

# PROBABILISTIC FORECASTING OF THE ARCTIC SEA ICE EDGE WITH CONTOUR MODELING

BY HANNAH M. DIRECTOR<sup>1</sup>, ADRIAN E. RAFTERY<sup>2</sup> AND CECILIA M. BITZ<sup>3</sup>

<sup>1</sup>*Department of Statistics, University of Washington, [direch@uw.edu](mailto:direch@uw.edu)*

<sup>2</sup>*Departments of Statistics and Sociology, University of Washington, [raftery@uw.edu](mailto:raftery@uw.edu)*

<sup>3</sup>*Department of Atmospheric Sciences, University of Washington, [bitz@uw.edu](mailto:bitz@uw.edu)*

Sea ice, or frozen ocean water, freezes and melts every year in the Arctic. Forecasts of where sea ice will be located weeks to months in advance have become more important as the amount of sea ice declines due to climate change, for maritime planning and other uses. Typical sea ice forecasts are made with ensemble models, physics-based models of sea ice and the surrounding ocean and atmosphere. This paper introduces Mixture Contour Forecasting, a method to forecast sea ice probabilistically using a mixture of two distributions, one based on postprocessed output from ensembles and the other on observed sea ice patterns in recent years. At short lead times, these forecasts are better calibrated than unadjusted dynamic ensemble forecasts and other statistical reference forecasts. To produce these forecasts, a statistical technique is introduced that directly models the sea ice edge contour, the boundary around the region that is ice-covered. Mixture Contour Forecasting and reference methods are evaluated for monthly sea ice forecasts for 2008–2016 at lead times ranging from 0.5–6.5 months using one of the European Centre for Medium-Range Weather Forecasts ensembles.

**1. Introduction.** Sea ice, or frozen ocean water, freezes and melts annually in response to seasonal changes in atmospheric and oceanic processes. Since the satellite record began in 1979, the amount of sea ice in the Arctic has declined rapidly (Comiso et al. (2008), Stroeve et al. (2012)). Continued reduction in sea ice is expected as the effects of climate change increase. Reduced sea ice cover allows for increased Arctic shipping (Smith and Stephenson (2013), Melia, Haines and Hawkins (2016)). Since waters without sea ice are more easily navigable than waters with sea ice, the importance of forecasting sea ice has increased in response. Reliable estimates of a ship’s probability of encountering sea ice are needed to plan maritime routes that avoid sea ice. In this paper we develop statistical methods to accurately predict the probability of encountering sea ice.

Sea ice concentration, or the percent of ice-covered area, has been derived from satellite measurements for a little over 40 years and is reported on a grid. For navigational purposes the concentration field can be reduced to a binary field indicating the presence or absence of sea ice. Prediction efforts focus on the location of the ice edge contour or the boundary line that separates ice-covered regions and open water. We follow the convention in sea-ice research of categorizing a grid box as ice covered if its concentration is at least 15%. This thresholding is necessary because of potentially low accuracy of satellite products at low concentrations. Concentration reduces from about 50% to near 0% over a small region, so the area classified as sea ice is only weakly affected by the exact threshold used.

Hierarchical spatiotemporal generalized linear models have been introduced for Arctic sea ice (Zhang and Cressie (2019, 2020)). However, many operational sea ice forecasts are informed by numerical prediction systems. These systems integrate systems of differential equations to represent the physical processes that drive sea ice formation and melting. These

systems are run multiple times with slightly varied initial conditions. The resulting collection of forecasts, called the ensemble, is skillful in predicting the total area or extent of sea ice at seasonal time scales in retrospective forecasts (e.g., Chevallier et al. (2013), Msadek et al. (2014), Sigmond et al. (2013), Wang, Chen and Kumar (2013)) and in current forecasts (Blanchard-Wrigglesworth et al. (2015)). Skill has also been shown at regional scales (Bushuk et al. (2017)) and for spatial fields for some ensembles at short lead times (Zampieri, Goessling and Jung (2018)).

However, errors in ensembles are common because the underlying systems of differential equations are only approximations of the true physical processes, initial conditions are not fully known and subgrid scale phenomena are not represented (Guemas et al. (2016), Blanchard-Wrigglesworth et al. (2015)). Forecasts can be biased and/or poorly calibrated. Bias means the average behavior is systematically predicted incorrectly, and poor calibration means that the range of predicted sea ice states does not reflect the observed variability.

Statistical postprocessing, or methods that incorporate or adjust information from ensemble forecasts, can be applied to address ensembles' weaknesses while maintaining much of the skill they provide. In this paper we develop Mixture Contour Forecasting (MCF), a post-processing method to improve the calibration of sea ice forecasts. First, a method for generating distributions of sea ice edge contours is developed. The mean location of the sea ice edge contour in these distributions is informed by the mean location of the sea ice edge contour obtained from ensembles. These generated contour distributions are then weighted with climatological information to account for the time-varying skill of ensemble forecasts and aspects of sea ice that cannot be represented with a contour boundary, such as holes in the sea ice. The MCF method gives better calibrated and more accurate forecasts than the unadjusted ensemble and better calibrated forecasts than existing postprocessing techniques.

The paper is organized as follows. In Section 2, a Bayesian model for generating distributions of the sea ice edge contour is introduced. This contour model is fit to observed ice edge contours from recent years. The distribution's prior is informed by the mean ice edge predicted from the ensemble. In Section 3, the contour model is combined with climatological information using a finite mixture model. In Section 4, the performance of MCF is compared with other post-processing and statistical forecasting techniques. Section 5 concludes with discussion.

**2. Contour model.** In this section we develop a Bayesian model for the distribution of sea ice edge contours. The method works by directly modeling contours as a sequence of connected points.

*2.1. Notation and setup.* A contour is the boundary line enclosing a defined area which here is the region that contains sea ice. A contour, denoted by  $S$ , can be represented as an ordered sequence of  $N$  spatial points,  $(S_1, \dots, S_N)$ , where each  $S_i$  is an  $(x, y)$  coordinate pair. Connecting  $S_i$  to  $S_{i+1}$  for all  $i = 1, \dots, N - 1$  and  $S_N$  to  $S_1$  encloses an area. To generate a distribution of contours, we need a way to generate realizations of  $S$ .

While the sea ice edge is often referred to as a single entity, it is actually a collection of edges defining multiple contiguous areas of sea ice. These contours are modeled separately. We focus on five regions shown in the map in Figure 1. We selected these regions by modifying an existing region mask (Cavalieri and Parkinson (2012)) obtained from the National Snow & Ice Data Center (2017). We excluded parts of the Arctic ocean where sea ice does not typically form in one contiguous section, since a contour model is not appropriate for these areas. For notational simplicity we do not subscript the regions and refer to the sea ice edge contour in a given region as  $S$ .

In typical regions, sea ice is formed in contiguous sections bordering land. In these regions,  $S$  is formed by a sequence of points that proceed from the coastline, into the ocean and back

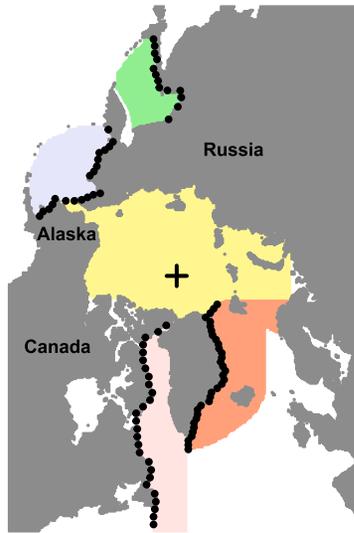


FIG. 1. Map of the Arctic ocean regions analyzed. Points indicate  $\mathbf{B}$ , the location of the fixed set of boundary points, in typical regions. The '+' symbol denotes the location of  $\mathbf{B}$  in the Central Arctic region.

to the coastline. We can reduce the number of points that need to be estimated by fixing a set of boundary points,  $\mathbf{B} = (B_1, \dots, B_n)$ , on land and considering how far into the ocean the contour extends at each boundary location. The subset of points in  $\mathbf{S}$  that must be fit are denoted by  $\tilde{\mathbf{S}}$  and are indexed  $1, \dots, n$ .

