Seasonal Prediction of Bottom Temperature on the Northeast U.S. Continental Shelf

Zhuomin Chen¹, Young-Oh Kwon¹, Ke Chen¹, Paula Fratantoni², Glen G. Gawarkiewicz¹, Terrence M. M Joyce³, Timothy J. Miller⁴, Janet A. Nye⁵, VINCENT SELLITTO SABA⁶, and Brian Stock⁴

¹Woods Hole Oceanographic Institution
²NOAA NMFS Northeast Fisheries Science Center
³WHOI
⁴NOAA NMFS, Northeast Fisheries Science Center
⁵University of North Carolina Chapel Hill
⁶National Oceanic and Atmospheric Administration, National Marine Fisheries Service

November 24, 2022

Abstract

The Northeast U.S. shelf (NES) is an oceanographically dynamic marine ecosystem and supports some of the most valuable demersal fisheries in the world. A reliable prediction of NES environmental variables, particularly ocean bottom temperature, could lead to a significant improvement in demersal fisheries management. However, the current generation of climate model-based seasonal-to-interannual predictions exhibit limited prediction skill in this continental shelf environment. Here we have developed a hierarchy of statistical seasonal predictions for NES bottom temperatures using an eddy-resolving ocean reanalysis dataset. A simple, damped local persistence prediction model produces significant skill for lead times up to ~6 months in the Mid-Atlantic Bight and up to ~11 months in the Gulf of Maine, although the prediction skill varies notably by season. Considering temperature from a nearby or upstream (i.e. more poleward) region as an additional predictor generally improves prediction skill, presumably as a result of advective processes. Large-scale atmospheric and oceanic indices, such as Gulf Stream path indices (GSIs) and the North Atlantic Oscillation index, are also tested as predictors for NES bottom temperatures. Only the GSI constructed from temperature observed at 200 m depth significantly improves the prediction skill relative to local persistence. However, the prediction skill from this GSI is not larger than that gained using models incorporating nearby or upstream shelf/slope temperatures. Based on these results, a simplified statistical model has been developed, which can be tailored to fisheries management for the NES.
Seasonal Prediction of Bottom Temperature on the Northeast U.S. Continental Shelf

Zhuomin Chen\textsuperscript{1,2}, Young-Oh Kwon\textsuperscript{1}, Ke Chen\textsuperscript{1}, Paula Fratantoni\textsuperscript{1,3}, Glen Gawarkiewicz\textsuperscript{1}, Terrence M. Joyce\textsuperscript{1}, Timothy J. Miller\textsuperscript{3}, Janet A. Nye\textsuperscript{4}, Vincent S. Saba\textsuperscript{5}, and Brian C. Stock\textsuperscript{3}

\textsuperscript{1}Physical Oceanography Department, Woods Hole Oceanographic Institution, Woods Hole, MA.
\textsuperscript{2}National Research Council Postdoctoral Associateship Program, hosted by NOAA NMFS, Northeast Fisheries Science Center, Woods Hole, MA.
\textsuperscript{3}NOAA NMFS, Northeast Fisheries Science Center, Woods Hole, MA.
\textsuperscript{4}University of North Carolina Chapel Hill, Institute of Marine Sciences, Morehead City, NC.
\textsuperscript{5}NOAA NMFS, Northeast Fisheries Science Center, Geophysical Fluid Dynamics Laboratory, Princeton University Forrestal Campus, Princeton, NJ.

Corresponding author: Zhuomin Chen (zchen@whoi.edu)

\textbf{Key Points:} (<140 characters)

- A local persistence model for bottom temperature provides prediction skill up to \textasciitilde11 months with strong variation by region and season.
- Including upstream or nearby bottom temperature as a predictor largely improves prediction skill, mainly attributed to ocean advection.
- An index of Gulf Stream position produces skillful predictions in fall.
Abstract (<250 words)

The Northeast U.S. shelf (NES) is an oceanographically dynamic marine ecosystem and supports some of the most valuable demersal fisheries in the world. A reliable prediction of NES environmental variables, particularly ocean bottom temperature, could lead to a significant improvement in demersal fisheries management. However, the current generation of climate model-based seasonal-to-interannual predictions exhibit limited prediction skill in this continental shelf environment. Here we have developed a hierarchy of statistical seasonal predictions for NES bottom temperatures using an eddy-resolving ocean reanalysis dataset. A simple, damped local persistence prediction model produces significant skill for lead times up to ~6 months in the Mid-Atlantic Bight and up to ~11 months in the Gulf of Maine, although the prediction skill varies notably by season. Considering temperature from a nearby or upstream (i.e. more poleward) region as an additional predictor generally improves prediction skill, presumably as a result of advective processes. Large-scale atmospheric and oceanic indices, such as Gulf Stream path indices (GSIs) and the North Atlantic Oscillation index, are also tested as predictors for NES bottom temperatures. Only the GSI constructed from temperature observed at 200 m depth significantly improves the prediction skill relative to local persistence. However, the prediction skill from this GSI is not larger than that gained using models incorporating nearby or upstream shelf/slope temperatures. Based on these results, a simplified statistical model has been developed, which can be tailored to fisheries management for the NES.

Plain Language Summary

Reliable prediction of bottom temperature could improve fisheries stock assessments on the Northeast U.S. shelf (NES), where some of the most commercially valuable fisheries located. However, for a variety of reasons the current generation of numerical climate models are unable to produce skillful predictions in this region. In this study, we have developed a collection of statistical models that produce seasonal predictions of NES bottom temperature. These include a simple model that only considers local persistence; a nonlocal model that also considers bottom temperature from an upstream or nearby region; and statistical prediction models that incorporate indices of large-scale atmospheric and oceanic variability like the North Atlantic Oscillation or position of the Gulf Stream (GS). The simple model provides significant skill for lead times up to ~6 months in the Mid-Atlantic Bight and up to ~11 months in the Gulf of Maine, although the prediction skill varies by season. The nonlocal model and the GS index model generally improve the prediction skill relative to the local model. However, the GS index model is not better than the nonlocal model. Based on these results, a simplified statistical model has been developed, which can be tailored to fisheries management on the NES.

1 Introduction

The Northeast U.S. shelf (NES), ranging from the Gulf of Maine (GoM) to Cape Hatteras (Figure 1a), is an oceanographically dynamic and highly productive marine ecosystem. It supports some of the most commercially valuable fisheries in the world, along with high gross revenues and employment. Located at the downstream end of an extensive coastal buoyancy-driven boundary current system, the NES receives cold and fresh arctic-origin water (Labrador shelf and slope waters), accumulated coastal riverine discharge, and ice melt that have been advected thousands of kilometers along the western boundary of the North Atlantic [Chapman and Beardsley, 1989; Loder et al., 1998; Townsend et al., 2006; Fratantoni and Pickart, 2007; Richaud et al., 2016]. Likewise, subtropical water masses, advected by the Gulf Stream (GS), its
meanders and rings, and the Atlantic temperate slope currents, interacting with the subpolar Labrador slope currents, also influence the composition of water masses within the NES region [Mountain, 2012; Greene et al., 2013; Gawarkiewicz et al., 2018].

The NES ecosystem structure and function has experienced significant changes as a result of climate-scale changes in the physical environment [Link et al., 2002; Greene and Pershing, 2007; Nye et al., 2009; Mills et al., 2013]. Shifts in marine species distribution, growth and recruitment have been directly linked to changes in ocean temperature [e.g., Pinsky et al., 2013; Kleisner et al. 2016; Miller et al., 2016; Miller et al., 2018]. Therefore, changing ocean temperature on the NES is an important factor that needs to be considered in marine resource management [e.g., Sullivan et al., 2005; Nye et al., 2011; Tommasi et al., 2017].

