Combining observations and numerical model results to improve estimates of hypoxic volume within the Chesapeake Bay, USA

Aaron J. Bever,[1,2] Marjorie A. M. Friedrichs,[1] Carl T. Friedrichs,[1] Malcolm E. Scully,[3] and Lyon W. J. Lanerolle[4]

[1] The overall size of the "dead zone" within the main stem of the Chesapeake Bay and its tidal tributaries is quantified by the hypoxic volume (HV), the volume of water with dissolved oxygen (DO) less than 2 mg/L. To improve estimates of HV, DO was subsampled from the output of 3-D model hindcasts at times/locations matching the set of 2004–2005 stations monitored by the Chesapeake Bay Program. The resulting station profiles were interpolated to produce bay-wide estimates of HV in a manner consistent with nonsynoptic, cruise-based estimates. Interpolations of the same stations sampled synoptically, as well as multiple other combinations of station profiles, were examined in order to quantify uncertainties associated with interpolating HV from observed profiles. The potential uncertainty in summer HV estimates resulting from profiles being collected over 2 weeks rather than synoptically averaged ~5 km^3. This is larger than that due to sampling at discrete stations and interpolating/extrapolating to the entire Chesapeake Bay (2.4 km^3). As a result, sampling fewer, selected stations over a shorter time period is likely to reduce uncertainties associated with interpolating HV from observed profiles. A function was derived that when applied to a subset of 13 stations, significantly improved estimates of HV. Finally, multiple metrics for quantifying bay-wide hypoxia were examined, and cumulative hypoxic volume was determined to be particularly useful, as a result of its insensitivity to temporal errors and climate change. A final product of this analysis is a nearly three-decade time series of improved estimates of HV for Chesapeake Bay.

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

1. Introduction

[2] The prevalence and spatial extent of low dissolved oxygen (DO) concentrations in estuarine and marine environments has been increasing for at least 50 years [*Diaz and Rosenberg*, 2008; *Gilbert et al.*, 2010]. Although hypoxia does occur naturally, as in the Baltic Sea [*Conley et al.*, 2009], the eastern Pacific Ocean [*Helly and Levin*, 2004], and fjord systems [*Nordberg et al.*, 2001], much of the recent increase in occurrence is believed to be caused by anthropogenic effects [*Cloern*, 2001; *Diaz and Rosenberg*, 2008], with a greater increase in the number of hypoxic coastal systems than of hypoxic open ocean systems [*Gilbert et al.*, 2010]. Examples include the Chesapeake Bay [*Officer et al.*, 1984], Gulf of Mexico [*Rabalais et al.*, 2007], Hood Canal [*Parker-Stetter and Horne*, 2009], Lake Erie [*Hawley et al.*, 2006], and Long Island Sound [*Lee and Lwiza*, 2008]. In this study we specifically focus on the presence of hypoxia in the Chesapeake Bay (hereafter also referred to as the Bay), a 12,000 km^2 estuary with a 170,000 km^2 watershed, relatively weak tidal mixing, and a central channel exceeding 25 m in depth, making it especially vulnerable to nutrient enrichment and hypoxia [*Linker et al.*, 2008].

[3] Hypoxia, defined as DO < 2.0 mg L^{-1}, was observed in the main channel of the Chesapeake Bay as early as the 1930s [*Newcombe and Horne*, 1938], and severe anoxia (DO < 0.2 mg L^{-1}) was observed at least as early as 1984 [*Seliger et al.*, 1984]. Hypoxia and anoxia are now observed every summer in large portions of the Bay [*Hagy et al.*, 2004]. In general, phytoplankton supply organic matter to the deep main stem (MS), which is isolated from
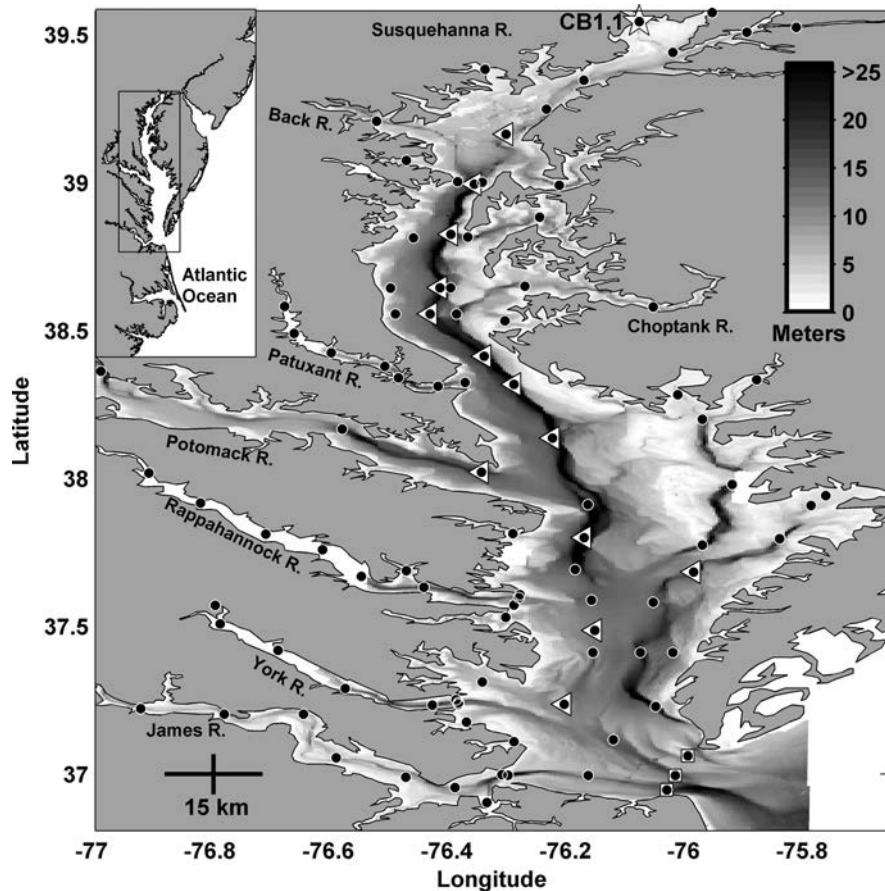
Figure 1. Bathymetry and spatial extent of Chesapeake Bay and its tributaries. The bathymetry is shaded. Circles are the CBP station locations; triangles mark the 13 stations used as a station set in this paper; squares show the CBP stations used for CH3D and ICM open boundary conditions; the star at the northern extent of the Bay marks station CB1.1. The aspect ratio of the Bay is stretched in the east-west direction, to better show the bathymetry and station locations.

oxygenated surface water by stratification in the spring [*Taft et al*., 1980]. The decay of organic matter in the isolated bottom water draws down the DO, leading to the observed hypoxic conditions [*Officer et al*., 1984; *Taft et al*., 1980]. Stronger winds increase lateral and vertical mixing in the fall [*Goodrich et al*., 1987], oxygenating the bottom water. The volume of low DO water also shows significant correlations to freshwater and nutrient input to the Bay and to wind direction, indicating that these factors may also be driving year-to-year variations in hypoxia [*Hagy et al*., 2004; *Murphy et al*., 2011; *Scully*, 2010b].

[4] The duration and extent of hypoxia in modern systems is generally estimated directly using measured DO (as in *DiMarco et al*. [2010], *Hagy et al*. [2004], *Lee and Lwiza* [2008], and *Officer et al*. [1984]); however, field observations of DO are inevitably limited in their spatial and/or temporal resolution. Current hypoxia observations within the Chesapeake Bay include DO profiles collected through the Chesapeake Bay Program (CBP; http://www.chesapeakeBay.net/data_waterquality.aspx). Roughly 30–60 stations are sampled monthly to semimonthly throughout the year, giving good spatial coverage of the entire Bay system (Figure 1 and Table 1). Vertical profiles of DO, temperature, salinity, chlorophyll *a*, pH, and nutrient data are collected at

the sampled stations. It takes anywhere from 7 to 14 days to collect the data from the stations sampled during a specific cruise using multiple vessels from two institutions; thus, the observations are not a synoptic representation of the system. The DO profiles are converted to different metrics that compare the amount of hypoxia from year to year and act as indicators of the health of the Bay, with the volume of hypoxic water currently being a widely used metric (as in *Hagy et al*. [2004] and *Murphy et al*. [2011]).

[5] As computing power has increased, models have become critical in efforts to extend DO observations in space and time and to determine the overall extent of hypoxia and the processes responsible for oxygen drawdown [*Cerco and Cole*, 1993; *Cerco and Noel*, 2004, 2005; *Hetland and DiMarco*, 2008; *Liu and Scavia*, 2010; *Pena et al*., 2010; *Scully*, 2010a]. For Chesapeake Bay alone, there are several different 3-D numerical model implementations investigating the driving processes behind observed DO. *Scully* [2010a] used the ChesROMS hydrodynamic model with a single-term DO formulation to show that hypoxia within the Chesapeake Bay was strongly influenced by wind direction during the summer months. *Cerco* [1995] and *Linker et al*. [2008] used the CBP model to investigate the effects of reduced nutrient loads on DO within

Table 1. All the Considered Chesapeake Bay Program Station Locations, With the Station Names, Depths (m), Latitude/Longitude, and the General Location Within the Estuary[a]

Name	Latitude	Longitude	Depth	Location
CB1.1	−76.081340	39.545113	6	STEM
CB2.1	−76.024666	39.440113	6	STEM
CB2.2	−76.174675	39.346775	13	STEM
CB3.1	−76.238010	39.248444	19	STEM
CB3.2	−76.306340	39.163445	12	STEM
CB3.3C	−76.359680	38.995113	26	STEM
CB3.3E	−76.346344	39.001778	9	STEM
CB3.3W	−76.388016	39.003445	9	STEM
CB4.1C	−76.399670	38.825115	33	STEM
CB4.1E	−76.371060	38.816505	25	TRIB
CB4.1W	−76.462730	38.813450	9	STEM
CB4.2C	−76.417730	38.644840	28	STEM
CB4.2E	−76.399950	38.644840	13	STEM
CB4.2W	−76.501340	38.643450	9	STEM
CB4.3C	−76.434670	38.556510	28	STEM
CB4.3E	−76.389670	38.556510	23	STEM
CB4.3W	−76.493010	38.556510	9	STEM
CCB4.4	−76.343000	38.413177	32	STEM
CB5.1W	−76.375500	38.325123	9	STEM
CB5.1	−76.293000	38.318455	36	STEM
CB5.2	−76.228000	38.136790	33	STEM
CB5.3	−76.168000	37.911793	29	STEM
CB5.4W	−76.294670	37.813465	5	STEM
CB5.4	−76.174660	37.800130	33	STEM
CB5.5	−76.189670	37.691800	23	STEM
CB6.1	−76.162160	37.588467	13	STEM
CB6.2	−76.156330	37.486800	14	STEM
CB6.3	−76.159660	37.411526	16	STEM
CB6.4	−76.207990	37.236526	12	STEM
CB7.1N	−75.974660	37.775127	33	STEM
CB7.1S	−76.057990	37.581240	17	STEM
CB7.1	−75.989660	37.683464	26	STEM
CB7.2E	−76.024660	37.411526	13	TRIB
CB7.2	−76.079660	37.411526	23	TRIB
CB7.3E	−76.053825	37.228752	27	STEM
CB7.3	−76.125210	37.116810	16	STEM
CB7.4N	−75.999374	37.062366	15	STEM
CB7.4	−76.020485	36.995700	17	STEM
CB8.1E	−76.034935	36.947370	19	STEM
CB8.1	−76.167720	36.995422	15	STEM
EE1.1	−76.249670	38.883450	13	TRIB
EE2.1	−76.274666	38.650120	9	TRIB
EE2.2	−76.308010	38.533455	14	TRIB
EE3.0	−76.016330	38.283455	8.2	TRIB
EE3.1	−75.974655	38.200123	13.6	TRIB
EE3.2	−75.924670	37.980103	28.6	TRIB
EE3.3	−75.766320	37.941790	3.7	TRIB
EE3.4	−75.791320	37.908460	7	TRIB
EE3.5	−75.844376	37.796516	29	STEM
ET1.1	−75.958000	39.575115	3	TRIB
ET2.1	−75.816320	39.525112	14.6	TRIB
ET2.3	−75.899666	39.508450	13.6	TRIB
ET4.2	−76.216340	38.991780	19.8	TRIB
ET5.2	−76.058000	38.580120	13.9	TRIB
ET6.2	−75.882990	38.333454	4	TRIB
LE1.1	−76.601620	38.425125	12	TRIB
LE1.2	−76.511060	38.378735	18	TRIB
LE1.3	−76.488010	38.340680	25.5	TRIB
LE1.4	−76.421340	38.312070	17	TRIB
LE2.2	−76.583015	38.166794	14.4	TRIB
LE2.3	−76.347725	38.021515	21	TRIB
LE3.1	−76.615920	37.759250	11	TRIB

Table 1. (continued)

Name	Latitude	Longitude	Depth	Location
LE3.2	−76.550470	37.669140	16	TRIB
LE3.3	−76.474420	37.688420	7	TRIB
LE3.4	−76.444860	37.631890	22	TRIB
LE3.6N	−76.283000	37.606800	4	TRIB
LE3.6S	−76.293000	37.572630	4	TRIB
LE3.6	−76.284670	37.596798	12	TRIB
LE3.7	−76.306610	37.530690	7	TRIB
LE4.1	−76.691250	37.418830	11	TRIB
LE4.2	−76.578110	37.290440	18	TRIB
LE4.3	−76.430940	37.233920	22	TRIB
LE5.1	−76.648330	37.202970	15	TRIB
LE5.2	−76.593060	37.056000	10	TRIB
LE5.3	−76.475440	36.990440	9	TRIB
LE5.4	−76.392750	36.954860	18	TRIB
LE5.5-W	−76.313280	36.999030	15	TRIB
LE5.5	−76.303000	36.996810	27	TRIB
LE5.6	−76.338360	36.904560	18	TRIB
RET1.1	−76.664130	38.490677	13	TRIB
RET2.4	−76.990530	38.362625	17.8	TRIB
RET3.1	−76.822200	37.917300	8	TRIB
RET3.2	−76.711950	37.811580	7	TRIB
RET4.2	−76.797150	37.571250	17	TRIB
RET4.3	−76.788890	37.508690	7	TRIB
RET5.2	−76.782190	37.202940	15	TRIB
TF1.7	−76.680240	38.581787	4	TRIB
TF2.1	−77.048590	38.706505	21	TRIB
TF3.3	−76.909280	38.018470	9	TRIB
TF5.6A	−76.923294	37.221817	8	TRIB
WE4.1	−76.346340	37.311806	6	TRIB
WE4.2N	−76.390500	37.251810	7	TRIB
WE4.2S	−76.386340	37.236810	6	TRIB
WE4.2	−76.386340	37.241810	18	TRIB
WE4.3	−76.373000	37.176810	6	TRIB
WE4.4	−76.293000	37.110146	8	TRIB
WT2.1	−76.341620	39.383442	2	TRIB
WT5.1	−76.524690	39.208443	16.2	TRIB
WT6.1	−76.474686	39.075110	6	TRIB

[a]The reported depth is the maximum reported in CBP profiles from each station. STEM and TRIB refer to the station location within the main stem of the Bay or one of the tributaries.

