Ocean Sci., 20, 1631–1656, 2024
https://doi.org/10.5194/os-20-1631-2024

Ocean Science Open Access EGU

Dynamically downscaled seasonal ocean forecasts for North American east coast ecosystems

Andrew C. Ross[1], Charles A. Stock[1], Vimal Koul[2], Thomas L. Delworth[1], Feiyu Lu[1], Andrew Wittenberg[1], and Michael A. Alexander[3]

[1]NOAA/OAR/Geophysical Fluid Dynamics Laboratory, 201 Forrestal Road, Princeton, NJ 08540, USA
[2]Princeton University Cooperative Institute for Modeling the Earth System, 201 Forrestal Road, Princeton, NJ 08540, USA
[3]NOAA/OAR/Physical Sciences Laboratory, 325 Broadway, Boulder, CO 80305, USA

Correspondence: Andrew C. Ross (andrew.c.ross@noaa.gov)

Received: 13 February 2024 – Discussion started: 21 February 2024
Revised: 21 October 2024 – Accepted: 23 October 2024 – Published: 10 December 2024

Abstract. Using a 1/12° regional model of the Northwest Atlantic Ocean (MOM6-NWA12), we downscale an ensemble of retrospective seasonal forecasts from a 1° global forecast model. To evaluate whether downscaling improved the forecast skill for surface temperature and salinity and bottom temperature, the global and downscaled forecasts are compared with each other and with a reference forecast of persistence using anomaly correlation. Both sets of forecasts are also evaluated on the basis of mean bias and ensemble spread. We find that downscaling significantly improved the forecast skill for monthly sea surface temperature anomalies in the Northeast US Large Marine Ecosystem, a region that global models have historically struggled to skillfully predict. The downscaled sea surface temperature (SST) predictions for this region were also more skillful than the persistence baseline across most initialization months and lead times. Although some of the SST prediction skill in this region stems from the recent rapid warming trend, prediction skill above persistence is generally maintained after removing the contribution of the trend, and patterns of skill suggestive of predictable processes are also preserved. While downscaling mainly improved the SST anomaly prediction skill in the Northeast US region, it improved bottom temperature and sea surface salinity anomaly skill across many of the marine ecosystems along the North American east coast. Although improvements in anomaly prediction via downscaling were ubiquitous, the effects of downscaling on prediction bias were mixed. Downscaling generally reduced the mean surface salinity biases found in the global model, particularly in regions with sharp salinity gradients (the Northern Gulf of Mexico and the Northeast US). In some cases, however, downscaling amplified the surface and bottom temperature biases found in the global predictions. We discuss several processes that are better resolved in the regional model and contribute to the improved skill, including the autumn reemergence of temperature anomalies and advection of water masses by coastal currents. Overall, the results show that a downscaled high-resolution model can produce improved seasonal forecast skill by representing fine-scale processes that drive predictability.

1 Introduction

The development of seasonal climate forecasting systems, which provide information about predicted climate conditions on timescales ranging from months to years, has sparked new applications to anticipate the response of marine ecosystems to predictable climate variability and provide information to inform the management of living marine resources (Jacox et al., 2020; Payne et al., 2017; Tommasi et al., 2017). Forecasts of sea surface temperatures from these seasonal forecast systems, which are based on global climate models with coupled atmosphere and ocean components, are skillful across many Large Marine Ecosystems (LMEs) (Hervieux et al., 2017; Stock et al., 2015). In many regions, global models are also capable of predicting the occurrence and intensity of marine heat waves (Jacox et al., 2022) and chlorophyll anomalies at the ocean surface (Park et al., 2019). A variety of processes contribute to predictabil-

ity in different regions, including the progression of coastal waves and the reemergence of prior anomalies (Jacox et al., 2020).

The skill of seasonal predictions for LMEs along the North American east coast (NAEC), including the Northeast US and Scotian Shelf LMEs, is typically much lower than many other regions in global forecast models (e.g., Stock et al., 2015). McAdam et al. (2022) suggested that the lack of skill may be due to coarse model resolution resulting in poor initialization or inability to maintain sharp fronts in the NAEC and other western boundary current regions. For the NAEC in particular, it is well known that models with resolution coarser than $1/10°$ are unable to accurately simulate the Gulf Stream and its separation from the continental shelf, typically resulting in severe temperature biases north and west of the current along the shelf and coast (Chassignet and Garaffo, 2001; Chassignet and Xu, 2017; Saba et al., 2016). Furthermore, compared to many other regions like the North American west coast, predictability for the east coast is hindered by its weaker teleconnections to the El Niño–Southern Oscillation (ENSO) and other predictable climate modes at seasonal timescales. The seasonal variability of ocean temperature anomalies in the North Atlantic basin is strongly connected to the North Atlantic Oscillation (NAO), but this mode is only moderately predictable at seasonal timescales (with dynamical model skill generally requiring large ensembles; Smith et al., 2020), and the contribution of NAO to ocean temperature predictability involves features and processes that are often poorly simulated in coarse-resolution models, including the modulation of the Gulf Stream (Sanchez-Franks et al., 2016; Shin and Newman, 2021) and the reemergence of anomalies from the previous winter sequestered below the mixed layer (Sukhonos and Alexander, 2023).

To the southwest, along the coast of the Gulf of Mexico, the predictability of ocean conditions is potentially enhanced by variability connected to ENSO (e.g., Alexander and Scott, 2002; Gomez et al., 2019). Idealized experiments have also shown the potential for 1–3 months of predictability of the Loop Current in the Gulf of Mexico (Dukhovskoy et al., 2023; Haley et al., 2023). High-resolution downscaled forecasts are likely necessary to realize this potential predictability; horizontal resolution of at least 10–25 km is necessary to simulate Loop Current eddies, for example (Oey et al., 2013).

Seasonal forecast models based on empirical relationships with observed temperature have shown some temperature prediction skill for the NAEC. Shin and Newman (2021) developed a linear inverse model to predict sea surface temperature (SST) and found that the model was significantly more skillful than a multi-model mean of global dynamical forecast models at predicting SST anomalies in the Northeast US and Scotian Shelf LMEs, particularly at longer lead times. Chen et al. (2021) found that bottom temperatures in NA coastal regions could be skillfully predicted by combining persistence of recent conditions with empirically determined advection from nearby regions. However, both of these models are univariate (they predict only surface or bottom temperature); are constrained by the availability and accuracy of historical observations; and may become unreliable if ocean currents, mixed layer depths, and other properties that influence the seasonal variability in ocean temperatures change in the future.

Dynamical downscaling of global model predictions with a high-resolution regional model could provide accurate seasonal predictions while avoiding many of the drawbacks of statistical/empirical predictions. In a pioneering application, Siedlecki et al. (2016) developed skillful dynamically downscaled forecasts of ocean temperature, oxygen, and acidification for the Northern California Current System. For the NAEC, dynamical downscaling has been used to produce short-term forecasts for the next few days (Wilkin et al., 2018) and long-term projections of the effects of climate change (Alexander et al., 2020; Han et al., 2019; Rutherford et al., 2024) but not to produce seasonal forecasts.

Dynamical downscaling does not guarantee increased forecast skill, however. Dynamically downscaled simulations often resemble the solution of the coarse model they are downscaling, which brings into question how much value they add (e.g., Ghantous et al., 2020). Although dynamical downscaling offers the possibility of running multiple ensemble members to provide information about forecast uncertainty, the substantial computational cost of high-resolution models used for downscaling has constrained the ensemble sizes of past climate downscaling efforts (Drenkard et al., 2021). In an analysis of seasonal forecasts for the California Current System, Brodie et al. (2023) found that although a small ensemble of downscaled forecasts outperformed a subset of the same size from a much larger ensemble of coarse-resolution forecasts, the full ensemble of coarse-resolution forecasts outperformed the small downscaled ensemble. Furthermore, in some cases, high ocean resolution has been found to actually degrade forecast skill (Sandery and Sakov, 2017; Thoppil et al., 2021).

In this study, we test the hypothesis that a high-resolution dynamical ocean model will improve the skill of seasonal forecasts for the NAEC relative to a lower-resolution global model. To do this, we dynamically downscale 29 years of retrospective forecast simulations produced by the Seamless System for Prediction and EArth System Research (SPEAR) model of the Geophysical Fluid Dynamics Laboratory (GFDL), which uses a coupled ocean component with $1°$ resolution. We downscale the SPEAR seasonal forecasts with a $1/12°$ model of the Northwest Atlantic Ocean built with the regional modeling capabilities of the MOM6 ocean model (MOM6-NWA12), which has recently been assessed for accuracy when simulating ocean temperatures, position and variability of the Gulf Stream, and other relevant ocean features along the NAEC in a reanalysis-forced historical simulation (Ross et al., 2023a). To the best of our knowledge, this is the first attempt to dynamically downscale seasonal ocean forecasts for the Northwest Atlantic Ocean.

In the sections that follow, we describe the methods used to downscale the SPEAR forecasts and evaluate the retrospective prediction skill (Sect. 2). The presentation of the model results (Sect. 3) focuses on seasonal prediction skill for surface and bottom temperature and surface salinity in the downscaled simulations relative to the global simulations. In the discussion (Sect. 4), we examine drivers of predictability and skill and discuss potential future improvements to the forecasting system. We conclude by noting the promising potential for these seasonal forecasts to improve the management of living marine resources along the NAEC.

2 Methods

2.1 Numerical models

The high-resolution regional model used in this study is a 1/12° MOM6-based model of the Northwest Atlantic Ocean (MOM6-NWA12; Ross et al., 2023a). Briefly, MOM6-NWA12 is a regional ocean–sea ice model with explicit tides that spans a wide domain, including the Caribbean Sea, Gulf of Mexico, and North American east coast (Fig. 1). The model configuration that we use here is essentially the same as the one evaluated by comparing a reanalysis-forced historical simulation of the model with observations in Ross et al. (2023a), including the same grid, vertical coordinates, parameterization choices, and parameter values. The only difference is that the version used here does not include coupled biogeochemistry. The model is forced at the surface by atmospheric data and by river discharge from land and at the ocean open boundaries. The wide extent of the model grid, which is a subset of the even broader North and Equatorial Atlantic grid of Chassignet and Xu (2017), was designed to reduce the influence of the model being downscaled on the coastal and shelf environments and the western boundary current system by placing the ocean open boundaries far from these regions of interest. In the reanalysis-forced simulation spanning 1993 to 2019, Ross et al. (2023a) found that the model was capable of accurately reproducing mean conditions of temperature, salinity, mixed layer depth, and other physical properties. The model also simulated many aspects of historical variability, although some of the recent extreme warming trends in the region were underestimated.

The high-resolution regional model was used to downscale lower-resolution global retrospective forecasts produced by GFDL's SPEAR model (Delworth et al., 2020). As described in Lu et al. (2020), the SPEAR retrospective forecasts were run using a coupled atmosphere–land and ocean–sea ice model with a nominal 0.5° resolution for the atmosphere and 1° resolution for the ocean (the SPEAR_MED configuration). The suite of retrospective forecasts from SPEAR starts in 1991 and extends to the present, with recent and ongoing forecasts contributed to the North American Multi-Model Ensemble (NMME). Each SPEAR retrospective seasonal forecast was run for 1 year. Earlier forecasts contained 15 ensemble members, while more recent runs contributed to the NMME contain 30. Ocean initial conditions for the SPEAR retrospective forecasts were obtained by assimilating the Optimum Interpolation Sea Surface Temperature v2 (OISSTv2) sea surface temperature product, vertical profiles from Argo floats, and several other sources using an ensemble adjustment Kalman filter; see Lu et al. (2020) for details.

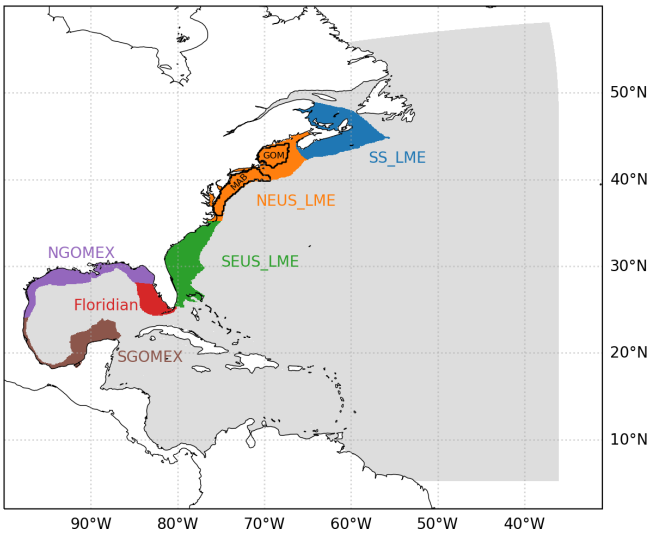

Figure 1. Map of the NWA12 ocean model domain (light gray shading) and the six marine ecosystems for which spatial averages were computed in the results: the Scotian Shelf (SS), Northeast US (NEUS), and Southeast US (SEUS) Large Marine Ecosystems (LMEs) and the Floridian and Northern and Southern Gulf of Mexico (NGOMEX and SGOMEX) ecoregions. Black outlines within the NEUS LME show two ecological production units: the Mid-Atlantic Bight (MAB) and the Gulf of Maine (GOM).

2.2 Downscaled retrospective forecast simulations

MOM6-NWA12 was initialized and used to run downscaled retrospective forecast simulations four times per year (beginning on the first day of the month at the start of each meteorological season – March, June, September, and December) for every year from 1993 through 2021. Initial conditions for the forecasts were derived from an analysis simulation of the regional MOM6-NWA12 model forced by the same atmosphere (ERA5; Hersbach et al., 2020), ocean (GLORYS12; Lellouche et al., 2021), and river (GloFAS; Alfieri et al., 2020) reanalyses as in Ross et al. (2023a). Additional details on the variables and boundary conditions used are provided in Ross et al. (2023a). One difference is that in the analysis simulation, temperature and salinity throughout the model domain were nudged (i.e., restored using Newtonian relaxation) towards monthly means from the GLORYS12 reanalysis with a 90 d damping timescale. Although the free-running simulation in Ross et al. (2023a) successfully recreated many

aspects of observed ocean variability and change, the addition of nudging helps maximize the accuracy of the initial conditions, while the moderate nudging timescale is intended to keep the simulation from being nudged too far away from the model's attractor and inducing a drift during the forecast when the nudging is removed. Instantaneous conditions saved at the beginning of each month of the historical simulation were used as initial conditions for the retrospective forecast experiments. For each initialization date, 10 ensemble members were integrated forward for 1 year. Each member used atmospheric forcing from a different SPEAR forecast member (discussed next) but began with the same initial conditions from the analysis simulation. Using the same initial conditions for each ensemble member will reduce the ensemble spread; however, as we show in the results (Sect. 3), the ensemble members quickly diverge after initialization and the effect of the identical initial conditions is minimal. We acknowledge, though, that deriving the initial conditions from a data assimilation process or a more sophisticated nudging method could improve the forecast spread and skill.

The downscaled forecasts were forced at the ocean surface with daily mean atmospheric data from the global SPEAR retrospective forecasts. Each of the 10 downscaled ensemble members was forced with the corresponding ensemble member from the SPEAR reforecasts. The forcing was derived directly from the SPEAR data; no correction for bias or drift was applied before downscaling. Preliminary experiments suggested that attempting to correct biases in the forcing yielded negligible changes in the downscaled forecast skill. The global SPEAR forecast simulations use a scheme that applies adjustments to the ocean tendencies based on typical analysis increments, and this reduces drift in the global model bias during the course of the forecast (Lu et al., 2020). A standard correction for bias and drift was applied to the downscaled forecasts after running the simulations (Sect. 2.3).

The regional MOM6-NWA12 model also requires ocean open boundary conditions, river discharge, and tidal forcing. Tidal forcing at the open boundaries and by the tidal potential, both of which can be accurately predicted far into the future with basic astronomy, were applied for the forecasts as in Ross et al. (2023a). For river discharge, which is substantially more challenging to forecast, we neglected any potential predictability and instead used a smoothed daily climatology of river discharge in the forecasts. The climatological mean for each day of the year was determined for each grid point in the 1993–2019 river forcing used in Ross et al. (2023a), which was derived from the GloFAS reanalysis (Harrigan et al., 2020). The climatologies were then smoothed with a ± 5 d triangular filter following Pegion et al. (2019) and Ross and Stock (2022).

The non-tidal open boundary conditions were also specified with monthly climatologies averaged over 1993–2019 from the daily GLORYS12 reanalysis temperature, salinity, velocity, and sea level data used in the analysis simulation. Climatology was used for the boundary data under the assumption that propagation of any predictable temperature or salinity signals from the model boundaries to the coastal and shelf regions of interest would occur over timescales longer than the 1-year length of the forecasts. For example, the propagation of anomalies from where the continental shelf intersects the northern model boundary to the Scotian Shelf and Gulf of Maine has been found to occur on timescales of 1–2 years (Brickman et al., 2018; Gonçalves Neto et al., 2021; Jackson et al., 2016; New et al., 2021). Preliminary experiments confirmed that the forcing used at the open boundaries made a negligible difference to the forecast skill as long as the forcing did not disrupt the Gulf Stream pathway. In addition to avoiding issues related to the Gulf Stream, using climatological boundary data also greatly simplifies the setup of the forecast simulations. Attempts to extend the forecasts beyond 1 year of prediction, however, would very likely require using the parent model forecasts as boundary conditions or devising some other way to allow predictable signals to enter through the boundary. The same is true for developing the forecasts to predict faster propagating sea level variations.

