 2017). However, O_2 at some monitoring stations oxygen still decreased at a
617 high rate of ~0.03 mg $L^{-1}yr^{-1}$, associated with rapid warming ~0.08 °C yr^{-1} (Staniec and Vlahos,
618 2017).
619
620 Although our finding that contemporary climate change has cancelled out – or largely
621 compensated for - potential benefits of nutrient management over the past three decades,
622 watershed management of nutrient and sediment pollutants has clearly contributed to the
623 recovery of some aspects Chesapeake Bay. For example, large-scale restoration of submersed

aquatic vegetation (Lefcheck et al., 2018) has been reported, as well as the improvement of regional water quality conditions (Zhang et al., 2018), including O_2. Some of these improvements may have resulted from upgrades to wastewater treatment plants discharging into tidal waters (Boynton et al. 2014), or reductions in direct atmospheric deposition, but we did not address those impacts on hypoxia. In the absence of warming, our simulations suggest that summer hypoxic volume would have decreased by 0.9 km^3 and hypoxia duration would be shortened by ~9 days (data not shown) over the past three decades. Despite warming, riverine nutrient load reductions facilitated an earlier termination of hypoxia and elevated bottom water O_2 in late summer water. Thus, in order to achieve the goal for hypoxia mitigation in the Chesapeake Bay in the context of climate change, additional watershed nutrient management will be required.

Acknowledgments. We are grateful to NSF (CBET-1360285) and NOAA Ocean Acidification Program (NOAA-OAP; Award NA15NOS4780184) for the financial support. This is UMCES contribution number xxxx.

References:

Andersen, J. H., Carstensen, J., Conley, D. J., Dromph, K., Fleming, Lehtinen, V., Gustafsson, B. G., ... & Murray, C. (2017). Long-term temporal and spatial trends in eutrophication status of the Baltic Sea. Biological Reviews, 92(1), 135-149.

Altieri, A. H. and Gedan, K. B. (2015). Climate change and dead zones. Global Change Biology, 21(4), pp. 1395–1406. doi: 10.1111/gcb.12754.

Basenback, N. (2019). Phenology of estuarine response to anthropogenic and climate drivers, a study of the Chesapeake Bay and Chester River Estuaries. University of Maryland College Park, Master Thesis.

Beck, M. W., & Murphy, R. R. (2017). Numerical and qualitative contrasts of two statistical models for water quality change in tidal waters. JAWRA Journal of the American Water Resources Association, 53(1), 197-219.

Bever, A. J., Friedrichs, M. A., Friedrichs, C. T., Scully, M. E., & Lanerolle, L. W. (2013). Combining observations and numerical model results to improve estimates of hypoxic volume within the Chesapeake Bay, USA. Journal of Geophysical Research: Oceans, 118(10), 4924-4944.

Boesch, D.F., W.C. Boicourt, R.I. Cullather, T. Ezer, G.E. Galloway, Jr., Z.P. Johnson, K.H. Kilbourne, M.L. Kirwan, R.E. Kopp, S. Land, M. Li, W. Nardin, C.K. Sommerfield, W.V. Sweet. 2018. Sea-level Rise: Projections for Maryland 2018, 27 pp. University of Maryland Center for Environmental Science, Cambridge, MD.

Boon, J. D., & Mitchell, M. (2015). Nonlinear change in sea level observed at North American tide stations. Journal of Coastal Research, 31(6), 1295-1305.

Borja, Á., Elliott, M., Carstensen, J., Heiskanen, A. S., & van de Bund, W. (2010). Marine management–towards an integrated implementation of the European Marine Strategy Framework and the Water Framework Directives. Marine pollution bulletin, 60(12), 2175-2186.

Boynton, W. R., Hodgkins, C. L. S., O'Leary, C. A., Bailey, E. M., Bayard, A. R., & Wainger, L. A. (2014). Multi-decade responses of a tidal creek system to nutrient load reductions: Mattawoman Creek, Maryland USA. Estuaries and coasts, 37(1), 111-127.

671 Boynton, W.R. & Kemp, W.M. (2008). Estuaries. In Nitrogen in the marine environment, ed.
672 D.G. Capone, D.A. Bronk, M.R. Mulholland and E.J. Carpenter, 809-866. Amsterdam:
673 Elsevier.
674 Brady, D. K. (1976). Are the Chesapeake Bay waters warming up? Chesapeake Science, 225-227.
675 Brady, D.C., Testa, J. M., Di Toro, D.M., Boynton, W. R., and Kemp, W. M. (2013). Sediment
676 flux modeling: Calibration and application for coastal systems. Estuarine, Coastal and
677 Shelf Science 117: 107-124. https://doi.org/10.1016/j.ecss.2012.11.003
678 Breitburg, D., Levin, L. A., Oschlies, A., Grégoire, M., Chavez, F. P., Conley, D. J., ... & Jacinto,
679 G. S. (2018). Declining oxygen in the global ocean and coastal
680 waters. Science, 359(6371), eaam7240.
681 Carstensen, J., Andersen, J. H., Gustafsson, B. G., & Conley, D. J. (2014). Deoxygenation of the
682 Baltic Sea during the last century. Proceedings of the National Academy of
683 Sciences, 111(15), 5628-5633.
684 Conley, D. J., Bonsdorff, E., Carstensen, J., Destouni, G., Gustafsson, B. G., Hansson, L. A.,
685 Rabalais, N. N., Voss, M. and Zillén, L. (2009a). Tackling hypoxia in the Baltic Sea: Is
686 engineering a solution? Environmental Science and Technology, 43(10), pp. 3407–3411.
687 doi: 10.1021/es8027633.
688 Conley, D. J., Bjorck, S., Bonsdorff, E., Carstensen, J., Destouni, G., Gustafsson, B. G., ... &
689 Müller-Karulis, B. (2009b). Hypoxia-related processes in the Baltic Sea. Environmental
690 Science & Technology, 43(10), 3412-3420.
691 Di Toro, D. M. (2001), Sediment flux modeling, 624 pp., Wiley-Interscience, New York.
692 Diaz, R. J. and Rosenberg, R. (2008). Spreading dead zones and consequences for marine
693 ecosystems. Science, 321(5891), pp. 926–929. doi: 10.1126/science.1156401.

694 Ding, H., & Elmore, A. J. (2015). Spatio-temporal patterns in water surface temperature from
695 Landsat time series data in the Chesapeake Bay, USA. Remote Sensing of
696 Environment, 168, 335-348.
697 Du, J., Shen, J., Park, K., Wang, Y. P. and Yu, X. (2018). Worsened physical condition due to
698 climate change contributes to the increasing hypoxia in Chesapeake Bay. Science of the
699 Total Environment, 630, pp. 707–717. doi: 10.1016/j.scitotenv.2018.02.265.
700 Duarte, C. M., Conley, D. J., Carstensen, J. and Sánchez-Camacho, M. (2009). Return to
701 Neverland: Shifting baselines affect eutrophication restoration targets. Estuaries and
702 Coasts, 32(1), pp. 29–36. doi: 10.1007/s12237-008-9111-2.
703 Egbert, G. D., & Erofeeva, S. Y. (2002). Efficient inverse modeling of barotropic ocean
704 tides. Journal of Atmospheric and Oceanic Technology, 19(2), 183-204.
705 Engelhart SE, Horton BP, Douglas BC, Peltier WR, Törnqvist TE (2009) Spatial variability of
706 late Holocene and 20th century sea level-rise along the Atlantic coast of the United State.
707 Geology,37: 1115-18.
708 Fairall, C. W., Bradley, E. F., Hare, J. E., Grachev, A. A., & Edson, J. B. (2003). Bulk
709 parameterization of air–sea fluxes: Updates and verification for the COARE
710 algorithm. Journal of climate, 16(4), 571-591.
711 Feng, Y., DiMarco, S. F., & Jackson, G. A. (2012). Relative role of wind forcing and riverine
712 nutrient input on the extent of hypoxia in the northern Gulf of Mexico. Geophysical
713 Research Letters, 39(9).
714 Fennel, K. and Testa, J. M. (2019). Biogeochemical Controls on Coastal Hypoxia. Annual
715 Review of Marine Science, 11(1), pp. 105–130. doi: 10.1146/annurev-marine-010318-
716 095138.

717 Filippino, K. C., Mulholland, M. R., & Bernhardt, P. W. (2011). Nitrogen uptake and primary
718 productivity rates in the Mid-Atlantic Bight (MAB). Estuarine, Coastal and Shelf
719 Science, 91(1), 13-23.
720 Giani, M., Djakovac, T., Degobbis, D., Cozzi, S., Solidoro, C., & Umani, S. F. (2012). Recent
721 changes in the marine ecosystems of the northern Adriatic Sea. Estuarine, Coastal and
722 Shelf Science, 115, 1-13.
723 Gilbert, R.O. 1987. Statistical Methods for Environmental Pollution Monitoring, Wiley, NY.
724 Greening, H., Janicki, A., Sherwood, E. T., Pribble, R. and Johansson, J. O. R. (2014).
725 Ecosystem responses to long-term nutrient management in an urban estuary: Tampa Bay,
726 Florida, USA. Estuarine, Coastal and Shelf Science. Elsevier Ltd, 151, pp. A1–A16. doi:
727 10.1016/j.ecss.2014.10.003.
728 Hagy, J. D., Boynton, W. R., Keefe, C. W. and Wood, K. V. (2004). Hypoxia in Chesapeake Bay,
729 1950-2001: Long-term change in relation to nutrient loading and river flow. Estuaries,
730 27(4), pp. 634–658. doi: 10.1007/BF02907650.
731 Haidvogel, D. B., Arango, H., Budgell, W. P., Cornuelle, B. D., Curchitser, E., Di Lorenzo, E., ...
732 & Levin, J. (2008). Ocean forecasting in terrain-following coordinates: Formulation and
733 skill assessment of the Regional Ocean Modeling System. Journal of Computational
734 Physics, 227(7), 3595-3624.
735 Harding, L. W., Gallegos, C. L., Perry, E. S., Miller, W. D., Adolf, J. E., Mallonee, M. E. and
736 Paerl, H. W. (2016) . Long-Term Trends of Nutrients and Phytoplankton in Chesapeake
737 Bay. Estuaries and Coasts, 39(3), pp. 664–681. doi: 10.1007/s12237-015-0023-7.
738 Hastie, T. & Tibshirani, R. (1986). Generalized additive models (with discussion). Statistical
739 Science, 1, 297-318.

