Journal of Marine Systems 173 (2017) 21–30

Contents lists available at ScienceDirect

Journal of Marine Systems

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

Assessing the quality of bottom water temperatures from the Finite-Volume Community Ocean Model (FVCOM) in the Northwest Atlantic Shelf region

Bai Li *[,1], Kisei R. Tanaka [1], Yong Chen, Damian C. Brady, Andrew C. Thomas

School of Marine Science, University of Maine, Orono, ME 04469, USA

ARTICLE INFO

Article history:
Received 2 November 2016
Accepted 2 April 2017
Available online 12 April 2017

Keywords:
Ocean circulation models
Skill assessment
Bottom temperature
Finite-Volume Community Ocean Model
Environmental Monitors on Lobster Traps (eMOLT)
Northwest Atlantic shelf

ABSTRACT

The Finite-Volume Community Ocean Model (FVCOM) is an advanced coastal circulation model widely utilized for its ability to simulate spatially and temporally evolving three-dimensional geophysical conditions of complex and dynamic coastal regions. While a body of literature evaluates model skill in surface fields, independent studies validating model skill in bottom fields over large spatial and temporal scales are scarce because these fields cannot be remotely sensed. In this study, an evaluation of FVCOM skill in modeling bottom water temperature was conducted by comparison to hourly *in situ* observed bottom temperatures recorded by the Environmental Monitors on Lobster Traps (eMOLT), a program that attached thermistors to commercial lobster traps from 2001 to 2013. Over 2×10^6 pairs of FVCOM-eMOLT records were evaluated by a series of statistical measures to quantify accuracy and precision of the modeled data across the Northwest Atlantic Shelf region. The overall comparison between modeled and observed data indicates reliable skill of FVCOM ($r^2 = 0.72$; root mean squared error = 2.28 °C). Seasonally, the average absolute errors show higher model skill in spring, fall and winter than summer. We speculate that this is due to the increased difficulty of modeling high frequency variability in the exact position of the thermocline and frontal zones. The spatial patterns of the residuals suggest that there is improved similarity between modeled and observed data at higher latitudes. We speculate that this is due to increased tidal mixing at higher latitudes in our study area that reduces stratification in winter, allowing improved model accuracy. Modeled bottom water temperatures around Cape Cod, the continental shelf edges, and at one location at the entrance to Penobscot Bay were characterized by relatively high errors. Constraints for future uses of FVCOM bottom water temperature are provided based on the uncertainties in temporal-spatial patterns. This study is novel as it is the first skill assessment of a regional ocean circulation model in bottom fields at high spatial and temporal scales in the Northwest Atlantic Shelf region.

1. Introduction

Quantitative ocean circulation models have been widely used by the scientific community to capture past, present and future climate-driven oceanographic profiles (Blumberg and Mellor, 1987; Stock et al., 2011). In general, ocean circulation models apply fundamental physical laws to numerically discretize model dynamics in time and three-dimensional space. While spatial and temporal resolutions of model outputs have improved considerably, commensurate with increases in computer processing capabilities, modeling uncertainties inherently emerge from processes of model development and parameterization and complexity of processes at different scales (Murphy et al., 1998).

The Finite Volume Community Ocean Model (FVCOM) is an ocean circulation model developed by collaborative efforts between the University of Massachusetts-Dartmouth and the Woods Hole Oceanographic Institution (Chen et al., 2006). FVCOM is one of the core models of the Northeast Coastal Ocean Forecast System (NECOFS), and widely used to investigate interannual variability of water properties, circulation, and geophysical conditions from global to estuarine scales (Chen et al., 2006). The model's unstructured-grid feature makes FVCOM well suited for examining the interannual variability in oceanographic and circulation patterns of inshore areas that are often characterized by complex coastlines and bathymetry. FVCOM has been configured to hindcast, nowcast, and forecast key ecosystem processes on the Northwest Atlantic Shelf (NAS; hereafter referred to as FVCOM-NAS) that include major Large Marine Ecosystems (LMEs) such as Scotian Shelf (SS), Gulf of Maine (GoM), Georges Bank (GB), and Mid Atlantic Bight (MAB) (Townsend et al., 2006). As of 2016, the FVCOM-NAS has been integrated for the time period 1 January 1978 to 31 December 2013, providing

* Corresponding author.
E-mail address: bai.li@maine.edu (B. Li).
[1] Both authors contributed equally to this work.

http://dx.doi.org/10.1016/j.jmarsys.2017.04.001
0924-7963/

hourly currents and hydrography information in the NAS region (Chen, 2015).

An increasing number of coupled biophysical models rely on FVCOM outputs to examine the impact of climate change and interannual variability on oceanographic conditions (e.g. Huret et al., 2007; Ji et al., 2008; Tanaka and Chen, 2015, 2016) and to facilitate the decision-making process for the management of marine resources in the NAS region. Recently, several bioclimate modeling efforts have been made for commercially important benthic and groundfish species such as scallops, cod, and American lobster (Ji et al., 2008; Li et al., 2015a; Tanaka and Chen, 2015, 2016). These biological and fisheries models are being developed downstream of NECOFS due to its ability to simulate bottom water constituents out of the view of remote sensing technology. However, rigorous skill assessment of FVCOM outputs for key benthic water properties such as bottom water temperature and salinity remain scarce, which may result in lack of confidence in coupled biophysical modeling studies based on the FVCOM outputs. Such skill assessment requires comprehensive comparison with observed data collected within the time and space domain of model integration.

In this study, using in situ hourly bottom temperature data collected throughout the NAS region, we compared observed and modeled bottom water temperatures to assess the accuracy and reliability of FVCOM. To our knowledge, this study provides the first systematic examination of the quality of modeled bottom water temperature from FVCOM-NAS over extensive temporal and spatial scales and over a large geographic area characterized by steep thermal gradients. Our results provide critical information on the quality of FVCOM outputs, identify potential issues associated with the FVCOM-predicted thermal fields, and potentially lead to better use of the data.

2. Material and methods

2.1. Study area

The NAS (Fig. 1) is a broad region extending >200 km offshore and supports some of the most productive fisheries in the world (Townsend et al., 2006; Fernandez et al., 2015). The coastal and shelf waters throughout the NAS region are strongly influenced by the large-scale circulation of the Northwest Atlantic Ocean (Loder et al., 1998), and characterized by a steep latitudinal temperature gradient (Townsend et al., 2006). The water properties in the region are also influenced by the supply of fresh water from Arctic outflow

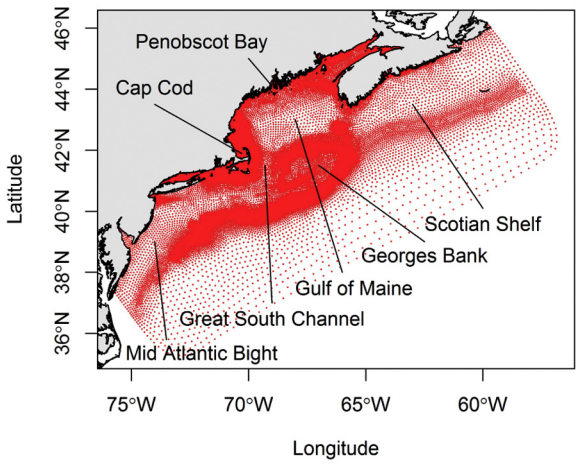


Fig. 1. Spatial domain and node locations (red dot) of the Finite-Volume Community Ocean Model (FVCOM) configured in the Northwest Atlantic Shelf (NAS) region. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

(Greene and Pershing, 2007). The region receives relatively fresher water from the southwesterly Labrador Current, the St. Lawrence River, and local smaller rivers (Townsend et al., 2006). The northward-flowing Gulf Stream brings warm and relatively saltier waters into the region (Xue et al., 2008). The physical oceanography of the Northwest Atlantic continental shelf is regulated by dynamics of the North Atlantic sub-polar and sub-tropical gyres, and the major current systems such as the Gulf Stream, Labrador Current, and adjoining shelf and slope water currents (Townsend et al., 2006; Townsend et al., 2015). Both Gulf Stream and Labrador Current systems are key components of the North Atlantic climate system, and their interannual and inter-decadal variability are linked to the North Atlantic Oscillation (NAO) (Hurrell, 1995). Bottom water properties in the NAS vary under the competing influence of warm and salty Warm Slope Water (8–12 °C and 34.7–35.5) that originates in North Atlantic Central Water, and cold and relatively fresh Labrador Slope Water (4–8 °C and 34.3–35) that flows southwest (Townsend et al., 2006). The two slope waters are mixed along the shelf break under NAO influence and play a vital role in determining the characteristics of the deep and bottom water properties in the region (Greene et al., 2013).

