Prospects for Seasonal Prediction of Summertime Trans-Arctic Sea Ice Path

MICHAEL WINTON,[a] MITCHELL BUSHUK,[a,b] YONGFEI ZHANG,[c] BILL HURLIN,[a] LIWEI JIA,[a,b] NATHANIEL C. JOHNSON,[a] AND FEIYU LU[a,b]

[a] *NOAA/Geophysical Fluid Dynamics Laboratory, Princeton, New Jersey*
[b] *University Corporation for Atmospheric Research, Boulder, Colorado*
[c] *Atmospheric and Oceanic Sciences Program, Princeton University, Princeton, New Jersey*

(Manuscript received 24 August 2021, in final form 28 December 2021)

ABSTRACT: The continuing decline of the summertime sea ice cover has reduced the sea ice path that must be traversed to Arctic destinations and through the Arctic between the Atlantic and Pacific Oceans, stimulating interest in trans–Arctic Ocean routes. Seasonal prediction of the sea ice cover along these routes could support the increasing summertime ship traffic taking advantage of recent low ice conditions. We introduce the minimum Arctic sea ice path (MIP) between Atlantic and Pacific Oceans as a shipping-relevant metric that is amenable to multidecadal hindcast evaluation. We show, using 1992–2017 retrospective predictions, that bias correction is necessary for the GFDL Seamless System for Prediction and Earth System Research (SPEAR) forecast system to improve upon damped persistence seasonal forecasts of summertime daily MIP between the Atlantic and Pacific Oceans both east and west of Greenland, corresponding roughly to the Northeast and Northwest Passages. Without bias correction, only the Northwest Passage MIP forecasts have lower error than a damped persistence forecast. Using the forecast ensemble spread to estimate a lower bound on forecast error, we find large opportunities for forecast error reduction, especially at lead times of less than 2 months. Most of the potential improvement remains after linear removal of climatological and trend biases, suggesting that significant error reduction might come from improved initialization and simulation of subannual variability. Using a different passive microwave sea ice dataset for calculating error than was used for data assimilation increases the raw forecast errors but not the trend anomaly forecast errors.

KEYWORDS: Arctic; Seasonal forecasting; Climate models; Data assimilation; Ensembles

1. Introduction

The Northern Hemisphere sea ice cover has declined in recent decades with the largest negative trends occurring in late summer when the ice edge is in the Arctic Ocean. This decline is well documented with passive microwave satellite observations and is simulated by climate models as a response to anthropogenic forcing, indicating that the downward trend will continue into the future (e.g., Notz and SIMIP Community 2020). The summer sea ice poses an impediment to trans-Arctic passage between the Atlantic and Pacific Oceans. Accompanying the reduction in ice extent, ice-free corridors have appeared in September in recent years according to the passive microwave observations. Ice-free trans-Arctic passages are of interest for marine transport because sea ice is a major hazard and expense for shipping, requiring stronger hulls and slower speeds. Significant Arctic Ocean ship traffic is responding to the reduced ice conditions (Eguíluz et al. 2016; Melia et al. 2017b). If the current downward trend in ice cover continues as expected, the sea ice that must be traversed on ocean passages between the Atlantic and Pacific Oceans will continue to decline and ice-free corridors will be available in late summer over increasing durations (Smith and Stephenson 2013; Mudryk et al. 2021).

Seasonal forecast systems are now being applied to the problem of supporting summertime Arctic shipping. For example, Melia et al. (2017a) demonstrated the possibility of skillfully predicting open-water trans-Arctic passageways. Koyama et al. (2021) review progress on Arctic ship routing and present a statistical approach to the problem. Ship routing to and through the Arctic Ocean involves important considerations in addition to the sea ice state including waves and severe weather (Inoue 2021). Beyond the presence or absence of ice, sea ice thickness is important for constraining ship speeds and the ship types that can operate over a route (Nakanowatari et al. 2018). Unfortunately, reliable summertime sea ice thickness observations are not currently available (Dawson et al. 2022). To maximize the potential for validation we set aside these considerations and focus on prediction of a shipping-relevant metric constructed from sea ice extent, a quantity with a multidecadal record of high-quality summertime passive microwave observations.

We will present model hindcasts of the satellite-observed minimum trans–Arctic Ocean sea ice path (MIP) from July through October. The MIP quantifies the ice distance—distance through sea ice concentrations greater than 15%—that must be traversed when passing between the Atlantic and Pacific Oceans through the Arctic Ocean. The error in an MIP forecast contributes to the uncertainty in Arctic Ocean crossing time because ship speeds are much slower in sea ice than in open water. The MIP can be calculated directly from daily sea ice concentration observations for forecast verification. We use a quasi-operational forecast system to make seasonal predictions of the daily MIPs, requiring passage west and east of Greenland (NW MIP and NE MIP), corresponding roughly to the Northwest and Northeast Passages, from initializations on the first of May through August. The Seamless System for Prediction and Earth System Research (SPEAR) seasonal forecast system, used in this study, has

Corresponding author: Michael Winton, michael.winton@noaa.gov

DOI: 10.1175/JCLI-D-21-0634.1

been shown capable of making skillful predictions of local daily summertime Arctic sea ice concentration and derived variables characterizing the open water period (Zhang et al. 2022), supporting its suitability for summertime MIP prediction.

Section 2 presents the SPEAR forecast system, the MIP calculation, and the simple reference forecasts used for comparison. Section 3 presents forecast skill evaluations using 1992–2017 retrospective predictions and an analysis of the forecast errors. We determine the placement of the SPEAR skill between the predictability limit and the skill of a simple reference forecast based on recent observations. Section 4 summarizes and discusses the conclusions.

2. Forecast system, MIP calculation, and comparison forecasts

A low atmosphere resolution version of the new SPEAR prediction system (SPEAR_LO) from the Geophysical Fluid Dynamics Laboratory (GFDL) is used for this study (Delworth et al. 2020). Bushuk et al. (2022) found that this lower-resolution atmosphere version of the SPEAR forecast system had Arctic sea ice prediction skill comparable to the standard SPEAR system (SPEAR_MED) used for supplying forecasts to the North American Multi-Model Ensemble (NMME; Kirtman et al. 2014) and Sea Ice Prediction Network (SIPN; Meier et al. 2021). The system uses recently developed component models such as the MOM6 ocean and SIS2 sea ice models, also used by GFDL CM4 and ESM4 climate and Earth system models (Adcroft et al. 2019; Held et al. 2019; Dunne et al. 2020) and a newly developed ocean data assimilation system (Lu et al. 2021). SPEAR_LO employs nominally 1° resolution atmosphere, ocean, and sea ice grids. A summary of the initialization procedures for the ocean, sea ice, land, and atmosphere is given in Bushuk et al. (2022). In addition to the standard version of SPEAR_LO we use an enhanced version, SPEAR_LO/iceDA, initialized with sea ice from an ice-ocean model ensemble incorporating ensemble Kalman filter assimilation of sea ice concentration observations into the sea ice component (Zhang et al. 2021). Zhang et al. (2022) show that sea ice concentration assimilation reduces short-lead-time summer ice concentration forecast errors in the coastal regions of interest for this study. Although sea ice thickness is not assimilated, it is constrained to some degree by the atmospheric forcing from the JRA-55 reanalysis used to force the ice-ocean model ensemble used for sea ice initialization (Tsujino et al. 2018). The atmosphere and land (and sea ice in standard SPEAR-LO) are initialized from a SPEAR ensemble run having SST, atmospheric temperature, humidity, and wind nudged to historical observations.

