JOURNAL OF GEOPHYSICAL RESEARCH: OCEANS, VOL. 118, 2919–2933, doi:10.1002/jgrc.20223, 2013

An assessment of the skill of real-time models of Mid-Atlantic Bight continental shelf circulation

John L. Wilkin[1] and Elias J. Hunter[1]

Received 8 January 2013; revised 26 April 2013; accepted 29 April 2013; published 13 June 2013.

[1] Prescribing open boundary conditions for regional coastal ocean models encounters the challenge of imposing information on sea level, velocity and tracers that characterize the unrepresented far field ocean. Deriving such information from a larger domain model without communicating information from the "nested" model back to the exterior model is "downscaling". We evaluate whether real-time models presently in operation for the Mid-Atlantic Bight (MAB) can deliver useful predictions of subtidal frequency currents and subsurface temperature and salinity for this downscaling purpose. The MAB is a broad continental shelf region where several models run in real time and there is a dense observational data set available for skill assessment. We examine seven real-time models that cover the MAB: three global models, and four regional models. A regional climatology is included as an eighth model. Skill metrics with respect to model bias, centered root mean square error and cross correlation are computed for temperature and salinity profile data from 16 autonomous underwater glider vehicle missions and four hydrographic voyages in 2010–2011. Two years of hourly HF-radar surface current observations that span the shelf are used to evaluate modeled mean surface currents and daily time scale variability in speed and direction. Skill metrics, with uncertainty estimates, are reported for inner and outer shelf subregions, and for stratified and unstratified seasons. A group of models is identified that offers useful skill for the purposes of providing open boundary data to inner shelf and estuary models for real-time applications.

Citation: Wilkin, J. L., and E. J. Hunter (2013), An assessment of the skill of real-time models of Mid-Atlantic Bight continental shelf circulation, *J. Geophys. Res. Oceans*, *118*, 2919–2933, doi:10.1002/jgrc.20223.

1. Introduction

[2] Hydrodynamic models are widely used in coastal oceanography to simulate the circulation of limited-area domains for the analysis of regional dynamics and studies of biogeochemistry, geomorphology, and ecosystem processes; for example, to deduce transport pathways for nutrients, sediments, pollutants, or larvae. When operated as real-time now-cast or forecast systems, coastal ocean and estuarine models offer predictions of the ocean state that assist decision making related to water quality and public health, coastal flooding, shipping, maritime safety, and numerous other applications.

[3] Simulations of limited geographic regions must be constrained at their perimeter by open boundary conditions. Depending upon the application, these lateral boundaries might logically be placed near the mouth of an estuary, somewhere midshelf, or in the adjacent deep sea. Having open ocean boundaries introduces two related challenges: (i) formulating computational algorithms that allow motions generated within the domain to escape while simultaneously imposing remote influences exterior to the domain, and (ii) identifying sources of information on sea level, velocity and tracers that accurately characterize the far field ocean so as to incorporate their influence into the computational boundary conditions. We do not consider the former issue here, save to mention that there is a rich literature that addresses aspects of the accuracy and stability of open ocean boundary conditions [*Blayo and Debreu*, 2005; *Orlanski*, 1976] and describes algorithmic approaches that perform adequately in many practical situations [*Herzfeld*, 2009; *Marchesiello et al*., 2001; *Mason et al*., 2010; *Nycander and Döös*, 2003; *Oddo and Pinardi*, 2008].

[4] Our focus here is on the second challenge: to appraise the accuracy of ocean state estimates available as boundary condition data for real-time regional coastal ocean models.

[5] It is reasonable to expect that comprehensive descriptions of regional temperature, salinity (T/S) and velocity might be obtained from larger domain hydrodynamic models that assimilate observations and are driven

[1]Institute of Marine and Coastal Sciences, Rutgers University, New Brunswick, New Jersey, USA.

Corresponding author: J. L. Wilkin, Institute of Marine and Coastal Sciences, Rutgers University, 71 Dudley Rd., New Brunswick, NJ 08901-8521, USA. (jwilkin@rutgers.edu)

2169-9275/13/10.1002/jgrc.20223

by skillful meteorological analyses or forecasts. By "comprehensive" we mean providing all necessary dynamic quantities, in three dimensions, on a regular schedule, in a timely manner for real-time applications [*De Mey and Proctor*, 2009]. Providing output from one model as the open boundary condition data to a second (usually higher resolution) "nested" model, without communicating any information from the nested model back to the exterior model, is essentially the "downscaling" problem.

[6] We evaluate the quality of ocean state information that is available in practice by focusing on shelf waters of the Middle Atlantic Bight (MAB) – a region with a diversity of real-time and operational models and a dense *in situ* observational data set for skill assessment. A further motivation for our focus on the MAB is that the U.S. Integrated Ocean Observing System (IOOS) has sponsored a Coastal and Ocean Modeling Test-bed (COMT) project which has as one of its themes the examination of models of the Chesapeake Bay with respect to simulating currents, temperature and salinity—these conditions being assumed physical oceanographic drivers relevant to the occurrence of hypoxia events that increasingly plague the Bay ecosystem. Most existing Chesapeake Bay models assume climatological T/S conditions at the open boundary and impose only tidal variability in currents and sea level; there is no imposed influence of mean currents or sea level gradients, or subtidal frequency variability in these, from the shelf ocean. The work described here contributes to the COMT Estuarine Hypoxia project by considering whether the caliber of existing real-time models is sufficient to provide open boundary conditions that improve upon climatological T/S, and deliver useful predictions of subtidal frequency currents.

[7] The majority of the modeling systems we examine assimilate satellite-observed surface temperature and sea level anomaly (altimetry). Most global operational models also utilize in situ T/S observations from ships of opportunity, drifters, and profiling floats. In the MAB, further data are available from the Mid-Atlantic Regional Association Coastal Ocean Observing System (MARACOOS; maracoos.org), which operates an extensive CODAR (Coastal Ocean Dynamics Applications Radar) network observing surface currents from the coast to near the shelf edge, and deploys autonomous underwater glider vehicles (AUGV) that acquire surface to bottom T/S data from an onboard conductivity temperature depth (CTD) sensor along transects throughout the MAB. None of the modeling systems evaluated here assimilates MARACOOS AUGV data, and only one assimilates CODAR data. While this is regrettable in the respect that potentially useful data presently go unused in these real-time systems, it does afford us a set of independent data in shelf waters against which the models can be assessed.

[8] In the next section, we describe MARACOOS data acquired in 2010 and 2011 that form the basis of our analysis, and the data processing applied to make the skill assessment as objective as possible. Section 2 presents synopses of the modeling systems evaluated; for readers familiar with this subject these are three operational global models (HYCOM, NCOM, and MERCATOR) and four U.S. East Coast regional models (COAWST, UMassHOPS, ESPreSSO and NYHOPS). A regional T/S climatology (MOCHA) is included as an eighth predictive "model". These abbreviations are explained fully, and references provided, in Section 2.2.

[9] Section 2 also introduces the skill metrics we selected for the comparison and the basis of our estimates of their uncertainty. Results follow in section 3. The discussion in section 4 reiterates possible limitations in the scope of the skill assessment methodology, but highlights that the models most likely to deliver useful real-time boundary conditions for inner shelf and estuarine models of the MAB are NYHOPS, ESPreSSO and NCOM for T/S, and NYHOPS, ESPreSSO, COAWST and MERCATOR for velocity.

2. Methods: Observations, Models, and Skill Metrics

2.1. MARACOOS Observations

2.1.1. Subsurface Temperature and Salinity from CTD

[10] The vertical stratification of MAB shelf waters has a strong seasonal cycle [*Castelao et al*., 2008a; *Chapman and Gawarkiewicz*, 1993] with surface temperatures reaching 25°C in summer in a mixed layer some 15–20 m thick across much of the 100 km wide shelf, while below this cooler waters of 8–10°C persist almost year round. Spatially, bottom temperature takes a local minimum on the outer shelf in a band termed the "cold pool" [*Houghton et al*., 1982] that is maintained by southward advection of water formed by winter mixing in the vicinity of Nantucket Shoals. Further seaward, at the shelf edge, strong horizontal gradients in T/S occur at a shelf-slope front that separates cooler, fresher shelf water from the warmer, saltier Slope Sea [*Csanady and Hamilton*, 1998] that occupies the region between the shelf and the Gulf Stream.

[11] Freshwater runoff at the coast fosters a persistent across-shelf salinity gradient; while after periods of high discharge in spring and summer a pronounced vertically stratified halocline forms [*Castelao et al*., 2010]. Abetted by upwelling winds that drive coastal waters offshore, salinity contrasts exceeding 2.5 have been observed across the halocline more than 60 km offshore [*Castelao et al*., 2008b; *Gong*, 2011]. Adding to the richness of the subsurface hydrography are Slope Water salinity intrusions [*Flagg et al*., 1994; *Lentz*, 2003]. Most common in summer, these occur at the pycnocline some 30 m below the surface, may be as little as 10 m thick or less with salinity anomalies of 0.5–1.0 [*Castelao et al*., 2010], and can extend tens of kilometers across the shelf. *Lentz* [2003] estimates these intrusions can endure for 90 days and over the course of a summer act to raise average salinity of the MAB by 0.3.

[12] Wind-driven mixing events (more so than surface cooling) have been observed to instigate the breakdown of vertical stratification in November [*Lentz et al*., 2003]. While the MAB shelf remains substantially isothermal throughout winter, across-shelf displacements of the foot of the shelf-slope front can introduce near bottom thermal inversions. The encroachment of this warm, salty water also acts to maintain year round vertical salt stratification on the outer shelf [*Castelao et al*., 2010].

[13] Reproducing the hydrographic variability described above is a demanding test of any coastal circulation model because it requires reasonable representation of numerous

processes, including air-sea fluxes, vertical turbulent mixing, baroclinic dynamics, river sources, and shelf-edge exchange, to list only a few. Therefore, we have formulated a set of performance metrics that compare model results to full water column observations of temperature and salinity spanning the entire MAB shelf. We consider this a more rigorous test than comparisons to sea surface temperature (SST), which in the majority of models here is constrained by imposed air temperatures in the meteorological forcing and assimilation of satellite observations.

[14] MARACOOS have augmented the repeated AUGV surveys described by *Castelao et al*. [2010] by introducing long (>30 days) deployments that traverse the full MAB, and missions targeting the New Jersey coast. In 2010–2011, 16 AUGV missions (Figure 1) achieved extensive coverage of the region. A quite different sampling pattern is accomplished by National Marine Fisheries Service Ecosystem Monitoring (ECOMON) voyages. Each quarterly survey visits some 120 stations with positions selected randomly for each voyage (Figure 1, +symbols). We used full water column CTD profile data from four ECOMON voyages in 2010–2011.