Variability in bottom temperature on the NES is strongly influenced by local oceanic processes, e.g., bottom circulation along the shelf and intrusions of warm and saline slope waters onto the shelf. In the Mid-Atlantic Bight (MAB), a distinctive seasonal bottom-trapped cold water mass - the Cold Pool, is maintained by the southwestward advection of winter remnant water formed in the vicinity of Nantucket Shoals and upstream Georges Bank (GB) and GoM water [e.g., Houghton et al., 1982; Lentz, 2017; Chen et al., 2018; Chen and Curchitser, 2020]. In the GoM, temperature variations at depth are influenced by two different processes: (1) advecitive inflow of fresh water from the Scotian Shelf (SS) and its modification through winter convective mixing [Mountain and Manning, 1994]; and (2) inflow of warmer and saltier slope waters through the Northeast Channel (NEC), consisting of varying proportions of Labrador slope water and Atlantic temperate slope water, each with their own seasonal property variations [Ramp et al., 1985; Smith et al., 2001; Greene et al., 2013]. Besides these advective processes, bottom waters on the NES are also influenced by cross-shelf exchange processes associated with GS warm-core rings and meanders. For example, an observed increase in the number of warm-core rings and westward shift in the GS destabilization point since the early 2000s are consistent with coastal warming trends on the NES [Andres et al., 2016; Gawarkiewicz et al., 2018; Gangopadhyay et al., 2019].

The NES ecosystem is sensitive to variations in the large-scale ocean and atmospheric circulation as indicated by meridional shifts in the GS path, changes in the Atlantic Meridional Overturning Circulation (AMOC), and the North Atlantic Oscillation (NAO). Distribution, recruitment, and biomass of some commercially important fish stock are closely related to changes in bottom temperature on the shelf and changes in the latitude of the GS [Nye et al., 2011; Xu et al., 2018]. In addition, changes in GS position lead variations in silver hake distribution by ~6 months, which implies predictability at seasonal to interannual time scales [Davis et al., 2017]. In situ measurements also suggest a direct link between the GS position and the temperature and salinity near the shelf break south of New England on shorter time scales [Gawarkiewicz et al. 2012]. Lucey and Nye [2010] indicated that the entire fish and macroinvertebrate assemblage on the NES respond to changes in water mass properties as indicated by the GS path index even when taking into account the impacts of fishing. Saba et al. [2015] also suggest that the GS position is associated with phytoplankton biomass on the shelf break, slope, and specific coastal regions of the MAB. Observations and modeling studies in the Northwest Atlantic (NWA) point to an inverse relationship between the AMOC and the position of the GS [Zhang, 2008] with a weaker AMOC related to a more northerly position of the GS [Joyce and Zhang, 2010; Zhang et al., 2011]. A northerly shift in the GS is then associated with warmer ocean temperature in the NWA [Frankignoul et al., 2001; Zhang and Vallis, 2007; Peña-
Molino and Joyce, 2008; Saba et al., 2016]. In addition, it is also shown that sea surface
temperature (SST) anomalies in the GoM correlated with the NAO can be traced upstream to the
Labrador Shelf four years earlier [Xu et al., 2015], which is also a source of predictability for the
temperature on the NES.

Many studies focused on the impacts of ocean warming on the NES have identified
historical and projected changes in the distribution of commercial [Kleisner et al., 2017; Selden
et al., 2018; McHenry et al., 2019] and recreational [Crear et al., 2020] species including two of
the most valuable commercial fisheries in the U.S., American lobster and sea scallops [Tanaka et
al., 2020]. Research has also focused on species responses to ocean temperature beyond
distribution shifts. For example, southern New England yellowtail flounder recruitment has been
linked to the variability of the size and duration of the MAB Cold Pool represented by the
average fall bottom temperature [Miller et al., 2016], and the position of the GS [Xu et al., 2018].
However, the relationship between the Cold Pool and recruitment is not entirely useful for
fisheries management because skillful annual forecasts of the Cold Pool are not available [Miller
et al., 2016]. The relationship between the GS and yellowtail flounder recruitment provided
skillful forecasts in only some years [Xu et al., 2018]. These studies emphasized the need for
skillful forecasts of ecologically relevant environmental variables that include estimates of
uncertainty, consideration of season and appropriate lags.

The current generation of climate model-based predictions of ocean temperature
emphasize basin-scale variations of surface temperatures on decadal scales, e.g., the El Niño-
Southern Oscillation [Stock et al., 2015], and shelf- and/or slope-scale dynamics are not well
captured due to the under-represented coastal processes (e.g., freshwater flux, instabilities, and
tides) and coarse resolution (~0.5°-2°) of those global forecast models [Stock et al., 2011].
Recent studies have found that among all U.S. Large Marine Ecosystems (LMEs), the current
global climate model forecast systems have realized only limited prediction skill in the SST
along the northeast coast (NES and SS) [Hervieux et al., 2019], compared with the west coast
[Jacox et al., 2018a, 2018b, 2019, and 2020]. Hence, there is a need for skillful predictions of
ocean temperature on the NES using alternative approaches, e.g., statistical models, along with
improved understanding of the causes of its variability.

In this study, we have developed a hierarchy of statistical models for the prediction of
ocean bottom temperature on the NES at seasonal scales. Our goal is to develop a tailored
approach to ocean prediction on the NES that meets the scientific needs of fisheries stock
assessments and management. This paper is organized as follows. In Section 2, we describe the
study area and the ocean reanalysis dataset used for constructing and validating the prediction
models. Section 3 describes the statistical prediction models, and methods for cross-validation
and prediction skill assessment. Results from the prediction models are described in Section 4,
followed by a discussion and summary in Section 5.

2 Study Area and Data

2.1 Study Area

The NES, the target region for this statistical prediction work, encompasses 47
Ecosystem Monitoring (EcoMon) strata defined by NOAA’s Northeast Fisheries Science Center
(NEFSC; Figure 1a) [Walsh et al., 2015]. Here, the 47 strata are grouped into 10 subregions to
be used for statistical prediction: box01-southern MAB (SMAB), box02-southern shelfbreak
(SSB), box03-northern MAB (NMAB), box04-northern shelfbreak (NSB), box05-Georges Bank (GB), box06-western shallow GoM (WSGoM), box07-deep GoM (DGoM), box08-Northeast Channel (NEC), box09-northern shallow GoM (NSGoM), and Box10-Browns Bank (BB), as shown in Figure 1a. Each subregion is chosen to include strata with consistent bottom temperature variability as determined by k-means clustering [Hartigan & Wong, 1979], which is an unsupervised machine learning technique that aggregates data based on the similarity of a chosen variable. The cluster analysis is performed separately for the strata belonging to the MAB and GB regions and the strata in the GoM region, owing to the distinct hydrographic characteristics in each region. The monthly mean bottom temperature anomalies calculated for each EcoMon strata based on the GLORYS12v1 ocean reanalysis dataset (1993-2018; Section 2.2) are used for the cluster analysis, with long-term linear trends removed.

Furthermore, we incorporate the 10 NES subregions defined by the cluster analysis into the shelf and slope boxes in the greater NWA domain considered by Chen et al. [2020], generating a new set of boxes that includes 19 shelf boxes and 8 slope boxes (Figure 1b). The bottom temperature anomaly calculated for each of these shelf and slope boxes will be considered as potential predictors as explained in Section 3.2.

**Figure 1.** The Northeast U.S. continental shelf (NES) and the Northwest Atlantic (NWA) continental shelf and slope. (a) The Ecosystem Monitoring strata (47 black polygons) and the grouped subregions (10 red polygons) on the NES. These 10 subregions are listed in the bottom, together with additional NWA shelf boxes. The blue dots represent locations for the observational data from the NOAA NEFSC hydrographic database (Text S1). Special geographical features on NES are represented by cyan markers and annotated in the upper left legend box. (b) The new set of shelf and slope boxes over the NWA created by replacing the NES boxes in the NWA shelf and slope boxes from Chen et al. [2020] with the new 10 clustered...
subregions from strata. The white dashed line represents the average position of the SSH-based Gulf Stream main axis [Chen et al., 2020] and the cyan dots represent the locations of temperature observations along the 15°C isotherm used to generate the 200 m temperature-based Gulf Stream Index [Joyce et al., 2009].

