JGR Oceans

RESEARCH ARTICLE
10.1029/2021JC017187

Special Section:
Advancing prediction of coastal marine ecosystems

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

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

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

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 exhibits 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 data set. A simple, damped local persistence prediction model produces significant skill for lead times up to ~5 months in the Mid-Atlantic Bight and up to ~10 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 are located. In this study, we have developed a collection of statistical models that produce seasonal predictions of NES bottom temperature with 1–12 months lead time. Variables considered in these prediction models include local persistence of bottom temperature from prior months, bottom temperature from an upstream or nearby region, and large-scale atmospheric and oceanic indices representing the North Atlantic Oscillation or position of the Gulf Stream (GS). Only considering local persistence provides significant skill for lead times up to ~5 months in the Mid-Atlantic Bight and up to ~10 months in the Gulf of Maine, although the skill varies by season. Using upstream or nearby bottom temperature and the GS index both generally improve the prediction skill. However, the GS index does not provide higher prediction skill than those upstream or nearby bottom temperatures. A simplified statistical model has been developed, which can be tailored to fisheries management on the NES.

Key Points:

- A local persistence model for bottom temperature provides prediction skill up to ~10 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

Supporting Information:

Supporting Information may be found in the online version of this article.

Correspondence to:

Z. Chen,
zchen@whoi.edu

Citation:

Chen, Z., Kwon, Y.-O., Chen, K., Fratantoni, P., Gawarkiewicz, G., Joyce, T. M., et al. (2021). Seasonal prediction of bottom temperature on the northeast U.S. continental shelf. *Journal of Geophysical Research: Oceans*, *126*, e2021JC017187. https://doi.org/10.1029/2021JC017187

Received 15 JAN 2021
Accepted 25 APR 2021

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

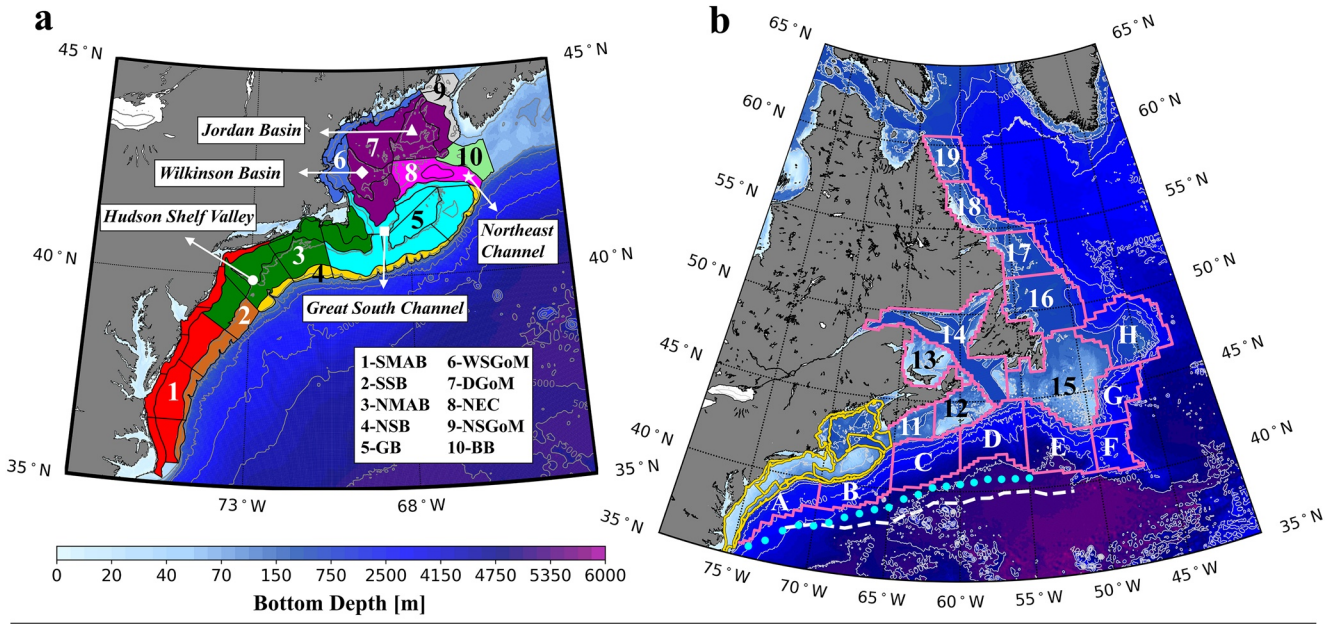

Figure 1. The Northeast U.S. continental shelf (NES) and the Northwest Atlantic (NWA) continental shelf and slope. (a) The Ecosystem Monitoring strata (47 black polygons) and the grouped 10 subregions (shaded in different colors) on the NES. Names of these 10 subregions are listed below, together with additional NWA shelf and slope boxes. Select geographical features are also annotated. (b) The boundaries of a new set of shelf (box01–19) and slope boxes (A–H) are shown in pink and yellow (target 10 subregions). Labels corresponding to the 10 newly clustered subregions (box01–10) are shown in (a). The white dashed line represents the average position of the sea surface height (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).

(Chapman & Beardsley, 1989; Fratantoni & Pickart, 2007; Loder, 1998; Richaud et al., 2016; Townsend et al., 2006). 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 (G. Gawarkiewicz et al., 2018; Greene et al., 2013; Mountain, 2012).

The NES ecosystem structure and function has experienced significant changes as a result of climate-scale changes in the physical environment (Greene & Pershing, 2007; Link et al., 2002; Mills et al., 2013; Nye et al., 2009). Shifts in marine species distribution, growth, and recruitment have been directly linked to changes in ocean temperature (e.g., Kleisner et al., 2016; Miller et al., 2016, 2018; Pinsky et al., 2013). The warming trend of ocean temperature further drives oxygen decline, which increases additional physiological stress on fish and benthic invertebrates (Brennan et al., 2016). Therefore, changing ocean temperature on the NES is an important factor that needs to be considered in marine resource management (e.g., Nye et al., 2011; Sullivan et al., 2005; Tommasi, Stock, Hobday, et al., 2017).

Variability in bottom temperature on the NES is strongly influenced by local oceanic processes, for example, 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., Chen & Curchitser, 2020; Chen et al., 2018; Houghton et al., 1982; Lentz, 2017). In the GoM, temperature variations at depth are influenced by two different processes: (a) advective inflow of relatively fresh water from the Scotian Shelf (SS) and its modification through winter convective mixing (Mountain & Manning, 1994) and (b) 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 (Greene et al., 2013; Ramp et al., 1985; Smith et al., 2001). Besides these advective processes, bottom waters on the NES are also influenced by cross-shelf exchange processes associated with GS warm-core rings (WCRs) and meanders. For example, an observed increase in the number of WCRs 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; Gangopadhyay et al., 2019; G. Gawarkiewicz et al., 2018).

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 shelfbreak south of New England on shorter time scales (G. G. 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 shelfbreak, 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 & 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; Peña-Molino & Joyce, 2008; Saba et al., 2016; Zhang & Vallis, 2007). 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 4 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; McHenry et al., 2019; Selden et al., 2018) 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 based on coarse-resolution (~0.5°–2°) global models exhibits in general more reliable skill for basin-scale variations of surface temperatures on interannual scales, for example, the El Niño–Southern Oscillation (Stock et al., 2015), while shelf- and/or slope-scale dynamics are not well captured due to the underrepresented coastal processes (e.g., freshwater flux, instabilities, and tides) primarily as a result of their coarse resolution (Stock et al., 2011). Recent studies have found that among all North American Large Marine Ecosystems, 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., 2019, 2020; Stock et al., 2015). In addition, due to the complexity of shelf- and slope-scale dynamics in this region, coarse-resolution climate models cannot simulate the bottom temperature realistically (Saba et al., 2016). Hence, there is a need for skillful predictions of ocean temperature on the NES using alternative approaches, for example, 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 data set 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, GB; box06, western shallow GoM (WSGoM); box07, deep GoM (DGoM); box08, 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 (BTAs) calculated for each EcoMon strata based on the monthly GLORYS12v1 ocean reanalysis data set (1993–2018; Section 2.2) are used for the cluster analysis, after removing the long-term linear trends (computed over the period of 1993–2018).

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 linearly detrended BTA is calculated for each of these newly generated shelf and slope boxes by removing the monthly climatology and linear trends (computed over the period of 1993–2018), and these anomalies will be considered as potential predictors in the prediction models as explained in Section 3.2.

2.2. The GLORYS12v1 Ocean Reanalysis Data Set

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 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 (Cabanes 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 (SSH), mixed layer depth, and ice parameters are archived at daily and monthly interval, available at the CMEMS.

We have compared the GLORYS12v1 reanalysis data set 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 (box09 and 10; see blue dots in Figure S1), neither evaluation of the reanalysis data set nor investigation of predictability in these regions is performed. Results suggest that the GLORYS12v1 data set is highly consistent with observations over the NES on both seasonal and interannual time scales (Text S1 and Figure S2). Hence, we use the bottom temperature from the GLORYS12v1 data set over the shelf regions

of the NWA together with the temperature at ~200 m depth over the slope and open ocean (35°N–65°N, 40°W–80°W). As discussed below, the exchange of water masses between the shelf and slope is mainly associated with boundary current variability such as GS meanders and WCRs. Temperature at ~200 m depth captures these processes well and 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 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 $i = 1,2,\ldots,8$, can be written as

$$\widehat{\mathrm{BTA}_i}(t) = \alpha \cdot \mathrm{BTA}_i(t-\tau) + c, \tag{1}$$

where t represents the target forecast month and τ is the lead time. This equation illustrates that BTA of a subregion could be approximated by its own historical BTA initialized from the month of $(t-\tau)$. The coefficients α 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, α and c are calculated via a k-fold cross-validation method, described in Section 3.5. Therefore, α and c are not constants but vary slightly between cross-validation segments.

3.2. The Persistence-Advection Model

In the local persistence model, the predictor is the BTA from the same box at the initialization month (τ months before target forecast month t). 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 persistence-advection model rewritten as

$$\widehat{\mathrm{BTA}_i}(t) = \alpha \cdot \mathrm{BTA}_j(t-\tau) + c, \tag{2}$$

where the $\mathrm{BTA}_j(t-\tau)$ represents the initialization at the month of $(t-\tau)$ using BTA from subregion j, and j is not necessarily equal to i. Similar to the local persistence model (Section 3.1), the whole k-fold cross-validation process is applied to all combinations of subregion i, j, forecast month t, and lead time τ. For a particular target predictand subregion i at the forecast target month t, any shelf and slope boxes (01–19 and A–H) at τ months lead can be chosen as the predictor. Among these 27 choices, the one that gives the highest prediction skill based on the anomaly correlation coefficient (ACC; Section 3.6) for each target month t and lead time τ 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.

One thing to note is that before settling on the model presented here, a multiple linear regression model was also considered by treating the nonlocal BTA as a separate additional predictor to the local persistence (cf., Jacox et al., 2019). This type of model forces the choice of a nonlocal predictor based on its relationship with the residual from local persistence. However, in most cases, the nonlocal predictor chosen by the model did not make physical sense, especially at shorter lead times when the local persistence term dominates.

Therefore, instead of a multiple regression model, our prediction model is developed in the form of a univariate regression model by combining the local persistence and advective BTA from adjacent regions into a single 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

$$\widehat{\mathrm{BTA}}_i(t) = \alpha \cdot X(t-\tau) + c; \quad X = \mathrm{GSI\ or\ NAOI}, \tag{3}$$

where X representing the GSI or NAOI.

We consider two different GSIs, both tracking meridional shifts in the GS position on monthly time scales from 1993 to 2018 (Figure S3). First, a GSI is calculated based on SSH following Pérez-Hernández and Joyce (2014), reflecting the position variability of the GS main axis between 72°W and 52°W based on the monthly mean ¼° satellite-based gridded SSH from the CMEMS. The other GSI is calculated based on the subsurface temperature at 200 m depth (T200) following Joyce et al. (2009), reflecting the position variability of the GS North Wall. Here, we use the EN4 quality-controlled gridded subsurface temperature (1° horizontal resolution; available from 1950s to present) (Good et al., 2013), instead of temperature profiles from the World Ocean Database as was used in Joyce et al. (2009). A comparison of the two temperature-based GSIs is included in Text S2 and Figure S3. The monthly NAOI, which is based on a rotated EOF analysis (Barnston & Livezey, 1987), is obtained from the National Weather Service Climate Prediction Center. The time period spans from 1993 to 2018 and a linear trend is removed from each of them prior to each of their use as a predictor in our statistical model.