[15] To characterize model performance in differing geographic regions and circulation regimes we consider three subregions indicated by color shading in Figure 1. The boundary between inner and outer shelf regions follows the 50 m isobath which is a zone of persistent T/S midshelf fronts that separate distinct regimes [*Ullman and Cornillon*, 2001]. The north-south division was a compromise between adding geographic refinement and maintaining sufficient data in each subregion to achieve statistically significant results. Region 1 thus combines both the Delaware and Chesapeake Bay entrances—there being insufficient observations to separate the zones further. Region 3 emphasizes the approaches to New York Harbor and the area influenced by the Hudson River plume and New Jersey coastal current dynamics [*Castelao et al*., 2008b; *Chant et al*., 2008; *Zhang et al*., 2009]. Region 2 between the midshelf and shelf-slope fronts intrudes shoreward at the Hudson Shelf Valley, but the absence of observations there leaves the analysis unaffected.

[16] The color-coded timelines in Figure 1 show the period of each AUGV deployment, with the ribbon colors set to indicate in which map subregion the vehicle was operating. The background dark and light shading indicates the summer stratified (May-October) and winter mixed (November-April) seasons, respectively, revealing the relative paucity of data during winter. In total there were 233 days of AUGV data during the 2 years.

[17] Webb Research *Slocum* electric gliders deployed by MARACOOS complete a vertical glide cycle in 5–20 min at a forward speed of about 0.3 ± 0.1 m s^{-1}. We averaged sequential groups of 6–24 CTD profiles, binned to 1 m standard depths, to derive average T/S profiles at roughly 1 km intervals. This averaging suppresses some of the instrument noise in individual profiles. None of the models

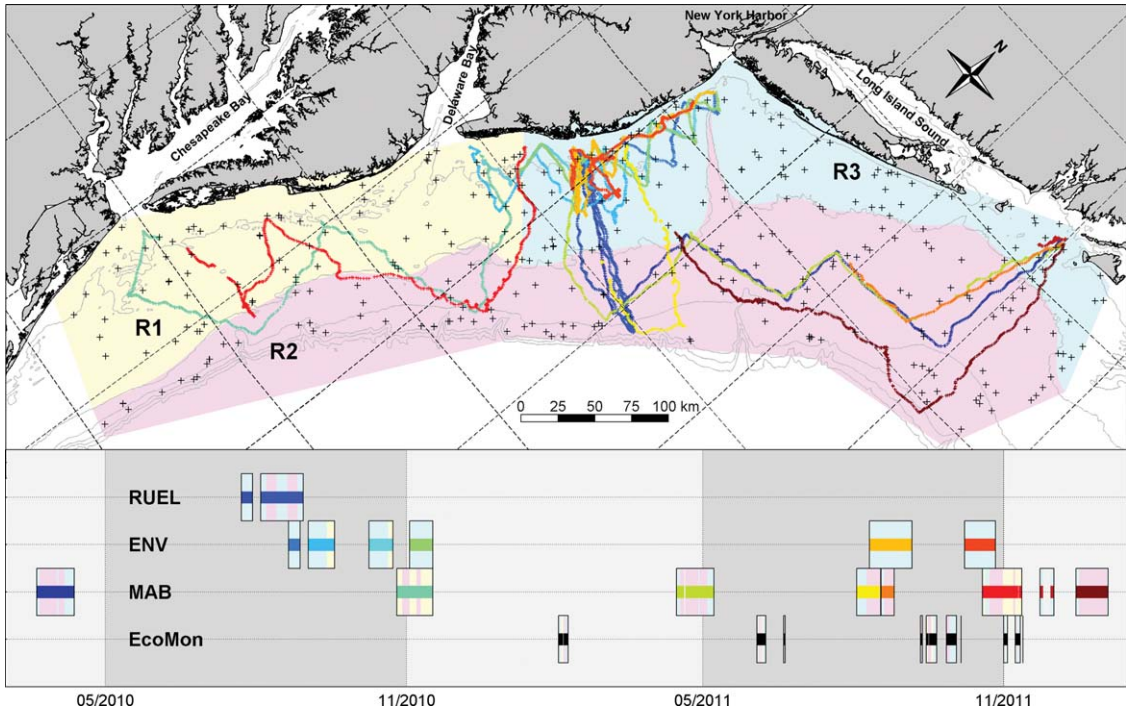

Figure 1. AUGV tracks (colored lines) and ECOMON CTD stations (+ symbols) in 2010–2011. Geographic subregions R1, R2, and R3 are shown by color shading. Timeline panel shows dates of observations grouped by project: Rutgers University Endurance Line (RUEL), New Jersey state environmental agencies (ENV), long range MARACOOS deployments (MAB), and NMFS stock assessment voyages (ECOMON). For each box on the timeline the central line color corresponds to a map track, and the background color the subregion in which the vehicle was operating at that time. Dark and light gray shading on timeline indicates summer and winter, respectively.

achieves effective horizontal resolution better than 2 km so greater data resolution would add little value. Moreover, we examined the along-track auto-correlation of individual AUGV CTD profiles and deduced a correlation length scale of about 24 km throughout much of the MAB, so to retain higher spatial resolution would sacrifice independence of the data. An example subset of the resulting binned profile data for the AUGV deployment of April 2010 is shown in the top row of Figure 2 as a function of along-track distance.

2.1.2. Surface Currents from HF-Radar

[18] The mean circulation in the MAB is largely consistent with simple dynamical models; a mean southwestward along-shelf flow is driven by a large-scale along-shelf pressure gradient and across-shelf buoyancy forcing associated primarily with the across-shelf salinity gradient [*Csanady*, 1976; *Lentz*, 2008]. Surface and bottom Ekman layers and geostrophic flow due to the along-shelf pressure gradient contribute to across-shelf currents [*Lentz*, 2008; *Zhang et al*., 2011]. Wind forcing imprints seasonal variability on the circulation. *Zhang et al*. [2009] showed that omitting the southward mean flow from the open boundary conditions of a model of the New York Bight significantly impacted the modeled circulation of the mid- and outer shelf affecting the dispersal of the Hudson River plume on time scales of several months.

[19] A lengthy archive of MARACOOS CODAR observations is available to test, at least at the ocean surface, whether real-time models capture observed subtidal shelf-wide currents. These data have previously been used to describe across-shelf transport of low salinity Hudson River water [*Castelao et al*., 2008b; *Zhang et al*., 2009], characterize seasonal variability in circulation in the New Jersey sector of the MAB [*Gong et al*., 2010], and analyze and predict Lagrangian trajectories [*Ullman et al*., 2006]. Since those studies were undertaken the MARACOOS network of long range (5 MHz) SeaSonde transmitters has expanded considerably, and from 2009 onward has provided near complete coverage from Cape Cod to Cape Hatteras to a range of roughly 100 km offshore [*Roarty et al*., 2010]. Radial velocities measured from multiple transmitter/receiver pairs are combined by optimal interpolation (OI) [*Kohut et al*., 2006] to give hourly total velocity vector observations on a 6 km spatial grid, with accompanying expected error estimates that depend on the geometry of the radar sites, range to target, and the number of radial data combined in the OI. An accepted value for the effective depth of measurements from 5 MHz radar is 2.4 m below the surface [*Ullman et al*., 2006]. From these data for 2010–2011 we have prepared a surface current data set for comparison to the models.

[20] The OI combining step uses covariance scales of 10 km across-shelf and 25 km along-shelf, so to ensure independent data we merge hourly data into 28 km bins. We exclude vectors with high expected error as an added quality control step. Near the coastline the transmitter/receiver geometry leads to low precision in the vector direction and we exclude data within 25 km of the coast as a further aggressive quality control step. Hourly data are binned to daily values.

[21] Daily binning potentially aliases energy from the semidiurnal tides, but for the four models that simulate tides we average the output in the same manner to make the comparison of velocity variability. For the three models that do not include tides a comparison between simple daily averages is reasonable.

[22] To downscale models to nested inner shelf domains a conventional approach is to construct open boundary conditions using independent harmonic tidal variability added to slowly time evolving boundary conditions from the outer domain model. Thus, some low-pass filtering strategy for the outer domain solution is required. We have effectively chosen daily boxcar filtering as that strategy for the purposes of this skill assessment.

2.2. Real-Time Models

[23] We evaluate seven hydrodynamic models of the MAB region that we know ran in real time during 2010–2011 and provide output openly accessible via the Internet. In all instances, we evaluate the fields the respective modeling groups elected to present as their "best time series" estimate of the ocean state.

[24] Some groups may have identified and remedied real-time configuration errors and superseded results with more skillful reanalyses, but we do not consider reanalyses here. So our assessment may not appraise skill as it is now. However, we assume that the real-time results were the genuine best efforts of the respective modeling groups at the time they were generated, and this approach is the only methodology we could envision to compare fairly all models. Groups who are nimble at catching faults in their real-time systems, and promptly remedying them, are rewarded in the skill assessment by virtue of our having made the model-data comparison over a lengthy 2 year period of observations.

[25] The models are introduced next. We have done our utmost to verify our synopses of model configurations with the creators, but accept responsibility for any misrepresentations of the modeling system features. Key elements of each model are given in Table 1, notably the open access THREDDS (Thematic Real-time Environmental Distributed Data Services) servers at which we accessed output. The adoption of THREDDS and CDM (Common Data Model) standards by most of the groups archiving these data greatly facilitated this project by enabling us to use software tools developed for interoperability of CDM formatted data to access the output of several diverse models with minimal code customization.

2.2.1. HYCOM

[26] The HYbrid Coordinate Ocean Model analyzed here is global with 1/12° horizontal resolution [*Chassignet et al*., 2009]. The real-time system is operated by the U.S. Navy NAVOCEANO center and uses the NCODA (Navy Coupled Ocean Data Assimilation) [*Cummings*, 2005] multivariate optimal interpolation (MVOI) scheme to assimilate satellite sea surface height (SSH) and SST, and temperature and salinity profiles from Argo floats and ships of opportunity. Surface meteorological forcing is from the U.S. Navy Operational Global Atmospheric Prediction System (NOGAPS) [*Rosmond et al*., 2002]. Tides are not simulated. River inflows are from a climatological monthly global database [*Vörösmarty et al*., 1996]. Vertical coordinates are a hybrid of isopycnal layers in deep water and a mixture of pressure and terrain following on continental