The forecasts are made with 15 ensemble members and the ensemble mean is used for the ensemble forecast. Using the ensemble mean results in some positive bias to the MIP forecast at small values as ensemble members increasingly encounter the nonnegativity constraint. The sea ice concentration data assimilation and primary forecast verification both use "NASA Team" algorithm sea ice concentrations observations (Cavalieri et al. 1996). To assess the impact of using the same dataset for both purposes we shall also present an alternative assessment of the model with the "Bootstrap"

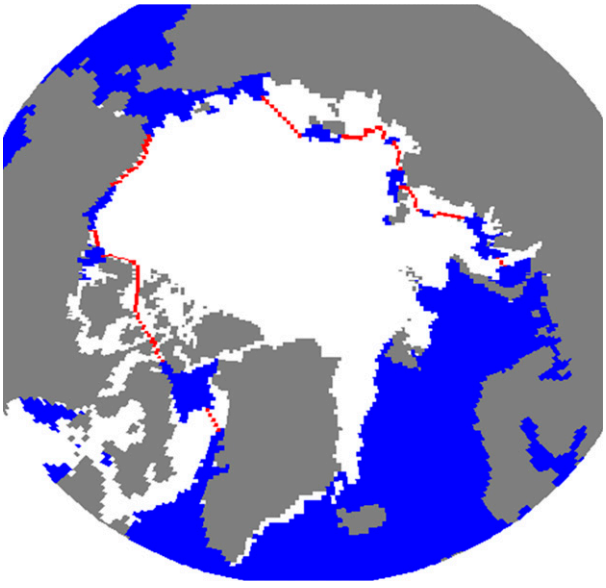

FIG. 1. An example of NE and NW MIPs (red) calculated using the 15 Jul 1992 sea ice concentration field. The region having sea ice concentration greater than 15% is shown in white. The MIPs are only defined in this region since open water contributes nothing to the path length.

sea ice concentration observations (Comiso 2017). The retrospective predictions are performed over the period 1992–2017 with initializations on the first of each month. We restrict our initializations to May–August because of the substantially higher skill of these forecasts relative to initializations in earlier months in many of the Arctic Ocean marginal seas (Bushuk et al. 2022).

We calculate the MIP between the Atlantic and Pacific, east and west of Greenland using Dijkstra's shortest path algorithm (Dijkstra 1959). Figure 1 shows an example of the NW and NE MIPs calculated for 15 July 1992. Dijkstra's algorithm finds the shortest path between two points, calculated by summing traversed edges between nodes in a graph. Here the nodes are interpreted as ocean grid cell centers and the edge lengths as the great circle distances between these centers. Only adjacent cells share edges. Moves between diagonally adjacent ocean cells are allowed as long as there is at least one ocean cell on the other diagonal—no strait is implied between diagonally adjacent ocean cells when the other diagonal is occupied by two land cells. The edge lengths are weighted by ice extent so that only cells with ice concentration greater than 15% contribute distance to the MIP. The use of ice extent, having the 15% threshold, rather than ice concentration, directly, is intended to make the calculation more robust to the confounding effects of surface melt ponds on summertime passive microwave observations. Thus, a path with a zero MIP traverses only ice concentrations less than 15%. The algorithm finds the MIP to all ocean points starting from a seasonally ice-free location in the Pacific Ocean at 58°N, 180°. The distinction between NW and NE MIPs is maintained by placing an artificial land strip southward from

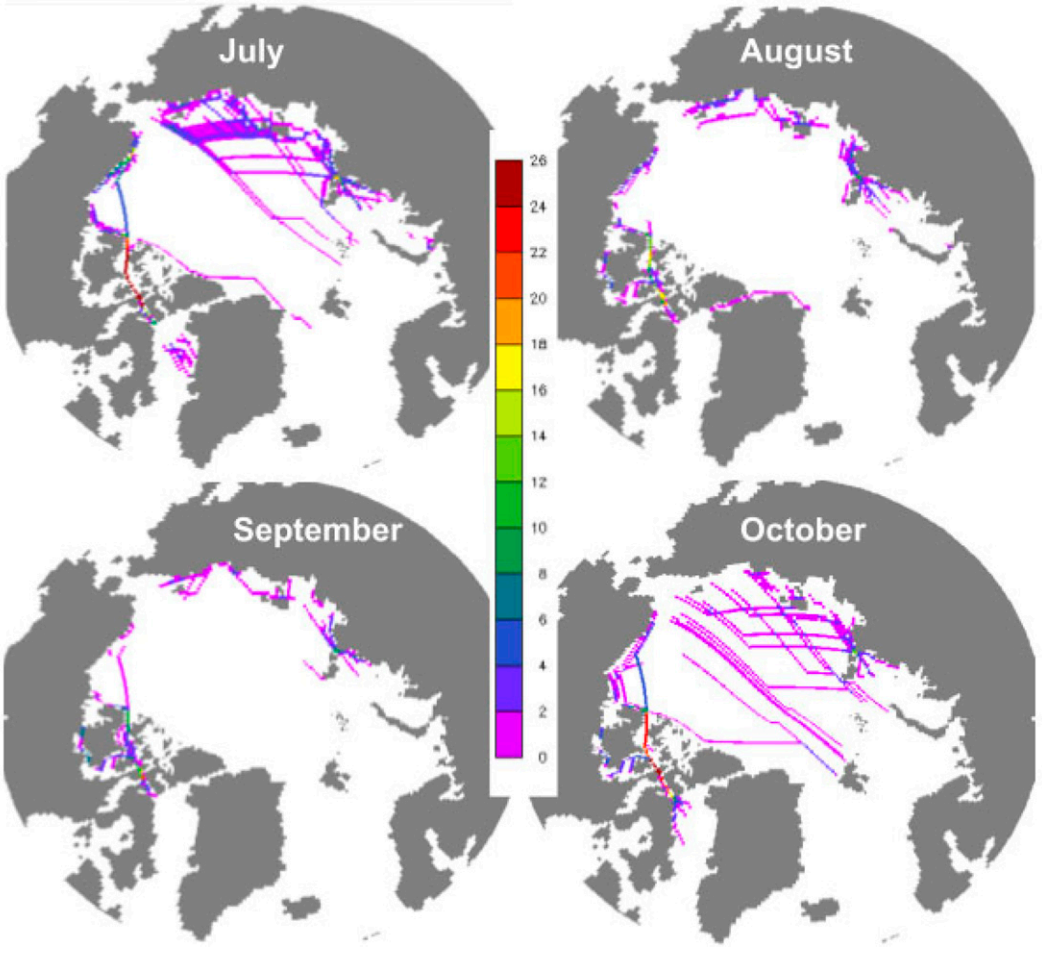

FIG. 2. MIP grid counts (number of times that a grid cell is visited by the MIP) for 15 July, 15 August, 15 September, and 15 October over the period 1992–2017. Note that only traversed points with ice cover greater than 15% are counted since the path is arbitrary in open-ocean regions.

the southern tip of Greenland. The NW and NE MIPs are taken at the seasonally ice-free locations west and east of this boundary, respectively. We use the length of the MIP for verification and not the detailed path itself. Agreement of MIP length is a necessary but not sufficient condition for detailed path agreement although the fine structure of the coastline and ice edge make it unlikely that two substantially different minimal paths will agree on length.