3.4. The Persistence-Advection Model With GSI or NAOI

We combine the persistence-advection 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:

$$\widehat{\mathrm{BTA}}_i(t) = \alpha \cdot X(t-\tau) + c; \quad X = \mathrm{BTA}_j, \mathrm{GSIs, or\ NAOI}. \tag{4}$$

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 nonstationary, 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: (a) 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; (b) fit the prediction model to the reanalysis time series over the training time period, that is, obtain its parameters α and c through linear regression; (c) apply the prediction model to calculate a predicted time series over the testing time period; (d) change the testing time period to another segment, repeating steps (b) and (c); and (e) 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 ACC and root mean square error (RMSE) between the cross-validated prediction ($\widehat{\mathrm{BTA}}$) and GLORYS12v1 BTA time series, which are considered as functions of the target forecast month (t) and lead time in months (τ):

$$\mathrm{ACC}(t,\tau) = \frac{\sum_{i=1}^{n}\left(\widehat{\mathrm{BTA}}\cdot \mathrm{BTA}\right)}{\sqrt{\left(\sum_{i=1}^{n}\widehat{\mathrm{BTA}}^{2}\right)\cdot\left(\sum_{i=1}^{n}\mathrm{BTA}^{2}\right)}}, \tag{5}$$

$$\mathrm{RMSE}(t,\tau) = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\widehat{\mathrm{BTA}}-\mathrm{BTA}\right)^{2}}, \tag{6}$$

where $\widehat{\mathrm{BTA}} = \widehat{\mathrm{BTA}}(t,\tau)$, $\mathrm{BTA} = \mathrm{BTA}(t)$, 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 τ) 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, for example, the $\mathrm{BTA}(t)$ in our case (D. C. Howell, 2009).

In addition, we also evaluate predictions using the Brier Score (BrS), which is a measure of the average squared forecast probability error relative to the observed probability for a probability category over the time period (Brier, 1950). The BrS can also be expressed as a function of the forecast month (t) and lead time (τ):

$$\mathrm{BrS}(t,\tau) = \frac{1}{n}\sum_{i=1}^{n}\left(\hat{P}-P\right)^{2}, \tag{7}$$

where $\hat{P}$ and P are the predicted and observed probabilities of a warmer (or cooler) event, respectively, and n is the number of years in the time series. Values of the BrS range from 0 to 1, with smaller BrS indicating better model performance. Following Hervieux et al. (2019), the probability categories we considered in this study are the upper and lower terciles of the temperature distribution, that is, $P = 1/3$ for the both categories.

To assess the probabilistic prediction skills of a prediction model in comparison to the baseline model (e.g., the local persistence model), we also calculated the Brier Skill Score (BSS), which compares the BrS of a prediction model to that of the local persistence model:

$$\mathrm{BSS}(t,\tau) = 1 - \frac{\mathrm{BrS}}{\mathrm{BrS}_{\text{baseline model}}}, \tag{8}$$

where $\mathrm{BrS} = \mathrm{BrS}(t,\tau)$ and $\mathrm{BrS}_{\text{baseline model}} = \mathrm{BrS}_{\text{local persistence model}}(t,\tau)$. A negative value of the BSS indicates poorer performance by the prediction model relative to the baseline model. If the BSS is zero, then the two models perform similarly, whereas a positive BSS value indicates the new model performs better than the baseline model.

4. Results

4.1. Local Persistence

On the NES, the baseline damped local persistence predictions of BTAs yield significant prediction skill in ACC (at 95% confidence level) for lead times up to several months (Figures 2a–2h). Generally, northern NES boxes (box05–08) exhibit longer persistence and smaller RMSE, compared to the southern shelf boxes (box01–04; Figures 2 and 3). For example, the BTAs persist up to ~5 months in the NMAB (box03) and up to ~10 months in the DGoM (box07), and the mean RMSE across all forecast months and lead time is ~0.9°C

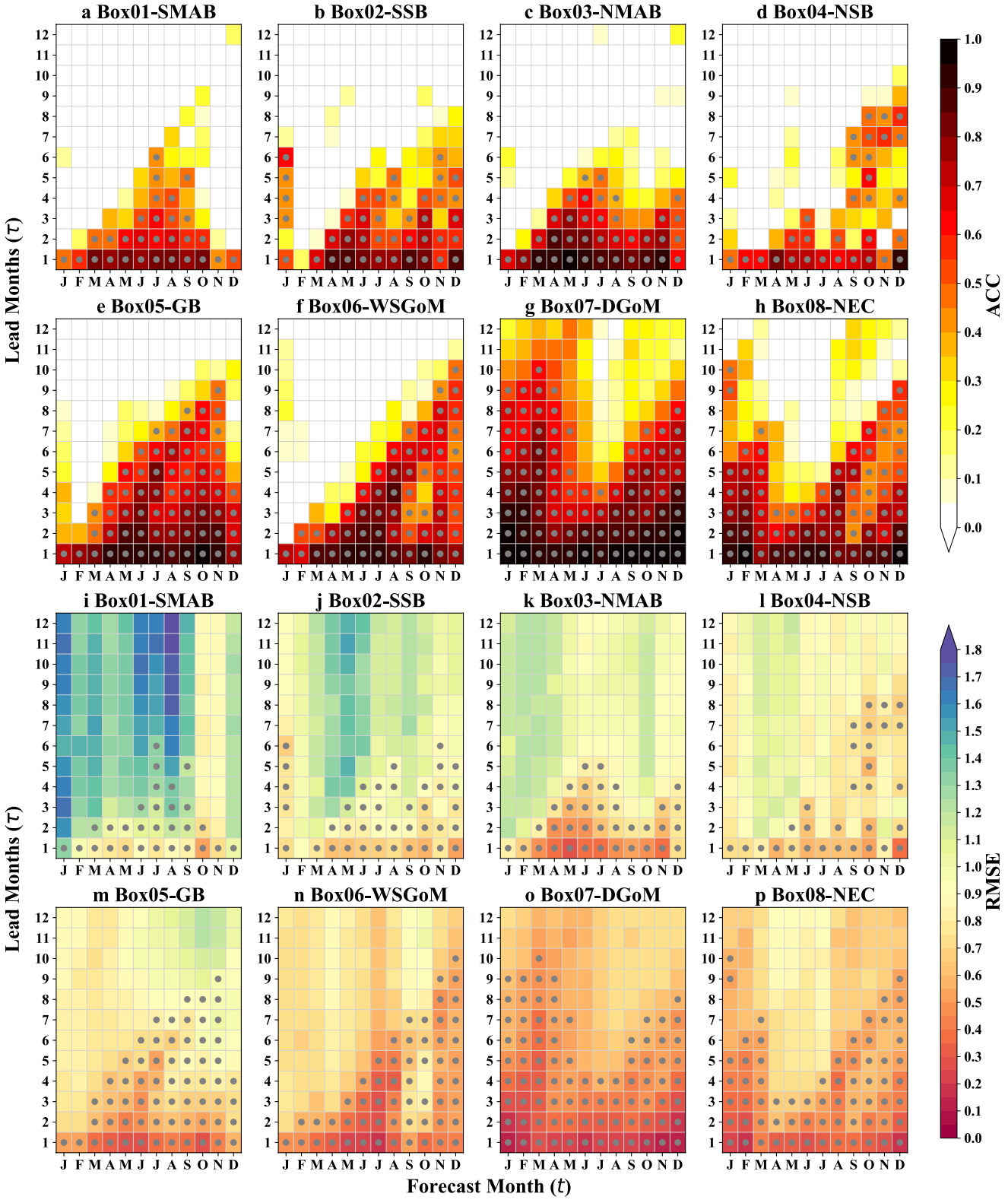

Figure 2. The prediction skill indicated by the anomaly correlation coefficient (ACC; a–h) and root mean square error (RMSE; i–p) of the damped local persistence model in each subregion (shelf box01–08) on the Northeast U.S. shelf. The *x*-axis represents target forecast month (t), and *y*-axis represents lead months (τ). The gray dots in each panel indicate that the ACC is significant at the 95% confidence level.

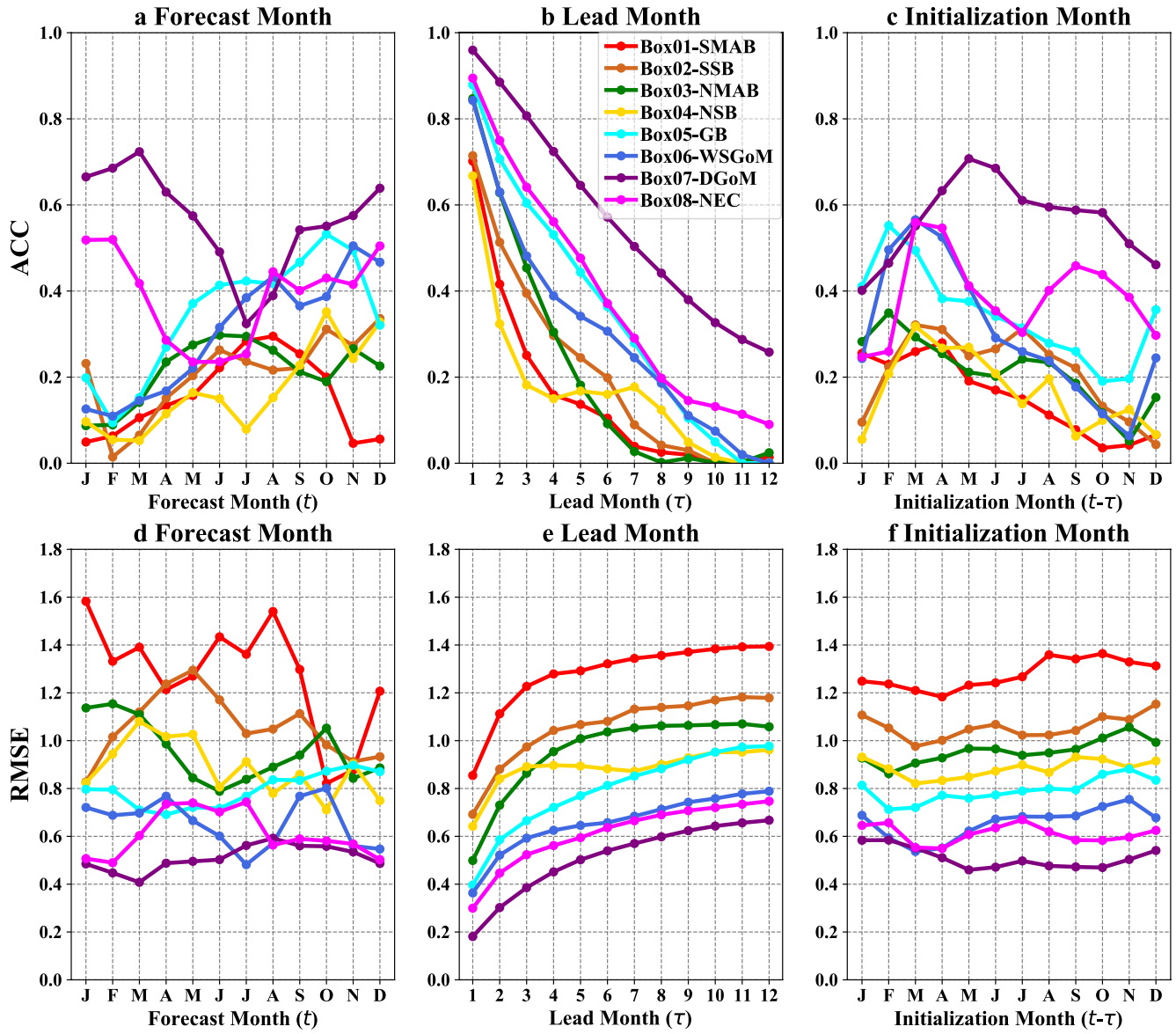

Figure 3. The prediction skill indicated by the anomaly correlation coefficient (ACC; a–c) and root mean square error (RMSE; d–f) of the damped local persistence model in each subregion on the Northeast U.S. shelf (represented by different colors), averaged for (a, d) each target forecast month (t), (b, e) lead time (τ), and (c, f) initialization month (t - τ) by averaging values in Figure 2 in the vertical, horizontal, or diagonal direction, respectively.

in the former and ~0.5°C in the latter. This difference is likely due to differences in the regional settings, for example, bottom depths, mixed layer depths, and proximity to the GS between the two regions.

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 box07 and 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 RMSE, the skill matrix in each subregion generally follows a seasonal pattern consistent with that for ACC (Figures 2i–2p), which suggests the higher the ACC, the lower the RMSE (Figure 3). Similarly, the probabilistic skills indicated by the BrS for upper and lower terciles exhibit the agreement in seasonal prediction patterns (Figures S4 and S5).

For the SMAB region (box01), 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 (Figures 2a and 2i), 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; J. 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–July. However, the persistence of BTA remains roughly the same (up to 3–4 months) in the following months (August-December) (Figures 2c and 2k), consistent with the establishment of the subsurface Cold Pool water (<10°C) typically from spring to fall (Chen et al., 2018).