2.3 Forecast post-processing

From monthly mean time series of both the downscaled NWA12 and parent SPEAR model output, we calculated area-weighted averages of surface and bottom temperature and surface salinity within the six marine ecoregions shown in Fig. 1: three Large Marine Ecosystems (LMEs) along the US east coast (Scotian Shelf, Northeast US Continental Shelf, and Southeast US Continental Shelf) and three areas for the North and South Gulf of Mexico and the West Florida Shelf derived from corresponding marine ecoregions of the world (Spalding et al., 2007). For a few analyses, we also calculated averages within two smaller ecological production units (EPUs) for the Gulf of Maine and Mid-Atlantic Bight. To ensure that model area averages were consistent with conditions on the shelf, the regions used for averaging were further reduced to include only model points with depths 200 m or shallower in the three shallow Marine Ecoregions and 600 m or shallower in the remaining LMEs and EPUs.

A climatology for each initialization month (March, June, September, and December), lead month (0–11 months), variable, and region or grid cell was calculated using the ensemble mean of the full 1993–2021 set of forecasts. Forecast anomalies were calculated by subtracting the respective climatology from the forecasts. When compared with observed anomalies, the forecast anomalies effectively have the lead-dependent bias removed. Although this approach is commonly used, calculating the model climatology with the full set of forecasts will introduce some artificial skill by incorporating information about the future that would have been unknown in the past (Risbey et al., 2021); however, we em-

phasize that this information is only used for the correction of bias and drift.

2.4 Forecast evaluation

To evaluate the forecasts, we primarily relied on data from the GLORYS12 ocean reanalysis to represent the observed conditions. This high-resolution reanalysis compares favorably with in situ observations relative to other reanalysis products (Amaya et al., 2023; Carolina Castillo-Trujillo et al., 2023) and has the advantage over purely observed datasets of providing complete coverage over time and space. We note, however, that the initial conditions for the downscaled forecasts were derived by nudging towards the GLORYS reanalysis, which may confer a skill advantage on the downscaled forecasts. In this paper, we often refer to the GLORYS reanalysis as "observations" because it has been extensively evaluated regionally and to indicate where methodologically another observational or state estimate dataset could be used. As with the forecasts, monthly climatologies were calculated for each variable in the GLORYS dataset during the 1993–2021 time period and were subtracted from the data to produce anomalies, and spatial averages were calculated within the different marine ecoregions in Fig. 1.

The downscaled retrospective forecasts produced by MOM6-NWA12 and the corresponding parent forecasts produced by SPEAR were compared with the reanalysis dataset using the Pearson correlation coefficient and the mean bias (the mean difference between the model and the observations). To determine how the forecast skill varied with the time of year when the forecast was initialized and the lead time, we calculated these metrics separately for each initialization month and lead time.

To quantitatively test the hypothesis that downscaling produces an increase in forecast skill, we compared the correlation coefficients from MOM6-NWA12 and SPEAR using the method of Steiger (1980). This method tests the difference between two correlation coefficients obtained from two sets of forecasts that are compared against a common set of observations and are correlated with each other. Neglecting this commonality between the two forecasts would result in overly conservative estimates of significance (Siegert et al., 2017). We used this to conduct a two-sided test with a null hypothesis that the difference between the correlation coefficients for NWA12 and SPEAR was 0. For a fair comparison between the two models, all evaluations of the SPEAR forecasts were based on only the first 10 ensemble members that were downscaled with MOM6-NWA12.

To determine if the model forecasts outperform a simple forecast, we also used the Steiger (1980) test to compare the observation–model correlations with the correlations between the observations and a set of persistence reference forecasts. The persistence forecasts were computed by taking the reanalysis monthly mean anomaly from the month before each forecast was initialized and assuming that this anomaly would remain constant throughout the forecast.

For a second comparison between the skill of the parent and downscaled model forecasts, we used the DelSole and Tippett (2016) random walk test to assess the difference in the absolute error in the two models (the absolute value of the difference between the model forecast and the observation). This test uses a time series of historical forecasts from two models. Starting at 0 at the beginning of the time series (year 1993 here), it adds 1 each time the first model (NWA12) has a lower absolute error than the second (SPEAR) and subtracts 1 each time the second model has a lower absolute error than the first. This produces a time series of the cumulative sign of the errors, and a consistent trend in the time series indicates consistent better performance by one model. Changes in relative performance can be indicated by changes in the trend of the time series. Under the null hypothesis that the model mean absolute errors are equal, the time series follows a random walk with a probability distribution given by DelSole and Tippett (2016). For this test, unlike the other comparisons, we pooled the forecasts over all initialization times. To simplify the presentation of the results, the forecasts from both models were resampled to seasonal (3-month) averages, and the test was run for the resulting four seasonal lead times.

Finally, to assess whether the ensemble members from both models provide reliable probability information (i.e., whether the typical ensemble spread is consistent with the typical error in the ensemble mean), the average spread of the ensemble members was compared with the root mean square error (RMSE) of the ensemble mean forecast. If the ensemble is well calibrated, these two metrics will be equal, while underdispersion and overconfidence will be indicated by a larger RMSE and overdispersion and underconfidence by a larger spread (Fortin et al., 2014). The RMSE of the ensemble mean was calculated using the anomalies, which removes the error due to mean bias. The ensemble spread was calculated as the square root of the average ensemble variance using Eq. (16) of Fortin et al. (2014):

$$S = \sqrt{\frac{1}{T}\sum_{t=1}^{T} s_t^2}. \tag{1}$$

Here, s_t^2 is the unbiased estimate of the variance of the ensemble:

$$s_t^2 = \frac{1}{E-1}\sum_{e=1}^{E}\left(\overline{X_t} - X_{t,e}\right)^2, \tag{2}$$

T is the number of forecast times, E is the number of ensemble members, and X is the ensemble of forecasts.

3 Results

Sea surface temperature prediction skill for both the global SPEAR and downscaled NWA12 models outperforms a

baseline forecast of persistence across many initialization months, lead times, and LMEs (Fig. 2). The highest correlations are generally found in the Northeast US LME, where both NWA12 (Fig. 2d) and SPEAR (Fig. 2e) exhibit a diagonal pattern of high skill corresponding to forecasts of autumn and winter SST anomalies. This pattern is consistent with the reemergence of early spring temperature anomalies from below the summer mixed layer in the autumn and winter (Sect. 4.1). NWA12 substantially improves on the SST prediction skill of the SPEAR model in the Northeast US LME (Fig. 2f). This improvement is the highest for forecasts initialized in September and at later lead times in June and December, where the SPEAR forecasts have correlations that are sometimes negative and worse than persistence. A possible cause of this error is discussed in Sect. 4.1.2–4.1.3. Both models also have high forecast correlations in the Scotian Shelf LME (Fig. 2a–c), although with a less clear pattern to the skill and few statistically significant improvements in the downscaled model. In the remaining four southern LMEs (Fig. 2g–r), the correlation coefficients are comparatively lower than those in the SS and NEUS, and the difference between the two models is negligible. Most forecasts have a higher correlation than persistence, however.

Forecast correlation coefficients are higher for bottom temperature (Fig. 3), which partially reflects its increased persistence. Most downscaled forecast correlations are higher than the persistence correlation, however, though the majority are not significantly higher. In the Northeast US (Fig. 3d–f), the pattern of forecast skill and downscaling improvement is similar to that seen for surface temperature. In other regions, NWA12 bottom temperature predictions have significantly higher skill than SPEAR for some cases where they did not for surface temperature. This is perhaps not unexpected as the downscaled NWA12 bottom temperature forecasts were initialized with data from the same GLORYS12 reanalysis used as validation, and the lower correlated SPEAR bottom temperature values partially reflect the persistence of differences between the SPEAR initialization and the GLORYS12 reanalysis. With this caution in mind, skill and improvement on the Scotian Shelf (Fig. 3a–c) resembles the Northeast US sea surface temperature pattern of skill for forecasts verifying in autumn and winter. This pattern only appears in the downscaled predictions; in the global model predictions, some of the anomaly correlations are significantly lower than persistence. In the Southeast US (Fig. 3g–i), the downscaled model has skill greater than persistence and SPEAR across a wide range of times, except for forecasts verifying in December, which may reflect poor prediction of the mixing of surface water to the bottom. A similar pattern of limited improvement in forecasts of winter bottom temperature in the downscaled model is seen in the Northern Gulf of Mexico (Fig. 3m–o). However, the three LMEs along the shores of the Gulf of Mexico (Fig. 3j–r) have improved downscaled forecast skill for forecasts verifying in the late summer and early autumn.

The patterns of correlation seen in surface and bottom temperature also generally appear in the correlation skill for sea surface salinity (Fig. 4). For example, the NEUS (Fig. 4d–f), Floridian (Fig. 4j–l), and NGOMEX (Fig. 4m–o) regions have a general pattern of higher correlation for forecasts verifying in the autumn and winter. Similarly, the SEUS LME (Fig. 4g–i) has lower skill for predictions of winter, which is consistent with lower skill of winter bottom temperature and the possible connection to unpredicted mixing. In most regions, the improvements of the downscaled NWA12 over the global SPEAR are higher for surface salinity than for temperature. Because the downscaled forecast river discharge is set to climatology and precipitation and evaporation are poorly predicted and/or negligible drivers of salinity, the forecast skill here must primarily come from better resolution of salinity anomalies in the initial conditions and better prediction of their advection and dispersion (Sect. 4.1.2) or better prediction of salinity advection by anomalous currents.

Over all times and regions, NWA12 generally has cooler SSTs than than SPEAR, which is apparent in how NWA12 worsens the cool biases in SPEAR (indicated by shades of purple in the right column of Fig. 5) and improves the warm biases (green colors). For the SS and NEUS LMEs, the bias has generally worsened in NWA12 relative to SPEAR despite the improved anomaly correlation, although the bias has improved in NWA12 for forecasts verifying in autumn and early winter when the downscaled forecasts have the most skill (Fig. 2). In these two regions (Fig. 5a, d), NWA12 has a pronounced cold bias that develops in spring and dissipates in summer. This bias as well as the warm autumn bias appear to be inherited from the global model, although the global biases are generally shifted 1–3 months earlier. It should be noted that SPEAR includes an ocean tendency adjustment (Lu et al., 2020) which helps keep biases in check during the course of the forecasts. The benefit of this adjustment is also evident from the lack of bias drift over lead times in both models; in other words, forecast biases depend mostly on verification date and not on lead time. Finally, whereas SST correlation in the southern three LMEs was similar in NWA12 and SPEAR, NWA12 typically has lower SST bias in these regions.

Bottom temperature biases also show a strong dependence on verification month, except in the Southeast US region (Fig. 6). NWA12 generally has a cool bias in the two northern LMEs and a warm bias elsewhere. The cool bias develops as the forecast simulations reach winter and persists for the remainder of the forecast (note that June initializations in Fig. 6a, d have near-zero bias for summer and autumn, but in December, initializations the following summer and autumn are biased cool). In the southern three LMEs (Fig. 6j, m, p), NWA12 has a warm bias in forecasts of summer and autumn, but this bias does not persist into the winter. NWA12 has substantially lower bottom temperature bias than SPEAR in the Southeast US LME (Fig. 6g–i), which has a narrow shelf and is dominated by the western boundary current. In

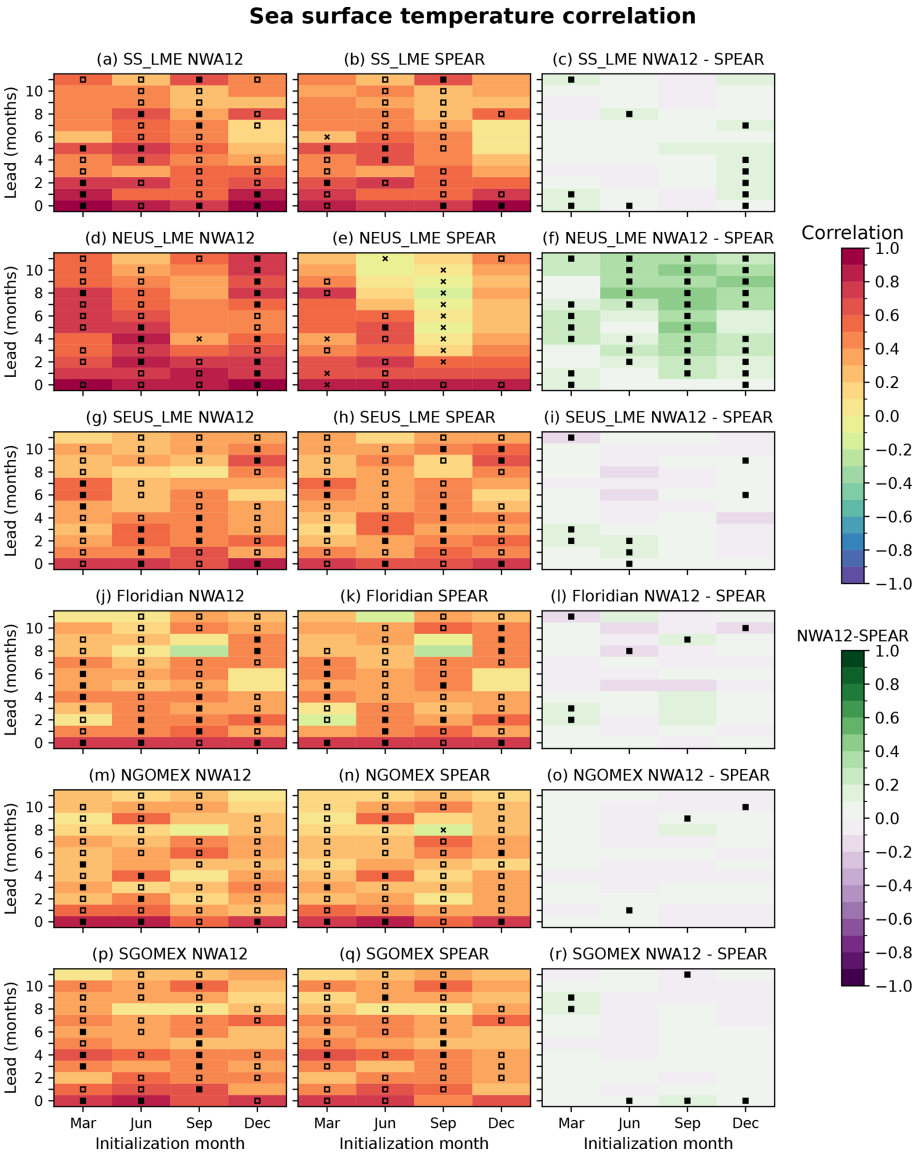

Figure 2. Comparison between sea surface temperature forecast–observation correlation coefficients for NWA12 (left column) and SPEAR (middle column) and the difference between the two (right column). In the left two columns, an open black square indicates a correlation that is greater than the persistence–observation correlation, a filled black square indicates that this difference is statistically significant, and a black x indicates a correlation that is significantly lower than the persistence–observation correlation. In the right column, a filled black square indicates that the anomaly correlation of NWA12 is significantly different than the anomaly correlation of SPEAR.

other regions, the differences between the global and downscaled prediction biases are more mixed.

Some of the largest and most consistent improvements from downscaling are found in the reduced sea surface salinity biases (Fig. 7). The reduced bias is most notable in the NEUS (Fig. 7d–f) and NGOMEX (Fig. 7m–o) LMEs, both regions of high river discharge and sharp coastal salinity gradients, where SPEAR underestimates the mean surface salinity. Across all regions, the downscaled model generally has a mild salty bias in contrast to SPEAR's fresh biases. It should be noted that the GLORYS12 reanalysis used as the observations in this comparison does not simulate salinity as well as it does temperature (Amaya et al., 2023; Carolina Castillo-Trujillo et al., 2023) and that some of the reduction in bias may be due to the use of this reanalysis in the derivation of the initial conditions for the downscaled forecasts.