2.2 The GLORYS12v1 Ocean Reanalysis Dataset

Observations near the bottom of the water column on the NWA continental shelf are limited and discontinuous in time and space, and even sparser over the continental slope [Richaud et al., 2016]. Thus, statistical prediction of the NES bottom temperature in this study is based on the GLORYS12v1 product [Lellouche et al., 2018], which is a global ocean eddy-resolving (1/12° horizontal resolution and 50 vertical levels) data assimilated hindcast from Mercator Ocean, covering the altimetry era (1993-2018) [Fernandez & Lellouche, 2018; Lellouche et al., 2018]. The base ocean model is the Nucleus for European Modelling of the Ocean [Madec, 2008] driven by European Center for Medium-Range Weather Forecasts Interim reanalysis [Dee et al., 2011]. Observations of delayed time sea level anomaly from all altimetric satellites, satellite sea surface temperature (SST) from Reynolds 0.25° Advanced Very-High-Resolution Radiometer-only [Reynolds et al., 2007], sea ice concentration from the Centre ERS d’Archivage et de Traitement [Girard-Ardhuin et al., 2008] and in situ temperature and salinity vertical profiles from Copernicus Marine Environment Monitoring Service (CMEMS) Coriolis Re-Analysis v4.1 database [Cabanès et al., 2013], are jointly assimilated by means of a reduced-order Kalman filter. A 3D-VAR scheme is employed to correct for slowly-evolving large-scale temperature and salinity biases. Variables such as temperature, salinity, currents, sea surface height, mixed layer depth and ice parameters are archived at daily and monthly interval, available at the CMEMS.

We have compared the GLORYS12v1 reanalysis dataset with in situ observations of both surface and bottom temperature collected by NOAA NEFSC. Because relatively few observations are available in NSGoM and BB boxes (shelf boxes 09 and 10; see blue dots in Figure 1a), neither evaluation of the reanalysis dataset nor investigation of predictability in these regions is performed. Results suggest the GLORYS12v1 dataset is highly consistent with observations over the NES on both seasonal and interannual time scales (Text S1 and Figure S1). Hence, we use the bottom temperature from the GLORYS12v1 dataset over the shelf regions of the NWA together with the temperature at ~200 m depth over the slope and open ocean (35°-65°N, 40°-80°W). Based on sill depths, it is assumed that slope waters deeper than ~200 m do not directly influence bottom temperatures on the NES. Thus, the temperature at ~200 m depth is considered in place of bottom values over the slope and open ocean. Detrended monthly bottom (or 200 m) temperature anomalies calculated within each subregion are used for our statistical prediction.

3 Statistical Prediction Models

Four statistical prediction models with a hierarchy of complexity are built to investigate the seasonal predictability of bottom temperature anomaly (BTA) on the NES.

3.1 Local Persistence Model
We first evaluate the predictability associated with the damped local persistence for each subregion, which can be considered a baseline for evaluating the skill of other prediction models. The local persistence model of BTA for a subregion \( i \) on the NES, where \( i = 1,2,\ldots,10 \), can be written as:

\[
\hat{BTA}_i(t) = \alpha \cdot BTA_i(t - \tau) + c
\]

where \( t \) represents the target forecast month and \( \tau \) is the lead time. The coefficients \( \alpha \) and \( c \) are determined from simple lagged linear auto-regression applied to the \( BTA_i \) time series from GLORYS12v1. To avoid overfitting the statistical models to the reanalysis time series, \( \alpha \) and \( c \) are calculated via a \( k \)-fold cross-validation method, described in Section 3.5. Therefore, \( \alpha \) and \( c \) are not constants, but vary slightly between cross validation segments.

3.2 The Generalized Persistence Model

In the local persistence model, the predictor is the BTA from the same box at the prediction lead time (\( \tau \) months before). The constraint on the choice of the predictor is relaxed in the second model, so that it can be chosen from any box on the NWA shelf and slope (Figure 1b), including the target predictand box itself. We call this second model, a generalized persistence model rewritten as:

\[
\hat{BTA}_i(t) = \alpha \cdot BTA_j(t - \tau) + c
\]

where \( j \) is not necessarily equal to \( i \). For a particular target predictand subregion \( i \) at the forecast target month \( t \), any shelf and slope boxes (01-19 and A-H) at \( \tau \) months lead can be chosen as the predictor. Among these 27 choices, the one that gives the highest prediction skill for each target month \( t \) and lead time \( \tau \) is chosen. This model is identical to the local persistence model when \( j = i \). Otherwise, the model is meant to capture the predictability associated with advection from box \( j \) to the target box \( i \). Therefore, lagged correlation of BTA time series between the two boxes should be positive, which is a criterion applied when selecting the predictor.

3.3 The Gulf Stream Index Model and NAO Index Model

We also considered indices of Gulf Stream position variability (GSIs) and the North Atlantic Oscillation Index (NAOI) as predictors for shelf bottom temperatures. The GSI or NAOI model can be written as:

\[
\hat{BTA}_i(t) = \alpha \cdot X(t - \tau) + c; X = GSI \text{ or } NAOI.
\]

where \( X \) representing the GSI or NAOI.

We consider two different GSIs, both tracking meridional shifts in the Gulf Stream position on monthly time scales from 1993-2018 (Figure S2). First, a GSI is calculated based on sea surface height (SSH) following Pérez-Hernández & Joyce [2014], reflecting the position variability of the Gulf Stream main axis between 72° and 52°W based on the monthly-mean 1/4° satellite-based gridded sea surface height (SSH) from the CMEMS. The other GSI is calculated based on the subsurface temperature at 200m depth (T200) following Joyce et al. [2009],
reflecting the position variability of the Gulf Stream North Wall. Here, we use the EN4 quality-controlled gridded subsurface temperature (1° horizontal resolution; available from 1950s - present) [Good et al., 2013], instead of temperature profiles from the World Ocean Database (WOD) as was used in Joyce et al. [2009]. A comparison of the two temperature-based GSIs are included in Text S2 and Figure S2.

The monthly NAOI, which is based on a rotated EOF analysis [Barnston and Livezey, 1987], is obtained from the National Weather Service Climate Prediction Center. The time period spans from 1993-2018 and a linear trend is removed prior to its use as a predictor in our statistical model.

3.4 The Generalized Persistence with GSI or NAOI Model

We combine the generalized persistence model and the GSI/NAOI model by considering the GSIs or NAOI in addition to BTAs from neighboring regions, allowing these indices to compete with shelf/slope bottom temperatures when predicting the NES bottom temperatures. This statistical prediction model exactly follows Equation 3, but includes a broader range of candidate predictors:

\[
\overline{BTA}_i(t) = \alpha \cdot X(t - \tau) + c; X = BTA_j, GSIs, or NAOI. \quad (4)
\]

3.5 k-fold Cross-Validation

We use k-fold cross-validation to avoid overfitting the prediction models to the GLORYS12v1 BTA time series. This is also a way to confirm that the statistical relationship between the predictor and predictand is stationary throughout the time series. If the statistical relationship is non-stationary, the prediction skill by the predictor will not be robust for the whole time series.

The basic idea of this method is to train the statistical model using a training time series while testing the relationship over the remaining time series. The detailed procedure is as follows:

1. evenly split the GLORYS12v1 BTA time series into k (k=12 in this study) segments, each spanning 2-3 years, identifying one segment as the testing time series and the remaining k-1 segments as training time period; (2) fit the prediction model to the reanalysis time series over the training time period, i.e. obtain its parameters \( \alpha \) and \( c \) through linear regression; (3) apply the prediction model to calculate a predicted time series over the testing time period; (4) change the testing time period to another segment, repeating steps (2) and (3); and (5) combine all the piecewise linearly predicted (i.e. cross-validated) time series segments chronologically to construct a full prediction for the entire time period. The full prediction time series is then compared to the observed (i.e. GLORYS12v1) time series for the entire time period as illustrated in Section 3.6.