2.2. FVCOM-NAS

FVCOM-NAS is a three-dimensional, unstructured, free surface, primitive equation model, which solves the governing momentum and thermodynamic equation through a second order finite-volume flux discrete scheme (Chen et al., 2006). Its computational capability ensures mass conservation on the individual control volumes and entire computational domain (Chen et al., 2006). The FVCOM employs a non-overlapping, unstructured triangular grid that incorporates the advantages of finite-element methods for geometric flexibility and finite-difference methods for computational efficiency (Chen et al., 2006). This makes the model well suited to simulating geophysical marine environments characterized by a complex and irregular coast (Chen et al., 2006).

While the FVCOM-NAS has been configured with three generations of model grids (G1–3: Chen, 2015), the skill assessment in this study was conducted with the G3 model grids with 48,450 nodes (Chen, 2015; Fig. 1). The unstructured FVCOM-G3 grid provides horizontal resolution ranging from as fine as ~20 m inshore to as coarse as ~10 km at the open boundary off the continental shelf (Chen et al., 2006). The FVCOM allows interpolation of missing oceanographic data (e.g. temperature and salinity data) at various temporal and spatial scales, which is one of its useful features. The FVCOM-NAS domain contains GoM, SS north to 45.2°N, and the MAB south to 39.1°N (Cowles et al., 2008; Fig. 1). Hourly bottom temperature data are modeled at 48,451 nodes. The FVCOM-NAS applies vertical grid discretization using a total of 45 terrain-following sigma layers (Chen et al., 2011). The bottom boundary layer has a variable thickness depending on given bathymetry (Chen et al., 2011). The FVCOM-NAS uses the USGS 15-arcsec digital bathymetry data set (Roworth and Signell, 1998) to determine water depth at each node with a minimum depth of 3 m in shallow coastal waters. The FVCOM-NAS incorporates assimilation of high-resolution satellite-derived sea surface temperature data and salinity observations at all surface nodes, as well as temperature fields on the open boundary (Cowles et al., 2008). The FVCOM-NAS began assimilation of the eMOLT-derived bottom temperature in 2008 (Manning et al., in review).

2.3. Environmental Monitors on Lobster Traps (eMOLT)

A collection of observed bottom temperatures provided by the eMOLT program served as the observational data set. The eMOLT program began in 2001 through an interdisciplinary-collaborative effort to monitor the physical environment of the GoM and the

Southern New England shelf (Manning and Pelletier, 2009). Using internally recording temperature probes attached to lobster traps, the eMOLT provides observed hourly bottom temperature data from 2001 to 2013 at 201 sites in the NAS region (Fig. 2). The spatial coverage of these study sites varied over time. The depth of these study sites varies from 0.2 to 356.6 m. The primary temperature probes are ONSET TidbiT Water Temperature Data Logger and Minilog II-T Temperature Data Logger with ±0.2 °C accuracy (Manning and Pelletier, 2009). The eMOLT provides an ideal array of bottom water property observations for the initialization, assimilation, and validation demands of ocean circulation models in the NAS region.

2.4. Pairing modeled and observed bottom temperatures

Modeled bottom water temperatures at hourly temporal resolution served as the subject of skill assessment in this study. For comparisons between modeled (FVCOM-NAS) and observed (eMOLT) bottom temperatures, modeled quantities were matched to in situ observations at hourly resolution at each eMOLT instrument location (Fig. 3). The observations with bottom water temperature <0 °C were excluded from this analysis. First, a hypothetical square with a side 0.01 decimal degree in length was simulated around each eMOLT instrument location to identify nearby FVCOM nodes. Second, an arithmetic mean of modeled quantities within the hypothetical square was paired with corresponding eMOLT observation at 1-h temporal resolution. A total of 2,124,867 pairs of FVCOM-eMOLT quantities were identified using this approach. The coupled FVCOM-eMOLT quantities were obtained throughout the FVCOM-NAS domain (Depth 0.2–356.6 m; Latitude 39.55–44.81°N; Longitude 73.05–63.86°W) over the period 2001–2013, allowing us to assess the skill of FVCOM at various temporal and spatial scales.

2.5. Statistical and quantitative measures for FVCOM-NAS bottom water temperature skill assessment

The literature suggests that a number of statistical and quantitative measures can be used to assess the skill of modeled estimates (Fitzpatrick, 2009; Stow et al., 2009).

2.5.1. Univariate comparison of predictions and observations

The following six statistical measures were used for the pair-wise comparison of modeled and observed data (Stow et al., 2009).

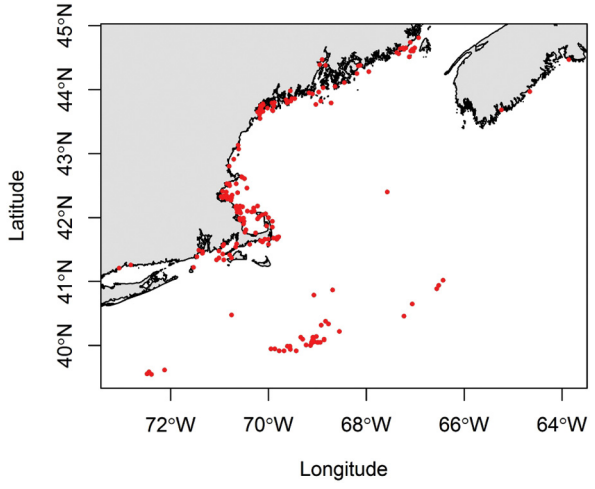


Fig. 2. Locations of the Environmental Monitoring on Lobster Traps (eMOLT) sites (n = 201) in the U.S. Northeast Continental Shelf and Nova Scotia, Canada.

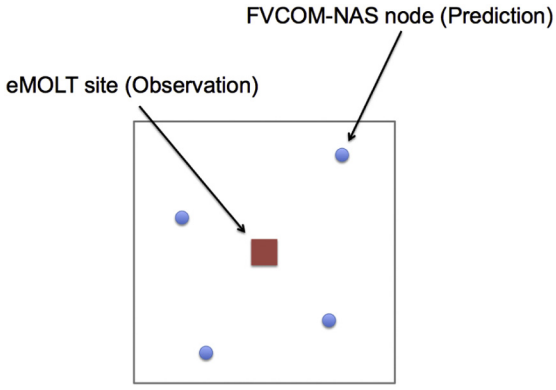


Fig. 3. A schematic of the process used to match predicted and observed data. A hypothetical square with sides of 0.01 decimal degrees in length was simulated around an Environmental Monitoring on Lobster Traps (eMOLT) site to identify nearby Finite Volume Community Ocean Model (FVCOM) nodes.

Correlation coefficient (r):

$$r = \frac{\sum_{i=1}^{n}(E_i-\overline{E})(F_i-\overline{F})}{\sqrt{\sum_{i=1}^{n}(E_i-\overline{E})^2\sum_{i=1}^{n}(F_i-\overline{F})^2}} \quad (1)$$

Average relative error (ARE):

$$ARE = \frac{\sum_{i=1}^{n}(F_i-E_i)}{n} \quad (2)$$

Average absolute error (AAE):

$$AAE = \frac{\sum_{i=1}^{n}|F_i-E_i|}{n} \quad (3)$$

Root mean square error (RMSE):

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(F_i-E_i)^2}{n}} \quad (4)$$

Reliability index (RI):

$$RI = exp\sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(log\frac{E_i}{F_i}\right)^2} \quad (5)$$

Modeling efficiency (ME):

$$ME = \frac{\sum_{i=1}^{n}(E_i-\overline{E})^2-\sum_{i=1}^{n}(F_i-\overline{F})^2}{\sum_{i=1}^{n}(E_i-\overline{E})^2} \quad (6)$$

where n is the number of data-model pairs (n = 2,124,867); E_i is the ith eMOLT observation; $\overline{E}$ represents the average of the observations; F_i represents the ith FVCOM prediction; and $\overline{F}$ is the average of the predictions.