Chesapeake Bay, ultimately driving management decisions which potentially may cost in the tens of billions of dollars [*Maryland*, 2012; *Commonwealth of Virginia*, 2010]. The authors are also currently using a combination of eight model implementations to investigate seasonal correlations between bottom DO and water column stratification.

[6] The goal of the present study is to use multiple model simulations combined with field measurements to better represent the spatial extent and temporal variability of DO within the Chesapeake Bay estuarine system. To this end, this study used a collection of four 3-D numerical hydrodynamic and DO model implementations, combined with DO measurements, to investigate optimal DO sampling strategies within the Chesapeake Bay. The main objectives of this paper are (1) to determine to what degree the current observational DO sampling strategy captures the total volume of hypoxic water within the Bay, (2) to determine if there is a minimum number of optimal stations that can be used to accurately calculate the hypoxic volume (HV) within the Bay, (3) to use results from the multiple

numerical models to generate a 29 year time series of likely improved estimates of hypoxic volume from observed DO concentrations, and (4) to investigate metrics other than hypoxic volume that could be used for characterizing total hypoxia within Chesapeake Bay.

2. Methods

[7] In this study, results from multiple numerical models are combined with dissolved oxygen measurements in the Chesapeake Bay in order to better understand hypoxic volume variability in the Bay. The hypoxic volumes calculated here only include portions of the Bay and its tributaries that have bottom depths greater than 2 m. Thus, we do not address the occurrence of hypoxic events generated in shallower waters resulting from diel variations in respiration and productivity, among other factors [i.e., *Breitburg*, 1990]. Because these shallow waters (<2 m) comprise only about 2.5 km^3 (~3%) of the total volume of the Bay, this omission results in a relatively small error in our total Bay hypoxic volumes. The occurrence of shallow water hypoxia, however, is one of the key indicators of ecosystem health and thus should be considered in future modeling studies attempting to better understand water quality in the Chesapeake Bay and its tributaries.

[8] A brief introduction to the hydrodynamic-DO models used in these analyses is provided below (section 2.1) as well as the calculations of hypoxic volume (section 2.2). A method to scale interpolated hypoxic volumes from discrete station sets to better match the true HV within the Bay is then discussed (section 2.3) and is followed by a discussion of other useful methods for quantifying hypoxia within the Bay (duration and spatial extent of hypoxia, cumulative hypoxic volume (CHV), section 2.4). This section concludes with an explanation of the target diagrams and other model skill metrics used to determine model proficiency (section 2.5).

2.1. Three-Dimensional Numerical Models

[9] The backbone of the numerical model estimates of HV are the hydrodynamic models that estimate the water characteristics and currents necessary for the DO calculations. Two 3-D hydrodynamic numerical models are used in this study: (1) The Chesapeake Bay Program's Curvilinear Hydrodynamics in Three Dimensions (CH3D) model and (2) the Regional Ocean Modeling System (ROMS). We term model simulations using different model grids, forcing conditions, or DO formulations as DO model "implementations." These DO model implementations were used to hindcast DO throughout the Bay for the calendar years 2004 and 2005. The years 2004 and 2005 were chosen to represent relatively wet and average years, respectively, and are close enough in time that the effects of long-term climate change and land-use change between the 2 years can be assumed to be negligible.

2.1.1. Hydrodynamic Models

[10] The CH3D model used a curvilinear boundary-fitted horizontal grid with 11,064 horizontal cells and a five-foot (1.52 m) Cartesian vertical grid with a maximum of 19 vertical layers [*Cerco et al.*, 2010; *Johnson et al.*, 1991; *Wang and Johnson*, 2000]. Turbulence was modeled using the *k*-ε

turbulence closure scheme. The open boundary with the Atlantic Ocean was set at the mouth of the Chesapeake Bay, with open boundary conditions based on CBP monitoring observations at three stations spanning the Bay mouth (Figure 1; P. Wang, personal communication, 2013). Freshwater flow from above the fall line was from the U.S. Geological Survey, and freshwater input below the fall line was from the CBP watershed model [*Donigian et al.*, 1994; *Linker et al.*, 2000]. Wind velocity was taken from the Thomas Point Light, Patuxent Naval Station, Richmond International Airport, Norfolk International Airport, and DC National Airport, then scaled to represent the magnitude over water and interpolated to the model grid [*Cerco et al.*, 2010].

[11] The ROMS model used a curvilinear horizontal grid and a stretched terrain-following vertical grid with 20 vertical layers [*Haidvogel et al.*, 2008; *Shchepetkin and McWilliams*, 2005]. ROMS results were based on both the Chesapeake Bay ROMS Community Model (ChesROMS) [*Xu et al.*, 2012] and the Chesapeake Bay Operational Forecast System (CBOFS) [*Lanerolle et al.*, 2009, 2011] implementations. Both versions of ROMS used the *k*-ω turbulence closure, and temperature and salinity open boundary conditions were based on the World Ocean Atlas 2001. Tidal forcing was based on the Advanced CIRCulation model (ADCIRC), with nontidal water levels accounted for using observations along the Atlantic Ocean coastline near the Bay mouth. Rivers were included using U.S. Geological Survey data. Meteorology, wind, and heat fluxes were a combination of model results and observations. A major difference between ChesROMS and CBOFS was the horizontal grid resolution; ChesROMS was 100 × 150 cells with about 1 to 7 km resolution; CBOFS was 332 × 291 cells with about 0.03 to 4 km resolution. In ChesROMS and CBOFS about 68% and 80% of the cells were on land, respectively, and were not included in the dynamic calculations.

2.1.2. DO Model Implementations

[12] Four DO model implementations were used in this analysis and are described as follows: (1) CH3D with the ecological Integrated Compartment Model (ICM), (2) CBOFS with a constant respiration rate DO formulation [*Scully*, 2013], (3) ChesROMS with the same constant respiration rate as in (2), and (4) ChesROMS with a depth-dependent respiration rate DO formulation.

[13] (1) CH3D + ICM: the ICM is a complex multicomponent ecological model that was forced offline with output from CH3D and has been extensively calibrated to the Chesapeake Bay [*Cerco*, 2000; *Cerco and Noel*, 2004; *Cerco et al.*, 2010; *Linker et al.*, 2000]. Simulations analyzed here (provided to us by the CBP) used 24 state variables in the water column (including physical variables, multiple algal and zooplankton groups, nitrogen, phosphorous, and silica) and a full sediment diagenesis component. The model computed algal biomass, nutrient cycling, and DO, as well as numerous additional constituents and processes (P. Wang, personal communication, 2013). For this study we were provided output from the CBP Modeling Team run number 379, which used the 11,064 horizontal grid cell version of ICM.

[14] (2) CBOFS + 1-Term and (3) ChesROMS + 1-Term: these two implementations used the ROMS

Table 2. Summary of the Different Methods of Calculating HV[a]

Hypoxic Volume Name	DO Source	DO Sample Timing	DO Sampling Locations	Notes
CBP Data HV	CBP Data	Monthly to bimonthly, ~2 week window	Entire Bay, variable stations sampled dependent on cruise	Uses all available data for a given cruise
Integrated 3-D	Models	Daily, instantaneous	Entire Bay	Total 3-D modeled HV
Absolute Match	Models	Same as CBP Data HV, ~ 2 week window	Same as CBP Data HV, entire Bay	Directly comparable to CBP Data HV
Spatial Match HV	Models	When CB1.1 was sampled, instantaneous	Same as CBP Data HV, entire Bay	Directly comparable to 3-D modeled HV
All Stations HV	Models	When CB1.1 was sampled, instantaneous	All possible CBP stations, entire Bay	Determines if more CBP stations need sampled
Main Stem HV (MS)	Models	Daily, instantaneous	10 stations in the main stem	Determines if only MS stations can capture HV
Main Stem + Flanks HV (MS + Flanks)	Models	Daily, instantaneous	10 main stem and 20 stations on the flanks of the main stem	Determines if only MS + Flanks stations can capture HV
Main Stem + Tributaries HV (MS + Tribs)	Models	Daily, instantaneous	10 main stem and 20 stations in the tributaries of the main stem	Determines if only MS + Tribs stations can capture HV
MS + Flanks + Tribs HV	Models	Daily, instantaneous	10 main stem, 20 flanks, and 20 tributary stations	Determines if MS + Flanks + Tribs stations can capture HV
Assumed Optimal HV	Models	Daily, instantaneous	32 stations throughout the Bay	Intelligently chosen attempting to capture low DO
CBP13 HV	Models	Daily, instantaneous	13 CBP stations in and around the main stem	Limits interpolated HV to only 13 CBP stations
CBP13 Observed HV	CBP Data	Monthly to bimonthly, ~2 day window	13 CBP stations in and around the main stem	Less uncertainty in HV than CBP Data HV
Upscaled CBP13 Observed HV	CBP Data	Monthly to bimonthly, ~2 day window	13 CBP stations in and around the main stem	Interpolated HV upscaled to better represent actual 3-D HV

[a]The DO sample timing gives both how frequently the HV estimates are calculated and the amount of time encompassed within the DO estimates used to calculate HV. The notes column gives extra information on each HV estimate.

hydrodynamic model and a simplified single-term DO formulation that specified a constant rate of oxygen drawdown due to respiration [*Scully*, 2010a]. As such, these implementations did not explicitly model nutrients, primary production, etc. Oxygen was added to the water at the surface assuming a constant piston velocity of 3 cm h^{-1}.

[15] (4) ChesROMS + DD: this implementation used the ChesROMS hydrodynamic model and a depth-dependent respiration rate. It was nearly identical to the constant respiration rate implementations, except the respiration rate increased with water depth such that the respiration rate was zero at the surface and increased linearly to 1.25 gO^2/m^3/d at 30 m water depth, and the surface oxygen concentration was set to saturation.

[16] Modeled DO results were provided by the original modelers in different output time steps. The full 3-D modeled DO fields were supplied as daily averaged output for the ICM model and daily instantaneous snapshots for the ROMS implementations. Time series at each of the Chesapeake Bay Program station locations were also provided and are daily averaged for the ICM model and hourly for the ROMS implementations.

2.2. Hypoxic Volume Calculations

[17] Hypoxic volume was first computed using the observed CBP profile station data (Table 2). Observations were grouped according to cruise, with each cruise encompassing 30–60 profiled stations that were collected within

±7 days of the sampling time at the Main Stem station, CB1.1. The CBP volumetric inverse distance squared interpolator program version 4.63 [*U.S. Environmental Protection Agency* (*USEPA*), 2003] was used to interpolate/extrapolate these profiles onto the entire Chesapeake Bay. This is the interpolation program the CBP has utilized for its estimation of water quality parameters throughout the Chesapeake Bay and its tributaries [*USEPA*, 2003]. The default interpolation options within the program were used in this study. 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 (a higher resolution grid was utilized in some of the smaller tributaries), using an inverse distance squared weighting of up to the nearest four surrounding vertical profiles, with a maximum allowed distance between the profiles and grid point of 25 km. If vertical profiles did not extend to the seabed, the DO value at the bottom of the profile was filled down to the seabed. The 16 cruises in both 2004 and 2005 (monthly in the fall/winter/spring, semimonthly in the summer) yielded 16 interpolated hypoxic volumes for each year. Because these HV estimates were based on the CBP interpolator plus profile data and not the numerical model simulations these are referred to as the *CBP Data HV*.

[18] Hypoxic volume was calculated from the output of each 3-D numerical model implementation using two different types of methodologies. The first method directly integrated the volumes of all grid cells with DO less than
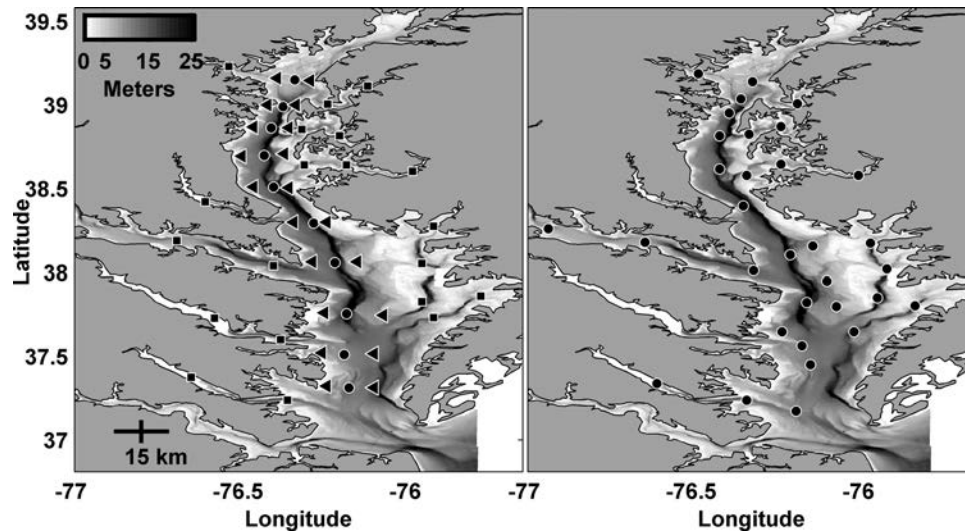
Figure 2. Locations of intelligently placed stations for efficiently capturing the 3-D hypoxic volume. (left) Locations of the (circles) Main Stem stations, (triangles) Flanks stations, and (squares) Tributary stations. (right) Locations of the Assumed Optimal stations.