740 Hastie, T. & Tibshirani, R. (1990). Generalized Additive Models. Chapman & Hall.
741 HELCOM (2007). HELCOM Baltic Sea action plan. Krakow, Poland, 15, 2007.
742 Hetland, R. D., & DiMarco, S. F. (2008). How does the character of oxygen demand control the
743 structure of hypoxia on the Texas–Louisiana continental shelf? Journal of Marine
744 Systems, 70(1-2), 49-62.
745 Hilton, T. W., Najjar, R. G., Zhong, L., & Li, M. (2008). Is there a signal of sea-level rise in
746 Chesapeake Bay salinity?. Journal of Geophysical Research: Oceans, 113(C9).
747 Irby, I. D., Friedrichs, M. A., Da, F., & Hinson, K. E. (2018). The competing impacts of climate
748 change and nutrient reductions on dissolved oxygen in Chesapeake Bay. Biogeosciences,
749 15, 2649–2668.
750 Isleib, R. P. E., J. J. Fitzpatrick, and J. Mueller (2007). The development of a nitrogen control
751 plan for a highly urbanized tidal embayment, Proceedings of the Water Environment
752 Federation, 5, 296–320.
753 Justić, D., Rabalais, N. N., & Turner, R. E. (2003). Simulated responses of the Gulf of Mexico
754 hypoxia to variations in climate and anthropogenic nutrient loading. Journal of Marine
755 Systems, 42(3-4), 115-126.
756 Kaushal, S. S., Likens, G. E., Jaworski, N. A., Pace, M. L., Sides, A. M., Seekell, D., ... &
757 Wingate, R. L. (2010). Rising stream and river temperatures in the United States.
758 Frontiers in Ecology and the Environment, 8(9), 461-466.
759 Kemp, W. M., Testa, J. M., Conley, D. J., Gilbert, D. and Hagy, J. D. (2009). Temporal
760 responses of coastal hypoxia to nutrient loading and physical controls. Biogeosciences,
761 6(12), pp. 2985–3008. doi: 10.5194/bg-6-2985-2009.
762 Kendall, M.G. 1975. Rank Correlation Methods, 4th edition, Charles Griffin, London.

763 Lake, S. J. and Brush, M. J. (2015). Modeling estuarine response to load reductions in a warmer
764 climate: York River Estuary, Virginia, USA. Marine Ecology Progress Series, 538(Ipcc
765 2007), pp. 81–98. doi: 10.3354/meps11448.
766 Langland, M. J., Moyer, D., & Blomquist, J. (2007). Changes in streamflow, concentrations, and
767 loads in selected nontidal basins in the Chesapeake Bay Watershed, 1985-2006 (No.
768 2007-1372). US Geological Survey.
769 Lefcheck, J. S., Orth, R. J., Dennison, W. C., Wilcox, D. J., Murphy, R. R., Keisman, J., Gurbisz,
770 C., Hannam, M., Brooke Landry, J., Moore, K. A., Patrick, C. J., Testa, J., Weller, D. E.
771 and Batiuk, R. A. (2018). Long-term nutrient reductions lead to the unprecedented
772 recovery of a temperate coastal region. Proceedings of the National Academy of
773 Sciences of the United States of America, 115(14), pp. 3658–3662. doi:
774 10.1073/pnas.1715798115.
775 Lennartz, S. T., Lehmann, A., Herrford, J., Malien, F., Hansen, H. P., Biester, H., & Bange, H.
776 W. (2014). Long-term trends at the Time Series Station Boknis Eck (Baltic Sea), 1957–
777 2013: does climate change counteract the decline in eutrophication?. Biogeosciences
778 (BG), 11, 6323-6339.
779 Li, M., Lee, Y. J., Testa, J. M., Li, Y., Ni, W., Kemp, W. M., & Di Toro, D. M. (2016). What
780 drives interannual variability of hypoxia in Chesapeake Bay: Climate forcing versus
781 nutrient loading? Geophysical Research Letters, 43(5), 2127-2134.
782 Li, M., Zhong, L., & Boicourt, W. C. (2005). Simulations of Chesapeake Bay estuary: Sensitivity
783 to turbulence mixing parameterizations and comparison with observations. Journal of
784 Geophysical Research: Oceans, 110(C12).

785 Li, M., Zhong, L., Boicourt, W. C., Zhang, S., & Zhang, D. L. (2006). Hurricane-induced storm
786 surges, currents and destratification in a semi-enclosed bay. Geophysical Research
787 Letters, 33(2).
788 Klee, R.J. (2017). Long Island Sound Hypoxia Season Review. http://www.iec-
789 nynjct.org/node/71
790 Mann, H.B. 1945. Non-parametric tests against trend, Econometrica 13:163-171.
791 Mee, L. (2006). Reviving dead zones. Scientific American, 295(5), 78-85.
792 Meier, H. E. M., Höglund, A., Eilola, K. and Almroth-Rosell, E. (2017). Impact of accelerated
793 future global mean sea level rise on hypoxia in the Baltic Sea. Climate Dynamics.
794 Springer Berlin Heidelberg, 49(1–2), pp. 163–172. doi:10.1007/s00382-016-3333-y.
795 Meier, H. M., Andersson, H. C., Eilola, K., Gustafsson, B. G., Kuznetsov, I., Müller-Karulis,
796 B., ... & Savchuk, O. P. (2011). Hypoxia in future climates: A model ensemble study for
797 the Baltic Sea. Geophysical Research Letters, 38(24).
798 Meire, L., Soetaert, K. E. R., & Meysman, F. J. R. (2013). Impact of global change on coastal
799 oxygen dynamics and risk of hypoxia. Biogeosciences, 10(4), 2633-2653.
800 Mesinger, F., DiMego, G., Kalnay, E., Mitchell, K., Shafran, P. C., Ebisuzaki, W., ... & Ek, M. B.
801 (2006). North American regional reanalysis. Bulletin of the American Meteorological
802 Society, 87(3), 343-360.
803 Miller KG, Kopp RE, Horton BP, Browning JV, Kemp AC (2013) A geological perspective on
804 sea-level rise and its impacts along the US mid-Atlantic coast. Earth's Future,1(1): 3-18.
805 Mountain, D. G. (2003). Variability in the properties of Shelf Water in the Middle Atlantic Bight,
806 1977–1999. Journal of Geophysical Research: Oceans, 108(C1).

807 Murphy, R. R., Kemp, W. M. and Ball, W. P. (2011) . Long-Term Trends in Chesapeake Bay
808 Seasonal Hypoxia, Stratification, and Nutrient Loading. Estuaries and Coasts, 34(6), pp.
809 1293–1309. doi: 10.1007/s12237-011-9413-7.
810 Murphy, R. R., Perry, E., Harcum, J., & Keisman, J. (2019). A Generalized Additive Model
811 approach to evaluating water quality: Chesapeake Bay case study. Environmental
812 Modelling & Software, 118, 1-13.
813 Ni, W., Li, M., Ross, A. C., & Najjar, R. G. (2019). Large projected decline in dissolved oxygen
814 in a eutrophic estuary due to climate change. Journal of Geophysical Research:
815 Oceans, 124(11), 8271-8289.
816 Obenour, D. R., Scavia, D., Rabalais, N. N., Turner, R. E., & Michalak, A. M. (2013).
817 Retrospective analysis of midsummer hypoxic area and volume in the northern Gulf of
818 Mexico, 1985–2011. Environmental science & technology, 47(17), 9808-9815.
819 O'Donnell, J., Dam, H. G., Bohlen, W. F., Fitzgerald, W., Gay, P. S., Houk, A. E., ... & Howard-
820 Strobel, M. M. (2008). Intermittent ventilation in the hypoxic zone of western Long
821 Island Sound during the summer of 2004. Journal of Geophysical Research:
822 Oceans, 113(C9).
823 Preston, B. L. (2004). Observed winter warming of the Chesapeake Bay estuary (1949–2002):
824 implications for ecosystem management. Environmental Management, 34(1), 125-139.
825 Qian, W., Gan, J., Liu, J., He, B., Lu, Z., Guo, X., ... & Dai, M. (2018). Current status of
826 emerging hypoxia in a eutrophic estuary: the lower reach of the Pearl River Estuary,
827 China. Estuarine, Coastal and Shelf Science, 205, 58-67.

828 Rabalais, N. N., Díaz, R. J., Levin, L. A., Turner, R. E., Gilbert, D. and Zhang, J. (2010).
829 Dynamics and distribution of natural and human-caused hypoxia. Biogeosciences, 7(2),
830 pp. 585–619. doi: 10.5194/bg-7-585-2010.
831 Rice, K. C., & Jastram, J. D. (2015). Rising air and stream-water temperatures in Chesapeake
832 Bay region, USA. Climatic Change, 128(1-2), 127-138.
833 Saraiva, S., Meier, H. E. M., Andersson, H., Höglund, A., Dieterich, C., Gröger, M., Hordoir, R.
834 and Eilola, K. (2019). Baltic Sea ecosystem response to various nutrient load scenarios in
835 present and future climates. Climate Dynamics. Springer Berlin Heidelberg, 52(5), pp.
836 3369–3387. doi: 10.1007/s00382-018-4330-0.
837 Scavia, D., Bertani, I., Obenour, D. R., Turner, R. E., Forrest, D. R., & Katin, A. (2017).
838 Ensemble modeling informs hypoxia management in the northern Gulf of
839 Mexico. Proceedings of the National Academy of Sciences, 114(33), 8823-8828.
840 Scully, M. E. (2010). Wind modulation of dissolved oxygen in Chesapeake Bay. Estuaries and
841 coasts, 33(5), 1164-1175.
842 Scully, M. E. (2013). Physical controls on hypoxia in Chesapeake Bay: A numerical modeling
843 study. Journal of Geophysical Research: Oceans, 118(3), 1239-1256.
844 Scully, M. E. (2016). The contribution of physical processes to inter-annual variations of
845 hypoxia in Chesapeake Bay: A 30-yr modeling study. Limnology and Oceanography,
846 61(6), pp. 2243–2260. doi: 10.1002/lno.10372.
847 Sen, Pranab Kumar (1968). Estimates of the regression coefficient based on Kendall's tau.
848 Journal of the American Statistical Association, 63 (324): 1379-1389.

849 Shchepetkin, A. F., & McWilliams, J. C. (2005). The regional oceanic modeling system (ROMS):
850 a split-explicit, free-surface, topography-following-coordinate oceanic model. Ocean
851 modelling, 9(4), 347-404.
852 Shen, C., J.M. Testa, M. Li, W-J Cai, G.G. Waldbusser, W. Ni, W.M. Kemp, J. Cornwall, B.
853 Chen, J. Brodur, and J. Su. (2019a). Controls on carbonate system dynamics in a coastal
854 plain estuary: a modeling study. Journal of Geophysical Research - Biogeosciences, 124,
855 https://doi.org/10.1029/2018JG004802.
856 Shen, C., J. Testa, W. Ni, W-J. Cai, M. Li, and W. Kemp. (2019b). Ecosystem metabolism and
857 carbon balance in Chesapeake Bay: A 30-year analysis using a coupled hydrodynamic-
858 biogeochemical model. Journal of Geophysical Research -Oceans, 124,
859 https://doi.org/10.1029/2019JC015296.
860 Shen, C., J.M. Testa, M. Li, W-J Cai. (2020). Understanding anthropogenic impacts on pH and
861 aragonite saturation in Chesapeake Bay: insights from a 30-year model study. Journal of
862 Geophysical Research – Biogeosciences, in press.
863 Shenk, G. W., & Linker, L. C. (2013). Development and application of the 2010 Chesapeake
864 Bay watershed total maximum daily load model. JAWRA Journal of the American Water
865 Resources Association, 49(5), 1042-1056.
866 Staniec, A., & Vlahos, P. (2017). Timescales for determining temperature and dissolved oxygen
867 trends in the Long Island Sound (LIS) estuary. Continental Shelf Research, 151, 1-7.
868 Taylor, K. E. (2001). Summerizing multiple aspects of model performance in a single diagram.
869 Journal of Geophysical Research, 106(D7), 7183–7192.
870 https://doi.org/10.1029/2000JD900719