The correlation coefficient (r) of Eq. (1) measures the magnitude of correlation and dependency between the modeled and observed data (Stow et al., 2009). The correlation coefficient can vary from −1 to 1, with negative values indicating an inverse relationship between the observed and predicted values and values close to 1 indicating excellent agreement.

The average relative error (ARE) of Eq. (2) measures the possible overall bias of FVCOM-modeled data, and average absolute error (AAE) and root mean square error (RMSE) of Eqs. (3) and (4) quantify the overall spatio-temporal bias and variability between predictions

Table 1
Summary of quantitative metrics statistics.

Correlation coefficient (r)	0.85
Average relative error (ARE)	0.04
Average absolute error (ABE)	1.56
Root mean square error (RMSE)	2.28
Reliability index (RI)	1.08
Modeling efficiency (ME)	0.71

and observations (Fitzpatrick, 2009). Values close to zero for these indices indicate a close match between modeled and observed data.

The reliability index (RI) quantifies the magnitude of differences between modeled and observed values in terms of average factors (Leggett and Williams, 1981). An RI of 2 indicates a model predicts the corresponding observation within a multiplicative factor of 2 on average, while RI closer to 1 indicates a better prediction (Stow et al., 2009).

The modeling efficiency (ME) quantifies the accuracy of model prediction relative to the average of the observation (Loague and Green, 1991; Nash and Sutcliffe, 1970). A negative ME or an ME close to zero indicates that the average of observations is a better predictor than an individual predictor. A ME near 1 implies close match between prediction and observation.

We also compared predictions and observations by site. We use the RMSE of bottom water temperature at each site to provide a spatial picture of reliability of FVCOM-NAS-modeled bottom water temperature.

2.5.2. Linear regression analysis

A regression analysis is a common statistical approach to complement the bivariate observed versus predicted comparison plot. A set of linear regression coefficients such as the coefficient of determination (r^2), slope (α), and intercept (β) obtained by minimizing the sum of the squares of the differences between modeled and the observed data can produce several criteria. Estimating α and β can assess how the bias may change within the observed range (e.g., the magnitude that α differs from 1) and a potential bias inherent in the predictions (e.g., changes in discrepancy between β and 0). The r^2, similar to r in Eq. (1), can measure the goodness-of-fit, which can account for the amount of the variance in the observed data that is explained by the model (Fitzpatrick, 2009). The following linear regression model was used to examine the bivariate observed versus predicted plot:

$$E = \alpha F + \beta \tag{7}$$

where E and F are observations and predictions respectively; α and β represent the slope coefficient and intercept of the regression model.

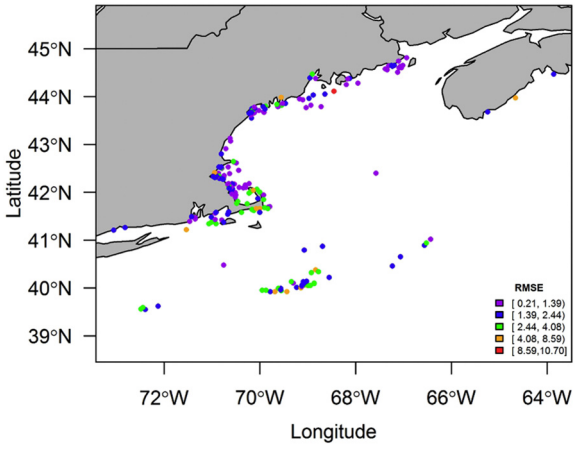


Fig. 4. Root mean square error (RMSE) map based on Environmental Monitoring on Lobster Traps (eMOLT) sites ($n = 201$).

An α of 1, β of 0, and r^2 of 1 imply that FVCOM-NAS has unbiased model skill.

2.5.3. Taylor diagram

Taylor diagrams have become a common tool in evaluating multiple aspects of the skill of different models (Wu et al., 2013; Miao et al., 2014). They provide a concise graphical summary of how closely patterns match each other in terms of correlation, centered RMSE, and standard deviation that represent the magnitude of their variations (Taylor, 2005). The similarity between modeled and observed bottom water temperature was summarized using standardized Taylor diagrams for four temporal and spatial criteria; month (January–December), year (2001 — 2013), depth (0.2–356.6 m), and distance offshore (0.0–2.9 decimal degree). The distance offshore was calculated as the shortest distance to the coastline from each eMOLT site. Fisher's natural breaks classification can divide a sequence of numeric values into multiple classes such that the sum of the squared deviations from the class means is minimal (Bivand et al., 2013), thus the classification method was used to divide the depth and distance offshore each into 10 classes by minimizing the intra-class variance and maximizing the inter-class variance (Bivand, 2013).

2.5.4. Time series analyses

The temporal similarity between the modeled and observed bottom water temperature time sequence was analyzed. A dynamic time warping (DTW) was used to compute the optimal alignment and subsequent cumulative distance between monthly averages of modeled and observed temperature time sequences. DTW is a time series comparison that can account for differences in time shift and scaling, and allows a non-linear comparison of two temporal sequences that may vary in frequency (Gu and Jin, 2006). Through the minimization of the distance between the modeled data and observed data, DTW can derive the least cumulative distance of the alignments between the two time sequences (Gu and Jin, 2006). The dtw R package was used to implement DTW analysis (Giorgino, 2009).

Modeled and observed time sequences of monthly bottom temperature anomalies were also compared to assess how well the FVCOM-NAS captures the "true" seasonal trend. Monthly climatological values were calculated by producing mean bottom water temperature in each calendar month between 2001 and 2013. Anomalies were calculated by

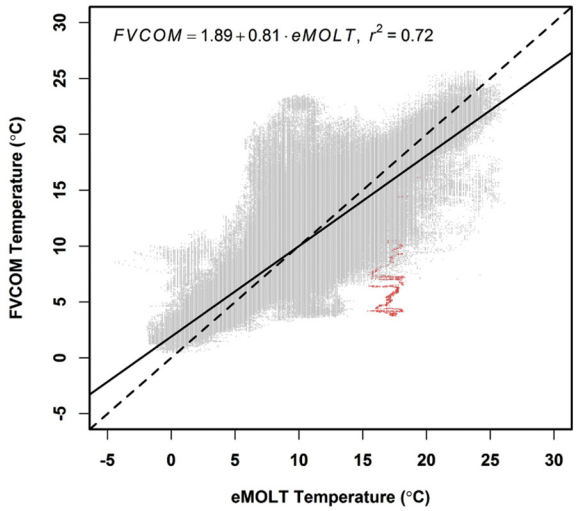


Fig. 5. The modeled (Finite Volume Community Ocean Model: FVCOM) versus observed (Environmental Monitoring on Lobster Traps: eMOLT) values. The linear regression for the model versus predicted value is plotted as a solid line. The dash line represents 1:1 line. The red points represent the potential non-random instrumental biases from WH02. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

subtracting climatological means from the monthly mean observations. A positive anomaly indicates that the temperature is warmer than the monthly average, while a negative anomaly indicates that the temperature is colder than the monthly average.

2.5.5. Non-parametric regression analysis

A generalized additive modeling (GAM) approach was implemented to evaluate whether the model skill of FVCOM-NAS exhibits any systematic biases at specific spatial and temporal scales. GAMs blend properties of generalized linear models and apply a flexible and automated approach to capture the nonlinear relationships between predictors and response variables (Hastie and Tibshirani, 1990). Using a GAM approach, the absolute error between prediction and observation was modeled as function of latitude (decimal degree), longitude (decimal degree), depth (m), distance offshore (decimal degree), year, and month. The GAM formulation with the six candidate predictor variables can be written;

$$g(E(y)) = \alpha + \sum_{i=1}^{p} f(x_i) + \varepsilon \quad (8)$$

where $g()$ denotes the link function that relates the expected value of y, $E(y)$, to the predictors; α denotes the intercept term; f denotes the non-parametric cubic spline smooth function; x_i denotes the i^{th} predictor variable; and ε is the residual error term. A GAM with a Gaussian family (identity link) was fitted and the proportion of deviance explained was used to measure how well the GAMs can explain the variance in the data. The *mgcv* R package was used to implement GAM analysis (Wood, 2011).