2 mg L^{-1} and thus represents the total volume of hypoxic water from the modeled 3-D DO fields. This calculation, which sums directly over all the model grid points, is referred to as *Integrated 3-D HV*. The second method involving model output utilizes modeled profiles only at select stations and thus mirrors the calculation of HV from the data by using the CBP interpolator described above. This second method, which includes several variations, is collectively referred to as *Interpolated Model HV* (Table 2).

[19] As part of the Interpolated Model HV approach, nine different sets of simulated DO station profiles were used with the CBP interpolator to formulate nine different estimates of hypoxic volume. The first set of simulated station profiles matched both the time and location of the CBP Data HV stations (Absolute Match HV). In contrast, the remaining eight sets of simulated station profiles were all synoptic, i.e., all stations were sampled instantaneously. Like the first case, Cases 2 and 3 (Spatial Match and All Stations) used actual CBP station locations. In contrast, Cases 4–8 (Figure 2) were derived from hypothetical station locations that might better be able to capture the full Integrated 3-D HV. These station sets were chosen based on model results showing where the bottom DO was frequently below 2 mg L^{-1} and where the variability in the bottom DO was high (Figure 3). The ninth and final case (CBP13 HV) includes 13 CBP station locations specifically chosen to capture key spatial variability in hypoxic volume in a major section of the Bay that is regularly sampled the most synoptically by the CBP (see Figure 1). Cases 1–3 use hourly model output for the ROMS implementations and daily for CH3D + ICM and are calculated for each CBP cruise, while Cases 4–9 use the daily model output and are calculated for every day of the model simulations, for direct comparison with the daily Integrated 3-D HV.

[20] Each of these nine sets of Interpolated Model HV was calculated using the DO fields from each of the four model implementations. In summary (Table 2):

[21] 1. The *Absolute Match HV* was computed using a subset of model output, which matched the CBP Data HV profiles from each cruise in both space and time (typically 30–60 CBP stations).

[22] 2. The *Spatial Match HV* was computed using a subset of model output which matched the locations of the CBP Data HV profiles (30–60 CBP stations) but assumed a synoptic snapshot such that all model profiles were taken at the same dates/times as the observed profile at CB1.1.

[23] 3. The *All Stations HV* was computed like Case 2, except that all CBP station locations were used (99 stations, Table 1).

[24] 4. The *Main Stem HV* was computed using 10 stations down the MS of the Bay (Figure 2a).

[25] 5. The *MS + Flanks HV* was computed using the Case 4 stations plus 20 stations along the flanks of the main stem.

[26] 6. The *MS + Tribs HV* was computed using the Case 4 stations plus 20 stations in the tributaries or near tributary mouths

[27] 7. The *MS + Flanks + Tribs HV* was computed using the Case 6 stations plus the 20 flank stations.

[28] 8. The *Assumed Optimal HV* was computed using a set of 32 stations that were a priori deemed to be optimal for completely capturing the total 3-D hypoxic volume (Figure 2b).

[29] 9. The *CBP13 HV* was computed using the set of 13 CBP station locations illustrated in Figure 1. Specifically, stations CB3.2, CB3.3C, CB4.1C, CB4.2C, CB4.3C, CB4.4, CB5.1, CB5.2, CB5.4, CB6.2, CB6.4, CB7.1, and LE2.3 were used.

[30] To determine if interpolated hypoxic volumes were overly sensitive to the exact location of the stations, new station sets were also created with each station location randomly shifted north/south and east/west using a Gaussian distribution with a standard deviation of 1.5 km. This process generated insignificant changes to the interpolated hypoxic volumes and is not discussed further.
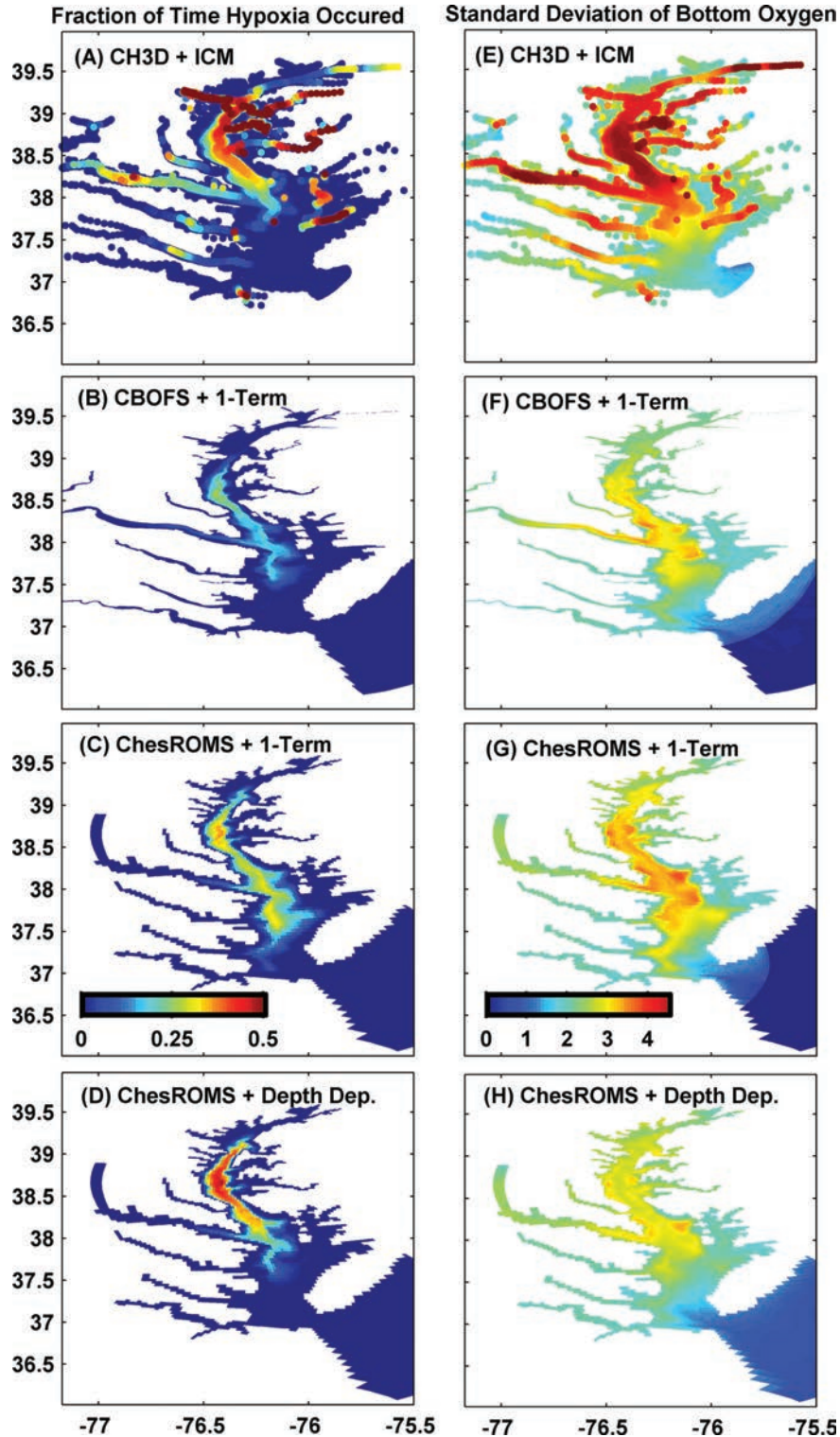
Figure 3. (left) The fraction of 2004 the bottom water column grid cell had DO less than 2 mg L^{-1}. (right) The standard deviation of DO in the bottom grid cell during 2004.

2.3. Upscaling the Interpolated Hypoxic Volumes

[31] A scaling method was developed to correct the Interpolated Model HV to better represent the Integrated 3-D HV. The first step in the upscaling of the Interpolated Model HV (*HVinterp*) cases was to define a dimensionless "correction factor" (*CF*), such that

$$HVscaled = (1 + CF)HVinterp,$$
$$\text{i.e., } CF = (HVscaled - HVinterp)/HVinterp, \quad (1)$$

where *HVscaled* is a more accurate, upscaled estimate of hypoxic volume. *CF* > 0 is needed if Interpolated Model HV underestimates 3-D Integrated HV, and *CF* < 0

is needed if Interpolated Model HV overestimates 3-D Integrated HV. Next, using output from all four numerical models over both years worth of CBP cruises, *CF* was determined for each cruise by equating *HVscaled* in (1) with that case's corresponding 3-D Integrated HV. It was seen that the correction factor, *CF*, was itself a clear (but with a large amount of scatter at lower HVs) function of the Interpolated Model HV in that *CF* became larger as Interpolated Model HV became smaller (see additional details in section 3).

[32] To account for this dependence of *CF* on Interpolated 3-D HV, an upscaling function was then fit to the many individual realizations of *CF* using

$$y = a(\log10(x)) + b, \tag{2}$$

where *x* was Interpolated Model HV (in km^3), *a* and *b* were best fit coefficients, and *y* was the best global fit for *CF*. When performing least squares fits to set the coefficients *a* and *b*, only hypoxic volumes greater than 2 km^3 were considered. Once best fit global choices for *a* and *b* had been identified, all Interpolated Model HV values were then scaled to their approximate 3-D HV values using the single relationship

$$HVscaled = (1 + [a(\log10(x)) + b])HVinterp. \tag{3}$$

[33] To prevent occasional excessive reductions in HV, the maximum that large Interpolated Model HV could be reduced by was set to 25% of the corresponding original Interpolated Model HV. Below 2 km^3, *CF* is simply set to 1. The 2 km^3 cutoff was chosen because it is large enough to occur every year and yet small enough that the *CF* does not become very large from the division by the interpolated HV (*HVinterp*) in equation (1). Also, below about 2 km^3 the scatter in the *CF* becomes increasingly large, implying this scaling method is not appropriate for low hypoxic volumes.

2.4. Other Metrics for Quantifying Hypoxia Severity

[34] In addition to hypoxic volume, five other measures of quantifying hypoxia severity were investigated. These include estimates of hypoxic area and hypoxic thickness, which were computed using the 3-D model output, as well as hypoxic duration, maximum hypoxic volume, and cumulative hypoxic volume, which provided a single value for each year.

[35] Hypoxic area was estimated from the model results by summing the horizontal surface area of each bottom grid cell that had DO less than 2 mg L^{-1}. As such, it was a representation of the 2-D area of the Bay undergoing hypoxic conditions. The Integrated 3-D HV was divided by the hypoxic area in order to calculate the average thickness of water that was hypoxic each day. The duration of hypoxia was defined as the amount of time the hypoxic volume was greater than 2 km^3. The calculation of the hypoxic duration was simply the number of days between the first and last hypoxic volume in each year greater than this 2 km^3 cutoff. If the hypoxic volume dropped below the threshold and then returned above the threshold, the amount of time below the threshold was still counted in the duration of hypoxia. This gives an estimate of the amount of time between the first and last instances of hypoxic volume greater than 2 km^3 in any given year.

[36] The cumulative hypoxic volume (km^3 days) integrated over a year conveniently incorporates the hypoxic volume and the duration of hypoxia into a single metric. The cumulative hypoxic volume was calculated by multiplying the hypoxic volumes by the length of time they occurred and then summing these products over a full year. The interpolated hypoxic volumes were treated as a step function, where the interpolated hypoxic volume was multiplied by the number of days between the date on which the hypoxic volume was interpolated and the date of the next set of profiles. A temporal interpolation was not done, because the 3-D output from the models showed that the total hypoxic volume did not behave in a manner that could be represented by interpolating between the cruises.

2.5. Methods for Model Skill Comparison

[37] Model skill was evaluated using the metric presented by *Wilmott* [1981],

$$S_W = 1 - \frac{\sum_{i=1}^{N} |X_{Ei} - X_{Ri}|^2}{\sum_{i=1}^{N} \left(|X_{Ei} - \overline{X_E}| + |X_{Ri} - \overline{X_R}|\right)^2} \tag{4}$$

where S_W is the Wilmott skill, *X* is the variable being compared, and $\overline{X}$ is the time-averaged value. Subscripts *E* and *R* represent estimated (usually modeled) and reference (generally observed) values, respectively. Perfect agreement between the estimates and the references yields a skill of one, and skill decreases toward zero as the two diverge.

[38] Graphical methods of evaluating model skill are helpful when trying to illustrate the relative skill of multiple model formulations. *Jolliff et al.* [2009] and *Hofmann et al.* [2011] provide a detailed description of target diagrams and their use in model skill assessment. In short, target diagrams show the total root-mean-squared difference (RMSD) as the distance from the origin, the bias between 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 estimates and the reference values, thus illustrating whether the estimated variability is higher (positive *X* axis) or lower (negative *X* axis) than the reference variability. The bias and ubRMSD are often normalized by the standard deviation of the reference values to aid in the comparison of different variables or years. After normalization, any point falling inside the circle of radius one performed better than simply using the mean of the reference values.

3. Results

[39] Verification of the modeled bottom DO is presented in section 3.1. Relative model skill in estimating the CBP Data HV is assessed by applying the same interpolation scheme to both the observed DO profile data and the model equivalents of these data (section 3.2). Once it has been ascertained that the models have skill in reproducing the observations, the models are used to estimate the ability of the CBP interpolator routine to capture the full 3-D hypoxic volume in the Bay. Various numbers and locations of select
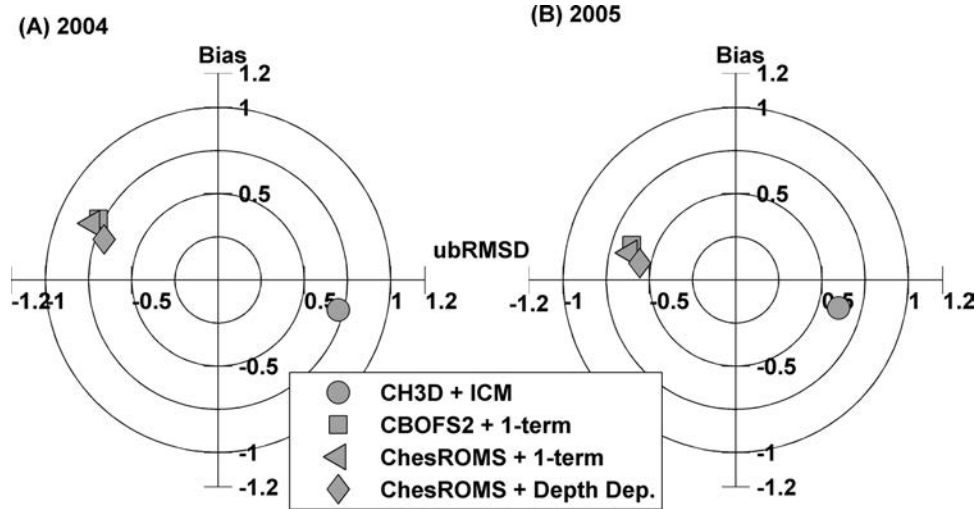
Figure 4. Normalized target diagrams comparing the bottom DO concentration from the models to the CBP monitoring observations.

stations are used as inputs to the interpolator with the aim of determining which stations are most critical to sample on each cruise (section 3.3). In section 3.4, an algorithm is applied that can be used to upscale the output of the interpolator to better capture the full hypoxic volume, even when only a subset of select stations are available for interpolation. In section 3.5, a variety of alternative measures of quantifying hypoxia severity are examined, several of which build on the upscaling presented earlier. Finally, these metrics are applied to 29 years of dissolved oxygen profile observations in the Chesapeake Bay (section 3.6).