We lay out an ordered series of parallel lines,  $\mathbf{L} = (L_1, \dots, L_n)$ , that cover the region, with one line for each  $B_i$ . Each  $L_i$  extends from its corresponding point  $B_i$  to the edge of the region. We assume that one point on the contour,  $\tilde{S}_i$ , lies on each line  $L_i$ . We let  $y_i$  be the line segment extending from point  $B_i$  on the coastline to point  $\tilde{S}_i$  on the contour. The set of all line segments is denoted by  $\mathbf{Y} = (y_1, \dots, y_n)$ . The contour is formed by connecting the points  $B_i$  to  $B_{i+1}$  for all  $i = 1, \dots, n - 1$ ,  $B_n$  to  $\tilde{S}_n$ ,  $\tilde{S}_i$  to  $\tilde{S}_{i-1}$  for all  $i = n, \dots, 2$  and  $\tilde{S}_1$  to  $B_1$ . The left panel of Figure 2 illustrates these variables for the Bering Sea region. The angle of all lines in  $\mathbf{L}$  is set to approximately match the direction the sea ice extends off the land.

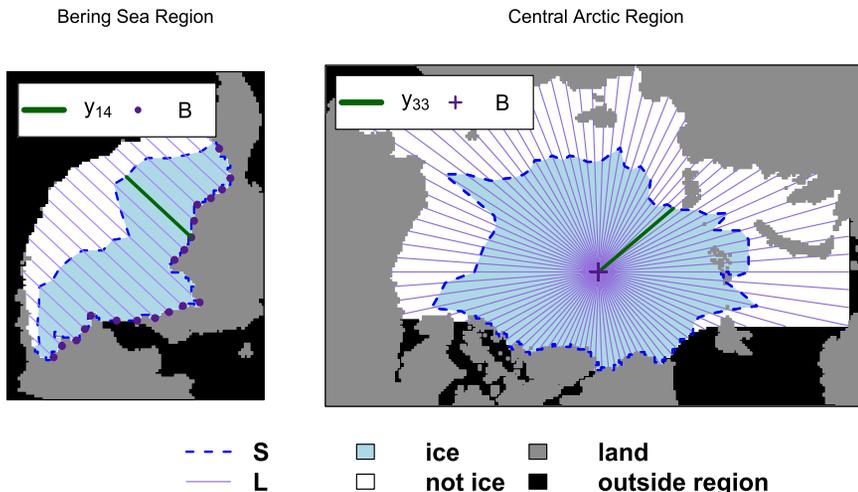


FIG. 2. Hypothetical sea ice edge contours ( $\mathbf{S}$ ), sets of fixed boundary points ( $\mathbf{B}$ ) and parallel lines ( $\mathbf{L}$ ) on which the points  $\tilde{\mathbf{S}}$  will be generated for the Bering Sea region (left) and for the Central Arctic region (right). The bold line designates the observed ice-covered line segment for the 14th (left) and 33rd (right) lines.



FIG. 3. Illustration of a hypothetical line segment  $L_i$  that crosses over a section of land. Line  $L_i$  starts at point  $B_i$ , denoted by a “+” sign, and ends at the black circle.

Sea ice in the Central Arctic region is not generally formed off a land boundary. To represent this region’s contour, we fix all the lines in  $\mathbf{L}$  to originate from a single fixed point rather than from a sequence of points. So, for this region  $B_i = B_j$  for all  $B_i, B_j \in \mathbf{B}$ . The lines extend at fixed angles evenly spaced around a circle, as illustrated in the right panel of Figure 2. In this case,  $\tilde{\mathbf{S}} = \mathbf{S}$  and  $n = N$ .

Note that, given  $\mathbf{B}$ ,  $\mathbf{Y}$  and the angles of all lines in  $\mathbf{L}$ , we have enough information to identify each  $\tilde{S}_i$ . We need only compute the length of each line segment  $y_i \in \mathbf{Y}$ . Each coordinate of the contour is then

$$(2.1) \quad \tilde{S}_i = B_i + (\|y_i\| \cos(\theta_i), \|y_i\| \sin(\theta_i)),$$

where  $\|\cdot\|$  denotes the length of the line segment and  $\theta_i$  is the angle of line  $L_i$ . Since  $\mathbf{B}$  and  $\mathbf{L}$  are fixed, we need only develop a statistical model for generating the length of the line segments in  $\mathbf{Y}$  to generate distributions of contours.

2.2. *Statistical model.* For the sea ice application, each line  $L_i$  is bounded below by zero and above by land or regions boundaries. Additionally, some  $L_i$  cross over land areas where sea ice cannot be observed. These constraints make it natural to model the proportion of each line that is ice covered, rather than the length of the line segments that compose  $\mathbf{Y}$  directly.

We now introduce notation for modeling proportions. These variables are illustrated in Figure 3. We split each  $L_i$  into sets of line segments that are in the ocean and sets of line segments that are on land. Let  $\mathbf{R}_i = \{R_{i,1}, \dots, R_{i,K_i}\}$  denote the  $K_i$  ocean line segments for  $L_i$ , and let  $\mathbf{H}_i = \{H_{i,1}, \dots, H_{i,K_i-1}\}$  denote the  $K_i - 1$  land line segments for  $L_i$ . Note that

$$(2.2) \quad \|L_i\| = \sum_{k=1}^{K_i} \|R_{i,k}\| + \sum_{k=1}^{K_i-1} \|H_{i,k}\| = \|\mathbf{R}_i\| + \|\mathbf{H}_i\|.$$

In the case where  $L_i$  just goes through ocean,  $\mathbf{R}_i = R_{i,1} = L_i$  and  $\mathbf{H}_i = \emptyset$ .

Since the line segments forming any  $\mathbf{H}_i$  cannot contain sea ice, we focus on modeling the proportion of the corresponding  $\mathbf{R}_i$  that are ice covered. More formally, let

$$(2.3) \quad \pi_i = \frac{\|y_i \cap \mathbf{R}_i\|}{\|\mathbf{R}_i\|},$$

where the numerator denotes the length of  $y_i$  that intersects the ocean line segments and the denominator denotes the total length of ocean line segments in  $L_i$ . The set of all proportions is denoted by  $\boldsymbol{\pi} = (\pi_1, \dots, \pi_n)$ .

We develop a model for  $\boldsymbol{\pi}$  that can be used to generate  $\mathbf{Y}$  and corresponding  $\tilde{\mathbf{S}}$ . For ease of modeling we transform the proportions to the real line. Let

$$(2.4) \quad \tilde{\pi}_i = \begin{cases} \text{logit}(\pi_i) & \text{for } \epsilon \leq \pi_i \leq 1 - \epsilon, \\ \text{logit}(\epsilon) & \text{for } \pi_i < \epsilon, \\ \text{logit}(1 - \epsilon) & \text{for } \pi_i > 1 - \epsilon, \end{cases}$$

where  $\text{logit}(x) = \log(x/(1 - x))$  and  $\epsilon$  is small. In our implementation  $\epsilon = 0.01$ . The set of transformed proportions,  $\tilde{\boldsymbol{\pi}}$ , are modeled using a multivariate normal distribution,

$$(2.5) \quad \tilde{\boldsymbol{\pi}} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma}),$$

where  $\boldsymbol{\mu}$  is an  $n \times 1$  mean vector and  $\boldsymbol{\Sigma}$  is an  $n \times n$  covariance matrix.

The data generating process for  $\tilde{\mathbf{S}}$  is then as follows. First, an underlying random vector,  $\tilde{\boldsymbol{\pi}}$ , is drawn. Then, each  $\tilde{\pi}_i$  is transformed back to a proportion, that is,  $\pi_i = \text{ilogit}(\tilde{\pi}_i)$  where  $\text{ilogit}(x) = \exp(x)/(1 + \exp(x))$ . The length of the corresponding  $y_i$  is computed from  $\pi_i$  as follows. Let  $D$  denote the maximum index of a line segment of  $\mathbf{R}_i$  that is fully ice covered,

$$(2.6) \quad D = \underset{d}{\operatorname{argmax}} \left\{ \frac{\sum_{k=1}^d \|\mathbf{R}_{i,k}\|}{\|\mathbf{R}_i\|} < \pi_i \right\}.$$

The length of  $y_i$  is then

$$(2.7) \quad \|y_i\| = \pi_i \|\mathbf{R}_i\| + \sum_{k=1}^{D-1} \|\mathbf{H}_{i,k}\|.$$

In other words,  $y_i$  is composed of  $\pi_i$  proportion of  $\mathbf{R}_i$  and all the line segments of  $\mathbf{H}_i$  that must be crossed to reach the  $D$ th segment of  $\mathbf{R}_i$ . For all  $i$ , the lengths of  $y_i$  are then used to compute  $\tilde{\mathbf{S}}_i$  using equation (2.1). Connecting the points in  $\tilde{\mathbf{S}}$ , along with the points in  $\mathbf{B}$ , if applicable, produces a generated contour.

