8-27-2018

Estimating Hypoxic Volume in the Chesapeake Bay Using Two Continuously Sampled Oxygen Profiles

Aaron J. Bever

Marjorie A.M. Friedrichs
Virginia Institute of Marine Science, marjy@vims.edu

Carl T. Friedrichs
Virginia Institute of Marine Science, carl.friedrichs@vims.edu

Malcolm E. Scully

Journal of Geophysical Research: Oceans

RESEARCH ARTICLE
10.1029/2018JC014129

Special Section:
The U.S IOOS Coastal and Ocean Modeling Testbed 2013-2017

Estimating Hypoxic Volume in the Chesapeake Bay Using Two Continuously Sampled Oxygen Profiles

Aaron J. Bever[1], Marjorie A. M. Friedrichs[2], Carl T. Friedrichs[2], and Malcolm E. Scully[3]

[1]Anchor QEA, LLC, San Francisco, CA, USA, [2]Virginia Institute of Marine Science, College of William & Mary, Gloucester Point, VA, USA, [3]Woods Hole Oceanographic Institution, Woods Hole, MA, USA

Key Points:
- Chesapeake Bay hypoxic volume is largely constrained by the bay's geometry and bathymetry
- Two continuously sampled profile locations yield better estimates of cumulative hypoxic volume than bimonthly data at numerous locations
- A Geometric hypoxic volume method provides a simple way to calculate hypoxic volume from two continuous vertical oxygen profiles

Correspondence to:
A. J. Bever,
abever@anchorqea.com

Citation:
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. *Journal of Geophysical Research: Oceans*, *123*, 6392–6407. https://doi.org/10.1029/2018JC014129

Received 29 APR 2018
Accepted 9 AUG 2018
Accepted article online 27 AUG 2018
Published online 12 SEP 2018

Abstract Low levels of dissolved oxygen (DO) occur in many embayments throughout the world and have numerous detrimental effects on biota. Although measurement of in situ DO is straightforward with modern instrumentation, quantifying the volume of water in a given embayment that is hypoxic (hypoxic volume (HV)) is a more difficult task; however, this information is critical for determining whether management efforts to increase DO are having an overall impact. This paper uses output from a three-dimensional numerical model to demonstrate that HV in Chesapeake Bay can be estimated well with as few as two vertical profiles. In addition, the cumulative hypoxic volume (HV_C; the total amount of hypoxia in a given year) can be calculated with relatively low uncertainty (<10%) if continuous DO data are available from two strategically positioned vertical profiles. This is because HV in the Chesapeake Bay is strongly constrained by the geometry of the embayment. A simple Geometric HV calculation method is presented and numerical model results are used to illustrate that for calculating HV_C, the results using two daily-averaged profiles are typically more accurate than those of the standard method that interpolates bimonthly cruise data. Bimonthly data produce less accurate estimates of HV_C because high-frequency changes in oxygen concentration, for example, due to regional-weather- or storm-induced changes in wind direction and magnitude, are not resolved. The advantages of supplementing cruise-based sampling with continuous vertical profiles to estimate HV_C should be applicable to other systems where hypoxic water is constrained to a specific area by bathymetry.

Plain Language Summary The Chesapeake Bay supports a diverse range of recreational and commercial fisheries. However, every summer the amount of oxygen dissolved in the bay's bottom water decreases to levels lethal for fish and shellfish, termed a "dead zone." To make informed management decisions and understand the health of the ecosystem, it is important to understand the variability in the dissolved oxygen and the amount of water with low dissolved oxygen. In this study, we used both observations collected from boats and three-dimensional computer models to demonstrate that the region of low dissolved oxygen is strongly constrained by the geometry of the Chesapeake Bay. Because of this geometric constraint, the amount of low dissolved oxygen water can be accurately calculated if continually monitored with only two vertical profiles. In contrast, the current method uses numerous locations sampled only twice per month. This study demonstrated that continuous vertical profiles at as few as two locations provide a better estimate of the volume of hypoxia in any given year. This means that the amount of hypoxic water can be estimated in real time using an ocean observing system that efficiently supplements the current monitoring program, because only a few continuous monitoring locations are needed.

1. Introduction

Hypoxia occurs when oxygen drawdown exceeds supply, leading to low levels of dissolved oxygen (DO). Hypoxia (DO<2 mg/L) occurs throughout the world (Diaz & Rosenberg, 2008; Gilbert et al., 2010) and specifically has been found on continental margins (Helly & Levin, 2004; Rabalais et al., 2007), estuaries and embayments (Officer et al., 1984; Parker-Stetter & Horne, 2009), fjords (Nordberg et al., 2001), and lacustrine settings (Hawley et al., 2006). The prevalence of systems that undergo hypoxic conditions has been increasing, particularly in coastal systems (Gilbert et al., 2010), which may in part be due to warming temperatures (Anderson et al., 2017; Breitburg et al., 2018; Du et al., 2018; Irby et al., 2018).

While the determination of whether a certain location experiences hypoxic conditions is relatively easy using modern instrumentation, quantifying the total volume of water experiencing hypoxic conditions is much more difficult. Accurately estimating the total amount of hypoxic volume (HV) integrated over the full year

Table 1
Frequently Used Acronyms

Acronym	Complete name	Description
3-D HV	Three-dimensional hypoxic volume	Hypoxic volume calculated by summing the volume of model grid cells
CBP	Chesapeake Bay Program	The Chesapeake Bay Program Partnership
DO	Dissolved oxygen	Concentration of dissolved oxygen in the water in mg/L
GHV	Geometric hypoxic volume	Hypoxic volume calculated using the Geometric method
HV	Hypoxic volume	Volume of water less than 2 mg/L
HV_C	Cumulative hypoxic volume	Total amount of hypoxia in a given year in km^3 days
IHV	Interpolated hypoxic volume	Hypoxic volume calculated using the inverse-distance volumetric interpolator

(i.e., the cumulative hypoxic volume [HV_C]) is even more challenging (Bever et al., 2013), yet is particularly critical for managers who need to determine whether management efforts, such as those aimed at reducing anthropogenic nutrient inputs, are resulting in an interannual trend of decreasing hypoxia (Anderson et al., 2017; Brady et al., 2018; Del Giudice et al., 2018; Irby et al., 2018; Scavia et al., 2017; Testa et al., 2017).

Hypoxia occurring in embayments, lakes, or fjords lends itself more easily to the calculation of the severity of hypoxic conditions in terms of HV than do continental shelf or open ocean systems because the hypoxic water is more constrained by bathymetry and the geographic extent of the system. For example, many fjords typically have a sill at the mouth that constrains the low DO water within the fjord (Nordberg et al., 2001). Lakes generally show hypoxia first occurring in the deepest portion of the lake and spreading from there (Hawley et al., 2006; Nurnberg, 1995, 2004), with the maximum volume of hypoxic water dictated by the size of the lake. Even the Chesapeake Bay, a drowned river valley estuary, has a sill near the mouth with hypoxia shown to be most prevalent in the deeper thalweg (Hagy et al., 2004; Officer et al., 1984). This tendency for low DO water to regularly occur in specific locations within the system facilitates quantitative estimates of the amount of water undergoing hypoxic conditions.

Vertical oxygen profiles are routinely used to estimate both the area and volume of hypoxic water in systems ranging from lakes and embayments (Bever et al., 2013; Buzzelli et al., 2002; Murphy et al., 2011; Nurnberg, 2004) to continental shelves (Grantham et al., 2004; Rabalais et al., 2002). Previous studies in some of these systems have used vertical profile data together with information about basin geometry to estimate the amount of hypoxia. For example, Buzzelli et al. (2002) estimated the spatial extent of hypoxia in the Neuse River Estuary using only vertical profiles of DO in the main channel of the estuary. Nurnberg (1995) estimated the duration and aerial extent of hypoxia in lakes by determining the depth of the 1 mg/L oxycline from vertical profiles and then estimating the area of hypoxic water based on the geometry of lakes. In the Chesapeake Bay, hypoxic water volume is typically estimated by interpolating many vertical oxygen profiles onto a three-dimensional (3-D) grid of the bay, and then summing the volume of grid cells with low levels of DO (i.e., Bever et al., 2013; Murphy et al., 2011). The vertical profiles are generally collected using cruise-based sampling strategies that lead to nonsynoptic sampling. In fact, Bever et al. (2013) showed that the uncertainty in the HV calculated from the coarse temporal resolution of cruise-based sampling can be as large as the uncertainty associated with interpolating vertical profiles throughout an embayment.

This paper uses a simple HV calculation method, similar to that used by Nurnberg (1995), to demonstrate that hypoxic water in the Chesapeake Bay is strongly constrained by the embayment geometry and bathymetry and that HV can be estimated relatively accurately using only two vertical profile locations. The use of as few as two vertical profile locations acts to greatly reduce the uncertainty due to temporal lags in cruise-based sampling strategies and facilitates cost-effective continuous HV monitoring, relative to continuously monitoring many stations. In addition, 3-D model output is used to demonstrate that the availability of continuous data removes errors associated with high-frequency events that tend to be under resolved by bimonthly cruise data. As a result, using two continuously sampled profiles provides better estimates of cumulative hypoxic volume than interpolating bimonthly cruise data.