Aggregated across all four initialization months and averaged into 3-month seasons, the random walk metric shows that NWA12 SST anomalies have a statistically significant lower mean error than SPEAR (positive random walk value) for all lead seasons in the Northeast US (Fig. 8b) and all but the last season in the Floridian region (Fig. 8d). Overall, the

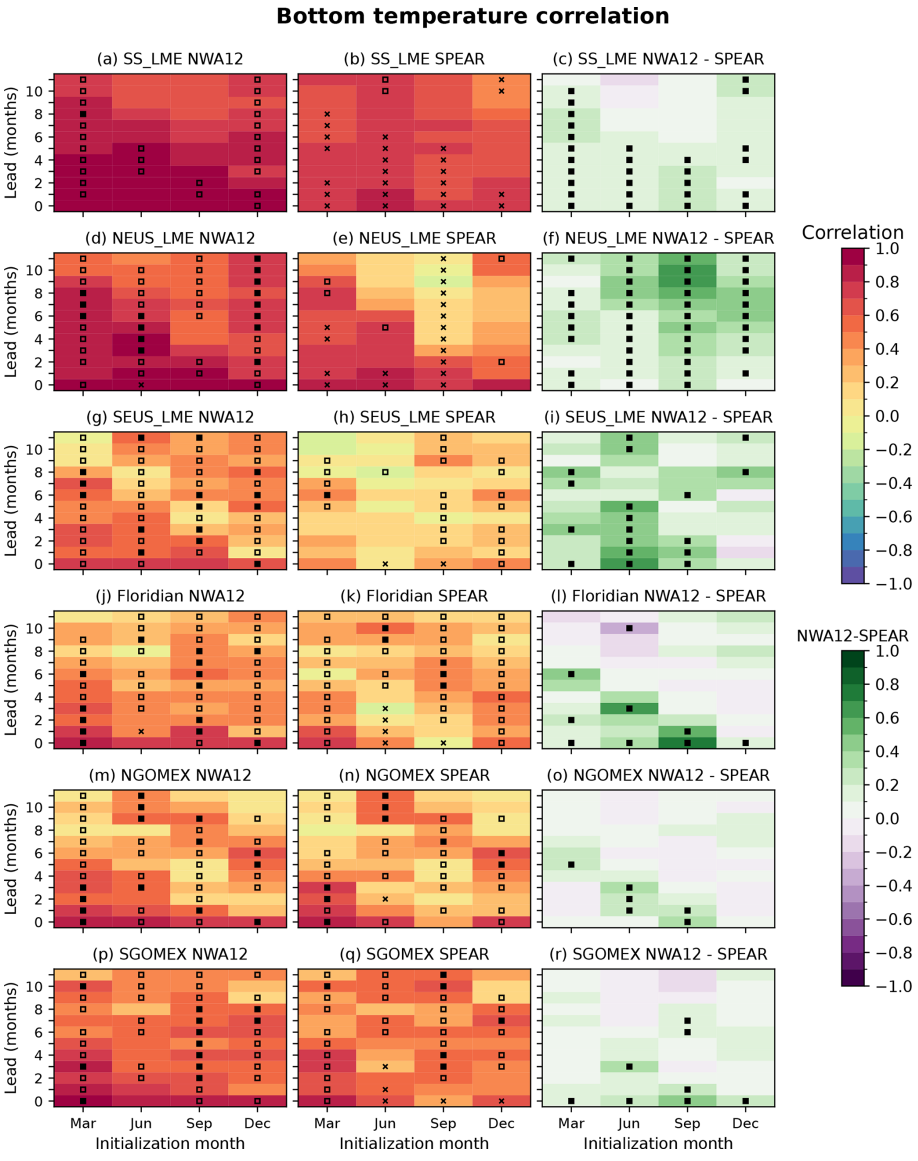

Figure 3. Comparison between bottom temperature forecast–observation correlation coefficients for NWA12 (left column) and SPEAR (middle column) and the difference between the two (right column). In the left two columns, an open black square indicates a correlation that is greater than the persistence–observation correlation, a filled black square indicates that this difference is statistically significant, and a black x indicates a correlation that is significantly lower than the persistence–observation correlation. In the right column, a filled black square indicates that the anomaly correlation of NWA12 is significantly different than the anomaly correlation of SPEAR.

mean absolute error (MAE) of predicted SST anomalies in the Northeast US is approximately 20 % lower in the downscaled NWA12 model than in the global model. NWA12 also has significantly lower error for the first season after initialization in the Southeast US (Fig. 8c) and marginally statistically significant improvements for the Southern Gulf of Mexico in the first and third seasons (Fig. 8f) although, in these cases, the magnitude of the improvement in the overall MAE is small. None of the tests conclude that the error in NWA12 is significantly worse (i.e., no final value of the test lies below the 90 % confidence interval for a random walk with 0 mean) although the final lead season nearly ends as significantly worse in the Southern Gulf of Mexico and was occasionally significantly worse in the NGOMEX. For the Northeast US and, to a lesser extent, the Scotian Shelf LMEs, the results suggest a possibility that the downscaled forecasts were no longer better than the global model forecasts after around 2016 (indicated by roughly flat lines); however, the sample size is too small for this result to be conclusive.

Despite the use of initial conditions that were identical for each of the 10 ensemble members, the downscaled SST forecasts have similar or better ensemble spread character-

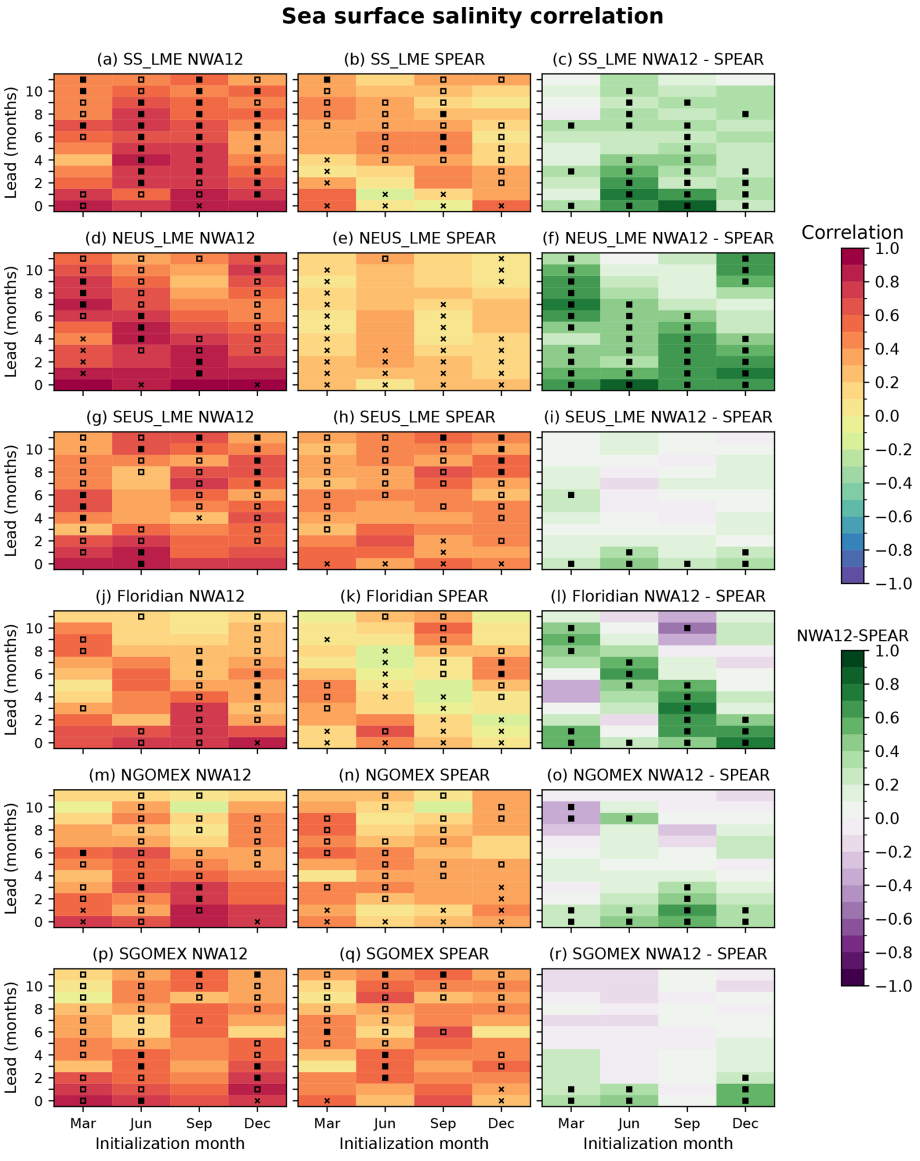

Figure 4. Comparison between sea surface salinity forecast–observation correlation coefficients for NWA12 (left column) and SPEAR (middle column) and the difference between the two (right column). In the left two columns, an open black square indicates a correlation that is greater than the persistence–observation correlation, a filled black square indicates that this difference is statistically significant, and a black x indicates a correlation that is significantly lower than the persistence–observation correlation. In the right column, a filled black square indicates that the anomaly correlation of NWA12 is significantly different than the anomaly correlation of SPEAR.

istics than the global model (Fig. 9). In all regions, both models have similar ensemble spread in the first forecast month, which indicates that the ensemble of atmospheric conditions provided by SPEAR is able to rapidly force the downscaled ensemble members to diverge from their identical initial conditions. Across most lead times in the two northern LMEs (Fig. 9a–b), NWA12 has both higher ensemble spread and lower RMSE than the SPEAR forecasts, although the RMSE generally remains greater than the ensemble spread (i.e., NWA12 remains overconfident but less so). Differences between the two models are smaller in the remaining four regions, aside from NWA12 having higher spread than SPEAR in the Southeast US and lower spread and RMSE than SPEAR in the Floridian region. Both models are also overconfident (RMSE greater than spread) in the Southeast US and Northern Gulf of Mexico regions. In the Southern Gulf of Mexico LME, only the first forecast month is substantially overconfident. However, the RMSE of both models rapidly converges to the RMSE of the climatology, which indicates no forecast skill.

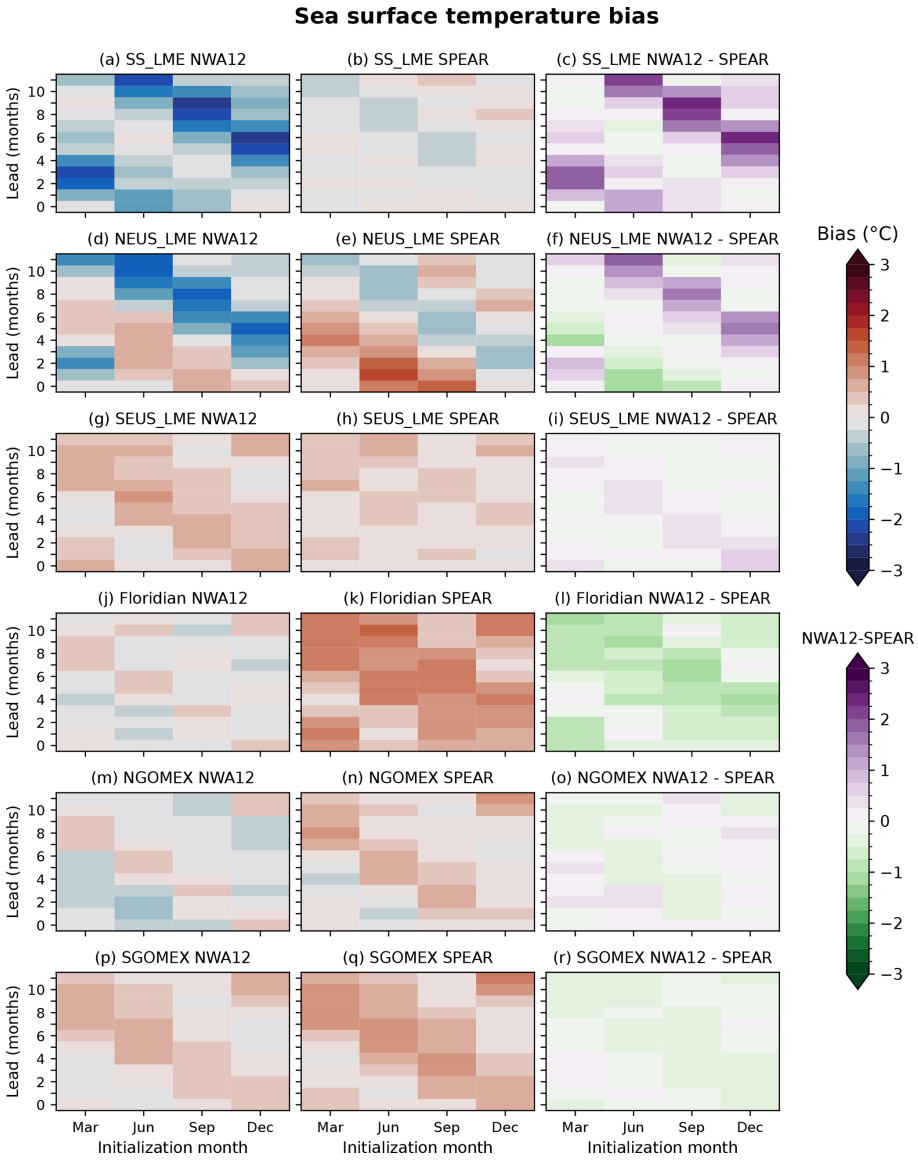

Figure 5. Comparison between sea surface temperature forecast–observation mean bias for NWA12 (left column) and SPEAR (middle column) and the difference between the magnitude of the mean bias in NWA12 and SPEAR (right column). Green shades in the right column indicate that NWA12 mean SST is closer to the observed mean than SPEAR, while purple shades indicate that the SPEAR mean is closer than NWA12.

4 Discussion

In this section, we examine three mechanisms that contribute to the prediction skill and how better resolution of these mechanisms improves the skill in the downscaled forecasts: the reemergence of SST anomalies, advection of salinity anomalies, and temperature and salinity anomalies driven by variability of the Gulf Stream (Sect. 4.1). This examination primarily focuses on the Northeast US region, where downscaling produced consistent improvements in surface and bottom temperature and surface salinity. After analyzing the mechanisms that contribute to prediction skill, we determine whether the skill would be reduced if we accounted for the effect of the predictable trend in SST (Sect. 4.2). Finally, we discuss whether there is an opportunity to increase the forecast skill by increasing the number of members in the model ensemble (Sect. 4.3).

4.1 Sources of model skill and improvement from downscaling

4.1.1 Reemergence of temperature anomalies

The reemergence mechanism, whereby temperature anomalies are deeply mixed during the winter, then insulated from

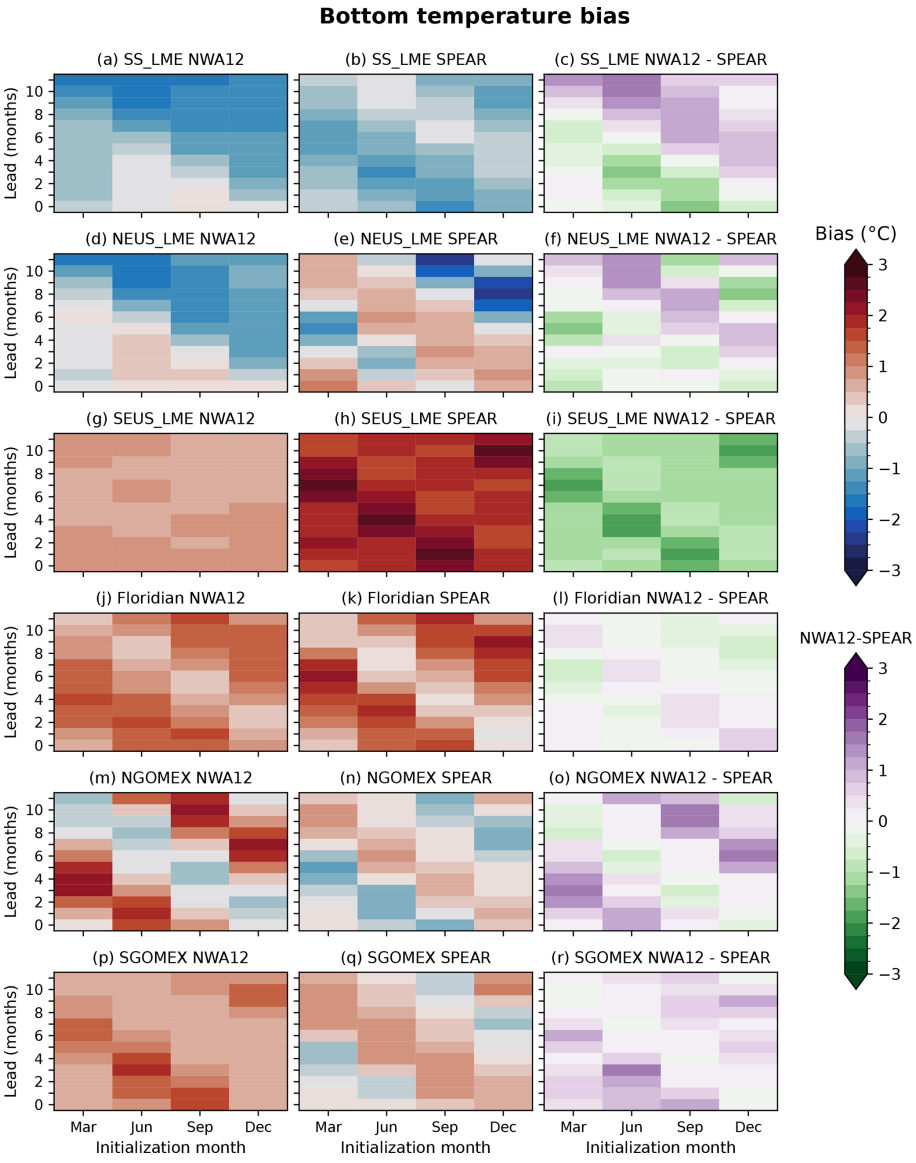

Figure 6. Comparison between bottom temperature forecast–observation mean bias for NWA12 (left column) and SPEAR (middle column) and the difference between the magnitude of the mean bias in NWA12 and SPEAR (right column). Green shades in the right column indicate that NWA12 mean bottom temperature is closer to the observed mean than SPEAR, while purple shades indicate that the SPEAR mean is closer than NWA12.

the atmosphere beneath the shallow summer mixed layer, and finally mixed back to the surface in the autumn and early winter as the mixed layer deepens again (e.g., Alexander et al., 1999), is one prominent source of long-term seasonal predictability in many regions (Jacox et al., 2020). In the results (Fig. 2), we found that SST forecasts for the Northeast and Southeast US LMEs were most skillful for predictions of autumn and winter SST anomalies even for forecasts initialized in March or the previous December, which is consistent with predictable reemergence of SST anomalies.