3. Results

3.1. Skill assessment metrics and root mean square error (RMSE) map

The selected metrics for FVCOM-NAS skill performance revealed a high correlation ($r = 0.87$) and a small overall difference between modeled and observed data (ARE = 0.04, AAE = 1.56, RMSE = 2.28; Table 1). The computed values of RI (1.08) and ME (0.71) suggested that overall, the FVCOM-NAS performed very well (Table 1). The RMSE at each eMOLT site varied from 0.2 to 10.7 °C. Clusters of large RMSEs (RMSE > 4) were observed in the area around Cape Cod and at the edge of continental shelf. The largest RMSE value was observed in eastern Penobscot Bay (Fig. 4).

3.2. Regression analysis

The regression analysis showed a significant positive relationship between observed and modeled data, while the slope (α) and intercept (β) coefficients were significantly different from 1 and 0, respectively ($\alpha = 0.81, \beta = 1.89, p < 0.05$) (Fig. 5). The regression slope indicated that FVCOM-NAS overestimated low bottom water temperature and underestimated high bottom temperatures. A series of strong cold FVCOM outliers was observed at eMOLT temperatures between 15 and 20 °C.

3.3. Taylor diagrams

Taylor diagrams summarized the variance, correlation coefficient, and RMSE for two spatial (depth and distance offshore) and two temporal (month and year) factors (Fig. 6). The Taylor diagram for month showed that the model skill in March yielded the highest correlation

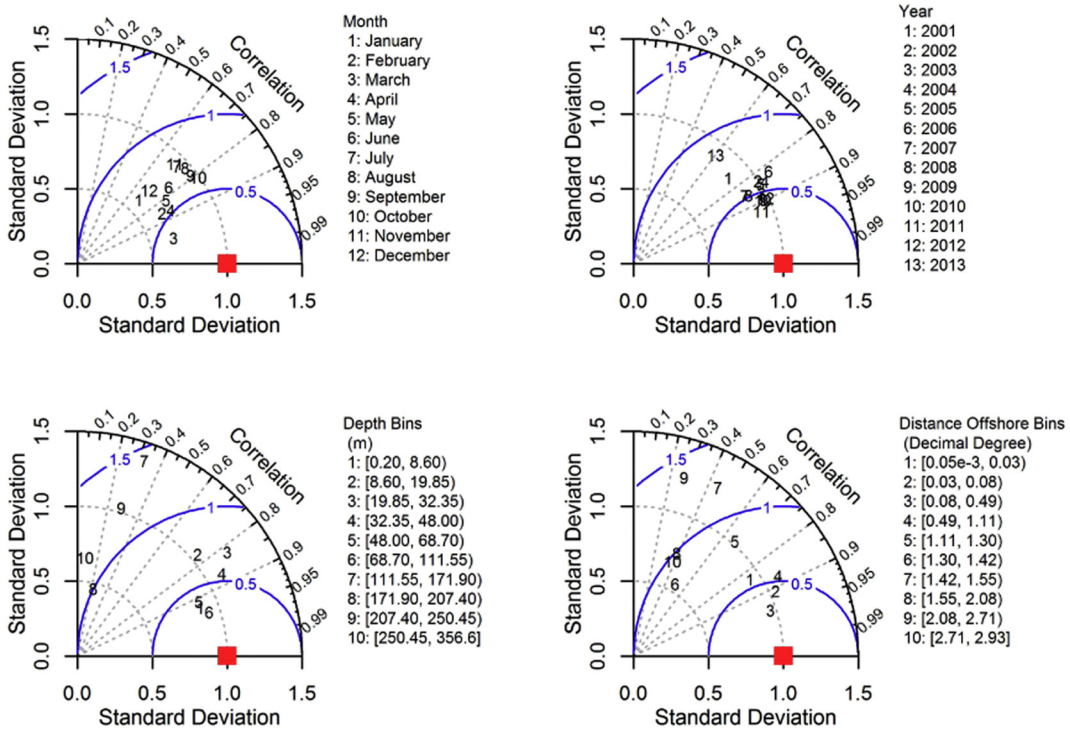


Fig. 6. Standardized Taylor diagrams for month (top left: January–December), year (top right: 2001–2013), depth bins (bottom left: 0.2–356.6 m), and distance offshore bins (bottom right: 0.00005–2.93 decimal degree). Total 2,220,402 modeled-observed match-ups were used to construct each Taylor diagram. The red square on the x-axis represents the standardized observation. The numbers on each diagram represent corresponding temporal or spatial bins of modeled data. The position of each number indicates the correlation (angular distance from the x-axis in gray), root mean square error (radial distance from the standardized observation in blue), and standard deviation (radial distance from the origin in gray) of predicted temperatures comparing with standardized observations. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

coefficient and lowest RMSE, while the model skill in July–August and November–January was characterized by relatively low correlation coefficient and high RMSE. However, the model skill in July–August captured variability similar to the observed data compared to the model skill in November–January. Overall, the model values across all the months showed less variability than the observed data.

The Taylor diagram for year indicated that the model skill in 2001 and 2013 yielded lower correlation coefficients and higher RMSEs compared to other years. The modeled data and observed data showed relatively equal variability.

The Taylor diagram for depth showed that the model skill in depths <111.55 m yielded relatively high correlations and low RMSE, while the model skill in deeper areas indicated lower correlation coefficients and higher RMSE.

The Taylor diagram for distance offshore showed that the model skill in nearshore areas (e.g. <1.11 decimal degree) was characterized with relatively high correlation coefficient and low RMSE. Furthermore, the model skill in nearshore areas showed less variability compared to the observed data.

3.4. Time series analyses

The DTW comparison of two temporal sequences revealed a strong similarity between the modeled and observed temporal sequences (Fig. 7). The least cumulative distance curve between two sequences closely followed the diagonal line, indicating a close match between the modeled and observed temporal sequences with similar variability. The modeled and observed monthly temperature anomalies showed similar temporal trends (Fig. 8). The FVCOM-NAS underestimated the bottom water temperature in the warmest months from 2005 to 2008, while the model overestimated the bottom water temperature after 2012. Most of FVCOM-NAS anomalies showed a tendency to underestimate during the colder months.

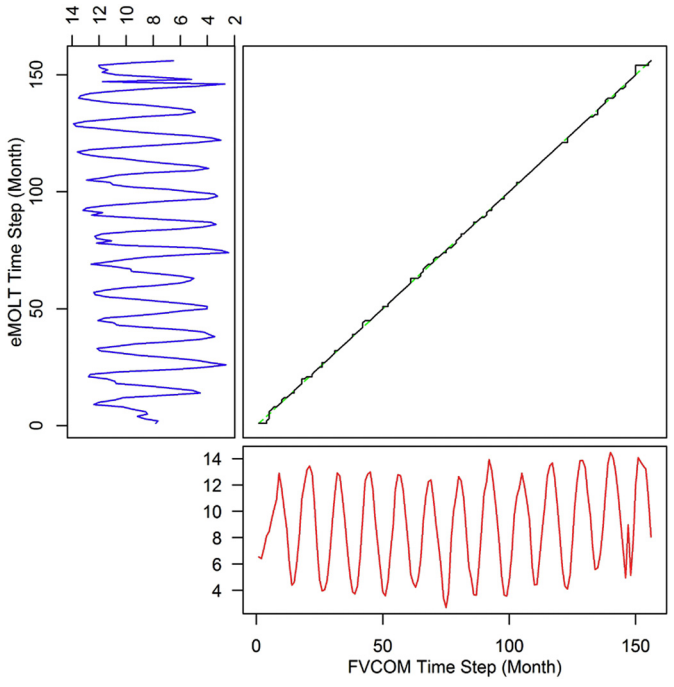


Fig. 7. A dynamic time warping (DTW) plot showing the least cumulative distance between observed (eMOLT: blue line) versus modeled (FVCOM: red line) time signals. The two time signals in the upper left and lower right panels represent monthly averaged observed and modeled bottom temperature versus monthly time steps. The x-axis and y-axis for upper left panel and lower right panel represents the time step (month), and bottom water temperature (°C) respectively. The black solid line represents the optimal alignments between the FVCOM and eMOLT time signals. A least cumulative distance curve that is close to the diagonal line (green dash line) indicates a close match between the two time signals. A cumulative distance curve under and over the diagonal line indicates that the lags between two time signals. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