2. Methods

This section details the study focus location along with the observations and numerical model used in the analyses. Three different methods are described for estimating hypoxic volume: (1) the 3-D HV was calculated by summing the volume of all hypoxic model grid cells, (2) the "Geometric HV" (GHV) was calculated based solely on the geometry of the embayment and a few vertical profiles, and (3) the "Interpolated HV" (IHV) was calculated based on the traditional method for estimating the HV by spatially interpolating vertical profiles (see Table 1 for a list of acronyms). The 3-D HV can only be used with numerical model output that provides continuous spatial and temporal estimates of DO throughout the bay; however, the other two methods, GHV and IHV, were used with both numerical model output and boat-based observations. The statistics and

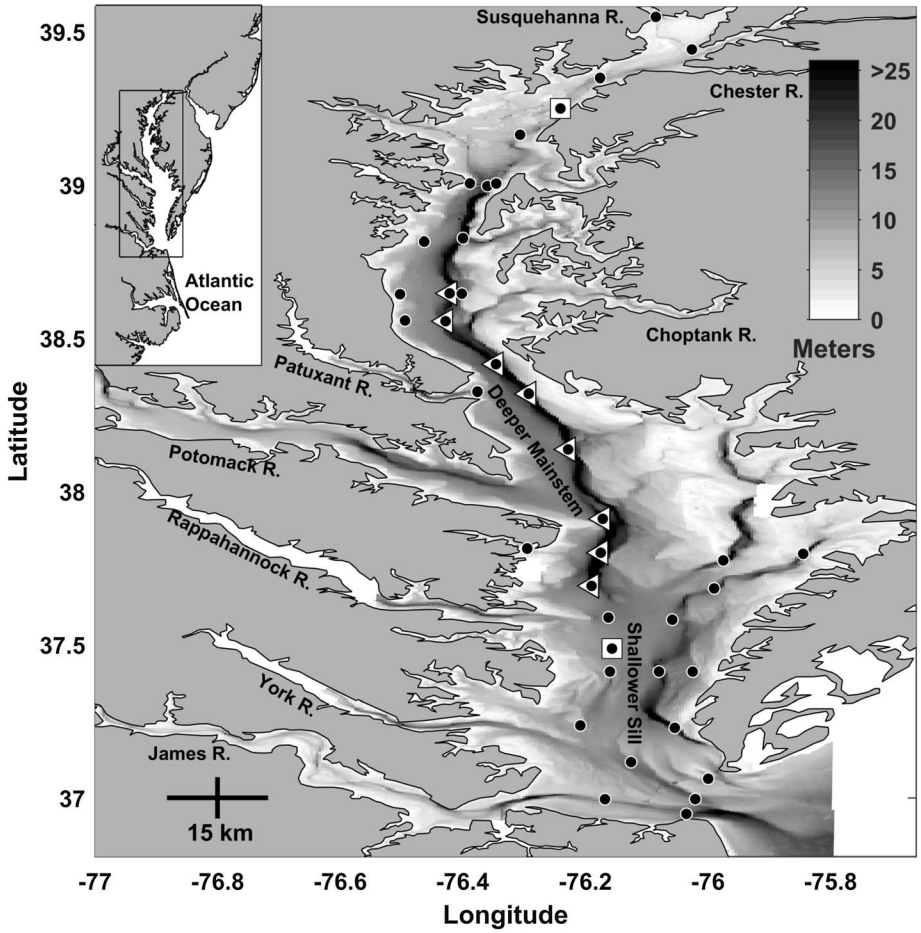

Figure 1. Bathymetry of Chesapeake Bay with major tributaries labeled. Circles represent the considered CBP station locations and triangles highlight the eight stations used in the detailed calculations of the GHV using the CBP data, progressing from north to south: CB4.2C, CB4.3C, CB4.4, CB5.1, CB5.2, CB5.3, CB5.4, and CB5.5 (Table 2). Squares show CB3.1 and CB6.2. The aspect ratio of the bay is stretched in the east-west direction to better show the bathymetry and station locations, with the scale bar indicating 15 km in the east-west and north-south directions.

methods used to evaluate these methods and to compute the cumulative hypoxic volumes computed for full years are also presented below.

2.1. Chesapeake Bay and Data Used

The Chesapeake Bay is a large tidal estuary that undergoes seasonally varying stratification and hypoxia (Officer et al., 1984). Numerous rivers supply freshwater and nutrients to the Chesapeake Bay, with the largest being the Susquehanna River, which discharges into the far northern portion of the bay. The bathymetry of the Chesapeake Bay is characterized by a relatively deep ($>$12 m) main channel, broad expanses of relatively shallow areas on the channel flanks, and a sill near the mouth of the bay that limits the exchange of estuarine and oceanic water (Figure 1). Phytoplankton supply organic matter to the deep main stem, which is isolated from oxygenated surface water by stratification, and from oxygenated oceanic water by the sill (Officer et al., 1984). Decay of the organic matter in the near-bottom water of the main stem results in the drawdown of DO and hypoxic conditions. The geometry of Chesapeake Bay and the method by which hypoxia is created generally constrains the hypoxia to the deeper main stem portion of the Bay (i.e., Officer et al., 1984). This study focuses on main stem hypoxia because it is the largest continuous expanse of hypoxia in the bay and occurs continuously for many months every year. Although hypoxia also occurs in the tributaries and shallows fringing the Bay (i.e., Breitburg, 1990) and affects water quality and biota in these areas (Muller et al., 2016), the tributaries and fringing shallows were not considered in this study because the hypoxia in these areas represents a relatively small water volume or occurs on diel time scales.

Virginia and Maryland state agencies and partners have collected vertical profiles of DO in the Chesapeake Bay since 1984 and made the data available through the Chesapeake Bay Program's (CBP) online data server (Chesapeake Bay Program, 2014), as part of the long-term Water Quality Monitoring Program. Roughly 30 to 60 stations in the main stem portion of the bay are sampled bimonthly in the warmer months (May to September) and monthly throughout the remainder of the year, with vertical profiles collected at about 1-m vertical resolution using two ships, with each ship from a different institution. The large number of stations provides excellent spatial coverage of the Chesapeake Bay, but leads to lag times of up to 2 weeks between the collection of data at the first and last stations. These data have been used previously by numerous scientists and agencies to estimate the extent of hypoxia in the Chesapeake Bay (i.e., Bever et al., 2013; Evans & Scavia, 2011; Hagy et al., 2004; Murphy et al., 2011). This study considered 39 CBP stations from the main stem of the Chesapeake Bay in the HV calculations (Figure 1).

2.2. Numerical Modeling and 3-D HV

Numerous 3-D numerical model implementations have been developed to study DO dynamics and hypoxia in the Chesapeake Bay (i.e., Cerco & Noel, 2013; Da et al., 2018; Feng et al., 2015; Irby et al., 2018; Li et al., 2015; Scully, 2013; Testa et al., 2017; Xia & Jiang, 2016) and have recently been compared and found to have similar skill in reproducing the mean and seasonal variability of DO (Irby et al., 2016). Here we used the implementation from Scully (2013), which is based on an implementation of the Regional Ocean Modeling System (ROMS) developed specifically for the Chesapeake Bay (ChesROMS; Xu et al., 2012). The ChesROMS grid has 20 sigma layers in the vertical, and orthogonal curvilinear horizontal coordinates with highest resolution (430 m) in the northern bay and lowest resolution at the open boundary in the Mid-Atlantic Bight. The stretched terrain-following vertical coordinates result in vertical resolutions of about 0.7 m near the bed, 1.7 m midwater, and 0.3 m in the surface layer in the main stem of the bay (at CBP station CB4.2C, which is 19.1 m deep in the model grid).

Following Scully (2013), DO was modeled over a 21-year period from 1985 to 2005 using a simplistic one-equation DO module abbreviated here as SRM (Simple Respiration rate Model). DO was treated as a passive tracer in the 3-D hydrodynamic model with a spatially and temporally constant oxygen consumption rate of 1.7×10^{-4} mmol $O_2 \cdot m^{-3} \cdot s^{-1}$. DO was added to the surface layer using a wind speed dependent piston velocity and the difference between the DO in the surface layer and the DO at saturation. DO of river inflow and the ocean boundary were set to saturation. DO was also not allowed to go negative, effectively setting a zero respiration rate in locations with anoxic conditions. As part of a multiple model comparison supported by the NOAA-IOOS Coastal and Ocean Modeling Testbed (Luettich et al., 2013, 2017), SRM has been shown to generate oxygen concentrations with similar accuracy to those generated from mechanistic coupled hydrodynamic-biogeochemical models (Irby et al., 2016). The SRM model (Bever et al., 2013) also accurately captures the timing of onset and breakup of hypoxia within the temporal resolution of the current bimonthly regional monitoring. However, the simplicity of the SRM limits the ability of the model to predict changes in the onset of hypoxia resulting from the impact of warming temperatures on phytoplankton and zooplankton that has recently been identified in the bay (see Irby et al., 2018).

A benefit of using numerical models is that HV estimates are available at all times and locations, and the model can be subsampled and processed in a similar manner to field-collected data to quantify errors that may be associated with such subsampling. For example, when using vertical profiles extracted from model results, there are no temporal lags between DO profiles at different locations, which eliminates uncertainty due to nonsynoptic sampling and allows for a direct evaluation of HV calculation methods based on vertical profiles. The "true" HV was calculated by integrating model output over the full 3-D DO field and is thus termed the "3-D HV." Specifically, the 3-D HV was calculated by summing the volume of each model grid cell that had a daily-averaged DO <2 mg/L.