In rare cases the generated values in  $\tilde{\mathbf{S}}$  will result in a contour that intersects itself. When these self-intersections occur, the contour fails to be a boundary around a single contiguous area. A small adjustment to the modeled line is made to the part(s) of the contours that have self-intersections. Each part of the contour that contains a self-intersection is replaced with an approximation that does not have a self-intersection obtained using the Douglas–Peucker algorithm (Douglas and Peucker (1973)). Details of this adjustment are given in the Supplementary Material A Section 3 (Director, Raftery and Bitz (2021)). The generated contour also, occasionally, comes very close to touching a region boundary or land without actually touching it. Since this is physically unrealistic,  $\|y_i\|$  values that are within half the nominal length of a grid box (12.5 nominal kilometers) of a region or land boundary are adjusted to exactly align with the region or land boundary. The total area involved in both adjustments is very small. These adjustments just ensure that any individual generated contour looks physically realistic.

**2.3. Parametric covariance.** To allow for efficient fitting of equation (2.5), we define a parametric covariance structure. In sea ice observations, the mean and covariance of the ice-covered proportion of each line varies substantially within and across regions. To represent these features well, we need a statistical model with a reasonably flexible covariance structure. The values of  $\tilde{\pi}_i$  and  $\tilde{\pi}_j$  tend to be more similar when  $L_i$  and  $L_j$  are close together. So, we structure our covariance in typical regions based on the differences between the indices of the lines in  $\mathbf{L}$ . We let  $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}(\boldsymbol{\sigma}, \kappa)$ , where  $\boldsymbol{\sigma} = (\sigma_1, \dots, \sigma_n)$  and  $\kappa > 0$ . The element,  $\boldsymbol{\Sigma}_{ij}$ , in the  $i$ th row and  $j$ th column of this covariance is

$$(2.8) \quad \boldsymbol{\Sigma}_{ij} = \sigma_i \sigma_j \exp\left(-\frac{|i - j|}{\kappa}\right),$$

where  $|\cdot|$  denotes the absolute value.

In the Central Arctic region, the lines are laid out in a circle so that the first and last lines are close to each other despite their indices being far apart. The difference between the indices of line  $i$  and line  $j$  does not necessarily correspond to the distance between lines  $L_i$  and  $L_j$ . So, we apply an alternative covariance function based on the difference between angles  $\theta_i$  and  $\theta_j$ . Various covariance functions based on differences between angles have been proposed (Gneiting (2013)). Like the other regions, we apply an exponential covariance structure where  $\boldsymbol{\sigma} = (\sigma_1, \dots, \sigma_n)$ ,  $\kappa > 0$ , and the element in the  $i$ th row and  $j$ th column of the covariance matrix is

$$(2.9) \quad \boldsymbol{\Sigma}_{ij} = \sigma_i \sigma_j \exp\left(-\frac{d(\theta_i, \theta_j)}{\kappa}\right),$$

where  $d(\theta_i, \theta_j) \in [0, \pi]$  is the smaller angle between  $\theta_i$  and  $\theta_j$ .

We find that an exponential covariance structure fits the data well. This covariance structure allows for the correlation to drop off rapidly as lines become farther apart while maintaining some nonzero correlation among all lines. Allowing for the latter behavior is needed, since some region-wide correlation would be expected given large-scale phenomena that could occur across a region, such as a particularly cold or warm month.

*2.4. Number of lines in  $L$ .* Setting  $n$ , the number of lines in  $L$ , involves tradeoffs between accuracy and computation time. With more lines, contours can be represented in more detail. However, computation increases with the corresponding increase in the size of the covariance matrix. We set  $n = 90$  in the Central Arctic region and scale  $n$  for other regions relative to how their area compares to the area of the Central Arctic region. More discussion of how  $n$  is selected is given in the Supplementary Material A Section 2 (Director, Raftery and Bitz (2021)).

*2.5. Prior distribution of the mean sea ice edge,  $\mu$ .* Parameters are estimated with a Bayesian approach. The ensemble forecast informs the likely location of the mean sea ice edge. Using the ensemble output, we place a strong prior on  $\mu$ , the mean sea ice edge. We adjust the ensemble output with contour shifting, a technique to reduce systematic bias, before using it in the prior (Director, Raftery and Bitz (2017)).

*2.5.1. Contour shifting.* We apply the contour shifting method introduced in Director, Raftery and Bitz (2017) for the Central Arctic region to all regions. We summarize the method using the notation in Section 2. For some historical training period preceding the forecast year, we compare the ice edge predicted by the ensemble mean forecast to the observed ice edge. For each year  $j$  and line  $L_i$  in a particular region, we record the length of the line segments extending from each point on the coastline,  $B_i$ , to the corresponding point on the observed ice edge contour,  $\tilde{S}_{i,j}^{\text{obs}}$ . We also record the lengths of the line segments from each point on the coastline,  $B_i$ , to the corresponding point on the ensemble mean ice edge,  $\tilde{S}_{i,j}^{\text{ens}}$ . We denote these lengths by  $\|y_{i,j}^{\text{obs}}\|$  and  $\|y_{i,j}^{\text{ens}}\|$ , respectively. Assuming linear change in these lengths over time, we estimate the length to which the sea ice will extend on line  $L_i$  at some new time point  $t$  for the observed ice edge,

$$(2.10) \quad \|\hat{y}_{i,t}^{\text{obs}}\| = \hat{\alpha}_i^{\text{obs}} + \hat{\beta}_i^{\text{obs}}t,$$

and the ensemble mean ice edge,

$$(2.11) \quad \|\hat{y}_{i,t}^{\text{ens}}\| = \hat{\alpha}_i^{\text{ens}} + \hat{\beta}_i^{\text{ens}}t.$$

Here,  $\hat{\alpha}_i^{\text{obs}}$ ,  $\hat{\alpha}_i^{\text{ens}}$ ,  $\hat{\beta}_i^{\text{obs}}$ , and  $\hat{\beta}_i^{\text{ens}}$  denote fitted regression coefficients. These regressions are fit using Huber M-estimation, a form of robust linear regression (Huber (2011)).

The difference between  $\|\hat{y}_{i,t}^{\text{obs}}\|$  and  $\|\hat{y}_{i,t}^{\text{ens}}\|$  gives the expected difference between the length predicted by the mean ensemble and the length that will be observed at time  $t$ . So, the forecasted length on line  $L_i$  at time  $t$  is expected to be

$$(2.12) \quad \|\hat{y}_{i,t}^{\text{CS}}\| = \|\hat{y}_{i,t}^{\text{ens}}\| + (\|\hat{y}_{i,t}^{\text{obs}}\| - \|\hat{y}_{i,t}^{\text{ens}}\|),$$

where the superscript CS indicates that the contour shifting adjustment has been made.

Each adjusted length,  $\|\hat{y}_{i,t}^{\text{CS}}\|$ , can be combined with the corresponding  $B_i$  and  $L_i$  values as in Section 2.1 to produce new ice edge contours. The resulting contours may have self-intersections. These self-intersections are corrected in the same way as when they occur with generated contours as described in Section 2.2 and in Supplementary Material A Section 3 (Director, Raftery and Bitz (2021)). Contours from ensembles that have been adjusted to reduce bias are referred to as contour shifted. These contours have reduced systematic error compared to the initial ensemble forecasts.