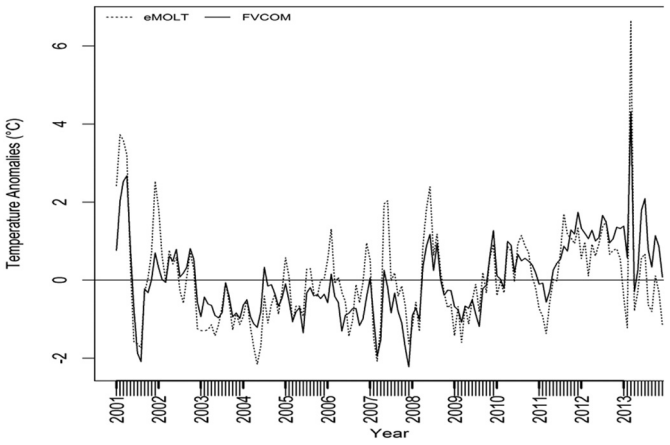


Fig. 8. Temperature anomalies of modeled (FVCOM: solid line) and observed (eMOLT: dash line) data from 2001 to 2013. Both trend lines represent monthly averages.

3.5. Non-parametric regression analysis

The GAM results revealed that the model skill of FVCOM-NAS exhibited some biases in both space and time (Fig. 9). All six candidate predictor variables were found to be statistically significant ($p < 0.001$), and the GAM accounted for 36.3% of the total deviance in the original data. Among years, the lowest and highest absolute errors were found in 2009 and 2013 respectively, while among months, November and July were associated with the lowest and highest absolute errors respectively. Absolute error decreased with depth, relatively gradually from 0 to 150 m, and then more strongly from 250 to 350 m. With distance from shore, absolute error was highest close to shore, decreased to a minimum ~1.0° from shore, and then increased again with continuing distance from shore. The absolute error decreased from low to high latitudes, most strongly in the southern portion of the study area (40–41°N), and increased from west to east, most strongly from the western boundary to ~69°W.

4. Discussion

4.1. Observational uncertainty

The skill assessments presented in this study inherently assume that the eMOLT data represents the "truth" (i.e. all temperature probes functioned properly throughout the study area during 2001–2013). However, instances of high variability associated with observed data as well as obvious outliers indicates potential non-random instrumental biases (Fig. 5). We identified 3 eMOLT sites that accounted for >90% of eMOLT – FVCOM-NAS discrepancies larger than 10 °C ($n = 8684$).

At both sites AB01 ($n = 4103$; 42.04°N; 70.13°W; −18.3 m) and DB01 ($n = 2532$, 44.11°N; 68.45°W; −16.5 m), the majority of the discrepancies larger than 10 °C occurred during summer months (June through September). AB01, located in shallow waters inside the Cape Cod Bay and operated from September 2001 through August 2013, recorded median observed and modeled temperatures of 8.1 °C and 19 °C respectively. AB01 has the most temperature variability among the all the eMOLT sites (Manning, personal communication), and constantly showed ~8–10 °C variations within a single tidal cycle (Appendix A). Capturing such large, high frequency natural temperature fluctuations was not what FVCOM-

NAS was designed to simulate. Similarly, FVCOM-NAS consistently overestimated bottom temperature at DB01, located at the east side of the entrance to Penobscot Bay and operated between August–October 2004, with median observed and modeled temperatures of 9.9 °C and 22.2 °C respectively (Appendix B). With DB01, we could not determine whether modeled or observed bottom temperature reflected the truth more accurately. However, it should be noted that observed temperatures below 10 °C would be unusual in June–September in this region. In this way, model residuals are useful diagnostic tools for analyzing data and not simply to validate models.

Another seasonal discrepancy larger than 10 °C occurred at WH02 ($n = 1429$; 41.54°N; 70.67°W; −1.8 m) in December and January with the median observed and modeled temperatures of 17.3 °C and 5.1 °C respectively. WH02 is a very shallow site (~1 m) operated between December 2001 through October 2002, where the eMOLT probe is potentially affected by exposure and ice during the winter months (Appendix C). It is likely that the difference between observed and modeled depth coupled with changing sea level due to tides contributed to some of those large discrepancies at WH02. Furthermore, the FVCOM-NAS skill is likely to be limited at such depths, as the model was not designed to resolve such shallow regions.

Given the regional oceanographic conditions surrounding these sites, we consider some of the in situ data with skepticism. However, these data that accounted for <5% of total sample size, remained in our final analysis as we could not ascertain whether the temperature probes malfunctioned or whether very localized processes caused these sites to stray so strongly from seasonal norms (Manning, personal communication).

Finally, there were a number of undocumented relocations of eMOLT temperature probes that likely affected the quality of observed data (Manning and Pelletier, 2009). While such observational uncertainty cannot be quantified or ignored, it is important to acknowledge that relatively low model skill observed in particular place or time could be due to potential systematic monitoring inaccuracies. Further skill assessment of bottom temperatures should aim to incorporate potential measurement uncertainties.

4.2. Spatiotemporal variations of FVCOM-NAS model skill

While the overall skill assessment metrics showed a strong correlation between modeled and observed data, the magnitude of model skill varied over both space and time scales. While the highest RMSE (10.7 °C) was found in eastern Penobscot Bay (Fig.4), the spatial RMSE pattern showed higher error in lower latitudes such as the area around Cape Cod around 41–42°N and at the edge of continental shelf around 40°N (Fig. 4). The GAM analysis and Taylor diagrams, which provided a more holistic view of spatial variation in model skill, identified lower model skill in shallower, inshore waters towards lower latitude (Figs. 6 and 9). Increased tidal mixing at higher latitude, which reduces stratification in winter, would have likely resulted in improved model accuracy. Overall, the complex tidally mixed coastal current water properties, highly variable tidal range, coupled with influxes of fresher waters from rivers (e.g. Penobscot Bay), and abrupt depth changes at