2.3. Geometric Hypoxic Volume (GHV)

Here we introduce a simple Geometric method to estimate HV. The method is similar to that used by Nurnberg (1995); in that, it is based on using the geometry of lakes and the thickness of the layer of hypoxic water, but we focus on a tidal embayment and on the volume of hypoxic water instead of the aerial extent of hypoxia. The overarching assumption of the method is that the hypoxic water initially

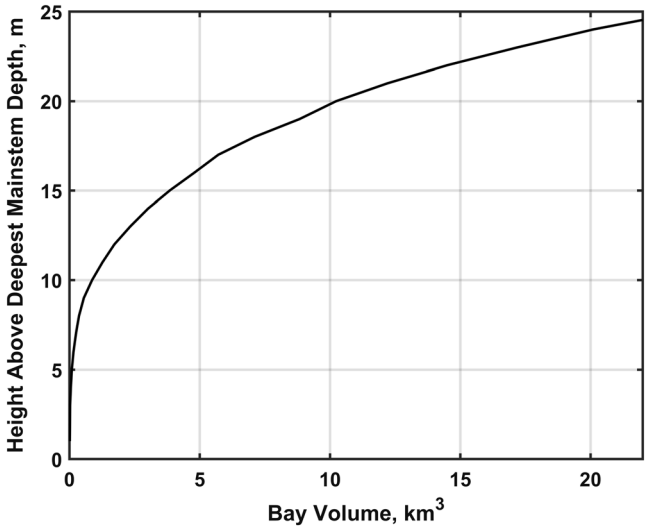

Figure 2. The volume of the main stem of the Chesapeake Bay as a function of the height above the maximum depth.

forms in the deepest portion of an embayment and progressively fills the embayment from the bottom up with a roughly horizontal upper surface as the amount of hypoxic water increases. If this assumption is reasonably valid, the HV can be estimated simply by knowing the volume of the embayment as a function of the height above the maximum depth and the thickness of the hypoxic water. In this study, the volume of the Chesapeake Bay versus the height above the maximum main stem depth was calculated using an approximately 400-m digital elevation model of the bay (Figure 2). (The spatial extent considered was limited to only the portion of the main stem where annual near-bottom hypoxia occurs, from about CB3.1 to CB6.2; as such, Figure 2 is not representative of the entire Chesapeake Bay.) A similar volume versus thickness curve was also generated for the numerical model grid. The steps used in the GHV method are the following:

1. Determine the volume of the bay as a function of the height above the maximum depth within the region that experiences hypoxia (i.e., Figure 2).
2. Use DO vertical profiles to determine the thickness of the less than 2 mg/L DO water relative to the maximum depth in the region of the bay that experiences hypoxia, which corresponds to the total thickness of the hypoxic water.
3. Use the curve developed in step 1 (Figure 2) and the thickness of the 2 mg/L water to estimate the volume of the embayment below the 2-mg/L surface for each vertical profile.
4. If using cruise-based vertical profiles, interpolate the time series of GHV at each station onto consistent dates and times, since the vertical profiles at each station are collected at different dates and times.
5. Average the volumes generated from each individual vertical profile to attain a single HV estimate for the bay.

In addition to the main assumption that the hypoxia occurs near the seabed and progressively extends toward the surface, the GHV method also assumes that the hypoxic water is predominantly found in the deeper portions of the embayment and not on channel-shoal margins or shallower flats regions. The method is still applicable even if there is some degree of tilt or undulation to the surface of the hypoxic water, especially if multiple vertical profiles are used. Previous studies have assumed flat DO surfaces in the across-channel direction (Buzzelli et al., 2002; Hagy et al., 2004). The results presented in this paper support this assumption and demonstrate that assuming generally flat DO surfaces in the along- and across-channel directions can be valid even in complex tidal estuarine systems such as the Chesapeake Bay. Because the GHV method occasionally generates unrealistically high HV estimates in response to extreme tilts in the DO surface, a threshold HV of 20 km^3 was specified, above which the method was assumed to have produced an incorrect estimate, and the HV estimate was not included in the averaging (step 5 above). The GHV was computed both from in situ vertical profile data and numerical model results. The in situ vertical profile data resulted in GHV time series with one to two GHV estimates each month. Hourly modeled output at the 39 CBP stations (Figure 1) was used to calculate daily-averaged DO vertical profiles that were used in the GHV method described above. This resulted in a daily estimate of GHV computed from the model results. A seven-day trailing mean also was applied to the daily GHV estimates from the model to smooth high-frequency variability that occurred when using only a few vertical profiles. A trailing mean was used because it would be applicable for use in estimating hypoxic volume in real time.

2.4. Interpolated Hypoxic Volume (IHV)

For comparison with the 3-D HV and GHV estimates described above, traditional IHVs were calculated by spatially interpolating vertical profiles of DO to the entire bay. For this study, IHV was calculated both in the same manner as those in Bever et al. (2013) using their recommended 13 CBP stations and the CBP volumetric inverse distance-squared interpolator program (U.S. Environmental Protection Agency, 2003) and using the same interpolator but only two stations. The default interpolation options within the program were generally used. Specifically, the DO profiles were interpolated onto a 1-km^2 grid with 0.5-m vertical resolution within

the main stem of the Chesapeake Bay for the years 1985 through 2013. The interpolation used an inverse distance-squared weighting of up to the nearest four surrounding vertical profiles and a maximum interpolation distance away from a vertical profile of 50 km. Unlike the GHV method, the IHV method does not require the use of the trailing mean to reduce high-frequency variability when using continuous modeled output.

2.5. Evaluating Hypoxic Volume Estimates

Target diagram statistics, the total root-mean-square difference (RMSD), and the coefficient of determination (r^2) were used to compare different estimates of the HV. A detailed description of target diagrams is given in Jolliff et al. (2009) and Hofmann et al. (2011). In short, target diagram statistics have the benefit of indicating the skill of one estimate at recreating another both graphically and quantitatively. When graphically displaying the model skill, target diagrams show the total RMSD as the distance from the origin, with the bias between the estimated and reference values on the *y* axis, and the unbiased RMSD (ubRMSD) on the *x* axis. The ubRMSD is multiplied by the sign of the difference between the standard deviations of the estimated and reference values to indicate if the estimated variability is lower (negative *x* axis) or higher (positive *x* axis) than the variability in the reference values. In this study, all target diagram statistics were normalized by the standard deviation of the reference values ($RMSD_N$) to aid in the comparison of different data sets. After this normalization, any point falling inside the circle of radius one ($RMSD_N < 1$) performed better than simply estimating the mean of the reference values. The comparison statistics were only calculated at times when the HV was nonzero.

The GHV was calculated individually using observations from 39 CBP stations. The time series of GHV at each station was then interpolated onto the date and time of the assumed IHV for each sampling cruise to compare the GHV with the IHV based on the CBP cruise data. This temporal interpolation was necessary because the nonsynoptic cruise-based sampling resulted in the observed vertical profiles, and thus the GHV, at different stations being all at different dates and times. However, when calculating the IHV, each cruise was assumed to occur at a single date and time. Any stations whose GHV estimates were not as skillful as simply assuming that the long-term average HV ($RMSD_N > 1$) were excluded from further analysis. This screening-level assessment was used as an objective method of reducing the number of considered stations in the detailed analyses of HV. This method of reducing the number of considered stations does not suggest that these stations are the only ones necessary to understand all aspects of main stem hypoxia in the bay. The stations that were relatively skillful at individually estimating the HV via the GHV method were analyzed both as single stations and as groups ranging from combinations of two stations to a combination of all the stations that remained in the analysis. The stations within each grouping with the lowest $RMSD_N$ between the GHV and the reference HV (either IHV or 3-D HV), were then selected as station groups for detailed analysis.

2.6. Quantifying the Annual Severity of Hypoxia

The cumulative HV (HV_C) is a metric that integrates the HV over a full calendar year (km^3 days) and is an effective metric that quantifies the overall severity of hypoxia in a given year, and thus can be conveniently used to examine interannual changes in the threat of hypoxia to living organisms in a given system (Bever et al., 2013). Here HV_C was calculated as the sum of the instantaneous HV times the length of time between each HV estimate:

$$HV_C = \Sigma_{i=1}^{n} HV_i * dt_i \quad (1)$$

where the HV_C is the cumulative hypoxic volume in a given year, *n* is the number of hypoxic volumes calculated in the year, HV_i is each calculated hypoxic volume, and dt_i is the length of time between successive HVs.

Cumulative hypoxic volume was calculated using the HVs from the three methods detailed above (GHV, IHV, 3-D HV). To quantify the errors in the HV_C estimates resulting from the use of bimonthly cruise data, Monte Carlo analyses were performed on the modeled HV time series. Specifically, the HV on random days in the first half and the second half of each month were subsampled from a daily HV time series and then used to compute the HV_C. The Monte Carlo analyses included 100,000 iterations, resulting in 100,000 HV time series and 100,000 estimates of HV_C. The "cruise dates" in the Monte Carlo analyses were not adjusted to account for timing constraints that would occur in a real field sampling program. That is, cruise dates were allowed to

Table 2
Statistics Detailing How Well the GHV Reproduced the IHV Using Bimonthly Vertical Profiles From the CBP Cruise Observations

Number of stations	Total RMSD (km^3)	r^2	Stations used							
1	2.19	0.73	CB5.2							
2	1.68	0.85	CB5.1	CB5.4						
3	1.40	0.91	CB4.2C	CB5.2	CB5.4					
4	1.24	0.90	CB4.2C	CB5.1	CB5.2	CB5.4				
5	1.29	0.91	CB4.2C	CB5.1	CB5.2	CB5.3	CB5.4			
6	1.25	0.91	CB4.2C	CB4.3C	CB5.1	CB5.2	CB5.3	CB5.4		
7	1.30	0.92	CB4.2C	CB4.3C	CB5.1	CB5.2	CB5.3	CB5.4	CB5.5	
8	1.30	0.91	CB4.2C	CB4.3C	CB4.4	CB5.1	CB5.2	CB5.3	CB5.4	CB5.5

be on consecutive days. This was done to avoid possibly biasing the Monte Carlo results based on subjective logic of when cruises can occur and to maintain random sampling dates.