871 Testa, J. M., Li, Y., Lee, Y. J., Li, M., Brady, D. C., Di, D. M., et al. (2014). Quantifying the
872 effects of nutrient loading on dissolved O 2 cycling and hypoxia in Chesapeake Bay
873 using a coupled hydrodynamic – biogeochemical model. Journal of Marine System., 139,
874 139–158. http://doi:10.1016/j.jmarsys.2014.05.018
875 Testa, J.M., Y. Li, Y.J. Lee, M. Li, D.C. Brady, D.M. DiToro, and W.M. Kemp. (2017).
876 Modeling physical and biogeochemical controls on dissolved oxygen in Chesapeake Bay:
877 lessons learned from simple and complex approaches. Chapter 6 in in *Modeling Coastal*
878 *Hypoxia - Numerical Simulations of Patterns, Controls and Effects of Dissolved Oxygen*
879 *Dynamics*, edited by D. Justic, K. Rose, R. Hetland, and K. Fennel. Springer International
880 Publishing AG, Switzerland, 95-118, doi: 10.1007/978-3-310-54571-4_5.
881 Testa, J. M., Murphy, R. R., Brady, D. C. and Kemp, W. M. (2018). Nutrient-and climate-
882 induced shifts in the phenology of linked biogeochemical cycles in a temperate estuary.
883 Frontiers in Marine Science, 5, pp. 1–15. doi: 10.3389/fmars.2018.00114.
884 Testa, J.M., Brady, D.C., Di Toro, D.M., Boynton, W.R., and Kemp, W.M. (2013). Sediment
885 flux modeling: Nitrogen, phosphorus and silica cycles. Estuarine, Coastal and Shelf
886 Science, 131:245-263.
887 Theil, H. (1950). A rank-invariant method of linear and polynomial regression analysis. I, II, III.
888 Nederl. Akad. Wetensch., Proc., 53: 386–392, 521–525, 1397–1412, MR 0036489.
889 Van Meter, K. J., Van Cappellen, P. and Basu, N. B. (2018) . Legacy nitrogen may prevent
890 achievement of water quality goals in the Gulf of Mexico. Science, 360(6387), pp. 427–
891 430. doi: 10.1126/science.aar4462.
892 Wilcox, Rand R. (2001). Fundamentals of Modern Statistical Methods: Substantially Improving
893 Power and Accuracy. Springer-Verlag, pp. 207-210.

894 Wilson, R. E., Bratton, S. D., Wang, J., & Colle, B. A. (2015). Evidence for directional wind
895 response in controlling inter-annual variations in duration and areal extent of summertime
896 hypoxia in western Long Island Sound. Estuaries and coasts, 38(5), 1735-1743.
897 Wood, S.A. (2006). Generalized Additive Models: An Introduction with R. Boca Raton, FL:
898 CRC press.
899 Xie, X., & Li, M. (2018). Effects of wind straining on estuarine stratification: A combined
900 observational and modeling study. Journal of Geophysical Research: Oceans, 123(4),
901 2363-2380.
902 Xin, M., Wang, B., Xie, L., Sun, X., Wei, Q., Liang, S., & Chen, K. (2019). Long-term changes
903 in nutrient regimes and their ecological effects in the Bohai Sea, China. Marine pollution
904 bulletin, 146, 562-573.
905 Zhai, W. D., Zhao, H. D., Su, J. L., Liu, P. F., Li, Y. W., & Zheng, N. (2019). Emergence of
906 summertime hypoxia and concurrent carbonate mineral suppression in the central Bohai
907 Sea, China. Journal of Geophysical Research: Biogeosciences, 124(9), 2768-2785.
908 Zhang, J., Gilbert, D., Gooday, A., Levin, L., Naqvi, W., Middelburg, J., ... & Oguz, T. (2009).
909 Natural and human-induced hypoxia and consequences for coastal areas: synthesis and
910 future development. Biogeosciences Discussions, 6(6).
911 Zhang, Q., Brady, D. C., Boynton, W. R. and Ball, W. P. (2015). Long-Term Trends of Nutrients
912 and Sediment from the Nontidal Chesapeake Watershed: An Assessment of Progress by
913 River and Season. Journal of the American Water Resources Association, 51(6), pp.
914 1534–1555. doi: 10.1111/1752-1688.12327.
915 Zhang, Q., Murphy, R. R., Tian, R., Forsyth, M. K., Trentacoste, E. M., Keisman, J., & Tango, P.
916 J. (2018). Chesapeake Bay's water quality condition has been recovering: Insights from a

multimetric indicator assessment of thirty years of tidal monitoring data. Science of the Total Environment, 637, 1617-1625.

Zhong, L., & Li, M. (2006). Tidal energy fluxes and dissipation in the Chesapeake Bay. Continental Shelf Research, 26(6), 752-770.

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

Zillén, L., Conley, D. J., Andrén, T., Andrén, E. and Björck, S. (2008). Past occurrences of hypoxia in the Baltic Sea and the role of climate variability, environmental change and human impact. Earth-Science Reviews. Elsevier B.V., 91(1–4), pp. 77–92. doi: 10.1016/j.earscirev.2008.10.001.

930 **Figure Captions**

931 Figure 1. (a) Map of Chesapeake Bay. The black stars mark NOAA tidal gauge stations, and the
932 red dots mark EPA Chesapeake Bay Program (CBP) monitoring stations. The eight major rivers
933 are highlighted in dark blue letters. (b) ROMS-RCA model grid. The yellow, light green, dark
934 green and purple regions indicate four subregions used in the following analysis: upper bay,
935 upper middle-bay, lower middle-bay and lower bay. The red and blue lines show the river and
936 ocean boundaries of the model.
937
938 Figure 2. (a) Relationship between winter-spring (January-May) Susquehanna River flow and
939 nitrate loading during 1985-2000 (black circles) and 2001-2016 (red squares). Monthly
940 NO_2+NO_3 (b) and PO_4 (c) concentrations (black lines) and detrended NO_2+NO_3 (b) and PO_4 (c)
941 concentration (red lines) in the Susquehanna River. (d) Monthly averaged Susquehanna River
942 discharge.
943
944 Figure 3. (a) Time series of water level at Duck, North Carolina used to force the ROMS model.
945 The blue line shows the hourly observations, the cyan line shows de-tided water level and the red
946 line is the linear trend. (b)-(f) Modeled (black line) and observed (grey dots) monthly averaged
947 water level at the NOAA gauge stations in Chesapeake Bay.
948
949 Figure 4. Taylor diagram comparing the modelled and observed water level (a) at NOAA tidal
950 gauge stations, and surface (b)/(d) and bottom (c)/(e) temperature/O_2 concentration at the CBP
951 monitoring stations.
952

953 Figure 5. (a)-(h) Modeled (hourly, blue lines) and observed (grey dots) surface/bottom
954 temperature at the CBP monitoring stations.
955
956 Figure 6. Modeled (4-hourly, blue line) and observed (grey dots) surface (right column) and
957 bottom (left column) O_2 concentration at the CBP monitoring stations.
958
959 Figure 7. Time series of modeled monthly (grey line) and GAM fitted (black line) surface (a) and
960 bottom (b) O_2 concentration averaged in the upper-mid bay. (c)-(j) Long-term O_2 residuals (grey
961 dots) and their linear trends (red line) in four subregions of Chesapeake Bay.
962
963 Figure 8. Variations in monthly hypoxic volume calculated from ROMS-RCA: May (a), June (b),
964 July (c), August (d), September (e). The red line marks a linear fit through the data.
965
966 Figure 9. Variations in seasonally-averaged bottom O_2 in the upper mid-bay calculated from
967 ROMS-RCA: Winter (a), Spring (b), Summer (c), Fall (d). The red line marks a linear fit through
968 the data.
969
970 Figure 10. The smooth term representing the long-term trend of surface (left) and bottom (right)
971 O_2 obtained from the hindcast model run and scenario model runs removing temperature increase,
972 sea level rise and nutrient management factors. O_2 is averaged over (a)(b) the upper bay; (c)(d)
973 the upper mid-bay; (e)(f) the lower mid-bay; (g)(h) the lower bay.
974

975 Figure 11. (a)-(d) Bottom O_2 differences between Period2 and Period1 in the hindcast run and
976 scenario runs. (e)-(h) O_2 differences between the scenario runs and the hindcast run in Period2.
977 The O_2 time series were low-pass filtered to remove short-term fluctuations.
978
979 Figure 12. Hypoxia onset (a) and breakup (c) timing at CB4.1C during 1985-2016 obtained from
980 the Base run (black solid cicle), the DtrTEMP run (red empty circle), the DtrSLR run (blue
981 empty circle) and the DtrNut run (green empty circle). The difference of hypoxia onset (b) and
982 breakup (d) timing between between the DtrTEMP, DtrSLR, DtrNut and Base runs.
983
984 Figure 13. (a) Monthly averaged water column respiration in mid-bay in Period 2 in the Base run
985 and scenario runs. (b) Difference of monthly averaged water column respiration in the mid-bay
986 in Period 2 between the Base run and scenario runs.
987
988
989 **Table Captions**
990 Table 1. Sen's slope and significance of M-K trend test of the observed and modeled monthly
991 water level at the NOAA gauge stations.
992
993 Table 2. Sen's slope and significance of M-K trend test of the observed and modeled annual
994 mean surface and bottom water temperature at the CBP stations.
995
996 Table 3. Sen's slope and significance of M-K trend test of the modeled monthly surface and
997 bottom water temperature and salinity at the upper, upper-mid, lower-mid and lower Chesapeake
998 Bay regions obtained from the Base run.