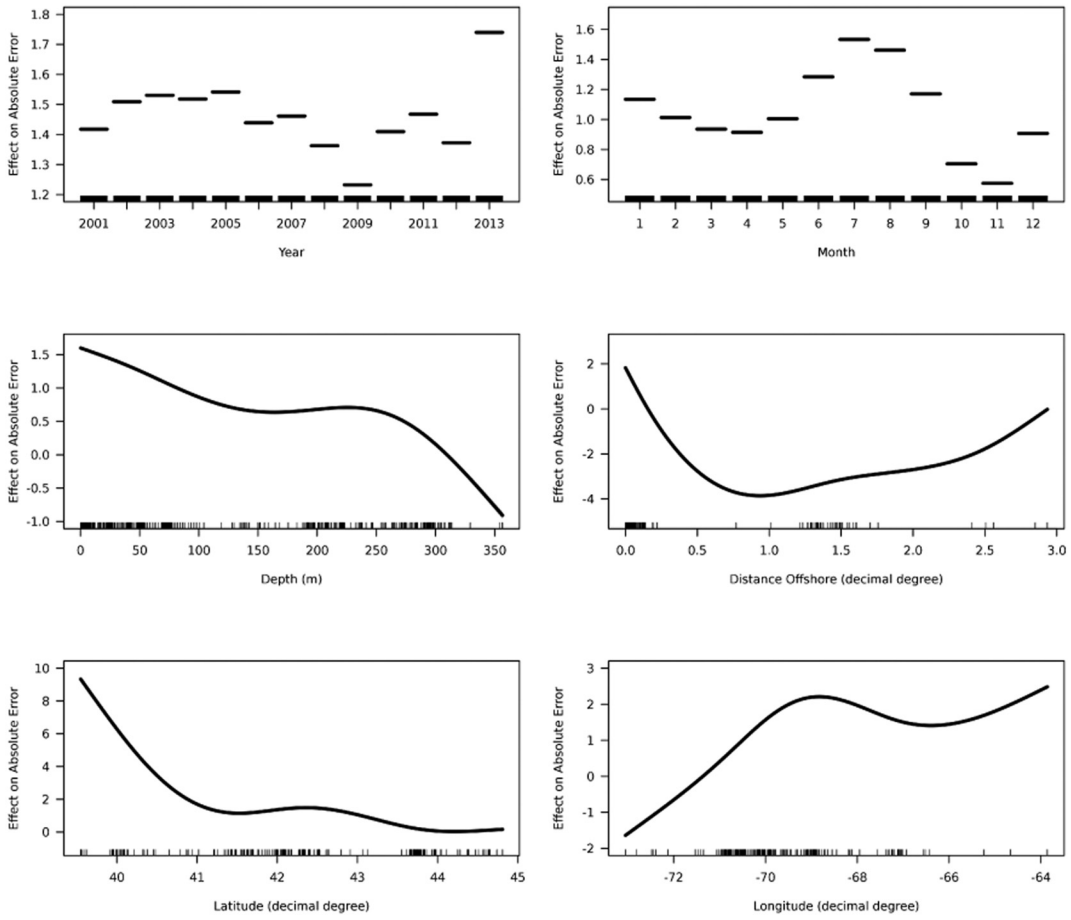


Fig. 9. Partial effects of spatial and temporal variables in the generalized additive model (GAM) of absolute error. The solid line represents the smoothed model fit. The 95% confidence interval is shown as a gray envelope, but is mostly within the width of the solid line. Tick marks on the x-axis indicate where samples occurred. Note that the scale of the y-axis differs from one panel to the next for display purposes.

the shelf edge have all likely contributed to the systematic spatial variation in the model skill (Figs. 4, 6, and 9). However, while the FVCOM-NAS has a sufficient coverage of the NAS region, the majority of the eMOLT sites were located in inshore waters. Consequently, we acknowledge that the skill assessment in offshore waters reflects sparser data and is thus less confident.

Both the Taylor diagrams and GAM analysis show that model skill was higher during fall, winter and spring (February–May, and October) and lower during summer (June–September) (Figs. 6 and 9). During summer, the study region develops strong vertical stratification, and many regions (e.g., Scotian Shelf, GoM, and Georges Bank) develop strong and persistent thermal fronts (Townsend et al., 2006). Lower model skill associated with summer months likely reflects increased difficulty in tracking the exact position of the thermocline and frontal zones, both of which are modulated by tides, internal waves and other high frequency events that may not be captured fully by the model. Using a data-assimilative high-resolution re-analysis database may improve FVCOM-NAS skill in capturing spatial-temporal patterns of stratification (Li et al., 2015b). While boundary conditions are also the sources of uncertainty in the regional circulation models, FVCOM-NAS has improved its existing boundary conditions through an open boundary configuration (Cowles et al., 2008). Future skill assessments for regional circulation models may incorporate a stochastic simulation analysis to assess the effect of the boundary conditions.

4.3. Research and management implications

Collection of bottom oceanographic data is often limited by significant logistical hurdles (e.g., cost, resources, diverse bathymetry, large spatial ranges). Scientists and resource managers, however, often require high spatial and temporal resolution data and low-cost tools to meet their research and management objectives. To this end, the FVCOM-NAS has been configured to provide geophysical properties of the NAS system, including bottom properties. The FVCOM-NAS was originally commissioned to support the Massachusetts State Ocean Plan (MEEA, 2014). Its bottom temperature estimates have been incorporated in many studies that examine the impact of climatic variability on economically important demersal and benthic species to facilitate decision-making for the management of marine resources in the NAS region (e.g. Li et al., 2015a; Tanaka and Chen, 2015, 2016). While advanced regional circulation models are increasingly used to inform effective ecosystem-based management in a highly complex and variable ocean environment (Stock et al., 2011), quantifying spatiotemporal variability of model skill is critical to identify the possible consequences of using the data indiscriminately. The spatiotemporal skill assessment results presented in this study provide guidance to stakeholders on how FVCOM-NAS bottom temperature outputs are best used by highlighting when and where the model skill is most reliable, or when/where they should be handled with caution. For example, bottom phenomena related to large scale oceanographic variability over seasonal cycles and interannual scales can quite reasonably be analyzed using FVCOM-NAS, while localized high-frequency dynamics in vertically stratified summer conditions may be less skillfully handled. Finally, the multiple spatiotemporal skill assessment criteria presented in this study can be used as priori criteria before the use of any global and regional ocean circulation models.

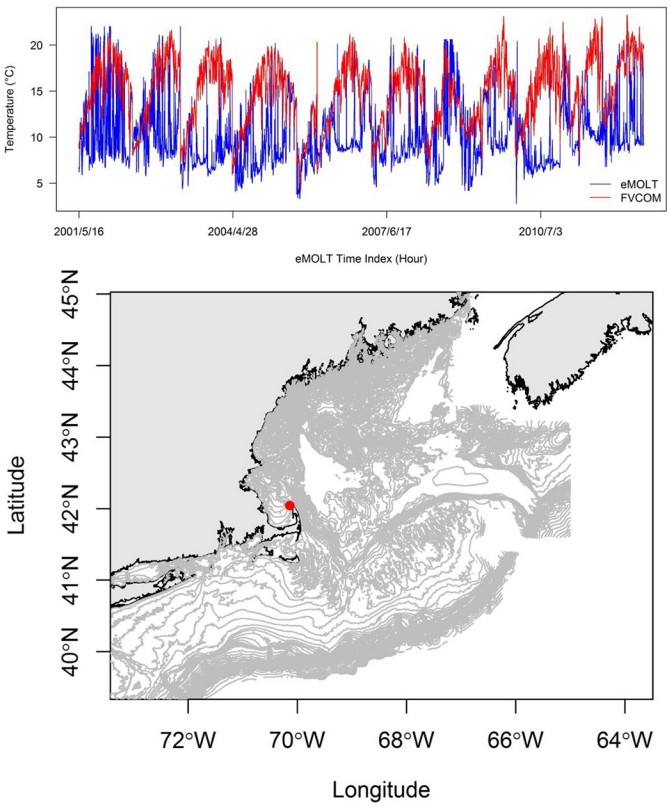
5. Conclusion

Using a series of quantitative methods and a large, unique database of bottom temperature records, this study provides a comprehensive skill assessment of FVCOM-NAS bottom temperature variability. Overall, high correlation and low discrepancy between modeled and observed data indicate that the FVCOM-NAS exhibited reliable model skill throughout its spatiotemporal domain. The observed variability in model skill addresses the questions regarding the quality of the FVCOM bottom water estimate at various spatial and temporal scales in the NAS region. The skill assessment measures used in this study can be applied to other modeled oceanographic variables to serve broader interests. This study provides FVCOM users opportunities to incorporate the spatially-varying magnitude of confidence in the model outputs.