The two shelfbreak regions (box02 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 (Figures 2b and 2d). As an example, the NSB region exhibits very limited persistence (1–3 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 (box05–08) exhibit generally longer persistence and smaller RMSE compared with the MAB subregions (Figures 2 and 3). These four subregions can be grouped into two categories based on the seasonal patterns of persistence. The first group includes the GB and WSGoM (box05 and 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, Figures 2e and 2f). 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. The prediction skill for forecasts in the fall (September-December) is not as consistently strong as the skill achieved in summer forecasts across multiple lead times, reflective of a transitional period between the prior summer and the following winter forecasts. The shift to shorter prediction time scales for winter forecasts (January-February) coincides with a similarly abrupt transition in the forcing over the region, that is, the onset of winter storms and cold outbreaks (e.g., J. Forsyth et al., 2018). In contrast, the DGoM and NEC regions both exhibit a shifted seasonality with minimum persistence in May and April, respectively (Figures 2g and 2h). It is likely that for these two deepest subregions, where winter mixed layers do not reach the bottom (based on the same GLORYS12v1 data set), the winter mixing affects the bottom temperature indirectly with delays, for example, through advection and convection from the shallower neighboring regions. Also, in comparison to the shallower regions, the breakdown 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 (Figures 2g and 2h).

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 (Figures 3b and 3e). On average, predictions in the DGoM based on local persistence gradually decrease (increase) in ACC (RMSE) from 0.96 (0.18°C) at 1-month lead to about 0.26 (0.67°C) for the 12-month lead (Figures 3b and 3e). Among all forecast months, winter months in DGoM have the highest prediction skill (ACC = 0.72, RMSE = 0.41°C in March) and summer months have the lowest prediction skill (ACC = 0.32, RMSE = 0.56°C in July; Figures 3a and 3d). 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 0.71 (RMSE = 0.46°C; Figures 3c and 3f).

The NEC region exhibits a similar BTA persistence pattern to the DGoM, although with decreased prediction skill and slightly shifted seasonality (Figures 2h and 2p). The highest prediction skill (ACC = 0.52, RMSE = 0.48°C) regardless of lead month is achieved in February, 1 month earlier than the DGoM region; and the lowest prediction skill (ACC = 0.24, RMSE = 0.55°C) is obtained 2 months earlier than the DGoM in May (Figures 3a and 3d). 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 March, which is also 2 months 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. The Persistence-Advection Model

The persistence-advection 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 persistence-advection model is identical to the local persistence model (i.e., $j = i$ in Equation 2). In other words, the persistence-advection 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 persistence-advection 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 persistence-advection model in all shelf subregions (Figure 4). On average, the NMAB shows the most improvement (~0.32) among the subregions, especially for October forecasts (~0.47, 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 (~0.11 on average), as the local persistence model already showed strong performance.

We find two types of predictability barriers (indicated by a line of minimum predictability), based on the prediction skill revealed by the persistence-advection model (Figures 5 and 6). One is along the diagonal, that is, dependent on the initialization month, and the other is aligned vertically, that is, dependent on the forecast month. Diagonal predictability barriers are visible in the NMAB, GB, and WSGoM (box03, 05, and 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 persistence-advection model (Figures 6b and 6e).

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 1 or 2 month(s) lead for nearly all forecast months (Figure 5a). For predictions with a 2–7 months lead, the neighboring SSB (box02), upstream NMAB and GB (box03 and 05), slope box F, and southern Gulf of St. Lawrence (SGSL; box13) are better predictors providing an increase of ~0.4 in prediction skill of ACC and an decrease of ~0.3°C in RMSE on average (Figures 5d and 5g). 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 (box19) (Figures 5d and 5g), with considerable improvement relative to local persistence also found in the BSS for lower and upper terciles (Figures 5k and 5m). 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 1 year for an anomaly signal to advect from the northern Labrador shelf to the SMAB (cf., 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 box12, ESS), and especially the slope regions B and C adjacent to the GB and SS, which exhibit the largest improvement with regard to ACC increase, RMSE decrease, and BSS values for both upper and lower terciles (Figures 5e, 5h, 5k, and 5n). Winter-spring forecasts (January-June) are also improved by the northernmost nonlocal BTA signals (Labrador shelf box18 and 19, slopes 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 (Figures 5c and 6a; also Figures 2e and 2f). Within these triangles, there is little improvement in prediction skill (<0.2) gained from nonlocal predictors, but

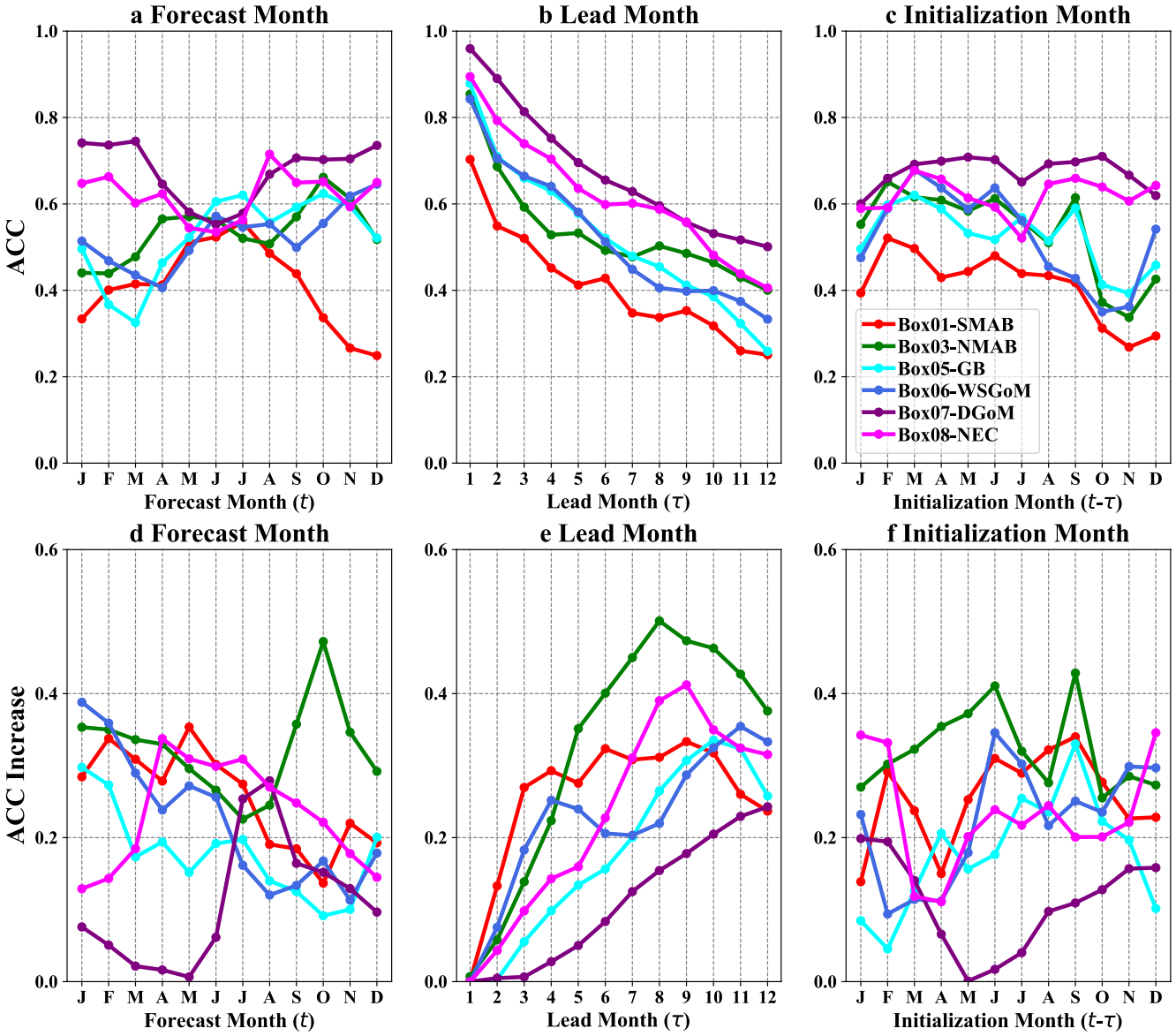

Figure 4. As in Figures 3a–3c, but for the prediction skills in anomaly correlation coefficient (ACC) of the persistence-advection model (a–c) and the prediction skill improvement (d–f) relative to the local persistence model in subregions of shelf box01, 03, 05, and 06–08 (represented by different colors),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

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, and ESS (box02, 04, and 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 (box07) and NEC (box09), and upstream GSL subregions (box13 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 (Figures 4 and 6). The largest improvements (~0.3 increase) are realized for summer and fall forecasts (August–December), when considering BTA in slopes C–E as nonlocal predictors (Figure 6e). 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 box11 (western SS) and slope

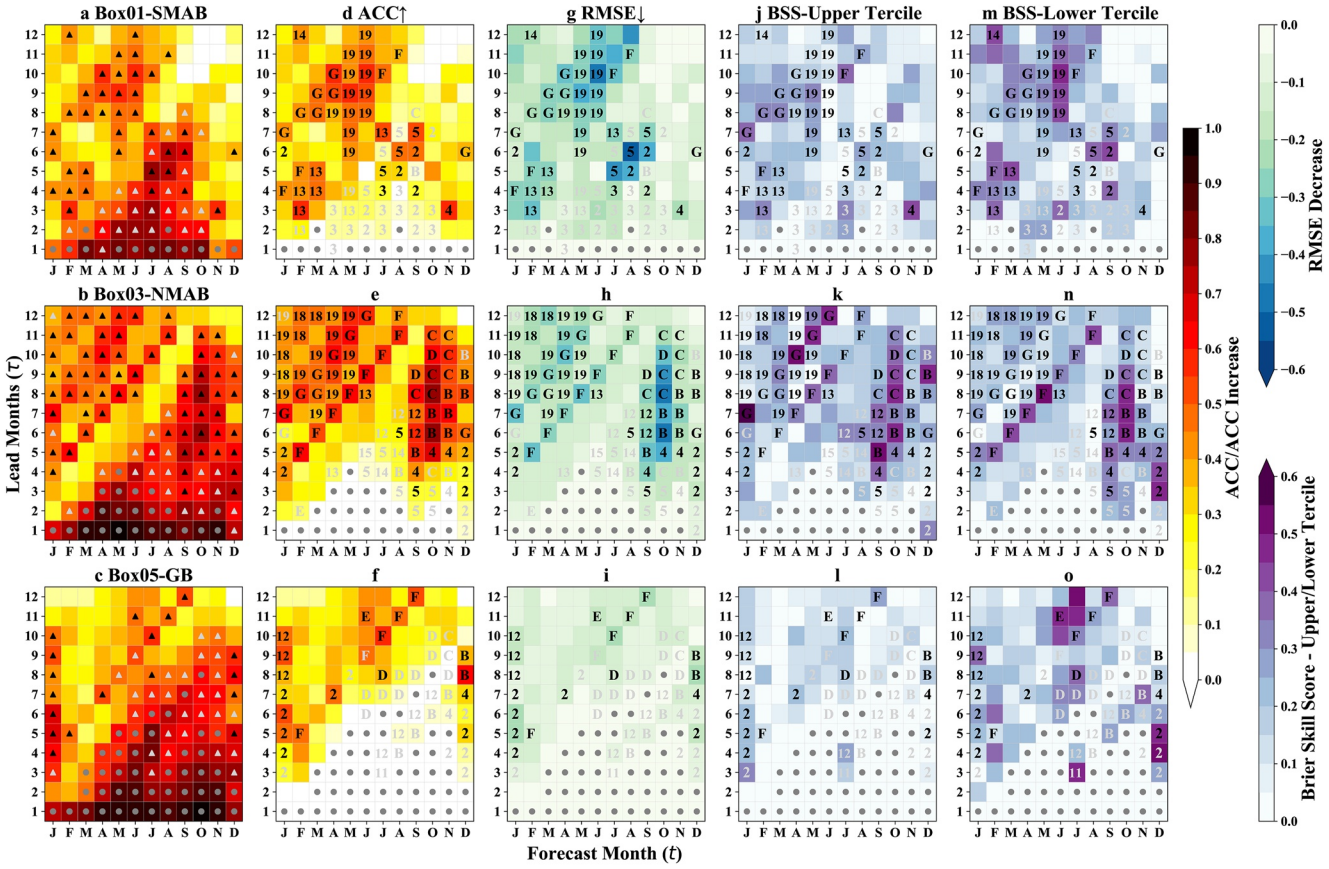

Figure 5. The prediction skill for the southern three subregions, that is, (top row) box01, southern Mid-Atlantic Bight (SMAB); (middle row) box03, northern Mid-Atlantic Bight (NMAB); and (bottom row) box05, Georges Bank (GB), represented by the anomaly correlation coefficient (ACC) prediction skill of the persistence–advection model (a–c), as well as the skill improvement relative to the local persistence model based on the ACC (i.e., persistence–advection minus local persistence) (d–f), root mean square error (RMSE; g–i), and the Brier Skill Score (BSS) for upper (j–l) and lower (m–o) terciles. The *x*-axis in each panel represents the forecast month (t) and *y*-axis the lead time (τ). The darker the color, the better the model performance in all panels. In panels (a)–(c), the markers (gray dots, light gray, and black triangles) indicate the ACC is significant at the 95% confidence level. In particular, the gray dot indicates the prediction skill of the persistence–advection model is identical to the local persistence model, while the triangle (either light gray or black) indicates that the persistence-advection model has higher prediction skill in ACC 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), 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 that the ACC improvement is significant at 95% confidence level, while those light gray numbers/characters are not. In panels (g)–(o), the gray dots and the predictor box numbers are based on the corresponding ACC, thus identical to those in (d)–(f).

boxes C–E (near the SS and the Tail of the Grand Banks) along the advective pathway of the slope currents (Figure 6f).