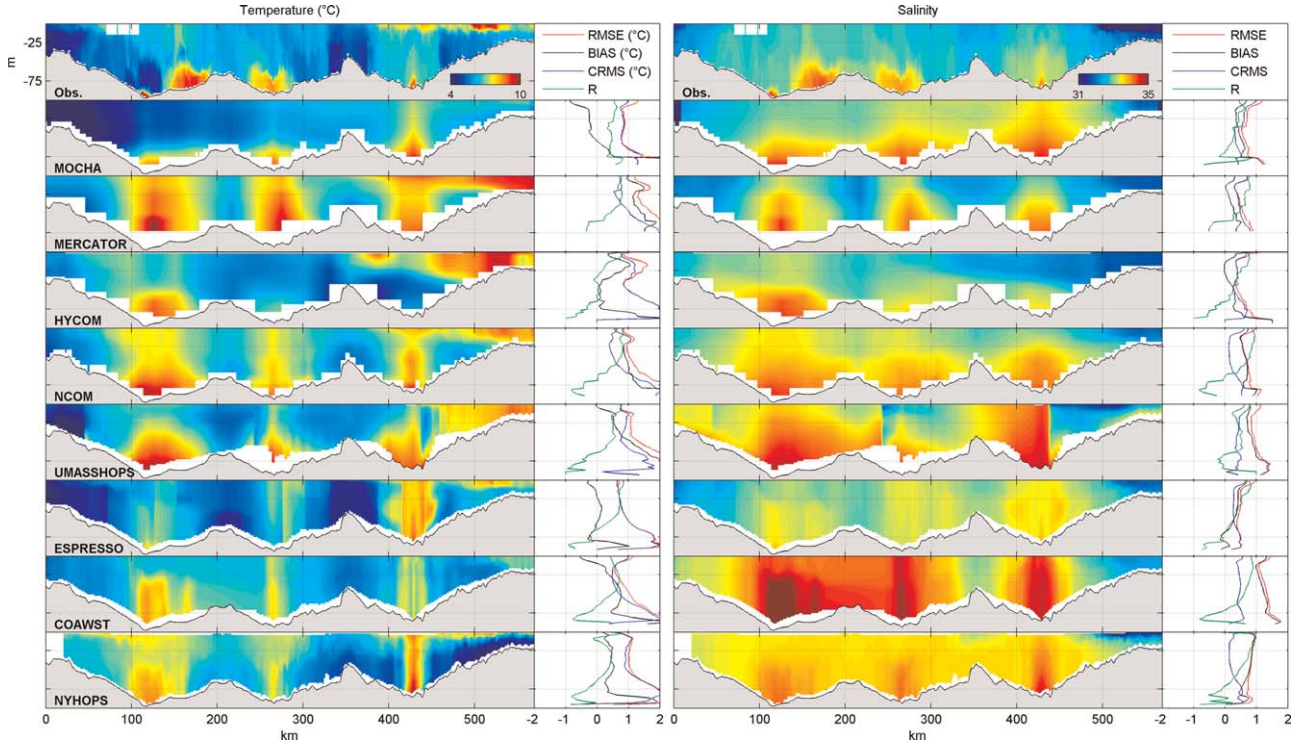

Figure 2. (left) Temperature and (right) salinity versus along-track distance and depth for the MAB AUGV deployment of 04/2010 (see Figure 1 for trajectory). Top row: reduced resolution observation set obtained following binning described in text. Other rows: Individual model results interpolated to the AUGV track at binned observation times. Model labels are as defined in Table 1. Summary line plots show root mean squared error (RMSE), mean bias (MB), centered root mean squared error (CRMS), and cross-correlation (R) versus depth for each model compared to this set of observations.

shelves [*Chassignet et al*., 2006]. There are 32 layers in total with typically 10 active layers on the MAB shelf. We access daily snapshots output at fixed vertical levels interpolated by the HYCOM group.

2.2.2. NCOM

[27] The Navy Coastal Ocean Model is a 1/8° resolution global system also operated by NAVOCEANO. Aspects of the code are described by *Barron et al*. [2006] and the real-time system is outlined by *Kara et al*. [2006]. NCOM assimilates essentially the same data as HYCOM but using the Navy's MODAS system [*Fox et al*., 2002] to prepare the analysis fields. A slow data insertion technique [*Rhodes et al*., 2002] introduces incremental adjustments over the analysis interval. The surface meteorology is NOGAPS. Tides are modeled. River inflows are monthly climatology [*Barron and Smedstad*, 2002]. The model has a hybrid vertical coordinate: 19 terrain following levels on the shelf stretched to give resolution better than 1 m at the surface; 21 fixed depth z levels in the deep. The open access output is 3 h interval snapshots interpolated to fixed vertical levels.

2.2.3. MERCATOR

[28] MERCATOR is based on the NEMO ocean model [*Madec*, 2012] and uses a z level coordinate system with partial cells to represent sloping bathymetry. The daily average output analyzed here is from the 1/12° resolution operational Atlantic Ocean model with 43 vertical levels. On the MAB shelf where the shelf-break parallels the 100 m isobath there are 12 vertical levels with thickness increasing from 5 m at the surface to 25 m at 100 m depth. Output for 2010–2011 assimilated SSH, SST and profile data using a reduced Kalman filter approach [*Brasseur et al*., 2005]. MERCATOR is forced with surface meteorology from the European Centre for Medium-range Weather Forecasts (ECMWF). Tides are not modeled. River inflows are *Dai and Trenberth* [2002] monthly climatology.

2.2.4. NYHOPS

[29] The New York Harbor Observing and Prediction System [*Blumberg et al*., 1999; *Georgas and Blumberg*, 2010] is based on the Princeton Ocean Model [*Blumberg and Mellor*, 1987]. The domain spans the MAB shelf to the 200 m isobath from Cape Cod to Maryland; it does not encompass the approaches to Chesapeake Bay. The curvilinear horizontal grid resolution is approximately 4.5 km at the inner shelf. There are 10 terrain following vertical levels. Outputs from a companion model of wind-wave growth, propagation and decay [*Schwab et al*., 1984] are used in a parameterization of wave-current effects on bottom drag [*Grant and Madsen*, 1979]. T/S observations inside New York Harbor are assimilated by optimal interpolation every 24 h at the initialization of a new forecast. Over 200 river and stream inflows are specified using 6 h forecasts from NOAA's Advanced Hydrological Prediction System [*McEnery et al*., 2005] plus mean inflows at more than 240 urban point sources. Meteorological forcing is from the NCEP (National Centers for Environmental

Table 1. Key Features of the Models Evaluated

Model	THREDDS[a] Catalog URL	Resolution in MAB	Output Interval	Surface Forcing	Tides	Waves[b]	Rivers	Open Boundary Data	Assimilation Data	Assimilation Method
HYCOM	http://tds.hycom.org/thredds/dodsC/glb_analysis.html	7 km 10 z levels in h<100 m	Daily snapshots	NOGAPS	No	No	Monthly climatology	Global	SSH, SST, T/S profiles	MVOI
NCOM	http://edac-dap.northerngulfinstitute.org/thredds/catalog.html	11 km 19 TF[c] levels in h<100m	3 h snapshots	NOGAPS	Yes	No	Monthly climatology	Global	SSH, SST, T/S profiles	Weighted sum
MERCATOR	Register at http://myocean.eu. Python script motu-client.py used to access data set PSY2V4R2	7 km 12 z levels in h<100m	Daily average	ECMWF	No	No	Monthly climatology	Global	SSH, SST, T/S profiles	SEEK filter
NYHOPS	http://colossus.dl.stevens-tech.edu/thredds/dodsC/fmrc/NYBight (best time series)	4.5 km 10 TF levels	10 min averages	NCEP NAM	Yes	Yes	Hydrology forecast & point sources	WOA1998 T/S, MDT[d]+ETSS sea level	NY Harbor T/S	OI
ESPreSSO	http://tds.marine.rutgers.edu/thredds/dodsC/roms/espresso/2009_da/his.html	5 km 36 TF levels	3 h snapshots	NCEP NAM	Yes	No	Daily observed	HYCOM, with MOCHA adjusted T/S	SSH, SST, CODAR	4D-VAR
UMassHOPS	http://aqua.smast.umassd.edu:8080/thredds/catalog/pe_shelf_ass/fmrc/catalog.html (best time series)	15 km 16 TF levels	3 h snapshots	NCEP GFS	No	No	None	No data radiation condition	SSH, SST	Feature model OI
COAWST	http://geoport.whoi.edu/thredds/dodsC/coawst_2_2/fmrc/coawst_2_2_best.ncd.html	5 km 16 TF levels	2 –h snapshots	NCEP NAM	Yes	Yes	None	HYCOM	HYCOM T/S	Nudging
MOCHA climatology	http://tds.marine.rutgers.edu/thredds/catalog/other/climatology/mocha/	5 km 50 z levels in h<100 m	Monthly	N/A	N/A	N/A	N/A	N/A	T/S profiles	Weighted least squares

[a]Thematic Real-time Environmental Distributed Data Service.
[b]Wave-current interactions.
[c]Terrain-following.
[d]Geostrophic mean dynamic topography.

Prediction) 12 km resolution North American Mesoscale model (NAM) [*Janjic*, 2004]. Conditions imposed at the open boundaries are World Ocean Atlas 1998 (WOA) [*Antonov et al*., 1998] monthly climatological T/S, mean geostrophic current and sea level, low frequency sea level variability from the National Oceanic and Atmospheric Administration (NOAA) Extratropical Storm Surge Model (ETSS), and harmonic tides [*Mukai et al*., 2002].

2.2.5. ESPreSSO

[30] The Experimental System for Predicting Shelf and Slope Optics uses the Regional Ocean Modeling System (ROMS; www.myroms.org). The model domain extends from Cape Cod to Cape Hatteras and the coast to beyond the shelf-break, with 5 km horizontal resolution and 36 terrain following vertical levels. Open boundary values are taken from global HYCOM with adjustments using MOCHA climatology (introduced below) to reduce bias in T/S, and the addition of harmonic tides [*Mukai et al*., 2002]. Meteorology forcing is NAM. Inflows for the seven largest rivers are from daily average USGS discharge data. Strong constraint four-dimensional variational (4D-Var) data assimilation [*Moore et al*., 2011] is used to incorporate satellite SSH from Jason-2, satellite SST from infrared and microwave radiometers, monthly MOCHA T/S climatology, and hourly CODAR surface currents [*Zavala-Garay et al*., 2012].

2.2.6. UMassHOPS

[31] The University of Massachusetts at Dartmouth has configured a version of the Harvard Ocean Prediction System [*Robinson et al*., 1988] for the MAB and Gulf of Maine [*Schmidt and Gangopadhyay*, 2012] out to beyond the Gulf Stream. An analysis of T/S and geostrophically balanced velocity for assimilation is computed using feature model sets for Gulf Stream meanders and rings [*Gangopadhyay et al*., 1997] and the Gulf of Maine and Georges Bank [*Gangopadhyay et al*., 2003], melded with WOA climatology for the MAB. Dynamic height is adjusted to the shelf bathymetry [*Brown et al*., 2007]. Horizontal resolution is 15 km. There are 16 terrain-following vertical levels. Tides are not modeled. The surface forcing is GFS [*NCEP*, 2003]. The *Orlanksi* [1976] open boundary radiation condition does not require external data.

2.2.7. COAWST

[32] The Coupled Ocean-Atmosphere-Waves and Sediment Transport model covers the U.S. east coast and Gulf of Mexico [*Warner et al*., 2010]. The ocean component is ROMS with 5 km spatial resolution and 16 vertical levels. COAWST is two-way coupled to a regional surface waves model (SWAN) [*Warner et al*., 2008] with 60 direction and 24 frequency bins. Waves modify currents via Stokes drift and bottom drag, and through the influence of wave breaking on mixed layer turbulence. NAM atmospheric forcing drives both models. Boundary conditions to SWAN are from WaveWatch III; boundary conditions to ROMS are HYCOM. There is assimilation in the form of nudging to HYCOM throughout the domain. Tides are modeled. There are no river inflows.