2.5.2. *Prior for mean proportions ice covered.* We can now incorporate this reduced bias form of the forecasted ensemble mean ice edge contour into the prior for  $\boldsymbol{\mu}$ . We use the following prior distribution:

$$(2.13) \quad \boldsymbol{\mu} \sim N(\boldsymbol{\mu}_0, \boldsymbol{\Lambda}_0),$$

where  $\boldsymbol{\mu}_0$  is an  $n \times 1$  mean vector informed by the ensemble forecast. Let

$$(2.14) \quad \pi_{i,t}^{\text{CS}} = \frac{\|y_{i,t}^{\text{CS}} \cap \mathbf{R}_i\|}{\|\mathbf{R}_i\|}$$

denote the proportion of  $\mathbf{R}_i$  that  $y_{i,t}^{\text{CS}}$  covers in the contour shifted ensemble mean ice edge. Then, for all  $i$ , let

$$(2.15) \quad \mu_{0,i} = \begin{cases} \text{logit}(\|\pi_{i,t}^{\text{CS}}\|) & \text{for } \epsilon \leq \pi_{i,t}^{\text{CS}} \leq 1 - \epsilon, \\ \text{logit}(\epsilon) & \text{for } \pi_{i,t}^{\text{CS}} < \epsilon, \\ \text{logit}(1 - \epsilon) & \text{for } \pi_{i,t}^{\text{CS}} > 1 - \epsilon. \end{cases}$$

The matrix  $\boldsymbol{\Lambda}_0$  is a  $n \times n$  diagonal covariance matrix. Details on setting  $\boldsymbol{\Lambda}_0$  are given in the Supplementary Material A Section 4.2 (Director, Raftery and Bitz (2021)).

2.6. *Prior for  $\boldsymbol{\Sigma}$ .* For the priors for the covariance parameters,  $\boldsymbol{\sigma}$  and  $\kappa$ , we only use information about the physical constraints. While ensembles could provide information about covariance, the variability of the ensembles we have analyzed do not align with the variability seen in observations.

Since standard deviation values are bounded below and considerable differences in variances exist for the  $\tilde{\pi}_i$  values, we select an independent uniform prior for each  $\sigma_i$  such that

$$(2.16) \quad \sigma_{0,i} \stackrel{\text{iid}}{\sim} \text{Unif}(\alpha_{\sigma,0}, \beta_{\sigma,0}).$$

Details on setting  $\alpha_{\sigma,0}$  and  $\beta_{\sigma,0}$  are given in the Supplementary Material A Section 4.3 (Director, Raftery and Bitz (2021)). With little information from which to anticipate how correlation decreases with distance, we use the following vague prior for  $\kappa$ ,

$$(2.17) \quad \kappa_0 \sim \text{Unif}(\alpha_{\kappa,0}, \beta_{\kappa,0}),$$

where  $\alpha_{\kappa,0} = 0.05$  and  $\beta_{\kappa,0} = 20$ . This prior ensures that  $\kappa$  remains positive.

2.7. *Posterior distribution.* To fit this model, we need a set of observed contours drawn from the same distribution. We treat the contours in the  $P$  years immediately preceding the forecast year as independent samples. With this approach we assume that the distribution of the contours is stationary over the  $P$ -year period. While this stationarity assumption is not strictly true given climate change, for decadal time scales the effects of the climate change trend on sea ice are small relative to year-to-year variability. Therefore, we fix  $P$  and assume these recent observations provide a reasonable basis on which to build a Bayesian model. We index the years with the subscripts  $j = 1, \dots, P$ . We denote the set of  $n$  observed proportions in year  $j$  by  $\tilde{\boldsymbol{\pi}}_j$ . The element  $\tilde{\pi}_{ij}$  is the proportion of  $\mathbf{R}_i$  that  $y_{ij}$  covers in year  $j$ .

Combining the likelihood for the observed proportions with the prior distributions introduced in Sections 2.5 and 2.6 gives the posterior distribution

$$(2.18) \quad \prod_{j=1}^P \{p(\tilde{\boldsymbol{\pi}}_j, \boldsymbol{\mu}, \boldsymbol{\sigma}, \kappa)\} p(\boldsymbol{\mu}) p(\boldsymbol{\sigma}) p(\kappa) = \prod_{j=1}^P \{N(\tilde{\boldsymbol{\pi}}_j | \boldsymbol{\mu}, \boldsymbol{\Sigma}(\boldsymbol{\sigma}, \kappa))\} N(\boldsymbol{\mu} | \boldsymbol{\mu}_0, \boldsymbol{\Lambda}_0) \\ \times \prod_{i=1}^n \{\text{Unif}(\sigma_i | \alpha_{\sigma,0}, \beta_{\sigma,0})\} \text{Unif}(\kappa | \alpha_{\kappa,0}, \beta_{\kappa,0}).$$

The posterior means of  $\boldsymbol{\mu}$  and  $\boldsymbol{\Sigma}$  can be used with equations (2.5), (2.8) and (2.9) to generate  $\tilde{\boldsymbol{\pi}}$ .

2.8. *Model fitting.* We sample from the posterior distribution in equation (2.18) for each region independently with Markov chain Monte Carlo (MCMC), using the observed sea ice in the preceding  $P$  years. We use Metropolis steps for updating each  $\mu_i$ ,  $\sigma_i$  and  $\kappa$ . Normal proposals are used for each parameter at each iteration centered at their current value. Log acceptance rates are given in Supplementary Material A Section 5.1 (Director, Raftery and Bitz (2021)). MCMC diagnostics are given in Supplementary Material A Section 5.2 (Director, Raftery and Bitz (2021)). Regions that are either completely filled with sea ice or contain no sea ice in all training years are omitted from model fitting. In such cases we predict full ice-cover or no sea ice, respectively. In some months of the year, the observed proportions at the start and/or end of the fixed boundary lines are 0 or 1 for all observed  $P$ . We fix these lines with proportions of 0 or 1 rather than fit them. This omission in fitting avoids estimating an excessively high  $\kappa$  due to perfect correlation among the lines in these sections. In the Central Arctic region, sets of lines bordering the Canadian Archipelago that have proportion 1 for all  $P$  years are fixed in the same way.

**3. Mixture contour forecasting.** The contour model described in Section 2 generally provides reasonable forecasts of the sea ice edge contour but does have some limitations for predicting sea ice presence. First, these forecasts only focus on the contour. While the vast majority of the sea ice is contained within contiguous areas within the main sea ice edge, small areas of sea ice sometimes form away from this main area. Areas of open water, called polynyas, also sometimes form within the main sea ice area. The contour model proposed in the previous section cannot represent these features. Second, forecasts of this type are tied to the existing ensemble forecast, so if the initial ensemble forecast is not very accurate, such as at long lead times, the resulting forecast will not be very skillful. We address these weaknesses by developing a mixture model that combines the contour model with a climatological forecast that has different strengths and weaknesses.

MCF produces a forecast distribution that is a mixture, or weighted average, of two component distributions: the contour model introduced in the previous section and a distribution that represents recent climatology. Here, we define the climatology forecast for each grid box as the proportion of times sea ice has been present in that grid box in the  $P$  years preceding the forecast year.

The climatology forecast has different attributes than the contour model. The climatology forecast can represent features such as polynyas and sea ice away from the main ice edge contours. However, this forecast's reliance on only the small number of observations in the past  $P$  years means that it does not capture all plausible ice edges. This weakness of the climatology forecast is greatest in the highly variable months around the sea ice minimum.

The weighting of the two models can be viewed as a simple case of ensemble Bayesian model averaging (Raftery et al. (2005)). The weight is estimated by maximum likelihood using observations and predictions from preceding years. Let  $w$  be the weight of the contour model and  $1 - w$  the weight of the climatology distribution. Also, let  $\gamma_{s,t}$  be a binary indicator of sea ice presence for grid box  $s$  and year  $t$  in the training period. Let  $g_p(\gamma_{s,t})$  and  $g_c(\gamma_{s,t})$  be the predicted Bernoulli probability of sea ice presence in grid box  $s$  at time  $t$  obtained from the contour model and the climatology, respectively. In the former case the predicted probability is the proportion of time that grid box  $s$  is within the area enclosed by the generated contours at time  $t$ . The predicted probability of sea ice presence at grid box  $s$  at time  $t$  is then

$$(3.1) \quad p(\gamma_{s,t}) = w g_p(\gamma_{s,t}) + (1 - w) g_c(\gamma_{s,t}).$$

Assuming that errors in space and time are independent, the corresponding log-likelihood is

$$(3.2) \quad l(w) = \sum_t \sum_s \log \{ w a_s g_p(\gamma_{s,t}) + (1 - w) a_s g_c(\gamma_{s,t}) \}.$$

The variable  $a_s$  is the proportion of the entire area that is in grid box  $s$ , that is,  $\sum_s a_s = 1$ . Use of  $a_s$  is needed, since the grid boxes do not all have the same area. Assuming spatial and temporal independence is almost certainly inaccurate, but Raftery et al. (2005) found in a similar case that results were not particularly sensitive to this assumption.