Overall, considering nonlocal BTA predictors in the persistence-advection 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 is further discussed in Section 5.1.

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 GS position (Figure 7). In the SMAB region, the T200-based GSI provides very limited prediction skill for all forecast months except the fall, that is, September and October, with lead time of 3–8 months. The values

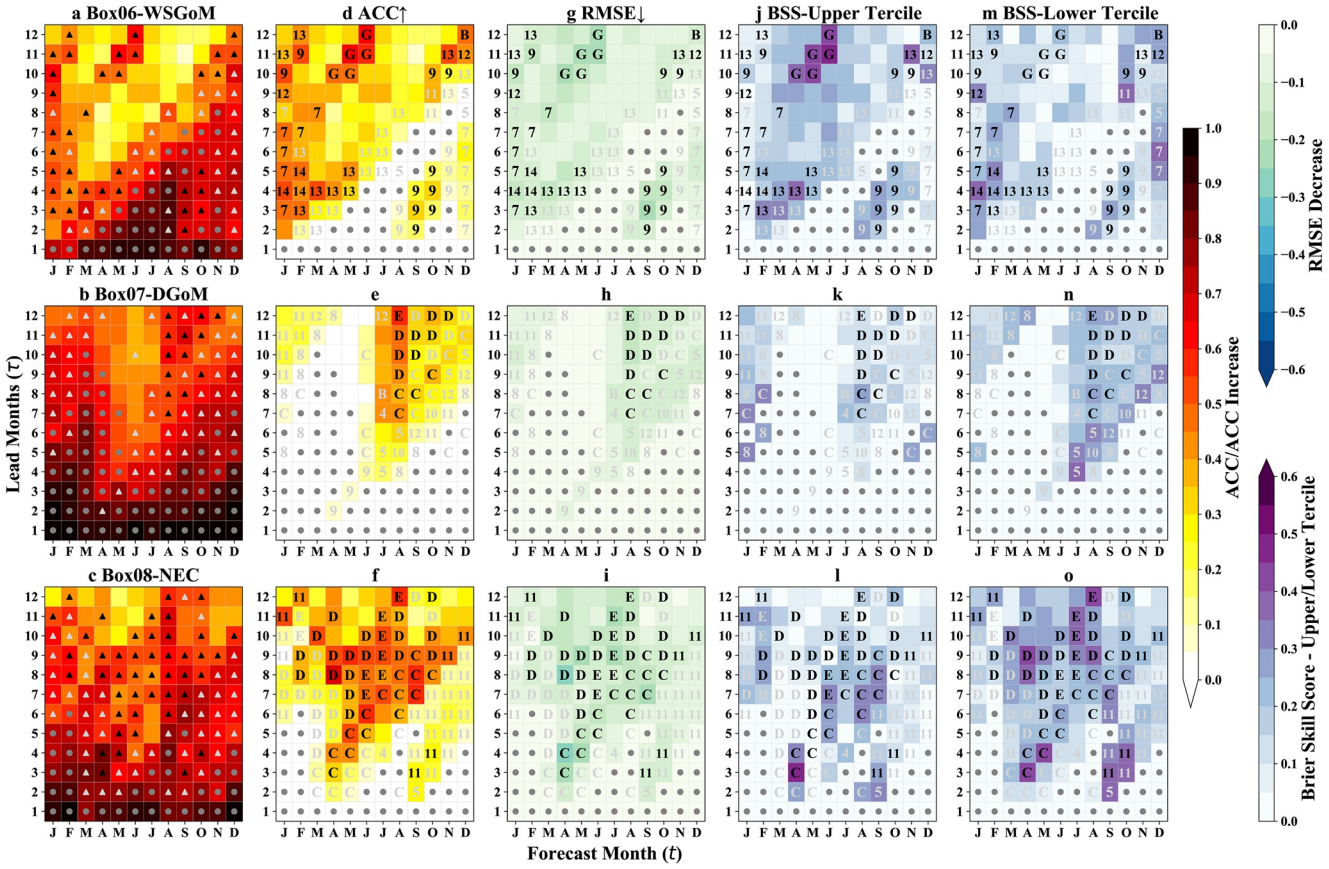

Figure 6. Same as Figure 5, but for the northern three subregions: (a) box06, western shallow Gulf of Maine (WSGoM); (b) box07, deep Gulf of Maine (DGoM); and (c) box08, Northeast Channel (NEC).

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 7a). In the NMAB region, the fall forecast months (September-November) also show significantly higher prediction skill relative to the local persistence model, especially with lead time of 4–10 months (Figure 7b).

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-month lead; Figure 6c). 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 GB (Figure 7c). This GSI signal initiating in April is also identified in the DGoM and NEC regions (Figures 7e and 7f). In contrast, BTA predictability in the WSGoM based on GSI is very limited (Figure 7d), 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 (Figures 7e and 7f). Significant increases in prediction skill are achieved by the GSI model for forecasts in the NEC region, when predicting the summer-fall (July-October) BTA at a lead of 7–10 months (Figure 7f).

We also tested two other indices—the SSH-based GSI and the NAOI for shelf BTA predictions (Figure S6). Interestingly, both of these indices provided very limited prediction skill for all forecast months at lead months less than 1 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

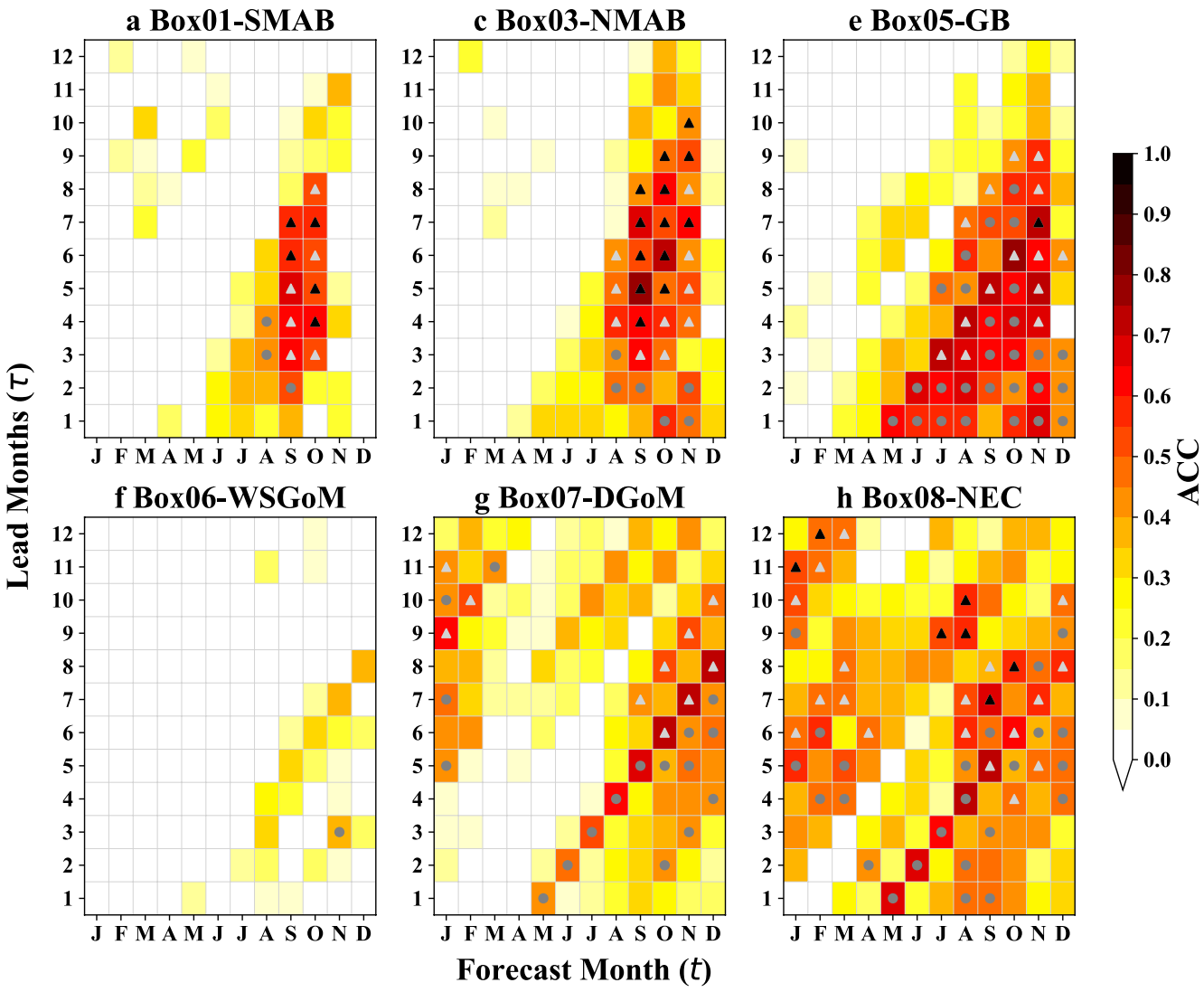

Figure 7. The prediction skill of the 200-m temperature-based Gulf Stream path index (GSI) model in subregions of (a) box01, southern Mid-Atlantic Bight (SMAB); (b) box03, northern Mid-Atlantic Bight (NMAB); (c) box05, Georges Bank (GB); (d) box06, western shallow Gulf of Maine (WSGoM); (e) box07, deep Gulf of Maine (DGoM); and (f) box08, Northeast Channel (NEC). Grids with markers (triangles or dots) on them mean that the anomaly correlation coefficient (ACC) is significant at the 95% confidence level. Specifically, gray 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.

T200-based GSI directly reflects the water temperature variability associated with the meridional shift of the GS North Wall, including the effects of WCRs and GS meanders, while the SSH-based GSI is not a direct measurement of temperature and primarily captures the meridional shift of GS main axis. In addition, due to 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 persistence-advection 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 persistence-advection model (Section 3.4). Results suggest that this persistence-advection with GSI model results in limited improvement relative to the original persistence-advection model, and the improved prediction skill is not significantly different at the 95% confidence level in nearly any NES subregions (Figure S7).

5. Summary and Discussion

The seasonal predictability of BTA on the NES was investigated based on a hierarchy of statistical prediction models, evaluated using deterministic and probabilistic metrics (ACC, RMSE, BrS, and BSS) 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 (persistence-advection model) leads 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 persistence-advection model. Overall, using nearby or upstream BTA's as predictors in the persistence-advection model provides better results than those using GSI.

In the remainder of the study, we will discuss various aspects of our statistical BTA prediction, including the physical interpretation of the persistence-advection 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 persistence-advection 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 persistence–advection model are primarily located in neighboring or upstream regions (Figures 5 and 6; 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 persistence-advection 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 data set 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 8 and 9).

The spatial pattern of maximum prediction skill for the NEC (Figures 8a–8d) highlights nonlocal predictors from the WSS (box11) and from the adjacent continental slope regions (within C–E, also see Figure 6f). At a lead time of 1 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 DGoM 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 8). The evolution of the maximum prediction skill pattern confirms our interpretation of the persistence-advection model for the NES in Section 4.2, that is, the nonlocal predictability is associated with the subsurface equatorward advection of the Labrador slope waters into the NEC along the continental slope.

As expected, prediction of BTAs in the DGoM (Figure S8) 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. The prediction skill in these nonlocal regions is lower than that of local persistence at shorter time lead (Figures S8a and S8b). While at longer time lead, BTAs on the slopes C–E exhibit slightly higher prediction skills for BTAs in the DGoM as suggested in Section 4.2, compared to the local persistence (Figures S8c and S8d).