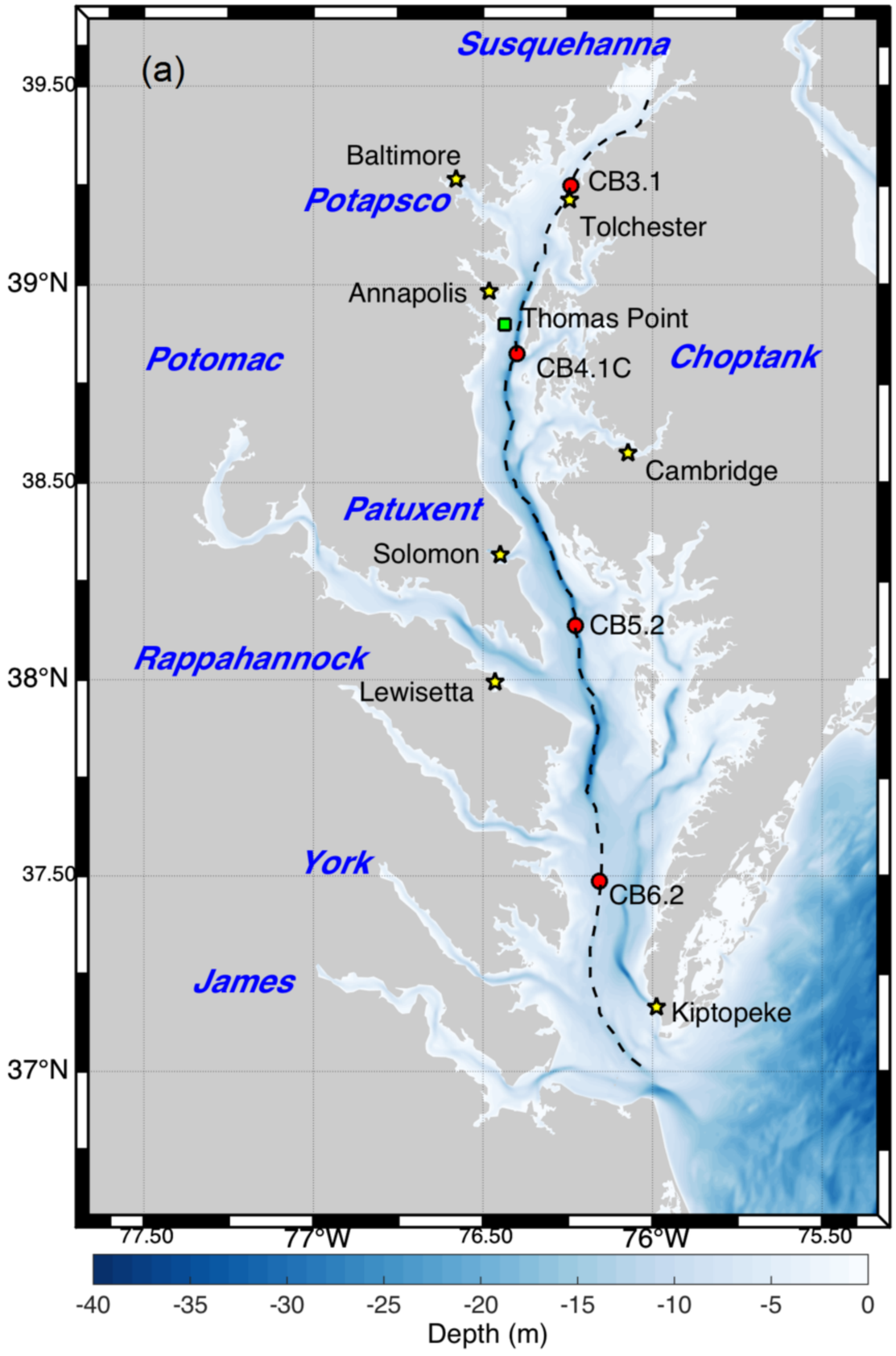
999

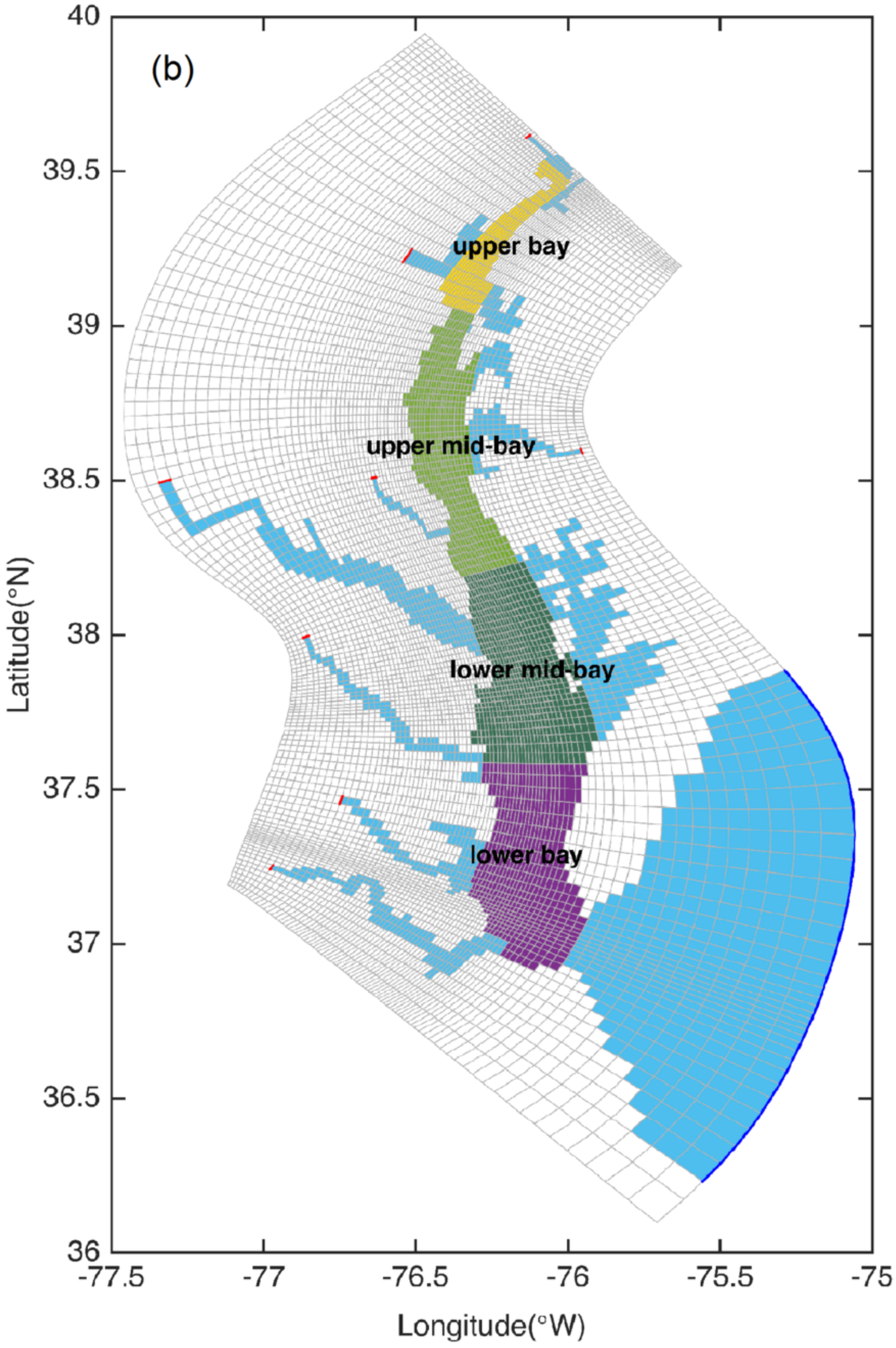
1000 Table 4. Sen's slope and significance of the modeled monthly surface and bottom O_2 at upper,

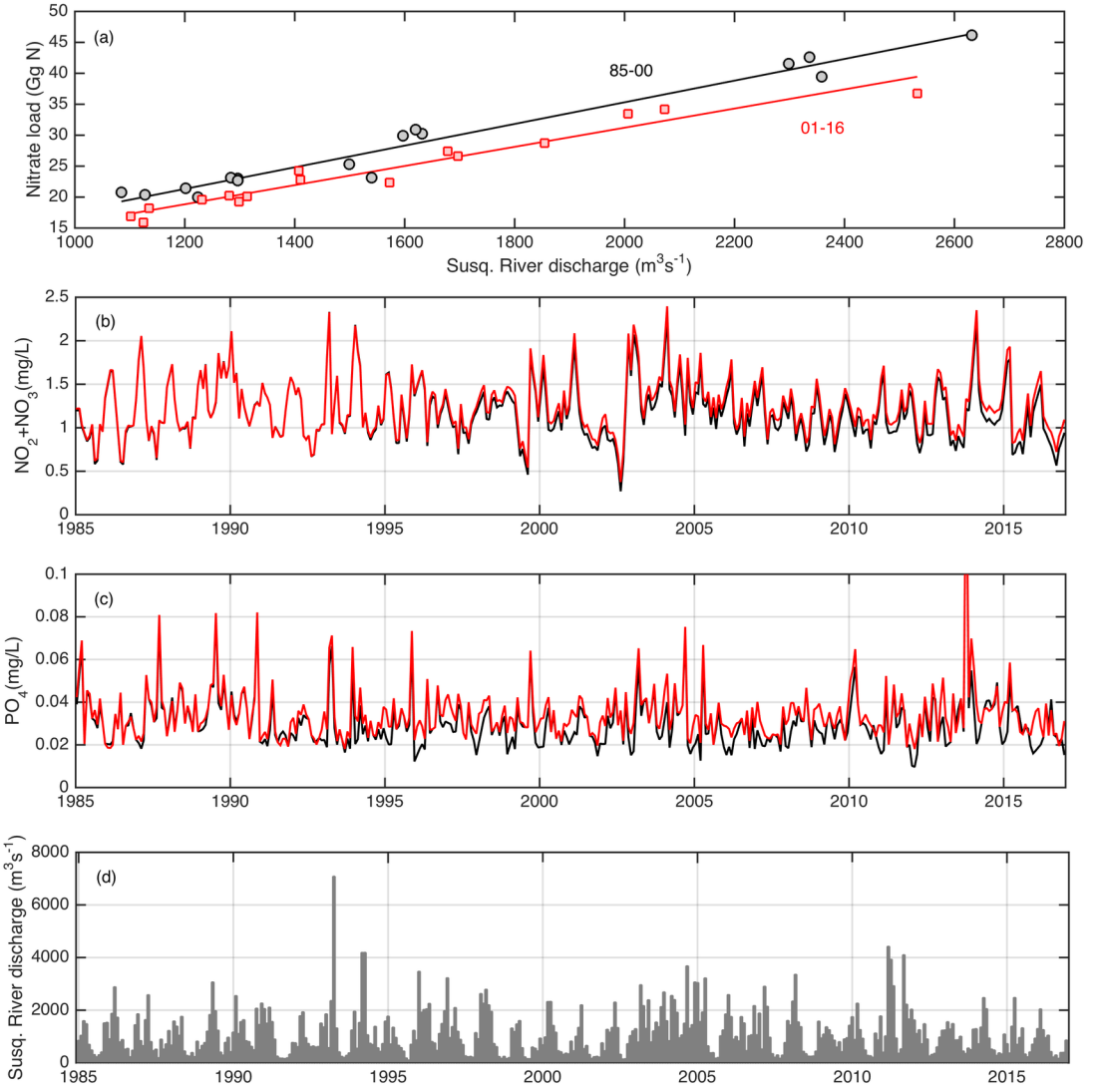
1001 upper-mid, lower-mid and lower Chesapeake Bay region obtained from the Base run.