3. Results

3.1. Comparison of Methods for Calculating HV Time Series Using Bimonthly CBP Cruise Data

The GHV and IHV methods were compared by first applying both techniques to the vertical DO profiles available from the bimonthly CBP cruise data. The GHV method was applied individually to all 39 CBP stations shown in Figure 1; eight of these produced GHV estimates that matched the standard IHV better than simply assuming the long-term average (Figure 1 and Table 2). The GHV resulting from grouping stations CB5.1 and CB5.4 produced the lowest $RMSD_N$ when compared to the IHV, relative to any other combination of two stations (Figure 3). The GHV calculated from these two stations reproduced the IHV similarly to the GHV with eight stations (r^2 = 0.85 and r^2 = 0.91, respectively), even though the IHV was estimated using 13 stations (Table 2). Given the large amount of uncertainty inherent in IHV estimates, upwards of 5 km^3 based on Bever et al. (2013, Figure 7), and that the difference between the GHV computed using two stations and

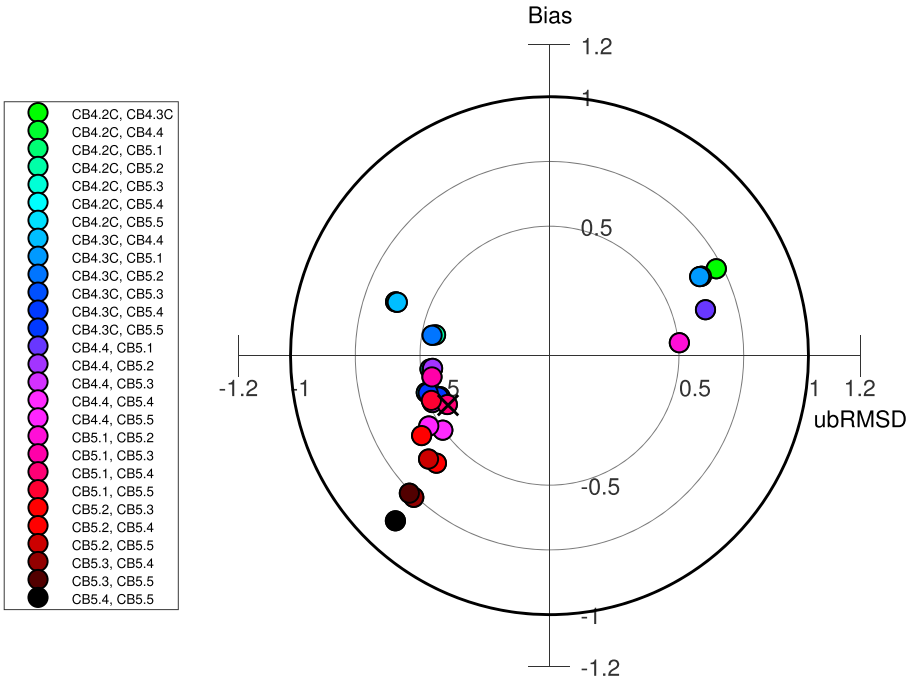

Figure 3. Target diagram showing how the GHV using each set of two stations compared to the IHV. Both the Geometric and Interpolated HVs were calculated using CBP cruise data. The set with the lowest $RMSD_N$ is marked with an X (CB5.1, CB5.4).

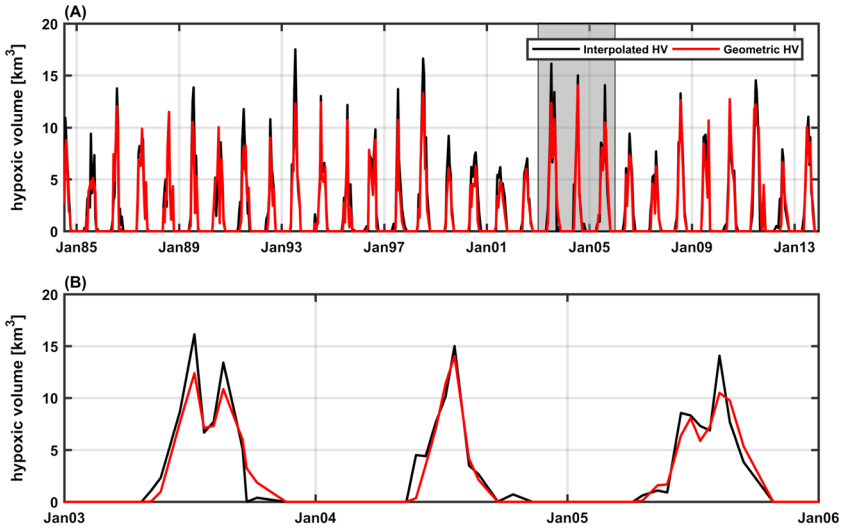

Figure 4. (a) IHV (using 13 stations) and (b) GHV (using two stations: CB5.1 and CB5.4) based on the CBP cruise data. The shading on (a) marks the duration shown in (b).

the IHV computed using 13 stations is typically less than 5 km^3 (Figure 4), these results suggest that these two methods generally produce comparable estimates of HV when bimonthly cruise data are used as inputs.

Little, if any, improvement in RMSD or r^2 occurred when more than four stations were used in the analysis. This lack of consistent improvement in the statistics partially results from the GHV estimates from each station being at a different date/time. Because the GHV method treats each profile as occurring at the exact date/time the data were collected, while the IHV method assumes that all the profiles occurred at exactly the same date/time, the GHV using cruise-based data cannot be expected to ever exactly match the IHV. In addition, the surface of the 2-mg/L DO surface is never completely flat, and as a result, each vertical profile will result in a somewhat different GHV estimate. Averaging a few GHV estimates produces an accurate estimate of the IHV, but simply including more stations does not result in the GHV estimate exactly converging to the IHV estimate. Therefore, the statistics do not consistently improve as more stations are added.

3.2. Comparison of Methods for Calculating HV Time Series Using Numerical Model Output

To further assess the suitability of using the GHV method for computing hypoxic volume time series, the same analysis described above was also performed on output from the numerical model, but this time using daily-averaged vertical profiles rather than instantaneous bimonthly profiles (Table 3 and Figure 5). In this idealized case, the "true" estimate of HV is defined as the 3-D HV, which, as described above, was calculated by summing the volume of each grid cell with oxygen concentrations $<$2 mg/L. Results indicate that the GHV computed from vertical profiles extracted from the numerical model output at seven stations (specifically

Table 3
Statistics Detailing How Well the GHV and IHV Methods Reproduced the 3-D HV Using Continuous Vertical Profiles From the Numerical Model Results

HV method	Number of stations	Total RMSD (km^3)	r^2	Stations used						
GHV	1	2.65	0.79	CB5.1						
GHV	2	1.57	0.89	CB5.2	CB4.2C					
GHV	3	1.65	0.88	CB5.1	CB5.2	CB4.1C				
GHV	4	1.65	0.88	CB5.1	CB5.2	CB5.1 W	CB4.3C			
GHV	5	1.69	0.89	CB5.1	CB5.2	CB5.1 W	CB4.3C	CB4.1C		
GHV	6	1.85	0.88	CB5.1	CB5.2	CB5.1 W	CB4.4	CB4.3C	CB4.1C	
GHV	7	2.01	0.88	CB5.1	CB5.2	CB5.1 W	CB4.4	CB4.3C	CB4.2C	CB4.1C
IHV	2	1.29	0.91	CB5.2	CB4.2C					
IHV	13	0.60	0.98	CB3.2	CB3.3C	CB4.1C	CB4.2C	CB4.3C	CB4.4	CB5.1
				CB5.2	CB5.4	CB6.2	CB6.4	CB7.1	LE2.3	

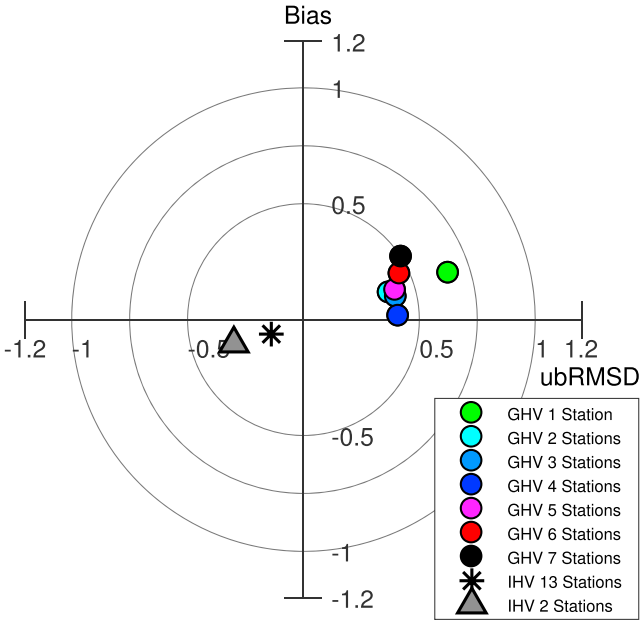

Figure 5. Target diagram based on numerical model results showing how the set of stations with the lowest $RMSD_N$ within each GHV grouping compared to the 3-D HV. The results of comparing IHV calculated using two and 13 stations to the 3-D HV are shown by the triangle and star. Stations in each group are listed in Table 3.

CB4.1C, CB4.2C, CB4.3C, CB4.4, CB5.1, CB5.1W, and CB5.2) estimated the 3-D HV better than simply assuming the mean of the 3-D HV ($RMSD_N < 1$). The grouping of stations CB5.2 and CB4.2C had the lowest $RMSD_N$ between the GHV and the 3-D HV of any of the combinations of two stations. Again, including combinations of more than two stations did not appreciably improve the comparison of the GHV to the 3-D HV (Table 3), and in some cases led to higher bias (Figure 5). The GHV time series closely reproduced the true 3-D HV time series both on interannual (Figure 6a) and seasonal time scales (Figure 6b).