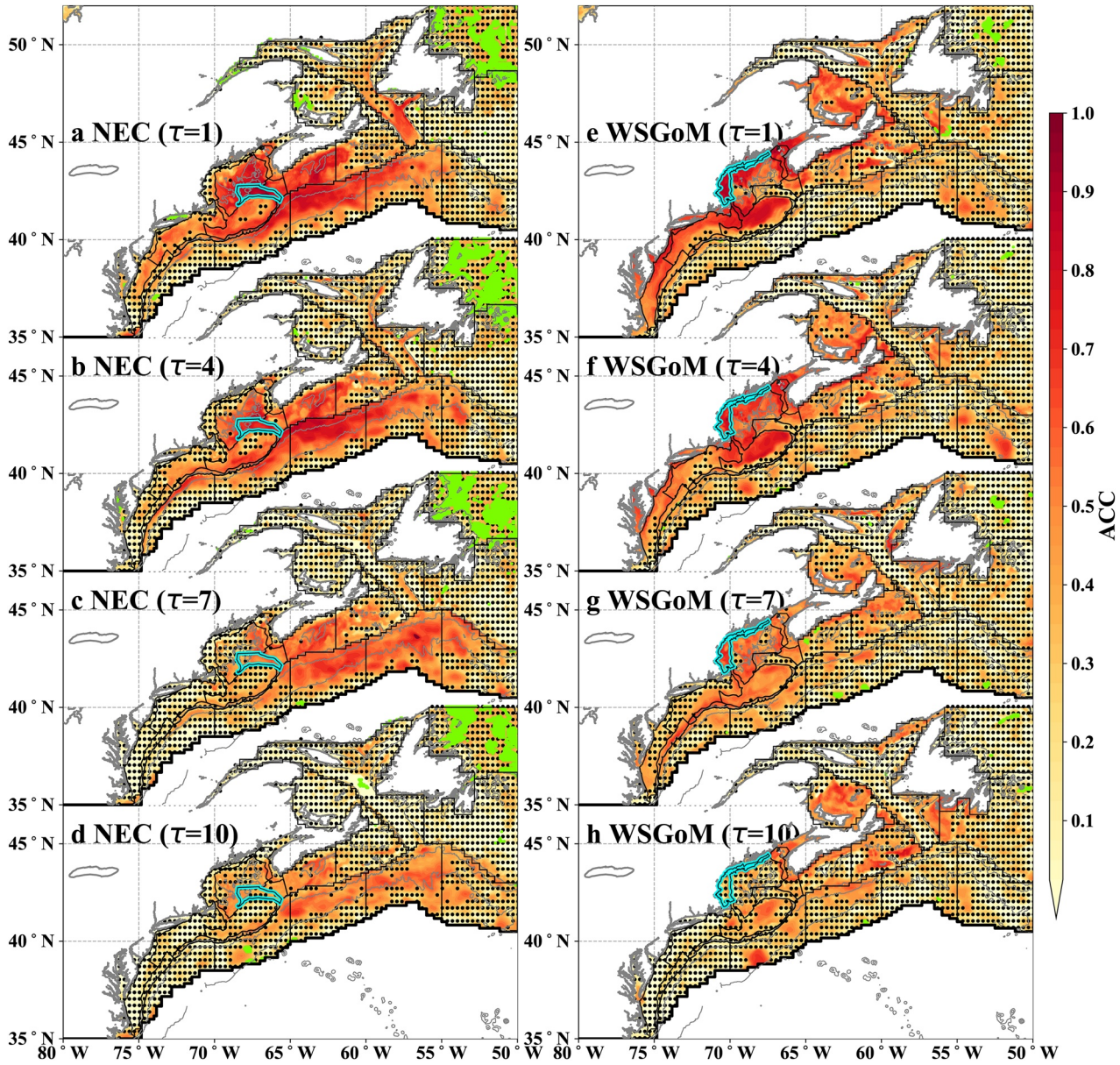

Figure 8. Spatial patterns of the maximum prediction skill when each grid point is used as a predictor in the persistence-advection model for the target box08, Northeast Channel (NEC; a–d) and target box06, western shallow Gulf of Maine (WSGoM; e–h) (highlighted by cyan lines) at a lead time of (a, e) 1 month, (b, f) 4 months, (c, g) 7 months, and (d, h) 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 black dots. For those locations that have prediction skills above 95% level, but whose bottom temperature anomaly negatively correlated with the target region, are covered with green dots (where are discarded when picking the best predictor). Bathymetry contours of 200, 2,000, and 4,000 m are plotted as solid gray lines.

Unlike the DGoM and NEC regions, the best nonlocal predictors for WSGoM (box06) are mostly located along the shallow shelf coastal boundary current regions, including NSGoM (box09), and southern Gulf of St. Lawrence (SGSL; box13) (Figures 8e–8h). Interestingly, the spatial patterns of maximum prediction skill for WSGoM look very different compared with those for the NEC at shorter lead times (Figures 8a–8d), 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

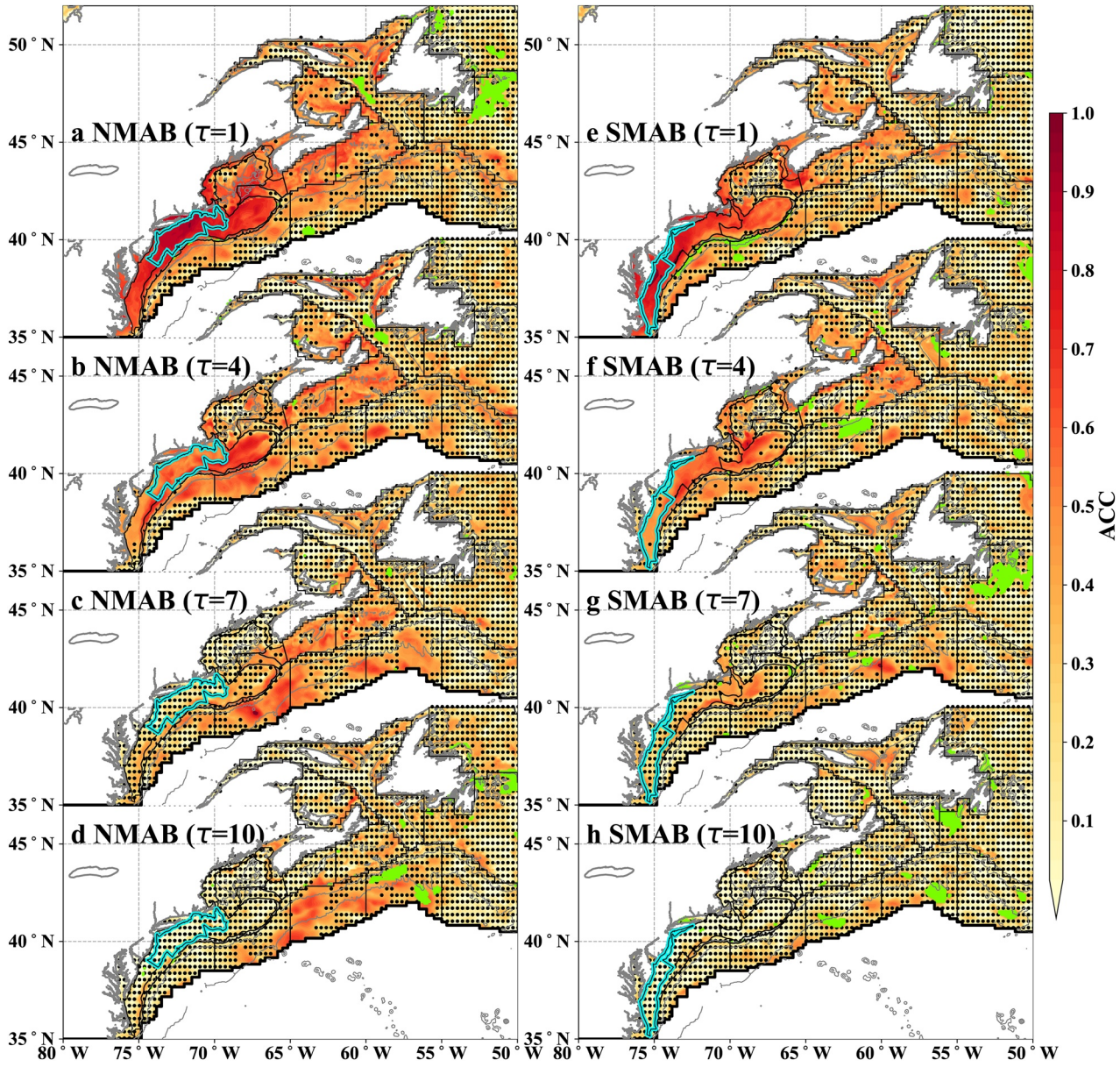

Figure 9. As in Figure 8, but for the target box03, northern Mid-Atlantic Bight (NMAB; a–d) and box01, southern Mid-Atlantic Bight (SMAB; e–h).

lead times, suggests that the BTAs in WSGoM are dominated by atmospheric forcing, which also affects the downstream regions at the same time.

The best nonlocal predictors for the MAB and GB regions include the shallow shelf regions in the MAB, GB, GoM, and SS (Figures 9 and S8e–S8h). For example, the entire MAB shelf and GB exhibit very high skill at shorter lead times for predicting BTA in the NMAB (box03) (Figures 9a and 9b), 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) (Figures 9b–9d), suggesting the importance of GS meanders (e.g., Andres, 2016; G. G. 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–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

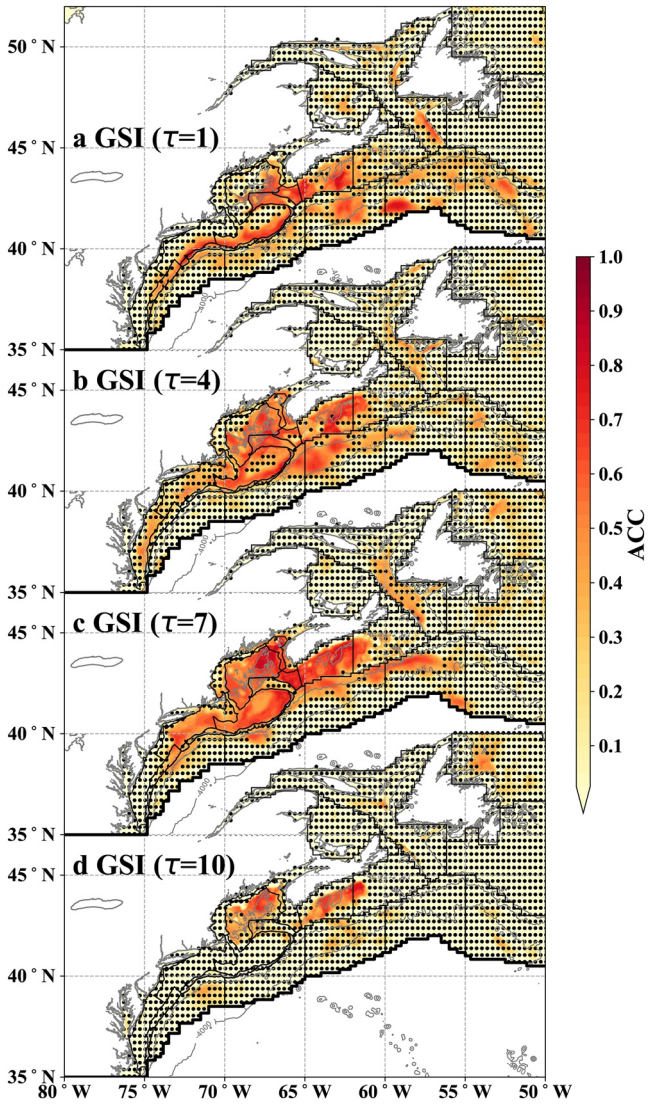

Figure 10. Similar to Figures 8a–8d, but for spatial patterns of prediction skill when the T200-based Gulf Stream path index (GSI) initiating in April is used as a predictor in the persistence–advection model for each spatial point at a lead time of (a) 1 month, (b) 4 months, (c) 7 months, and (d) 10 months.

the peak months for WCR formation in the slope sea, occurring from May to July (Gangopadhyay et al., 2019).

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 northern NES subregions, except MAB subregions 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 that 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 DGoM 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–November) BTAs on the WSS, DGoM, and GB have the most notable response to the GS meridional shift in spring. The fact that the persistence–advection model already incorporates these regions in the pathways of the GS-related signal as the nonlocal predictors explain why the T200-GSI does not provide additional significant improvement over the original persistence-advection model.

Other large-scale indices, for example, 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 4 years based on the NOAA Optimum Interpolation SST version 2 data set over the period of 1982–2010, indicating the NAO signal takes about 4 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 data sets, which indicates the nonstationarity 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 1993–2018 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 1993–2018. Results suggest only the NEC BTA has negative correlations ($r = -0.56$ to 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.

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., Chen et al., 2020; J. S. T. Forsyth et al., 2015; Pershing et al., 2015), 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 persistence-advection model (Section 4.2) in order to quantify the increase in prediction skill due to long-term warming (Tommasi, Stock, Alexander, et al., 2017). Results suggest that including the linear trends increases the prediction skill by roughly 0.15 in the GB, 0.13 in the WSGoM and NEC regions, and about 0.1 in the DGoM and MAB subregions, and these increases have distinct seasonal preferences (Figure S9). For example, the prediction skill in the SMAB is improved by about 0.32 for October forecasts but results in almost no improvement in June. Similarly, the prediction skill for the GB is improved by roughly 0.25 in winter/late fall (December-March), while the prediction skill for summer BTAs only improves by about 0.05 (Figure S9a). The skill gained from including the linear trend is larger at longer lead times, increasing by up to 0.27 at 12-month lead for GB (Figure S9b). When the linear trend is included, the prediction skill from the persistence–advection model reaches 0.50 in the SMAB to 0.76 in the DGoM on average.