To quantitatively evaluate the observed and predicted reemergence of SST anomalies, we first calculated the correlation between detrended SST anomalies in December and March and anomalies in subsequent months for the Northeast and Southeast US LMEs and two smaller ecological production units (EPUs) within the NEUS LME (Fig. 10). These region-average anomalies were calculated only for areas deeper than 60 m to avoid including nearshore regions that are too shallow to develop deep anomalies that reemerge. We also calculated the index of reemergence developed by Geiss et al. (2020), which is the difference between the actual lagged correlation and the lagged correlation expected from a best-fit first-order autoregressive process. Reemergence is indicated by a lagged correlation that is higher than

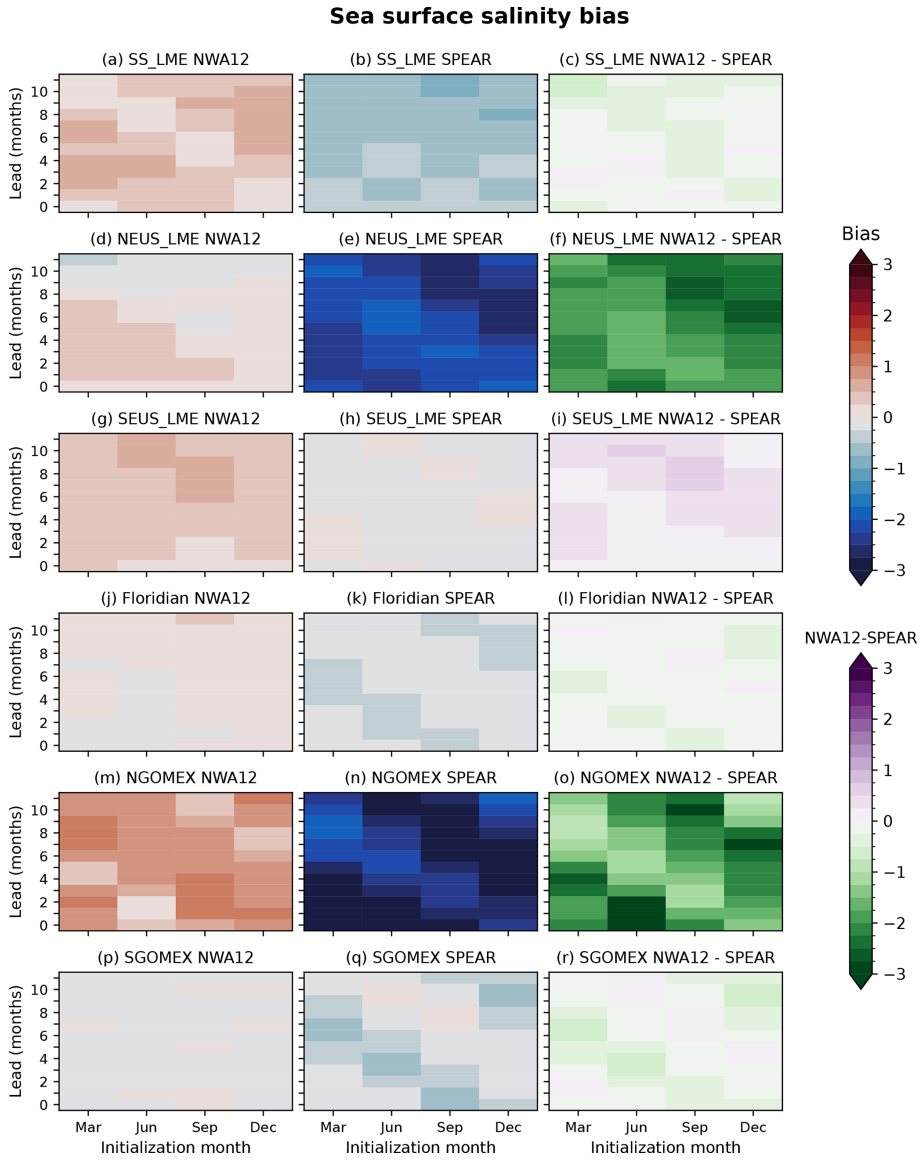

Figure 7. Comparison between sea surface salinity forecast–observation mean bias for NWA12 (left column) and SPEAR (middle column) and the difference between the magnitude of the mean bias in NWA12 and SPEAR (right column). Green shades in the right column indicate that NWA12 mean surface salinity is closer to the observed mean than SPEAR, while purple shades indicate that the SPEAR mean is closer than NWA12.

expected; in Fig. 10, we plot black triangles when the difference is greater than 0.1. This index is similar to the approach developed by Byju et al. (2018), who developed an index using the difference between lagged correlations with winter temperature in the following 6 and 12 months, except that it can identify reemergence in any month. Here we fit the autoregressive process to data from 1–5 months after the start month by minimizing the RMSE of the autoregression over data from these months. For the forecasts, the 10 ensemble members were pooled together rather than using the ensemble mean.

In the GLORYS reanalysis, reemergence is evident in the Northeast US LME as a significant increase in correlation in October or November after a minimum in summer (Fig. 10a–b). Over the two EPUs within the Northeast US, the reemergence is strongest in the Gulf of Maine (GOM; Fig. 10c–d). In the Southeast US LME (Fig. 10g–h), March SSTs have a similar reemergence signature, with a weaker magnitude but a longer persistence into winter.

In the global SPEAR model, winter SST anomalies in the Northeast US tend to persist for too long throughout the summer, especially in the Mid-Atlantic Bight (MAB). March SST anomalies in the Southeast US also persist for too long

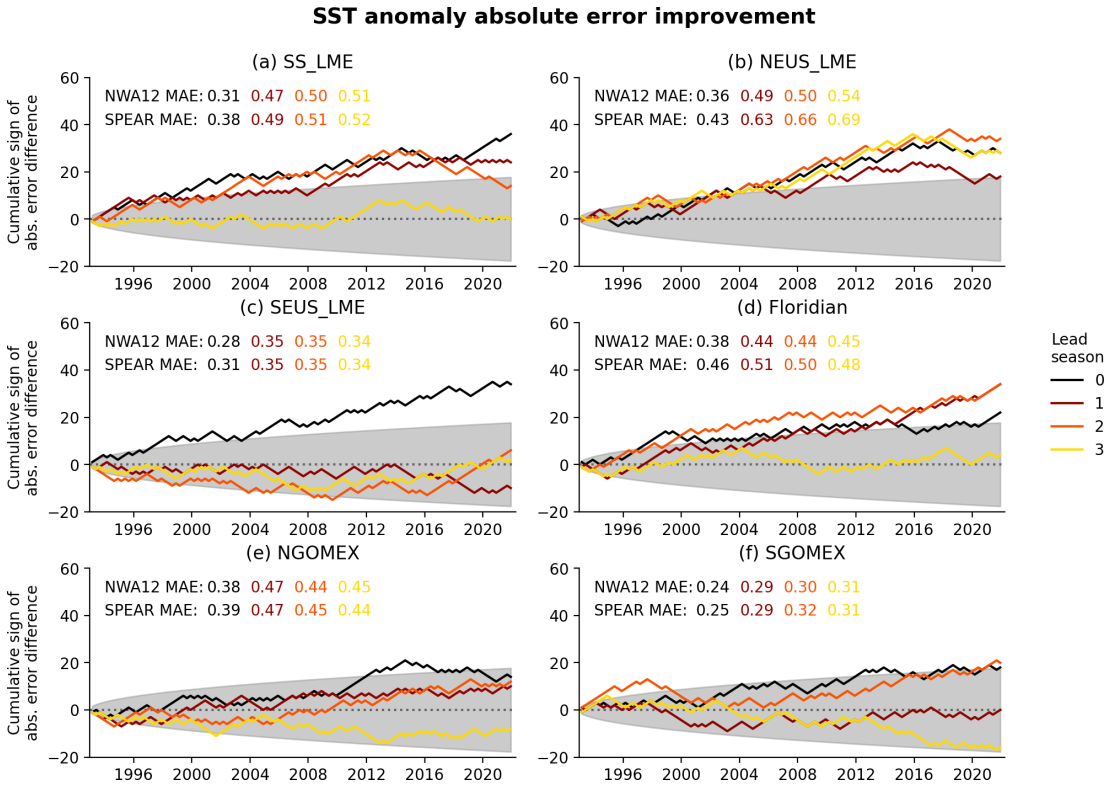

Figure 8. Cumulative sign of the difference between the absolute errors in SST forecasts from NWA12 and SPEAR. Forecasts have been aggregated into 3-month seasonal averages. The gray-shaded region indicates where the difference in absolute error between the two models is not statistically significant at $\alpha = 0.1$. Annotations in the top left of each panel give the overall mean absolute error (MAE) of the model forecasts for each lead time.

in the global model, whereas December anomalies are eroded too rapidly in the summer. However, in the majority of cases, the global model forecasts still maintain a signal of reemergence (black triangles in Fig. 10) although, in some cases, the SPEAR model reemergence is delayed compared to the GLORYS reanalysis.

Compared to the global SPEAR forecasts, the downscaled NWA12 forecasts generally have lagged SST correlations in Fig. 10 that are closer to the GLORYS reanalysis and are not overly persistent. Although downscaling improves the lagged correlation, it does not completely eliminate all biases in reemergence; NWA12 is similar to SPEAR in terms of months when significant reemergence is detected and the magnitude of the reemergence signal. For example, compared to the reanalysis, both models have a much stronger reemergence of December SST anomalies in the MAB.

The presence of reemergence in the forecasts is consistent with the reemergence of SST forecast skill in Fig. 2, particularly for the downscaled predictions. In the Northeast US, the skill of forecasts initialized in both March and December peaks in the subsequent November, which matches the timing of the forecast reemergence of SST anomalies in Fig. 10. In the Southeast US, March-initialized forecast skill has a minor reemergence in October, which matches the weak reemergence of SST in the observations and forecasts. December-initialized skill, however, has a large peak in September, which does not match any modeled or observed signal of reemergence. The most notable bias that appears in the downscaled results is a tendency for the strongest reemergence signal to appear 1–2 months late in the Northeast US LME and the two EPUs.

4.1.2 Advection of water masses

The Nova Scotia Current is a major advective pathway for cool, fresh surface water to enter the Northeast US LME (Du et al., 2021; Grodsky et al., 2017). In Fig. 11, we show the correlation between June sea surface salinity averaged over a box on the Scotian Shelf just outside of the Northeast US LME (black squares in the left panels of Fig. 11) and sea surface salinity in following months. The panels in the top row, from the analysis simulation, show evidence that salinity anomalies starting on the Scotian Shelf are advected to the eastern Gulf of Maine in 1–3 months and continue counterclockwise around the Gulf before beginning to reach the Mid-Atlantic Bight after around 7 months in, which is consistent with the well-known direction and speed of the circulation in the Gulf of Maine (Smith et al., 2001). Although

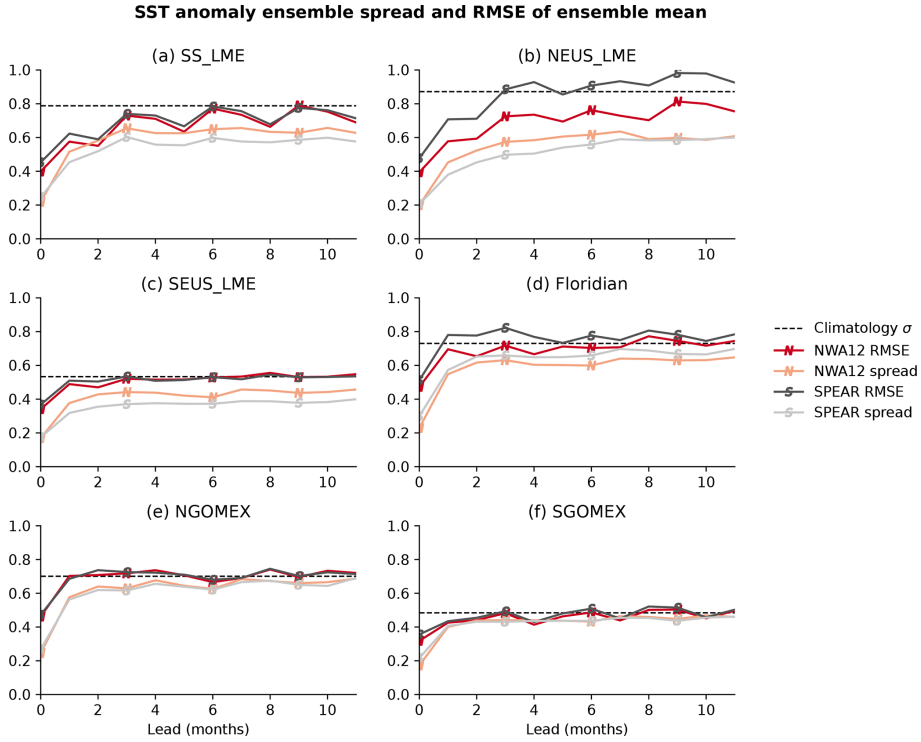

Figure 9. RMSE of NWA12 and SPEAR ensemble mean SST anomaly forecasts (dark red and gray, respectively), spread of the model ensemble members (light red and gray), and standard deviation of the observed anomalies (dashed black line). Letter markers on the lines mark whether the line is for NWA12 (N) or SPEAR (S).

the Nova Scotia Current is strongest in the winter (Han et al., 1997), we found the strongest advection signal for salinity anomalies by starting with June Scotian Shelf salinity, perhaps due to the better maintenance of surface properties during stratified summer conditions; the influence of other currents, such as transport through the Northeast Channel, that are stronger in the spring and summer (Smith et al., 2001); or the process of advection by the cyclonic circulation in the Gulf of Maine that strengthens during the summer and autumn and reaches a maximum in December (Xue et al., 2000).

In the second row of Fig. 11, we investigate whether the downscaled forecast model can simulate this advection process by performing a similar analysis, comparing lead-0 Scotian Shelf salinity from forecasts initialized in June with forecast sea surface salinity in subsequent months. The forecasts do capture the advection to the Eastern Gulf of Maine in the first few months. After lead 5 (November), the advection seems to stall, with correlations remaining high in the Gulf of Maine but not increasing in the MAB. Nevertheless, the ability to roughly capture the lagged correlations associated with this coastal advection likely contributes to the forecast skill for surface salinity and temperature for a wide range of lead times in forecasts initialized in June (in addition to the reemergence process, which contributes to predictability around leads 3–6).

By comparison, the 1° SPEAR model poorly predicts the initial advection of salinity anomalies from the Scotian Shelf. Although SPEAR does correctly predict that a salinity anomaly from the Scotian Shelf advects to the Northeast US LME, the advection is too fast (note that an anomaly reaches Georges Bank by lead 3, September, in SPEAR versus around lead 7, January, in the analysis simulation), and the low correlations indicate poor preservation of the advected anomaly. On the other hand, the diffuse and low correlation in SPEAR is fairly consistent with the analysis simulation by lead 7.

4.1.3 Gulf Stream variability and trends driven by climate change

Over the last 2 decades, most of the Northwest Atlantic Ocean within the model domain has warmed substantially faster than the rest of the global ocean, and the warming has been particularly rapid along the Northeast US continental shelf (Glenn et al., 2015; Pershing et al., 2015; Seidov et al., 2021; Wang et al., 2023). The rapid Northeast US warming is consistent with the effect of projected slowing of the overturning circulation associated with anthropogenic climate change (Caesar et al., 2018; Saba et al., 2016) and has been attributed to several factors, including a northward shift of the western portion of the Gulf Stream (Chi et al.,

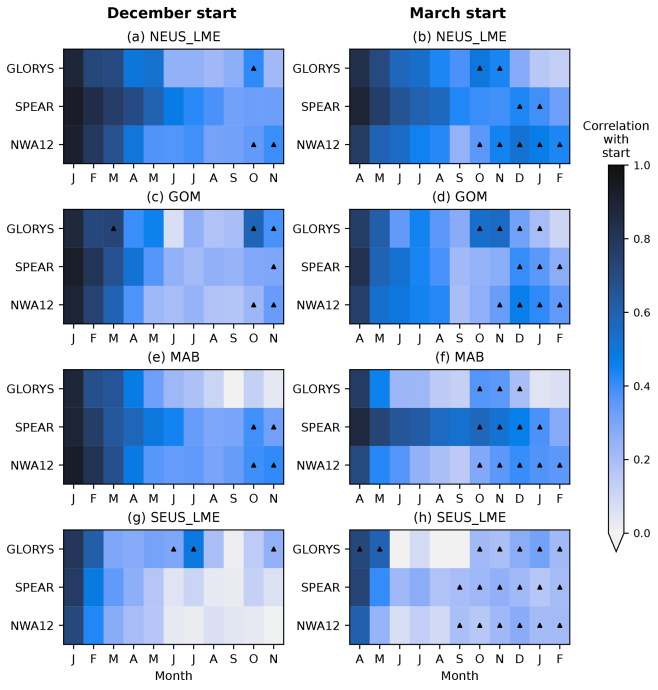

Figure 10. Correlations between December (left panels) and March (right panels) SST anomalies in different regions and SST anomalies in subsequent months. Black triangles are plotted where the Geiss et al. (2020) index exceeds 0.1; these triangles indicate where the correlation is greater than expected from a first-order autoregressive process.