The numerical model output was also used to compare the results of the IHV and the 3-D HV. When using continuous model output at the 13 stations in the main stem of the bay, the IHV had higher r^2 (0.98; Table 3) when compared to the 3-D HV than any of the GHV calculations (highest $r^2 = 0.89$; Table 3). Interestingly, the IHV also worked very well when using only two stations in the interpolator. In fact, the inverse distance-squared IHV time series using only two stations reproduced the 3-D HV time series ($r^2 = 0.91$; Table 3) nearly as well as the IHV time series using 13 stations ($r^2 = 0.98$) and performed very similarly to the GHV method using two stations ($r^2 = 0.89$). The IHV time series closely reproduced the 3-D HV time series, both on interannual time scales (Figure 7a) and seasonal time scales (Figure 7b), regardless of whether two or 13 stations were used (Figures 7b and 7c).

Overall, these results indicate that when continuous information is available, the GHV method using two station locations (Table 3 and Figure 6) is nearly as accurate at reproducing the 3-D HV as the IHV method using two stations (Table 3 and Figure 7). Furthermore, when using the IHV method, using information extracted at only two stations from the model results is comparable in accuracy to using the optimal 13 stations (Bever et al., 2013).

3.3. Comparison of Methods for Calculating Interannual Variability of HV_C Using Numerical Model Output

When assessing whether hypoxia is improving from year to year, cumulative hypoxic volume (HV_C) is often used as an annual metric (Bever et al., 2013; Irby et al., 2018; Li et al., 2015); here HV_C was computed in three

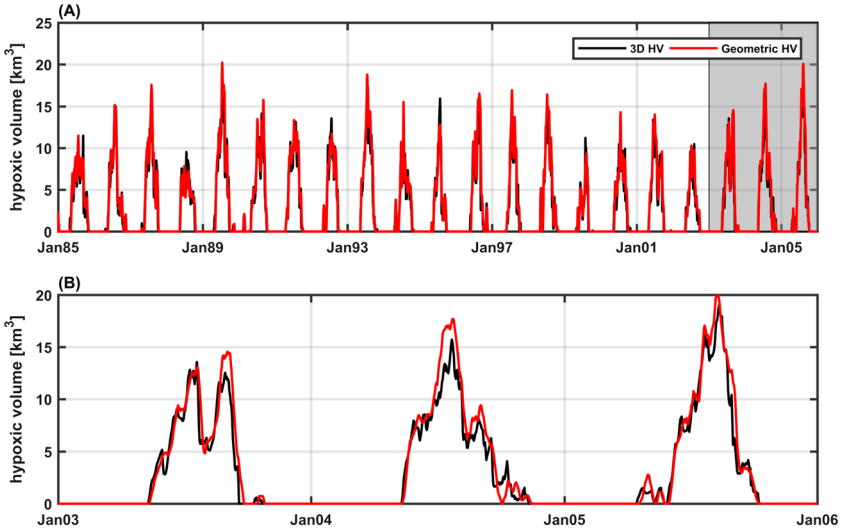

Figure 6. The (a) 3-D HV and (b) GHV estimated from continuous vertical profiles extracted from numerical model results. The GHV estimates were based on using two stations (CB5.2 and CB4.2C). The shading on (a) marks the duration shown in (b).

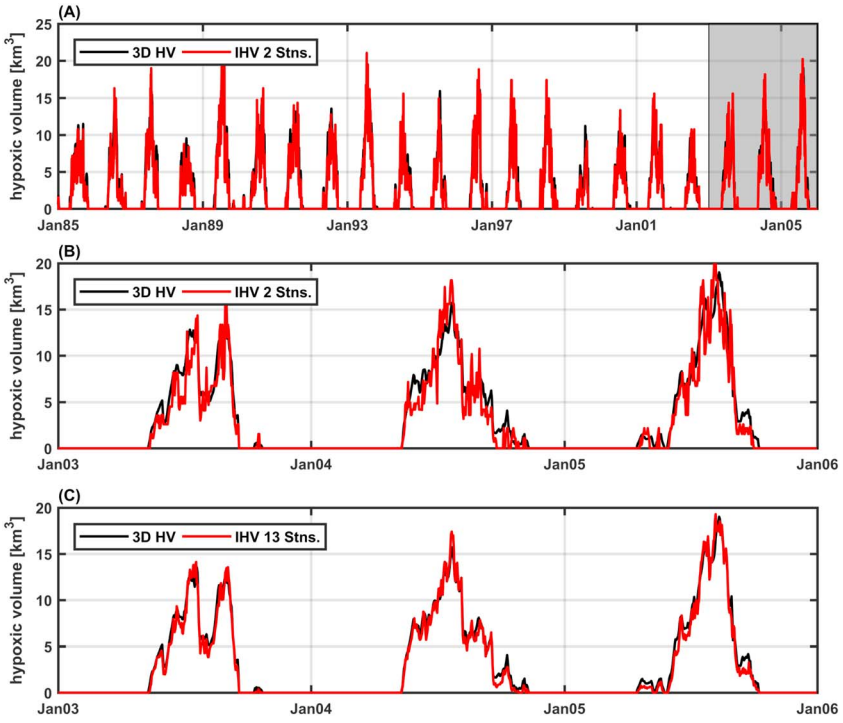

Figure 7. The 3-D HV and IHV estimated from continuous vertical DO profiles extracted from the numerical model results. The IHV estimates were based on using sets of (a and b) two stations and (c) 13 stations. The shading on (a) marks the duration shown in (b) and (c).

different ways from the numerical model output (1985 through 2005), and in each case is compared to that derived from the 3-D HV also derived from the model output (Figures 8 and 9 and Table 4). The 3-D HV estimate of HV_C varied by roughly a factor of 3, with the best year being 1999 (477 km^3 days) and the worst year being 1989 (1378 km^3 days; Figure 8). When the GHV method was applied to model output extracted continuously at two to four stations, the GHV results reproduced the 3-D HV quite closely (Figures 8 and 9a), with an RMSD of only 69 to 102 km^3 days (6.7% to 9.8%; Table 4). In general, the results using the GHV method with four stations produced HV_C estimates that were only slightly lower than those obtained using two or three stations.

Vertical profiles of observed DO have historically been available in the Chesapeake Bay at most bimonthly (rather than continuously, as assumed in Figure 9a). To quantify and isolate the impact of the cruise sampling dates on the HV_C estimates, HV_C was also computed by applying the GHV method to bimonthly vertical profiles of DO. This analysis assumed that data were available everywhere throughout the bay but only at the times of the CBP cruises (Table 4 and Figure 9b, light blue circles). Comparing these results to those of the 3-D HV revealed that using the bimonthly CBP sampling rather than continuous data resulted in errors ranging from an underestimate of 128 km^3 days (in 1996) to an overestimate of 236 km^3 days (in 2001). In contrast, during other years, such as in 1989, the specific cruise timing resulted in very low HV_C errors. To obtain a more representative estimate of the potential errors due to the cruise sampling dates, a Monte Carlo analysis was performed on the sampling dates (see section 2.6). In this case the range of HV_C errors ranged from −281 to 333 km^3 days. Interestingly, this Monte Carlo analysis (Figure 9b, dark blue circles) produced an RMS percent error that was slightly lower than when the actual dates of the CBP cruises (Figure 9b, cyan circles) were used (i.e., 7.0 versus 9.2%, respectively; Table 4).

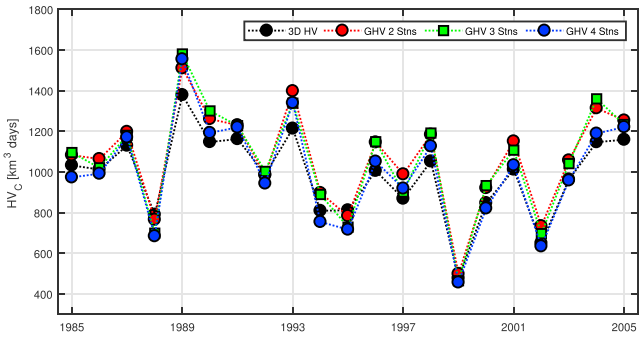

Figure 8. Cumulative HV time series calculated from the GHV method using continuous data from two, three, and four stations, compared to that computed from the 3-D HV.

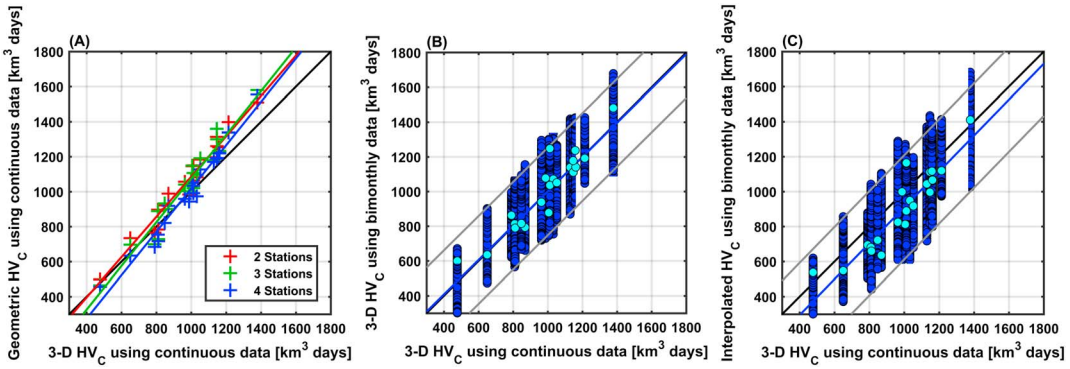

Figure 9. Scatter plots of HV_C calculated from model output for 21 years (1985 through 2005). Comparison of the "true" 3-D HV_C with (a) HV_C calculated using the GHV method with continuous vertical profile data from two, three, and four stations. (b) The 3-D HV_C calculated using bimonthly data. (c) HV_C calculated using the IHV method assuming data are available bimonthly at 13 stations; this represents the standard method currently used for the Chesapeake Bay. In (b) and (c), light blue circles represent results for the actual times of the CBP cruises; dark blue circles represent Monte Carlo analysis of sampling dates (see section 2.6), and blue lines represent the least squares best fit through the dark blue circles; gray lines represent 95% confidence intervals. In each plot, the 1:1 relationship is shown as a solid black line.