3.1. Verification of Modeled Bottom DO

[40] Model skill at reproducing the observed bay-wide DO distributions is here limited to determining how well the models reproduced DO on a monthly to bimonthly time period. All four of the model implementations proved to be skillful at reproducing the seasonal trend in bottom DO throughout the Chesapeake Bay in both 2004 and 2005

(Figure 4). Although higher frequency observations of deep water DO were not available for systematic model-data comparison in the context of this present study, the models examined here all displayed tidal to weekly variations in the DO field that were qualitatively consistent with higher frequency DO observations collected at other times [i.e., *Sanford et al*., 1990].

3.2. HV Computed From Interpolation of Observed DO (CBP Data HV) Versus Interpolation of Simulated DO (Absolute Match HV)

[41] Hypoxic volume was first computed by applying the same interpolation scheme to both the observations (CBP Data HV) and to the model equivalents of those observations (Absolute Match HV). In general, the model-generated Absolute Match HV successfully reproduced the mean CBP Data HV for all four model implementations (Figure 5). Interestingly, the ROMS-based models all overestimate the temporal variability of HV (i.e., their output
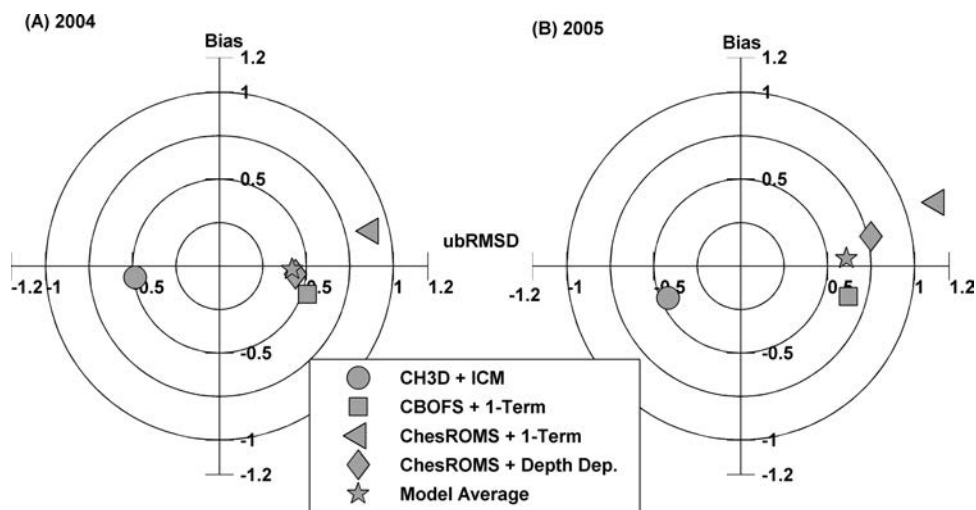

Figure 5. Normalized target diagrams comparing the hypoxic volumes calculated from the models (Absolute Match HV) to those calculated from the observations (CBP Data HV).

Table 3. Maximum, Mean, and Standard Deviation of Hypoxic Volume[a]

Year	Source of Station DO	Oxygen Formulation	Max	Jan–Dec HV (km³)			May–Aug HV (km³)		
				Mean	Standard Deviation	Wilmott Skill	Mean	Standard Deviation	Wilmott Skill
2004	CH3D-ICM	Full ecological Model	9.3	2.3	2.9	0.92	4.2	2.9	0.85
	CBOFS2	1-Term	13.6	2.1	3.9	0.94	4.1	4.9	0.90
	ChesROMS	1-Term	19.7	3.3	5.8	0.89	6.6	7.0	0.82
	ChesROMS	Depth-dependent respiration	11.4	2.4	3.5	0.95	4.7	3.7	0.91
	Model average	NA	13.5	2.5	3.9	0.93	5.0	4.4	0.93
	Observations	NA	10.0	2.6	3.4	NA	5.0	3.4	NA
2005	CH3D-ICM	Full ecological model	8.9	2.6	3.2	0.92	3.5	3.0	0.88
	CBOFS2	1-Term	12.8	2.7	4.4	0.90	5.3	4.5	0.88
	ChesROMS	1-Term	22.0	4.9	7.4	0.82	9.6	8.1	0.67
	ChesROMS	Depth-dependent respiration	15.0	4.1	5.6	0.89	7.5	6.1	0.78
	Model average	NA	14.7	3.6	5.1	0.92	6.8	5.5	0.84
	Observations	NA	11.2	3.4	3.9	NA	6.3	3.5	NA

[a]The Wilmott skill of the four models in terms of their ability to reproduce the hypoxic volume computed from observed station DO profiles (CBP Data HV) using simulated station DO profiles (Absolute Match HV). Statistics for the average of the four models are also presented.

all falls on the right-hand side of the target diagrams), whereas CH3D + ICM underestimates this variability (and is on the left-hand side of the target diagrams; Figure 5).

[42] Most symbols lie well within the outer circle of the target diagrams (Figure 5) and have high Wilmott skill values (Table 3). Although CH3D + ICM performed similarly well in 2004 and 2005, the ROMS-based models all performed better in 2004 than in 2005. Not surprisingly, skill scores were also higher when the full year time series (January–December) was used, rather than just the summer month time series (1 June to 30 September), since all models were able to successfully reproduce the near zero winter hypoxic volumes. Although all models performed reasonably well in terms of reproducing the seasonal cycle of HV, it is interesting to note that when the 2004 time series of hypoxic volumes from the four models were averaged, the resulting HV generally matched the CBP Data HV better than did individual models (Table 3 and Figure 5).

3.3. Integrated 3-D HV Versus Interpolated Model HV

[43] Hypoxic volumes computed from summing the volume of all grid cells with DO <2 mg L^{-1} (i.e., Integrated 3-D HV, shown by continuous blue lines in Figure 6) were almost always greater than or equal to those interpolated from the model equivalents of the data (Absolute Match HV, black circles in Figure 6), with the maximum differences being nearly 50% of the total hypoxic volume (up to 7 km³). The method used here to calculate the Absolute Match HV mimics that used by the CBP when the inverse distance squared interpolator is applied to cruise data collected over 1–2 weeks. There are two distinct contributions to this mismatch between model output sampled and interpolated as if it were a series of station profiles and the "true" time series of hypoxic volume output by each model: (i) spatial interpolation error and (ii) temporal interpolation error.

[44] The spatial error in hypoxic volume associated with interpolating over a finite number of stations was isolated by utilizing model output to instantaneously "sample" all stations from a given cruise period at a single moment during the cruise rather than matching the timing of each individual DO profile. These instantaneous (i.e., synoptic) station profiles were then entered into the CBP interpolator to generate the time series termed Spatial Match HV (red circles in Figure 6). The difference between the Spatial Match HV and the Integrated 3-D HV for the corresponding moments in time is then an estimate of the spatial error associated with the interpolation. The magnitude of this spatial interpolation error contribution among the four models for May–August 2004 and 2005, indicated by the series of blue dots in Figure 7, has a mean value of 2.4 km³. Surprisingly, the spatial error associated with the All Station HV, which synoptically utilized all 99 profile stations sampled at any time by the CBP from 2004 to 2005, was not statistically different from those computed by interpolation simulated DO from only ~30–60 stations (Spatial Match HV).

[45] The maximum size of the potential temporal interpolation error associated with the length of time over which profiles are taken (typically about 10 days; see horizontal black lines in Figure 6) was determined by continually computing the instantaneous Spatial Match HV over the date range of each cruise. (Contributions to this range were calculated daily for the CH3D + ICM model and every 3 h for the ROMS models.) The total range of the Spatial Match HV over the course of the cruise period was then equal to the maximum potential temporal error in the Absolute Match HV. This potential temporal error is shown by the vertical red lines in Figure 6 and by the red dots in Figure 7. The average potential temporal error for Absolute Match HV among the four models for May–August 2004 and 2005 was 4.7 km³, which is larger than the average spatial error. Thus, it is clear that the time which elapses over the course of a given cruise can significantly affect the HV estimates produced by the interpolation.

[46] As described above, hypoxic volume computed by the interpolation generally underestimates the total hypoxic volume within the Bay as represented by the dynamic models. In an attempt to determine whether alternate hypothetical station locations might do a better job of capturing the total hypoxic volume when using the CBP interpolator, HV was interpolated from six more sets of various station locations (Cases 4–9 described in section 2.2, Table 2), and all
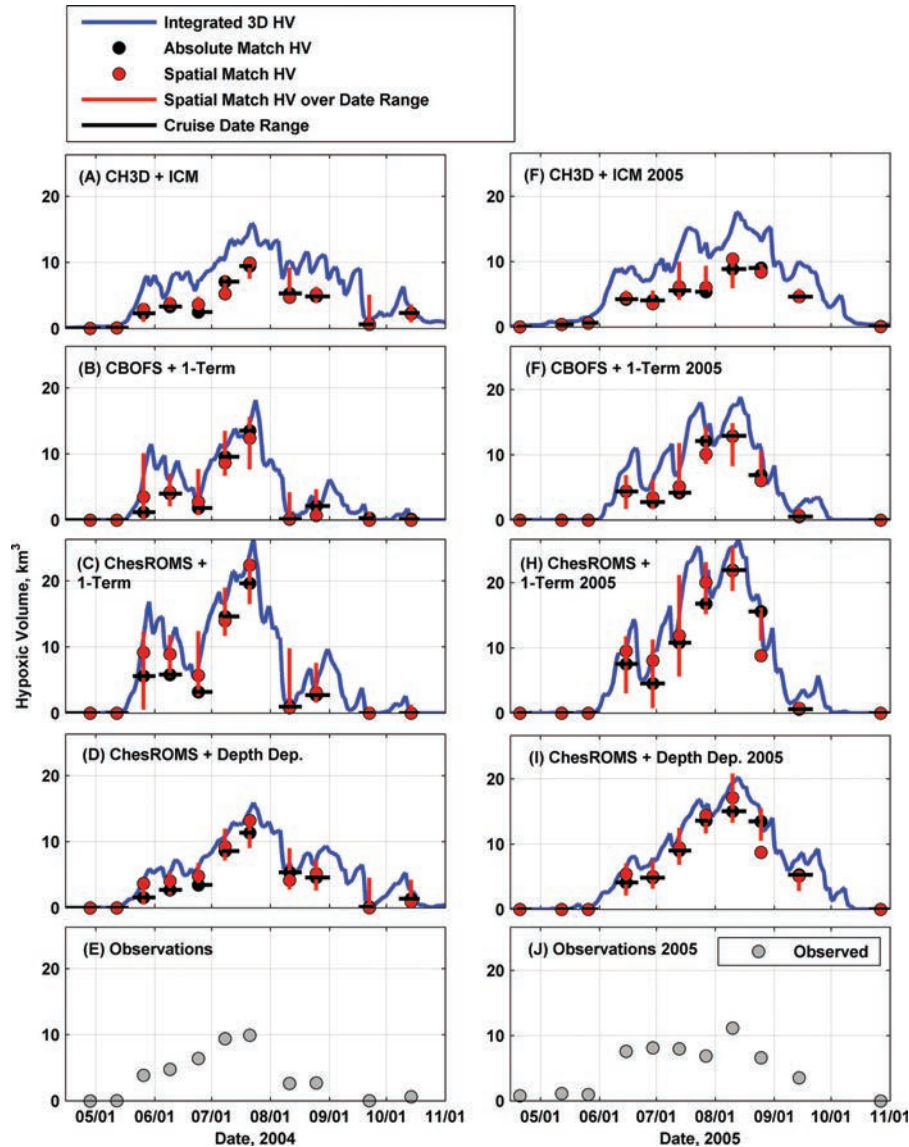
Figure 6. Comparisons of different methods for calculating hypoxic volumes in (A–E) 2004 and (F–J) 2005: the Integrated 3-D HV (continuous blue line), the Absolute Match HV (black circles), and the Spatial Match HV (red circles). See text and Table 2 for HV definitions. Horizontal lines (black) show the range of time over which the observed profiles were sampled, and vertical lines (red) show the range of Spatial Match HV computed over these date ranges.

were compared to their corresponding "true" Integrated 3-D HV. Before upscaling (see section 3.4), none of these six station sets consistently reproduced the Integrated 3-D HV most closely (Figure 8). These results suggest that the interpolation of 10–20 well-placed stations was equally as skillful as the 99 existing CBP stations in terms of reproducing total Bay HV, and thus, eliminating spatial errors (in the absence of upscaling) would likely require many hundreds of synoptically observed profile stations.

[47] We also explored whether differences in the volumes of the grids themselves could be systematically influencing the comparison of hypoxic volumes. It is conceivable that the dynamic models might consistently produce overly large Integrated 3-D HV estimates if their grid volumes were all significantly larger than the volume represented by the interpolator's grid. Although the various

model bathymetries do differ, these differences do not explain the consistently mismatched integrated versus interpolated HV. The 3-D model grid volume is smaller than that of the interpolator grid in one case (CH3D), nearly identical to the interpolator grid volume in another case (ChesROMS), and larger than the interpolator grid volume in the third case (CBOFS). Yet application of the interpolator to output from every model similarly underestimated the corresponding model's Integrated 3-D HV.