The MIP algorithm is applied to the daily satellite sea ice concentration observations interpolated onto the SPEAR ocean/sea ice grid to obtain our forecast verification data. Analysis is always performed on these daily MIPs—we never apply the nonlinear MIP algorithm to averaged sea ice concentrations. Figure 2 shows the number of times that given grid cells have contributed to the MIP on the 15th of the months July–October over our forecast evaluation period, 1992–2017. August–October are the months with the largest trans-Arctic ship traffic (Eguíluz et al. 2016) and so form the core target period of interest for our predictions. The NW MIPs and the NE MIPs in August and September are primarily coastal, taking advantage of coastal openings in the ice to reduce their length, while the July and October NE MIPs can either be coastal or more direct, passing near the pole. The large variation of October NE MIPs makes their forecasting particularly challenging. We note that the abrupt angles taken in some of the transpolar MIPs are due to the limited angles available when paths are required to move between adjacent grid cell centers. The rectangular grid allows a maximum of eight headings in four orientations, a constraint that does not depend on resolution. Some of the paths might be shortened by using a more complex algorithm that permits moves between nonadjacent cell centers. However, the inclusion of other variables, such as sea ice thickness, in the cost calculation would favor the optimality of the grid-adjacency Dijkstra algorithm. Anticipating that other variables will be included in the future, we employ the simplest Dijkstra algorithm despite the suboptimality of some of its trajectories.

When analyzing MIPs, we shall make use of detrending to separate subannual variability from longer-term changes. MIPs are positive definite, and this can lead to the unphysical

situation in which the trend estimate of an MIP is negative. A zero MIP forecast will give a trend anomaly forecast error even if the observed MIP is also zero when the forecast and observed negative trend estimates are different. To avoid this, we take the MIP trend estimate to be the maximum of the standard linear estimate and zero for the analyses presented in our study. Thus, the MIP trend anomaly m' is given by

$$m' = m - \max(a + bt, 0). \quad (1)$$

where m is the MIP and a and b are the intercept and slope, respectively, of the MIP regression on time t.

To estimate the potential for improvement of the numerical predictions, we will compare our forecast errors with the forecast ensemble spread. This spread provides an estimate of the forecast error lower bound, the intrinsic uncertainty that cannot be reduced by model or initialization improvement (e.g., Wang et al. 2013). The forecast ensemble standard deviation is approximately the root mean square error (RMSE) of a forecast of a hypothetical "truth" ensemble member by the mean of the other members (Bushuk et al. 2019). For the upper bound, the forecasts are considered usefully skillful to the extent that they have lower error than a reference forecast that might be employed by hypothetical practitioners making use of recent experience. This reference forecast is the sum of the average MIP over the previous 5 years for a particular target day and a damped anomaly, the deviation of the MIP from its previous 5-yr average on the day of the forecast. The 5-yr climatology follows a changing climate but does not assume linear change, which may not be warranted, especially if there is a natural variability contribution to Arctic sea ice decline (Swart et al. 2015; Zhang 2015; Ding et al. 2017). We expect a persistence forecast to be better than the climatological forecast for short lead times based on the observed persistence of sea ice extent anomalies (Walsh et al. 2019) and the improvement in short-lead-time forecast skill accompanying improvement in sea ice concentration initialization (Zhang et al. 2022). The forecast reverts to climatology over longer time scales by weighting the anomaly persistence contribution to decline exponentially with an MIP anomaly autocorrelation decay time scale determined from observations.

3. Forecast skill and error analysis

Bushuk et al. (2022) showed that the SPEAR system makes skillful seasonal predictions of summertime regional ice cover anomalies in the regions traversed by the MIPs shown in Fig. 2 with lead times of a few months, often more skillfully than a persistence forecast. Does this imply that the SPEAR system will also make skillful MIP predictions? The close relationships between the regional ice cover and MIP anomalies, shown in Fig. 3, support this possibility. The figure shows that the NW MIP is well correlated with ice cover in the Canadian Archipelago and Beaufort Sea, and the NE MIP is well correlated with the ice cover in the Kara, Laptev, and East Siberian Seas over summer and early autumn (July–October). Both the downward trend and trend anomalies contribute to this correlation. The correlation between NW MIP and its regional seas extent anomalies is 0.90

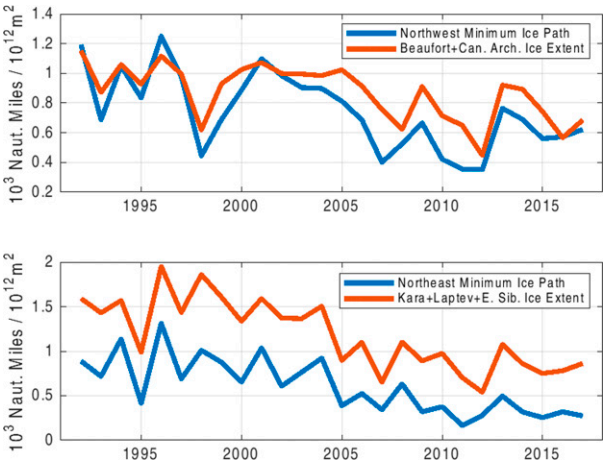

FIG. 3. July–October average (top) NW and (bottom) NE MIPs and comparable total regional ice extents. The regions are defined as in Bushuk et al. (2017). The correlations between the observed MIPs and sea ice extent in the local seas are 0.90 (NW) and 0.95 (NE), reduced to 0.84 (NW) and 0.88 (NE) after linearly detrending.

and reduced only to 0.84 when these series are first detrended. The corresponding correlations for the NE MIP and ice extent in its regional seas are 0.95 and 0.88. Therefore, subannual fluctuations account for the bulk of these close relationships and the skillful prediction of regional ice extent on subannual time scales presents an opportunity for seasonal forecasting of the NW and NE MIPs. Persistence of ice thickness anomalies after melt onset has been established as a mechanism for good prediction skill for summertime ice extent (Bonan et al. 2019; Bushuk et al. 2020).

As a preliminary to a broader evaluation later, we investigate SPEAR_LO/iceDA's 1 June-initialized forecasts and forecast errors. Figure 4 compares the observed and forecast MIPs from 1 June through the first week of November. The observations show declining NW and NE MIPs with considerable variability. Periods of zero MIP appear sporadically early in the record and more frequently in recent years. The Northeast Passage is open (zero NE MIP) for some part of the summer during the last ten years of this record. The forecasts capture the downward MIP trends but underestimate the periods of ice-free passage. This bias could be improved by using the ensemble median rather than mean as the forecast at the expense of larger overall errors due to the larger variance of sample medians relative to means. Both the NW and NE MIP forecasts are high-biased early in the season during melt back, with the NW MIP forecast having the larger bias. The NE MIP forecast is also low biased late in the season during freeze-up. The observations show some shorter, weather time scale, MIP variations that are averaged out of SPEAR_LO/iceDA's ensemble mean forecasts.