5.4. Simplification of the Persistence-Advection Model

The most skillful predictions from the persistence-advection 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 to 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 persistence–advection prediction model (Figure 11). For example, for the NEC, we identify the combined slope region offshore of the SS (slope regions C and D) in addition to local persistence. The selection rule is to keep only those physically reasonable predictors that maximize the persistence–advection. The chosen nonlocal predictors for each subregion are GB (box05) for SMAB, slope box B for NMAB, and GB, NSGoM (box09) for WSGoM, and the combined slopes C and D for DGoM and NEC, as shown in Table S1. The seasonal prediction skill for the simplified prediction model in each subregion and the difference relative to the original persistence–advection 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 (1–3 months in the MAB regions and 1–4 months in the GoM regions) compared with BTA, and with fewer regional differences (Figure S10). 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

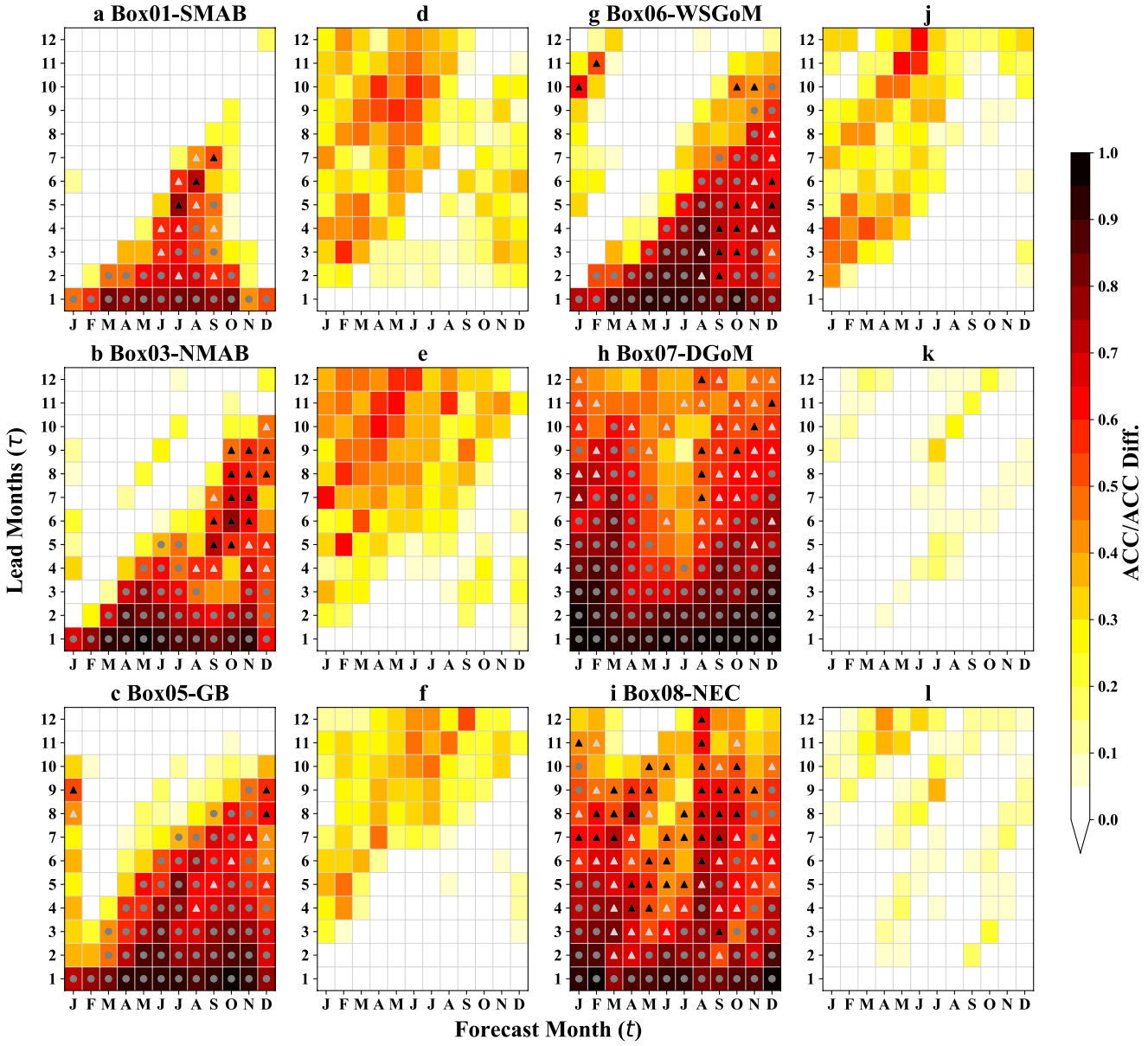

Figure 11. As in Figure 7, but for the simplified persistence–advection model (only selecting one or two best nonlocal predictors for each Northeast U.S. shelf subregion: box05-Georges Bank (GB) for target box01-southern Mid-Atlantic Bight (SMAB); slope box B for target box03-northern Mid-Atlantic Bight (NMAB) and target box05-GB; box09-northern shallow Gulf of Maine (NSGoM) for target box06-western shallow Gulf of Maine (WSGoM), and the combined slope boxes C and D for target box07-deep Gulf of Maine (DGoM) and target box08-Northeast Channel (NEC)) and the difference from the original persistence–advection model with values meaning the decreased prediction skills in comparison to the original persistence–advection model.

significant even at a lead time of 1 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, for example, freshwater input from rivers and atmospheric variability, that influence SSTA on the NES. Among all eight regions, the GB and the NSB have the longest local persistence (up to ~4 months), however, because of the lack of tidal mixing and the challenge in resolving shelfbreak dynamics in the reanalysis data set, 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.

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 NMAB, WSGoM, and DGoM are significant in November only. Unlike the BTA prediction in the SMAB, SSTA has almost no predictability from the T200-based GSI. Similarly, incorporating the T200-GSI as an extra predictor in the persistence–advection 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, for example, 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 elusive because of some key challenges. A critical component of fisheries stock assessment models is 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 (a) the relationship between the oceanographic variable and a population process must be established, (b) the oceanographic variable must also be observed or forecasted with reasonable skill and (c) 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 spatiotemporal 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 & 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 S2). Here, prediction skill ranged from 0.2 to 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 S2; Figures 5 and 6). 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 S2). 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 (Gaichas et al., 2016; Punt et al., 2014, 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 nontarget species bycatch (Hazen et al., 2018; Hobday et al., 2019; Howell et al., 2008, 2015) could potentially be more effective with prediction products developed herein.

Data Availability Statement

The GLORYS12v1 ocean reanalysis data set 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 data set 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 (https://resources.marine.copernicus.eu/?option=com_csw&view=details&product_id=SEALEVEL_GLO_PHY_L4_REP_OBSERVATIONS_008_047). The subsurface temperature EN4 data set 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). The regression coefficients α and c of the persistence–advection prediction model in each subregion on the NES are publicly available at Figshare (https://doi.org/10.6084/m9.figshare.14445690.v1).

Acknowledgments
This work was supported by NOAA's Climate Program Office's Modeling, Analysis, Predictions, and Projections (MAPP) Program (NA17OAR4310111, NA19OAR4320074), and Climate Program Office's Climate Variability and Predictability (CVP) Program (NA20OAR4310482). We acknowledge our participation in MAPP's Marine Prediction Task Force.

References

Andres, M. (2016). On the recent destabilization of the Gulf Stream path downstream of Cape Hatteras. *Geophysical Research Letters*, *43*, 9836–9842. https://doi.org/10.1002/2016GL069966

Barnston, A. G., & Livezey, R. E. (1987). Classification, seasonality and persistence of low-frequency atmospheric circulation patterns. *Monthly Weather Review*, *115*(6), 1083–1126. https://doi.org/10.1175/1520-0493(1987)115<1083:csapol>2.0.co;2

Brennan, C. E., Blanchard, H., & Fennel, K. (2016). Putting temperature and oxygen thresholds of marine animals in context of environmental change: A regional perspective for the Scotian Shelf and Gulf of St. Lawrence. *PloS One*, *11*(12), e0167411.

Bretherton, C. S., Widmann, M., Dymnikov, V. P., Wallace, J. M., & Bladé, I. (1999). The effective number of spatial degrees of freedom of a time-varying field. *Journal of Climate*, *12*(7), 1990–2009. https://doi.org/10.1175/1520-0442(1999)012<1990:tenosd>2.0.co;2

Brier, G. W. (1950). Verification of forecasts expressed in terms of probability. *Monthly Weather Review*, *78*(1), 1–3. https://doi.org/10.1175/1520-0493(1950)078<0001:vofeit>2.0.co;2

Cabanes, C., Grouazel, A., Von Schuckmann, K., Hamon, M., Turpin, V., Coatanoan, C., et al. (2013). The CORA dataset: Validation and diagnostics of in-situ ocean temperature and salinity measurements. *Ocean Science*, *9*(1), 1–18. https://doi.org/10.5194/os-9-1-2013

Caesar, L., Rahmstorf, S., Robinson, A., Feulner, G., & Saba, V. (2018). Observed fingerprint of a weakening Atlantic Ocean overturning circulation. *Nature*, *556*(7700), 191–196. https://doi.org/10.1038/s41586-018-0006-5

Castelao, R., Glenn, S., Schofield, O., Chant, R., Wilkin, J., & Kohut, J. (2008). Seasonal evolution of hydrographic fields in the central Middle Atlantic Bight from glider observations. *Geophysical Research Letters*, *35*, L03617. https://doi.org/10.1029/2007GL032335

Chapman, D. C., & Beardsley, R. C. (1989). On the origin of shelf water in the middle Atlantic Bight. *Journal of Physical Oceanography*, *19*(3), 384–391. https://doi.org/10.1175/1520-0485(1989)019<0384:otoosw>2.0.co;2

Chen, K., He, R., Powell, B. S., Gawarkiewicz, G. G., Moore, A. M., & Arango, H. G. (2014). Data assimilative modeling investigation of Gulf Stream Warm Core Ring interaction with continental shelf and slope circulation. *Journal of Geophysical Research: Oceans*, *119*, 5968–5991. https://doi.org/10.1002/2014JC009898

Chen, Z., Curchitser, E., Chant, R., & Kang, D. (2018). Seasonal variability of the cold pool over the Mid-Atlantic Bight continental shelf. *Journal of Geophysical Research: Oceans*, *123*, 8203–8226. https://doi.org/10.1029/2018JC014148

Chen, Z., & Curchitser, E. N. (2020). Interannual variability of the Mid-Atlantic Bight cold pool. *Journal of Geophysical Research: Oceans*, *125*, e2020JC016445. https://doi.org/10.1029/2020JC016445

Chen, Z., Kwon, Y. O., Chen, K., Fratantoni, P., Gawarkiewicz, G., & Joyce, T. M. (2020). Long-term SST variability on the Northwest Atlantic continental shelf and slope. *Geophysical Research Letters*, *47*, e2019GL085455. https://doi.org/10.1029/2019GL085455

Crear, D. P., Watkins, B. E., Saba, V. S., Graves, J. E., Jensen, D. R., Hobday, A. J., & Weng, K. C. (2020). Contemporary and future distributions of cobia, *Rachycentron canadum*. *Diversity and Distributions*, *26*(8), 1002–1015. https://doi.org/10.1111/ddi.13079

Davis, X. J., Joyce, T. M., & Kwon, Y.-O. (2017). Prediction of silver hake distribution on the Northeast U.S. shelf based on the Gulf Stream path index. *Continental Shelf Research*, *138*, 51–64. https://doi.org/10.1016/j.csr.2017.03.003

Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., et al. (2011). The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Quarterly Journal of the Royal Meteorological Society*, *137*(656), 553–597. https://doi.org/10.1002/qj.828

Dunn, D. C., Maxwell, S. M., Boustany, A. M., & Halpin, P. N. (2016). Dynamic ocean management increases the efficiency and efficacy of fisheries management. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(3), 668–673. https://doi.org/10.1073/pnas.1513626113

Fernandez, E., & Lellouche, J. M. (2018). Product user manual for the global ocean physical reanalysis product GLORYS12V1. *Copernicus Product User Manual*, *4*, 1–15.