To maximize this log-likelihood, we use the expectation-maximization algorithm (Dempster, Laird and Rubin 1977). This optimization algorithm can be applied in situations where, if some unobserved quantity were known, estimation of the variable(s) of interest would be simple. In this case, estimating  $w$  would be easy if we knew for every grid box and time point whether the climatology model or the contour model estimated the observed sea ice presence more accurately. So, we introduce the latent variable  $z_{p,s,t}$  which has value 1 if the contour model is the best forecast for grid box  $s$  in year  $t$  and 0 otherwise. The variable  $z_{c,s,t}$  is defined analogously for climatology. Note that only one of the parameters  $z_{c,s,t}$  or  $z_{p,s,t}$  could truly be 1; but, for estimation, these parameters can take any value in the interval  $[0, 1]$ . Also, note  $\hat{z}_{p,s,t} = 1 - \hat{z}_{c,s,t}$ . Then, the E-step is

$$(3.3) \quad \hat{z}_{p,s,t}^{(j)} = \frac{w^{(j-1)} a_s g_p(\gamma_{s,t})}{w^{(j-1)} a_s g_p(\gamma_{s,t}) + (1 - w^{(j-1)}) a_s g_c(\gamma_{s,t})},$$

and the M-step is

$$(3.4) \quad w^{(j)} = \frac{\sum_t \sum_s a_s \hat{z}_{p,s,t}^{(j)}}{\sum_t \sum_s a_s}$$

for the  $j$ th iteration. To avoid degeneracies, any  $(s, t)$  pairs where  $g_p(\gamma_{s,t}) = g_c(\gamma_{s,t})$  are omitted from this maximization. Therefore, the denominator in equation (3.4) may be unequal to the number of years in the training period.

#### 4. Method evaluation.

4.1. *Model outputs and observations.* All postprocessing methods are applied to the fifth generation of the European Centre for Medium-Range Weather Forecasts (ECMWF) seasonal forecasting system (SEAS5) (Johnson et al. (2019), European Centre for Medium-Range Weather Forecasts (2017)). The relevant model output can be downloaded from the Copernicus Climate Change Service Climate Data Store (Copernicus Climate Change Service (2019)). Among a set of publicly available ensembles without postprocessing, ECMWF has been shown to generally be the most skillful (Zampieri, Goessling and Jung (2018)). The 25-member ensemble ECMWF forecasts are initialized monthly and extend to 215 days. Model output was regridded to the National Snow and Ice Data Center Polar stereographic grid with an approximately 25 km by 25 km grid (National Snow & Ice Data Center (2017)) using a nearest-neighbors method (Zhuang (2018)). Daily model output was averaged to monthly to match observations.

We evaluate forecast accuracy by comparing predictions to a monthly-averaged concentration produced from the National Aeronautics and Space Administration satellites Nimbus-7 SMMR and DMSP SSM/I-SSMIS (Comiso (2017)). These data are downloaded from the National Snow and Ice Data Center. Sea ice presence is defined as having concentration of at least 0.15.

We evaluate forecasting skill for monthly-averaged sea ice at lead times of 0.5 months to 6.5 months in the year 2008–2016. We report lead times treating the monthly mean as the halfway point within a month. For example, the 0.5-month lead forecast for January refers to the average of the first 31 days of a forecast initialized on January 1st. Grid boxes that are coded as land in the observations, the ensemble or the *IceCast* R package (Director, Raftery and Bitz (2021)) are treated as land.

TABLE 1

*Summary of forecast types evaluated. Probabilistic forecasts give estimates in the interval  $[0, 1]$ , and binary forecasts indicate predicted sea ice presence*

Forecast	Probabilistic	Binary
Ensemble	proportion of ensemble members predicting sea ice	indicator of whether median ensemble member predicts sea ice
Contour	ensemble mean forecast adjusted with contour shifting and calibrated by generating contours	ensemble mean forecast adjusted with contour shifting
Climatology	proportion of observations in the 10 years preceding the forecast year that contain sea ice	indicator of whether at least five of the 10 years preceding the forecast year contained sea ice
Mixture Contour Forecast (MCF)	forecast formed by weighting probability densities from climatology and the contour model	indicator of whether forecast formed by weighting probability densities from climatology and the contour model predicts sea ice with $p \geq 0.5$
Trend Adjusted Quantile Mapping (TAQM)	ensemble postprocessed using technique in <a href="#">Dirkson, Merryfield and Monahan (2019)</a>	NA
Damped Persistence	NA	indicator of whether predicted concentration from a damped persistence forecast is at least 0.15 (modified from <a href="#">Wayand, Bitz and Blanchard-Wrigglesworth (2019)</a> )

The forecasts previously described are summarized in lines 1–4 of Table 1. Beginning in 1993, all years preceding the forecast year are used in fitting contour shifting. A 10-year rolling window is used to fit the statistical model for generating contours and in the climatology forecast weighted in MCF. A three-year rolling window is used to determine the weights in MCF. Performance accuracy is generally insensitive to moderate changes in these training lengths. (See Supplementary Material A Section 6 ([Director, Raftery and Bitz \(2021\)](#)).) One hundred contours are generated for each forecast.

**4.2. Reference forecasts.** We compare our results to two additional reference forecasts summarized in lines 5–6 of Table 1. Trend-adjusted quantile mapping (TAQM) is another recently developed statistical postprocessing method for sea ice ([Dirkson, Merryfield and Monahan \(2019\)](#)). TAQM fits a parametric probability distribution to ensemble model output and applies a specialized version of quantile mapping to produce probabilistic forecasts of concentration. TAQM does not predict the probability of sea ice presence directly, but [Dirkson, Merryfield and Monahan \(2019\)](#) do use the resulting distribution of the concentration to predict the probability of sea ice presence. We also compute a damped persistence forecast in a manner similar to [Wayand, Bitz and Blanchard-Wrigglesworth \(2019\)](#). Exact implementation details are given in Supplementary Material A Section 7 ([Director, Raftery and Bitz \(2021\)](#)).

**4.3. Visualizing forecasts.** Uncertainty information is needed for maritime planning to adequately evaluate risks and benefits. Like [Gneiting, Balabdaoui and Raftery \(2007\)](#), we consider accurate model calibration to be vital for probabilistic forecasts. We illustrate the importance of calibration in this context with Figure 4, which shows samples of four probabilistic forecasts for September 2008. The corresponding observed sea ice edge for September

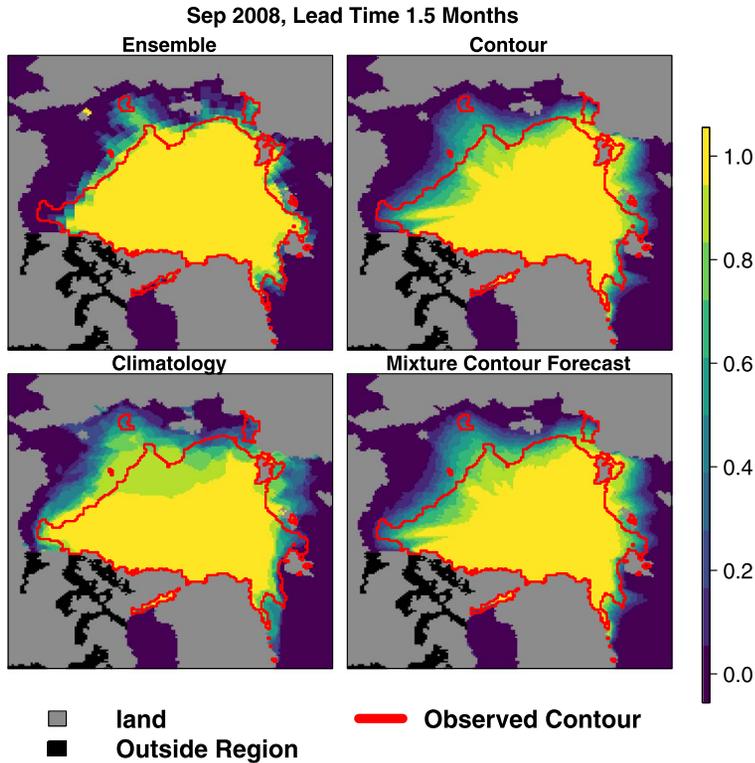


FIG. 4. Forecasts of the probability of sea ice presence for September 2008 using different methods. The forecasts are described in Table 1. The bold line is the observed sea ice edge contour.