The forecasts shown in Fig. 4 clearly have significant correlation with the observations, both having correlations greater than 0.9 that remain above 0.5 after detrending. We will forgo a significance analysis because these statistics have little use for practitioners. Rather we focus, in the error comparison below,

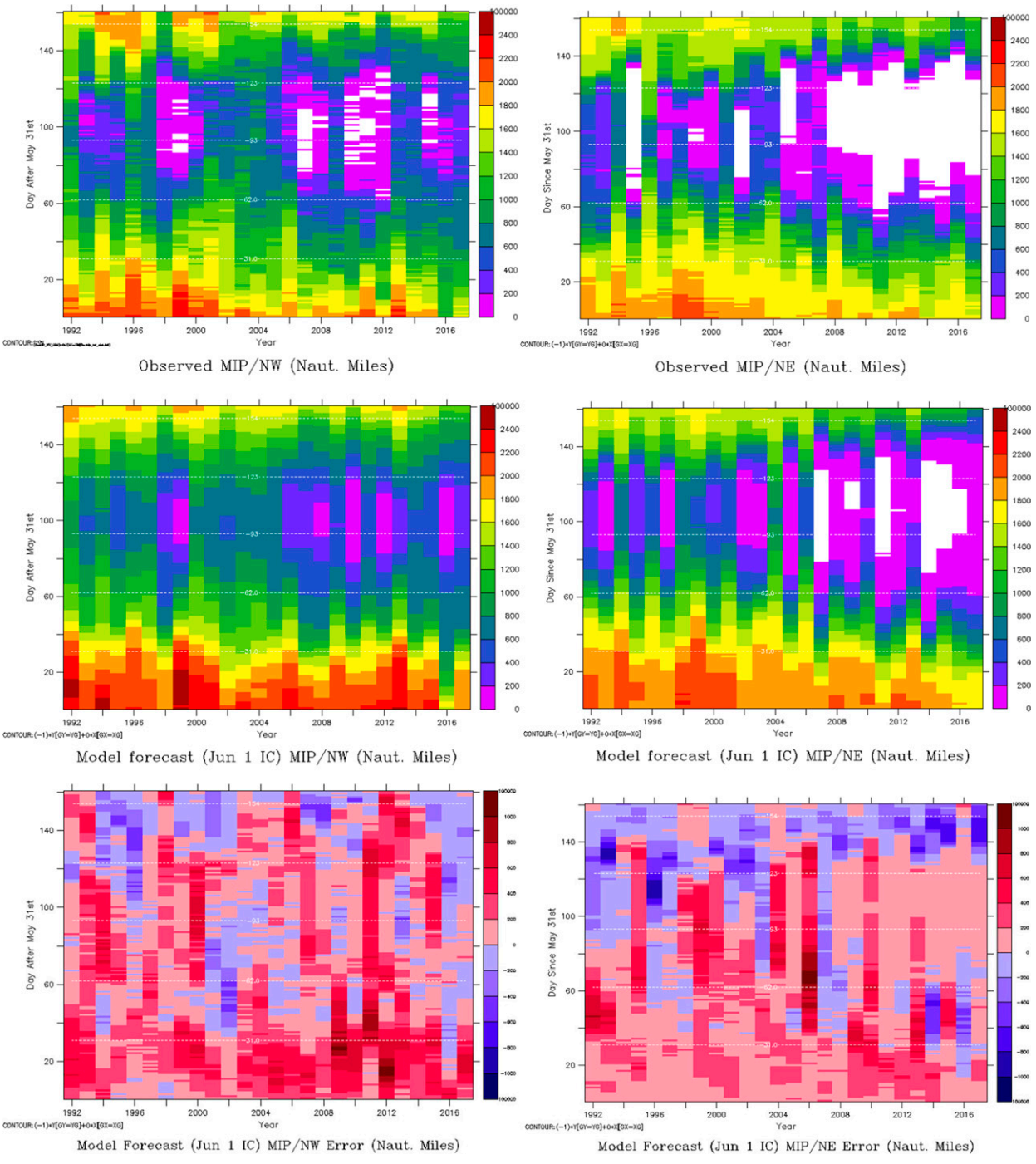

FIG. 4. The (left) NW and (right) NE MIPs from (top) satellite observations and (middle) as forecast from 1 Jun initialization with SPEAR_LO/iceDA, along with (bottom) forecast errors for 1 Jun–7 Nov. Zero MIPs are unshaded. White horizontal dotted lines indicate the first day of the months for reference. The 160-day seasonal RMSE of the MIP forecasts are 305 (NW) and 277 (NE) n mi.

on comparing the characteristic error (RMSE) of the forecasts with those that might be made by a practitioner using recent experience. We consider the forecasts to be usefully skillful if they have lower RMSEs than the damped persistence reference forecast. The MIP RMSE is directly relevant to uncertainty of the time and cost of crossing the Arctic Ocean because vessel speeds are much larger in open water, approaching 20 kt (1 kt $\approx$ 0.51 m s^{-1}), than in ice, at approximately 5 kt, and do not incur the cost of icebreaker escort. Using these characteristic speeds, a 200 n mi (1 n mi = 1.852 km) RMSE would translate to an uncertainty in traversal and icebreaker time of more than 1 day. The 160-day seasonal RMSEs of the

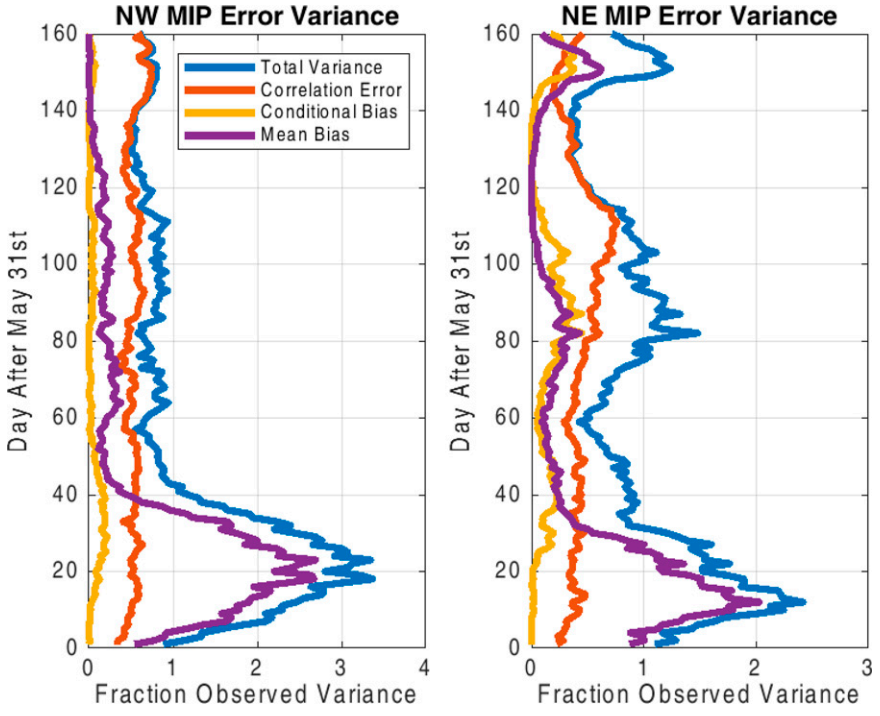

FIG. 5. The 1 Jun initialized forecast normalized error variance terms from Eq. (2) for (left) NW and (right) NE MIPs.