3.6 Assessment of BTA Predictions

The skill of BTA prediction is assessed using the anomaly correlation coefficient (ACC) between the cross-validated prediction (\( \overline{BTA} \)) and GLORYS12v1 BTA time series, which is considered as a function of the target forecast month (\( t \)) and lead time in months (\( \tau \)): 
\[ ACC(t, \tau) = \frac{\sum_{k=1}^{n} (\overline{BTA}_k - BTA)^2}{\sqrt{\sum_{k=1}^{n} \overline{BTA}_k^2} \cdot \sqrt{\sum_{k=1}^{n} BTA^2}} \] (6)

where \( \overline{BTA} = \overline{BTA}(t, \tau) \), \( BTA = BTA(t) \), and \( n \) is the number of years in the time series.

Statistical significance of ACC is tested at the 95% confidence level via Student’s t-test, considering the effective degrees of freedom to account for autocorrelation [Bretherton et al., 1999]. Statistical significance for the difference between two ACCs (resulting from two different statistical prediction models for the same \( t \) and \( \tau \)) is also tested at the 95% confidence level following Steiger [1980]’s method, taking into account the fact that the two correlations are based on a common time series, e.g., the \( BTA(t) \) in our case [Howell, 2009].

4 Results

4.1 Local Persistence

On the NES, the baseline damped local persistence predictions of BTAs yield significant prediction skill (at 95% confidence level) for lead times up to several months (Figure 2). Generally, northern NES boxes (boxes 05-08) exhibit longer persistence, compared to the southern shelf boxes (boxes 01-04; Figure 3). For example, the BTAs persist up to ~4 months in the NMAB (box 03) and up to ~11 months in the DGoM (box 07). This difference is likely due to differences in the regional settings, e.g., bottom depths, mixed layer depths, and proximity to the Gulf Stream between the two regions.

Figure 2. The prediction skill represented by the anomaly correlation coefficient (ACC) of the
The damped local persistence model in each subregion (shelf box 01-08) on the NES. The x-axis represents target forecast month ($t$), and y-axis represents lead months ($\tau$). The grey dots indicate that the ACC is significant at the 95% confidence level.

All eight subregions exhibit notable seasonal differences in prediction skill, with the forecast months corresponding to the shortest/longest persistence varying from region to region. In general, persistence is minimal for winter forecasts (January-March) in every subregion except the DGoM and NEC (shelf boxes 07-08; Figure 3a), presumably a consequence of deep winter mixing driven by storms and atmospheric cooling. In contrast, persistence gradually increases during spring and summer forecasts, reflecting the preservation of winter BTAs underneath a stratified upper layer.

For the SMAB region (box 01), the shortest persistence (<1 month) of BTA from prior months is in the forecast month of January. Persistence increases in the following months and reaches its longest value (~6 months) in July (Figure 2a), meaning that the BTA established in January can persist into July. Later in the summer, the persistence decreases to 5 months in August and September, followed by a sudden drop in October-December, which is likely associated with the onset of wind mixing in the fall [e.g., Castelao et al., 2008; Forsyth et al., 2018]. The NMAB shows a similar pattern of persistence for the winter and spring forecasts, with a minimum persistence of 1 month in January and longest persistence of 4 months in May to July. However, the persistence of BTA remains roughly the same (up to 3-4 months) in the following months (August-December) (Figure 2c), consistent with the establishment of the subsurface Cold Pool water (<10°C) typically from spring to fall [Chen et al., 2018].

![Figure 3](image-url) ACC prediction skill for the damped local persistence model in each subregion on the NES (represented by different colors), averaged for (a) each target forecast month ($t$), (b) lead time ($\tau$), and (c) initialization month ($t - \tau$) by averaging values in Figure 2 in the vertical, horizontal, or diagonal direction, respectively.

The two shelfbreak regions (box 02 and 04), which are narrow strips aligned with the shelfbreak in the MAB, show irregular seasonal persistence patterns that are difficult to interpret and may not be robust (Figure 2b and 2d). As an example, the northern shelfbreak region
exhibits very limited persistence (1-2 months) across all forecast months, with isolated elevated skill for fall predictions with lead times of 4-8 months. Previous work in the region estimates very short time scales for predictability based on mooring data from the Nantucket Shoals Flux Experiment [Yuan et al., 2004]. We will not examine predictability in these regions using the other statistical models.

As previously mentioned, BTAs in the northern boxes (boxes 05-08) exhibit generally longer persistence compared with the MAB subregions (Figure 2e-h). These four subregions can be grouped into two categories based on the seasonal patterns of persistence. The first group includes the GB and WSGoM (boxes 05-06), where both exhibit the shortest persistence (~1 month) for forecasts in January/February (similar to the southern boxes) and the longest persistence for forecasts in late fall (reaching 9-10 months lead time, Figure 2e-f). This result suggests the importance of late-winter/early-spring temperature in determining the BTA in the following summer and fall in these two regions. In contrast, the DGoM and NEC regions both exhibit a shifted seasonality with minimum persistence in May and April, respectively (Figure 2g-h). It is likely that for these two deepest subregions, the winter mixing affects the bottom temperature indirectly with delays, e.g., through advection and convection from the shallower neighbouring regions. Also, in comparison to the shallower regions, the break down of the surface stratification in the fall has limited impact on the persistence of BTA in the DGoM and NEC. As a result, the anomaly signal can last from early spring to the end of the following winter (Figure 2g-h).

The prediction skill may be summarized by averaging the skill shown in Figure 2 in the vertical, horizontal, and diagonal directions, respectively (Figure 3). The results further illustrate the seasonality and regional differences discussed above. Overall, the DGoM stands out when comparing the prediction skill of the local persistence model relative to lead time (Figure 3b). On average, predictions in the DGoM based on local persistence gradually decrease from 0.95 prediction skill at one-month lead to about 0.3 for the 12-month lead (Figure 3b). Among all forecast months, winter months in DGoM have the highest prediction skill (0.77 in March) and summer months have the lowest prediction skill (0.34 in July) (Figure 3a). For the DGoM, BTA initiating in May is the most stable and constant predictor for BTA predictions across the rest of the months, providing an average prediction skill of about 0.72 (Figure 3c).

The NEC region exhibits a similar BTA persistence pattern to the DGoM, although with decreased prediction skill and slightly shifted seasonality (Figure 2h). The highest prediction skill (~0.52) regardless of lead month is achieved in February, one month earlier than the DGoM region; and the lowest prediction skill (0.17) is obtained two months earlier than the DGoM in May (Figure 3a). This is mainly because the NEC region is less isolated than the DGoM, influenced more directly by changes on the slope. Similarly, the most stable and constant BTA signal initiates from April, which is also one-month earlier than that in DGoM (Figure 3c). Prediction skill in the remaining subregions is minimum in early winter (January or February), and the BTA initiating in February or March is also the most stable and constant predictor for the following months in these regions.

4.2 Generalized Persistence Model

The generalized persistence model allows BTA signals from other regions to compete
with the local persistence to achieve more skillful predictions (Section 3.2), presumably through advection. When local persistence dominates, the generalized model is identical to the local persistence model (i.e. \( j = i \) in Equation 2). In other words, the generalized model will either improve the prediction skill or it will be equal to the local persistence model.

We begin by comparing the prediction skill of the generalized persistence model with the local persistence model from three different perspectives (relative to the forecast month, lead month, and initialization month). The results suggest that a significant improvement in prediction skill has been achieved using the generalized persistence model in all shelf subregions (Figure 4). On average, the NMAB shows the most improvement (~0.32) among the subregions, especially for October forecasts (~0.43, averaged over all lead months) and for lead times of 8 months (~0.50, averaged over all forecast months). The DGoM shows the least improvement (~0.1 on average), as the local persistence model already showed strong performance.