2.2.8. MOCHA

[33] The Mid-Atlantic Climatological Hydrographic Analysis [*Fleming and Wilkin*, 2010] is a three-dimensional regional climatology of temperature and salinity based on a comprehensive collection of historical in situ observations in the MAB and Slope Sea. Though not a prognostic hydrodynamic model, MOCHA nevertheless offers a prediction of the ocean state at any time; namely, the long-term average of prior observations. MOCHA uses weighted least squares to produce monthly analyses on a 4.5 km horizontal grid. In shelf waters ($<$ 100 m depth) there are 22 vertical levels. None of the 2010–2011 glider observations used in the skill assessment were included the MOCHA. MOCHA does not provide velocity.

2.3. Model Versus Observation Skill Metrics

[34] We use conventional measures of model-observation difference to quantity model skill with respect to AUGV T/S data: root mean squared error (RMSE), mean bias (MB), and centered root mean squared (CRMS) error. Denoting a set of observational values as $\mathbf{o}$, corresponding model predictions as $\mathbf{m}$, using an over bar to denote the mean of the set, and a prime to denote perturbations from the mean, i.e., $\mathbf{m}' = \mathbf{m} - \overline{\mathbf{m}}$, these error metrics are defined as:

$$\mathrm{RMSE} = \left[\overline{(\mathbf{m}-\mathbf{o})^2}\right]^{\frac{1}{2}}, \tag{1}$$

$$\mathrm{MB} = \overline{\mathbf{m}} - \overline{\mathbf{o}}, \tag{2}$$

$$\mathrm{CRMS} = \left[\overline{(\mathbf{m}'-\mathbf{o}')^2}\right]^{\frac{1}{2}}. \tag{3}$$

[35] On a Cartesian graph with axes MB and CRMS, RMSE is the distance from the origin. Treating CRMS as negative when model variance is less than observed, *Jolliff* [2009] introduced the so-called *target* diagram. We observe no patterns in the sign of MB that inform our discussion, so for clarity we plot absolute value of both MB and CRMS. This effectively transforms all values to the first quadrant of a target diagram.

[36] CRMS can be expanded in terms of variance and correlation as

$$\mathrm{CRMS}^2 = S_m^2 + S_o^2 - 2S_mS_o\mathrm{R}, \tag{4}$$

where S_m and S_o are model and data standard deviation, and R is the cross-correlation coefficient $\mathrm{R} = \overline{\mathbf{m}'\mathbf{o}'}/(S_mS_o)$. Exploiting the law of cosines, *Taylor* [2001] observed that for a point plotted in polar coordinates with radius S_m and azimuth $\cos^{-1}\mathrm{R}$, CRMS is the distance to the point $(S_o, 0)$. These are widely termed *Taylor* diagrams; it is common practice to normalize values by the observation standard deviation S_o. Taylor diagrams do not depict MB, so it must be reported separately.

[37] For vector surface current data from CODAR, complex values are formed from the velocity components: $\mathbf{o} = \mathbf{u}_o + i\mathbf{v}_o$ and $\mathbf{m} = \mathbf{u}_m + i\mathbf{v}_m$. We present vector MB $\overline{\mathbf{m}} - \overline{\mathbf{o}}$ and the complex vector correlation $\hat{\mathrm{R}} = \overline{\mathbf{m}'\mathbf{o}'^*}/(S_mS_o)$, where the asterisk denotes complex conjugate, by computing the mean of time series of velocity data at each location. For complex data the standard deviations are e.g., $S_m = \overline{\mathbf{m}'\mathbf{m}'^*}^{1/2}$. It follows that $\hat{\mathrm{R}}$ is complex; its phase represents consistent directional errors in velocity variability. In practice, this has little spatial structure so we map only the

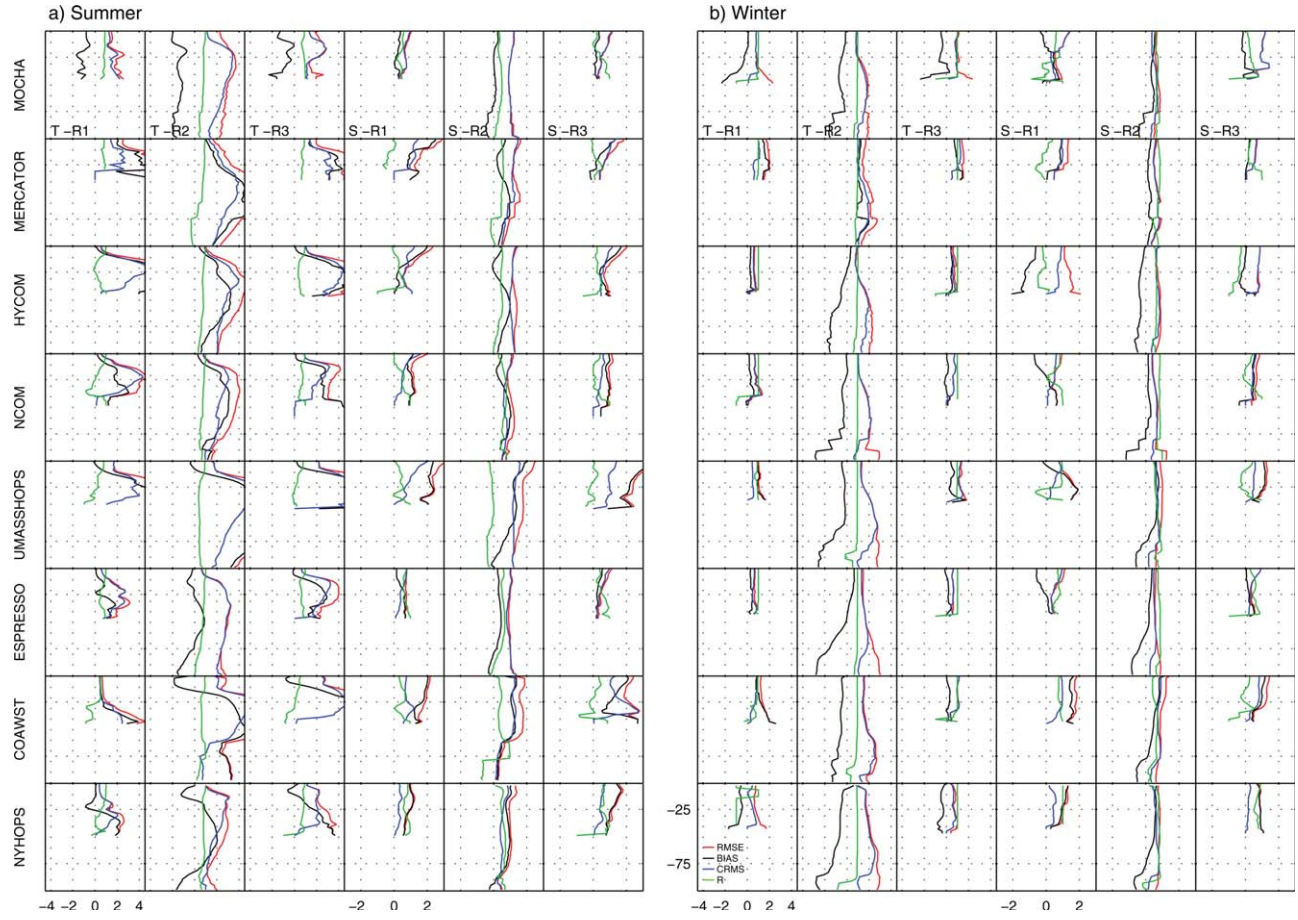

Figure 3. Vertical profiles (0–100 m depth) of skill metrics for each model (rows) for (a) summer and (b) winter. Columns are grouped by (left) temperature and (right) salinity in each panel with columns for the three subregions (Figure 1: R1, R2, and R3). Red: Root mean squared error RMSE; black: mean bias MB; blue: centered root mean squared error CRMS; green: Cross correlation R. Except for R, values are in the observation units: °C for T, no units for S. Columns without *x* axis tick labels have the same values as to their left.

magnitude. We depict skill of the velocity direction in the subregions using polar histograms.

[38] Individual AUGV deployments and ECOMON voyages naturally give differing values for any given skill metric; likewise for temporal subsets of the CODAR data. This uncertainty in the skill scores must be contemplated when concluding whether differing models perform consistently better or worse, or are indistinguishable. We characterize uncertainty in the T/S skill scores in two ways. First, in Taylor diagrams we show results for separate AUGV deployments to indicate how much the same model metric can differ for independent subsets of observations. Second, in MB versus CRMS plots we show the 95 percentile range of an ensemble derived using random sampling with replacement of entire deployments or voyages from the combined set of AUGV and ECOMON data. Conventional bootstrap approaches to estimating uncertainty would do this slightly differently: by randomly sampling individual depth profiles from the entire set irrespective of mission. We chose our methodology out of concern that particularly long AUGV deployments could skew the comparison toward models that happened to have scored especially well at those times, or in the regions they sampled. By this we are acknowledging that our along-track binning procedure is unlikely to have reduced the data to a set of truly independent profiles, and that distinct deployments are decidedly *more independent* (if we may be permitted this loose expression of the concept of independence).

3. Results

3.1. Subsurface Temperature and Salinity

[39] Line plots in Figure 2 show vertical profiles of the skill metrics averaged across each standard depth for this AUGV deployment. RMSE, MB, and CRMS are in T/S observation units. A number of patterns emerge. Several models are biased warm and slightly salty. Skill is highest near the sea surface (high R; low CRMS) presumably due in part to the constraints of imposed atmospheric conditions and assimilation of satellite data (in the case of temperature). Errors generally increase with depth. The scores are erratic at the deepest depths where the data show warm, salty Slope Sea water encroaching onto the shelf—though these metrics may be unreliable given there are only a handful of observations below 75 m depth.