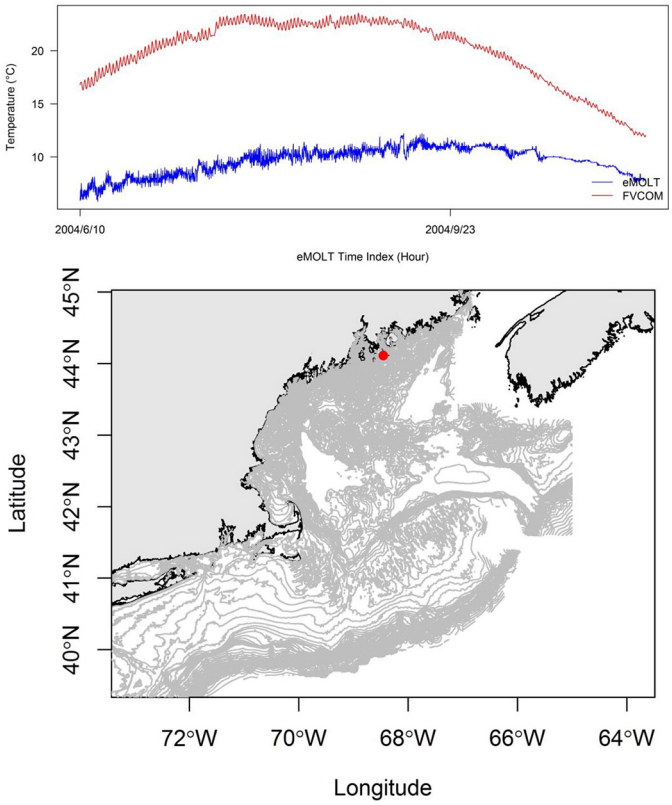
Acknowledgements

We thank Dr. Changsheng Chen and the modelers at the Marine Ecosystem Dynamics Modeling Laboratory, School for Marine Science and Technology, University of Massachusetts-Dartmouth (MEDML/SMAST/UMASSD) for providing the FVCOM hindcast database. This database was produced by SeaPlan (1978–2010) and NERACOOS (2011–2015) in cooperation with the MEDML/SMAST/UMASSD. We thank Mr. James Manning at NOAA Northeast Fisheries Science Center for graciously providing access to the eMOLT data and for thoughtful discussions about the data. We thank Mr. Steve Cousins at the Advanced Computing Group at the University of Maine for providing us high computer processing capability. This work was funded by NSF's Coastal SEES #OCE-1325484 (YC and ACT), NASA #NNX-16 AG59G (ACT), and NSF #11A-1355457 to Maine EPSCoR at the University of Maine (DCB).

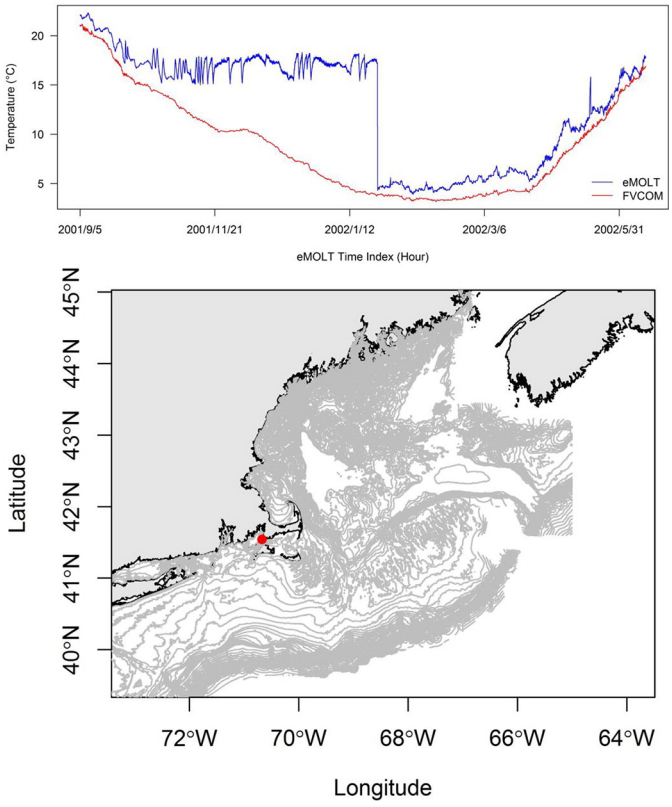
Appendix



Appendix A. Time series of observed (eMOLT: blue) and modeled (FVCOM: red) bottom temperatures at site AB01 (2001–2013; $n = 36{,}690$; −18.3 m) in the top panel. The location of site AB01 is shown in the bottom panel.



Appendix B. Time series of observed (eMOLT: blue) and modeled (FVCOM: red) bottom temperatures at site DB01 during the period of June–November 2004 (n = 3819; −16.5 m). The location of site DB01 is shown in the bottom panel.



Appendix C. Time series of observed (eMOLT: blue) and modeled (FVCOM: red) bottom temperatures at site WH02 (2001–2002; n = 10,491; −1.8 m). The location of site WH02 is shown in the bottom panel.

Appendix D

Parametric coefficients and approximate significance of smooth terms obtained by the generalized additive model containing Year, Month, Depth, Distance Offshore, Latitude, and Longitude. Of the 13 categorical values for Year and 12 categorical values for Month, the first category class is considered a baseline and adopts a value of zero

Parametric coefficients				Approximate significance of smooth terms		
	Estimate	Std. error	p-Value		edf	p-Value
(Intercept)	0.108	0.004	<0.001	s(depth)	3.998	<0.001
2002	0.091	0.003	<0.001	s(distance offshore)	3.995	<0.001
2003	0.113	0.003	<0.001	s(latitude)	4.004	<0.001
2004	0.102	0.003	<0.001	s(longitude)	3.999	<0.001
2005	0.126	0.003	<0.001	*edf: estimated degree of freedom		
2006	0.022	0.003	<0.001			
2007	0.043	0.003	<0.001			
2008	−0.054	0.003	<0.001			
2009	−0.185	0.003	<0.001			
2010	−0.008	0.003	<0.05			
2011	0.049	0.003	<0.001			
2012	−0.046	0.003	<0.001			
2013	0.324	0.003	<0.001			
February	−0.121	0.005	<0.001			
March	−0.198	0.004	<0.001			
April	−0.219	0.004	<0.001			
May	−0.129	0.003	<0.001			
June	0.150	0.003	<0.001			
July	0.399	0.003	<0.001			
August	0.327	0.003	<0.001			
September	0.037	0.003	<0.001			
October	−0.430	0.004	<0.001			
November	−0.559	0.005	<0.001			
December	−0.227	0.004	<0.001			

References

Bivand, R., 2013. Package "classInt": choose univariate class interval. R Package Version 0.1-21. 2013 . https://CRAN.R-project.org/package=classInt.

Bivand, R.S., Pebesma, E., Gomez-Rubio, V., 2013. Applied Spatial Data Analysis with R. second ed. Springer, New York, NY.

Blumberg, A.F., Mellor, G.L., 1987. A description of a three-dimensional coastal ocean circulation model. In: Heaps, N.S. (Ed.), Three-Dimensional Coastal Ocean Models. American Geophysical Union, Washington, D.C., pp. 1–16.

Chen, C., 2015. Northeast US 30+ year hindcast on GOM3 grid (1978–2013). The School for Marine Science and Technology, University of Massachusetts, Dartmouth Accessed on 28 May 2015. URL. http://www.smast.umassd.edu:8080/thredds/hindcasts.html.

Chen, C., Beardsley, R.C., Cowles, G.W., 2006. An unstructured-grid, finite-volume coastal ocean model (FVCOM) system. Oceanography 19, 78–89.

Chen, C., Huang, H., Beardsley, R.C., Xu, Q., Limeburner, R., Cowles, G.W., Sun, Y., Qi, J., Lin, H., 2011. Tidal dynamics in the Gulf of Maine and New England Shelf: an application of FVCOM. J. Geophys. Res. 116, C12010. http://dx.doi.org/10.1029/2011JC007054.

Cowles, G.W., Lentz, S.J., Chen, C., Xu, Q., Beardsley, R.C., 2008. Comparison of observed and model-computed low frequency circulation and hydrography on the New England shelf. J. Geophys. Res. Oceans 113:1–17. http://dx.doi.org/10.1029/2007JC004394.

Fernandez, I.J., Schmitt, C.V., Birkel, S.D., Stancioff, E., Pershing, A.J., Kelley, J.T., Runge, J.A., Jacobson, G.L., Mayewski, P.A., 2015. Maine's Climate Future: 2015 Update. Orono, ME.

Fitzpatrick, J.J., 2009. Assessing skill of estuarine and coastal eutrophication models for water quality managers. J. Mar. Syst. 76:195–211. http://dx.doi.org/10.1016/j.jmarsys.2008.05.018.

Giorgino, T., 2009. Computing and visualizing dynamic time warping alignments in R: the dtw package. J. Stat. Softw. 31, 1–24.