To assess the errors in HV_C associated with the current standard practice of using bimonthly cruise data and the CBP's Interpolator program to estimate HV_C, annual estimates were also computed applying the IHV method to bimonthly model output extracted at the 13 CBP stations (rather than assuming data are available everywhere throughout the bay as in Figure 9b). This comparison (Figure 9c) reveals that the HV_C uncertainty associated with the current bimonthly methodology applied to 13 stations has an RMS percent error of roughly 13%, whether or not the Monte Carlo sampling analysis is applied (Table 4). The range of errors associated with the Monte Carlo HV_C estimates based on the standard bimonthly interpolator methodology is −426 to +326 km^3 days. The range of estimates using the actual CBP sampling dates is somewhat smaller (−233 to +153 km^3 days), but still considerably larger than that obtained with the GHV method (−28 to +184 km^3 days).

4. Discussion

4.1. Comparison of HV Resulting From the Geometric HV and the Interpolated HV Methods

Estimates of HV computed from the CBP cruise data calculated using the simple Geometric method were within the uncertainty in the Interpolated HV estimates, indicating that the GHV estimates are similar in accuracy to the HV estimates based on the CBP's standard methodology using the inverse distance-weighted interpolator. The accuracy of the Geometric method using only two to three CBP stations highlights that the location and area of hypoxic water is being constrained by the geometry and bathymetry of the Chesapeake Bay. Nonsynoptic sampling of the CBP vertical profiles likely leads to temporal uncertainty in the IHVs

Table 4
Statistics for Comparing Different Sets of Cumulative Hypoxic Volume Estimates With the 3-D HV_C Estimate Derived From the Numerical Model

HV method	Number of stations	Figure number	Temporal frequency	Minimum difference (km^3 days)	Maximum difference (km^3 days)	RMS difference (km^3 days)	RMS %difference
GHV	2	Figures 8 and 9a	Continuous	−28	184	102	9.9%
GHV	3	Figures 8 and 9a	Continuous	−91	213	104	9.8%
GHV	4	Figures 8 and 9a	Continuous	−106	176	69	6.7%
GHV	Data everywhere	Figure 9b	Bimonthly CBP dates	−128	236	79	9.2%
GHV	Data everywhere	Figure 9b	Bimonthly Monte Carlo dates	−281	333	65	7.0%
IHV	13	Figure 9c	Bimonthly CBP dates	−233	153	120	13.2%
IHV	13	Figure 9c	Bimonthly Monte Carlo dates	−426	326	118	13.0%

estimated using cruise-based station sets (Bever et al., 2013) that is as large as the spatial uncertainty associated with the simplistic Geometric method. Further analysis using synoptic samples or stationary continuous monitoring systems would be necessary to better understand the uncertainty in the HV calculated using the CBP cruise data, regardless of the HV calculation method.

The analysis using the numerical model results strengthened the hypothesis that the geometry of the Chesapeake Bay strongly constrains the location of hypoxia and that the HV can be estimated relatively accurately using only a few vertical profiles. Using the numerical model results removed any uncertainty due to temporal differences in sampling from the estimated HVs. That is, the model results allowed for a direct examination of using only a few stations to estimate the HV while including only spatial uncertainties resulting from sparse vertical profiles. This analysis demonstrated that when the temporal uncertainty due to nonsynoptic sampling was removed, the Geometric method performed nearly as well as the interpolation method at recreating the 3-D HV, that is, at recreating the true modeled HV (Figure 5 and Table 3). The analysis also established that the interpolation method using only two stations was not drastically worse at reproducing the 3-D HV than the interpolation method using 13 stations. The relative accuracy of the interpolation method using only two stations relative to 13 stations further demonstrated that the HV in the Chesapeake Bay can be relatively accurately estimated using only a few vertical profiles.

4.2. Advantages and Disadvantages of Using Two Stations to Estimate Hypoxic Volume

The determination of the area or volume of hypoxia in many systems is currently quantified using data from cruise-based sampling (DiMarco et al., 2010; Murphy et al., 2011) that may contain significant time lags between vertical profiles that add uncertainty to the calculations (Bever et al., 2013). However, the demonstration that the location of hypoxia in the Chesapeake Bay is constrained by embayment geometry results in the applicability of using only a few vertical profiles to relatively accurately estimate the HV. This use of relatively few vertical profiles helps alleviate the problem of increasing uncertainty in estimating HV resulting from nonsynoptic sampling. Even when using cruise-based measurements, the transit time from the first to the last station location is reduced when only sampling a limited number of stations compared to sampling many stations (Bever et al., 2013). The fact that the hypoxic water is constrained by the embayment geometry in the Chesapeake Bay also facilitates continuous real-time monitoring of the HV in the embayment, because relatively few automated vertical profilers in strategic locations would be needed. This conclusion is similar to that noted in DiMarco et al. (2010), who state that geographic controls on the location of hypoxia "… may have important implications for coastal management strategies by identifying physically controlled hot spots in the central and western Louisiana shelf where monitoring efforts can be most efficiently targeted."

Only sampling a few locations and extrapolating to bay-wide estimates does result in potential disadvantages. There are inevitably environmental conditions in a field setting that will cause the estimate of HV using very few stations to fail and predict either unrealistically high HV or too low HV. For example, it has been shown that in the Chesapeake Bay strong winds may force a sloping 2-mg/L DO surface in the direction of the shoals (i.e., Sanford et al., 1990; Scully, 2010), which may at times invalidate the assumption of a relatively flat surface. Another problem could arise if the monitoring stations are outside the region experiencing hypoxia, leading to the estimation of zero HV even when hypoxia is present. This could affect the accurate observation of the timing of onset and breakup of hypoxia. The use of only a few stations also provides much less data on the lateral and longitudinal extent of hypoxia in the embayment, and the overall extent would then also need to be estimated based on the geometry of the embayment. Also, the Geometric conceptual model of DO and the findings on the utility of using two locations to estimate the extent of hypoxia in the Chesapeake Bay are likely not applicable to other variables routinely measured during the long-term Water Quality Monitoring Program, such as nutrient concentrations and chlorophyll *a*.

In a field setting, as the number of vertical profiles is reduced, tides and meteorologically driven undulation of the surface of the 2 mg/L water will lead to increased violations of the underlying assumptions of using only a few vertical profilers and have a larger influence on the HV calculations. However, calculating daily-averaged HV, as was done in this study when using the model results, reduces the influence of tides on the estimated HV. With multiple automated vertical profiles used in the calculations, a tidal analysis may be another option for potentially reducing the uncertainty in the HV caused by tidal influences on the surface of the hypoxic water.

4.3. Applicability to Real-Time Hypoxic Volume Monitoring

While it has been well documented that hypoxia in lakes generally fills the lake from the deepest depth toward the surface (i.e., Hawley et al., 2006; Nurnberg, 1995), this study demonstrated that even in a relatively complex tidal and estuarine system the location of hypoxic water was constrained by the bathymetry and the geometry of the embayment. Automated observing systems could be installed in a few strategic locations that minimize effects on ship traffic and navigation and yet provide substantial continuous information on the vertical structure of DO. Vertical DO profiles at regular time intervals could then be converted to daily estimates of HV by either interpolating throughout the embayment or by a geometric approach. The analysis presented in this paper suggests that when computing HV from two vertical profiles, the inverse distance-weighted IHV method has a similar skill to that of the GHV method (r^2 = 0.91 and r^2 = 0.89, respectively; Table 3); however, the GHV method may be preferable in some instances because it is far simpler to implement.

A further benefit of using automated continuous vertical profilers that sample on a regular time interval is that the frequent DO vertical profiles would allow for the validation of model predictions that HV varies widely on a time scale that is much shorter than boat-based sampling is traditionally collected. Data on the short time scale (hourly to daily) DO and estimated HV would provide a very valuable model calibration and validation tool that could improve numerical model predictions of both instantaneous DO concentration and of HV. Improved understanding of the temporal dynamics of both DO and HV and reduced model uncertainties would translate directly into better informed management decision making and greater confidence in scientific studies.

4.4. Reducing Uncertainties in Annual Metric of Hypoxia Severity

Although HV time series over the course of a single year are invaluable for quantifying the hypoxia at any one time and investigating seasonal variations in hypoxia, further analysis is necessary to investigate interannual variability in hypoxia and/or rank individual years in terms of the severity of hypoxia. Cumulative HV, average summer HV, and maximum annual HV have been used as different metrics to quantify the severity of hypoxia for a given year in the Chesapeake Bay (Bever et al., 2013; Hagy et al., 2004; Kemp et al., 2005; Scavia et al., 2006). Here we focus on the cumulative HV, as this metric has the advantage of incorporating the impact of hypoxic volume in late spring, which is becoming more common with warming bay temperatures (Irby et al., 2018), yet is neglected using the average summer (average HV from 1 June to 30 September) and maximum annual hypoxic volume metrics. The cumulative HV will also capture the breakup of hypoxia in the fall, a time period when the response to nutrient reductions may be able to be observed (Testa et al., 2018; Zhou et al., 2014).