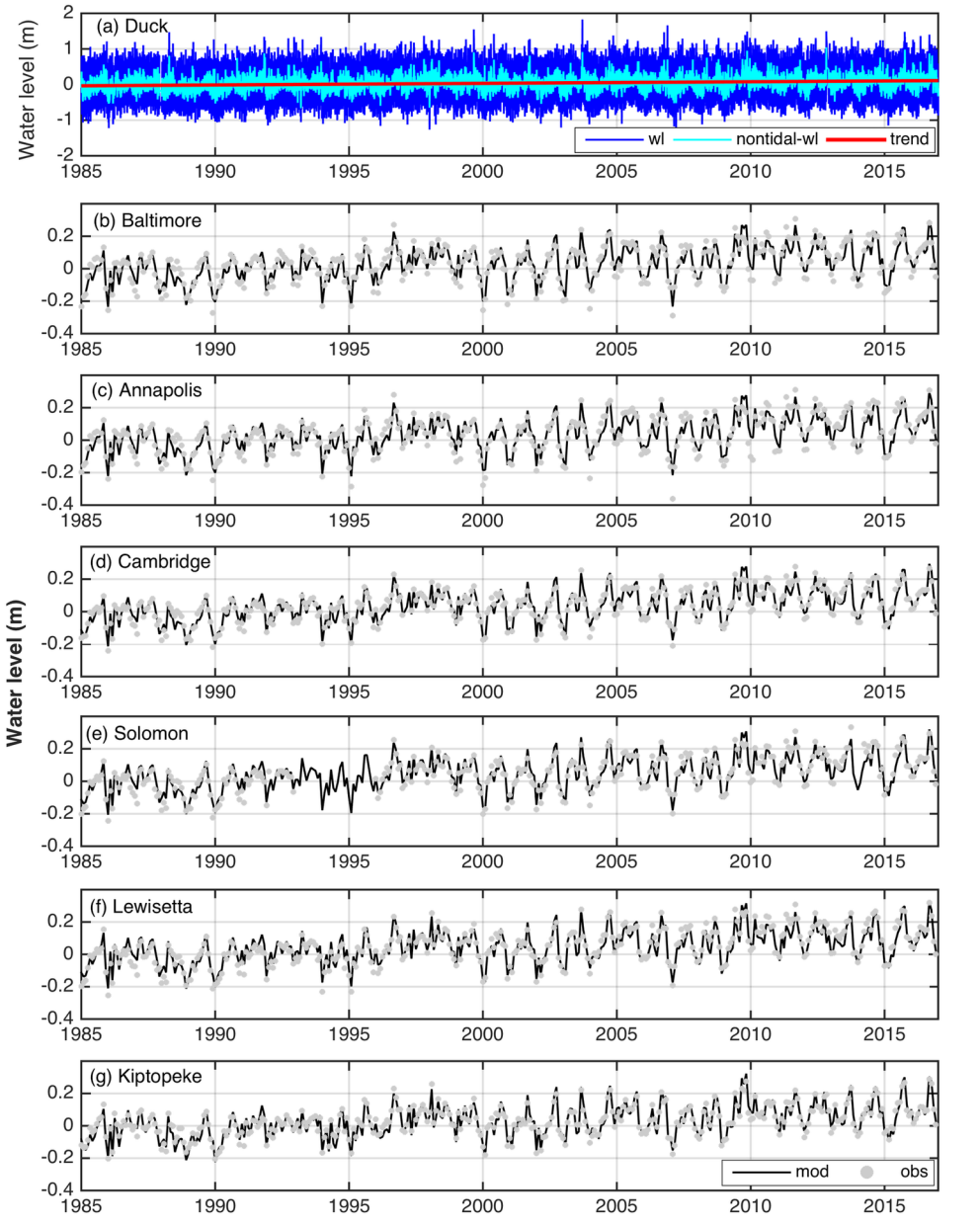
1002
1003

1004

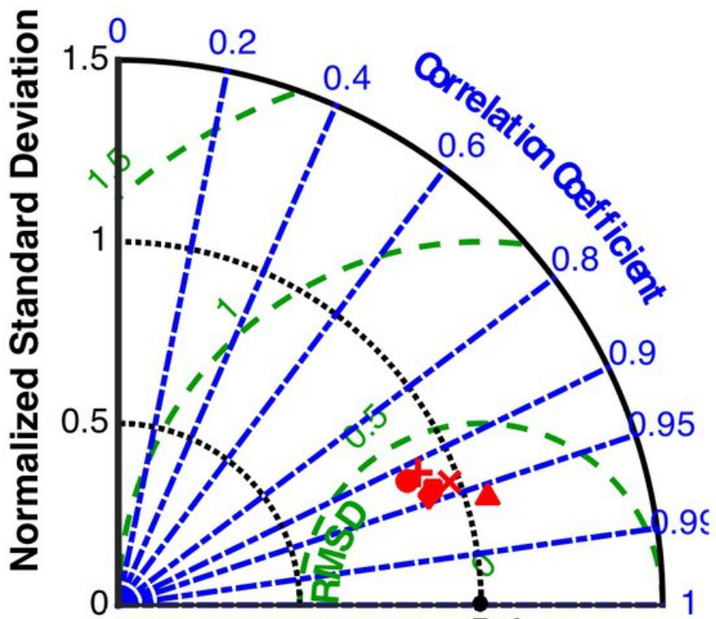

(a)
Susquehanna
Baltimore
CB3.1
Potapsco
Tolchester
Annapolis
Thomas Point
Potomac
CB4.1C
Choptank
Cambridge
Patuxent
Solomon
CB5.2
Rappahannock
Lewisetta
York
CB6.2
James
Kiptopeke
39.50
39°N
38.50
38°N
37.50
37°N
77.50
77°W
76.50
76°W
75.50
-40
-35
-30
-25
-20
-15
-10
-5
0
Depth (m)


(b)
upper bay
upper mid-bay
lower mid-bay
lower bay
40
39.5
39
38.5
38
37.5
37
36.5
36
Latitude(°N)
-77.5
-77
-76.5
-76
-75.5
-75
Longitude(°W)

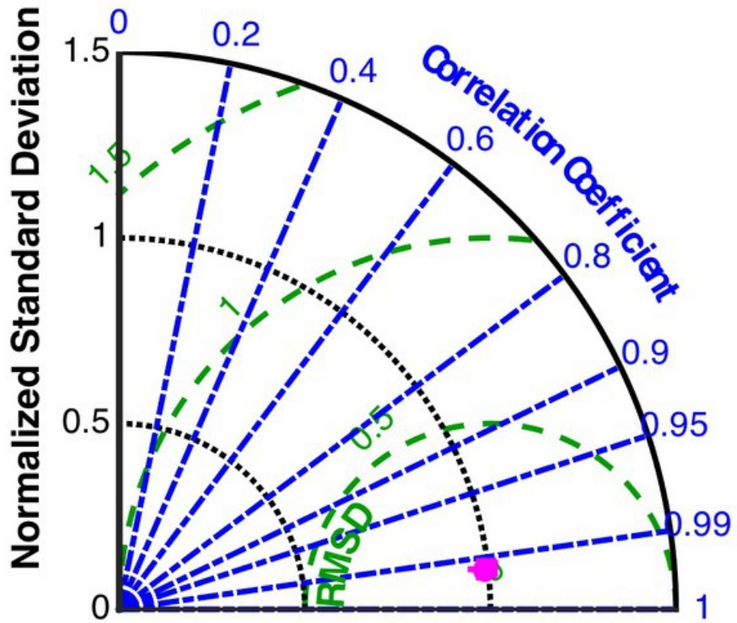

(a)
Nitrate load (Gg N)
Susq. River discharge (m3s-1)
85-00
01-16
(b)
NO2+NO3 (mg/L)
(c)
PO4 (mg/L)
(d)
Susq. River discharge (m3s-1)


(a) Duck
wl
nontidal-wl
trend
(b) Baltimore
(c) Annapolis
(d) Cambridge
(e) Solomon
(f) Lewisetta
(g) Kiptopeke
mod
obs
Water level (m)
Water level (m)
1985
1990
1995
2000
2005
2010
2015

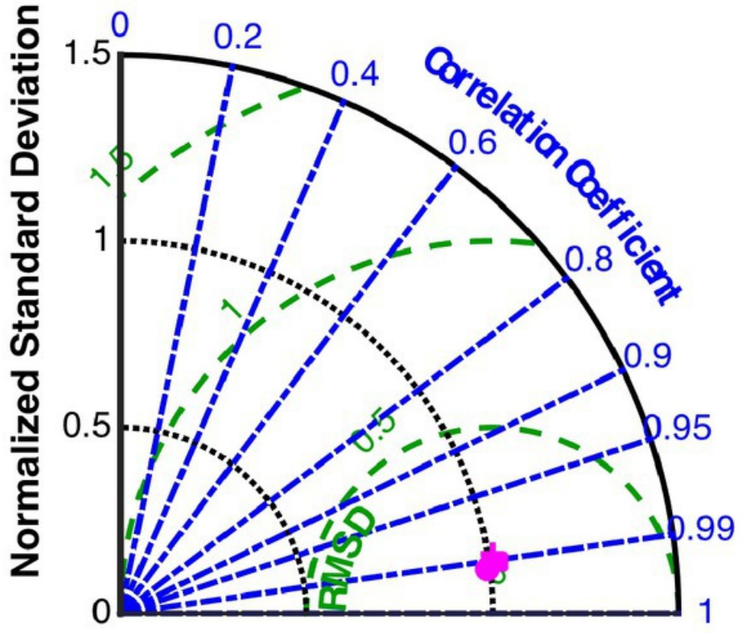
(a) NOAA stations: water level



(b) CBP stations: surface temperature



(c) CBP stations: bottom temperature

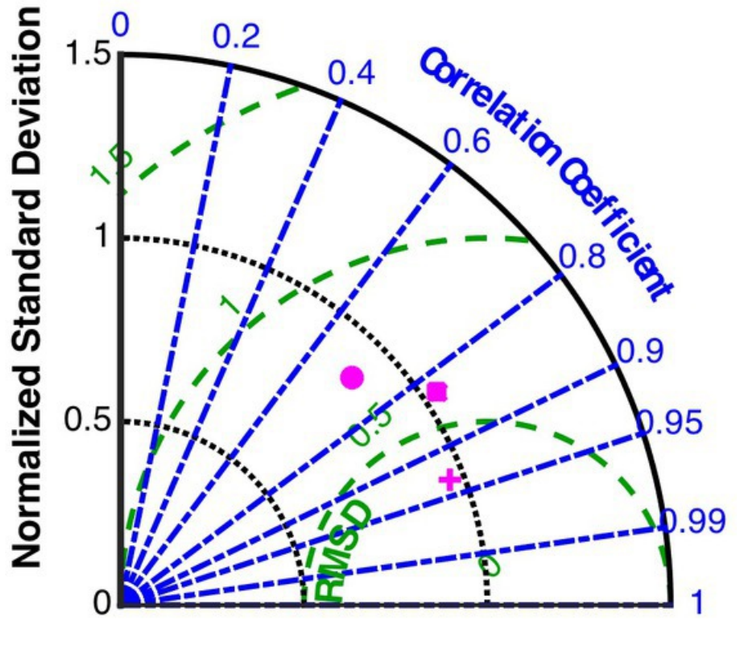


Reference

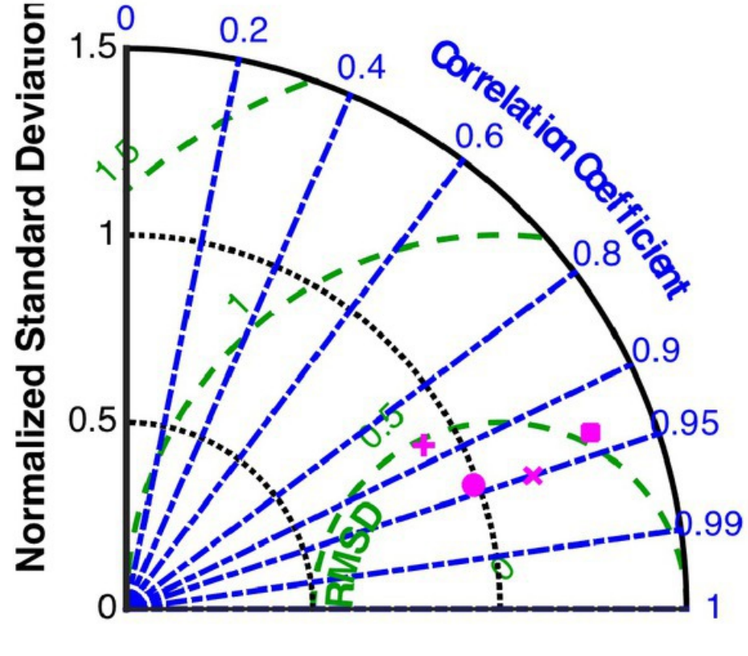
- Baltimore
- Annapolis
- Cambridge
- Solomon
- Lewisetta
- Kiptopeke