Forsyth, J., Gawarkiewicz, G., Andres, M., & Chen, K. (2018). The interannual variability of the breakdown of fall stratification on the New Jersey shelf. *Journal of Geophysical Research: Oceans*, *123*, 6503–6520. https://doi.org/10.1029/2018JC014049

Forsyth, J. S. T., Andres, M., & Gawarkiewicz, G. G. (2015). Recent accelerated warming of the continental shelf off New Jersey: Observations from the CMV Oleander expendable bathythermograph line. *Journal of Geophysical Research: Oceans*, *120*, 2370–2384. https://doi.org/10.1002/2014JC010516

Frankignoul, C., de Coëtlogon, G., Joyce, T. M., & Dong, S. (2001). Gulf Stream variability and ocean-atmosphere interactions. *Journal of Physical Oceanography*, *31*(12), 3516–3529. https://doi.org/10.1175/1520-0485(2002)031<3516:gsvaoa>2.0.co;2

Fratantoni, P. S., & Pickart, R. S. (2007). The western North Atlantic shelfbreak current system in summer. *Journal of Physical Oceanography*, *37*(10), 2509–2533. https://doi.org/10.1175/jpo3123.1

Gaichas, S. K., Seagraves, R. J., Coakley, J. M., DePiper, G. S., Guida, V. G., Hare, J. A., et al. (2016). A framework for incorporating species, fleet, habitat, and climate interactions into fishery management. *Frontiers in Marine Science*, *3*, 105. https://doi.org/10.3389/fmars.2016.00105

Gangopadhyay, A., Gawarkiewicz, G., Silva, E. N. S., Monim, M., & Clark, J. (2019). An observed regime shift in the formation of warm core rings from the Gulf Stream. *Scientific Reports*, *9*(1), 1–9. https://doi.org/10.1038/s41598-019-48661-9

Gawarkiewicz, G. G., Todd, R. E., Plueddemann, A. J., Andres, M., & Manning, J. P. (2012). Direct interaction between the Gulf Stream and the shelfbreak south of New England. *Scientific Reports*, *2*, 553. https://doi.org/10.1038/srep00553

Gawarkiewicz, G., Todd, R., Zhang, W., Partida, J., Gangopadhyay, A., Monim, M.-U.-H., et al. (2018). The changing nature of shelf-break exchange revealed by the OOI Pioneer Array. *Oceanography*, *31*(1), 60–70. https://doi.org/10.5670/oceanog.2018.110

Girard-Ardhuin, F., Ezraty, R., & Croizé-Fillon, D. (2008). Arctic and Antarctic sea ice concentration and sea ice drift satellite products at Ifremer/CERSAT. *Quarterly Newsletter*, *31*.

Good, S. A., Martin, M. J., & Rayner, N. A. (2013). EN4: Quality controlled ocean temperature and salinity profiles and monthly objective analyses with uncertainty estimates. *Journal of Geophysical Research: Oceans*, *118*, 6704–6716. https://doi.org/10.1002/2013JC009067

Greene, C. H., Meyer-Gutbrod, E., Monger, B. C., McGarry, L. P., Pershing, A. J., Belkin, I. M., et al. (2013). Remote climate forcing of decadal-scale regime shifts in Northwest Atlantic shelf ecosystems. *Limnology & Oceanography*, *58*(3), 803–816. https://doi.org/10.4319/lo.2013.58.3.0803

Greene, C. H., & Pershing, A. J. (2007). OCEANS: Climate drives sea change. *Science*, *315*, 1084–1085. https://doi.org/10.1126/science.1136495

Haltuch, M. A., & Punt, A. E. (2011). The promises and pitfalls of including decadal-scale climate forcing of recruitment in groundfish stock assessment. *Canadian Journal of Fisheries and Aquatic Sciences*, *68*(5), 912–926. https://doi.org/10.1139/f2011-030

Hartigan, J. A., & Wong, M. A. (1979). Algorithm AS 136: A K-means clustering algorithm. *Applied Statistics*, *28*(1), 100–108. https://doi.org/10.2307/2346830

Hazen, E. L., Scales, K. L., Maxwell, S. M., Briscoe, D. K., Welch, H., Bograd, S. J., et al. (2018). A dynamic ocean management tool to reduce bycatch and support sustainable fisheries. *Science Advances*, *4*(5), eaar3001. https://doi.org/10.1126/sciadv.aar3001

Hervieux, G., Alexander, M. A., Stock, C. A., Jacox, M. G., Pegion, K., Becker, E., et al. (2019). More reliable coastal SST forecasts from the North American multimodel ensemble. *Climate Dynamics*, *53*(12), 7153–7168. https://doi.org/10.1007/s00382-017-3652-7

Hobday, A. J., Hartog, J. R., Manderson, J. P., Mills, K. E., Oliver, M. J., Pershing, A. J., & Siedlecki, S. (2019). Ethical considerations and unanticipated consequences associated with ecological forecasting for marine resources. *ICES Journal of Marine Science*, *76*(5), 1244–1256.

Houghton, R. W., Schlitz, R., Beardsley, R. C., Butman, B., & Chamberlin, J. L. (1982). The Middle Atlantic Bight cold pool: Evolution of the temperature structure during summer 1979. *Journal of Physical Oceanography*, *12*(10), 1019–1029. https://doi.org/10.1175/1520-0485(1982)012<1019:tmabcp>2.0.co;2

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

Howell, E. A., Hoover, A., Benson, S. R., Bailey, H., Polovina, J. J., Seminoff, J. A., & Dutton, P. H. (2015). Enhancing the TurtleWatch product for leatherback sea turtles, a dynamic habitat model for ecosystem-based management. *Fisheries Oceanography*, *24*(1), 57–68. https://doi.org/10.1111/fog.12092

Howell, E. A., Kobayashi, D. R., Parker, D. M., Balazs, G. H., & Polovina, J. J. (2008). TurtleWatch: A tool to aid in the bycatch reduction of loggerhead turtles *Caretta caretta* in the Hawaii-based pelagic longline fishery. *Endangered Species Research*, *5*(2–3), 267–278. https://doi.org/10.3354/esr00096

Jacox, M. G., Alexander, M. A., Siedlecki, S., Chen, K., Kwon, Y.-O., Brodie, S., et al. (2020). Seasonal-to-interannual prediction of North American coastal marine ecosystems: Forecast methods, mechanisms of predictability, and priority developments. *Progress in Oceanography*, *183*, 102307. https://doi.org/10.1016/j.pocean.2020.102307

Jacox, M. G., Alexander, M. A., Stock, C. A., & Hervieux, G. (2019). On the skill of seasonal sea surface temperature forecasts in the California current system and its connection to ENSO variability. *Climate Dynamics*, *53*(12), 7519–7533. https://doi.org/10.1007/s00382-017-3608-y

Joyce, T. M., Kwon, Y.-O., & Yu, L. (2009). On the relationship between synoptic wintertime atmospheric variability and path shifts in the Gulf Stream and the Kuroshio Extension. *Journal of Climate*, *22*(12), 3177–3192. https://doi.org/10.1175/2008jcli2690.1

Joyce, T. M., & Zhang, R. (2010). On the path of the Gulf Stream and the Atlantic meridional overturning circulation. *Journal of Climate*, *23*(11), 3146–3154. https://doi.org/10.1175/2010jcli3310.1

Kleisner, K. M., Fogarty, M. J., McGee, S., Barnett, A., Fratantoni, P., Greene, J., et al. (2016). The effects of sub-regional climate velocity on the distribution and spatial extent of marine species assemblages. *PLoS One*, *11*(2), e0149220. https://doi.org/10.1371/journal.pone.0149220

Kleisner, K. M., Fogarty, M. J., McGee, S., Hare, J. A., Moret, S., Perretti, C. T., & Saba, V. S. (2017). Marine species distribution shifts on the U.S. Northeast Continental Shelf under continued ocean warming. *Progress in Oceanography*, *153*, 24–36. https://doi.org/10.1016/j.pocean.2017.04.001

Lellouche, J.-M., Greiner, E., Le Galloudec, O., Garric, G., Regnier, C., Drevillon, M., et al. (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. https://doi.org/10.5194/os-14-1093-2018

Lentz, S. J. (2017). Seasonal warming of the Middle Atlantic Bight Cold Pool. *Journal of Geophysical Research: Oceans*, *122*, 941–954. https://doi.org/10.1002/2016JC012201

Link, J. S., Brodziak, J. K. T., Edwards, S. F., Overholtz, W. J., Mountain, D., Jossi, J. W., et al. (2002). Marine ecosystem assessment in a fisheries management context. *Canadian Journal of Fisheries and Aquatic Sciences*, *59*(9), 1429–1440. https://doi.org/10.1139/f02-115

Loder, J. W. (1998). The coastal ocean off northeastern North America: A large-scale view. *The Sea*, *11*, 105–138.

Lucey, S., & Nye, J. (2010). Shifting species assemblages in the Northeast US continental shelf large marine ecosystem. *Marine Ecology Progress Series*, *415*, 23–33. https://doi.org/10.3354/meps08743

Madec, G. (2008). NEMO ocean engine. In *Note du Pole de Modelisation*. Paris: Institut Pierre-Simon Laplace (IPSL).

Maxwell, S. M., Hazen, E. L., Lewison, R. L., Dunn, D. C., Bailey, H., Bograd, S. J., et al. (2015). Dynamic ocean management: Defining and conceptualizing real-time management of the ocean. *Marine Policy*, *58*, 42–50. https://doi.org/10.1016/j.marpol.2015.03.014

McHenry, J., Welch, H., Lester, S. E., & Saba, V. (2019). Projecting marine species range shifts from only temperature can mask climate vulnerability. *Global Change Biology*, *25*(12), 4208–4221. https://doi.org/10.1111/gcb.14828

Miller, T. J., Hare, J. A., & Alade, L. A. (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. https://doi.org/10.1139/cjfas-2015-0339

Miller, T. J., O'Brien, L., & Fratantoni, P. S. (2018). Temporal and environmental variation in growth and maturity and effects on management reference points of Georges Bank Atlantic cod. *Canadian Journal of Fisheries and Aquatic Sciences*, *75*(12), 2159–2171. https://doi.org/10.1139/cjfas-2017-0124

Mills, K. E., Pershing, A. J., Brown, C. J., Chen, Y., Chiang, F.-S., Holland, D. S., et al. (2013). Fisheries management in a changing climate: Lessons from the 2012 ocean heat wave in the Northwest Atlantic. *Oceanography*, *26*(2), 191–195. https://doi.org/10.5670/oceanog.2013.27

Mills, K. E., Pershing, A. J., & Hernández, C. M. (2017). Forecasting the seasonal timing of Maine's lobster fishery. *Frontiers in Marine Science*, *4*, 337. https://doi.org/10.3389/fmars.2017.00337

Mountain, D. G. (2012). Labrador slope water entering the Gulf of Maine—Response to the North Atlantic Oscillation. *Continental Shelf Research*, *47*, 150–155. https://doi.org/10.1016/j.csr.2012.07.008

Mountain, D. G., & Manning, J. P. (1994). Seasonal and interannual variability in the properties of the surface waters of the Gulf of Maine. *Continental Shelf Research*, *14*(13–14), 1555–1581. https://doi.org/10.1016/0278-4343(94)90090-6

NEFSC. (2020). *Southern New England Mid-Atlantic Winter Flounder 2020 Assessment update report*. Retrieved from https://apps-nefsc.fisheries.noaa.gov/saw/sasi/sasi_report.php

Nye, J. A., Joyce, T. M., Kwon, Y.-O., & Link, J. S. (2011). Silver hake tracks changes in Northwest Atlantic circulation. *Nature Communications*, *2*(1), 1–6. https://doi.org/10.1038/ncomms1420

Nye, J. A., Link, J. S., Hare, J. A., & Overholtz, W. J. (2009). Changing spatial distribution of fish stocks in relation to climate and population size on the Northeast United States continental shelf. *Marine Ecology Progress Series*, *393*, 111–129. https://doi.org/10.3354/meps08220

Payne, M. R., Hobday, A. J., MacKenzie, B. R., Tommasi, D., Dempsey, D. P., Fässler, S. M., et al. (2017). Lessons from the first generation of marine ecological forecast products. *Frontiers in Marine Science*, *4*, 289. https://doi.org/10.3389/fmars.2017.00289

Peña-Molino, B., & Joyce, T. (2008). Variability in the Slope Water and its relation to the Gulf Stream path. *Geophysical Research Letters*, *35*, L03606. https://doi.org/10.1029/2007GL032183

Pérez-Hernández, M. D., & Joyce, T. M. (2014). Two modes of Gulf Stream variability revealed in the last two decades of satellite altimeter data. *Journal of Physical Oceanography*, *44*(1), 149–163. https://doi.org/10.1175/jpo-d-13-0136.1

Pershing, A. J., Alexander, M. A., Hernandez, C. M., Kerr, L. A., Le Bris, A., Mills, K. E., et al. (2015). Slow adaptation in the face of rapid warming leads to collapse of the Gulf of Maine cod fishery. *Science*, *350*(6262), 809–812. https://doi.org/10.1126/science.aac9819

Pinsky, M. L., Worm, B., Fogarty, M. J., Sarmiento, J. L., & Levin, S. A. (2013). Marine taxa track local climate velocities. *Science*, *341*(6151), 1239–1242. https://doi.org/10.1126/science.1239352