In this study we demonstrated that using information from two continuous vertical profiles in the Chesapeake Bay reproduced cumulative HV more reliably than using bimonthly cruise data from many stations distributed throughout the Bay (Figure 9 and Table 4). This is because high frequency variability in the bay hydrodynamics, such as rapid stratification and destratification, substantially impacts hypoxic volume (Li et al., 2015; Scully, 2010, 2013, and references therein) and is often missed when using bimonthly snapshots of the bay's oxygen field. The results shown here demonstrate that errors in cumulative HV induced by bimonthly rather than continuous sampling are larger than the errors in cumulative HV induced by sampling at two stations rather than at many stations. This is because the temporal variability of oxygen concentrations in the bay (on time scales of several days) is greater than the spatial (horizontal) variability of oxygen concentrations on the bay (on spatial scales of tens of kilometers, that is, the distance between CBP stations). Because the results shown here demonstrate that cumulative HV was more successfully captured using continuous data from only a few stations as compared to using bimonthly data from 13 stations, we posit that using a few automated vertical profilers would supplement the current long-term Water Quality Monitoring Program data and allow for an improved estimate of the interannual change in the severity of hypoxia. However, even with optimal locations, continuous monitoring in the main stem would provide little information on environmental conditions in the tributaries. Continuous monitoring of DO would also not provide any information on many other critically important environmental variables that are observed during the long-term Water Quality Monitoring Program and used by agencies and researchers (i.e., Harding et al., 2016; Patrick & Weller, 2015). Because of this, continuous monitoring could supplement but not replace the current long-term Water Quality Monitoring sampling plan.

Future work is required to determine the optimal locations for continuous monitoring and to develop formal methods for combining information from cruise-based station data and continuous vertical profile data. The large set of main stem Water Quality Monitoring Program stations was objectively subset for detailed analyses by comparing the GHV to the IHV (cruise data) and the GHV to the 3-D HV (model results). However, the retained subset of stations and the combinations of two to eight stations may not be optimal when considering all aspects of hypoxia. This study minimized the $RMSD_N$ to objectively reduce the number of considered stations and to determine the station sets; however, the bias could have been minimized instead. Figure 3 highlights how choices on which combination of stations to use impact the HV and the HV_C. There is some north-south bias (Figure 3) where stations located farther north tend to have positive bias (larger average GHV than IHV) and stations located farther south have negative bias. Minimizing the bias instead of the $RMSD_N$ would have resulted in different, but possibly equally appropriate, station sets. Station sets with minimized bias would have likely better matched the HV_C. However, minimizing the bias could result in HV time series that are very different but have identical means, whereas our choice of minimizing the $RMSD_N$ accounts for differences in both the average value and the variability between two HV time series. Continuous monitoring locations should be decided after considering many aspects of hypoxia and the bay, such as capturing the onset and breakdown of hypoxia that may change due climate change or nutrient reductions (Irby et al., 2018; Testa et al., 2018; Zhou et al., 2014), accurately estimating daily and yearly metrics of the severity of hypoxia, integrating continuous monitoring with the Water Quality Monitoring Program data, and accounting for constraints on sampling locations resulting from navigational concerns. The precise choice of the best locations for deploying a few continuous sampled profiles would also benefit by comparing output from multiple numerical models, including biogeochemical formulations more complex than that contained in SRM (Irby et al., 2016).

4.5. Applicability to Other Systems

The relatively similar characteristic geometry of many embayments worldwide of a deeper area behind a sill (i.e., Lee & Lwiza, 2008; Nordberg et al., 2001) means that hypoxic water in other systems may also be constrained by embayment geometry and bathymetry. Even though the hydrodynamics of estuarine systems can be locally variable on tidal time scales, this study demonstrated that the spatial distribution of DO in the deeper portion of an estuary was not variable enough to invalidate the use of a few continuous profiles for estimating HV. The sensitivity of the calculated HV to local tidal time scale variations in the DO was especially reduced when considering daily-averaged hypoxic volume. As such, fjord type systems with a sill near the mouth, such as Hood Canal in Puget Sound (Newton et al., 2011), Kiljo Fjord in Sweden (Nordberg et al., 2001), and Long Island Sound in New York (Lee & Lwiza, 2008), may also have the location of hypoxia strongly constrained by the geometry of the embayment and be candidates for the quantification of hypoxia using only a few vertical profile locations. Systems in which monitoring programs are being developed or modified especially should consider the use of a few continuous monitoring locations. However, factors not routinely present in the Chesapeake Bay could affect the use of only a few stations for characterizing HV in these systems. One such factor is midwater oxygen minima, such as those observed in Hood Canal (Parker-Stetter & Horne, 2009), where a low DO layer in the middle of the water column is separated from the hypoxic bottom water by a layer of relatively high DO water.

5. Conclusions

Through analyses using both cruise-based dissolved oxygen profiles and 3-D numerical model results, we demonstrated that the location of hypoxic water can be strongly constrained by embayment bathymetry and geometry even in a relatively complex tidal and estuarine system. This conclusion should not be limited to the focus location of this study, but rather should be applicable to many systems worldwide. This tendency for low-oxygen water to regularly occur in the deepest location within the system, and progressively fill the embayment from the bottom up, facilitates an estimate of hypoxic volume based on the use of only a few vertical oxygen profiles. The Geometric HV method introduced here is based on this tendency and provides a useful supplement to cruise-based spatial interpolation procedures, given that it provides similar accuracy and is quite easy to calculate. As few as two continuously sampled vertical profiles could provide hypoxic volume estimates in real time via an automated observing system.

In anthropogenically impacted systems such as the Chesapeake Bay, a goal of monitoring hypoxia is to determine whether multiyear management efforts to improve water quality are having a quantifiable effect. On these longer time scales, cumulative hypoxic volume, that is, the integrated volume of hypoxic water over a full year, can be used for assessment. In this study, we demonstrated that the interannual variability of this metric in the Chesapeake Bay can be determined more reliably from continuous data at two locations than from bimonthly sampling at many stations distributed throughout the bay. This is because bimonthly sampling misses critical high-frequency events that stratify or mix the water column and have large effects on the DO distribution throughout the bay. Further work is currently underway to (1) determine optimal locations for vertical profiles based on the location of hypoxia and observational constraints such as shipping traffic and (2) develop methodologies for combining information from cruise-based station data and continuous vertical profile data to generate optimal estimates of hypoxic volume.

Acknowledgments
This paper is the result of research funded in part by NOAA's U.S. Integrated Ocean Observing System Program Office as a subcontract to the Virginia Institute of Marine Science under award NA13NOS0120139 to the Southeastern University Research Association. This work was performed using High Performance Computing facilities at the College of William & Mary, which were provided by contributions from the National Science Foundation, the Commonwealth of Virginia Equipment Trust Fund, and the Office of Naval Research. We would like to thank our other members of the Coastal Ocean Modeling Testbed Chesapeake Bay Hypoxia team (Raleigh Hood, Hao Wang, and Isaac Irby) for providing input on the results of this study. Jeni Keisman provided us with the Chesapeake Bay Program interpolator. Model output is publicly available through the THREDDS server associated with the IOOS Coastal Modeling Testbed site: https://comt.ioos.us/projects/cb_hypoxia. The cruise-based data used in this study are publicly available through the Chesapeake Bay Program online data server at http://data.chesapeakebay.net/WaterQuality. This is Virginia Institute of Marine Science contribution 3771.

References

Anderson, J. H., Carstensen, J., Conley, D. J., Dromph, K., Fleming-Lehtinen, V., Gustafsson, B. G., et al. (2017). Long-term temporal and spatial trends in eutrophication status of the Baltic Sea. *Biological Reviews*, *92*(1), 135–149. https://doi.org/10.1111/brv.12221

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. *Journal of Geophysical Research: Oceans*, *118*, 4924–4944. https://doi.org/10.1002/jgrc.20331

Brady, D. C., DePinto, J. V., Chapra, S. C., Di Toro, D. M., Friedrichs, M. A. M., Gray, M. W., et al. (2018). Scientific and Technical Advisory Committee: Chesapeake Bay Water Quality and Sediment Transport Model (WQSTM) Review, STAC Publication Number 18–002 (40 pp.). Edgewater, MD.

Breitburg, D., Levin, L. A., Oschlies, A., Grégoire, M., Chavez, F. P., Conley, D. J., et al. (2018). Declining oxygen in the global ocean and coastal waters. *Science*, *359*(6371), eaam7240. https://doi.org/10.1126/science.aam7240

Breitburg, D. L. (1990). Near-shore hypoxia in the Chesapeake Bay: Patterns and relationships among physical factors. *Estuarine, Coastal and Shelf Science*, *30*(6), 593–609. https://doi.org/10.1016/0272-7714(90)90095-9

Buzzelli, C. P., Luettich, R. A., Powers, S. P., Peterson, C. H., McNinch, J. E., Pinckney, J. L., & Paerl, H. W. (2002). Estimating the spatial extent of bottom-water hypoxia and habitat degradation in a shallow estuary. *Marine Ecology Progress Series*, *230*, 103–112. Retrieved from http://www.jstor.org/stable/24865097, https://doi.org/10.3354/meps230103

Cerco, C. F., & Noel, M. R. (2013). Twenty-one-year simulation of Chesapeake Bay water quality using the CE-QUAL-ICM eutrophication model. *Journal of the American Water Resources Association*, *49*(5), 1119–1133. https://doi.org/10.1111/jawr.12107

Chesapeake Bay Program (2014). Chesapeake Bay Program Data Hub. (Accessed October 2014). Retrieved from http://www.chesapeakebay.net/data

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. *Journal of Geophysical Research: Oceans*, *123*, 5004–5025. https://doi.org/10.1029/2018JC014009

Del Giudice, D., Zhou, Y., Sinha, E., & Michalak, A. M. (2018). Long-term phosphorus loading and springtime temperatures explain interannual variability of hypoxia in a large temperate Lake. *Environmental Science & Technology*, *52*(4), 2046–2054. https://doi.org/10.1021/acs.est.7b04730