3.4. Capturing Total Hypoxic Volume in the Bay by Upscaling Interpolated HV Estimates

[48] The CBP13 station set was ultimately chosen for upscaling to the 3-D hypoxic volume for three reasons. First, the limited number of stations made for a reduced sampling time during observational cruises (~2.3 days)
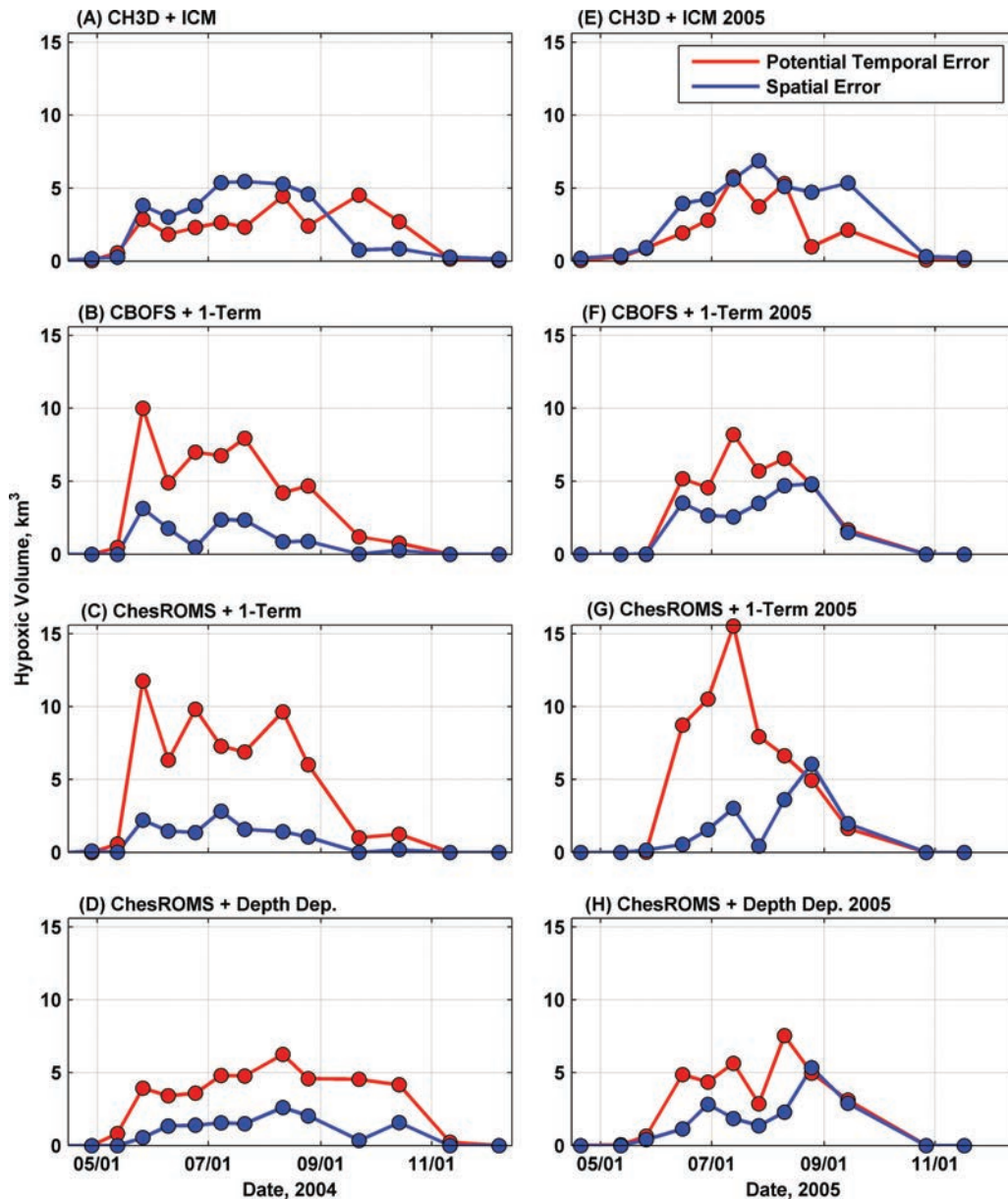
Figure 7. Potential uncertainty in the calculation of hypoxic volume from discrete stations. Red circles mark the range of the interpolated Spatial Match HV over each sampling period (i.e., the length of red bars on Figure 6; potential temporal error). Blue circles represent the difference between the Integrated 3-D HV and the interpolated Spatial Match HV (i.e., the distance between the blue line and red circles in Figure 6; spatial error).

and therefore limited potential errors due to nonsynopticity. Second, adding tens of additional synoptic stations, even when placed at locations where the variance in hypoxia is high, did not notably improve the ability of the various Interpolated Model HV cases to match Integrated 3-D HV (Figure 8). Third, because the scaling was developed using actual CBP station locations, rather than using hypothetical station locations as in Cases 4–8, it could easily be applied as a correction factor for the CBP hypoxic volumes computed directly from station observations, including cruises over past decades.

[49] The correction factor (see equation (1)) required to scale each Interpolated Model HV to match Integrated 3-D HV was a function of Interpolated Model HV for each of the four models in 2004 and 2005 (Figure 9). When Interpolated Model HV was less than about 10 km^3, Integrated 3-D HV was strongly underpredicted (i.e., a positive CF was needed), whereas when Interpolated Model HV was greater than about 20 km^3, Integrated 3-D HV was slightly overpredicted (favoring a slightly negative CF). This general trend held for all models for both 2004 and 2005 (Figure 9), although the Interpolated Model HV values for CH3D + ICM were all approximately ≤10 km^3 and thus always underpredicted the "true" Integrated 3-D HV. When all the Interpolated Model HV cases were combined for both years for all four models, the least-squares best fit
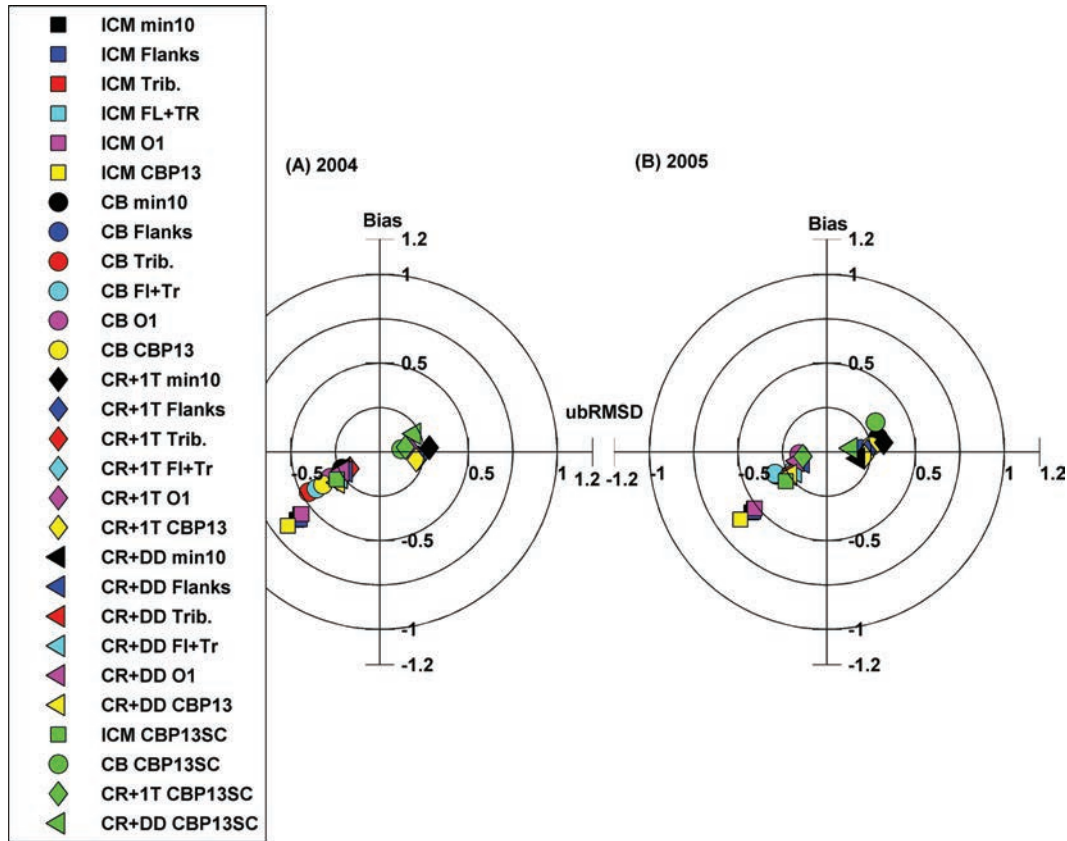
Figure 8. Normalized target diagrams showing how well different combinations of stations reproduced the Integrated 3-D HV. Station sets corresponded to: MS10, 10 stations in the main stem of the Bay; Flanks, the 10 Main Stem stations plus the stations along the flanks of the main stem, Trib., the 10 Main Stem stations plus those in the tributaries; Fl + Tr, the 10 Main Stem stations plus the flanks and tributary stations; AO, the Assumed Optimal station locations for capturing hypoxia; CBP13, the set of 13 CBP stations; CBP13SC, the set of 13 CBP stations scaled to better match the total 3-D hypoxic volume.

for the coefficients a and b for the global correction factor, $y = a(\log 10(HVinterp)) + b$ (equation (2)), were $a = -1.22$ and $b = 1.52$, where Interpolated Model HV is in km^3, and the best fit correction factor (black line in Figure 10) was found to cross zero at 18 km^3. In other words, the global correction factor was slightly negative for Interpolated Model HV >18 km^3 and became positive for Interpolated Model HV <18 km^3.

[50] After applying the correction factor, the scaled hypoxic volumes more successfully reproduced the Integrated 3-D HV. This is shown by the target diagram (Figure 8), where the scaled CBP13 case (in green) for each model better matches Integrated 3-D HV than does the corresponding unscaled CBP13 HV (in yellow). These scaling coefficients worked nearly equally well for the CBP13 set of stations, the main stem set of 10 stations, and a set of 11 CBP stations in the main stem of the Bay; thus, these best fit scaling coefficients were not overly sensitive to the specific CBP13 station set.

3.5. Results Using Other Metrics for Quantifying Hypoxia Severity

[51] Additional metrics that can be used to quantify the severity of hypoxia include hypoxic area, hypoxic thick-

ness, the duration of hypoxia, peak hypoxic volume, and cumulative hypoxic volume. The duration of hypoxia, peak hypoxic volume, and cumulative hypoxic volume metrics have the advantage of providing a single annual value with which to describe hypoxia. Analysis of 3-D model output indicated that the variation in hypoxic volumes among the ROMS implementations was largely the result of different spatial extents of hypoxia, because the average thicknesses from the three ROMS models were similar (Figure 11). The CH3D + ICM model estimated a generally thinner hypoxic layer than the ROMS-based implementations. Although useful as an indicator of extent of benthic environments subject to hypoxic stress, hypoxic area (and resulting hypoxic thickness) are not direct outputs of the CBP interpolator and thus were not a primary focus of this present study.

[52] The duration of hypoxia (times with hypoxia >2 km^3) varied based on the HV estimates used in the calculation, with a longer duration from the Integrated 3-D DO fields than from DO fields calculated from the discrete Interpolated Models (Figure 12). On average, the Integrated 3-D DO fields from the four hydrodynamic + DO models showed 44 and 40 more days of hypoxia than the Spatial Match for 2004 and 2005, respectively. These differences
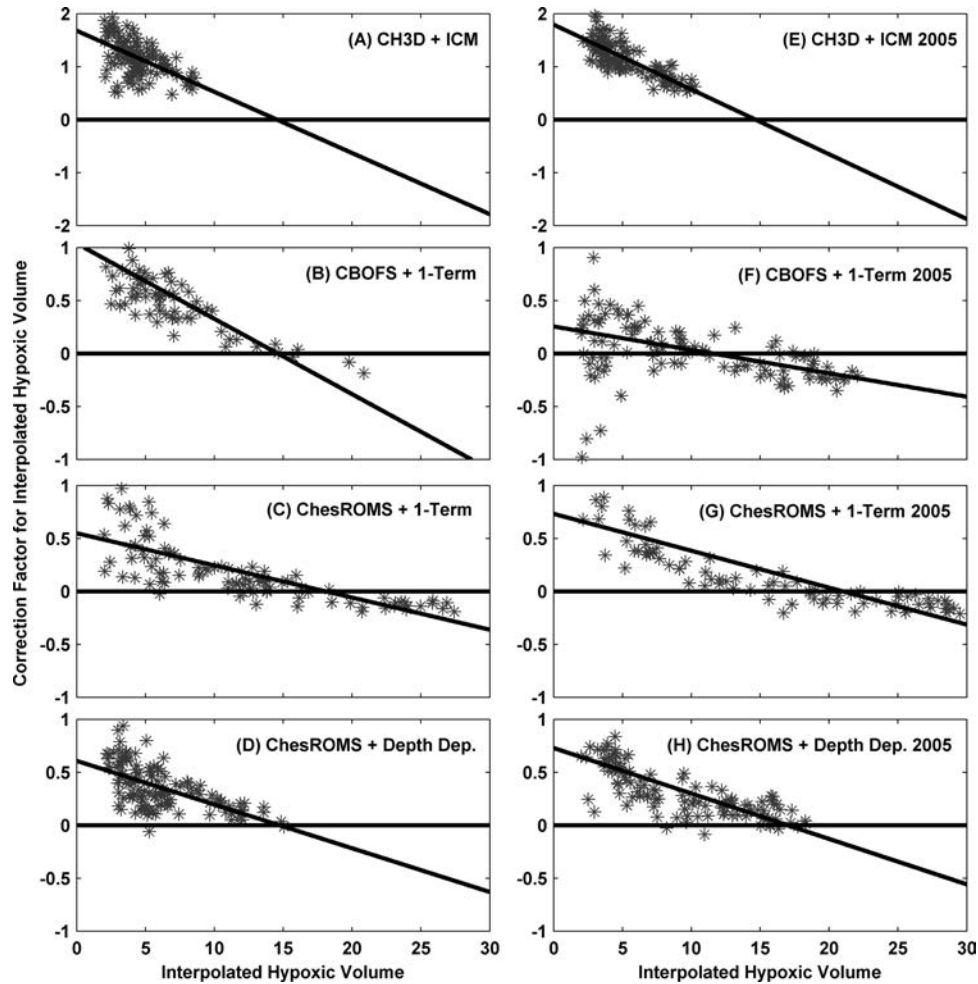
Figure 9. Stars indicate the individual correction factors required for each Interpolated Model HV to match the Integrated 3-D HV for the CBP13 station set. Each of the four models is shown separately for 2004 and 2005. The linear trend is shown in each case for reference.
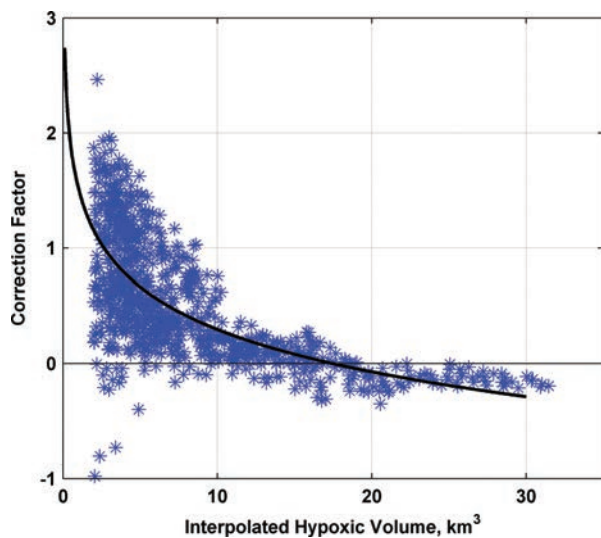