2008 is also plotted for reference. Figure 4 illustrates the types of forecasting errors that can occur when forecasts are not calibrated. Specifically, events with low predicted probability occur more often than expected, and/or events with high predicted probability occur less often than expected.

For the contour model and MCF, the observed contour is almost entirely contained within regions with positive probability and has only small areas where sea ice is predicted with probability 1 but sea ice is not present in observations. The MCF forecast is slightly more variable than the contour model, reflecting its weighting with climatology. Since  $w$  is high for September at a 1.5-month lead time, the difference between the contour model and MCF is small. In cases where  $w$ , the weight on the contour model, is lower, the difference between the contour model and MCF may be larger. For the climatology and ensemble forecasts, the observed contour goes through some regions with probability 0, suggesting that these forecasts are not sufficiently variable. Discrepancies like these between the forecasted probability and what will likely occur makes maritime planning and risk mitigation difficult.

**4.4. Assessing calibration.** We now evaluate model calibration for the probabilistic forecasts with reliability diagrams. These diagrams plot the forecasted probability of observing sea ice against the proportion of times sea ice was observed. A perfectly calibrated forecast would have all points on the  $y = x$  line, that is, grid boxes forecasted to contain sea ice with a given probability contain sea ice the same proportion of the time. So, the closer the points lie to the  $y = x$  line, the better calibrated the forecast is.

Shipping varies seasonally in the Arctic, with more shipping in months around the annual sea ice minimum in September (Ellis and Brigham (2009)). We emphasize performance in these peak shipping months. In Figure 5, we show the reliability diagrams for the peak-shipping months for the probabilistic forecasts. Predictions from MCF are substantially better

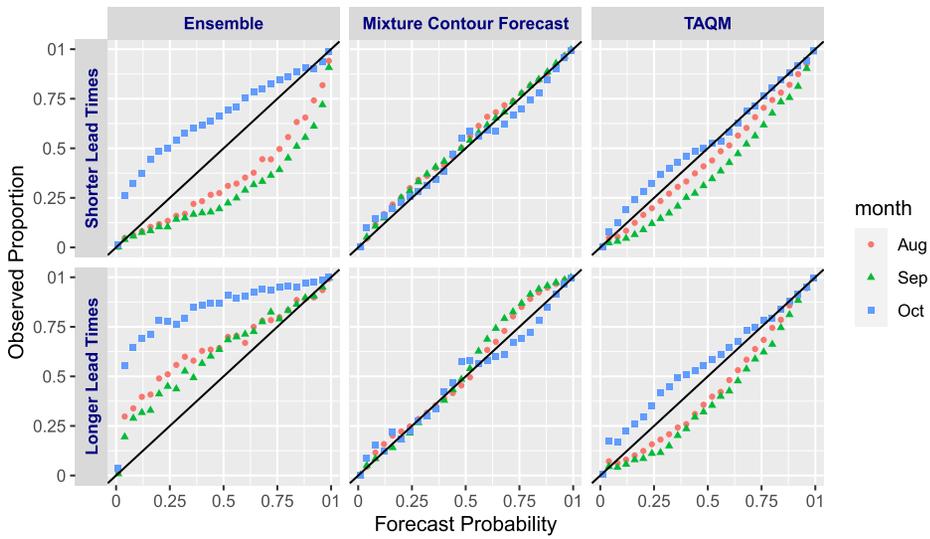


FIG. 5. Average proportion of the time sea ice was present against the predicted probability of sea ice presence for the ECMWF forecasts: unadjusted (left), after post-processing with MCF (middle) and TAQM (right). Forecasts are grouped into lead times of 0.5 and 1.5 months (top) and 2.5–6.5 months (bottom). Perfect calibration would give all points on the  $y = x$  line.

calibrated than the ensemble and better calibrated than TAQM during these months, especially at short lead times. Supplementary Material A Figures 5–6 show that MCF always improves calibration over the unadjusted ensemble and generally improves calibration compared to TAQM (Director, Raftery and Bitz (2021)). We note that MCF has been designed specifically for predicting sea ice presence, while TAQM addresses the more general goal of forecasting concentration. The difference in calibration performance between MCF and TAQM highlights the benefit for maritime planning of having a method focused exclusively on predicting sea ice presence. TAQM remains valuable for its broader applicability.

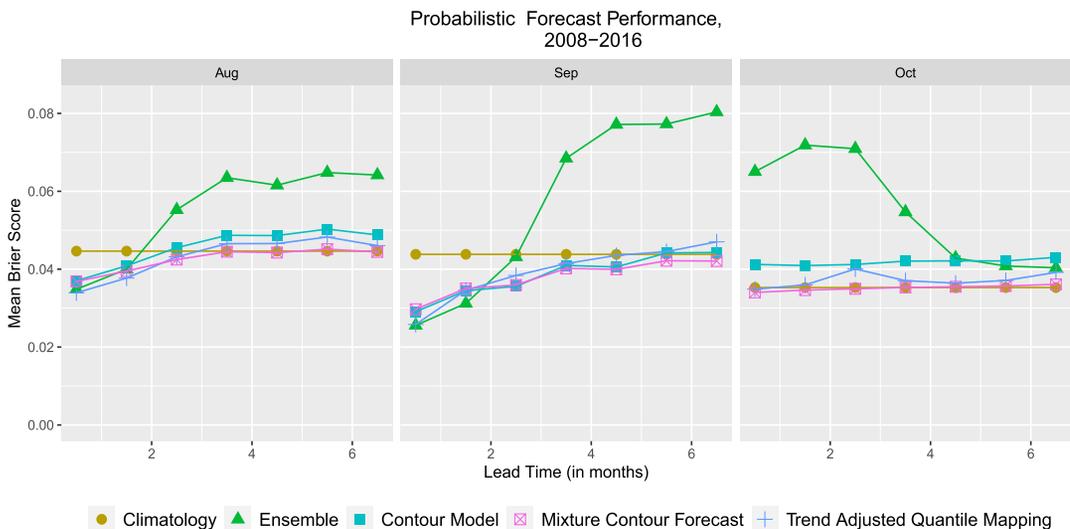


FIG. 6. Average Brier scores by month for the test years 2008–2016 for the probabilistic forecasts. The Brier Score for each grid box is weighted based on its area. Forecasts are described in Table 1.

4.5. *Assessing accuracy.* We evaluate forecast accuracy using Brier scores (Brier (1950)). We compute average area-weighted Brier scores over the  $T = 9$  years in the test set as

$$(4.1) \quad \frac{\sum_t \sum_s a_s (f_{s,t} - o_{s,t})^2}{T},$$

where  $f_{s,t}$  and  $o_{s,t}$  denote the forecast and observation in grid box  $s$  in year  $t$ . The value  $a_s$  is the proportion of the total area that is in grid box  $s$ . The observed value is 1 when concentration is at least 0.15 and 0 otherwise. For probabilistic forecasts,  $f_{i,j} \in [0, 1]$ , and for binary forecasts,  $f_{i,j} \in \{0, 1\}$ .

In Figure 6, we plot the average Brier score in peak-shipping months by lead time for the probabilistic forecasts. The ensemble forecasts typically have increasing Brier scores as lead time increases. Our contour model generally improves forecast accuracy, and MCF improves accuracy further. TAQM also generally improves accuracy of forecasts.

With MCF as lead time increases, performance converges to equal or better performance than climatology. Supplementary Material A Figure 7 shows the average weight placed on the contour model for the years in the test set (Director, Raftery and Bitz (2021)). High weights occur at short lead times, reflecting that the ensemble typically has the most skill shortly after it is initialized. High weights also occur in months around the sea ice minimum in September. These are periods of high year-to-year variability; so, climatology tends to perform poorly, and the ensemble's ability to simulate evolving physical conditions becomes more important.