2021; W.-Z. Zhang et al., 2020) and increased shedding of warm eddies (Gangopadhyay et al., 2019, 2020; Silver et al., 2023). Recent warming trends have been the greatest in the summer and autumn (Friedland et al., 2023), the same seasons when the Gulf Stream is closest to the continental shelf (Du et al., 2021) and sheds the most warm core eddies (Silver et al., 2023).

In the Northeast US LME, the greatest benefit of downscaling for forecasts of surface temperature and salinity was found in forecasts initialized in September and verifying over the next several months (Figs. 2, 4). When broken down into smaller regions (not shown), this improvement was greatest in the MAB region, which is the nearest to the Gulf Stream. Our hypothesis is that some of the improvement from downscaling stems from the improved simulation of Gulf Stream variability and trends and the resulting impacts on surface temperature and salinity both in the initial conditions and during the forecast.

To test this hypothesis, we begin by examining SST trends in the observations and in the two different sets of forecasts (Fig. 12). For the forecasts, we evaluate the trend across initialization years from forecasts of a given lead time (or verification month); note that the trend is the same for the sea surface temperature and the lead-dependent SST anomaly. Consistent with Friedland et al. (2023), the linear warming trend in the observations during 1993–2022 was the greatest during late summer and autumn although the confidence intervals are broad. Trends in the downscaled forecasts are slightly higher than observed and are fairly consistent regardless of initialization or lead time, again with confidence intervals that overlap the other forecasts and observations in all cases (Fig. 12a). In the parent SPEAR forecasts, however, most SST trends are lower than the observations, particularly in winter and spring and in forecasts initialized in summer and autumn when the predicted SST trends are actually negative (Fig. 12b). The trend is the most negative in September-initialized forecasts of April SST. The periods of erroneously low SST trends overlap with the periods of low or negative SST forecast skill in the SPEAR model (Fig. 2).

In Fig. 13, we examine the spatial patterns of SST trends in the Northeast US and Scotian Shelf regions from observations and September-initialized forecasts from the NWA12 and SPEAR models. The downscaled NWA12 forecasts have the right SST trends during the initial month and continue to correctly simulate the trends over the course of the forecasts. In forecasts of September (lead 0), the SPEAR forecasts have been correctly initialized with warming SSTs in the Gulf of Maine, but SSTs in the initial month have been incorrectly cooling in the Mid-Atlantic Bight. This cooling trend becomes worse in subsequent forecast months and extends along the shelf break north of the typical Gulf Stream position in February.

Finally, in Fig. 14, we examine the trends of geostrophic current speeds in the AVISO satellite altimetry product provided by the Copernicus Marine Service (https://doi.org/10.48670/moi-00148, EU Copernicus Marine Service Product, 2023a) and the two forecast models. A northward shift of the Gulf Stream is evident in the satellite observations in Fig. 14a, d, and g as a line of increased current speed immediately north of a line of decreased speed, particularly west of 67° W, which is both consistent with other observations (Chi et al., 2021; W.-Z. Zhang et al., 2020) and simulated remarkably well in the downscaled forecasts (Fig. 14b, e, h). The parent forecasts, however, have weak current speed trends that are more consistent with a modest southward shift of the Gulf Stream (Fig. 14c, f, i).

These results support our hypothesis that the improved forecasts in the downscaled model partially stem from improved simulation of the Gulf Stream and its impact on variability and trends in the adjacent coast and shelf region. This ability to match the observed Gulf Stream position and variability was also found by Ross et al. (2023a) in a reanalysis-forced historical simulation and reflects well on the capabilities of the NWA12 model. We emphasize that this result shows the benefits of high-resolution simulations rather than a deficiency specifically with the SPEAR model. The Gulf Stream is well known to require ocean resolution of $1/10°$ or finer to simulate properly (Chassignet and Garaffo, 2001; Chassignet and Xu, 2017) and the observed shift in the Gulf Stream occurred over roughly a degree of latitude

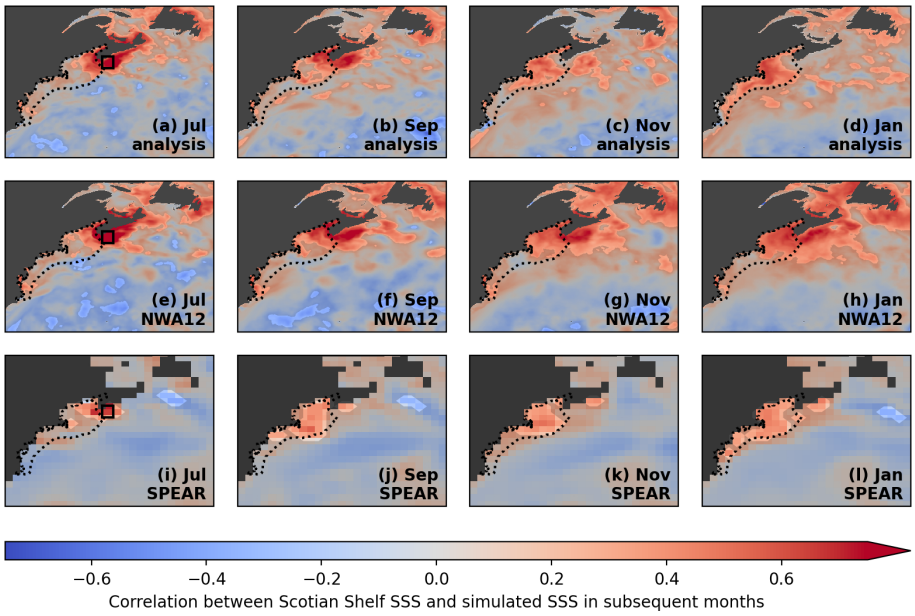

Figure 11. Correlation between June sea surface salinity in the southwest Scotian Shelf (within the black box in panels **a**, **e**, and **i**) and sea surface salinity 1, 3, 5, and 7 months later from the analysis simulation **(a–d)**, the June-initialized downscaled forecasts **(e–h)**, and the June-initialized SPEAR forecasts **(i–l)**. The dotted lines show the boundary of the Northeast US LME. Areas overlaid with light-gray shading are not statistically significant at the 90 % level using the Bretherton et al. (1999) method.

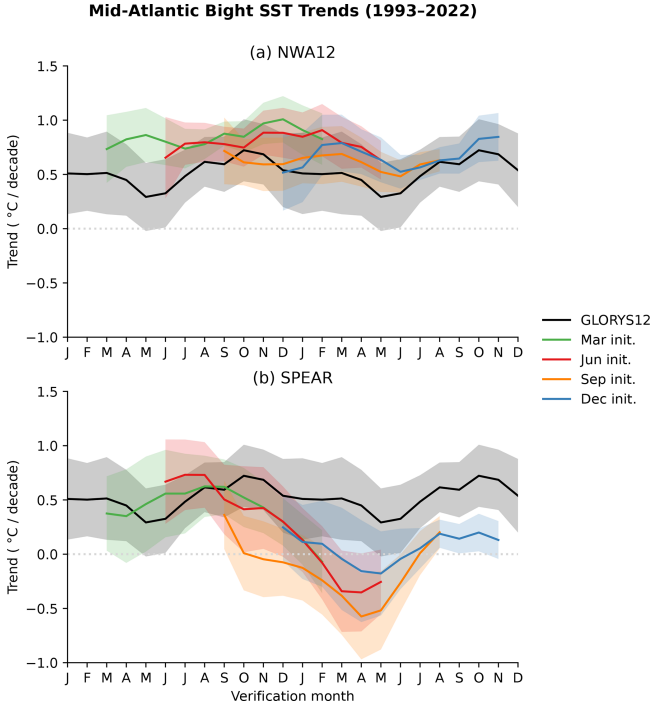

Figure 12. Linear trend of the Mid-Atlantic Bight average SST as a function of the forecast initialization month and lead time. The shading denotes 90 % confidence intervals.

(Wang et al., 2022), so there is no reason to expect that any model with a resolution on the order of 1° for the ocean would be able to accurately simulate the Gulf Stream shift or the resulting temperature and salinity changes.

To illustrate the benefits of resolving Gulf Stream variability at the event scale, in Figs. 15 and 16, we show output from the nudged historical simulation and downscaled forecasts, respectively, for a pronounced warming event along the Northeast US and Scotian Shelf in September to December of 2020. Conditions in September featured a weak warm anomaly south of Newfoundland and west of the Grand Banks, which strengthened over time and was joined by a second warm anomaly to the west in November (top row of Fig. 15). Both anomalies were associated with northward excursions of the Gulf Stream (middle row) and resulted in increased temperature and salinity in the Northeast Channel (point 1, bottom-left panel) and near the Scotian Shelf (point 2, bottom-right panel). In the Northeast Channel, temperature and salinity was near-average at all depths during September (indicated by same-colored plus symbols that are near each other). As time progressed, the salinity increased at all depths, deep waters became warmer, and near-surface waters cooled but much less so than usual. By December (square symbols), surface conditions were approximately 2° warmer and 0.5 units saltier than average, and water at 105 m depth was over 3° warmer and slightly less than 0.5 units saltier than average. Near the Scotian Shelf (point 2), a similar pattern was observed, with the most anomalous conditions found in October. At depth, the temperature and salinity increase from September to October was consistent with intrusion water from the Gulf Stream (marked by GS), and

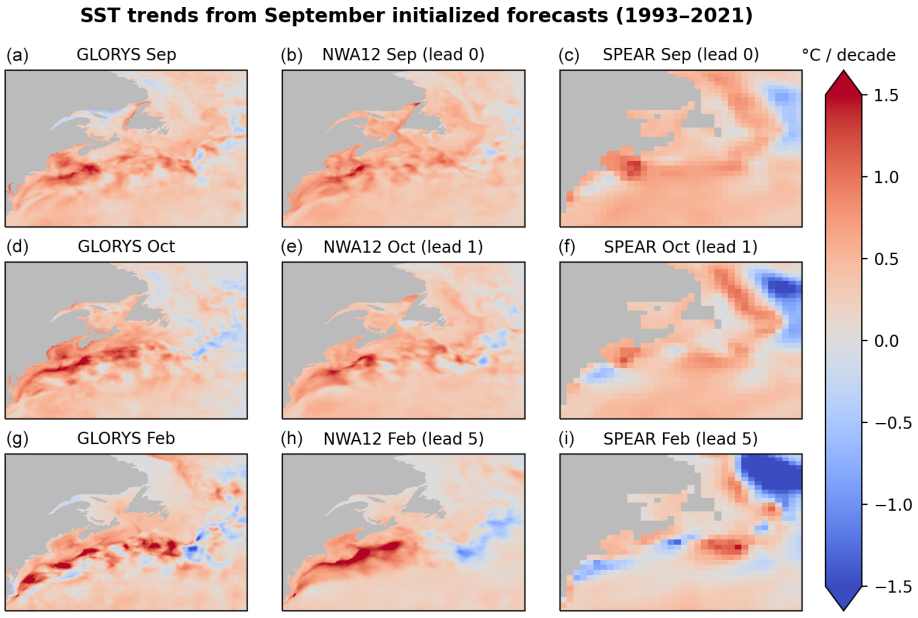

Figure 13. Observed linear SST trends from the GLORYS12 reanalysis in September, October, and February **(a, d, g)**, and SST trends in September-initialized forecasts that verify in the same months from the NWA12 model **(b, e, h)** and the SPEAR model **(c, f, i)**.

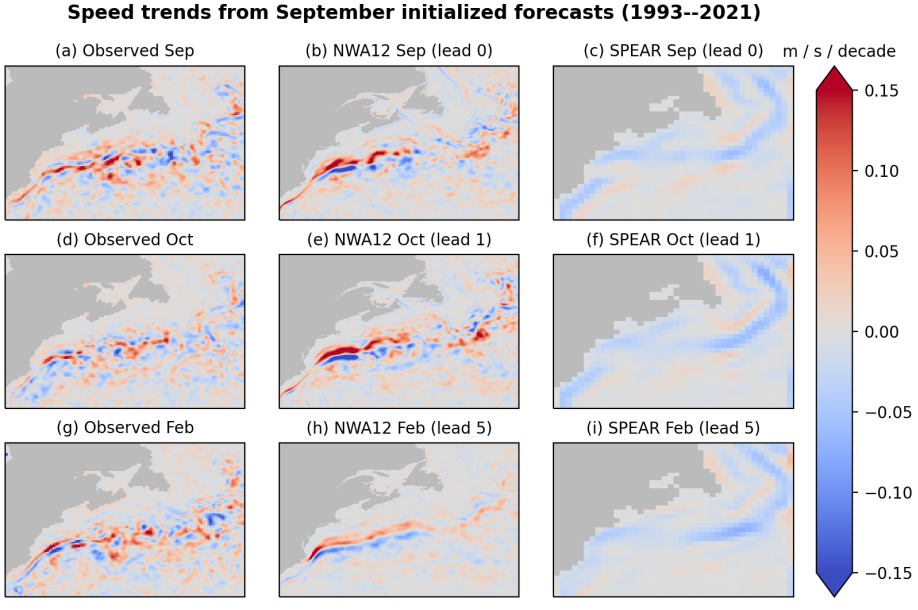

Figure 14. Observed linear geostrophic current speed trends from the AVISO satellite altimetry product in September, October, and February **(a, d, g)**, and speed trends in September-initialized forecasts that verify in the same months from the NWA12 model **(b, e, h)** and the SPEAR model **(c, f, i)**.

October conditions were over 4° warmer and 1 unit saltier than average.

This event was predicted remarkably well by the ensemble mean of the downscaled forecasts initialized at the beginning of September 2020 (Fig. 16). The mean forecast correctly predicted the eastern warm SST anomaly in October 2020 joined by a western warm anomaly connected to a northward Gulf Stream fluctuation in November 2020. The forecast also correctly predicted increasing salinity at all levels and warming at depth in the Northeast Channel (bottom-left panel) and advection of Gulf Stream-like water off the Scotian Shelf in October (bottom-right panel). Note that this plot shows the raw forecast without the lead-dependent climatology subtracted as in other plots.

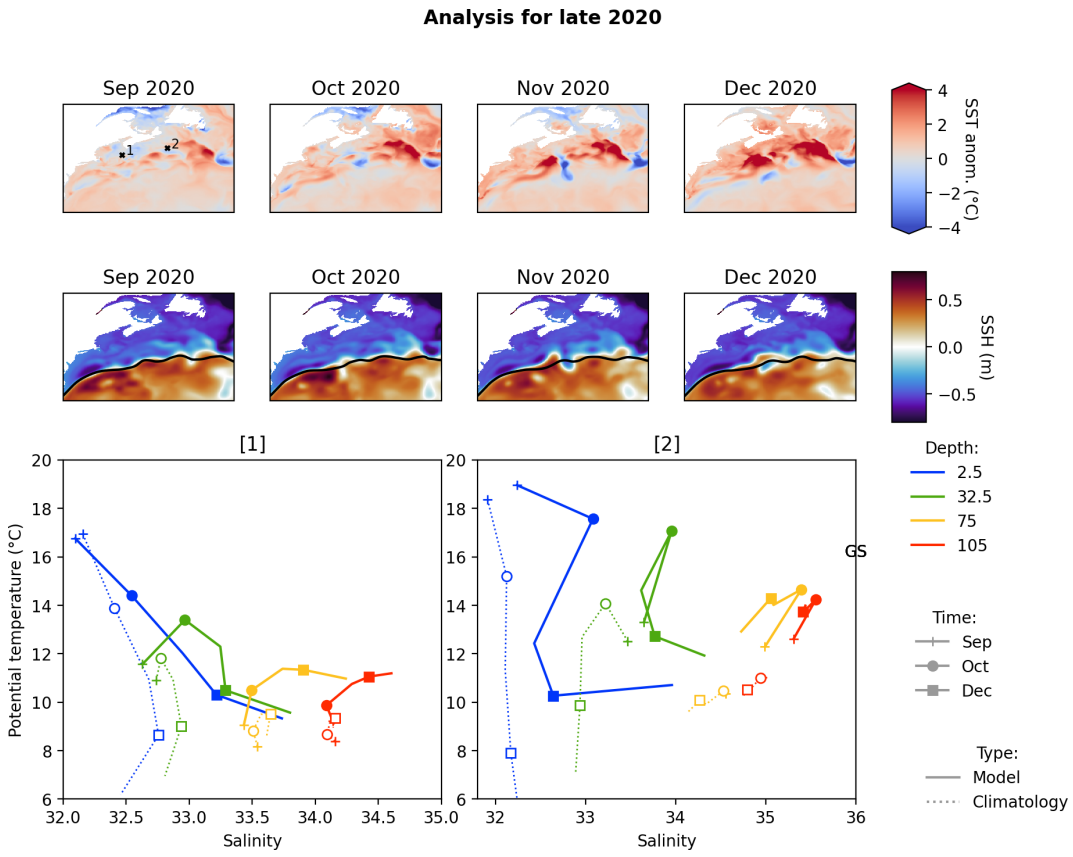

Figure 15. Temperature, salinity, and sea surface height during the period from September 2020 to January 2021 from the nudged downscaled historical simulation. The top row shows SST anomalies. The middle row shows sea surface height, with the climatological position of the 0 cm height contoured in black. The bottom row shows temperature–salinity diagrams for the two locations marked with dots in the top-left panel.