Figure 10. Stars indicate the individual correction factors required for each Interpolated Model HV to match the Integrated 3-D HV for the CBP13 station set. The global best fit log function (see equation (2)) is shown by the continuous black curve.

in hypoxic duration arose from both the Interpolated Model HV cases underestimating the 3-D hypoxic volume and the dates of cruises not corresponding with the actual beginning or end of hypoxia. The CBP13 Interpolated HV was calculated using the daily model output, yet still underestimated the duration of hypoxia. Upscaling of Interpolated HV helped correct the former shortcoming; thus, hypoxic duration based on CBP13SC was consistently closest to the "true" Integrated 3-D duration. However, if calculated at the same dates and times as the CBP Data, upscaling could not inherently correct for potentially unfortunate cruise timing near the actual beginning or end of hypoxia.

[53] The cumulative hypoxic volumes have a single annual value, a benefit over the hypoxic volume time series. Unlike the annual peak hypoxic volume, for example, these cumulative values should also have reduced uncertainties over a single hypoxic volume computed from a given cruise. The Absolute Match and Spatial Match cumulative hypoxic volumes were similar, on average, disagreeing by only 10% or less at the end of the year (Figure 13). Nonetheless, the cumulative hypoxic volume from all the Interpolated Model HV sets underestimated the real cumulative hypoxic volume from the 3-D fields. After scaling, however, most of the new cumulative hypoxic volumes nearly
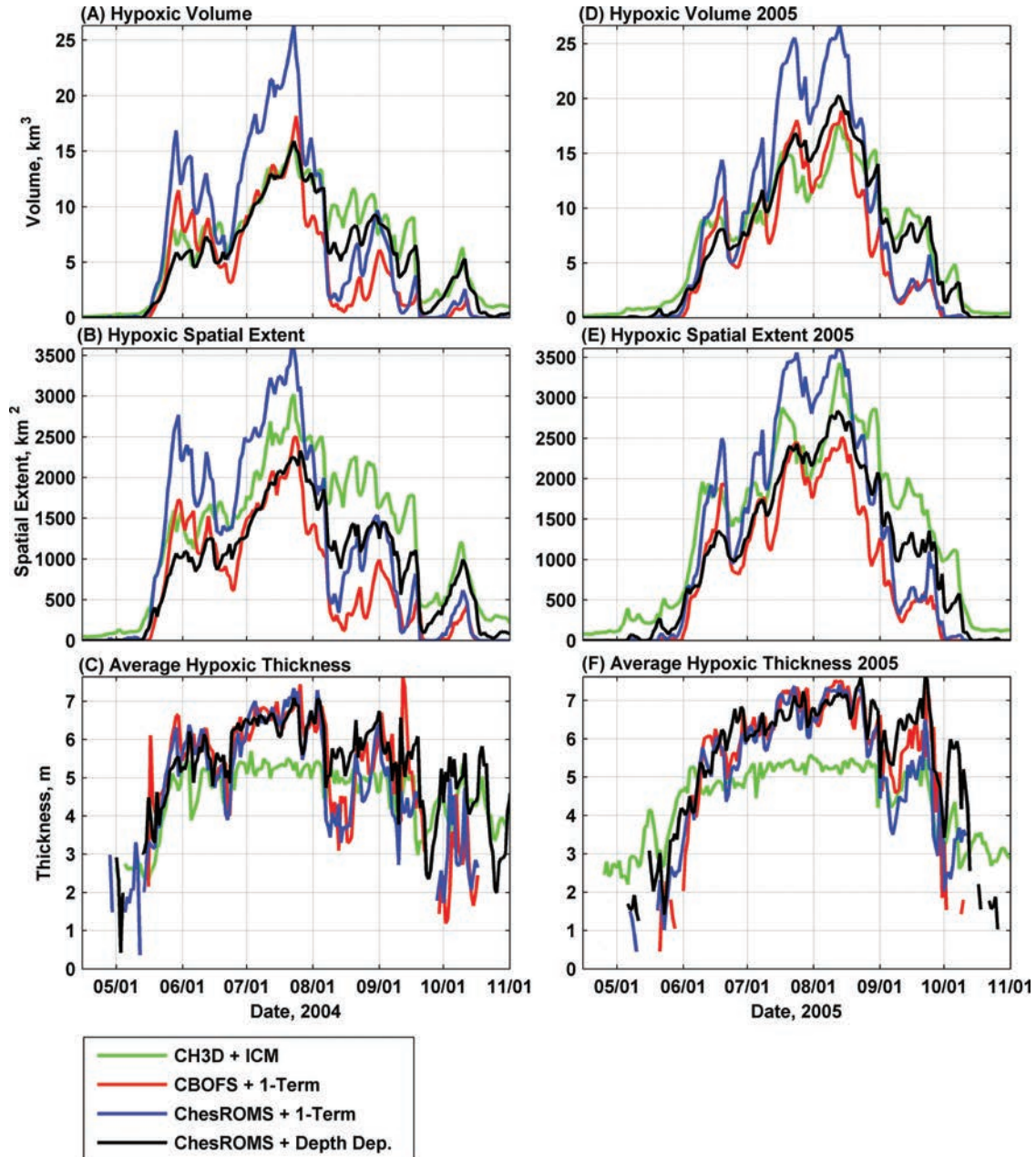
Figure 11. (A and D) Hypoxic volume from the 3-D fields, (B and E) hypoxic area, and (C and F) the average thickness of hypoxia from each model for 2004 (Figures 11A–11C) and 2005 (Figures 11D–11F).

matched the 3-D values, with an average mismatch of only 11.5% of the 3-D cumulative hypoxic volumes, compared to 28% before scaling. Much of this mismatch was caused by the CH3D + ICM model, because the stations did not represent the 3-D hypoxic volume as well in this model.

3.6. Application of Upscaling to Metrics for Chesapeake Bay Hypoxia From 1984 to 2012

[54] Finally, the upscaling method derived above (section 3.4) was applied to several decades of DO profile station data collected by the CBP. Potentially useful long-term hypoxia metrics that were likely improved by this upscaling included time series of semimonthly/monthly hypoxic volume (which visually highlights the maximum annual HV), annual cumulative hypoxic volume (which can smooth away extreme events of short duration), the annual duration of hypoxia, and the average summer hypoxic volume (Figure 14). It is interesting to note from Figure 14 that those years with the very highest or lowest maximum HV do not correspond on a one-to-one basis with those with the very highest or lowest cumulative HV.

[55] Supporting text files are provided with the online version of this manuscript that contain time series of many different interpolated HVs and cumulative HV from the CBP Data.

4. Discussion

[56] By combining results from multiple models and determining the results that are consistent between all the
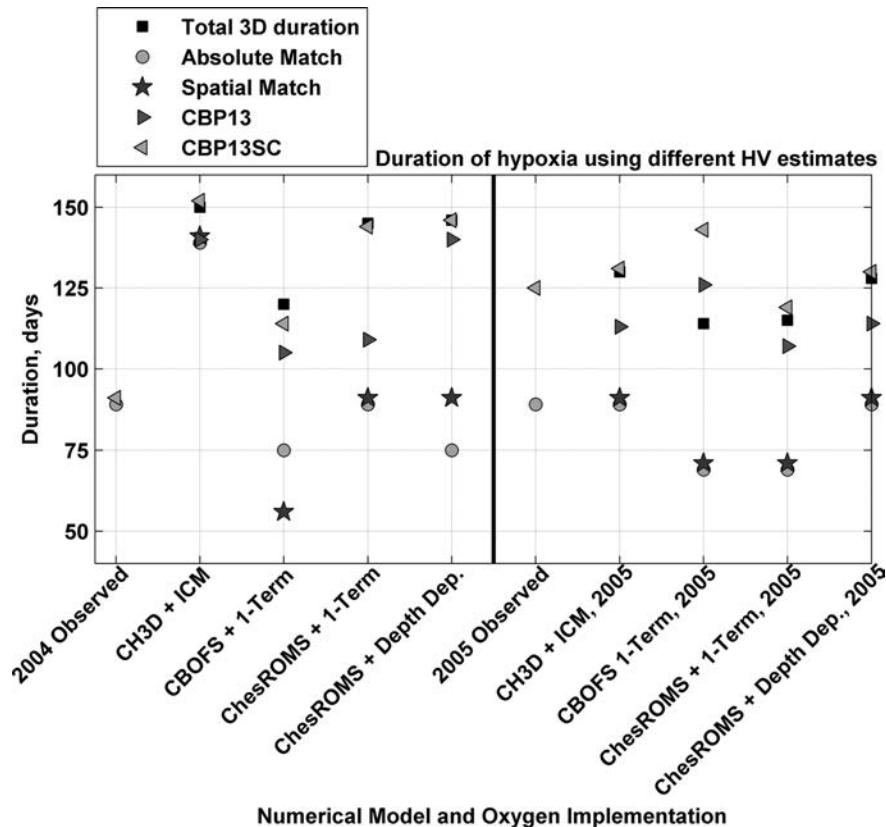
Figure 12. The duration of hypoxic volume greater than 2 km^3 for the different models and sampling approaches.

models, information was gained to better understand the uncertainty in the available observations, methods for improving the interpolated observations were developed, and additional metrics for evaluating hypoxia were suggested. The analysis presented in this paper benefits from the simultaneous use of two different core hydrodynamic models and three different dissolved oxygen formulations, each sampled for hypoxic volume in 10 different ways. The approach applied here highlights the advantages of multiple models for both interpreting observations and simulating environmental scenarios. The use of multiple models is a valuable tool, which should be used going forward to investigate how systems function and to improve observational sampling strategies.

4.1. Limitations in Observation-Based Estimates of Hypoxic Volume, Alternative Sampling Strategies, and Improvement Through Upscaling

[57] Utilization of multiple model simulations suggests that the inverse distance squared interpolation method underestimated the summer hypoxic volume by up to 50%, due to both spatial and temporal errors associated with sampling constraints. Because of the strong time variation inherent to hypoxia in Chesapeake Bay, the temporal errors associated with a 2 week sampling cruise could be as large as ~10 km^3 and are likely to exceed the uncertainties associated with finite spatial sampling. It is important to recognize that the above estimates of potential temporal error in observed HV are dependent on the temporal variability produced by individual hydrodynamic + DO models used. As

shown by their positions on the right-hand side of the target diagrams in Figure 5, the three ROMS models overestimated the actual variability in HV observed in the monthly to bimonthly CBP Data HV, whereas the CH3D + ICM model underestimated the variability of the observed CBP Data HV. Thus, the temporal errors the ROMS models predict to be present in the interpolated HV observations may be higher than the true temporal error, while the temporal errors predicted by the CH3D + ICM model are likely to be lower than the true temporal error.

[58] By choosing to recalculate the hypoxic volume from the CBP Data using only profiles in the set of 13 CBP stations close to the main stem (see Figure 1) and then upscaling the hypoxic volume, the limitations inherent in the present CBP interpolation method are at least partially corrected. During Chesapeake Bay Program cruises, the 13 CBP stations are sampled relatively quickly (averaging ~2.3 days), so their use, as opposed to using station data collected over a 10 day cruise, helps reduce temporal error. The correction factor produced by the best fit logarithmic scaling function was based on the difference in interpolated model output from these 13 stations and the "true" hypoxic volume calculated from spatially integrating over every model grid point. In this way the tendency of the interpolation to underestimate HV due to spatial limitations was also at least partially corrected.