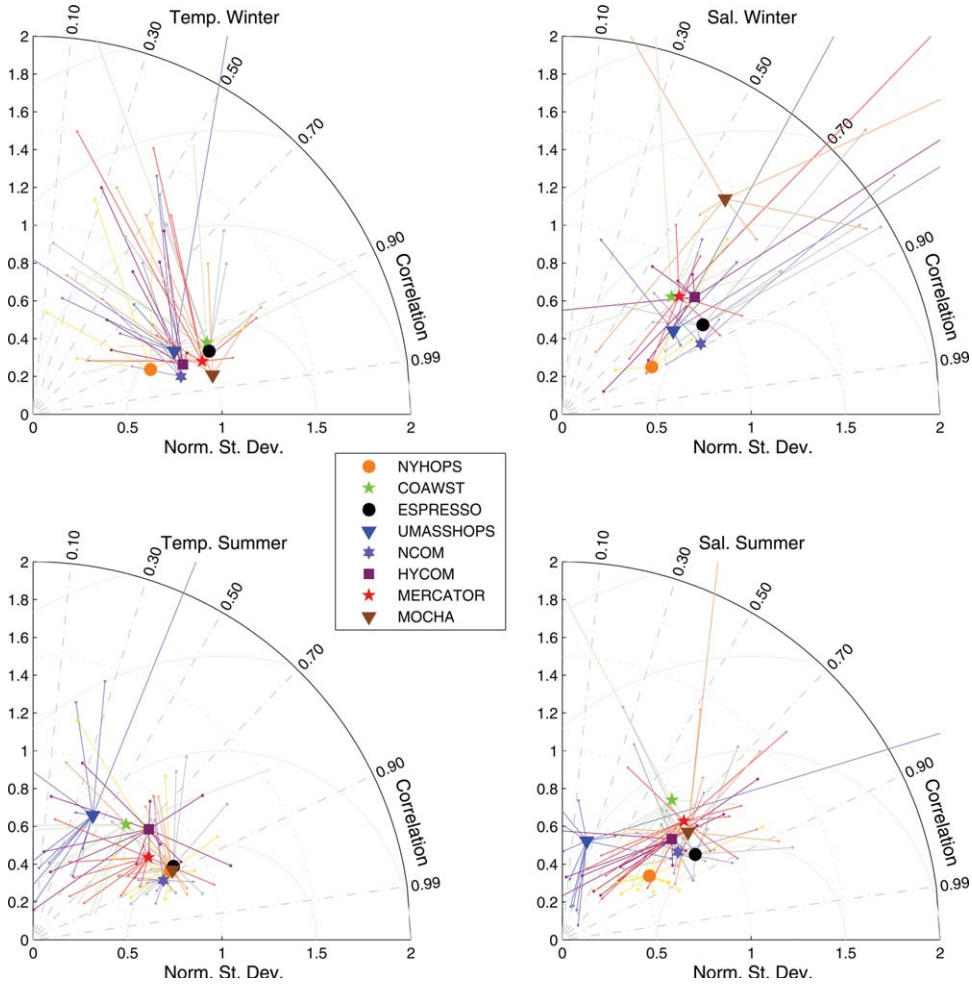

Figure 4. Normalized Taylor diagrams for models (see color key in legend) compared to AUGV observations of temperature (left) and salinity (right) for winter (top row) and summer (bottom row). Large symbols show aggregate of statistics across all deployments; light lines connect to small color symbols showing the contributing statistics from the individual AUGV deployments to illustrate the spread in statistics between deployments. In a normalized Taylor diagram radius is model standard deviation error divided by observation standard deviation, azimuth is arccos of cross correlation (R), and distance to point (1,0) on the abscissa is Centered RMS. Light lines that reach axis limits connect to values with negative R or extreme standard deviation error that are not plotted.

[40] Figure 3 presents the same profiles of skill metrics computed across all AUGV deployments and ECOMON voyages, but grouped according to subregions R1–R3, and winter or summer (Figure 1). Model errors are consistently less in winter than summer, but Taylor diagrams (Figure 4) show this distinction vanishes when scores are normalized by observation variance. Again, the majority of models are biased warm and salty. In summer, temperature errors are low near the surface and both RMSE and bias typically take a mid-depth maximum. This would be consistent with the models imprecisely tracking variability in the depth of the strong summer thermocline. For winter temperature, there is a tendency for RMSE to increase steadily with depth without a local maximum. For the inner shelf subregions R1 and R3, ESPreSSO and NYHOPS appear to have the better performance for temperature, but are inferior to MOCHA climatology.

[41] For summer salinity, ESPreSSO appears to give the best results in R1 and R3, with NYHOPS and NCOM little different. The other models have pronounced summertime high salinity biases at the surface (exceeding three units in the case of UMassHOPS and COAWST) though these biases decrease in magnitude with depth. CRMS in winter salinity exhibits a similar but less pronounced trend of being greater at the sea surface and decreasing with depth. There is rather less difference between the models in the outer shelf subregion R2, in either season, though the same three models ESPreSSO, NYHOPS, and NCOM generally outperform the others.

[42] While there are patterns to the vertical structure of the skill metrics, these are largely consistent between models such that if we were to compare only near surface, or mid-depth, or near seafloor results, it would not fundamentally alter a ranking of models. Consequently, we proceed

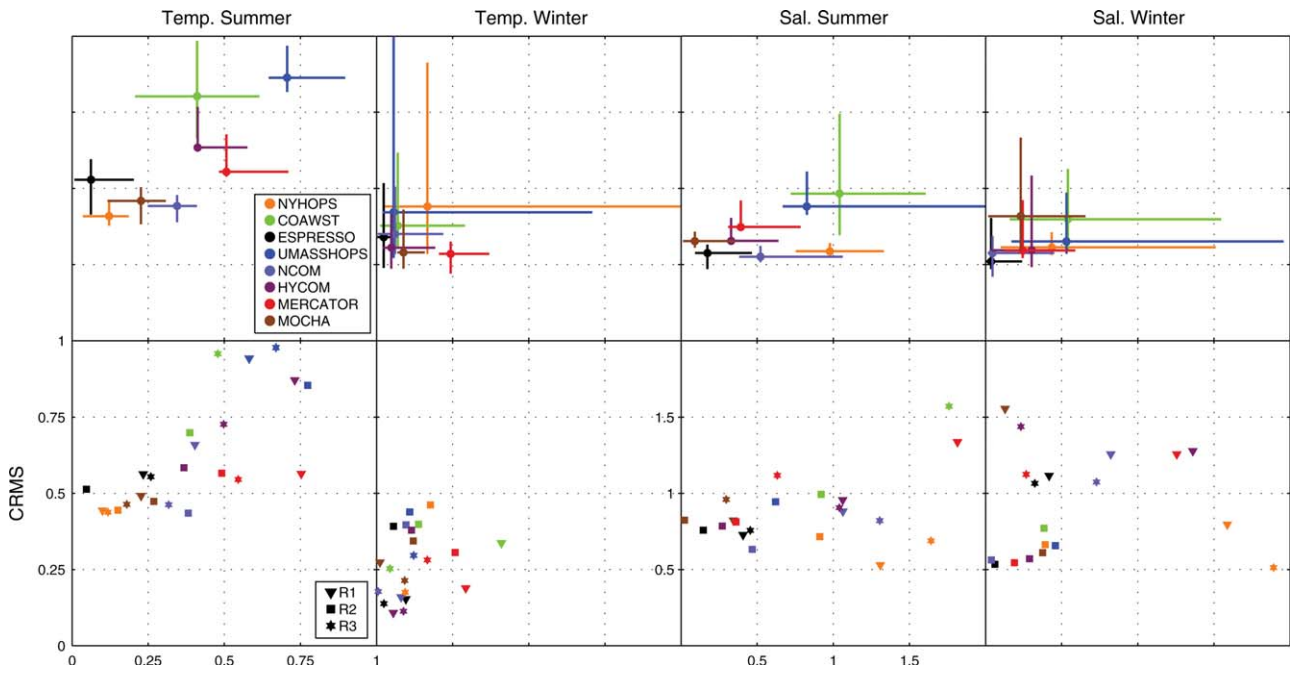

Figure 5. Mean bias (MB) and centered root mean square error (CRMS) of each model (see legend for model color key) with respect to AUGV and ECOMON CTD observations of temperature (left two columns) and salinity (right two columns), calculated separately for summer and winter seasons. Bottom row: Values for individual subregions R1–R3 (see Figure 1). Top row: Values for all sub-regions combined; horizontal and vertical bars depict 95% confidence limits (see text).

by averaging the skill metrics over depth for further model comparison.

[43] Figure 4 presents Taylor diagrams for T and S grouped by season and with coordinates normalized by S_o. Each model is depicted in a distinct color by a graphical object that somewhat resembles a sea urchin: results calculated for all AUGV and ECOMON data aggregated together are marked by a large symbol (the "body"); the spines that emanate from this connect to the results calculated separately for each deployment/voyage. The broad spread of the spines indicates how much results can vary for the same metric computed from independent sets of observations, and highlights that there is uncertainty in the analysis.

[44] On first thought the body of the urchin might be expected to be centered on the pattern of spines—being as it is based on the aggregation of all individual deployment data. This is not the case because statistics for each AUGV deployment use means ($\overline{\mathbf{o}}$, $\overline{\mathbf{m}}$) computed for that deployment alone, which differs from the mean of the aggregation, so variance and correlation are defined with respect to different means. The aggregated data can include a trend or long period variability within the season (exacerbated by sampling clustered toward late fall and early spring; Figure 1) whereas for the individual statistics this low-frequency variability is relegated to the MB. (Recall MB is not depicted in a Taylor diagram). For temperature, we see correlations tend to be greater for the aggregated data than for most of the individual deployments. It is rather easier to model low frequency temperature variability than capture submesoscale variance along an AUGV trajectory, so this is a caution against assuming aggregate skill scores apply equally at higher frequencies. This effect can occur in any skill assessment that utilizes time series that resolve a long period cycle. This is why we present both individual and aggregate scores in Figure 4. Note that we cannot average the statistics computed for individual deployments and plot an ensemble result because the geometry in the law of cosines holds only when $\overline{\mathbf{o}}$ and $\overline{\mathbf{m}}$ are held constant.

[45] Overlooking the uncertainty in the Taylor diagram metrics betrayed by the lengthy urchin spines, some patterns emerge from the aggregate data skill scores. Recalling that points closest to (1,0) have the least CRMS error we see that models NCOM, ESPreSSO, and NYHOPS generally perform best overall.

[46] An exception is winter temperature when MERCATOR has the lowest CRMS, and COAWST and HYCOM have CRMS less than NYHOPS. However, none of these achieve the skill of MOCHA climatology. Also, for winter salinity, UMassHOPS has comparable CRMS to NYHOPS, and MOCHA climatology performs poorly.

[47] In summer, NCOM, ESPreSSO, and NYHOPS have the best correlation for both T and S though the normalized variance for NYHOPS salinity is low, as it is in winter. ESPreSSO is consistently closest to the unit radius indicating its variance matches observations. The more skillful models tend to have somewhat less spread in the individual metrics (i.e., shorter spines).