Figure 4. As in Figure 3, but for the prediction skills of the generalized persistence model (a-c) and the prediction skill improvement (d-f) relative to the local persistence model in subregions of shelf box01, 03, 05, and 06-08 (represented by different colors), averaged for each forecast month (a, d), lead month (b, e), and initialization month (c, f).

We find two types of predictability barriers (indicated by a line of minimum predictability), based on the prediction skill revealed by the generalized persistence model
(Figure 5). One is along the diagonal, i.e. dependent on the initialization month, and the other is aligned vertically, i.e. dependent on the forecast month. Diagonal predictability barriers are visible in the NMAB, GB, and WSGoM (boxes 03, 05, 06), all of which indicate that BTA from any shelf/slope box initiating in the fall months (October-December) does not improve the prediction skill for these three target regions. This predictability barrier likely reflects the effect of strong mixing in late fall/early winter eroding existing anomalies. The vertical predictability barrier is identified in the DGoM, with very limited prediction skill for the May-June forecast months, which was already apparent in the local persistence model (Figure 2g). As noted, predictability in the DGoM did not improve much with the generalized persistence model (Figure 5h and k).

Most of the nonlocal predictors are located along upstream advective paths and/or within neighboring subregions. In the SMAB region, local persistence only dominates at one or two month(s) lead for nearly all forecast months (Figure 5a and d). For predictions with a 2-6 month lead, the neighboring SSB (box 02), upstream GB (box 05), eastern SS (ESS; box 12) and southern Gulf of St. Lawrence (SGSL; box 13), are better predictors providing an increase of ~0.4 in prediction skill on average. For predictions with a lead of 7-12 months, the best predictor for winter-spring is slope box G (west of the Grand Banks) initiated in the preceding June/July, and the best predictor for May/June is BTA from the northern Labrador shelf (box 19) (Figure 5d). However, not all of these predictors can be explained by advection, as the lead months are not always consistent with advection time scales. For example, it normally takes more than one year for an anomaly signal to advect from the northern Labrador shelf to the SMAB [c.f. Xu et al., 2015], longer than the 7-11 months lead time revealed by the model (Figure 5d).

Local persistence also dominates at shorter lead times (less than 4 months) in the NMAB region. Nonlocal predictors improve the skill of late-summer and fall forecasts (August-December) by ~0.6 on average. Similarly, the most effective predictors include neighboring and upstream shelf regions (box02-SSB, box04-NSB, box05-GB, and box 12-ESS), and slope regions B and C adjacent to the GB and SS (Figure 5e). Winter-spring forecasts (January-June) are also improved by the northernmost nonlocal BTA signals (Labrador shelf boxes 16-19, slope F and G), which again may not be necessarily interpreted as advective signals.

Fall forecasts in the GB and WSGoM regions are dominated by local persistence, with lead months up to 8 or 9 months, forming a nearly right-triangular pattern (Figure 5c and 5g; also Figure 2e-f). Within these triangles, there is little improvement in prediction skill (<0.2) gained from nonlocal predictors, but distinct improvement is achieved for winter and spring forecasts. The best nonlocal predictors mainly include neighboring or upstream boxes. For the GB region, BTA from shelf subregions SSB, NSB, ESS (boxes 02, 04, 12) and slope boxes B, E, and F are better predictors relative to local persistence (Figure 5f). For the WSGoM region, the nonlocal predictors are in the neighboring GoM shelf boxes - DGoM (box 07) and NEC (box 09), and upstream GSL subregions (box 13 and 14), and slope G off the Tail of the Grand Banks (Figure 5j), which are all located along the mean advective pathway of the Labrador Current.

The DGoM has the largest dominance and highest prediction skill from local persistence among all six subregions, and has the minimum improvement in prediction skill when considering nonlocal BTA predictors (Figure 4 and 5). The largest improvements (~0.3 increase) are realized for summer and fall forecasts (August-December), when considering BTA in nearby
GB (box 05), upstream ESS (box 12) and slope D and E as nonlocal predictors (Figure 5k). In the NEC, the prediction skill has been increased by up to ~0.6 relative to the local persistence. These nonlocal predictors also suggest the role of advection, with contributions from upstream shelf box 11 (western SS) and slope boxes C, D, and E (near the SS and the Tail of the Grand Banks) along the advective pathway of the slope currents (Figure 5l).

Overall, considering nonlocal BTA predictors in the generalized persistence model significantly improves prediction skill relative to the local persistence model, and the best nonlocal predictors are predominantly located in neighboring boxes or upstream regions, indicative of an advective pathway. Interpretation of these nonlocal predictors are further discussed in Section 5.1.

Figure 5. The prediction skill of the generalized persistence model in subregions of (a) box01-SMAB, (b) box03-NMAB, (c) box05-GB, (g) box06-WSGoM, (h) box07-DGoM, and (i) box08-NEC, and the corresponding prediction skill increase relative to the local persistence model.
(generalized persistence minus local persistence) in these subregions, with x-axis in each panel representing the forecast month (t) and y-axis the lead time (\tau). In panels (a)-(c) and (g)-(i), the markers (grey dots, light grey and black triangles) indicate the ACC is significant at the 95% confidence level. In particular, the grey dot indicates the prediction skill of the generalized persistence model is identical to the local persistence model, while the triangle (either light grey or black triangles) indicates the generalized persistence model has higher prediction skill than the local persistence model. The black triangle further indicates the difference in ACC between the two models is significant at the 95% confidence level. In panels (d)-(f) and (j)-(l), the corresponding predictor box, ranging from 19 shelf boxes (1-19) and 8 slope boxes (A-H) is indicated for those statistically significant cases with a nonlocal predictor. The black predictor box numbers/characters indicate the ACC improvement is significant at 95% confidence level, while those light grey numbers/characters are not.

4.3 Prediction with Gulf Stream and NAO Indices

We also consider using large-scale oceanic and atmospheric indices as potential predictors for NES bottom temperatures (Section 3.3). We first use the T200-based GSI as the sole predictor for shelf bottom temperatures. The GSI model suggests that the BTA on the NES has a seasonally dependent relationship to the Gulf Stream position (Figure 6). In the SMAB region, the T200-based GSI provides very limited prediction skill for all forecast months except the fall, i.e. September and October, with lead times of 3-8 months. The values of the statistical significant ACC are about 0.55 on average, which are only slightly higher than those of the local persistence model, and most of the increased ACC values are not significantly different from those of local persistence at the 95% confidence level (Figure 6a). In the NAB region, the fall forecast months (September-November) also show significantly higher prediction skill relative to the local persistence model (Figure 6b). In addition, there is a distinctive GSI signal initiating in April, that becomes a statistically significant predictor when predicting the May-to-November BTA in the NAB (Figure 6b). This GSI signal initiating in April is also identified in the GB, DGoM, and NEC regions (Figure 6c, e, and f).

In the GB region, the GSI is a good predictor for summer and fall forecast months across multiple initiating months and lag times, although only one case leads to statistically significant improvements relative to the local persistence model (November forecast based on the GSI in April, 7 months lead; Figure 6c). In contrast, BTA predictability in the WGoM based on GSI is very limited (Figure 6d), suggesting that water properties in shallow regions will be influenced more by the coastal circulation and local atmospheric forcing than by a meridional shift in the GS.

As previously mentioned, the April GSI is a statistically significant predictor for forecasting BTA in the DGoM and NEC from May to the following March. However, the prediction skill is not significantly different from that of local persistence in both regions (Figure 6e-f). Significant increases in prediction skill are achieved by the GSI model for forecasts in the NEC region, when predicting the summer-fall (July to October) BTA at a lead of 7-10 months (Figure 6f).

We also tested two other indices – the SSH-based GSI and the NAOI for shelf BTA predictions (Figure S3). Interestingly, both of these indices provided very limited prediction skill
for all forecast months at lead months less than one year. Especially when comparing the two GSIs, the SSH-based GSI does not show any statistical relationships as the T200-based GSI does with the NES BTAs. We suspect this is because the T200-based GSI reflects the meridional shift of the GS North Wall, including the effects of warm core rings (WCRs) and GS meanders, while the SSH-based GSI captures the meridional shift of GS main axis. Due to closer proximity, the T200-based GSI is more relevant to the NES temperatures. Similarly, the NAOI offers very limited prediction skill, except in the SMAB and NMAB regions for winter forecasts (January and December) at 1-2 months lead, which implies the impact of NAO on local atmospheric forcing over the shallow regions in MAB.