Greene, C.H., Meyer-Gutbrod, E., Monger, B.C., McGarry, L.P., Pershing, A.J., Belkin, I.M., Fratantoni, P.S., Mountain, D.G., Pickart, R.S., Proshutinsky, A., Ji, R., Bisagni, J.J., Hakkinen, S.M.A., Haidvogel, D.B., Wang, J., Head, E., Smith, P., Reid, P.C., Conversi, A., 2013. Remote climate forcing of decadal-scale regime shifts in Northwest Atlantic shelf ecosystems. Limnol. Oceanogr. 58:803–816. http://dx.doi.org/10.4319/lo.2013.58.3.0803.

Greene, C.H., Pershing, A.J., 2007. Climate drives sea change. Science 315 (5815): 1084–1085. http://dx.doi.org/10.1126/science.1136495.

Gu, J., Jin, X., 2006. A simple approximation for dynamic time warping search in large time series database. Intelligent Data Engineering and Automated Learning–IDEAL 2006, 841–848.

Hastie, T., Tibshirani, R., 1990. Generalized additive models. 43. CRC Press.

Huret, M., Runge, J.A., Chen, C., Cowles, G., Xu, Q., Pringle, J.M., 2007. Dispersal modeling of fish early life stages: sensitivity with application to Atlantic cod in the western Gulf of Maine. Mar. Ecol. Prog. Ser. 347:261–274. http://dx.doi.org/10.3354/meps06983.

Hurrell, J.W., 1995. Decadal trends in the north atlantic oscillation: regional temperatures and precipitation. Science 269 (5224):676–679. http://dx.doi.org/10.1126/science.269.5224.676.

Ji, R., Davis, C., Chen, C., Beardsley, R., 2008. Influence of local and external processes on the annual nitrogen cycle and primary productivity on Georges Bank: a 3-D biological-physical modeling study. J. Mar. Syst. 73:31–47. http://dx.doi.org/10.1016/j.jmarsys.2007.08.002.

Leggett, R.W., Williams, L.R., 1981. A reliability index for models. Ecol. Model. 13:303–312. http://dx.doi.org/10.1016/0304-3800(81)90034-X.

Li, B., Cao, J., Chang, J., Wilson, C., Chen, Y., 2015a. Evaluation of effectiveness of fixed-station sampling for monitoring American lobster settlement. North Am. J. Fish. Manag. 35:942–957. http://dx.doi.org/10.1080/02755947.2015.1074961.

Li, Y., Fratantoni, P.S., Chen, C., Hare, J.A., Sun, Y., Beardsley, R.C., Ji, R., 2015b. Spatio-temporal patterns of stratification on the Northwest Atlantic shelf. Prog. Oceanogr. 134: 123–137. http://dx.doi.org/10.1016/j.pocean.2015.01.003.

Loague, K., Green, R.E., 1991. Statistical and graphical methods for evaluating solute transport models: overview and application. J. Contam. Hydrol. 7:51–73. http://dx.doi.org/10.1016/0169-7722(91)90038-3.

Loder, J.W., Boicourt, W.C., Simpson, J.H., 1998. Western ocean boundary shelves. In: Robinson, A.R., Brink, K.H. (Eds.), The sea: The Global Coastal Ocean: Interdisciplinary Regional Studies and Syntheses. John Wiley & Sons, Inc., pp. 3–27.

Manning, J.P., Pelletier, E., 2009. Environmental monitors on lobster traps (eMOLT): long-term observations of New England's bottom-water temperatures. J. Oper. Oceanogr. 2:25–33. http://dx.doi.org/10.1080/1755876X.2009.11020106.

Massachusetts Energy and Environmental Affairs, 2014. Massachusetts ocean management plan. Baseline Assessment and Science Framework. Boston, Massachusetts. 2.

Miao, C., Duan, Q., Sun, Q., Huang, Y., Kong, D., Yang, T., Ye, A., Di, Z., Gong, W., 2014. Assessment of CMIP5 climate models and projected temperature changes over Northern Eurasia. Environ. Res. Lett. 9:55007. http://dx.doi.org/10.1088/1748-9326/9/5/055007.

Murphy, J.M., Sexton, D.M.H., Barnett, D.N., Jones, G.S., Webb, M.J., Collins, M., Stainforth, D.A., 1998. Quantification of modelling uncertainties in a large ensemble of climate change simulations. Nature 396:1–5. http://dx.doi.org/10.1038/nature02770.1.

Nash, J.E., Sutcliffe, J.V., 1970. River flow forecasting through conceptual models part I — a discussion of principles. J. Hydrol. 10:282–290. http://dx.doi.org/10.1016/0022-1694(70)90255-6.

Roworth, E.T., Signell, R.P., 1998. Construction of digital bathymetry for the Gulf of Maine. Open File Rep. 98–801.

Stock, C.A., Alexander, M.A., Bond, N.A., Brander, K.M., Cheung, W.W.L., Curchitser, E.N., Delworth, T.L., Dunne, J.P., Griffies, S.M., Haltuch, M.A., Hare, J.A., Hollowed, A.B., Lehodey, P., Levin, S.A., Link, J.S., Rose, K.A., Rykaczewski, R.R., Sarmiento, J.L., Stouffer, R.J., Schwing, F.B., Vecchi, G.A., Werner, F.E., 2011. On the use of IPCC-class models to assess the impact of climate on living marine resources. Prog. Oceanogr. 88:1–27. http://dx.doi.org/10.1016/j.pocean.2010.09.001.

Stow, C.A., Jolliff, J., McGillicuddy Jr., D.J., Doney, S.C., Allen, J.I., Friedrichs, M.A.M., Rose, K.A., Wallhead, P., 2009. Skill assessment for coupled biological/physical models of marine systems. J. Mar. Syst. 76:4–15. http://dx.doi.org/10.1016/j.jmarsys.2008.03.011.

Tanaka, K., Chen, Y., 2015. Spatiotemporal variability of suitable habitat for American lobster (*Homarus americanus*) in Long Island Sound. J. Shellfish Res. 34:531–543. http://dx.doi.org/10.2983/035.034.0238.

Tanaka, K., Chen, Y., 2016. Modeling spatiotemporal variability of the bioclimate envelope of *Homarus americanus* in the coastal waters of Maine and New Hampshire. Fish. Res. 177:137–152. http://dx.doi.org/10.1016/j.fishres.2016.01.010.

Taylor, K.E., 2005. Taylor Diagram Primer.Available at. http://www.atmos.albany.edu/daes/atmclasses/atm401/spring_2016/ppts_pdfs/Taylor_diagram_primer.pdf (last access: 21 March 2017). :pp. 1–4 http://dx.doi.org/10.1029/2000JD900719.

Townsend, D.W., Pettigrew, N.R., Thomas, M.A., Neary, M.G., Mcgillicuddy, D.J., Donnell, J.O., 2015. Water masses and nutrient fluxes to the Gulf of Maine. J. Mar. Res. 73: 93–122. http://dx.doi.org/10.1038/141548c0.

Townsend, D.W., Thomas, A.C., Mayer, L.M., Thomas, M.A., Quinlan, J.A., 2006. Oceanography of the northwest Atlantic Continental Shelf (1, W). In: Robinson, A.R., Brink, K.H. (Eds.), The Sea: The Global Coastal Ocean: Interdisciplinary Regional Studies and Syntheses. Harvard University Press, Cambridge, USA, pp. 119–168.

Wood, S.N., 2011. Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. J. R. Stat. Soc. Ser. B (Stat Methodol.) 73:3–36. http://dx.doi.org/10.1111/j.1467-9868.2010.00749.x.

Wu, R., Chen, J., Wen, Z., 2013. Precipitation-surface temperature relationship in the IPCC CMIP5 models. Adv. Atmos. Sci. 30:766–778. http://dx.doi.org/10.1007/s00376-012-2130-8.

Xue, H., Incze, L., Xu, D., Wolff, N., Pettigrew, N., 2008. Connectivity of lobster populations in the coastal Gulf of Maine. Part I: Circulation and larval transport potential. Ecol. Model. 210:193–211. http://dx.doi.org/10.1016/j.ecolmodel.2007.07.024.