Fig. 4 MIP forecasts, 305 n mi (Northwest) and 277 n mi (Northeast), are consequential by this standard.

To compare opportunities for forecast improvement, we analyze the 1 June initialized forecast errors using a breakdown of the normalized squared errors due to Murphy (1988). A slight rearrangement of Murphy's Eq. (12) gives

$$s_e^2/s_o^2 = (1 - r_{\mathrm{fo}}{}^2) + \left[r_{\mathrm{fo}} - (s_f/s_o)\right]^2 + \left[(m_f - m_o)/s_o\right]^2, \quad (2)$$

where s_e is the forecast RMSE, m_f and m_o are the climatological means of the forecast and observed values, s_f and s_o are their climatological standard deviations, and r_{fo} is their correlation with each other. The first term on the right-hand side represents error due to lack of correlation between the forecast and observations. We will refer to this term as the correlation error. The second and third terms are due to the forecast conditional and mean biases, respectively. These two terms represent errors that can be removed by linear adjustment to the forecast assuming sufficient forecast and observational data are available to constrain the relationship.

Figure 5 shows the Eq. (2) normalized error variance terms and totals for the 1 June–initialized NW and NE MIP forecasts. The normalized error variances above one in June mean that the error variance exceeds the observed variance for both NW and NE MIPs over this period. The NE MIP error also occasionally rises above the observed variance in later months. Consistent with Fig. 4, the mean bias variance dominates the June errors. After June, the correlation error variance becomes the most important term. This term has little variation over time and, generally, is about half the observed variance. The conditional bias contributes the least of the three terms. From these error variance budgets, after June, improvements to the correlation error seem the most promising avenue for reducing errors. However, part of this term is intrinsic error, an unavoidable consequence of the chaotic nature of climate anomaly evolution. We must estimate this irreducible error so as to quantify the opportunity to reduce the correlation error. The correlation error is the error that remains after linear removal of biases and can be considered an upper bound estimate of the irreducible error. We note that some of the forecast errors stemming from biases may not be amenable to linear removal, such as when permanent ice-covered or ice-free conditions are forecast in a region that experiences ice cover variability. These biases would manifest partly as correlation error. Correlation error can also stem from initialization deficiencies. In the perfect model context, two historical forcing ensemble members predict each other without mean or conditional bias, but have little correlation of subannual anomalies, and hence, large correlation error.

Figure 6 shows the standard deviation of the forecast ensemble, an estimate of the irreducible error, alongside the forecast error, the trend anomaly forecast error and the renormalized correlation error from Fig. 5. These error estimates are consistent in the sense that the errors left after making the linear corrections to the forecasts lie between the forecast error and the ensemble spread. We note that the forecast ensemble spread estimate of irreducible error relies upon the ability of the model to correctly simulate variability that is unpredictable due to initial condition uncertainty

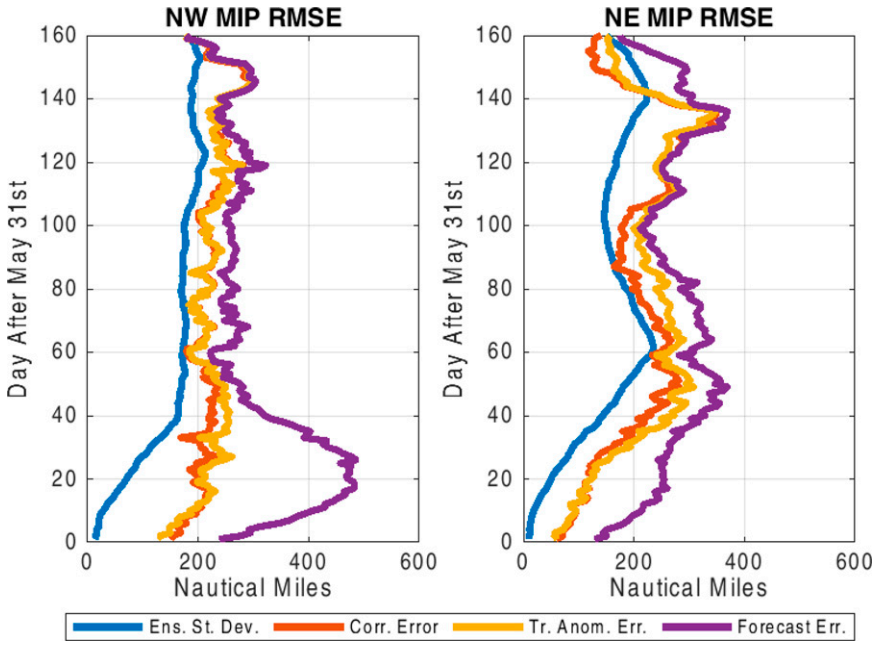

FIG. 6. The 1 Jun initialized forecast Eq. (2) correlation term RMSE (red) in comparison with forecast ensemble standard deviation (blue), trend anomaly forecast error (yellow), and full forecast error (purple) for (left) NW and (right) NE MIPs.

(Kumar et al. 2014). The SPEAR_LO systems, in particular, might underestimate this uncertainty at weather forecasting time scales because a single atmospheric state is used as a nudging target for preparing the ensemble initializations. However, weather time scale variability contributes little to the observed MIP variance in Fig. 4. Five-day smoothing reduces the observed MIP variances by only 5% (NW) and 3% (NE). The two forecasts obtained by linear correction—by trend fit correction (yellow lines) or linearly correcting forecast errors (orange lines)—track each other with occasional small relative improvements from the linear forecast correction. The linear forecast correction can reduce errors in subannual variability in addition to the long-term mean and trend but the trend fit correction gives clearer guidance for model improvement since mean, trend, and subannual variation can have different physical causes.

We are left with two imperfect measures of the lower bound on forecast error: the error remaining after linear correction, which may retain some nonlinear and reducible correlation error, and the ensemble spread, which might overestimate predictability. Kumar et al. (2014) note that the latter condition might come about if a model simulates excessive anomaly persistence. We can check for this condition by comparing the model anomaly autocorrelation function with the observed. Because initialization might introduce autocorrelation between ensemble members through drift—an effect that does not affect the model spread—we use a large ensemble of historical SPEAR-LO simulations (Delworth et al. 2020) for the autocorrelation comparison. Figure 7 shows the 30-member ensemble average of the autocorrelation of daily anomalies from the 1990–2020 trend for the NW and NE MIPs as well as the average of the NASA Team and Bootstrap autocorrelations for 1992–2017 trend anomalies. Although the observed autocorrelation is very noisy, model and observations show the same basic structures with anomaly persistence growing then shrinking over the 160-day period starting 1 June for both NW and NE MIPs.