As the generalized persistence model and the T200-based GSI model provide significant improvement in prediction skill relative to the local persistence model, we combine them by treating the T200-based GSI as another box in the generalized persistence model (Section 3.4). Results suggest that this further generalized model results in limited improvement relative to the original generalized persistence model, and the improved prediction skill is not significantly different at the 95% confidence level in any NES subregions (Figure S4).

Figure 6. The prediction skill of the 200 m temperature-based GSI model in subregions of (a) box01-SMAB, (b) box03-NMAB, (c) box05-GB, (d) box06-WSGoM, (e) box07-DGoM, and (f) box08-NEC. Grids with markers (triangles or dots) on them means that the ACC is significant at the 95% confidence level. Specifically, grey dots represent the ACC that is not larger than that of
the local persistence model; while triangles denote a higher ACC relative to the local persistence model and black triangles indicate this increase in ACC is significant at the 95% confidence level.

5 Summary and Discussion

The seasonal predictability of BTA on the NES was investigated based on a hierarchy of statistical prediction models, evaluated relative to the local persistence model. The predictors considered were the local BTA (local persistence model) and nonlocal BTAs (i.e. from nearby or upstream shelf and slope subregions), the T200- and SSH-based GSIs (representing the GS meridional shift), and the NAOI (representing basin-wide atmospheric variability over the North Atlantic). Among these, the combination of local and nonlocal BTAs (generalized persistence model) lead to significant improvements in prediction skill relative to the local persistence model. The use of T200-based GSI also improved the prediction skill of BTAs on NES, but the improvements were confined to fall forecast months and were not significantly better than the skill achieved by the generalized persistence model. Overall, using nearby or upstream BTA’s as predictors in the generalized persistence model provides better results than those using GSI.

In the remainder of the paper, we will discuss various aspects of our statistical BTA prediction, including the physical interpretation of the generalized persistent model results, role of the large-scale variability, additional predictability from the linear long-term trends, and comparison with the SST predictability. As our original motivation of the study is to address the scientific needs of NOAA’s NEFSC, we will conclude the paper by discussing the implications of our BTA prediction results on the fisheries management applications along with suggesting a further simplified generalized prediction model for the potential end-users.

5.1 Interpretation of Nonlocal BTA Predictors

For all of the NES subregions, the best nonlocal predictors in the generalized persistence model are primarily located in neighboring or upstream regions (Figure 5; Section 4.2), suggesting advective pathways along the NWA shelf or slope to the NES. In order to explore this interpretation further, we modified the generalized persistence model to predict the BTA for a given target NES subregion (as before), considering the BTA time series at each grid point in the GLORYS12v1 dataset as candidate predictors, rather than regional average BTAs calculated within each box. The goal is to visualize the pathways associated with the predictability in each target region. To do that, we consider the maximum prediction skill achieved at each grid point for a target region at different lead times, highlighting the maximum potential for prediction, regardless of the forecast month. In other words, for a given lead time, we plot the maximum skill achieved at each grid point among the 12 possible values from 12 forecast months. Doing so also allows us to examine any subjectivity in the statistical models associated with the definition of the shelf and slope boxes. In the following, we discuss the source regions of prediction for the NEC, WSGoM, and NMAB at a lead time of 1, 4, 7, and 10 months (Figures 7-9).

The spatial pattern of maximum prediction skill for the NEC (Figure 7) highlights nonlocal predictors from the WSS (box 11) and from the adjacent continental slope regions (within C, D, and E, also see Figure 5i). At a lead time of one month, results clearly suggest that the highest prediction skill (>0.9) comes from the NEC itself, which agrees with the results in
Section 4.2 that local persistence dominates at shorter lead-times. The spatial maps also show that the deep GoM and WSS basins as well as the slope regions offshore of the SS and GB also contribute to skillful predictions of BTAs in the NEC at 1-month lead. At a lead time of 4 months, the prediction skill within the NEC decreases signaling a decrease in local persistence, while skill remains high in the slope region off the SS. One of the more intriguing features that has emerged from this analysis is that the region of best predictors gradually shifts upstream along the slope with increasing lead time, ultimately reaching the Tail of the Grand Banks (Figure 7). The evolution of the maximum prediction skill pattern confirms our interpretation of the generalized persistence model for the NES in Section 4.2, i.e. the nonlocal predictability is associated with the subsurface equatorward advection of the Labrador slope waters into the NEC along the continental slope.

**Figure 7.** Spatial patterns of the maximum prediction skill when each grid point is used as a predictor in the generalized model for the target box 08-NEC (in cyan boundary) at a lead time of (a) 1 month, (b) 4 months, (c) 7 months, and (d) 10 months. Note that the maximum skills are chosen among the predictions for 12 forecast months at the given lead time for each grid point. In each panel, locations that have prediction skills *not significant* at 95% level are masked with gray dots. For those locations that have prediction skills above 95% level, but whose BTA negatively correlated with the target region, are covered with cyan dots (where are discarded when picking the best predictor). Bathymetry contours of 50, 200, 1000, 2000, and 4000 m are plotted as solid black lines, and the 200 m contour is thickened.
As expected, prediction of BTAs in the DGoM (Figure S5) exhibit patterns similar to those described for the NEC, with connections to predictors in the deep basins on the SS and over the slope off the SS. However, the prediction skill in these nonlocal regions is lower than that of local persistence. Additionally, BTAs on GB (box 05), in the NEC (box 08), and in the NSGoM (box 09) also exhibit significant prediction skill for BTAs in the DGoM as suggested in Section 4.2, but the ACCs are slightly less compared to the local persistence.

Unlike the DGoM and NEC regions, the best nonlocal predictors for WSGoM (box 06) are mostly located along the shallow shelf coastal boundary current regions, including NSGoM (box 09), and southern Gulf of St. Lawrence (SGSL; box 13) (Figure 8). Interestingly, the spatial patterns of maximum prediction skill for WSGoM looks mutually exclusive to those for the NEC at shorter lead times (Figure 7), indicating two totally different source waters for the two regions. Source waters for the WSGoM can be traced to the fresh coastal waters along the inner-shelf off Nova Scotia. However, advection influences cannot explain the pattern of high prediction skill in the downstream shallow regions like GB, Nantucket Shoals, and inner MAB shelf. The covariability with the WSGoM, as well as disappearance of the advective signal at longer lead times, suggest that the BTAs in WSGoM are dominated by atmospheric forcing, which also affects the downstream regions at the same time.

Figure 8. As in Figure 7, but for the WSGoM.

The best nonlocal predictors for the MAB and GB regions include the shallow shelf
regions in the MAB, GB, GoM, and SS. For example, the entire MAB shelf and GB exhibit very high skill at shorter lead times for predicting BTA in the NMAB (Box 03) (Figure 9a-b), consistent with the advective pathways for shelf waters as well as their coherent temperature variations. In addition, at longer lead times, the most skillful predictors for BTAs in the NMAB and GB include the slope region offshore of GB (slope box B) (Figure 9b-d), suggesting the importance of GS meanders [e.g., Andres, 2016; Gawarkiewicz et al., 2012] and rings [Chen et al., 2014; Zhang & Gawarkiewicz, 2015]. The highest prediction skill obtained in the slope region off the GB is mainly for fall forecast months (September to December) with a lead time of 4-9 months (not shown). This indicates that spring and summer subsurface temperature anomalies on the slope are associated with BTAs on the NMAB in the following fall. This is consistent with the peak months for WCR formation in the slope sea, occurring from May-July [Gangopadhyay et al., 2019].

Figure 9. As in Figure 7, but for the NMAB.