[48] Ranking models by correlation or CRMS gives reasonably consistent results so we turn our attention to MB in Figure 5. Values are again normalized by observation standard deviation. Panels in the bottom row show MB versus CRMS for each model (marker color) computed for data aggregated over all deployments, grouped by season,

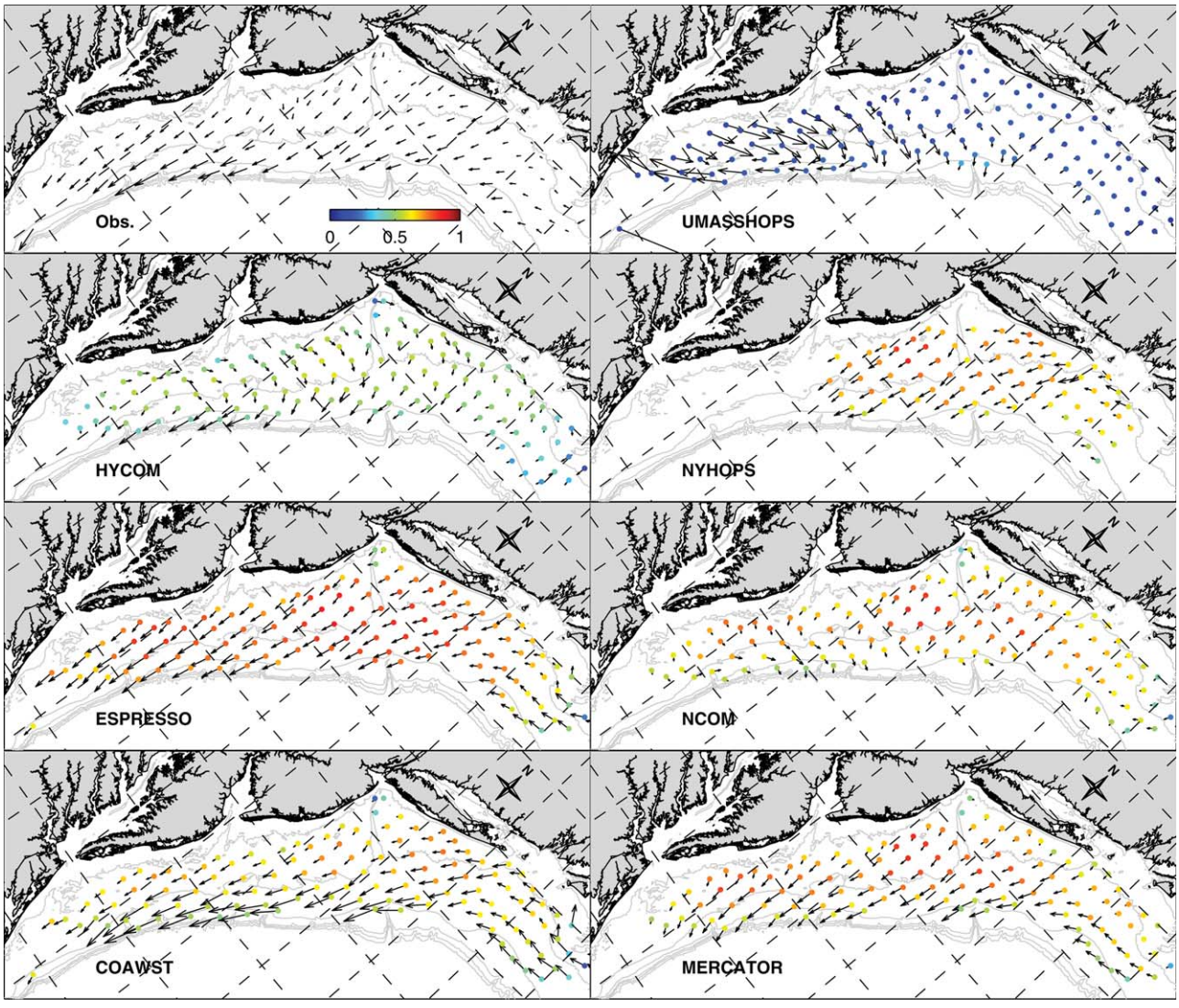

Figure 6. Mean (2010–2011) surface velocity from HF-radar (top left) and the respective models at 2 m below surface. For the models, colored dots show the magnitude of the complex correlation between daily variability of HF-radar and model.

and further split according to subregion R1–R3 (marker shape). Symbols with the same color tend to cluster together so we see no tendency for any model to perform especially well in one region over others. Subregion 2 results (square symbols) plot closer to the origin (with the exception of winter temperature) suggesting that the outer shelf is better modeled than the inner shelf.

[49] With no clear distinction between the models by subregion we consider statistics computed for the entire shelf in the top row of Figure 5. Crosses indicate 95% confidence limits derived as described in section 2.3. It seems no model can achieve CRMS error less than one quarter of the observation standard deviation for either T or S, while (except for summer temperature) no model CRMS is ever worse than half of S_o. With the exception of summer salinity for just a few models, all results fall above the unit slope line indicating CRMS contributes more than MB to overall mean squared error (distance from origin). The model with the smallest MB is ESPreSSO in all cases, though it is defeated by MOCHA climatology for summer salinity. MERCATOR and HYCOM have lower salinity bias than NYHOPS in both seasons. NYHOPS winter temperature performance is relatively poor, while summer temperature is extremely good, results that are consistent with Figure 4.

[50] A general observation, consistent with the Taylor diagrams, is that the models with generally greater skill (NYHOPS, ESPreSSO and NCOM) have less uncertainty in the metrics.

3.2. Surface Currents

[51] Mean velocity vectors are shown in Figure 6 along with color symbols indicating the magnitude of the vector correlation $\hat{R}$ for daily variability. COAWST, MERCATOR, ESPreSSO, and NYHOPS mean currents most closely resemble the observations (but remember ESPreSSO assimilates these data). These four models have the highest correlation, although COAWST is slightly less than the others. NCOM has a poor representation of the

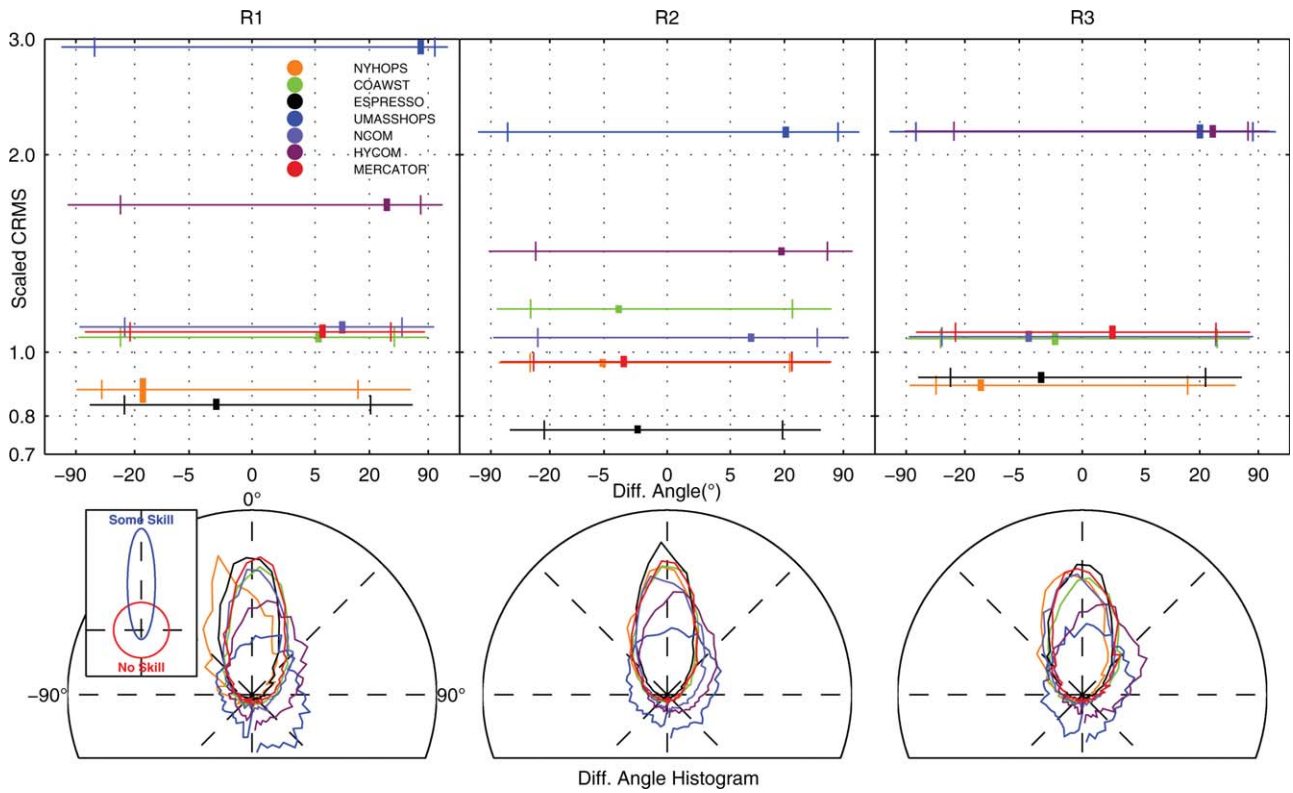

Figure 7. Velocity variability statistics for the sub-regions R1, R2, and R3 (see Figure 1) arranged by columns from left to right. Bottom row: polar histograms of model minus data direction error; positive angle for model current to the right of observed. Top row: Average for all points in each sub-region of angle error (abscissa) and centered RMS of speed (ordinate). The extent of each bar is the 90-percentile spread of direction error; small tick marks are 70 percentile; heavy tick mark is mean direction error. Height of center tick is 90 percentile spread of CRMS speed.

mean flow but is competitive in correlation. UMassHOPS performance is lackluster.

[52] Figure 7 aggregates results for subregions R1–R3 depicting summary statistics for speed variability and direction errors. Plots in the bottom row are polar histograms of direction errors (model minus data) with the mean velocity retained in the difference. The closer the major axis of the roughly elliptical shape aligns with 0° the more accurate is the direction on average, while the more eccentric the ellipse the more consistent is the agreement. We summarize the results further in the top row; each model is represented with a horizontal bar the breadth of which spans the 90 percentile range of direction error, and small tick marks show the 70 percentile limits. The vertical axis is the normalized CRMS error of velocity magnitude. The heavy tick mark on the bar shows the mean direction error, and its vertical span is the 90 percentile spread of CRMS. For example, a heavy tick mark at [5°, 0.9] would say that on average the model velocity is 5° to the right of observed, and RMS error in speed is 90% of the observed standard deviation; a perfect model would plot at [0°, 0]. Both axes are stretched.

[53] ESPreSSO assimilates the HF-radar data so it is not surprising it achieves the smallest mean directional error in R1 and R2, and close to smallest in R3. COAWST and MERCATOR perform well everywhere in direction, closely followed by NCOM which has smaller direction errors than NYHOPS in R1 and R3. NYHOPS performs best in terms of CRMS in speed on the inner shelf (R1, R3). MERCATOR and NYHOPS are comparable on the outer shelf (R2). In all subregions, HYCOM and UMassHOPS are much too energetic and have mean direction errors exceeding 20°.

4. Conclusions

[54] We have used 2 years of observations of surface velocity and subsurface temperature and salinity to evaluate best-estimate analysis fields from seven hydrodynamic modeling systems that deliver real-time predictions of the MAB continental shelf. An eighth "model" is a regional climatology (T/S but not velocity).