The Fig. 7 average autocorrelations, including both NW MIP and NE MIP, are 0.52 for the SPEAR_LO model and 0.47 for the observations. Using Eq. (2), we can translate these numbers into an estimate of the ratio of the perfect model MIP RMSE to its true minimum, assuming persistence is the source of predictability. To obtain the fractional underestimate of RMSE, we take the ratio of square roots of Eq. (2) for the model and observed autocorrelations, noting that the bias terms will be zero for this perfect model calculation. The spurious reduction of the RMSE of the model relative to observations due to the excess autocorrelation, estimated this way, is about 3% $\{=[(1-0.52^2)/(1-0.47^2)]^{1/2}-1\}$. We conclude that the SPEAR_LO's excess autocorrelation should cause its ensemble spread to only slightly underestimate the minimum possible RMSE.

The observed autocorrelations are also useful for calibrating our reference forecast. We will use an autocorrelation decay time scale that gives the same average autocorrelation as the observational autocorrelation shown in Fig. 7 over the 160-day period starting 1 June. Therefore our damped persistence forecast, $\mathrm{MIP}_{\mathrm{DP}}(t_0, t)$ at time t_0 for MIP at time t, becomes

$$\mathrm{MIP}_{\mathrm{DP}}(t_0,t) = \mathrm{MIP}_{\mathrm{CLIM}}(t) + \{1-\exp[-(t-t_0)/\tau]\} \times [\mathrm{MIP}(t_0) - \mathrm{MIP}_{\mathrm{CLIM}}(t_0)], \quad (3)$$

where τ is 56 days. By using a uniformly damped persistence as our reference forecast, we have not taken advantage of the

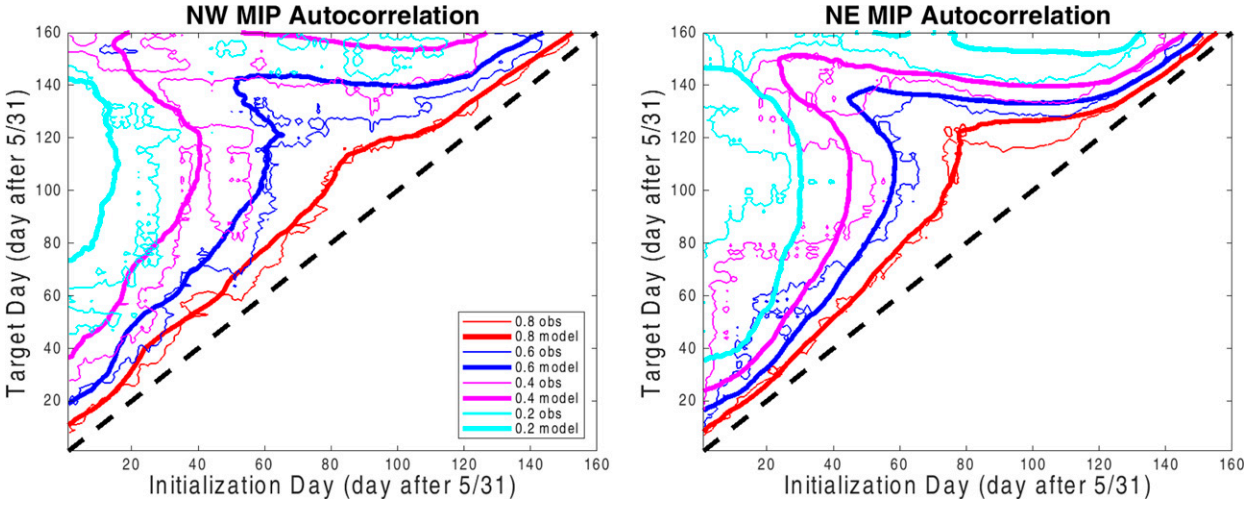

FIG. 7. Autocorrelation of (left) NW and (right) NE MIP trend anomalies, showing the averages for 1990–2020 SPEAR-LO 30-member ensemble of historical runs autocorrelations (thick) and for the average of NASA Team and Bootstrap 1992–2017 derived MIP autocorrelations (thin).

temporal variations in persistence that are suggested by Fig. 7 but have also avoided overfitting these noisy observations.

Now we examine the RMSEs for our four initializations, using average daily values over four target months. Figure 8 shows the RMSE of daily forecasts for the months of July–October initialized on the first days of May–August. Here we compare our SPEAR_LO/iceDA forecast using sea ice concentration data assimilation (Zhang et al. 2022) with the standard SPEAR_LO forecast made without sea ice concentration assimilation and our reference damped persistence forecast based on observations. We also include the error of the SPEAR_LO/iceDA trend anomaly forecast, which linearly removes the effect of climatological and trend biases, and the SPEAR_LO/iceDA forecast ensemble spread to provide a lower bound on the numerical forecast error.

Zhang et al. (2022) showed that sea ice concentration assimilation improved forecasts primarily over lead times less than one month. The impact of this improvement can be seen in Fig. 8 by comparing the 1 July and 1 August initialized forecasts for those months: the concentration assimilating initial conditions (red bars) improves the standard SPEAR forecast (blue bars) in all four cases, consistent with the Zhang et al. (2022) finding of improved coastal SIC forecasts with 1 July and 1 August initializations. The damped persistence forecast (black bars) is also competitive in these short-lead-time forecasts, providing the best forecasts in three of the four cases. The forecast ensemble spread (green bars) is considerably less than these forecast errors, indicating that a large opportunity for improvement remains. Removing the mean and trend biases only recovers a small part of this improvement (magenta bars). Improved initialization and simulation of subannual anomalies might play a role in the remaining potential improvement. The opportunity for improvement, as indicated by the reduction in error from the damped persistence forecast to the forecast ensemble spread, decreases to lead times of 2 and 3 months, remaining nearly constant thereafter. In addition to the short lead improvement, SPEAR_LO/IceDA has slightly lower overall RMSE than SPEAR_LO so we use it as our model reference. It produces lower RMSE than the observation-based reference forecasts in 11 of the 15 cases for NW MIP but only 7 of the 15 cases for NE MIP.

Figure 9 gives a summary view of forecast errors, comparing the average over all initializations and forecast lead times shown in Fig. 8. The raw model forecast improves upon the reference forecast only for the NW MIP, while the bias corrected model improves upon both the NW MIP and NE MIP reference forecasts. However, the ensemble spread suggests the possibility of making forecasts with greater than 80 n mi average RMSE reductions relative to the raw model forecasts. The trend anomaly forecast error indicates that more than half of the potential improvement remains after linear correction of mean and trend biases.

Figure 9 also shows RMSEs for the reference and raw model forecast taken relative to the alternative Bootstrap passive microwave observations in order to show the impact of errors in the observations used for data assimilation or statistical calibration. The impact is larger for the NW MIP, increasing the RMSEs by about 40 n mi. The raw model forecast remains better than damped persistence for the NW MIP but worsens relative to damped persistence for the NE MIP. After including the effects of observational uncertainty the raw model skill is still better than the observation-based reference NW MIP forecast but is now slightly worse for the NE MIP. The trend anomaly forecast error is, unexpectedly, slightly improved when taken relative to the Bootstrap observations. These comparisons indicate that errors in passive microwave observations are important for mean and trend biases but not for the prediction of subannual anomalies.