5.2 Predictability of the Large-scale Indices

Among the three large-scale indices, the T200-based GSI provides the most skillful predictions for the NES bottom temperature, with specific forecast and initialization months. For example, the T200-based GSI initiating in April produces the most skillful forecasts of BTA in nearly all of the NES subregions, except SMAB and WSGoM. This is very interesting since it suggests that there is a common response of bottom temperature among these NES subregions to
the GS meridional shift in a particular month. To visualize the spatial distribution as well as to
examine any subjectivity associated with the definition of the shelf and slope boxes, we
calculated the prediction skill when the T200-based GSI in April is used as a predictor for each
spatial point in this region, at lead times of 1, 4, 7, and 10 months (Figure 10).

Spatial maps suggest the WSS consistently has the most significant prediction skill
produced by the April GSI, across all lead months. At 1-month lead (i.e. for the target forecast
month of May), mainly the outer-shelf (>60-m isobath) from the SMAB to WSS and a small area
of the slope off WSS exhibit notable response to the April GS variability, with a maximum
prediction skill of 0.82 on the WSS (Figure 10a). Three months later for the forecast month of
August (GSI lead by 4 months), prediction skills of BTA are enhanced in the WSS and adjacent
slope with high skill expanding into the deep GoM basins, while in MAB the enhanced skill
becomes insignificant moving onshore (Figure 10b). High skill in the GoM and WSS continues
and affects the GB and NMAB at 7-month lead prediction (Figure 10c). For forecasts of shelf
BTA in February (GSI lead by 10 months), GSI-related prediction skill is largely weakened, and
the WSS and Jordan basin are the only two regions that still have significant response to the April
GS variability. This is not surprising since vertical mixing strongly influences BTA during
winter months. In contrast, winter mixing does not directly influence the BTA in the deep basins
of the GoM and WSS, and the forecast skill is higher in these regions (Figure 10d).

Over all, the T200-GSI-related predictability exhibits distinct spatial and seasonal
preferences. In particular, summer-to-fall (August to November) BTA on the WSS, DGoM, and
GB have the most notable response to the GS meridional shift in spring. The fact that the
generalized prediction model already incorporates these regions in the pathways of the GS-
related signal as the nonlocal predictors explains why the T200-GSI does not provide additional
significant improvement over the original generalized persistence model.

Other large-scale indices, e.g. NAOI, are also reported to provide predictability for
temperature on the NES, however, on longer time scales and over different time periods. Xu et al.
[2015] found that the NAOI is negatively correlated ($r=-0.47$) with SST in the GoM at a lead
time of four years based on the NOAA Optimum Interpolation SST version 2 dataset over the
period of 1982 to 2010, indicating the NAO signal takes about four years to propagate from the
Labrador Sea to the GoM. However, this statistical relationship no longer exists over the period
of this study (1993-2018) using the same datasets, which indicates the non-stationarity of the
NAO impact on the NES temperature over the recent decades. Besides, Mountain [2012] found
that the percentage of Labrador Slope Water entering the GoM at depth through the NEC is
negatively correlated with the NAOI with a lag of 2 years, based on temperature/salinity
measurements for the period of 1964-2008. Again, our longer-term (1-5 lead years) analysis over
the period of 1993 to 2018 does not suggest any statistically significant relationships between the
NAOI and BTA in either GoM or NEC (not shown).

We also tested the predictability of annual-mean BTA on the NES using the annual
AMOC index [Caesar et al., 2018] over the period of 1993 to 2018. Results suggest only the
NEC BTA has negative correlations ($r=-0.56--0.59$) with the AMOC index at lags of 0-2 years
(not shown). However, the prediction skill measures below the 95% confidence level following
our assessment method. In addition, the predictability associated with the AMOC index may also
be correlated with those from the GSI or the nonlocal upstream shelf/slope boxes.
Figure 10. Similar to Figure 7, but for spatial patterns of prediction skill when the T200-based GSI initiating in April is used as a predictor in the generalized model for each spatial point at a lead time of (a) 1 month, (b) 4 months, (c) 7 months, and (d) 10 months.

5.3 Prediction Skills with Linear Trends

The statistical models developed in this study are all based on linearly detrended BTA time series, in order to avoid spurious correlations due to linear trends. However, since the NES region has experienced notable warming over the past few decades [e.g., Forsyth et al., 2015; Pershing et al., 2015; Chen et al., 2020], we could also expect some prediction skill to result from this trend. Therefore, we have added the observed linear trend back into the time series predicted by the generalized persistence model (Section 4.2) in order to quantify the increase in prediction skill due to long-term warming. Results suggest that including the linear trends increases the prediction skill by roughly 0.2 in the GB and NEC regions, 0.15 in the DGoM and WSGoM regions, and about 0.1 in the NMAB and SMAB regions, and these increases have distinct seasonal preferences (Figure S6). For example, the prediction skill in the SMAB is improved by about 0.35 for October forecasts, but results in almost no improvement in March. Similarly the prediction skill for the GB/NEC is improved by roughly 0.25 in winter/fall, while the prediction skill for summer BTAs only improves by about 0.1 (Figure S6a). The skill gained from including the linear trend is larger at longer lead times, increasing by up to 0.3 at 12 month lead for GB (Figure S6b). When the linear trend is included, the prediction skill from the generalized prediction model reaches 0.52 in the SMAB to 0.78 in the DGoM on average.
5.4 Simplification of the Generalized Prediction Model

The most skillful predictions from the generalized model involve a variety of nonlocal predictors which perform best at different target months and lead times. However, in order for the model to be useful to end-users, it is desirable to simplify the model to include fewer predictors that can produce reasonable predictions for most target months and lead times. By combining information from Sections 4.2 and 5.1, we are able to identify 1-2 best nonlocal predictors to be used in addition to local persistence for a simplified version of the generalized persistence prediction model (Figure 11). For example, for the NEC we identify the combined slope region offshore of the SS (slope region C+D) in addition to local persistence. The selection rule is to keep only those physically reasonable predictors that maximize the generalized persistence. The chosen nonlocal predictors for each subregion are: SGSL (box 13) for SMAB, ESS (box 12) for NMAB, NGSL (box 14) for GB, SGSL (box 13) for WSGoM, ESS (box 12) for DGoM, and the combined slope C and D for NEC. The seasonal prediction skill for the simplified prediction model in each subregion and the difference relative to the original generalized persistence model are shown in Figure 11. The smallest reduction in skill is observed in the DGoM and NEC (~0.2) while the largest decreases are observed in the simplified NMAB model, especially in the upper left corner (behind the diagonal predictability barrier) of the prediction skill matrix (Figure 11e). Although the simplified model loses some skill, it still provides significant improvements over the local persistence, and it is easy to adapt.

5.5 Comparison with SST Predictability

We apply the same statistical models to the seasonal prediction of SSTA on the NES. Results suggest that the local persistence of SSTA is much more limited in time (1-2 months in the MAB regions and 2-4 months in the GoM regions) compared with BTA, and with fewer regional differences (Figure S7). This is presumably due to the strong influence of atmospheric forcing, which is correlated across larger spatial scale than oceanic processes and is less persistent. Another difference is that the SSTA prediction skill pattern is more irregular than that of BTA. For some forecast months in late summer or fall, prediction skill is not significant even at a lead time of one month. Further, the prediction skill abruptly decreases with increasing lead time over most of the year, and isolated peaks appear frequently at 6-11 months lead. These irregular patterns highlight the complex physical processes, e.g., freshwater input from rivers and atmospheric variability, that influence SST on the NES. Among all eight regions, the GB has the longest local persistence (up to ~5 months) and the northern Shelfbreak (NSB) is second, however, because of the lack of tidal mixing and the challenge in resolving shelfbreak dynamics in the reanalysis dataset, the result for these two regions should be considered with caution.