[59] The model results presented here suggest sampling fewer stations more often can give a better estimate of the extent of deep water hypoxia. This is likely because the strong variability of hypoxia in the Chesapeake Bay is not
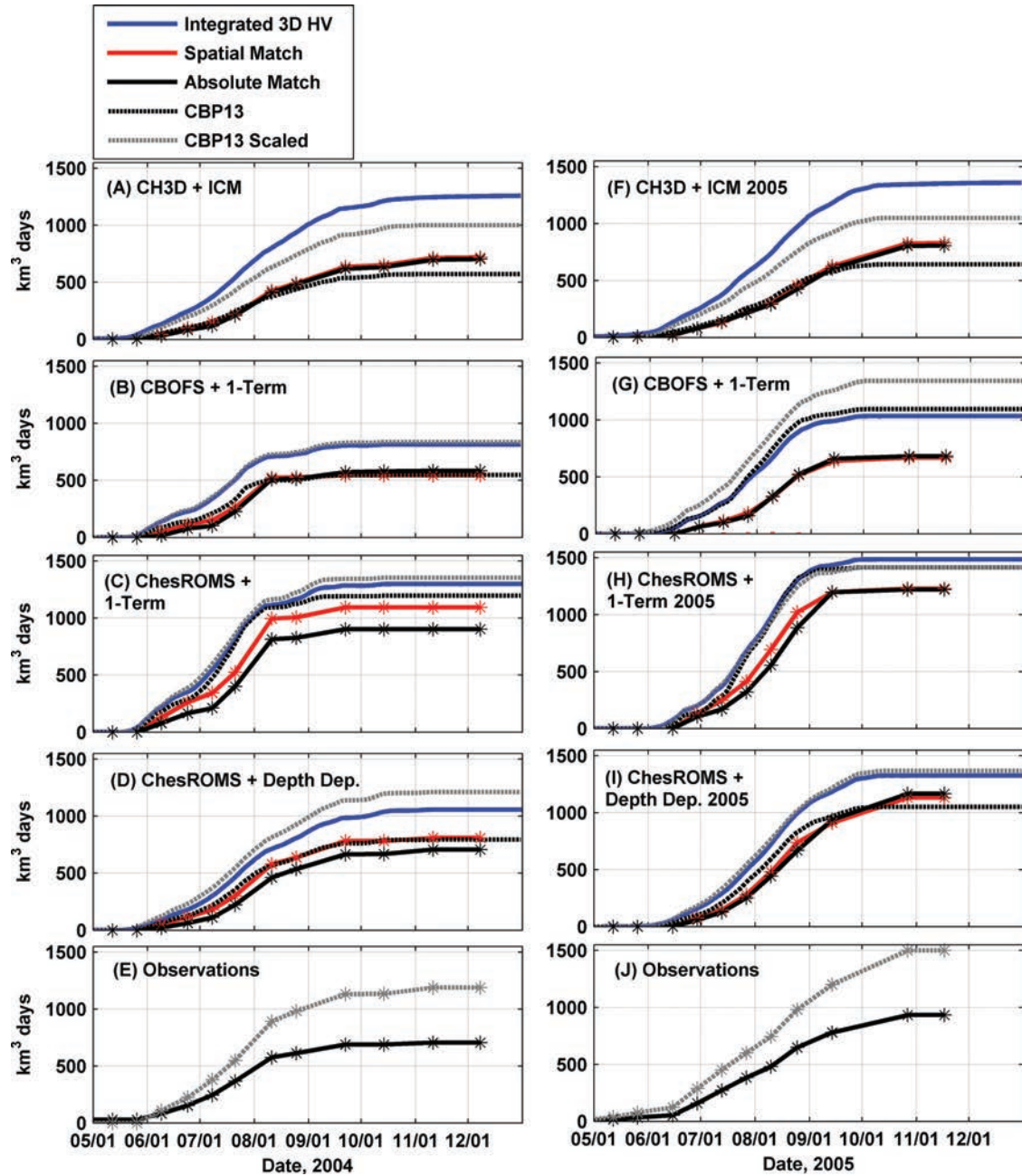
Figure 13. The cumulative hypoxic volume days (km³ days) calculated from the Integrated 3-D fields (blue line) and the Absolute Match (black), Spatial Match (red), CBP13 (black dotted), and the scaled CBP13 (gray dotted) interpolated hypoxic volumes, computed for each model for (a–d) 2004 and (f–i) 2005. Stars on the lines mark the dates of cruises where the HV is calculated. Nonstarred lines represent more continuous daily HV calculations. Also shown are the cumulative hypoxic volumes computed from the observations for the CBP13 and the scaled CBP13 stations for (e) 2004 and (j) 2005.

simply a function of longitudinal location or depth within the estuary. The work presented here and by *Sanford et al.* [1990] showed the DO within the Chesapeake Bay was highly variable in time, as has also been shown in other systems [*DiMarco et al.*, 2010; *Lee and Lwiza*, 2008]. For example, the models showed high temporal variability in response to periodic advection of low oxygen water out of the main channel due to wind forcing [*Sanford et al.*, 1990; *Scully*, 2010a]. This spatially coherent advection of DO

results in a large amount of variability in the 3-D hypoxic volumes between, and during, each CBP profiling cruise. Sampling fewer stations more often and more synoptically is likely a more efficient strategy for capturing this coherent temporal variability as opposed to sampling more stations less often and less synoptically.

[60] Work in the Gulf of Mexico also shows how models can be used to improve DO sampling strategies. *DiMarco et al.* [2010] and *Hetland and DiMarco* [2008] showed that
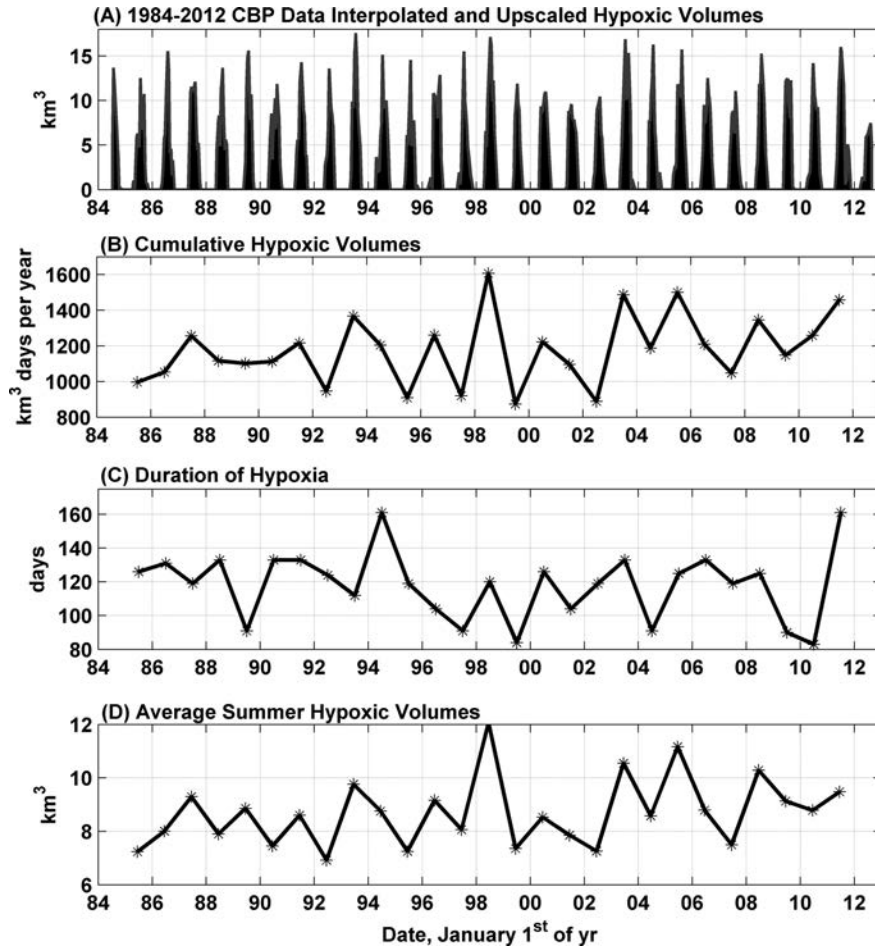
Figure 14. (a) Hypoxic volumes calculated from 13 CBP stations, using the CBP hypoxic volume interpolator program, and then upscaled by the developed scaling method. (b) The cumulative hypoxic volume days, (c) the duration of hypoxic volume greater than 2 km^3, and (d) the average summer hypoxic volume based on these scaled hypoxic volumes. The average summer hypoxic volume is the average of all HVs between 1 June and 30 September of a given year.

a spatially and temporally coarse sampling strategy on the Louisiana-Texas shelf does not optimally capture hotspots of low DO and thus does not represent the overall HV well. Their model results demonstrate that DO monitoring observations in the Gulf of Mexico need to be closer in space and time to better capture the true HV.

4.2. Importance of Using Multiple Metrics in Evaluating Long-Term Variability in Chesapeake Bay Hypoxia

[61] The large temporal variability in both the location of low bottom DO and the total hypoxic volume make metrics determining the amount of hypoxia based on only one or two stations or a few hypoxic volumes unlikely to fully characterize the extent of hypoxia within a single year, let alone multiple years. In locations where hypoxia varies in space and time, metrics need to be crafted such that they are part of a system that captures the overall severity of hypoxia from year to year. The system needs to take into account the spatial and temporal complexity of the natural environment and the inherent errors associated with interpolating observations. Ideal metrics should also capture changes to hypoxia caused by climate change. For exam-

ple, metrics based on a specific date range may inadequately capture hypoxia if there is a climatic induced shift in the timing or duration of hypoxia. It should be very clear when generating and reporting a system of metrics from the observations and/or model results as to what exactly these numbers represent. Currently, there are a number of different methods being used to calculate hypoxic volumes and associated metrics (based on *Hagy et al.* [2004], *Murphy et al.* [2011], *Scavia et al.* [2006], *Testa et al.* [2008], and P. Tango (personal communication, 2013)).

[62] We put forth using the yearly maximum hypoxic volume, the yearly cumulative hypoxic volume, and the yearly duration of hypoxia as ways to characterize the overall extent and likely impact of hypoxia within the Chesapeake Bay. The accuracy of all three of these metrics has likely been improved by the upscaling method presented in this paper. These three metrics have the benefit of each being a single value for an entire season and all being relevant to ecological health. The diversity of using these three metrics simultaneously can also be used to better reflect the diversity of ecological sensitivity to hypoxic intensity, extent, and duration. It is important to recognize that years with the most intense instantaneous hypoxia (i.e., largest

Table 4. Yearly Hypoxia Metrics From the CBP Profiles[a]

Year	Maximum HV (km^3)		Cumulative HV (km^3 days)		Duration of HV >2 km^3 (days)	
	All Profiled Stations	CBP13 Scaled	All Profiled Stations	CBP13 Scaled	All Profiled Stations	CBP13 Scaled
1985	8.2	11.7	536	883	98	126
1986	11.7	15.1	641	956	91	131
1987	9.4	11.2	801	1137	119	119
1988	8.5	13.0	630	1000	133	133
1989	11.0	15.2	701	1025	77	91
1990	8.7	11.0	636	990	119	133
1991	9.5	13.6	729	1100	92	133
1992	9.8	12.9	566	850	78	105
1993	16.4	17.6	907	1276	91	112
1994	9.4	14.6	686	1080	91	161
1995	10.7	14.0	523	807	105	119
1996	8.7	12.1	737	1141	90	104
1997	10.5	15.1	491	822	42	56
1998	12.2	17.1	984	1483	120	120
1999	7.1	11.0	528	779	70	70
2000	7.5	10.1	739	1092	105	126
2001	6.5	8.7	642	971	104	104
2002	7.3	9.5	465	784	70	70
2003	15.2	16.7	1036	1387	119	133
2004	10.0	16.0	703	1091	91	91
2005	11.2	15.3	932	1367	91	125
2006	8.3	11.7	670	1084	84	119
2007	7.1	10.2	588	928	91	119
2008	17.3	14.8	938	1234	77	112
2009	8.8	11.6	699	1050	90	90
2010	9.9	13.5	775	1138	83	83
2011	12.0	15.7	976	1355	117	161

[a]The years of 1984 and 2012 did not have complete data during the preparation of this manuscript and have been left off of this table.

peak volume) do not necessarily correspond to the years with greatest hypoxic duration or greatest cumulative hypoxic volume (Figure 14 and Table 4). Although there is general agreement between the yearly ups and downs in these three metrics, they do not directly follow each other, making a characterization based on a single one insufficient to determine even the relative magnitudes of the others.

[63] The dynamics of hypoxia in diverse regions, including the Chesapeake Bay, Gulf of Mexico, Hood Canal, and Long Island Sound, has been shown to be linked to physical and biological forcing, such as riverine nutrient input varying between wet and dry years [*Bianchi et al.*, 2010; *Hagy et al.*, 2004; *Paulson et al.*, 2006], mixing and lateral advection due to winds [*Sanford et al.*, 1990; *Scully*, 2010b], stratification or overturning of the water column [*Bianchi et al.*, 2010; *Lee and Lwiza*, 2008; *Warner et al.*, 2001], and reduced DO solubility with increasing water temperature [*Najjar et al.*, 2010; *Scully*, 2013], all of which may be influenced by climate change. Using the yearly maximum hypoxic volume, the yearly cumulative hypoxic volume, and the yearly duration of hypoxia to characterize the severity of low DO has the additional benefit of incorporating changes in hypoxia caused by climate change. The duration of hypoxia can account for an increase in the amount of time such systems are hypoxic without being biased by temporal shifts in the timing of hypoxia. The cumulative hypoxic volume metric integrates over the entire year and thus will directly account for possible increases

(or decreases) in the duration of hypoxia, while also being insensitive to shifts in the timing of hypoxia.

[64] Of these three metrics, the annual cumulative hypoxic volume is likely the one that can be estimated with the most confidence. As a single metric for characterizing interannual variability in the severity of hypoxia, cumulative hypoxic volume is more attractive than the maximum hypoxic volume, duration of hypoxia, or average summer hypoxic volume. The annual maximum hypoxic volume, even when upscaled from the CBP 13 HV observations, still suffers from temporal errors associated with variations in hypoxic volume that can occur on time scales less than 2.3 days. The annual hypoxic duration metric also still suffers from temporal errors after upscaling, given that the timing of the monthly CBP cruises in spring or fall may or may not correspond well with the first or last appearance of significant hypoxia. The annual cumulative hypoxic volume, on the other hand, is relatively insensitive to temporal observational errors, both in terms of nonsynopticity in sampling and in terms of exactly capturing the first or last appearance of modest levels of hypoxia ($\sim$2 km^3). Another advantage of the cumulative hypoxic volume approach is its inherent ability to incorporate the effects of climate change. *Najjar et al.* [2010] and *Scully* [2013] show climate change induced warming of the Bay could have large negative implications on DO. If the hypothesized continued warming of the Chesapeake Bay causes intense hypoxia to occur substantially earlier or end substantially later, then

that will be immediately seen in the cumulative hypoxic volume. The common CBP approach of averaging hypoxic volume over several specific summer cruises (but not integrating beyond this temporal window; see supporting information) will not account for the full severity of hypoxic volume expansion associated with climate change.