4.2 Effect of long-term warming trends

Because of the rapid increase in SSTs in the Northeast US region, a portion of the model surface and bottom temperature forecast skill could simply come from the consistently warming initial conditions and the greenhouse gas forcing in the atmospheric forecasts. The persistence forecast used as a baseline for comparison in the results would include the effects of warming initial conditions, but it would not include warming that varies by season or the (likely negligible) component of the warming trend that occurs during the course of each 1-year forecast. To analyze the effect that discounting the predictable trend would have on the skill of the downscaled SST forecasts for the Northeast US LME, in Fig. 17, we compare the anomaly correlation for the raw downscaled ensemble mean with the correlation after removing the linear trend from the forecasts and observations. Trends were calculated separately for the forecasts and observations, and separately for each month and each initialization month. The correlation of the NWA12 model predictions with the observations is substantially reduced after detrending the forecasts and observations, especially after the first 1–2 months of lead time, indicating that the linear trend is a substantial part of the forecast skill. However, the linear trend also contributes to the apparent skill of the persistence forecast, and the difference between the model and persistence skill generally remains the same whether or not detrending is applied. Patterns of correlation indicative of mechanisms behind prediction skill, such as the reemergence of skill in forecasts verifying in autumn (Sect. 4.1.1) also generally remain after detrending. We also note that the trends that were removed were fit to the full time series, which is considered an "unfair" method since this trend would not have been known in the past. Bushuk et al. (2019) developed a "fair" method of detrending that iterates over each retrospective forecast and removes the trend calculated only using data available before the start of each forecast. If this method was applied here, the skill of the detrended forecasts would increase. However, we only show the unfair method of detrending since the correlation metric is also an unfair calculation that uses means and standard deviations calculated from the full time series. Finally, we also emphasize that the origin of the forecast skill, whether from a linear trend or dynamic variability,

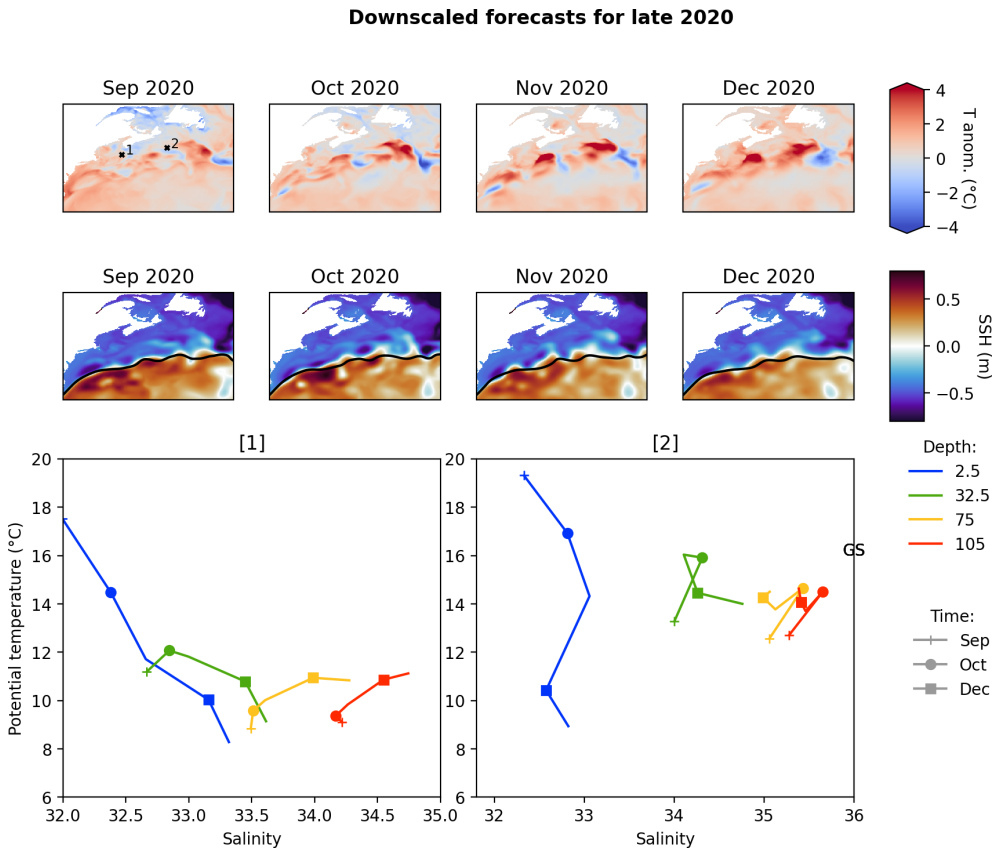

Figure 16. Temperature, salinity, and sea surface height during the period from September 2020 to January 2021 from the ensemble mean of downscaled forecasts initialized at the beginning of September 2020. The top row shows SST anomalies. The middle row shows sea surface height, with the climatological position of the 0 cm height contoured in black. The bottom row shows temperature–salinity diagrams for the two locations marked with dots in the top-left panel. The bottom row shows raw forecast values without adjustment for the lead-dependent bias.

is likely unimportant for most marine resource management purposes.

4.3 Skill versus ensemble size

In this study, we evaluated the skill of two 10-member forecast ensembles. Other high-resolution ocean prediction and projection studies have typically relied on fewer ensemble members or only one prediction, often due to the computational costs of running many ensemble members (Drenkard et al., 2021). Jacox et al. (2020) found quickly diminishing returns as the number of ensemble member forecasts of the California Current System SST was increased. On the other hand, features with a low signal-to-noise ratio, such as the NAO, may only be skillfully predicted with much larger ensembles (Dunstone et al., 2016; Scaife and Smith, 2018; Strommen and Palmer, 2019). To assess the impact of ensemble size on seasonal forecast skill in the NWA12 model, we created 1000 bootstrap resamples (sampling from the 10 ensemble members for each initialization and lead time with replacement) of the Northeast US. LME forecasts of the surface and bottom temperature anomaly and surface salinity anomaly for ensembles consisting of 1, 2, 4, and 10 members. For each ensemble size, we calculated the median correlation coefficient for the ensemble mean and the median raw ensemble continuous ranked probability score (CRPS; Bröcker, 2012), which evaluates the absolute error in the empirical probability distribution of the ensemble. Both metrics, shown in Fig. 18, indicate diminishing marginal improvements in skill (higher correlation and lower CRPS) for all three variables as the ensemble size is increased; the improvement from a single member to two members is substantial, while the improvement from 4 to 10 members is minor. This is consistent with the expected effect of ensemble size on the estimation of the mean and CRPS (Leutbecher, 2018), and the results suggest that, at least for predictions of these temperature and salinity anomalies, an ensemble with approximately four members would provide a reasonable compromise between computational costs and prediction skill. However, we expect that forecasts of rare events, such as marine heat waves, or nonlinear variables, such as ocean chlorophyll, would still benefit from larger ensem-

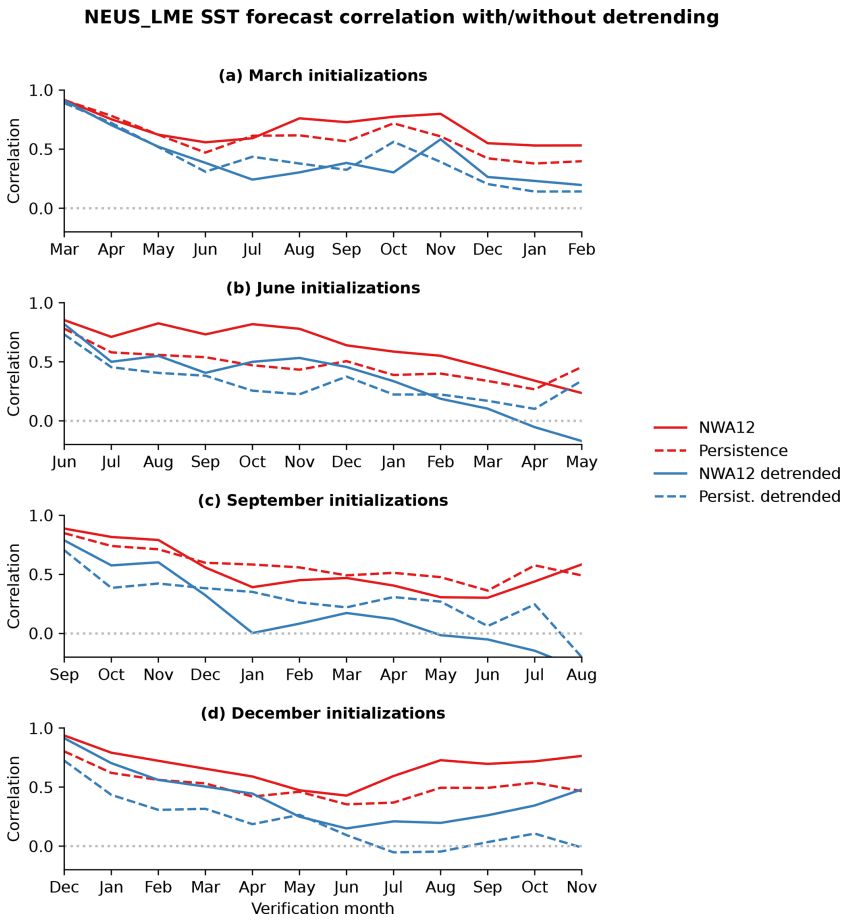

Figure 17. Correlation of downscaled Northeast US LME SST forecasts with (red) and without (blue) detrending of the forecasts and observations for the downscaled forecasts (solid lines) and the anomaly persistence reference forecast (dashed lines).

bles (Jacox et al., 2020). A larger ensemble size may also increase skill more if it was a multi-model ensemble, consisting of downscaled predictions forced by different models and/or using different models for downscaling rather than just an ensemble of multiple runs from the same model. For example, Brodie et al. (2023) found significantly higher skill in an ensemble of 73 members from six different coarse-resolution models compared to a much smaller ensemble of three coarse or downscaled models.

4.4 Role of model resolution

The analysis showed that the high-resolution regional model had significantly higher forecast skill than the global model in many cases and that this skill comes from several sources including better representation of reemergence, advection of water masses, and Gulf Stream variability and trends. Given the experimental design used in this study, however, where the high-resolution model uses initial conditions from a different higher-resolution source, it is difficult to determine how much of the increased forecast skill comes from the higher-resolution initial conditions and how much comes from evolving the initial conditions forward in time with higher resolution. In an analysis for the US west coast, Jacox et al. (2023) examined two sets of downscaled retrospective forecasts, one initialized from a high-resolution reanalysis (similar to the present study) and the other initialized from the coarse-resolution parent model. Initializing from the high-resolution reanalysis yielded generally negligible improvements in forecast skill for surface and bottom temperature aside from in the first month. The high-resolution reanalysis did significantly improve the skill of sea surface heights in their analysis. However, it is worth noting that improved (bias-corrected) atmospheric forcing was also included in their model runs initialized from the high-resolution product. Overall, additional experiments are needed to conclusively determine the role of the resolution of the initial conditions in downscaled seasonal forecast skill.

As we noted in Sect. 2.4, the skill assessment could have been biased in favor of the high-resolution model, which was initialized with the same GLORYS reanalysis used as for the observations in the assessment. On the other hand, the GLORYS reanalysis has been repeatedly found to closely match

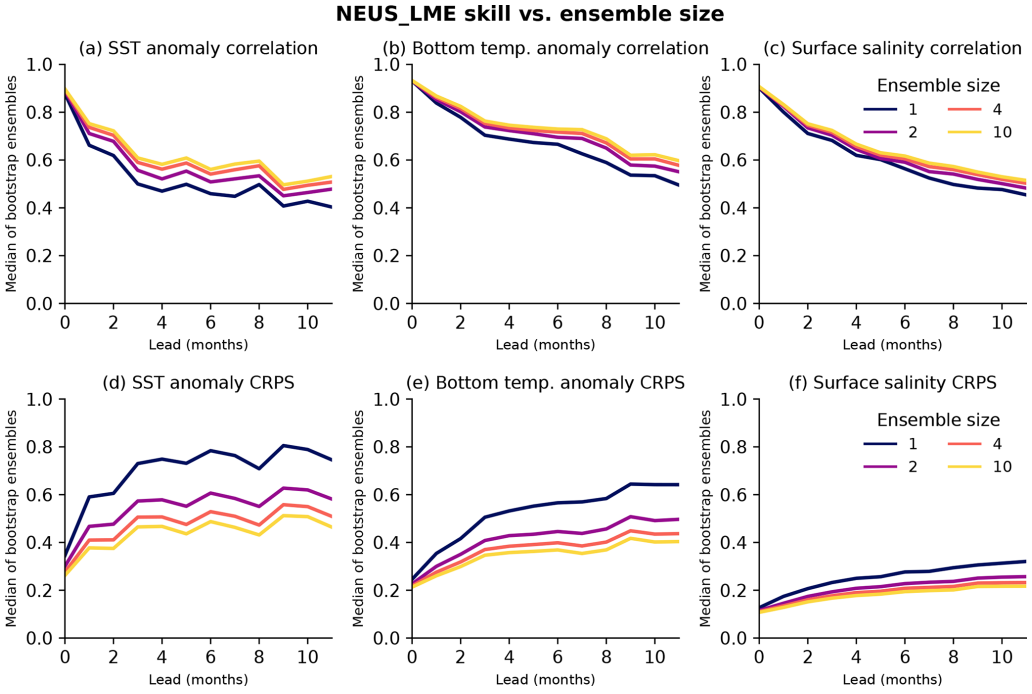

Figure 18. Correlation **(a–c)** and ensemble continuous ranked probability score (CRPS; **d–f**) for NWA12 as a function of the size of the model ensemble (colors). Plotted values are the medians of 1000 bootstrap samples of a given ensemble size from the full 10-member ensemble.

in situ observations (Amaya et al., 2023; Carolina Castillo-Trujillo et al., 2023), which would suggest that comparing against the GLORYS reanalysis should be similar to comparing against in situ observations. To determine whether any bias could be an issue, we repeated the assessment of forecast SST anomaly correlation using the OISST dataset (Reynolds et al., 2007) instead of the GLORYS reanalysis (Fig. S1 in the Supplement). Even though the SPEAR model derived its initial conditions by assimilating data from OISST, there is no meaningful difference between the forecast skill relative to GLORYS or OISST. In fact, in many cases, the downscaled model has slightly higher prediction skill if OISST is used as the source of observations. A lack of sensitivity to the dataset used for verification was also found by Jacox et al. (2023), who downscaled seasonal forecasts for the US west coast.

5 Conclusions

The high-resolution downscaled seasonal predictions yielded a significant improvement to forecast skill for temperature and salinity anomalies in the Northeast US Large Marine Ecosystem and maintained or improved skill in other regions along the North American east coast. Although this initial experiment focused on predicting physical ocean conditions, the dynamical drivers of the forecast skill suggest that skillful predictions of biogeochemical features and connections to living marine resources will also be possible. For example, similar to the reemergence of temperature anomalies that was skillfully predicted in the downscaled forecasts, Park et al. (2019) found that nitrate anomalies in the Northwest Atlantic also persist below the summer mixed layer and reemerge in the winter and lead to predictable chlorophyll anomalies in the following spring. Furthermore, the predictable advection of salinity by the cyclonic coastal circulation in the Gulf of Maine (Fig. 11) suggests the potential to predict the advection of harmful algal blooms by the same current (Li et al., 2014; Y. Zhang et al., 2020). The forecasts presented here are the first step towards building connections between predictable ocean temperature and salinity, biogeochemistry, and living marine resources along the North American east coast to develop critically needed large-scale marine ecosystem predictions (Link et al., 2023).

Although the downscaled predictions of temperature and salinity anomalies were skillful in many cases, some nontrivial biases in the mean temperature and salinity were present even in the downscaled predictions. These ocean biases could be reduced by correcting biases in the atmospheric forcing used to drive the regional model, and experimenting with methods to correct these biases is a goal of future research. Reducing biases in ocean mixing, whether by correcting the wind forcing or adjusting the model parameterizations, may be especially important for reducing bottom temperature biases. Correcting for biases in the forcing should reduce biases in the ocean predictions but may not substantially improve the prediction skill of anomalies (as found by Jacox et al., 2023, along the US west coast). However, reducing ocean

temperature and salinity biases will likely be important for obtaining accurate predictions with coupled biogeochemical model components.

Other improvements to the basic forecast model configuration developed in this study may also be necessary to expand the skill and capabilities of the model. For example, using forecasts of river discharge rather than a climatology of river discharge may be necessary for predicting nearshore salinity biogeochemistry. Similarly, applying predicted conditions for the ocean open boundaries rather than using a climatology, could be necessary for predicting quickly propagating sea level anomalies. Finally, full data assimilation rather than nudging towards a reanalysis could improve prediction skill through better initial conditions and also yield better probabilistic forecasts and predictions of extreme events.

Code availability. The source code for each component of the MOM6-NWA12 model has been archived by Ross et al. (2023a) at https://doi.org/10.5281/zenodo.7893349 (Ross et al., 2023b).