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

DiMarco, S. F., Chapman, P., Walker, N., & Hetland, R. D. (2010). Does local topography control hypoxia on the eastern Texas-Louisiana shelf? *Journal of Marine Systems*, *80*(1-2), 25–35. https://doi.org/10.1016/j.jmarsys.2009.08.005

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. *Science of the Total Environment. 630*, 707–717. https://doi.org/10.1016/j.scitotenv.2018.02.265

Evans, M. A., & Scavia, D. (2011). Forecasting hypoxia in the Chesapeake Bay and Gulf of Mexico: Model accuracy, precision, and sensitivity to ecosystem change. *Environmental Research Letters*, *6*(1), 1–11. https://doi.org/10.1088/1748-9326/6/1/015001

Feng, Y., Friedrichs, M. A. M., Wilkin, J., Tian, H., Yang, Q., Hofmann, E. E., et al. (2015). Quantifying Chesapeake Bay nitrogen fluxes using a land-estuarine ocean biogeochemical modeling system: Model description, evaluation and budgets. *Journal of Geophysical Research: Biogeosciences*, *120*, 1666–1695. https://doi.org/10.1002/2015JG002931

Gilbert, D., Rabalais, N. N., Diaz, R. J., & Zhang, J. (2010). Evidence for greater oxygen decline rates in the coastal ocean than in the open ocean. *Biogeosciences*, *7*(7), 2283–2296. https://doi.org/10.5194/bg-7-2283-2010

Grantham, B. A., Chan, F., Nielsen, K. J., Fox, D. S., Barth, J. A., Huyer, A., et al. (2004). Upwelling-driven nearshore hypoxia signals ecosystem changes in the Northeast Pacific. *Nature*, *429*(6993), 749–754. https://doi.org/10.1038/nature02605

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. *Estuaries and Coasts*, *39*(3), 664–681. https://doi.org/10.1007/s12237-015-0023-7

Hawley, N., Johengen, T. H., Rao, Y. R., Ruberg, S. A., Beletsky, D., Ludsin, S. A., et al. (2006). Lake Erie hypoxia prompts Canada-U.S. study. *Eos, Transactions of the American Geophysical Union*, *32*(8), 313. https://doi.org/10.1029/2006EO320001

Helly, J. J., & Levin, L. A. (2004). Global distribution of naturally occurring marine hypoxia on continental margins. *Deep Sea Research Part I: Oceanographic Research Papers*, *51*(9), 1159–1168. https://doi.org/10.1016/j.dsr.2004.03.009

Hofmann, E. E., Cahill, B., Fennel, K., Friedrichs, M. A. M., Hyde, K., Lee, C., et al. (2011). Modeling the dynamics of continental shelf carbon. *Annual Review of Marine Science*, *3*(1), 93–122. https://doi.org/10.1146/annurev-marine-120709-142740

Irby, I. D., Friedrichs, M. A. M., Da, F., & Hinson, K. (2018). The competing impacts of climate change and nutrient reductions on dissolved oxygen in Chesapeake Bay. *Biogeosciences*, *15*, 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., et al. (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. *Journal of Marine Systems*, *76*(1-2), 64–82. https://doi.org/10.1016/j.jmarsys.2008.05.014

Kemp, W. M., Boynton, W. R., Adolf, J. E., Boesch, D. F., Boicourt, W. C., Brush, G., et al. (2005). Eutrophication of Chesapeake Bay: Historical trends and ecological interactions. *Marine Ecology Progress Series*, *303*, 1–29. https://doi.org/10.3354/meps303001

Lee, Y. J., & Lwiza, K. M. M. (2008). Characteristics of bottom dissolved oxygen in Long Island sound, New York. *Estuarine, Coastal and Shelf Science*, *76*(2), 187–200. https://doi.org/10.1016/j.ecss.2007.07.001

Li, Y., Li, M., & Kemp, W. M. (2015). A budget analysis of bottom-water dissolved oxygen in Chesapeake Bay. *Estuaries and Coasts*, *38*(6), 2132–2148. https://doi.org/10.1007/s12237-014-9928-9

Luettich, R. A. Jr., Wright, L. D., Nichols, C. R., Baltes, R., Friedrichs, M. A. M., Kurapov, A., et al. (2017). A test bed for coastal and ocean modeling. *Eos*, *98*. https://doi.org/10.1029/2017EO078243

Luettich, R. A., Wright, L. D., Signell, R., Friedrichs, C., Friedrichs, M. A. M., Harding, J., et al. (2013). Introduction to special section on the US IOOS coastal and ocean modeling testbed. *Journal of Geophysical Research: Oceans*, *118*, 6319–6328. https://doi.org/10.1002/2013JC008939

Muller, A. C., Muller, D. I., & Muller, A. (2016). Resolving spatiotemporal characteristics of the seasonal hypoxia cycle in shallow estuarine environments of the Severn River and South River, MD, Chesapeake Bay, USA. *Heliyon*, *2*(9), e00157. https://doi.org/10.1016/j.heliyon.2016.e00157

Murphy, R. R., Kemp, W. M., & Ball, W. P. (2011). Long-term trends in Chesapeake Bay seasonal hypoxia, stratification, and nutrient loading. *Estuaries and Coasts*, *34*(6), 1293–1309. https://doi.org/10.1007/s12237-12011-19413-12237

Newton, J. A., Bassin, C., Devol, A., Richey, J., Kawase, M., & Warner, M. (2011). Hood canal dissolved oxygen program integrated assessment and modeling report: I. Overview and results synthesis Rep. (31 pp.). Seattle, WA.

Nordberg, K., Filipsson, H. L., Gustafsson, M., Harland, R., & Roos, P. (2001). Climate, hydrographic variations and marine benthic hypoxia in Koljo Fjord, Sweden. *Journal of Sea Research*, *46*(3-4), 187–200. https://doi.org/10.1016/S1385-1101(01)00084-3

Nurnberg, G. K. (1995). Quantifying anoxia in lakes. *Limnology and Oceanography*, *40*(6), 1100–1111. https://doi.org/10.4319/lo.1995.40.6.1100

Nurnberg, G. K. (2004). Quantified hypoxia in lakes and reservoirs. *The Scientific World*, *18*(4), 299–306. https://doi.org/10.1080/07438140209353936

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

Parker-Stetter, S. L., & Horne, J. K. (2009). Nekton distribution and midwater hypoxia: A seasonal, diel prey refuge? *Estuarine, Coastal and Shelf Science*, *81*(1), 13–18. https://doi.org/10.1016/j.ecss.2008.09.021

Patrick, C. J., & Weller, D. E. (2015). Interannual variation in submerged aquatic vegetation and its relationship to water quality in subestuaries of Chesapeake Bay. *Marine Ecology Progress Series*, *537*, 121–135. https://doi.org/10.3354/meps11412

Rabalais, N. N., Turner, R. E., Gupta, B. K. S., Boesch, D. F., Chapman, P., & Murrell, M. C. (2007). Hypoxia in the northern Gulf of Mexico: Does the science support the plan to reduce, mitigate, and control hypoxia? *Estuaries and Coasts*, *30*(5), 753–772. https://doi.org/10.1007/BF02841332

Rabalais, N. N., Turner, R. E., & Wiseman, W. J. (2002). Gulf of Mexico hypoxia, a.k.a. "the dead zone". *Annual Review of Ecology and Systematics*, *33*(1), 235–263. https://doi.org/10.1146/annurev.ecolsys.33.010802.150513

Sanford, L. P., Sellner, K. G., & Breitburg, D. L. (1990). Covariability of dissolved oxygen with physical processes in the summertime Chesapeake Bay. *Journal of Marine Research*, *48*(3), 567–590. https://doi.org/10.1357/002224090784984713

Scavia, D., Bertanio, I., Obenour, D. R., Turner, R. E., Forrest, D. R., & Katin, A. (2017). Ensemble modeling informs hypoxia management in the northern Gulf of Mexico. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(33), 8823–8828. https://doi.org/10.1073/pnas.1705293114

Scavia, D., Kelly, E. L. A., & Hagy, J. D. (2006). A simple model for forecasting the effects of nitrogen loads on Chesapeake Bay hypoxia. *Estuaries and Coasts*, *29*(4), 674–684. https://doi.org/10.1007/BF02784292

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

Scully, M. E. (2013). Physical controls on hypoxia in Chesapeake Bay: A numerical modeling study. *Journal of Geophysical Research: Oceans*, *118*, 1239–1256. https://doi.org/10.1002/jgrc.20138

Testa, J. M., Murphy, R. R., Brady, D. C., & Kemp, W. M. (2018). Nutrient- and climate-induced shifts in the phenology of linked biogeochemical cycles in a temperate estuary. *Frontiers in Marine Science*, *5*. https://doi.org/10.3389/fmars.2018.00114

Testa, M. T., Clark, B., Dennison, W. C., Donovan, E. C., Fisher, A. W., Ni, W., et al. (2017). Ecological forecasting and the science of hypoxia in Chesapeake Bay. *Bioscience*, *67*(7), 614–626. https://doi.org/10.1093/biosci/bix048

USEPA (2003). Ambient water quality criteria for dissolved oxygen, water clarity and chlorophyll *a* for the Chesapeake Bay and its tidal tributaries Rep. (343 pp.). U.S. Environmental Protection Agency Region III, Chesapeake Bay Program Office, Annapolis, MD.

Xia, M., & Jiang, L. (2016). Application of an unstructured grid-based water quality model to Chesapeake Bay and its adjacent coastal ocean. *Journal of Marine Science and Engineering*, *4*(3). https://doi.org/10.3390/jmse4030052

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. *Estuaries and Coasts*, *35*(1), 237–261. https://doi.org/10.1007/s12237-011-9423-5

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