Punt, A. E., A'mar, T., Bond, N. A., Butterworth, D. S., de Moor, C. L., De Oliveira, J. A. A., et al. (2014). Fisheries management under climate and environmental uncertainty: Control rules and performance simulation. *ICES Journal of Marine Science*, *71*(8), 2208–2220. https://doi.org/10.1093/icesjms/fst057

Punt, A. E., Butterworth, D. S., de Moor, C. L., De Oliveira, J. A. A., & Haddon, M. (2016). Management strategy evaluation: Best practices. *Fish and Fisheries*, *17*(2), 303–334. https://doi.org/10.1111/faf.12104

Ramp, S. R., Schlitz, R. J., & Wright, W. R. (1985). The deep flow through the northeast Channel, Gulf of Maine. *Journal of Physical Oceanography*, *15*(12), 1790–1808. https://doi.org/10.1175/1520-0485(1985)015<1790:tdfttn>2.0.co;2

Reynolds, R. W., Smith, T. M., Liu, C., Chelton, D. B., Casey, K. S., & Schlax, M. G. (2007). Daily high-resolution-blended analyses for sea surface temperature. *Journal of Climate*, *20*(22), 5473–5496. https://doi.org/10.1175/2007jcli1824.1

Richaud, B., Kwon, Y.-O., Joyce, T. M., Fratantoni, P. S., & Lentz, S. J. (2016). Surface and bottom temperature and salinity climatology along the continental shelf off the Canadian and U.S. East Coasts. *Continental Shelf Research*, *124*, 165–181. https://doi.org/10.1016/j.csr.2016.06.005

Saba, V. S., Griffies, S. M., Anderson, W. G., Winton, M., Alexander, M. A., Delworth, T. L., et al. (2016). Enhanced warming of the northwest Atlantic Ocean under climate change. *Journal of Geophysical Research: Oceans*, *121*, 118–132. https://doi.org/10.1002/2015JC011346

Saba, V. S., Hyde, K. J. W., Rebuck, N. D., Friedland, K. D., Hare, J. A., Kahru, M., & Fogarty, M. J. (2015). Physical associations to spring phytoplankton biomass interannual variability in the U.S. Northeast Continental Shelf. *Journal of Geophysical Research: Biogeosciences*, *120*, 205–220. https://doi.org/10.1002/2014JG002770

Selden, R. L., Batt, R. D., Saba, V. S., & Pinsky, M. L. (2018). Diversity in thermal affinity among key piscivores buffers impacts of ocean warming on predator-prey interactions. *Global Change Biology*, *24*(1), 117–131. https://doi.org/10.1111/gcb.13838

Smith, P. C., Houghton, R. W., Fairbanks, R. G., & Mountain, D. G. (2001). Interannual variability of boundary fluxes and water mass properties in the Gulf of Maine and on Georges Bank: 1993–1997. *Deep Sea Research Part II: Topical Studies in Oceanography*, *48*(1–3), 37–70. https://doi.org/10.1016/s0967-0645(00)00081-3

Steiger, J. H. (1980). Tests for comparing elements of a correlation matrix. *Psychological Bulletin*, *87*(2), 245–251. https://doi.org/10.1037/0033-2909.87.2.245

Stock, C. A., Alexander, M. A., Bond, N. A., Brander, K. M., Cheung, W. W., Curchitser, E. N., et al. (2011). On the use of IPCC-class models to assess the impact of climate on living marine resources. *Progress in Oceanography*, *88*(1–4), 1–27. https://doi.org/10.1016/j.pocean.2010.09.001

Stock, C. A., Pegion, K., Vecchi, G. A., Alexander, M. A., Tommasi, D., Bond, N. A., et al. (2015). Seasonal sea surface temperature anomaly prediction for coastal ecosystems. *Progress in Oceanography*, *137*, 219–236. https://doi.org/10.1016/j.pocean.2015.06.007

Sullivan, M. C., Cowen, R. K., & Steves, B. P. (2005). Evidence for atmosphere-ocean forcing of yellowtail flounder (*Limanda ferruginea*) recruitment in the Middle Atlantic Bight. *Fisheries Oceanography*, *14*(5), 386–399. https://doi.org/10.1111/j.1365-2419.2005.00343.x

Tanaka, K. R., Torre, M. P., Saba, V. S., Stock, C. A., & Chen, Y. (2020). An ensemble high-resolution projection of changes in the future habitat of American lobster and sea scallop in the Northeast US continental shelf. *Diversity and Distributions*, *26*(8), 987–1001.

Tommasi, D., Stock, C. A., Alexander, M. A., Yang, X., Rosati, A., & Vecchi, G. A. (2017). Multi-annual climate predictions for fisheries: An assessment of skill of sea surface temperature forecasts for large marine ecosystems. *Frontiers in Marine Science*, *4*, 201. https://doi.org/10.3389/fmars.2017.00201

Tommasi, D., Stock, C. A., Hobday, A. J., Methot, R., Kaplan, I. C., Eveson, J. P., et al. (2017). Managing living marine resources in a dynamic environment: The role of seasonal to decadal climate forecasts. *Progress in Oceanography*, *152*, 15–49. https://doi.org/10.1016/j.pocean.2016.12.011

Townsend, D. W., Thomas, A. C., Mayer, L. M., Thomas, M. A., & Quinlan, J. A. (2006). Oceanography of the Northwest Atlantic Continental Shelf. In A. R. Robinson & K. H. Brink (Eds.), *The sea: The global coastal ocean: Interdisciplinary regional studies and syntheses* (pp. 119–168). Cambridge, MA: Harvard University Press.

Walsh, H. J., Richardson, D. E., Marancik, K. E., & Hare, J. A. (2015). Long-term changes in the distributions of larval and adult fish in the Northeast US shelf ecosystem. *PLoS One*, *10*(9), e0137382. https://doi.org/10.1371/journal.pone.0137382

Xu, H., Kim, H.-M., Nye, J. A., & Hameed, S. (2015). Impacts of the North Atlantic Oscillation on sea surface temperature on the Northeast US Continental Shelf. *Continental Shelf Research*, *105*, 60–66. https://doi.org/10.1016/j.csr.2015.06.005

Xu, H., Miller, T. J., Hameed, S., Alade, L. A., & Nye, J. A. (2018). Evaluating the utility of the Gulf Stream Index for predicting recruitment of Southern New England-Mid Atlantic yellowtail flounder. *Fisheries Oceanography*, *27*(1), 85–95. https://doi.org/10.1111/fog.12236

Yuan, G. C., Lozier, M., Pratt, L., Jones, C., & Helfrich, K. (2004). Estimating the predictability of an oceanic time series using linear and nonlinear methods. *Journal of Geophysical Research*, *109*, C08002. https://doi.org/10.1029/2003JC002148

Zhang, R. (2008). Coherent surface-subsurface fingerprint of the Atlantic meridional overturning circulation. *Geophysical Research Letters*, *35*, L20705. https://doi.org/10.1029/2008GL035463

Zhang, R., Delworth, T. L., Rosati, A., Anderson, W. G., Dixon, K. W., Lee, H. C., & Zeng, F. (2011). Sensitivity of the North Atlantic Ocean circulation to an abrupt change in the Nordic Sea overflow in a high resolution global coupled climate model. *Journal of Geophysical Research: Oceans*, *116*, C12024. https://doi.org/10.1029/2011JC007240

Zhang, R., & Vallis, G. K. (2007). The role of bottom vortex stretching on the path of the North Atlantic western boundary current and on the northern recirculation gyre. *Journal of Physical Oceanography*, *37*(8), 2053–2080. https://doi.org/10.1175/jpo3102.1

Zhang, W. G., & Gawarkiewicz, G. G. (2015). Dynamics of the direct intrusion of Gulf Stream ring water onto the Mid-Atlantic Bight shelf. *Geophysical Research Letters*, *42*, 7687–7695. https://doi.org/10.1002/2015GL065530

References From the Supporting Information

ASMFC. (2018). *Northern shrimp benchmark stock assessment and peer review report* (pp. 356). Washington, DC: Atlantic States Marine Fisheries Commission.

ASMFC. (2020). *American lobster benchmark stock assessment and peer review report*. Washington, DC: Atlantic States Marine Fisheries Commission.

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

Cao, J., Thorson, J. T., Richards, R. A., & Chen, Y. (2017). Spatiotemporal index standardization improves the stock assessment of northern shrimp in the Gulf of Maine. *Canadian Journal of Fisheries and Aquatic Sciences*, *74*(11), 1781–1793.

Chen, Z., Curchitser, E., Chant, R., & Kang, D. (2018). Seasonal variability of the cold pool over the Mid-Atlantic Bight continental shelf. *Journal of Geophysical Research: Oceans*, *123*, 8203–8226. https://doi.org/10.1029/2018JC014148

Chi, L., Wolfe, C. L., & Hameed, S. (2019). The distinction between the Gulf Stream and its North Wall. *Geophysical Research Letters*, *46*, 8943–8951. https://doi.org/10.1029/2019GL083775

Fernandez, E., & Lellouche, J. (2018). Product user manual for the Global Ocean Physical Reanalysis product GLOBAL_REANALYSIS_PHY_001_030 (Report CMEMS-GLO-PUM(001-030), Vol. 15).

Hennen, D. R., Mann, R., Munroe, D. M., & Powell, E. N. (2018). Biological reference points for Atlantic surfclam (*Spisula solidissima*) in warming seas. *Fisheries Research*, *207*, 126–139.

Joyce, T. M., Deser, C., & Spall, M. A. (2000). The relation between decadal variability of subtropical mode water and the North Atlantic Oscillation. *Journal of Climate*, *13*(14), 2550–2569.

Joyce, T. M., Kwon, Y.-O., Seo, H., & Ummenhofer, C. C. (2019). Meridional Gulf Stream shifts can influence wintertime variability in the North Atlantic storm track and Greenland blocking. *Geophysical Research Letters*, *46*, 1702–1708. https://doi.org/10.1029/2018GL081087

Joyce, T. M., Kwon, Y.-O., & Yu, L. (2009). On the relationship between synoptic wintertime atmospheric variability and path shifts in the Gulf Stream and the Kuroshio Extension. *Journal of Climate*, *22*(12), 3177–3192.

Miller, A. S., Shepherd, G. R., & Fratantoni, P. S. (2016). Offshore habitat preference of overwintering juvenile and adult Black Sea Bass, *Centropristis striata*, and the relationship to year-class success. *PLoS One*, *11*(1), e0147627.

Miller, T. J., Hare, J. A., & Alade, L. A. (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.

Miller, T. J., O'Brien, L., & Fratantoni, P. S. (2018). Temporal and environmental variation in growth and maturity and effects on management reference points of Georges Bank Atlantic cod. *Canadian Journal of Fisheries and Aquatic Sciences*, *75*(12), 2159–2171.

NEFSC. (2014). 58th Northeast Regional Stock Assessment Workshop (58th SAW) assessment report (US Department of Commerce, Northeast Fisheries Science Center Reference Document 14-04 pp. 784). Retrieved from https://repository.library.noaa.gov/view/noaa/4719

NEFSC. (2020). *Southern New England Mid-Atlantic winter flounder 2020 assessment update report*. Retrieved from https://apps-nefsc.fisheries.noaa.gov/saw/sasi/sasi_report.php

O'Leary, C. A., Miller, T. J., Thorson, J. T., & Nye, J. A. (2019). Understanding historical summer flounder (*Paralichthys dentatus*) abundance patterns through the incorporation of oceanography-dependent vital rates in Bayesian hierarchical models. *Canadian Journal of Fisheries and Aquatic Sciences*, *76*(8), 1275–1294.

O'Leary, C. A., Thorson, J. T., Miller, T. J., & Nye, J. A. (2020). Comparison of multiple approaches to calculate time-varying biological reference points in climate-linked population-dynamics models. *ICES Journal of Marine Science*, *77*(3), 930–941.

Pershing, A. J., Alexander, M. A., Hernandez, C. M., Kerr, L. A., Le Bris, A., Mills, K. E., et al. (2015). Slow adaptation in the face of rapid warming leads to collapse of the Gulf of Maine cod fishery. *Science*, *350*(6262), 809–812.

Richards, R. A., Fogarty, M. J., Mountain, D. G., & Taylor, M. H. (2012). Climate change and northern shrimp recruitment variability in the Gulf of Maine. *Marine Ecology Progress Series*, *464*, 167–178.

Tanaka, K. R., Cao, J., Shank, B. V., Truesdell, S. B., Mazur, M. D., Xu, L., & Chen, Y. (2019). A model-based approach to incorporate environmental variability into assessment of a commercial fishery: A case study with the American lobster fishery in the Gulf of Maine and Georges Bank. *ICES Journal of Marine Science*, *76*(4), 884–896.

Xu, H., Miller, T. J., Hameed, S., Alade, L. A., & Nye, J. A. (2018). Evaluating the utility of the Gulf Stream Index for predicting recruitment of Southern New England-Mid Atlantic yellowtail flounder. *Fisheries Oceanography*, *27*(1), 85–95.