Supplementary Material A Figures 1–2 show that TAQM and MCF have similar overall accuracy but that the pattern of their performance by lead time and month varies (Director, Raftery and Bitz (2021)). For peak shipping months, MCF outperforms TAQM. For the shortest lead time of 0.5 months, the damped persistence forecast performs best, but its skill decays rapidly with lead time. The performance of the damped persistence forecast indicates that there could be a role for the current observed state of the sea ice in forecasting but that the role would need to be restricted to very short lead times. In summary, MCF provides the best calibrated forecasts year round and accurate forecasts during peak-shipping months.

We also briefly assess binary forecasts (see Supplementary Material A Figures 3–4 (Director, Raftery and Bitz (2021))). The contour shifted forecast improves accuracy compared to the ensemble. This result indicates contour shifting reduces some systematic bias in typical ensembles. Binary MCF performs similarly to the contour shifted forecast. However, MCF substantially outperforms the contour shifted ensemble when the ensemble forecast is poor. This case illustrates that the adaptive weighting provided by MCF is valuable when issuing binary forecasts as well as probabilistic forecasts.

**5. Discussion.** We have introduced the MCF method for issuing sea ice forecasts. MCF forecasts are probabilistic and calibrated. Their predicted probability of sea ice presence at a given location approximately matches the proportion of times sea ice will be observed at these locations. For most lead times and forecast months, probabilistic MCF forecasts are also as or more accurate than the initial ECMWF ensemble and the other postprocessed and statistical forecasts.

Because MCF provides calibrated and relatively accurate forecasts, MCF could increase operational sea ice forecasting skill and, thereby, improve maritime planning in the Arctic. As Arctic routes are planned, the risk of encountering sea ice where it is not expected must be weighed against the cost savings of a shorter route. Vessels in the Arctic have an ice classification regulating where they can legally and safely travel. For vessels that are easily damaged in sea ice, any sea ice poses great risk. In contrast, ships that are designed to travel safely through sea ice may gain speed and efficiency by avoiding sea ice but do not need to completely avoid it.

Sea ice predictions can be misclassified in two ways: predicting the presence of sea ice when it is absent and predicting the absence of sea ice when it is present. Our model evaluation weights both types of misclassification errors equally. However, accounting for the different costs of the two types of error could be useful. MCF enables planning routes using decisions rules that incorporate the probability of sea ice presence. For example, a ship that has high risk of damage when traversing sea ice might elect to only consider routes through areas with very low probability of sea ice.

We have also developed a framework for directly modeling contours. While forecasts could likely be made with field-based geostatistical models (e.g., Zimmerman and Stein (2010)) or by identifying the exceedance level contours estimated from fields (Bolin and Lindgren (2015), French and Hoeting (2016)), these approaches may have limitations for this application. Most of the error in sea ice forecasts occurs in the region where a rapid transition from fully ice-covered regions to open water occurs (Tietsche et al. (2014)). Whether sea ice will be found in grid boxes in the interior of the sea ice region and far from the sea ice edge is essentially known in advance. So placing the majority of the computational cost and modeling effort on the boundary is valuable. MCF provides a framework for modeling that could be extended to other situations where the boundary is of interest.

As implemented in this paper, estimates of the covariance of the sea ice edge are based on the covariance estimated from the preceding years. These estimates are, therefore, independent of the covariance of the sea ice edge in the ensemble members. However, the ensemble could plausibly give information about the expected covariance that could not be obtained from past observations. For example, sea ice is expected to continue to become thinner. Thinner sea ice is more affected by variation in meteorological conditions, so the variance of sea ice extent will likely increase (Holland, Bailey and Vavrus (2011)). Effects like these are captured by the ensemble but are not in past observations. As such, incorporating the covariance in the ensemble could further improve forecast skill. However, the spread of the ensemble does not align with observed variability, and the relationship between variability in observations and variability in the ensemble is inconsistent both spatially and temporally. Assessment of when the ensemble covariance is informative and how it relates to the observed covariance is needed before it will be feasible to incorporate the ensemble covariance into MCF.

The ECMWF ensemble used in Section 4 is not the only ensemble prediction system. Since MCF does not use any specific features of the ECMWF ensemble, the postprocessing techniques developed in this paper could be directly applied to other ensembles. However, model biases and calibration issues vary, so exact performance would need to be assessed. Different ensembles also vary in which forecast months they perform well and vary more by the extent to which skill declines with lead time (Zampieri, Goessling and Jung (2018)). So, extending MCF to use multiple ensemble members, as has been done with other forecasting methods (e.g., Dirkson, Denis and Merryfield (2019), Raftery et al. (2005)), could provide further skill.

Our analysis of sea ice forecasting highlights situations where statistical postprocessing provides value. Many aspects of physical processes are known to evolve following established equations. Such information can only be crudely approximated with an observational data-driven approach. On the other hand, physical models are often biased or poorly calibrated, and statistical post-processing methods can be effective in remedying these problems. Combining the strengths of physical and statistical modeling can create predictions that are more accurate than either modeling framework alone.

**Acknowledgments.** We thank Nicholas Wayand for assistance with data processing and Arlan Dirkson for addressing questions on Trend-Adjusted Quantile Mapping. We also thank the Associate Editor and two anonymous reviewers for suggestions on the manuscript. Results were produced using the *IceCast* R package available in Supplementary Material B

(Director, Raftery and Bitz (2021)) or online <https://github.com/hdirector/IceCast>. The IceCast R package relies on the *rgeos* (Bivand and Rundel (2020)), *sp* (Pebesma and Bivand (2005)) and *Rcpp* (Eddelbuettel et al. (2011)) R packages, among others. Relevant code is also available in Supplementary Material B (Director, Raftery and Bitz (2021)) and online at <https://github.com/hdirector/ProbSeaIce>. The figures were primarily made with the *Tidyverse* R packages (Wickham et al. (2019)). Any opinions, findings and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

**Funding.** This work was supported by the National Oceanic and Atmospheric Administration’s Climate Program Office, Climate Variability and Predictability Program through grant NA15OAR4310161.

The first author was partially supported by the National Science Foundation Graduate Research Fellowship under Grant No. DGE-1256082.

## SUPPLEMENTARY MATERIAL

**Supplementary Material A: Additional information** (DOI: [10.1214/20-AOAS1405SUPPA](https://doi.org/10.1214/20-AOAS1405SUPPA); .pdf). Additional figures and MCF details.

**Supplementary Material B: IceCast R package and R scripts** (DOI: [10.1214/20-AOAS1405SUPPB](https://doi.org/10.1214/20-AOAS1405SUPPB); .zip). R package for bias reduction and probabilistic sea ice forecasting and R scripts for producing the figures and tables in this manuscript.

## REFERENCES

- BIVAND, R. and RUNDEL, C. (2020). *rgeos*: Interface to geometry engine—open source (‘GEOS’). R package version 0.5-5.
- BLANCHARD-WRIGGLESWORTH, E., CULLATHER, R. I., WANG, W., ZHANG, J. and BITZ, C. M. (2015). Model forecast skill and sensitivity to initial conditions in the seasonal sea ice outlook. *Geophys. Res. Lett.* **42** 8042–8048.
- BOLIN, D. and LINDGREN, F. (2015). Excursion and contour uncertainty regions for latent Gaussian models. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **77** 85–106. [MR3299400 https://doi.org/10.1111/rssb.12055](https://doi.org/10.1111/rssb.12055)
- BRIER, G. W. (1950). Verification of forecasts expressed in terms of probability. *Mon. Weather Rev.* **78** 1–3.
- BUSHUK, M., MSADEK, R., WINTON, M., VECCHI, G. A., GUDGEL, R., ROSATI, A. and YANG, X. (2017). Skillful regional prediction of Arctic sea ice on seasonal timescales. *Geophys. Res. Lett.* **44** 4953–4964.
- CAVALIERI, D. J. and PARKINSON, C. L. (2012). Arctic sea ice variability and trends, 1979–2010. *Cryosphere* **6** 881.
- CHEVALLIER, M., SALAS Y MÉLIA, D., VOLDOIRE, A., DÉQUÉ, M. and GARRIC, G. (2013). Seasonal forecasts of the pan-Arctic sea ice extent using a GCM-based seasonal prediction system. *J. Climate* **26** 6092–6104.
- COMISO, J. (2017). Bootstrap sea ice concentrations from Nimbus-7 SMMR and DMSP SSM/I-SSMIS. Version 3. NASA National Snow and Ice Data Center Distributed Active Archive Center, Boulder, Colorado, USA.
- COMISO, J. C., PARKINSON, C. L., GERSTEN, R. and STOCK, L. (2008). Accelerated decline in the Arctic sea ice cover. *Geophys. Res. Lett.* **35**.
- COPERNICUS CLIMATE CHANGE SERVICE (2019). Copernicus climate change service climate data store. Available at <https://cds.climate.copernicus.eu>.
- DEMPSTER, A. P., LAIRD, N. M. and RUBIN, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *J. Roy. Statist. Soc. Ser. B* **39** 1–38. With discussion. [MR0501537](https://doi.org/10.1111/rssb.12055)
- DIRECTOR, H. M., RAFTERY, A. E. and BITZ, C. M. (2017). Improved sea ice forecasting through spatiotemporal bias correction. *J. Climate* **30** 9493–9510.
- DIRECTOR, H. M., RAFTERY, A. E. and BITZ, C. M. (2021). Supplement to “Probabilistic forecasting of the Arctic sea ice edge with contour modeling.” <https://doi.org/10.1214/20-AOAS1405SUPPA>, <https://doi.org/10.1214/20-AOAS1405SUPPB>
- DIRKSON, A., DENIS, B. and MERRYFIELD, W. J. (2019). A multimodel approach for improving seasonal probabilistic forecasts of regional Arctic sea ice. *Geophys. Res. Lett.*