Data availability. All model output that was analyzed in this paper has been published at https://doi.org/10.5281/zenodo.10642294 (Ross, 2024). The datasets used for comparison with the model and the DOI where the data can be downloaded are as follows: GLORYS12 reanalysis (https://doi.org/10.48670/moi-00021, EU Copernicus Marine Service Product, 2023b) and global ocean gridded sea surface heights (https://doi.org/10.48670/moi-00148, EU Copernicus Marine Service Product, 2023a).

Supplement. The supplement related to this article is available online at: https://doi.org/10.5194/os-20-1631-2024-supplement.

Author contributions. TLD, FL, and AW developed the SPEAR global prediction model. MAA contributed to the discussion on the reemergence of SST anomalies. VK contributed to the methods for forecast evaluation. ACR prepared the initial draft of the manuscript with guidance from CAS. All coauthors participated in discussions during various stages of the model development and evaluation and read and approved the final version of the paper.

Competing interests. The contact author has declared that none of the authors has any competing interests.

Disclaimer. Publisher's note: Copernicus Publications remains neutral with regard to jurisdictional claims made in the text, published maps, institutional affiliations, or any other geographical representation in this paper. While Copernicus Publications makes every effort to include appropriate place names, the final responsibility lies with the authors.

Acknowledgements. We appreciate comments from three anonymous reviewers and internal reviews of a draft of this manuscript by Mitchell Bushuk and Michael Jacox. Computational resources for this work were provided by allocations on NOAA's Research and Development High-Performance Computing System provided by the Geophysical Fluid Dynamics Laboratory and NOAA's Climate, Ecosystems and Fisheries Initiative. This study has been conducted using EU Copernicus Marine Service Information: https://doi.org/10.48670/moi-00021 (EU Copernicus Marine Service Product, 2023b) and https://doi.org/10.48670/moi-00148 (EU Copernicus Marine Service Product, 2023a).

Financial support. This research has been supported by the NOAA Climate Program Office.

Review statement. This paper was edited by Anne Marie Treguier and reviewed by K. Andrew Peterson and two anonymous referees.

References

Alexander, M. and Scott, J.: The Influence of ENSO on Air-Sea Interaction in the Atlantic, Geophys. Res. Lett., 29, 46-1–46-4, https://doi.org/10.1029/2001GL014347, 2002.

Alexander, M. A., Deser, C., and Timlin, M. S.: The Reemergence of SST Anomalies in the North Pacific Ocean, J. Climate, 12, 2419–2433, https://doi.org/10.1175/1520-0442(1999)012<2419:TROSAI>2.0.CO;2, 1999.

Alexander, M. A., Shin, S.-i., Scott, J. D., Curchitser, E., and Stock, C.: The Response of the Northwest Atlantic Ocean to Climate Change, J. Climate, 33, 405–428, https://doi.org/10.1175/JCLI-D-19-0117.1, 2020.

Alfieri, L., Lorini, V., Hirpa, F. A., Harrigan, S., Zsoter, E., Prudhomme, C., and Salamon, P.: A Global Streamflow Reanalysis for 1980–2018, J. Hydrol. X, 6, 100049, https://doi.org/10.1016/j.hydroa.2019.100049, 2020.

Amaya, D. J., Alexander, M. A., Scott, J. D., and Jacox, M. G.: An Evaluation of High-Resolution Ocean Reanalyses in the California Current System, Prog. Oceanogr., 210, 102951, https://doi.org/10.1016/j.pocean.2022.102951, 2023.

Bretherton, C. S., Widmann, M., Dymnikov, V. P., Wallace, J. M., and Bladé, I.: The Effective Number of Spatial Degrees of Freedom of a Time-Varying Field, J. Climate, 12, 1990–2009, https://doi.org/10.1175/1520-0442(1999)012<1990:TENOSD>2.0.CO;2, 1999.

Brickman, D., Hebert, D., and Wang, Z.: Mechanism for the Recent Ocean Warming Events on the Scotian Shelf of Eastern Canada, Cont. Shelf Res., 156, 11–22, https://doi.org/10.1016/j.csr.2018.01.001, 2018.

Bröcker, J.: Evaluating Raw Ensembles with the Continuous Ranked Probability Score, Q. J. Roy. Meteor. Soc., 138, 1611–1617, https://doi.org/10.1002/qj.1891, 2012.

Brodie, S., Pozo Buil, M., Welch, H., Bograd, S. J., Hazen, E. L., Santora, J. A., Seary, R., Schroeder, I. D., and Jacox, M. G.: Ecological Forecasts for Marine Resource Management during Climate Extremes, Nat. Commun., 14, 7701, https://doi.org/10.1038/s41467-023-43188-0, 2023.

Bushuk, M., Msadek, R., Winton, M., Vecchi, G., Yang, X., Rosati, A., and Gudgel, R.: Regional Arctic Sea–Ice Prediction: Potential versus Operational Seasonal Forecast Skill, Clim. Dynam., 52, 2721–2743, https://doi.org/10.1007/s00382-018-4288-y, 2019.

Byju, P., Dommenget, D., and Alexander, M. A.: Widespread Reemergence of Sea Surface Temperature Anomalies in the Global Oceans, Including Tropical Regions Forced by Reemerging Winds, Geophys. Res. Lett., 45, 7683–7691, https://doi.org/10.1029/2018GL079137, 2018.

Caesar, L., Rahmstorf, S., Robinson, A., Feulner, G., and Saba, V.: Observed Fingerprint of a Weakening Atlantic Ocean Overturning Circulation, Nature, 556, 191–196, https://doi.org/10.1038/s41586-018-0006-5, 2018.

Carolina Castillo-Trujillo, A., Kwon, Y.-O., Fratantoni, P., Chen, K., Seo, H., Alexander, M. A., and Saba, V. S.: An Evaluation of Eight Global Ocean Reanalyses for the Northeast U.S. Continental Shelf, Prog. Oceanogr., 219, 103126, https://doi.org/10.1016/j.pocean.2023.103126, 2023.

Chassignet, E. P. and Garaffo, Z. D.: Viscosity parameterization and the Gulf Stream separation. From Stirring to Mixing in a Stratified Ocean, Proc. 'Aha Huliko'a Hawaiian Winter Workshop, 16–19 January 2001, Honolulu, HI, University of Hawai'i at Manoa, 37–41, https://n2t.net/ark:/13960/t6g22rz6s (last access: 6 December 2024), 2001.

Chassignet, E. P. and Xu, X.: Impact of Horizontal Resolution (1/12° to 1/50°) on Gulf Stream Separation, Penetration, and Variability, J. Phys. Oceanogr., 47, 1999–2021, https://doi.org/10.1175/JPO-D-17-0031.1, 2017.

Chen, Z., Kwon, Y.-O., Chen, K., Fratantoni, P., Gawarkiewicz, G., Joyce, T. M., Miller, T. J., Nye, J. A., Saba, V. S., and Stock, B. C.: Seasonal Prediction of Bottom Temperature on the Northeast U.S. Continental Shelf, J. Geophys. Res.-Oceans, 126, e2021JC017187, https://doi.org/10.1029/2021JC017187, 2021.

Chi, L., Wolfe, C. L. P., and Hameed, S.: Has the Gulf Stream Slowed or Shifted in the Altimetry Era?, Geophys. Res. Lett., 48, e2021GL093113, https://doi.org/10.1029/2021GL093113, 2021.

DelSole, T. and Tippett, M. K.: Forecast Comparison Based on Random Walks, Mon. Weather Rev., 144, 615–626, https://doi.org/10.1175/MWR-D-15-0218.1, 2016.

Delworth, T. L., Cooke, W. F., Adcroft, A., Bushuk, M., Chen, J.-H., Dunne, K. A., Ginoux, P., Gudgel, R., Hallberg, R. W., Harris, L., Harrison, M. J., Johnson, N., Kapnick, S. B., Lin, S.-J., Lu, F., Malyshev, S., Milly, P. C., Murakami, H., Naik, V., Pascale, S., Paynter, D., Rosati, A., Schwarzkopf, M., Shevliakova, E., Underwood, S., Wittenberg, A. T., Xiang, B., Yang, X., Zeng, F., Zhang, H., Zhang, L., and Zhao, M.: SPEAR: The Next Generation GFDL Modeling System for Seasonal to Multidecadal Prediction and Projection, J. Adv. Model. Earth Sy., 12, e2019MS001895, https://doi.org/10.1029/2019MS001895, 2020.

Drenkard, E. J., Stock, C., Ross, A. C., Dixon, K. W., Adcroft, A., Alexander, M., Balaji, V., Bograd, S. J., Butenschön, M., Cheng, W., Curchitser, E., Lorenzo, E. D., Dussin, R., Haynie, A. C., Harrison, M., Hermann, A., Hollowed, A., Holsman, K., Holt, J., Jacox, M. G., Jang, C. J., Kearney, K. A., Muhling, B. A., Buil, M. P., Saba, V., Sandø, A. B., Tommasi, D., and Wang, M.: Next-Generation Regional Ocean Projections for Living Marine Resource Management in a Changing Climate, ICES J. Mar. Sci., 78, 1969–1987, https://doi.org/10.1093/icesjms/fsab100, 2021.

Du, J., Zhang, W. G., and Li, Y.: Variability of Deep Water in Jordan Basin of the Gulf of Maine: Influence of Gulf Stream Warm Core Rings and the Nova Scotia Current, J. Geophys. Res.-Oceans, 126, e2020JC017136, https://doi.org/10.1029/2020JC017136, 2021.

Dukhovskoy, D. S., Chassignet, E. P., Bozec, A., and Morey, S. L.: Assessment of Predictability of the Loop Current in the Gulf of Mexico from Observing System Experiments and Observing System Simulation Experiments, Front. Mar. Sci., 10, 1153824, https://doi.org/10.3389/fmars.2023.1153824, 2023.

Dunstone, N., Smith, D., Scaife, A., Hermanson, L., Eade, R., Robinson, N., Andrews, M., and Knight, J.: Skilful Predictions of the Winter North Atlantic Oscillation One Year Ahead, Nat. Geosci., 9, 809–814, https://doi.org/10.1038/ngeo2824, 2016.

EU Copernicus Marine Service Product: Global Ocean Gridded L4 Sea Surface Heights And Derived Variables Reprocessed 1993 Ongoing, Mercator Ocean International [data set], https://doi.org/10.48670/moi-00148, 2023a.

EU Copernicus Marine Service Product: Global Ocean Physics Reanalysis, Mercator Ocean International [data set], https://doi.org/10.48670/moi-00021, 2023b.

Fortin, V., Abaza, M., Anctil, F., and Turcotte, R.: Why Should Ensemble Spread Match the RMSE of the Ensemble Mean?, J. Hydrometeorol., 15, 1708–1713, https://doi.org/10.1175/JHM-D-14-0008.1, 2014.

Friedland, K. D., Record, N. R., Pendleton, D. E., Balch, W. M., Stamieszkin, K., Moisan, J. R., and Brady, D. C.: Asymmetry in the Rate of Warming and the Phenology of Seasonal Blooms in the Northeast US Shelf Ecosystem, ICES J. Mar. Sci., 80, 775–786, https://doi.org/10.1093/icesjms/fsad007, 2023.

Gangopadhyay, A., Gawarkiewicz, G., Silva, E. N. S., Monim, M., and Clark, J.: An Observed Regime Shift in the Formation of Warm Core Rings from the Gulf Stream, Sci. Rep., 9, 12319, https://doi.org/10.1038/s41598-019-48661-9, 2019.

Gangopadhyay, A., Gawarkiewicz, G., Silva, E. N. S., Silver, A. M., Monim, M., and Clark, J.: A Census of the Warm-Core Rings of the Gulf Stream: 1980–2017, J. Geophys. Res.-Oceans, 125, e2019JC016033, https://doi.org/10.1029/2019JC016033, 2020.

Geiss, A., Marchand, R., and Thompson, L.: The Influence of Sea Surface Temperature Reemergence on Marine Stratiform Cloud, Geophys. Res. Lett., 47, e2020GL086957, https://doi.org/10.1029/2020GL086957, 2020.

Ghantous, M., Ayoub, N., De Mey-Frémaux, P., Vervatis, V., and Marsaleix, P.: Ensemble Downscaling of a Regional Ocean Model, Ocean Model., 145, 101511, https://doi.org/10.1016/j.ocemod.2019.101511, 2020.

Glenn, E., Comarazamy, D., González, J. E., and Smith, T.: Detection of Recent Regional Sea Surface Temperature Warming in the Caribbean and Surrounding Region, Geophys. Res. Lett., 42, 6785–6792, https://doi.org/10.1002/2015GL065002, 2015.

Gomez, F. A., Lee, S.-K., Hernandez, F. J., Chiaverano, L. M., Muller-Karger, F. E., Liu, Y., and Lamkin, J. T.: ENSO-induced Co-Variability of Salinity, Plankton Biomass and Coastal Currents in the Northern Gulf of Mexico, Sci. Rep., 9, 178, https://doi.org/10.1038/s41598-018-36655-y, 2019.

Gonçalves Neto, A., Langan, J. A., and Palter, J. B.: Changes in the Gulf Stream Preceded Rapid Warming of the Northwest

Atlantic Shelf, Communications Earth & Environment, 2, 74, https://doi.org/10.1038/s43247-021-00143-5, 2021.

Grodsky, S. A., Reul, N., Chapron, B., Carton, J. A., and Bryan, F. O.: Interannual Surface Salinity on Northwest Atlantic Shelf, J. Geophys. Res.-Oceans, 122, 3638–3659, https://doi.org/10.1002/2016JC012580, 2017.

Haley, P. J., Mirabito, C., Doshi, M., and Lermusiaux, P. F. J.: Ensemble Forecasting for the Gulf of Mexico Loop Current Region, in: OCEANS 2023 – MTS/IEEE U.S. Gulf Coast, 1–10, IEEE, Biloxi, MS, USA, ISBN 9798218142186, https://doi.org/10.23919/OCEANS52994.2023.10337035, 2023.

Han, G., Hannah, C. G., Loder, J. W., and Smith, P. C.: Seasonal Variation of the Three-Dimensional Mean Circulation over the Scotian Shelf, J. Geophys. Res.-Oceans, 102, 1011–1025, https://doi.org/10.1029/96JC03285, 1997.

Han, G., Ma, Z., Long, Z., Perrie, W., and Chassé, J.: Climate Change on Newfoundland and Labrador Shelves: Results From a Regional Downscaled Ocean and Sea-Ice Model Under an A1B Forcing Scenario 2011–2069, Atmos.-Ocean, 57, 3–17, https://doi.org/10.1080/07055900.2017.1417110, 2019.

Harrigan, S., Zsoter, E., Alfieri, L., Prudhomme, C., Salamon, P., Wetterhall, F., Barnard, C., Cloke, H., and Pappenberger, F.: GloFAS-ERA5 operational global river discharge reanalysis 1979–present, Earth Syst. Sci. Data, 12, 2043–2060, https://doi.org/10.5194/essd-12-2043-2020, 2020.

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., De Chiara, G., Dahlgren, P., Dee, D., Diamantakis, M., Dragani, R., Flemming, J., Forbes, R., Fuentes, M., Geer, A., Haimberger, L., Healy, S., Hogan, R. J., Hólm, E., Janisková, M., Keeley, S., Laloyaux, P., Lopez, P., Lupu, C., Radnoti, G., de Rosnay, P., Rozum, I., Vamborg, F., Villaume, S., and Thépaut, J.-N.: The ERA5 Global Reanalysis, Q. J. Roy. Meteor. Soc., 146, 1999–2049, https://doi.org/10.1002/qj.3803, 2020.

Hervieux, G., Alexander, M. A., Stock, C. A., Jacox, M. G., Pegion, K., Becker, E., Castruccio, F., and Tommasi, D.: More Reliable Coastal SST Forecasts from the North American Multimodel Ensemble, Clim. Dynam., 53, 7153–7168, https://doi.org/10.1007/s00382-017-3652-7, 2017.

Jackson, L. C., Peterson, K. A., Roberts, C. D., and Wood, R. A.: Recent Slowing of Atlantic Overturning Circulation as a Recovery from Earlier Strengthening, Nat. Geosci., 9, 518–522, https://doi.org/10.1038/ngeo2715, 2016.

Jacox, M. G., Alexander, M. A., Siedlecki, S., Chen, K., Kwon, Y.-O., Brodie, S., Ortiz, I., Tommasi, D., Widlansky, M. J., Barrie, D., Capotondi, A., Cheng, W., Di Lorenzo, E., Edwards, C., Fiechter, J., Fratantoni, P., Hazen, E. L., Hermann, A. J., Kumar, A., Miller, A. J., Pirhalla, D., Pozo Buil, M., Ray, S., Sheridan, S. C., Subramanian, A., Thompson, P., Thorne, L., Annamalai, H., Aydin, K., Bograd, S. J., Griffis, R. B., Kearney, K., Kim, H., Mariotti, A., Merrifield, M., and Rykaczewski, R.: Seasonal-to-Interannual Prediction of North American Coastal Marine Ecosystems: Forecast Methods, Mechanisms of Predictability, and Priority Developments, Prog. Oceanogr., 183, 102307, https://doi.org/10.1016/j.pocean.2020.102307, 2020.

Jacox, M. G., Alexander, M. A., Amaya, D., Becker, E., Bograd, S. J., Brodie, S., Hazen, E. L., Pozo Buil, M., and Tommasi, D.: Global Seasonal Forecasts of Marine Heatwaves, Nature, 604, 486–490, https://doi.org/10.1038/s41586-022-04573-9, 2022.