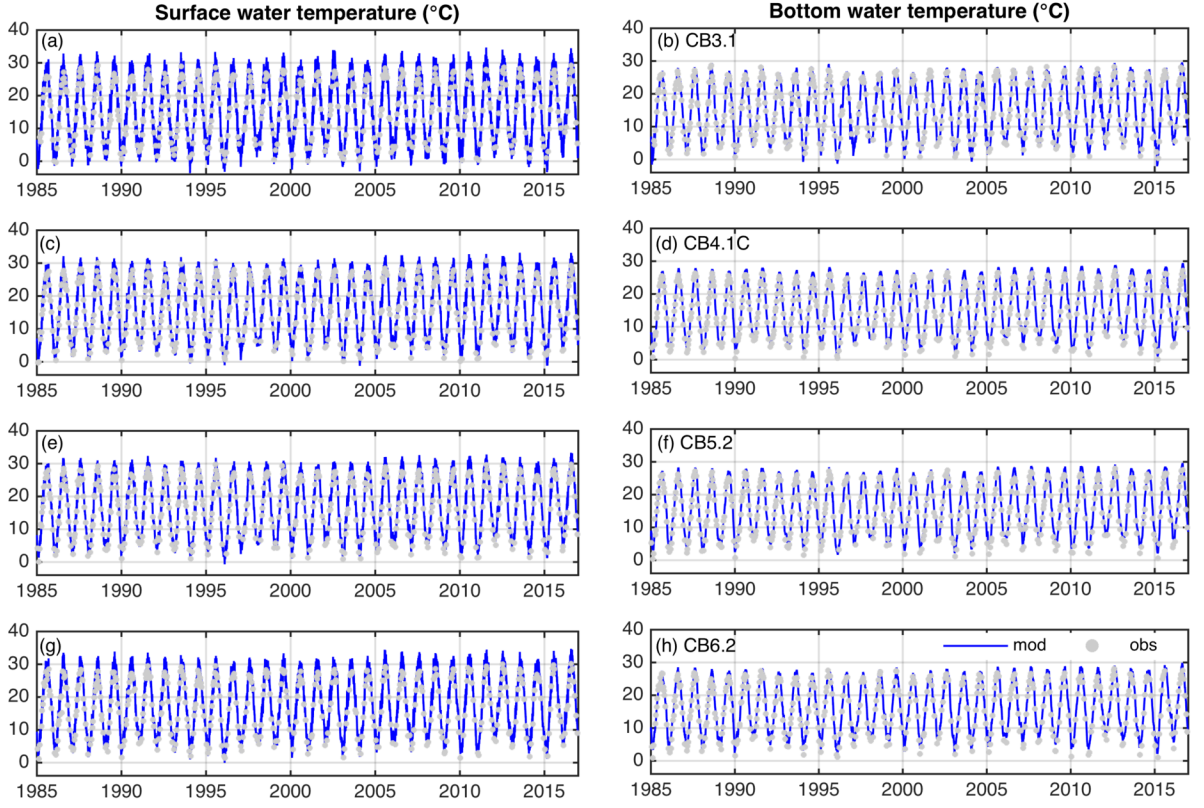
- CB3.1
- CB4.1C
- CB5.2
- CB6.2

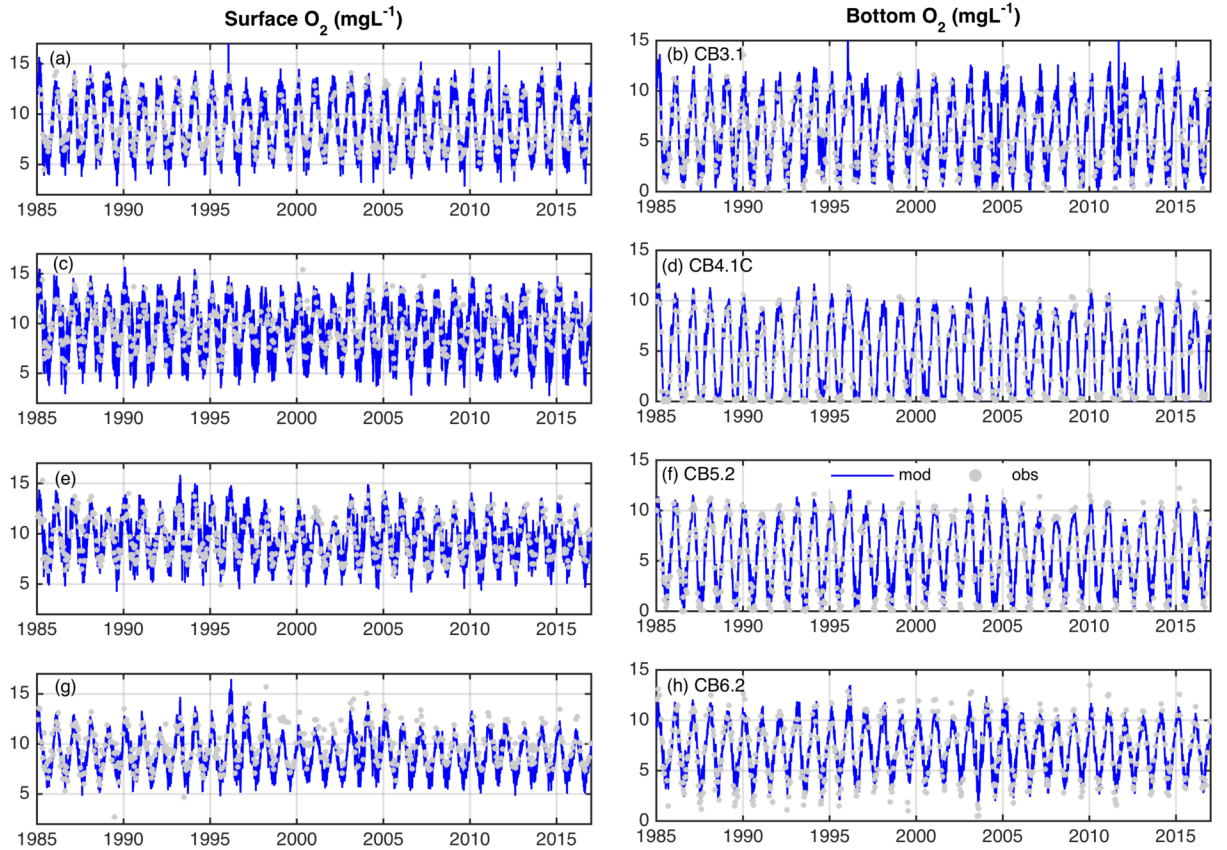
(d) CBP stations: surface DO

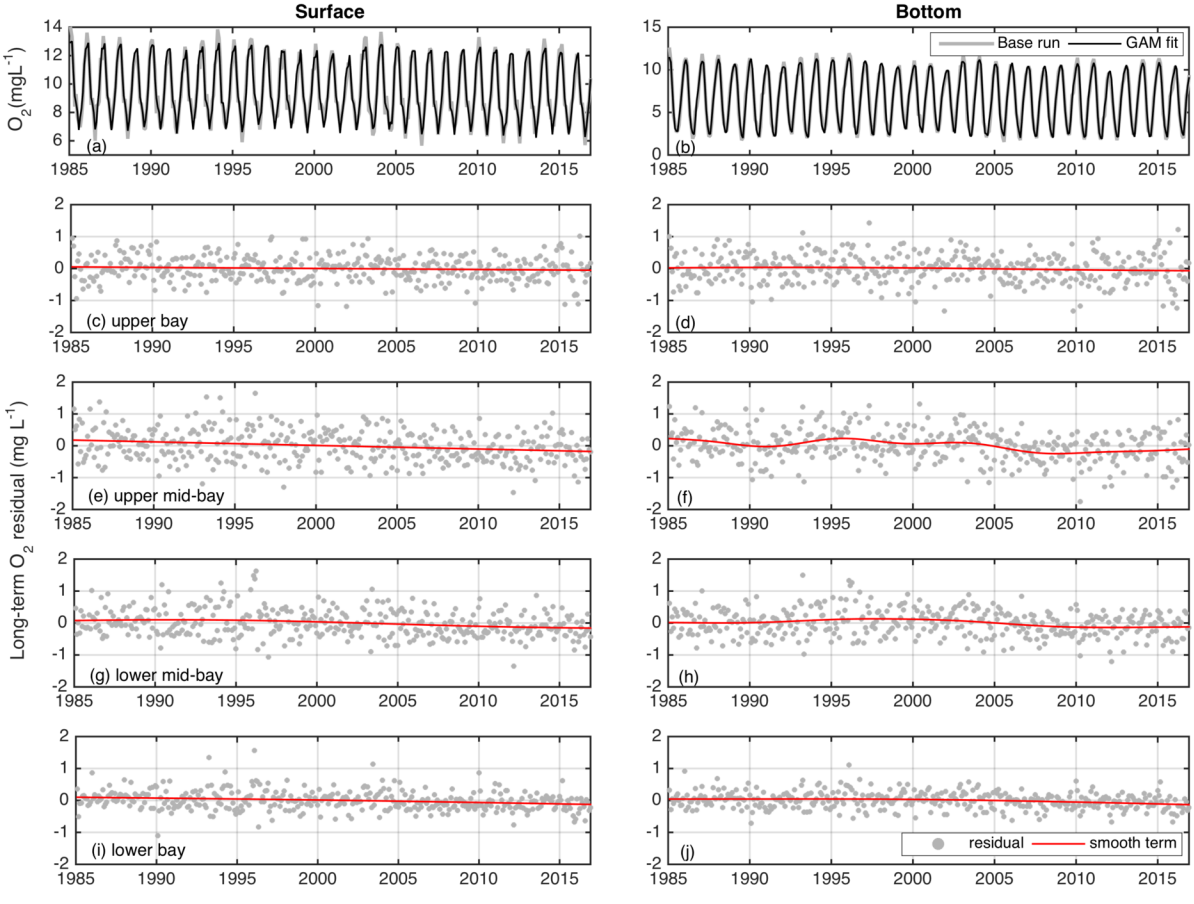


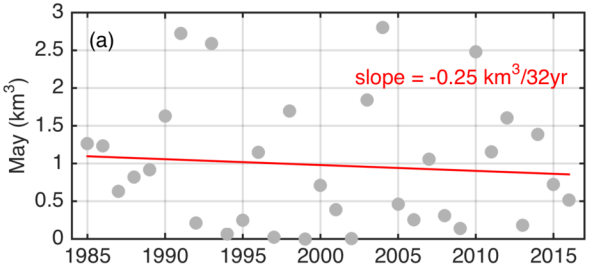
(e) CBP stations: bottom DO

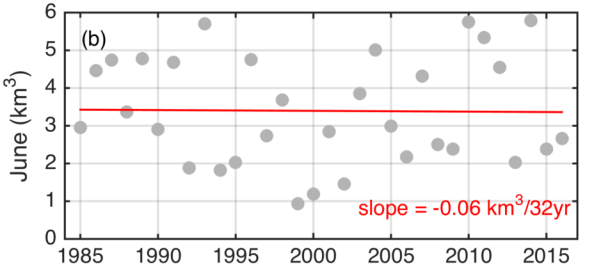


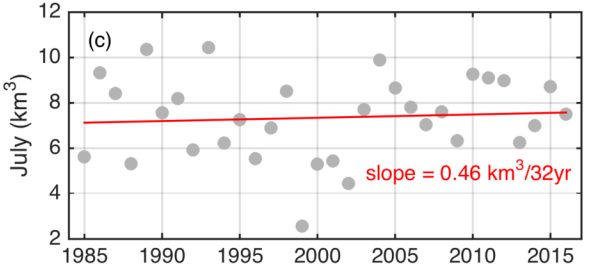

Surface water temperature (°C)
Bottom water temperature (°C)
(a)
(b) CB3.1
(c)
(d) CB4.1C
(e)
(f) CB5.2
(g)
(h) CB6.2
mod
obs
0
10
20
30
40
1985
1990
1995
2000
2005
2010
2015