[55] The skill metrics are conventional: bias, root mean squared error, centered RMS and cross correlation. The data set is large enough to evaluate separately the outer shelf, and northern and southern inner shelf, but this distinction does not find that particular models have regional strengths or weaknesses. Nor does separation of results by season lead to any difference in a ranking of overall skill. Relative model skill is quite consistent.

[56] Errors are less in winter for T/S for all models but not when normalized by observation variance showing that lower wintertime errors stem from there being less dynamic range in the ocean state in vertically well-mixed conditions.

Other common patterns are that skill is greater at the surface and seafloor for summer T/S, and least in the middle of the water column, while winter skill is generally best at the surface and diminishes with depth.

[57] Averaging over the water column but still separating the results into the contributions from bias and centered RMS we find that CRMS dominates the error. It is never less than 25% of observed standard deviation. Bias is always less than this, and even more so in winter than summer.

[58] In Taylor diagrams, which discriminate between correlation and CRMS, we find aggregating all data gives higher correlations than when glider deployments are considered separately. We speculate that long period variability is intrinsically more predictable (e.g., the seasonal cycle) so that when variability is defined with respect to a longer time average this somewhat inflates the skill. Consequently, models may be rather less skillful at the mesoscale to submesoscale than metrics derived from a year or more of data would suggest. The 233 days of T/S data available to us for 2010–2011, very unevenly distributed by season, were not sufficient to take this aspect of the analysis further.

[59] For temperature and salinity, three of the models, NCOM, ESPreSSO, and NYHOPS, tend to perform better than the other dynamic models for most metrics in the majority of situations (season, region, and vertical structure) though there are certainly occasional exceptions to this ranking. Of note is that MOCHA climatology is as good or better than several of the dynamic models in many situations.

[60] While the assessment of T/S skill was made with respect to data that were not assimilated by any of the systems, the velocity data *were* assimilated by ESPreSSO. This limits what we can state about its velocity skill. We find that the models NYHOPS, MERCATOR, and COAWST simulate annual mean inner shelf-surface velocity that closely resembles the HF-radar mean, while the other three models decidedly do not. In terms of daily variability in current speed and direction, and vector cross correlation, NYHOPS, MERCATOR, COAWST, and NCOM perform well (in addition to ESPreSSO). In the inner shelf regions (R1 and R3) COAWST and MERCATOR have the better direction skill and NYHOPS the best variance among the models that do not assimilate velocity data.

[61] The scope of this model assessment was dictated in large measure by the data available for corroboration, but also by our suppositions as to the ocean state information most needed as boundary conditions for models of the MAB inner shelf and estuaries. Our conjecture is that three-dimensional T, S, and velocity are all important.

[62] We could not assess subsurface velocity. Only NYHOPS and COAWST include the direct effect of atmospheric pressure gradients on sea level so we have not considered storm surge. Nor did we develop a metric for tidal sea level variability, since estuary/inner shelf models are generally quite skillful in this respect by use of independent tidal harmonic boundary conditions.

[63] Of the seven models we consider, NYHOPS, ESPreSSO, and NCOM generally exhibit superior performance when considering 3-D T/S. Their skill is comparable to, but often eclipsed by, MOCHA climatology. Surface velocities modeled by COAWST, MERCATOR, and NYHOPS very closely resemble CODAR observations and are little different from the output of the CODAR-assimilating system ESPreSSO. Whether these represent reliable sources of *three-dimensional* velocity remains uncertain. The across-shelf density gradient of the MAB shelf, which is dominated by salinity, makes a modest but non-negligible contribution to vertical shear in the along-shelf mean flow [*Lentz*, 2008]. The extent to which some models poorly simulate salinity might bring into question their subsurface velocity skill.

[64] NOAA supports two Operational Forecast Systems (OFS) in the MAB for the Chesapeake Bay (CBOFS) [*Lanerolle et al*., 2011] and Delaware Bay (DBOFS) [*Schmalz*, 2011]. As presently configured, T/S open boundary conditions imposed along arcs roughly 100 km from the mouth of each of these estuaries are derived from the NCEP Real-Time Ocean Forecast System (RTOFS) implementation of the HYCOM system evaluated here. Coastal tide gauge data merged with ETSS are used for subtidal frequency sea level. The mean inflow/outflow and subtidal frequency velocity variability is assumed to be zero, yet mean velocity open boundary conditions were shown to have a significant impact in a model by *Zhang et al*. [2009] of the New Jersey inner shelf. Thus, there is scope for improving the far-field ocean state information used by present NOAA operational estuary models.

[65] We conclude this discussion with some speculation on the elements that might comprise a more skillful modeling system than any single model considered here.

[66] The COAWST and NYHOPS models are both coupled to surface wave models and incorporate wave-related processes in the surface and bottom boundary layers. It is quite possible that these added dynamics contribute to their skill in simulating surface current variability and indicate a fruitful direction for model system development. But surface velocity skill in COAWST is not significantly different from the global model MERCATOR. Features of MERCATOR affecting the MAB momentum balance that might be superior to other models could be the treatment of sea ice and rivers that set buoyancy inflows from the Labrador Sea, the mean dynamic topography used in SSH assimilation, or ECMWF meteorology forcing. It would take further comparison of the global systems to unravel this. The absence of rivers in COAWST almost certainly upsets the across-shelf density gradient and works against mean current skill.

[67] The modest edge in velocity skill that ESPreSSO has over COAWST and MERCATOR suggests that assimilation of CODAR is worth retaining, and improving upon, in future system development.

[68] The bias in ESPreSSO T/S is the lowest of all models potentially indicating the value of the assimilating climatology to diminish errors inherited from boundary conditions or internal to the modeling system.

[69] All models exhibit a midwater minimum in summertime T/S skill suggesting that steps to improve the simulation of surface mixed layer depth variability are in order, but from the results here we cannot conclude whether this would be helped most by enhanced vertical resolution, improved air-sea buoyancy fluxes, or changes to the methods by which vertical turbulent mixing is parameterized.

[70] The modeling systems we identify as offering useful skill for the purposes of providing open boundary data to inner shelf and estuary models for real-time applications in the MAB are NCOM, ESPreSSO, NYHOPS, and MOCHA climatology for T/S, and NYHOPS, ESPreSSO, COAWST, and MERCATOR for velocity.

[71] It is somewhat unfair to appraise global models with respect to such a regionally limited set of shallow coastal observations. But most of the regional models have used global models in some form for open boundary data, so it is of interest to see how much downscaling has improved upon the global model skill through enhanced resolution, dynamics, or data assimilation. That the global models are competitive in our skill assessment suggests there is scope for further progress in downscaling methods that could deliver yet more skillful boundary data for inner shelf and estuary models.

[72] **Acknowledgments.** We thank the groups whose results we analyzed for making their research model results freely and readily accessible to the oceanographic community in real time. The majority of observational data used were acquired through the effort of scientists and technical staff at MARACOOS member institutions, funded by numerous U.S. federal and state programs. NOAA IOOS funded the Coastal and Ocean Modeling Test-bed. We thank Javier Zavala-Garay and Julia Levin for helpful discussions throughout the project, Richard Signell for assisting with software development to facilitate interoperability of the analysis tools, and Marge Innovera for assistance with the uncertainty estimates.

References

Antonov, J. I., S. Levitus, T. P. Boyer, M. E. Conkright, T. O'Brien, and C. Stephens (1998), World ocean atlas 1998, *Volume 1: Temperature of the Atlantic ocean, NOAA Atlas NESDIS 28*, 166 pp., U. S. Govern. Print. Off., Washington, D. C.

Barron, C. N., and L. F. Smedstad (2002), Global river inflow within the Navy Coastal Ocean Model, Proceedings *OCEANS '02 MTS/IEEE Conference*, 29–31 Oct. 2002, Vol. 3, pp.1472–1479, doi: 10.1109/OCEANS.2002.1191855.

Barron, C. N., A. B. Kara, P. Martin, R. C. Rhodes, and L. F. Smedstad (2006), Formulation, implementation and examination of vertical coordinate choices in the global Navy Coastal Ocean Model (NCOM), *Ocean Modell.*, *11*, 347–375, doi:310.1016/j.ocemod.2005.1001.1004.

Blayo, E., and L. Debreu (2005), Revisiting open boundary conditions from the point of view of characteristic variables, *Ocean Modell.*, *9*, 231–252.

Blumberg, A. F., and G. Mellor (1987), A description of a three-dimensional coastal ocean circulation model, in *Three-Dimensional Coastal Ocean Models, Coastal and Estuarine Sciences 4*, edited by N. Heaps, pp. 1–16, AGU, Washington, D. C.

Blumberg, A. F., L. A. Khan, and J. P. St. John (1999), Three-dimensional hydrodynamic model of New York Harbor region, *J. Hydraul. Eng.*, *125*, 799–816.

Brasseur, P., et al. (2005), Data assimilation for marine monitoring and prediction: The MERCATOR operational assimilation systems and the MERSEA developments, *Q. J. R. Meteorol. Soc.*, *131*(613), 3561–3582, doi:3510.1256/qj.3505.3142.

Brown, W. S., A. Gangopadhyay, F. L. Bub, Z. Yu, G. Strout, and A. R. Robinson (2007), An operational circulation modeling system for the Gulf of Maine/Georges Bank Region, Part 1: Basic elements, *IEEE J. Oceanic Eng.*, *32*, 807–822.

Castelao, R., S. Glenn, O. Schofield, R. Chant, J. Wilkin, and J. Kohut (2008a), Seasonal evolution of hydrographic fields in the central Middle Atlantic Bight from glider observations, *Geophys. Res. Lett.*, *35*, L03617, doi:03610.01029/02007GL032335.

Castelao, R. M., O. Schofield, S. Glenn, R. J. Chant, and J. Kohut (2008b), Cross-shelf transport of fresh water on the New Jersey Shelf, *J. Geophys. Res.*, *113*, C07017, doi:07010.01029/02007JC004241.

Castelao, R., S. Glenn, and O. Schofield (2010), Temperature, salinity and density variability in the central Middle Atlantic Bight, *J. Geophys. Res.*, *115*, C10005, doi:10010.11029/12009JC006082.

Chant, R. J., et al. (2008), Dispersal of the Hudson River Plume in the New York Bight: Synthesis of observational and numerical studies during LaTTE, *Oceanography*, *21*(4), 148–161.

Chapman, D., and G. Gawarkiewicz (1993), On the establishment of the seasonal pycnocline in the Middle Atlantic Bight, *J. Phys. Oceanogr.*, *23*(11), 2487–2492.

Chassignet, E., et al. (2006), Generalized vertical coordinates for eddy-resolving global and coastal ocean forecasts, *Oceanography*, *19*, 20–31.

Chassignet, E., et al. (2009), U.S. GODAE: Global ocean prediction with the HYbrid Coordinate Ocean Model (HYCOM), *Oceanography*, *22*(2), 64–75.

Csanady, G. (1976), Mean circulation in shallow seas, *J. Geophys. Res.*, *81*(30), 5389–5399.

Csanady, G., and P. Hamilton (1998), Circulation of slopewater, *Cont. Shelf Res.*, *8*, 565–624.