4. Summary and discussion

Using 1992–2017 retrospective subseasonal-to-seasonal predictions, we have evaluated the skill of the GFDL SPEAR

FIG. 8. Monthly averages of daily SPEAR (top) NW and (bottom) NE MIP forecast RMSEs for July–October with initializations on the first days of May–August with (red) and without (blue) sea ice concentration assimilation are compared with damped anomaly persistence forecasts (black). The SPEAR_LO/iceDA trend anomaly forecast errors (magenta) and ensemble RMSEs (green) are also shown to help to gauge the potential impact of forecast system error reductions.

forecast system at forecasting the minimum Arctic sea ice path (MIP) between the Atlantic and Pacific Oceans west and east of Greenland, corresponding roughly to the Northwest and Northeast Passages, respectively, at daily frequency over the summertime. Without mean and trend bias corrections, only the NW MIP forecasts have lower summertime average RMSE than a damped persistence forecast based on observed sea ice concentrations. Mean and trend bias corrections reduce errors below damped persistence, but considerable errors remain. Using the forecast ensemble spread as a lower bound on the error, we find large opportunities for forecast improvement, especially at lead times less than two months. The larger part of the overall model forecast error is not amenable to linear removal. These errors must consist of some combination of the nonlinear influence of biases and imperfect representation of subannual anomalies, either in their initialization or simulation. We estimate the additional error due to uncertainty in the sea ice concentration observations

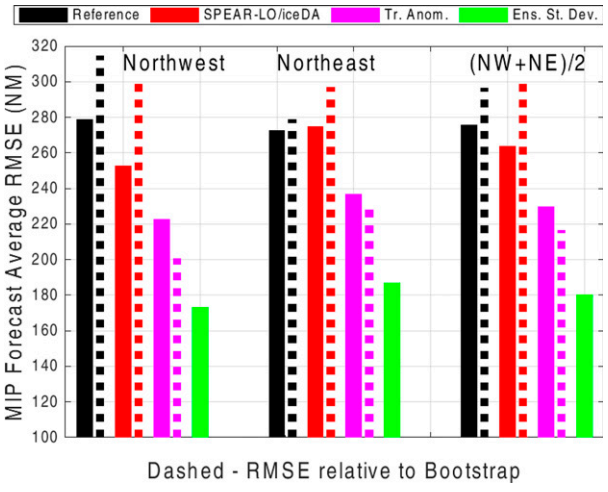

FIG. 9. Averages of Fig. 8 SPEAR MIP forecast RMSEs and ensemble spread in comparison with a damped persistence reference forecast. Thin dashed lines show the MIP RMSEs calculated relative to MIPs using Bootstrap sea ice concentration observations.

used for data assimilation and calibration by using a different passive microwave dataset (Bootstrap) for validation, finding that raw model forecast RMSEs increase and worsen somewhat relative to the observation-based reference forecast. However, the trend anomaly forecasts are slightly improved by this cross validation.

The MIP forecasts can be improved most directly by reducing biases, either using forecast postprocessing (e.g., Dirkson et al. 2019) or by improving the forecast system. Previous analyses have identified opportunities for improving summertime skill in the Arctic peripheral seas traversed by the MIPs. Bushuk et al. (2022) found that statistical predictors based on simulated thickness and concentration had better correlation with September ice cover in the Beaufort Sea than the SPEAR_MED forecasts at short lead times, indicating that SPEAR biases are degrading forecasts in that region. Good correlations of summer sea ice thickness and later concentrations in the Arctic peripheral seas (Bonan et al. 2019) suggest that improving summertime thickness initialization is a promising pathway to improved MIP forecast skill.

In this study, the MIP calculations and forecasts were intended to be sufficiently accurate for assessing skill and characterizing errors. We noted that our MIP calculation only finds the minimal path under the constraint of traversal via adjacent numerical grid centers. A more complex algorithm could reduce these paths or, more practically, it could include other influences on cost such as ice thickness. Even though it significantly affects maritime transport, we have omitted ice thickness in the current study so as to enhance the potential for validation since reliable summertime thickness observations are not currently available (Dawson et al. 2022). The passive microwave sea ice concentration observations, while offering the longest record for calibration and validation, are not best for operational use (Partington et al. 2003). Multisensor sea ice analyses should be preferred for Arctic ship routing. We showed that significant additional error is associated with uncertainty attached to the assimilated passive microwave observations. We expect differences between passive microwave and multisensor sea ice observations to be at least this large and, consequently, the associated error we diagnose to be a minimum. The verification observations were interpolated onto the model's coarse grid, and no accounting was made of the ocean depths or strait widths traversed by the MIPs. These may be constraining factors for larger vessels. Remedying these deficiencies in the current analysis would improve the relevance of the forecasts to Arctic operations.

Acknowledgments. The authors acknowledge the developers of the GFDL SPEAR forecast system used for this study and thank Will Cooke, Songmiao Fan, and three anonymous reviewers for helpful comments on the paper. The University Corporation for Atmospheric Research is sponsored by the National Science Foundation.

Data availability statement. The SPEAR reforecast data are available at https://figshare.com/articles/dataset/IceDA_reforecast_SIC/19684290.

REFERENCES

Adcroft, A., and Coauthors, 2019: The GFDL Global Ocean and Sea Ice Model OM4.0: Model description and simulation features. *J. Adv. Model. Earth Syst.*, **11**, 3167–3211, https://doi.org/10.1029/2019MS001726.

Bonan, D. B., M. Bushuk, and M. Winton, 2019: A spring barrier for regional predictions of summer Arctic sea ice. *Geophys. Res. Lett.*, **46**, 5937–5947, https://doi.org/10.1029/2019GL082947.

Bushuk, M., R. Msadek, M. Winton, G. A. Vecchi, R. G. Gudgel, A. Rosati, and X. Yang, 2017: Summer enhancement of Arctic sea ice volume anomalies in the September-ice zone. *J. Climate*, **30**, 2341–2362, https://doi.org/10.1175/JCLI-D-16-0470.1.

——, ——, ——, ——, X. Yang, A. Rosati, and R. G. Gudgel, 2019: Regional Arctic sea–ice prediction: Potential versus operational seasonal forecast skill. *Climate Dyn.*, **52**, 2721–2743, https://doi.org/10.1007/s00382-018-4288-y.

——, M. Winton, D. B. Bonan, E. Blanchard-Wrigglesworth, and T. L. Delworth, 2020: A mechanism for the Arctic sea ice spring predictability barrier. *Geophys. Res. Lett.*, **47**, e2020GL088335, https://doi.org/10.1029/2020GL088335.

——, and Coauthors, 2022: Mechanisms of regional Arctic sea ice predictability in two dynamical seasonal forecast systems. *J. Climate*, **35**, 4207–4231, https://doi.org/10.1175/JCLI-D-21-0544.1.

Cavalieri, D. J., C. L. Parkinson, P. Gloersen, and H. J. Zwally, 1996: Sea ice concentrations from Nimbus-7 SMMR and DMSP SSM/I-SSMIS passive microwave data, version 1. NSIDC, accessed 14 September 2020, https://doi.org/10.5067/8GQ8LZQVL0VL.

Comiso, J. C., 2017: Bootstrap sea ice concentrations from *Nimbus-7* SMMR and DMSP SSM/I-SSMIS, version 3. NSIDC, accessed 19 October 2021, https://doi.org/10.5067/7Q8HCCWS4I0R.

Dawson, G., and Coauthors, 2022: A 10-year record of Arctic summer sea ice freeboard from CryoSat-2. *Remote Sens. Environ.*, **268**, 112744, https://doi.org/10.1016/j.rse.2021.112744.