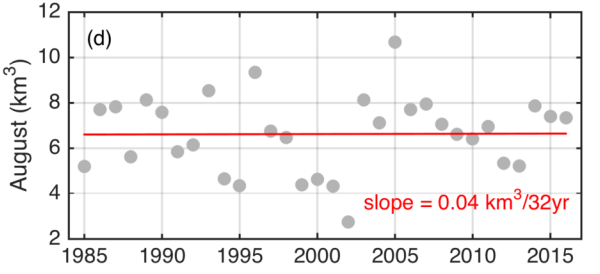

Surface O_2 (mgL^{-1})
Bottom O_2 (mgL^{-1})
(a)
(b) CB3.1
(c)
(d) CB4.1C
(e)
(f) CB5.2
mod
obs
(g)
(h) CB6.2
15
10
5
0
1985
1990
1995
2000
2005
2010
2015

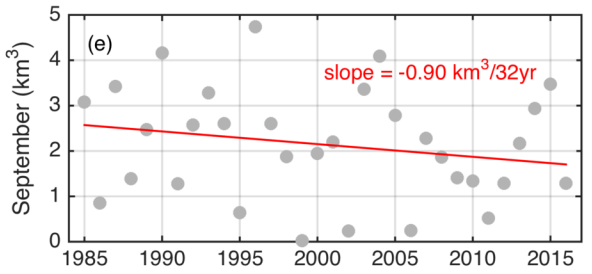

Surface
Bottom
O2 (mgL-1)
Base run
GAM fit
(a)
(b)
(c) upper bay
(d)
(e) upper mid-bay
(f)
(g) lower mid-bay
(h)
(i) lower bay
(j)
Long-term O2 residual (mg L-1)
residual
smooth term

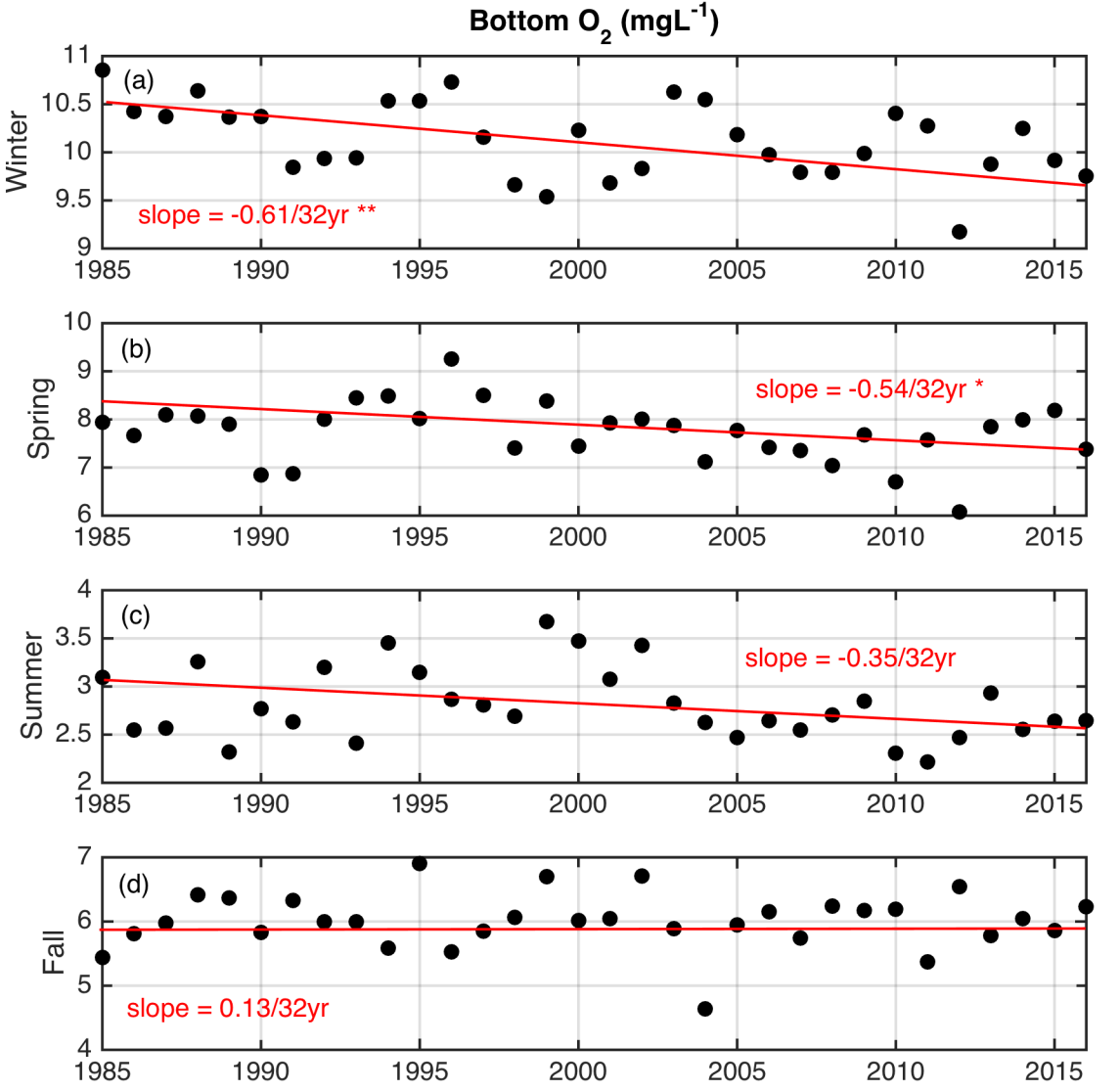

(a)
slope = -0.25 km3/32yr
May (km3)
3
2.5
2
1.5
1
0.5
0
1985
1990
1995
2000
2005
2010
2015

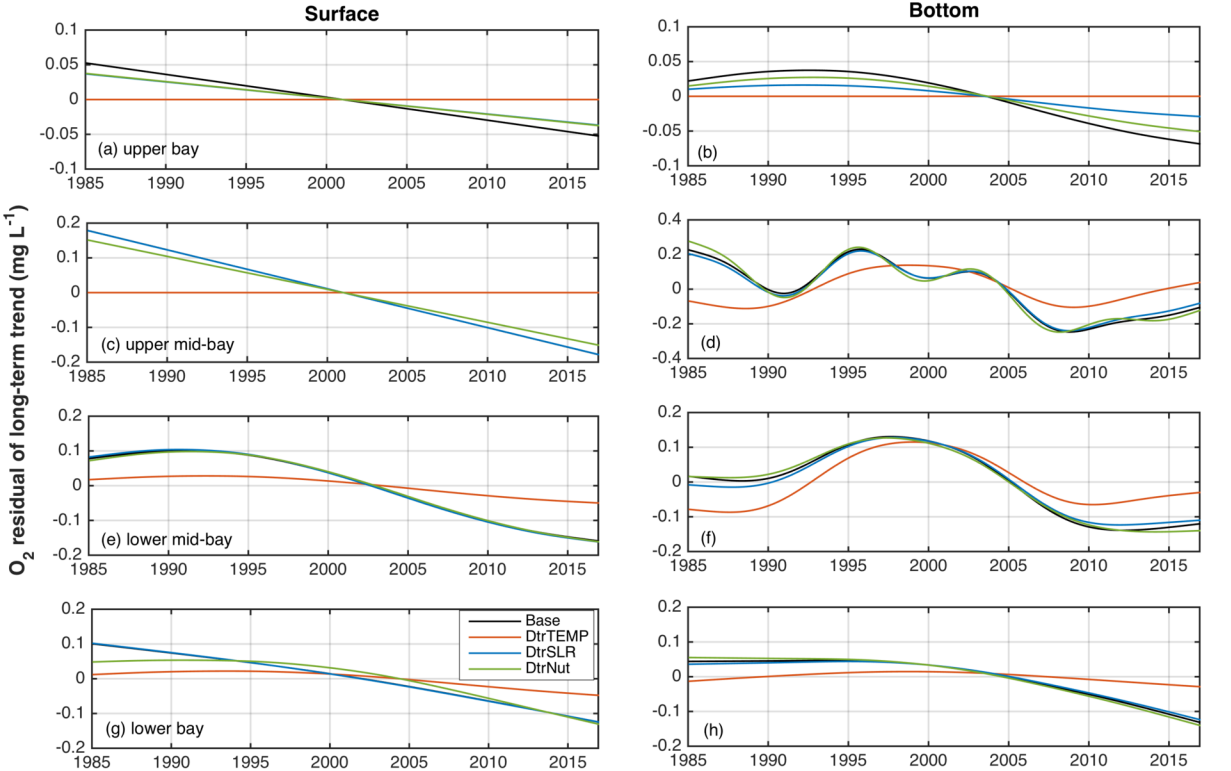

(b)
slope = -0.06 km3/32yr
June (km3)
6
5
4
3
2
1
0
1985
1990
1995
2000
2005
2010
2015