Cummings, J. (2005), Operational multivariate ocean data assimilation, *Q. J. R. Meteorol. Soc.*, *131*, 3583–3604.

Dai, A., and K. Trenberth (2002), Estimates of freshwater discharge from continents: latitudinal and seasonal variations, *J. Hydrometeorol.*, *3*, 660–687.

De Mey, P., and R. Proctor (2009), Assessing the value of GODAE products in coastal and shelf seas, *Ocean Dyn.*, *59*, 1–2.

Flagg, C., R. Houghton, and L. Pietrafesa (1994), Summertime thermocline salinity maximum intrusions in the Mid-Atlantic Bight, *Deep Sea Res. Part II*, *41*(2–3), 325–340.

Fleming, N., and J. Wilkin (2010), MOCHA: A 3-D climatology of the temperature and salinity of the Middle Atlantic Bight, *Eos Trans. AGU, 91, Ocean Sci. Meet. Suppl.*, Abstract PO35G-08.

Fox, D., C. N. Barron, M. R. Games, M. Booda, G. Peggion, and J. V. Gurley (2002), The modular ocean data analysis system, *Oceanography*, *15*, 22–28.

Gangopadhyay, A., A. R. Robinson, and H. Arango (1997), Circulation and dynamics of the western north atlantic. Part I: Multiscale feature models, *J. Atmos. Oceanic Technol.*, *14*, 1314–1332.

Gangopadhyay, A., A. R. Robinson, P. Haley, W. Leslie, C. Lozano, J. Bisagni, and Z. Yu (2003), Feature-oriented regional modeling and simulations in the Gulf of Maine and Georges Bank, *Cont. Shelf Res.*, *23*, 317–353.

Georgas, N., and A. F. Blumberg (2010), *Establishing confidence in marine forecast systems: The design and skill assessment of the New York Harbor Observation and Prediction System, version 3 (NYHOPS v3), paper presented at Eleventh International Conference in Estuarine and Coastal Modeling (ECM11)*, Am. Soc. of Civil Eng., Seattle, Wash., 4–6 Nov.

Gong, D. (2011), Mesoscale variability on the New Jersey Shelf: Effects of topography, seasons, winds, and offshore forcing on circulation, hydrography, and transport, PhD thesis, 188 pp., Rutgers Unive., New Brunswick, N. J.

Gong, D., J. Kohut, and S. Glenn (2010), Seasonal climatology of wind-driven circulation on the New Jersey Shelf, *J. Geophys. Res.*, *115*, C04006, doi:04010.01029/02009JC005520.

Grant, W., and O. Madsen (1979), Combined wave and current interaction with a rough bottom, *J. Geophys. Res.*, *84*, 1797–1808, doi:1710.1029/JC1084iC1704p01797.

Herzfeld, M. (2009), The role of numerical implementation on open boundary behaviour in limited area ocean models, *Ocean Modell.*, *27*, 18–32.

Houghton, R., R. Schlitz, R. Beardsley, B. Butman, and J. Chamberlin (1982), The Middle Atlantic Bight cold pool: Evolution of the temperature structure during summer 1979, *J. Phys. Oceanogr.*, *12*(10), 1019–1029.

Janjic, Z. (2004), The NCEP WRF core, paper presented at 20th Conference on Weather Analysis and Forecasting/16th Conference on Numerical Weather Prediction, Amer. Meteorol. Soc., Seattle, Wash. [Available at http://ams.confex.com/ams/84Annual/techprogram/paper_70036.htm, accessed 23 May 2013.].

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*, 64–82, doi:10.1016/j.jmarsys.2008.1005.1014.

Kara, A. B., C. N. Barron, P. Martin, L. F. Smedstad, and R. C. Rhodes (2006), Validation of interannual simulations from the 1/8° global Navy Coastal Ocean Model (NCOM), *Ocean Modell.*, *11*, 376–398, doi:310.1016/j.ocemod.2005.1001.1003.

Kohut, J., H. Roarty, and S. Glenn (2006), Characterizing observed environmental variability with HF Doppler radar surface current mappers

and acoustic Doppler current profilers: Environmental variability in the coastal ocean, *IEEE J. Ocean Eng.*, *31*, 876–884.

Lanerolle, L., R. C. Patchen, and F. Aikman III (2011), The second generation Chesapeake Bay Operational Forecast System (CBOFS2): Model development and skill assessment, *NOAA Tech. Rep. NOS CS 29*, *77 pp.*, Off. Coast Surv., Natl. Oceanic and Atmos. Admin., Silver Spring, Md.

Lentz, S. (2003), A climatology of salty intrusions over the continental shelf from Georges Bank to Cape Hatteras, *J. Geophys. Res.*, *108*(C10), 3326, doi:10.1029/2003JC001859.

Lentz, S. (2008), Observations and a model of the mean circulation over the Middle Atlantic Bight Continental Shelf, *J. Phys. Oceanogr.*, *38*(6), 1203.

Lentz, S., K. Shearman, S. Anderson, A. Plueddemann, and J. Edson (2003), Evolution of stratification over the New England shelf during the coastal mixing and optics study, August 1996–June 1997, *J. Geophys. Res.*, *108*(C1), 3008, doi:10.1029/2001JC001121.

Madec, G. (2012), *NEMO ocean engine*, 357 pp., Inst. Pierre-Simon Laplace (IPSL), France.

Marchesiello, P., J. McWilliams, and A. Shchepetkin (2001), Open boundary conditions for long-term integration of regional oceanic models, *Ocean Modell.*, *3*(1–2), 1–20.

Mason, E., J. Molemaker, A. Shchepetkin, F. Colas, J. McWilliams, and P. Sangrà (2010), Procedures for offline grid nesting in regional ocean models, *Ocean Modell.*, *35*, 1–15.

McEnery, J., J. Ingram, Q. Duan, T. Adams, and L. Anderson (2005), NOAA's advanced hydrologic prediction service, *Bull. Am. Meteorol. Soc.*, *86*(3), 375–385.

Moore, A. M., H. Arango, G. Broquet, B. Powell, A. T. Weaver, and J. Zavala-Garay (2011), The Regional Ocean Modeling System (ROMS) 4-dimensional variational data assimilations systems, Part I—System overview and formulation, *Prog. Oceanogr.*, *91*(34–39).

Mukai, A. Y., J. J. Westerink, R. A. Luettich, and D. Mark (2002), Eastcoast 2001, A tidal constituent database for the western North Atlantic, Gulf of Mexico and Caribbean Sea, *Tech. Rep. ERDC/CHL TR-02–24*, U.S. Army Corps of Engineers, Engineer Research and Development Center, Coastal Hydraulics Lab., Vicksburg MS. 196 pp.

NCEP (2003), The GFS atmospheric model, *NCEP Off. Note 442*, 14 pp., Environ. Model. Cent., Camp Springs Md.

Nycander, J., and K. Döös (2003), Open boundary conditions for barotropic waves, *J. Geophys. Res.*, *108*(C5), 3168, doi:3110.1029/2002JC001529.

Oddo, P., and N. Pinardi (2008), Lateral open boundary conditions for nested limited area models: A scale selective approach, *Ocean Modell.*, *20*, 134–156.

Orlanski, I. (1976), A simple boundary condition for unbounded hyperbolic flows, *J. Comput. Phys.*, *21*, 251–269.

Rhodes, R. C., et al. (2002), Navy real-time global modeling systems, *Oceanography*, *15*, 29–43.

Roarty, H., et al. (2010), Operation and application of a regional high frequency radar network in the Mid Atlantic Bight, *Mar. Technol. Soc. J.*, *44*, 133–145.

Robinson, A. R., M. Spall, and N. Pinardi (1988), Gulf stream simulations and the dynamics of ring and meander processes, *J. Phys. Oceanogr.*, *18*, 1811–1854.

Rosmond, T., J. Teixeira, M. Peng, T. Hogan, and R. Pauley (2002), Navy Operational Global Atmospheric Prediction System (NOGAPS): Forcing for ocean models, *Oceanography*, *15*, 99–108.

Schmalz, R. A. (2011), Three-dimensional hydrodynamic model developments for a Delaware River and Bay Nowcast/Forecast System, *NOAA Tech. Rep. NOS CS 28*, 179 pp., Off. Coast Surv., Natl. Oceanic and Atmos. Admin., Silver Spring, Md.

Schmidt, A., and A. Gangopadhyay (2012), An operational ocean circulation prediction system for the western North Atlantic: Hindcasting during July-September of 2006, Cont. Shelf Res., doi: http://www.sciencedirect.com.library.proxy.mbl.edu/science/article/pii/S0278434312002440.

Schwab, D., J. Bennett, P. Liu, and M. Donelan (1984), Application of a simple numerical wave prediction model to Lake Erie, *J. Geophys. Res.*, *89*, 3586–3592.

Taylor, K. (2001), Summarizing multiple aspects of model performance in a single diagram, *J. Geophys. Res*, *106*(D7), 7183–7192.

Ullman, D., and P. Cornillon (2001), Continental shelf surface thermal fronts in winter off the northeast US coast, *Cont. Shelf Res.*, *21*(11–12), 1139–1156.

Ullman, D., J. O'Donnell, J. Kohut, T. Fake, and A. Allen (2006), Trajectory prediction using HF radar surface currents: Monte Carlo simulations of prediction uncertainties, *J. Geophys. Res.*, *111*, C12005, doi:12010.11029/12006JC003715.

Vörösmarty, C., B. Fekete, and B. A. Tucker (1996), Global River Discharge Database (RivDIS v1.0), *Volumes 0 through 6, in A Contribution to IHP-V Theme 1*, Tech. Doc. in Hydrol. Ser., U. N. Educ., Sci. and Cult. Organ., Paris.

Warner, J., N. Perlin, and E. Skyllingstad (2008), Using the model coupling toolkit to couple earth system models, *Environ. Modell. Software*, *23*(10–11), 1240–1249.

Warner, J. C., B. Armstrong, R. He, and J. B. Zambon (2010), Development of a Coupled Ocean–Atmosphere–Wave–Sediment Transport (COAWST) modeling system, *Ocean Modell.*, *35*, 230–244, doi:210.1016/j.ocemod.2010.1007.1010.

Zavala-Garay, J., J. Wilkin, and J. Levin (2012), Data assimilation in coastal oceanography: IS4DVAR in the Regional Ocean Modeling System (ROMS), in *Advanced Data Assimilation for Geosciences*, edited by E. Blayo and M. Bocquet, Oxford Univ. Press, Oxford, U. K.

Zhang, W. G., J. L. Wilkin, and R. C. Chant (2009), Modeling the pathways and mean dynamics of river plume dispersal in New York Bight, *J. Phys. Oceanogr.*, *39*, 1167–1183.

Zhang, W., G. Gawarkieiwicz, D. McGillicuddy Jr., and J. Wilkin (2011), Climatological mean circulation at the New England shelf break, *J. Phys. Oceanogr.*, *41*, 1874–1893.