5. Summary and Conclusions

[65] Hypoxia in Chesapeake Bay is a significant ecological stressor that was observed as early as 1930 [*Newcombe and Horne*, 1938] and is currently a focus of regulations aimed at environmentally restoring the Chesapeake Bay, estimated to cost in the tens of billions of dollars [*Maryland*, 2012; *Virginia*, 2010]. Management plans are partly based on reducing the size of this Chesapeake Bay "dead zone," but little information is available on how reliably point measurements of DO can be scaled up to estimate the total volume of hypoxic water in the Bay. There is no doubt that the Chesapeake Bay Program's biweekly to monthly observations of water quality at dozens of stations in the Bay and its tributaries represent an invaluable long-term data set for characterizing DO. However, in complex systems limitations associated with the logistics of cruise-based sampling introduce both spatial and temporal errors when these observations are interpolated to represent the true hypoxic volume. Output from highly resolved, 3-D numerical models provide a useful tool for better understanding the errors introduced by interpolation of these discrete observed profiles.

[66] Application of the CBP interpolation program to profiles of DO "sampled" from 3-D model output at the times and locations matching CBP observation stations indicates that this observational approach of sampling and interpolating DO tends to underestimate true hypoxic volume during summer by several km^3. The component of this error associated with limitations in spatial sampling for summer 2004 and 2005 was found to be 2.4 km^3 on average. However, the potential temporal error in hypoxic volume associated with interpolation of station data collected over 1–2 weeks was even larger, approaching 5 km^3 on average. Thus, this study suggests that quickly sampling a select number of stations along and adjacent to the main stem of the Bay may optimize the sampling strategy for characterizing time-varying hypoxic volume within the Chesapeake Bay. Similarly quick sampling strategies at select stations could work to optimally estimate HV in other relatively deep and narrow estuarine hypoxic zones. Furthermore, adding stationary vertical profiler time series to the sampling strategy [i.e., *Newton et al.*, 2011] would greatly improve the understanding and monitoring of hypoxia in the Chesapeake Bay by allowing for a characterization of the time variation in the DO from field data, verifying the models' results on the time variation in the DO field and estimating the severity of hypoxia in real time.

[67] Based on model output from 2004 to 2005, a scaling factor was derived that can be used to upscale interpolations of observed hypoxic volumes to likely better represent the full 3-D hypoxic volumes within the Chesapeake Bay. Although future work may further refine this scaling method, the scaling factor derived here likely improved observation-based estimates of instantaneous hypoxic volume, maximum hypoxic volume, cumulative hypoxic volume, and the duration of hypoxia for CBP cruises extending from 1984 to 2012. Among these metrics for bay-wide hypoxia, cumulative hypoxic volume is likely the least sensitive to temporal errors brought about by nonsynoptic sampling. In addition, cumulative hypoxic volume is arguably more useful than average summer hypoxia, another metric often used for management purposes to quantify the health of the Bay, as it includes hypoxic conditions from key spring and fall months which is when increases in hypoxia due to climate change may be most significant.

[68] Numerical models have evolved to the point that they can be very beneficial for refining and giving insight into observational sampling schemes of societal relevance that drive environmental policy. By using multiple numerical models, we have improved the understanding of a long-term data set and generated better estimates of metrics that estimate and track the health of the Chesapeake Bay.

References

Bianchi, T. S., S. F. diMarco, J. H. C. Jr., R. D. Hetland, P. Chapman, J. W. Day, and M. A. Allison (2010), The science of hypoxia in the northern Gulf of Mexico: A review, *Sci. Total Environ.*, *408*, 1471–1484.

Breitburg, D. L. (1990), Near-shore hypoxia in the Chesapeake Bay: Patterns and relationships among physical factors, *Estuarine Coastal Shelf Sci.*, *30*, 593–609.

Cerco, C. F. (1995), Response of Chesapeake Bay to nutrient load reductions, *J. Environ. Eng.*, *121*(8), 549–557.

Cerco, C. F. (2000), Phytoplankton kinetics in the Chesapeake Bay eutrophication model, *Water Qual. Ecosyst. Model.*, *1*, 5–49.

Cerco, C. F., and T. Cole (1993), Three-dimensional eutrophication model of Chesapeake Bay, *J. Environ. Eng.*, *119*(6), 1006–1025.

Cerco, C. F., and M. R. Noel (2004), The 2002 Chesapeake Bay eutrophication model, Rep. EPA 903-R-04-004, 349 pp., Chesapeake Bay Program Off., U.S. Environ. Prot. Agency, Annapolis, Md.

Cerco, C. F., and M. R. Noel (2005), Incremental improvements in Chesapeake Bay Environmental Model Package, *J. Environ. Eng.*, *131*(5), 745–754.

Cerco, C. F., S. C. Kim, and M. R. Noel (2010), The 2010 Chesapeake Bay Eutrophication Model: A report to the US Environmental Protection Agency Chesapeake Bay Program, report, 228 pp., U.S. Army Eng. Res. and Dev. Cent., Vicksburg, Miss.

Cloern, J. E. (2001), Our evolving conceptual model of the coastal eutrophication problem, *Mar. Ecol. Prog. Ser.*, *210*, 223–253.

Commonwealth of Virginia (2010), Chesapeake Bay TMDL Phase I Watershed Implementation Plan: Revision of the Chesapeake Bay Nutrient and Sediment Reduction Tributary Strategy, 133 pp., Richmond, VA.

Conley, D. J., et al. (2009), Hypoxia-related processes in the Baltic Sea, *Environ. Sci. Technol.*, *43*(10), 3412–3420.

Diaz, R. J., and R. Rosenberg (2008), Spreading dead zones and consequences for marine ecosystems, *Science*, *321*, 926–929.

DiMarco, S. F., P. Chapman, N. Walker, and R. D. Hetland (2010), Does local topography control hypoxia on the eastern Texas-Louisiana shelf?, *J. Mar. Syst.*, *80*, 25–35.

Donigian, A. S., B. R. Bicknell, A. S. Patwardhan, L. C. Linker, C. Chang, and R. Reynolds (1994), Chesapeake Bay program watershed model application to calculate Bay nutrient loadings: Final findings and recommendations, Rep. EPA 903-R-94-042, 283 pp., U.S. EPA Chesapeake Bay Program, Annapolis, Md.

Gilbert, D., N. N. Rabalais, R. J. Diaz, and J. Zhang (2010), Evidence for greater oxygen decline rates in the coastal ocean than in the open ocean, *Biogeosciences*, *7*, 2283–2296.

Goodrich, D. M., W. C. Boicourt, P. Hamilton, and D. W. Pritchard (1987), Wind-induced destratification in Chesapeake Bay, *J. Phys. Oceanogr.*, *17*, 2232–2240.

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

Haidvogel, D. B., et al. (2008), Regional ocean forecasting in terrain-following coordinates: Model formulation and skill assessment, *J. Comput. Phys.*, *227*, 3595–3624.

Hawley, N., T. H. Johengen, Y. R. Rao, S. A. Ruberg, D. Beletsky, S. A. Ludsin, B. J. Eadie, D. J. Schwab, T. E. Croley, and S. B. Brandt (2006), Lake Erie hypoxia prompts Canada-U.S. study, *Eos Trans. AGU*, *32*(8), 313.

Helly, J. J., and L. A. Levin (2004), Global distribution of naturally occurring marine hypoxia on continental margins, *Deep Sea Res., Part I*, *51*, 1159–1168.

Hetland, R. D., and S. F. DiMarco (2008), How does the character of oxygen demand control the structure of hypoxia on the Texas-Louisiana continental shelf?, *J. Mar. Syst.*, *70*, 49–62.

Hofmann, E. E., et al. (2011), Modeling the dynamics of continental shelf carbon, *Annu. Rev. Mar. Sci.*, *3*, 93–122.

Johnson, B. H., R. E. Heath, B. B. Hsieh, K. W. Kim, and H. L. Butler (1991), Users's guide for a three-dimensional numerical hydrodynamic, salinity, and temperature model of Chesapeake Bay, Rep. HL-91-20, 81 pp., Dep. of the Army, Vicksburg, Miss.

Jolliff, J. K., J. C. Kindle, I. Shulman, B. Penta, M. A. M. Friedrichs, R. Helber, and R. A. Arnone (2009), Summary diagrams for coupled hydrodynamic-ecosystem model skill assessment, *J. Mar. Syst.*, *76*(2009), 64–82.

Lanerolle, L. W., R. C. Patchen, and F. A. Aikman (2009), The second generation Chesapeake Bay Operational Forecast System (CBOFS2): A ROMS-based modeling system, paper presented at the Eleventh International Conference on Estuarine and Coastal Modeling, Am. Soc. of Civ. Eng., Seattle, Wash.

Lanerolle, L. W., R. C. Patchen, and F. Aikman (2011), The second generation Chesapeake Bay Operational Forecast System (CDOFS2): Model development and skill assessment, report, 77 pp., U.S. Dept. of Commerce, National Oceanic and Atmospheric Administration, National Ocean Service, Office of Coast Survey, Coast Survey Development Laboratory, Silver Spring, Md.

Lee, Y. J., and K. M. M. Lwiza (2008), Characteristics of bottom dissolved oxygen in Long Island Sound, New York, *Estuarine Coastal Shelf Sci.*, *76*, 187–200.

Linker, L. C., G. W. Shenk, R. L. Dennis, and J. S. Sweeney (2000), Cross-media models of the Chesapeake Bay watershed and airshed, *Water Qual. Ecosyst. Model.*, *1*, 91–122.

Linker, L. C., G. Shenk, P. Wang, and R. Batiuk (2008), Integration of modelling, research and monitoring in the Chesapeake Bay Program, in *The Management of Water Quality and Irrigation Technologies*, edited by J. Albiac and A. Dinar, pp. 41–60, Earthscan, London.

Liu, Y., and D. Scavia (2010), Analysis of the Chesapeake Bay hypoxia regime shift: Insights from two mechanistic models, *Estuaries Coasts*, *33*, 629–639.

Maryland (2012), Maryland's Phase II Watershed Implementation Plan for the Chesapeake Bay TMDL, *NOAA Tech. Rep. NOS CS 29*, 84 pp.

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

Najjar, R. G., et al. (2010), Potential climate-change impacts on the Chesapeake Bay, *Estuarine Coastal Shelf Sci.*, *86*, 1–20.

Newcombe, C. L., and W. A. Horne (1938), Oxygen-poor waters of the Chesapeake Bay, *Science*, *88*(2273), 80–81.

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

Nordberg, K., H. L. Filipsson, M. Gustafsson, R. Harland, and P. Roos (2001), Climate, hydrographic variations and marine benthic hypoxia in Koljo Fjord, Sweden, *J. Sea Res.*, *46*, 187–200.

Officer, C. B., R. B. Biggs, J. L. Taft, L. E. Cronin, M. A. Tyler, and W. R. Boynton (1984), Chesapeake Bay anoxia: Origin, development, and significance, *Science*, *223*(4631), 22–27.

Parker-Stetter, S. L., and J. K. Horne (2009), Nekton distribution and mid-water hypoxia: A seasonal, diel prey refuge?, *Estuarine Coastal Shelf Sci.*, *81*, 13–18.

Paulson, A. J., C. P. Konrad, L. M. Frans, M. Noble, C. Kendall, E. G. Josberger, R. L. Huffman, and T. D. Olsen (2006), Freshwater and saline loads of dissolved inorganic nitrogen to Hood Canal and Kynch Cove, western Washington, U.S. Geol. Surv. Sci. Invest. Rep., 2006–5106, 92.

Pena, M. A., S. Katsev, T. Oguz, and D. Gilbert (2010), Modeling dissolved oxygen dynamics and hypoxia, *Biogeosciences*, *7*, 933–957.

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

Sanford, L. P., K. G. Sellner, and D. L. Breitburg (1990), Covariability of dissolved oxygen with physical processes in the summertime Chesapeake Bay, *J. Mar. Res.*, *48*, 567–590.

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

Scully, M. E. (2010a), Wind modulation of dissolved oxygen in Chesapeake Bay, *Estuaries Coasts*, *33*(5), 1164–1175.

Scully, M. E. (2010b), The importance of climate variability to wind-driven modulation of hypoxia in Chesapeake Bay, *J. Phys. Oceanogr.*, *40*(6), 1435–1440.

Scully, M. E. (2013), Physical controls on hypoxia in Chesapeake Bay: A numerical modeling study, *J. Geophys. Res.*, *118*, doi:10.1002/jgrc.20138, in press.

Seliger, H. H., J. A. Boggs, and W. H. Biggley (1984), Catastrophic anoxia in the Chesapeake Bay in 1984, *Science*, *228*(4695), 70–73.

Shchepetkin, A. F., and J. C. McWilliams (2005), The Regional Oceanic Modeling System (ROMS): A split-explicit, free-surface, topography-following-coordinate oceanic model, *Ocean Model.*, *9*, 347–404.

Taft, J. L., W. R. Taylor, E. O. Hartwig, and L. R (1980), Seasonal oxygen depletion in Chesapeake Bay, *Estuaries*, *3*(4), 242–247.

Testa, J. M., W. M. Kemp, W. R. Boyton, and J. D. Hagy (2008), Long-term changes in water quality and productivity in the Patuxent River Estuary: 1985 to 2003, *Estuaries Coasts*, *31*, 1020–1037.

U.S. Environmental Protection Agency (USEPA) (2003), Ambient water quality criteria for dissolved oxygen, water clarity and chlorophyll a for the Chesapeake Bay and its tidal tributaries, *Rep. EPA-903-R-03-002*, 343 pp., Region III, Chesapeake Bay Program Off., Annapolis, Md.

Wang, H. V., and B. H. Johnson (2000), Validation and application of the second generation three dimensional hydrodynamic model of Chesapeake Bay, *Water Qual. Ecosyst. Model.*, *1*, 51–90.

Warner, M. J., J. A. Newton, and M. Kawse (2001), Recent studies of the overturning circulation in Hood Canal, in *Proceedings of the 2001 Puget Sound Research Conference*, Puget Sound Action Team, Olympia, WA, 9 pp.

Wilmott, C. J. (1981), On the validation of models, *Phys. Geogr.*, *2*, 184–194.

Xu, J., W. Long, J. D. Wiggert, L. W. J. Lanerolle, C. W. Brown, R. Murtugudde, and R. R. Hood (2012), Climate forcing and salinity variability in Chesapeake Bay, USA, *Estuaries Coasts*, *35*(1), 237–261.