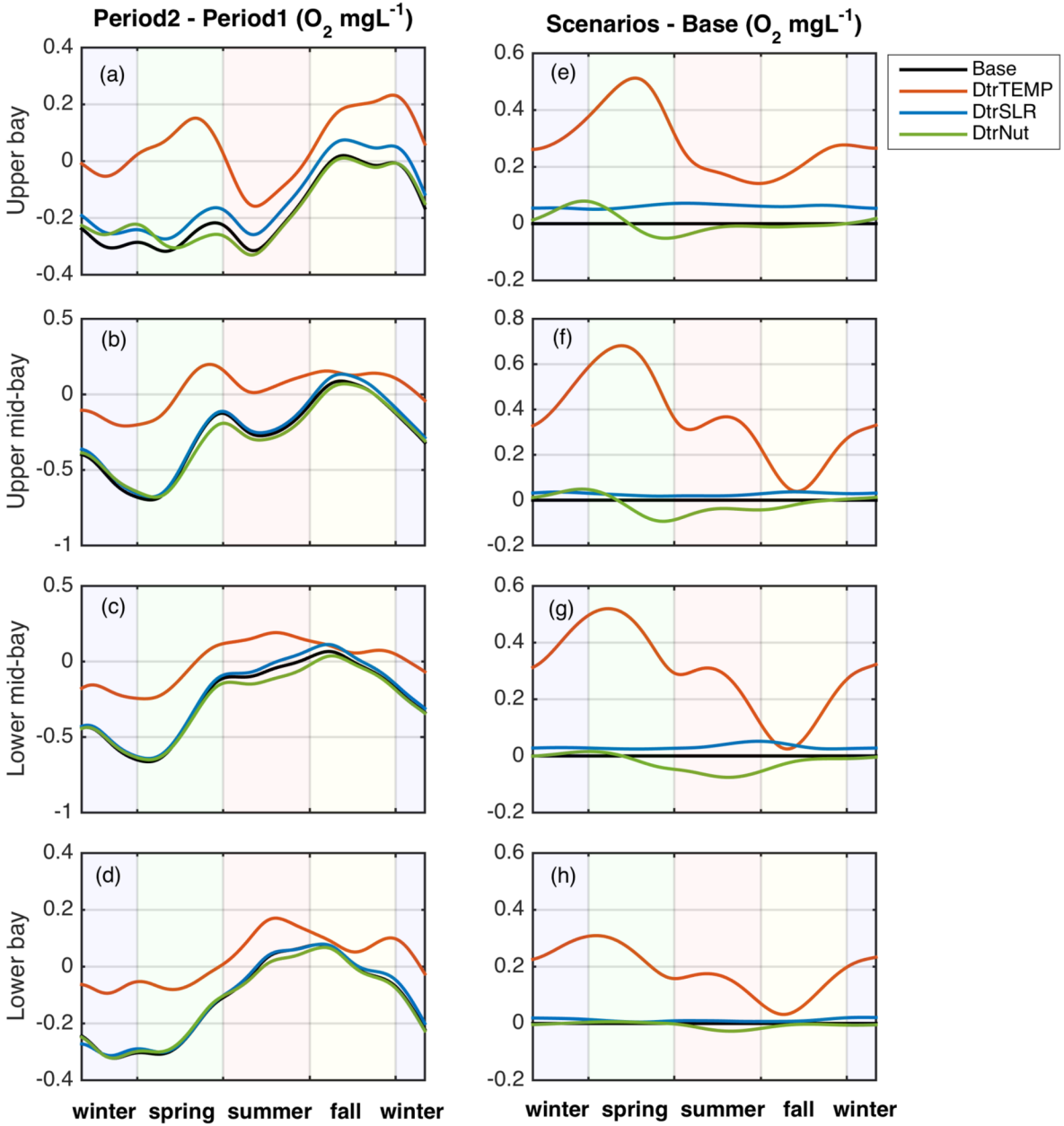

(c)
slope = 0.46 km3/32yr
July (km3)
12
10
8
6
4
2
1985
1990
1995
2000
2005
2010
2015

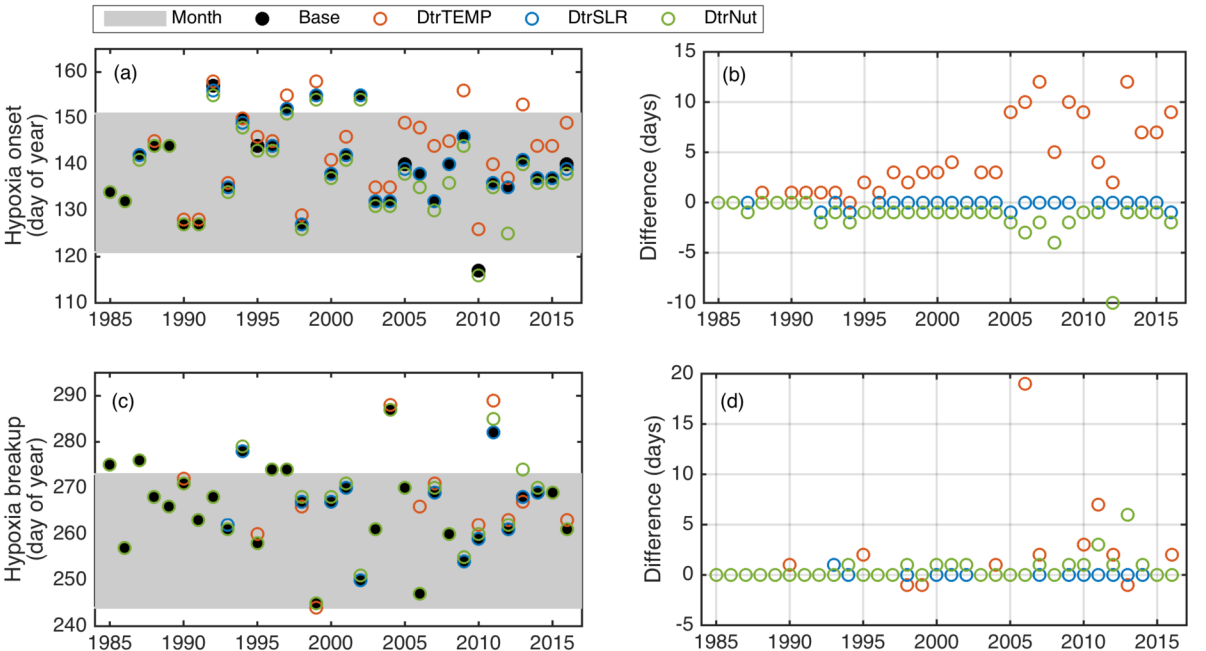

(d)
slope = 0.04 km3/32yr
August (km3)
12
10
8
6
4
2
1985
1990
1995
2000
2005
2010
2015

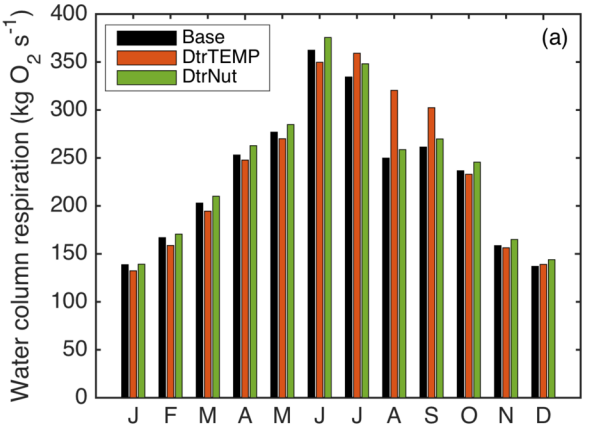

(e)
slope = -0.90 km3/32yr
September (km3)
5
4
3
2
1
0
1985
1990
1995
2000
2005
2010
2015


Bottom O2 (mgL-1)
(a)
Winter
slope = -0.61/32yr **
(b)
Spring
slope = -0.54/32yr *
(c)
Summer
slope = -0.35/32yr
(d)
Fall
slope = 0.13/32yr
1985
1990
1995
2000
2005
2010
2015


Surface
Bottom
O2 residual of long-term trend (mg L-1)
(a) upper bay
(b)
(c) upper mid-bay
(d)
(e) lower mid-bay
(f)
(g) lower bay
(h)
Base
DtrTEMP
DtrSLR
DtrNut


Period2 - Period1 (O_2 mgL^{-1})
Scenarios - Base (O_2 mgL^{-1})
Base
DtrTEMP
DtrSLR
DtrNut
(a)
(b)
(c)
(d)
(e)
(f)
(g)
(h)
Upper bay
Upper mid-bay
Lower mid-bay
Lower bay
winter
spring
summer
fall
winter


Month
Base
DtrTEMP
DtrSLR
DtrNut
(a)
Hypoxia onset (day of year)
160
150
140
130
120
110
1985
1990
1995
2000
2005
2010
2015
(b)
Difference (days)
15
10
5
0
-5
-10
(c)
Hypoxia breakup (day of year)
290
280
270
260
250
240
(d)
20
15
10
5
0
-5


Base
DtrTEMP
DtrNut
(a)
Water column respiration (kg O2 s-1)
400
350
300
250
200
150
100
50
0
J F M A M J J A S O N D

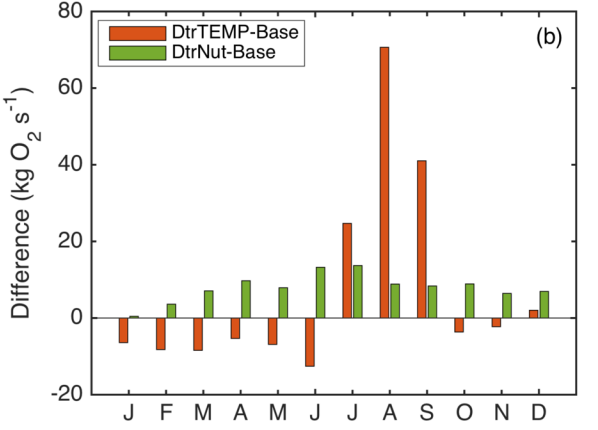

DtrTEMP-Base
DtrNut-Base
(b)
Difference (kg O2 s-1)
80
60
40
20
0
-20
J F M A M J J A S O N D

Table 1. Sen's slope and significance of M-K trend test of the observed and modeled monthly water level at the NOAA gauge stations.

	obs mm/yr	obs p-value	mod mm/yr	mod p-value	RMSE (m)
Baltimore	4.3	<0.01	4.7	<0.01	0.045
Annapolis	4.9	<0.01	4.7	<0.01	0.046
Cambridge	5.1	<0.01	4.8	<0.01	0.035
Solomon	6.2	<0.01	4.8	<0.01	0.038
Lewisetta	6.6	<0.01	4.8	<0.01	0.038
Kiptopeke	4.4	<0.01	4.9	<0.01	0.028

Table 2. Sen's slope and significance of M-K trend test of the observed and modeled annual mean surface and bottom water temperature at the CBP stations.

	surface				Bottom			
	obs		mod		Obs		mod	
	°C/32yr	p-value	°C/32yr	p-value	°C/32yr	p-value	°C/32yr	p-value
CB3.1	1.18	0.0053	1.24	0.0018	1.02	0.0322	1.35	0.0031
CB4.1C	0.87	0.0570	1.78	0.0005	0.56	0.1028	1.45	0.0003
CB5.2	0.88	0.0173	1.95	0.0002	0.75	0.0664	1.43	0.0005
CB6.2	1.51	0.0022	1.60	0.0004	0.97	0.0098	1.30	0.0016

Table 3. Sen's slope and significance of M-K trend test of the modeled monthly surface and bottom water temperature and salinity at the upper, upper-mid, lower-mid and lower Chesapeake Bay regions obtained from the Base run.

		Temperature (°C)		Salinity (psu)	
		MK-p	Sen-32yr	MK-p	Sen-32yr
surface	upper	<0.01	1.31	<0.01	0.30
	up-mid	<0.01	1.80	0.0591	0.30
	low-mid	<0.01	1.77	0.0135	0.30
	lower	<0.01	1.23	0.0365	0.12
bottom	upper	<0.01	1.34	0.0214	0.24
	up-mid	<0.01	1.65	0.0228	0.30
	low-mid	<0.01	1.57	0.0168	0.20
	lower	<0.01	1.08	0.0789	0.16

Table 4. Sen's slope and significance of the modeled monthly surface and bottom O_2 at upper, upper-mid, lower-mid and lower Chesapeake Bay region obtained from the Base run.

		O_2-Base (mg L^{-1})		O_2-DtrTEMP (mg L^{-1})	
		MK-p	Sen-32yr	MK-p	Sen-32yr
surface	upper	<0.01	-0.11	<0.01	1.35E-08
	up-mid	<0.01	-0.36	<0.01	-1.74E-08
	low-mid	<0.01	-0.33	<0.01	-0.09
	lower	<0.01	-0.22	<0.01	-0.08
bottom	upper	<0.01	-0.13	<0.01	2.01E-11
	up-mid	<0.01	-0.41	0.02	0.03
	low-mid	<0.01	-0.19	0.11	0.02
	lower	<0.01	-0.17	<0.01	-0.02
surface saturation	upper	<0.01	-0.23	0.17	0.06
	up-mid	<0.01	-0.33	0.37	0.04
	low-mid	<0.01	-0.31	0.55	0.02
	lower	<0.01	-0.20	0.77	0.01

Warming was the dominant driver of long term changes in O_2 and hypoxia in the Chesapeake Bay over the past 30 years