Delworth, T. L., and Coauthors, 2020: SPEAR: The next generation GFDL modeling system for seasonal to multidecadal

prediction and projection. *J. Adv. Model. Earth Syst.*, **12**, e2019MS001895, https://doi.org/10.1029/2019MS001895.

Dijkstra, E. W., 1959: A note on two problems in connection with graphs. *Numer. Math.*, **1**, 269–271, https://doi.org/10.1007/BF01386390.

Ding, Q., and Coauthors, 2017: Influence of high-latitude atmospheric circulation changes on summertime Arctic sea ice. *Nat. Climate Change*, **7**, 289–295, https://doi.org/10.1038/nclimate3241.

Dirkson, A., B. Denis, and W. Merryfield, 2019: A multimodel approach for improving seasonal probabilistic forecasts of regional Arctic sea ice. *Geophys. Res. Lett.*, **46**, 10 844–10 853, https://doi.org/10.1029/2019GL083831.

Dunne, J. D., and Coauthors, 2020: The GFDL Earth System Model version 4.1 (GFDL-ESM 4.1): Overall coupled model description and simulation characteristics. *J. Adv. Model. Earth Syst.*, **12**, e2019MS002015, https://doi.org/10.1029/2019MS002015.

Eguíluz, V., J. Fernández-Gracia, X. Irigoien, and C. Duarte, 2016: A quantitative assessment of Arctic shipping in 2010–2014. *Sci. Rep.*, **6**, 30682, https://doi.org/10.1038/srep30682.

Held, I. M., and Coauthors, 2019: Structure and performance of GFDL's CM4.0 climate model. *J. Adv. Model. Earth Syst.*, **11**, 3691–3727, https://doi.org/10.1029/2019MS001829.

Inoue, J., 2021: Review of forecast skills for weather and sea ice in supporting Arctic navigation. *Polar Sci.*, **27**, 100523, https://doi.org/10.1016/j.polar.2020.100523.

Kirtman, B. P., and Coauthors, 2014: The North American Multimodel Ensemble: Phase-1 seasonal-to-interannual prediction; phase-2 toward developing intraseasonal prediction. *Bull. Amer. Meteor. Soc.*, **95**, 585–601, https://doi.org/10.1175/BAMS-D-12-00050.1.

Koyama, T., T. Nakanowatari, and J. Inoue, 2021: Information retrieval for Northern Sea Route (NSR) navigation: A statistical approach using the AIS and TOPAZ4 data. *Polar Sci.*, **27**, 100626, https://doi.org/10.1016/j.polar.2020.100626.

Kumar, A., P. Peng, and M. Chen, 2014: Is there a relationship between potential and actual skill? *Mon. Wea. Rev.*, **142**, 2220–2227, https://doi.org/10.1175/MWR-D-13-00287.1.

Lu, F., and Coauthors, 2021: GFDL's SPEAR seasonal prediction system: Initialization and ocean tendency adjustment (OTA) for coupled model predictions. *J. Adv. Model. Earth Syst.*, **12**, e2020MS002149, https://doi.org/10.1029/2020MS002149.

Meier, W., and Coauthors, 2021: 2020 sea ice outlook post-season report. Sea Ice Prediction Network, https://www.arcus.org/sipn/sea-ice-outlook/2020/post-season.

Melia, N., K. Haines, E. Hawkins, and J. J. Day, 2017a: Towards seasonal Arctic shipping route predictions. *Environ. Res. Lett.*, **12**, 084005, https://doi.org/10.1088/1748-9326/aa7a60.

——, ——, and ——, 2017b: Future of the sea: Implications from opening Arctic sea routes. Government Office for Science Rep., 39 pp., https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/634437/Future_of_the_sea_-_implications_from_opening_arctic_sea_routes_final.pdf.

Mudryk, L. R., and Coauthors, 2021: Impact of 1, 2 and 4°C of global warming on ship navigation in the Canadian Arctic. *Nat. Climate Change*, **11**, 673–679, https://doi.org/10.1038/s41558-021-01087-6.

Murphy, A. H., 1988: Skill scores based on the mean square error and their relationships to the correlation coefficient. *Mon. Wea. Rev.*, **116**, 2417–2424, https://doi.org/10.1175/1520-0493(1988)116<2417:SSBOTM>2.0.CO;2.

Nakanowatari, T., and Coauthors, 2018: Medium-range predictability of early summer sea ice thickness distribution in the East Siberian Sea based on the TOPAZ4 ice–ocean data assimilation system. *Cryosphere*, **12**, 2005–2020, https://doi.org/10.5194/tc-12-2005-2018.

Notz, D., and Coauthors, 2020: Arctic sea ice in CMIP6. *Geophys. Res. Lett.*, **47**, e2019GLO86749, https://doi.org/10.1029/2019GL086749.

Partington, K., T. Flynn, D. Lamb, C. Bertoia, and K. Dedrick, 2003: Late twentieth century Northern Hemisphere sea-ice record from U.S. National Ice Center ice charts. *J. Geophys. Res.*, **108**, 3343, https://doi.org/10.1029/2002JC001623.

Smith, L. C., and S. R. Stephenson, 2013: New trans-Arctic shipping routes navigable by midcentury. *Proc. Natl. Acad. Sci. USA*, **110**, E1191–E1195, https://doi.org/10.1073/pnas.1214212110.

Swart, N. C., J. C. Fyfe, E. Hawkins, J. E. Kay, and A. Jahn, 2015: Influence of internal variability on Arctic sea-ice trends. *Nat. Climate Change*, **5**, 86–89, https://doi.org/10.1038/nclimate2483.

Tsujino, H., and Coauthors, 2018: JRA-55 based surface dataset for driving ocean–sea-ice models (JRA55-do). *Ocean Modell.*, **130**, 79–139, https://doi.org/10.1016/j.ocemod.2018.07.002.

Walsh, J. E., J. S. Stewart, and F. Fetterer, 2019: Benchmark seasonal prediction skill estimates based on regional indices. *Cryosphere*, **13**, 1073–1088, https://doi.org/10.5194/tc-13-1073-2019.

Wang, W., M. Chen, and A. Kumar, 2013: Seasonal prediction of Arctic sea ice external from a coupled dynamical forecast system. *Mon. Wea. Rev.*, **141**, 1375–1394, https://doi.org/10.1175/MWR-D-12-00057.1.

Zhang, R., 2015: Mechanisms for low-frequency variability of summer Arctic sea ice extent. *Proc. Natl. Acad. Sci. USA*, **112**, 1422296112, https://doi.org/10.1073/pnas.1422296112.

Zhang, Y.-F., M. Bushuk, M. Winton, B. Hurlin, X. Yang, T. Delworth, and L. Jia, 2021: Assimilation of satellite-retrieved sea ice concentration and prospects for September predictions of Arctic sea ice. *J. Climate*, **34**, 2107–2126, https://doi.org/10.1175/JCLI-D-20-0469.1.

——, and Coauthors, 2022: Subseasonal-to-seasonal sea ice forecast skill improvement from sea ice concentration assimilation. *J. Climate*, **35**, 4233–4252, https://doi.org/10.1175/JCLI-D-21-0548.1.