Jacox, M. G., Buil, M. P., Brodie, S., Alexander, M. A., Amaya, D. J., Bograd, S. J., Edwards, C. A., Fiechter, J., Hazen, E. L., Hervieux, G., and Tommasi, D.: Downscaled Seasonal Forecasts for the California Current System: Skill Assessment and Prospects for Living Marine Resource Applications, PLOS Climate, 2, e0000245, https://doi.org/10.1371/journal.pclm.0000245, 2023.

Lellouche, J., Greiner, E., Bourdallé-Badie, R., Garric, G., Melet, A., Drévillon, M., Bricaud, C., Hamon, M., Le Galloudec, O., Regnier, C., Candela, T., Testut, C., Gasparin, F., Ruggiero, G., Benkiran, M., Drillet, Y., and Le Traon, P.: The Copernicus Global 1/12° Oceanic and Sea Ice GLORYS12 Reanalysis, Front. Earth Sci., 9, 698876, https://doi.org/10.3389/feart.2021.698876, 2021.

Leutbecher, M.: Ensemble Size: How Suboptimal Is Less than Infinity?, Q. J. Roy. Meteor. Soc., https://doi.org/10.1002/qj.3387, 2018.

Li, Y., He, R., and Manning, J. P.: Coastal Connectivity in the Gulf of Maine in Spring and Summer of 2004–2009, Deep-Sea Res. Pt. II, 103, 199–209, https://doi.org/10.1016/j.dsr2.2013.01.037, 2014.

Link, J. S., Thur, S., Matlock, G., and Grasso, M.: Why We Need Weather Forecast Analogues for Marine Ecosystems, ICES J. Mar. Sci., 80, 2087–2098, https://doi.org/10.1093/icesjms/fsad143, 2023.

Lu, F., Harrison, M. J., Rosati, A., Delworth, T. L., Yang, X., Cooke, W. F., Jia, L., McHugh, C., Johnson, N. C., Bushuk, M., Zhang, Y., and Adcroft, A.: GFDL's SPEAR Seasonal Prediction System: Initialization and Ocean Tendency Adjustment (OTA) for Coupled Model Predictions, J. Adv. Model. Earth Sy., 12, e2020MS002149, https://doi.org/10.1029/2020MS002149, 2020.

McAdam, R., Masina, S., Balmaseda, M., Gualdi, S., Senan, R., and Mayer, M.: Seasonal Forecast Skill of Upper-Ocean Heat Content in Coupled High-Resolution Systems, Clim. Dynam., 58, 3335–3350, https://doi.org/10.1007/s00382-021-06101-3, 2022.

New, A. L., Smeed, D. A., Czaja, A., Blaker, A. T., Mecking, J. V., Mathews, J. P., and Sanchez-Franks, A.: Labrador Slope Water Connects the Subarctic with the Gulf Stream, Environ. Res. Lett., 16, 084019, https://doi.org/10.1088/1748-9326/ac1293, 2021.

Oey, L.-Y., Ezer, T., and Lee, H.-C.: Loop Current, Rings and Related Circulation in the Gulf of Mexico: A Review of Numerical Models and Future Challenges, in: Geophysical Monograph Series, edited by: Sturges, W. and Lugo-Fernandez, A., American Geophysical Union, Washington, D.C., 31–56, ISBN 978-1-118-66616-6, ISBN 978-0-87590-426-9, https://doi.org/10.1029/161GM04, 2013.

Park, J.-Y., Stock, C. A., Dunne, J. P., Yang, X., and Rosati, A.: Seasonal to Multiannual Marine Ecosystem Prediction with a Global Earth System Model, Science, 365, 284–288, 2019.

Payne, M. R., Hobday, A. J., MacKenzie, B. R., Tommasi, D., Dempsey, D. P., Fässler, S. M. M., Haynie, A. C., Ji, R., Liu, G., Lynch, P. D., Matei, D., Miesner, A. K., Mills, K. E., Strand, K. O., and Villarino, E.: Lessons from the First Generation of Marine Ecological Forecast Products, Front. Mar. Sci., 4, 289, https://doi.org/10.3389/fmars.2017.00289, 2017.

Pegion, K., Kirtman, B. P., Becker, E., Collins, D. C., LaJoie, E., Burgman, R., Bell, R., DelSole, T., Min, D., Zhu, Y., Li, W., Sinsky, E., Guan, H., Gottschalck, J., Metzger, E. J., Barton, N. P., Achuthavarier, D., Marshak, J., Koster, R. D., Lin, H., Gagnon, N., Bell, M., Tippett, M. K., Robertson, A. W., Sun, S., Benjamin, S. G., Green, B. W., Bleck, R., and Kim, H.: The Subseasonal Experiment (SubX): A Multi-Model Subseasonal Prediction Experiment, B. Am. Meteorol. Soc., 100, 2043–2060, https://doi.org/10.1175/BAMS-D-18-0270.1, 2019.

Pershing, A. J., Alexander, M. A., Hernandez, C. M., Kerr, L. A., Le Bris, A., Mills, K. E., Nye, J. A., Record, N. R., Scannell, H. A., Scott, J. D., Sherwood, G. D., and Thomas, A. C.: Slow Adaptation in the Face of Rapid Warming Leads to Collapse of the Gulf of Maine Cod Fishery, Science, 350, 809–812, https://doi.org/10.1126/science.aac9819, 2015.

Reynolds, R. W., Smith, T. M., Liu, C., Chelton, D. B., Casey, K. S., and Schlax, M. G.: Daily High-Resolution-Blended Analyses for Sea Surface Temperature, J. Climate, 20, 5473–5496, https://doi.org/10.1175/2007JCLI1824.1, 2007.

Risbey, J. S., Squire, D. T., Black, A. S., DelSole, T., Lepore, C., Matear, R. J., Monselesan, D. P., Moore, T. S., Richardson, D., Schepen, A., Tippett, M. K., and Tozer, C. R.: Standard Assessments of Climate Forecast Skill Can Be Misleading, Nat. Commun., 12, 4346, https://doi.org/10.1038/s41467-021-23771-z, 2021.

Ross, A.: Model output for "Dynamically downscaled seasonal ocean forecasts for North American East Coast ecosystems", in: Ocean Science, Zenodo [data set], https://doi.org/10.5281/zenodo.10642294, 2024.

Ross, A. C. and Stock, C. A.: Probabilistic Extreme SST and Marine Heatwave Forecasts in Chesapeake Bay: A Forecast Model, Skill Assessment, and Potential Value, Front. Mar. Sci., 9, 896961, https://doi.org/10.3389/fmars.2022.896961, 2022.

Ross, A. C., Stock, C. A., Adcroft, A., Curchitser, E., Hallberg, R., Harrison, M. J., Hedstrom, K., Zadeh, N., Alexander, M., Chen, W., Drenkard, E. J., du Pontavice, H., Dussin, R., Gomez, F., John, J. G., Kang, D., Lavoie, D., Resplandy, L., Roobaert, A., Saba, V., Shin, S.-I., Siedlecki, S., and Simkins, J.: A high-resolution physical–biogeochemical model for marine resource applications in the northwest Atlantic (MOM6-COBALT-NWA12 v1.0), Geosci. Model Dev., 16, 6943–6985, https://doi.org/10.5194/gmd-16-6943-2023, 2023a.

Ross, A. C., Stock, C. A., Adcroft, A., Curchitser, E., Hallberg, R., Harrison, M. J., Hedstrom, K., Zadeh, N., Alexander, M., Chen, W., Drenkard, E. J., du Pontavice, H., Dussin, R., Gomez, F., John, J. G., Kang, D., Lavoie, D., Resplandy, L., Roobaert, A., Saba, V., Shin, S.-I., Siedlecki, S., and Simkins, J.: Model source code for "A high-resolution physical-biogeochemical model for marine resource applications in the Northwest Atlantic (MOM6-COBALT-NWA12)", Zenodo [code], https://doi.org/10.5281/zenodo.7893349, 2023b.

Rutherford, K., Fennel, K., Garcia Suarez, L., and John, J. G.: Uncertainty in the evolution of northwestern North Atlantic circulation leads to diverging biogeochemical projections, Biogeosciences, 21, 301–314, https://doi.org/10.5194/bg-21-301-2024, 2024.

Saba, V. S., Griffies, S. M., Anderson, W. G., Winton, M., Alexander, M. A., Delworth, T. L., Hare, J. A., Harrison, M. J., Rosati, A., Vecchi, G. A., and Zhang, R.: Enhanced Warming of the Northwest Atlantic Ocean under Climate Change, J. Geophys. Res.-Oceans, 121, 118–132, https://doi.org/10.1002/2015JC011346, 2016.

Sanchez-Franks, A., Hameed, S., and Wilson, R. E.: The Icelandic Low as a Predictor of the Gulf Stream North Wall Position, J. Phys. Oceanogr., 46, 817–826, https://doi.org/10.1175/JPO-D-14-0244.1, 2016.

Sandery, P. A. and Sakov, P.: Ocean Forecasting of Mesoscale Features Can Deteriorate by Increasing Model Resolution towards the Submesoscale, Nat. Commun., 8, 1–8, https://doi.org/10.1038/s41467-017-01595-0, 2017.

Scaife, A. A. and Smith, D.: A Signal-to-Noise Paradox in Climate Science, npj Climate and Atmospheric Science, 1, 28, https://doi.org/10.1038/s41612-018-0038-4, 2018.

Seidov, D., Mishonov, A., and Parsons, R.: Recent Warming and Decadal Variability of Gulf of Maine and Slope Water, Limnol. Oceanogr., 66, 3472–3488, https://doi.org/10.1002/lno.11892, 2021.

Shin, S.-I. and Newman, M.: Seasonal Predictability of Global and North American Coastal Sea Surface Temperature and Height Anomalies, Geophys. Res. Lett., 48, e2020GL091886, https://doi.org/10.1029/2020GL091886, 2021.

Siedlecki, S. A., Kaplan, I. C., Hermann, A. J., Nguyen, T. T., Bond, N. A., Newton, J. A., Williams, G. D., Peterson, W. T., Alin, S. R., and Feely, R. A.: Experiments with Seasonal Forecasts of Ocean Conditions for the Northern Region of the California Current Upwelling System, Sci. Rep., 6, 27203, https://doi.org/10.1038/srep27203, 2016.

Siegert, S., Bellprat, O., Ménégoz, M., Stephenson, D. B., and Doblas-Reyes, F. J.: Detecting Improvements in Forecast Correlation Skill: Statistical Testing and Power Analysis, Mon. Weather Rev., 145, 437–450, https://doi.org/10.1175/MWR-D-16-0037.1, 2017.

Silver, A., Gangopadhyay, A., Gawarkiewicz, G., Fratantoni, P., and Clark, J.: Increased Gulf Stream Warm Core Ring Formations Contributes to an Observed Increase in Salinity Maximum Intrusions on the Northeast Shelf, Sci. Rep., 13, 7538, https://doi.org/10.1038/s41598-023-34494-0, 2023.

Smith, D. M., Scaife, A. A., Eade, R., Athanasiadis, P., Bellucci, A., Bethke, I., Bilbao, R., Borchert, L. F., Caron, L.-P., Counillon, F., Danabasoglu, G., Delworth, T., Doblas-Reyes, F. J., Dunstone, N. J., Estella-Perez, V., Flavoni, S., Hermanson, L., Keenlyside, N., Kharin, V., Kimoto, M., Merryfield, W. J., Mignot, J., Mochizuki, T., Modali, K., Monerie, P.-A., Müller, W. A., Nicolí, D., Ortega, P., Pankatz, K., Pohlmann, H., Robson, J., Ruggieri, P., Sospedra-Alfonso, R., Swingedouw, D., Wang, Y., Wild, S., Yeager, S., Yang, X., and Zhang, L.: North Atlantic Climate Far More Predictable than Models Imply, Nature, 583, 796–800, https://doi.org/10.1038/s41586-020-2525-0, 2020.

Smith, P. C., Houghton, R. W., Fairbanks, R. G., and Mountain, D. G.: Interannual variability of boundary fluxes and water mass properties in the Gulf of Maine and on Georges Bank: 1993–1997, Deep-Sea Res. Pt. II, 48, 37–70, https://doi.org/10.1016/S0967-0645(00)00081-3, 2001.

Spalding, M. D., Fox, H. E., Allen, G. R., Davidson, N., Ferdaña, Z. A., Finlayson, M., Halpern, B. S., Jorge, M. A., Lombana, A., Lourie, S. A., Martin, K. D., McManus, E., Molnar, J., Recchia, C. A., and Robertson, J.: Marine Ecoregions of the World: A

Bioregionalization of Coastal and Shelf Areas, BioScience, 57, 573–583, https://doi.org/10.1641/B570707, 2007.

Steiger, J. H.: Tests for Comparing Elements of a Correlation Matrix, Psychol. Bull., 87, 245–251, https://doi.org/10.1037/0033-2909.87.2.245, 1980.

Stock, C. A., Pegion, K., Vecchi, G. A., Alexander, M. A., Tommasi, D., Bond, N. A., Fratantoni, P. S., Gudgel, R. G., Kristiansen, T., O'Brien, T. D., Xue, Y., and Yang, X.: Seasonal Sea Surface Temperature Anomaly Prediction for Coastal Ecosystems, Prog. Oceanogr., 137, 219–236, https://doi.org/10.1016/j.pocean.2015.06.007, 2015.

Strommen, K. and Palmer, T. N.: Signal and Noise in Regime Systems: A Hypothesis on the Predictability of the North Atlantic Oscillation, Q. J. Roy. Meteor. Soc., 145, 147–163, https://doi.org/10.1002/qj.3414, 2019.

Sukhonos, P. A. and Alexander, M. A.: The Reemergence of the Winter Sea Surface Temperature Tripole in the North Atlantic from Ocean Reanalysis Data, Clim. Dynam., 61, 449–460, https://doi.org/10.1007/s00382-022-06581-x, 2023.

Thoppil, P. G., Frolov, S., Rowley, C. D., Reynolds, C. A., Jacobs, G. A., Joseph Metzger, E., Hogan, P. J., Barton, N., Wallcraft, A. J., Smedstad, O. M., and Shriver, J. F.: Ensemble Forecasting Greatly Expands the Prediction Horizon for Ocean Mesoscale Variability, Communications Earth & Environment, 2, 1–9, https://doi.org/10.1038/s43247-021-00151-5, 2021.

Tommasi, D., Stock, C. A., Hobday, A. J., Methot, R., Kaplan, I. C., Eveson, J. P., Holsman, K., Miller, T. J., Gaichas, S., Gehlen, M., Pershing, A., Vecchi, G. A., Msadek, R., Delworth, T., Eakin, C. M., Haltuch, M. A., Séférian, R., Spillman, C. M., Hartog, J. R., Siedlecki, S., Samhouri, J. F., Muhling, B., Asch, R. G., Pinsky, M. L., Saba, V. S., Kapnick, S. B., Gaitan, C. F., Rykaczewski, R. R., Alexander, M. A., Xue, Y., Pegion, K. V., Lynch, P., Payne, M. R., Kristiansen, T., Lehodey, P., and Werner, F. E.: Managing Living Marine Resources in a Dynamic Environment: The Role of Seasonal to Decadal Climate Forecasts, Prog. Oceanogr., 152, 15–49, https://doi.org/10.1016/j.pocean.2016.12.011, 2017.

Wang, Z., Yang, J., Johnson, C., and DeTracey, B.: Changes in Deep Ocean Contribute to a "See-Sawing" Gulf Stream Path, Geophys. Res. Lett., 49, e2022GL100937, https://doi.org/10.1029/2022GL100937, 2022.

Wang, Z., Boyer, T., Reagan, J., and Hogan, P.: Upper Oceanic Warming in the Gulf of Mexico between 1950 and 2020, J. Climate, 1, 1–32, https://doi.org/10.1175/JCLI-D-22-0409.1, 2023.

Wilkin, J., Levin, J., Lopez, A., Hunter, E., Zavala-Garay, J., and Arango, H.: A Coastal Ocean Forecast System for U.S. Mid-Atlantic Bight and Gulf of Maine, in: New Frontiers in Operational Oceanography, edited by: Chassignet, E. P., Pascual, A., Tintoré, J., and Verron, J., GODAE OceanView, ISBN 978-1-72054-997-0, https://doi.org/10.17125/gov2018.ch21, 2018.

Xue, H., Chai, F., and Pettigrew, N. R.: A Model Study of the Seasonal Circulation in the Gulf of Maine, J. Phys. Oceanogr., 30, 1111–1135, https://doi.org/10.1175/1520-0485(2000)030<1111:AMSOTS>2.0.CO;2, 2000.

Zhang, W.-Z., Chai, F., Xue, H., and Oey, L.-Y.: Remote Sensing Linear Trends of the Gulf Stream from 1993 to 2016, Ocean Dynam., 70, 701–712, https://doi.org/10.1007/s10236-020-01356-6, 2020.

Zhang, Y., Chen, C., Xue, P., Beardsley, R. C., and Franks, P. J. S.: A View of Physical Mechanisms for Transporting Harmful Algal Blooms to Massachusetts Bay, Mar. Pollut. Bull., 154, 111048, https://doi.org/10.1016/j.marpolbul.2020.111048, 2020.