- DIRKSON, A., MERRYFIELD, W. J. and MONAHAN, A. H. (2019). Calibrated probabilistic forecasts of Arctic sea ice concentration. *J. Climate* **32** 1251–1271.
- DOUGLAS, D. H. and PEUCKER, T. K. (1973). Algorithms for the reduction of the number of points required to represent a digitized line or its caricature. *Cartographica* **10** 112–122.
- EDDELBUETTTEL, D., FRANÇOIS, R., ALLAIRE, J., USHEY, K., KOU, Q., RUSSEL, N., CHAMBERS, J. and BATES, D. (2011). rcpp: Seamless R and C++ integration. *J. Stat. Softw.* **40** 1–18.
- ELLIS, B. and BRIGHAM, L. (2009). Arctic marine shipping assessment 2009 report.
- EUROPEAN CENTRE FOR MEDIUM-RANGE WEATHER FORECASTS (2017). ECMWF SEAS5 user guide. Available at [https://www.ecmwf.int/sites/default/files/medialibrary/2017-10/System5\\_guide.pdf](https://www.ecmwf.int/sites/default/files/medialibrary/2017-10/System5_guide.pdf).
- FRENCH, J. P. and HOETING, J. A. (2016). Credible regions for exceedance sets of geostatistical data. *Environmetrics* **27** 4–14. MR3457233 <https://doi.org/10.1002/env.2371>
- GNEITING, T. (2013). Strictly and non-strictly positive definite functions on spheres. *Bernoulli* **19** 1327–1349. MR3102554 <https://doi.org/10.3150/12-BEJSP06>
- GNEITING, T., BALABDAOUI, F. and RAFTERY, A. E. (2007). Probabilistic forecasts, calibration and sharpness. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **69** 243–268. MR2325275 <https://doi.org/10.1111/j.1467-9868.2007.00587.x>
- GUEMAS, V., BLANCHARD-WRIGGLESWORTH, E., CHEVALLIER, M., DAY, J. J., DÉQUÉ, M., DOBLAS-REYES, F. J., FUČKAR, N. S., GERME, A., HAWKINS, E. et al. (2016). A review on Arctic sea-ice predictability and prediction on seasonal to decadal time-scales. *Q. J. R. Meteorol. Soc.* **142** 546–561.
- HOLLAND, M. M., BAILEY, D. A. and VAVRUS, S. (2011). Inherent sea ice predictability in the rapidly changing Arctic environment of the community climate system model, version 3. *Clim. Dyn.* **36** 1239–1253.
- HUBER, P. J. (2011). *Robust Statistics*. Springer, Berlin.
- JOHNSON, S. J., STOCKDALE, T. N., FERRANTI, L., BALMASEDA, M. A., MOLteni, F., MAGNUSSON, L., TIETSCHKE, S., DECREMER, D., WEISHEIMER, A., BALSAMO, G. et al. (2019). SEAS5: The new ECMWF seasonal forecast system. *Geosci. Model Dev.* **12**.
- MELIA, N., HAINES, K. and HAWKINS, E. (2016). Sea ice decline and 21st century trans-Arctic shipping routes. *Geophys. Res. Lett.* **43** 9720–9728.
- MSADEK, R., VECCHI, G. A., WINTON, M. and GUDGEL, R. G. (2014). Importance of initial conditions in seasonal predictions of Arctic sea ice extent. *Geophys. Res. Lett.* **41** 5208–5215.
- NATIONAL SNOW & ICE DATA CENTER (2017). Region mask for the Northern Hemisphere. Available at [http://nsidc.org/data/polar-stereo/tools\\_masks.html](http://nsidc.org/data/polar-stereo/tools_masks.html).
- PEBESMA, E. J. and BIVAND, R. S. (2005). Classes and methods for spatial data in R. *R News* **5** 9–13.
- RAFTERY, A. E., GNEITING, T., BALABDAOUI, F. and POLAKOWSKI, M. (2005). Using Bayesian model averaging to calibrate forecast ensembles. *Mon. Weather Rev.* **133** 1155–1174.
- SIGMOND, M., FYFE, J. C., FLATO, G. M., KHARIN, V. V. and MERRYFIELD, W. J. (2013). Seasonal forecast skill of Arctic sea ice area in a dynamical forecast system. *Geophys. Res. Lett.* **40** 529–534.
- SMITH, L. C. and STEPHENSON, S. R. (2013). New trans-Arctic shipping routes navigable by midcentury. *Proc. Natl. Acad. Sci. USA* **110** E1191–E1195.
- STROEVE, J. C., SERREZE, M. C., HOLLAND, M. M., KAY, J. E., MALANIK, J. and BARRETT, A. P. (2012). The Arctic’s rapidly shrinking sea ice cover: A research synthesis. *Clim. Change* **110** 1005–1027.
- TIETSCHKE, S., DAY, J. J., GUEMAS, V., HURLIN, W. J., KEELEY, S. P. E., MATEI, D., MSADEK, R., COLLINS, M. and HAWKINS, E. (2014). Seasonal to interannual Arctic sea ice predictability in current global climate models. *Geophys. Res. Lett.* **41** 1035–1043.
- WANG, W., CHEN, M. and KUMAR, A. (2013). Seasonal prediction of Arctic sea ice extent from a coupled dynamical forecast system. *Mon. Weather Rev.* **141** 1375–1394.
- WAYAND, N. E., BITZ, C. M. and BLANCHARD-WRIGGLESWORTH, E. (2019). A year-round subseasonal-to-seasonal sea ice prediction portal. *Geophys. Res. Lett.* **46** 3298–3307.
- WICKHAM, H., AVERICK, M., BRYAN, J., CHANG, W., D’AGOSTINO MCGOWAN, L., FRANÇOIS, R., GROLEMUND, G., HAYES, A., HENRY, L. et al. (2019). Welcome to the tidyverse. *J. Open Sour. Softw.* **4** 1686.
- ZAMPIERI, L., GOESSLING, H. F. and JUNG, T. (2018). Bright prospects for Arctic sea ice prediction on sub-seasonal time scales. *Geophys. Res. Lett.* **45** 9731–9738.
- ZHANG, B. and CRESSIE, N. (2019). Estimating spatial changes over time of Arctic sea ice using hidden  $2 \times 2$  tables. *J. Time Series Anal.* **40** 288–311. MR3946154 <https://doi.org/10.1111/jtsa.12425>
- ZHANG, B. and CRESSIE, N. (2020). Bayesian inference of spatio-temporal changes of Arctic sea ice. *Bayesian Anal.* **15** 605–631. MR4097811 <https://doi.org/10.1214/20-BA1209>
- ZHUANG, J. (2018). xESMF: Universal regridded for geospatial data.
- ZIMMERMAN, D. L. and STEIN, M. (2010). Classical geostatistical methods. In *Handbook of Spatial Statistics. Chapman & Hall/CRC Handb. Mod. Stat. Methods* 29–44. CRC Press, Boca Raton, FL. MR2730951 <https://doi.org/10.1201/9781420072884-c3>