As with the BTA predictions, considering predictors of SSTA from other regions does improve the prediction skill, especially for predictions in the GoM in winter and fall (not shown). However, the increased prediction skill (by ~0.2 on average) is not comparable to the skill achieved in the BTA models and the improved skill patterns are still irregular.
Figure 11. As in Figure 7, but for the simplified generalized persistence model (only selecting one or two best nonlocal predictors for each NES subregion: box 13-SGSL for SMAB, box12-ESS for NMAB, box14-NGSL for GB, box13-SGSL for WSGoM, box12-ESS for DGoM, and the combined slope C and D for NEC) and the difference from the original generalized persistence model with values meaning the decreased prediction skills in comparison to the original generalized persistence model.

Like the BTA models, SSTA prediction models show a distinct response to the T200-based GSI, and very limited response to the other two indices (SSH-based GSI and the NAOI), although the prediction skill patterns are different (not shown). Among all subregions, SSTA in the GB and NEC regions have the most significant response to the T200-based GSI, mainly in the late summer-fall forecast months at 2-9 months lead, largely consistent with the BTA prediction in the same regions. By comparison, predictions based on the T200 GSI in the WSGoM and DGoM are significant in September and November only. Unlike the BTA
prediction in the SMAB and NMAB, SSTA has almost no predictability from the T200-based
GSI. Similarly, incorporating the T200-GSI as an extra predictor in the generalized persistence
model does not improve the prediction skill significantly at the 95% confidence level.

Overall, SSTA on the NES exhibit much lower seasonal predictability than the BTA,
regardless of the statistical model chosen. Future work exploring prediction skill from other
predictors, e.g., air-sea heat fluxes, would be beneficial.

5.6 Living Marine Resource Applications

While great progress has been made to incorporate environmental factors into fisheries
management and in particular, stock assessments, operationalizing this process has remained
equivocal because of some key challenges. A critical component of fisheries stock assessment
models are the short-term (i.e. 1-3 years) population forecasts and catch advice for management.
Operationalizing an assessment model with an environmental effect requires that (1) the
relationship between the oceanographic variable and a population process must be established,
(2) the oceanographic variable must also be forecasted with reasonable skill and (3) have
sufficient lags to be useful for management advice. The degree of predictability of the
oceanographic variable needed to improve management advice from an assessment model also
depends on the spatio-temporal scale of the oceanographic variable and intrinsic biological
factors (e.g., strength of relationship, life history, population process affected, stock status,
exploitation history, length of time-series, and initial value of the oceanographic variable; Haltuch and Punt, 2011), and therefore is case-specific. Great progress has been made in many
regions understanding the effects of environmental variables on population-level processes
particularly with the advent of ocean observation technology, and the Northeast U.S. is no
exception (Table S1). Here, prediction skill ranged from 0.2-0.9 for specific combinations of
region, forecast month, lead time, and fish stock that could be used for Northeast U.S. stock
assessments (Table S1, Figure 5). In general, there has been a critical need for predictions of
bottom variables, and greater lead times for bottom temperature predictions that the work above
provides.

Two types of lags commonly occur that may increase the feasibility of conducting
environment-linked fisheries assessments by reducing the lead time necessary for predictions.
First, there is often a lag between the terminal year of fish population data and when the
assessment is conducted. This arises because age-based assessments rely on labor-intensive
sampling of fishery and survey catch to determine fish age, weight, and maturity. In contrast, in
situ and remotely sensed physical oceanographic data can often be processed in near real-time,
and therefore can be collected much closer to the assessment date. As an example, the Southern
New England-Mid Atlantic winter flounder assessment conducted in September 2020 used
biological data through 2019 [NEFSC 2020]. The temperature-recruitment relationship depends
on January-March estuary bottom temperature, which means that the temperature in the first
projection year, 2020, has already been observed by the assessment date. The January-March
estuary bottom temperature in the second projection year, 2021, would need to be forecasted
using data through August 2020. Thus, a 5-7 month bottom temperature forecast enables a 2-year
population projection (Table S1). Second, there is also often a lag between the oceanographic
variable and its effect on the population process. For example, recruitment in year t frequently
depends on the spawning biomass and environmental conditions in the previous year, t-1. These
lags will be case-specific, and therefore the performance of predictions from an environment-linked assessment should ideally be tested using simulations in a management strategy evaluation [Punt et al., 2014; Gaichas et al., 2016; Punt et al., 2016].

Early ecological forecasts based solely on lags between the oceanography and ecology and that were constrained to using SST were hindered by lead times of near real time to ~4 months [Payne et al., 2017]. Prediction products like those developed here can potentially extend lead times to 12 months in some seasons and locations on the NES. In addition to the substantial advance in improving fisheries stock assessment, ocean temperature predictions can also be used to make and improve ecological forecasts. Existing ecological forecasts that predict opening of fisheries [Mills et al., 2017], support dynamic spatial management [Dunn et al., 2016; Maxwell et al., 2015] and that aim to reduce marine mammal and other non-target species bycatch [Howell et al., 2008, 2015; Hazen et al., 2018; Hobday et al., 2019] could potentially be more effective with prediction products developed herein.

Acknowledgments, Samples, and Data

This work was supported by NOAA's Climate Program Office's Modeling, Analysis, Predictions, and Projections (MAPP) program (NA17OAR4310111, NA19OAR4320074, and NA20OAR4310482). We acknowledge our participation in MAPP's Marine Prediction Task Force. The GLORYS12v1 ocean reanalysis dataset is available at the Copernicus Marine Environment Monitoring Service (CMEMS) (https://resources.marine.copernicus.eu/?option=com_csw&view=details&product_id=GLOBAL_REANALYSIS_PHY_001_030). The NOAA NEFSC hydrographic dataset is in the World Ocean Database maintained by NOAA National Centers for Environmental Information (https://www.ncei.noaa.gov/products/world-ocean-database). The SSH data are available at CMEMS (http://marine.copernicus.eu/services-portfolio/access-to-products/?option=com_csw&view=details&product_id=SEALEVEL_GLO_PHY_L4_REP_OBSERVATIONS_008_047). The subsurface temperature EN4 dataset is available at the Met Office Hadley Centre (https://www.metoffice.gov.uk/hadobs/en4/). The monthly NAO index is obtained from the National Weather Service Climate Prediction Center (https://www.cpc.ncep.noaa.gov/products/precip/CWlink/pna/nao.shtml). The data of NOAA Optimum Interpolation SST used in this study are available at NOAA Earth System Research Laboratory (https://www.esrl.noaa.gov/psd/data/gridded/data.noaa.oisst.v2.highres.html).

References


Dunn, D. C., S. M. Maxwell, A. M. Boustany, and P. N. Halpin (2016), Dynamic ocean management increases the efficiency and efficacy of fisheries management. *Proceedings of the National Academy of Sciences*, 113(3), 668-673.


Howell, D. C. (2009), Statistical methods for psychology, Nelson Education.


Lellouche, J.-M., E. Greiner, O. Le Galloudec, G. Garric, C. Regnier, M. Drevillon, M. Benkiran, C.-E. Testut, R. Bourdalle-Badie, and F. Gasparin (2018), Recent updates to the
Copernicus Marine Service global ocean monitoring and forecasting real-time 1/12° high-resolution system, Ocean Science, 14(5), 1093-1126.


McHenry, J., H. Welch, S. E. Lester, and V. Saba (2019), Projecting marine species range shifts from only temperature can mask climate vulnerability, Global change biology, 25(12), 4208-4221.

Miller, T. J., J. A. Hare, and L. A. Alade (2016), A state-space approach to incorporating environmental effects on recruitment in an age-structured assessment model with an application to southern New England yellowtail flounder, Canadian Journal of Fisheries and Aquatic Sciences, 73(8), 1261-1270.


References From the Supporting Information


Atlantic States Marine Fisheries Commission, Washington, D.C., USA

Bell, R.J., A. Wood, J. Hare, D. Richardson, J. Manderson and T. Miller (2018), Rebuilding in the face of climate change. Canadian Journal of Fisheries and Aquatic Sciences, 75(9), 1405–1414.


