Space-Time Contour Models for Sea Ice Forecasting

Hannah M. Director

A dissertation
submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington

2020

Reading Committee:

Adrian E. Raftery, Chair

Cecilia M. Bitz

Donald B. Percival

Program Authorized to Offer Degree:
Statistics

University of Washington

Abstract

Space-Time Contour Models for Sea Ice Forecasting

Hannah M. Director

Chair of the Supervisory Committee:
Adrian E. Raftery
Departments of Statistics and Sociology

This dissertation develops statistical methods for modeling contours. Particular emphasis is placed on forecasting the sea ice edge contour, or the boundary around ocean areas that are ice-covered. Current sea ice forecasts are largely based on dynamic ensembles. These physics-based prediction systems numerically solve differential equations to approximate possible evolutions of sea ice and its surrounding environment. While these dynamic ensemble forecasts provide information about future sea ice, they have systematic differences from observations and incorrect variability. I develop two methods to improve forecasts of the sea ice edge contour in the Arctic. I first introduce Contour-Shifting, a statistical method to anticipate and correct systematic errors in forecasts of the sea ice edge contour produced by dynamic ensembles. I then propose Mixture Contour Forecasting, a method to generate sea ice edge contours that have variability similar to observations. These generated contours can be used to predict the probability of sea ice presence weeks-to-months in advance. Both Contour-Shifting and Mixture Contour Forecasting represent the sea ice edge contour directly as a connected sequence of points. I extend this approach of modeling the points on contours directly with the development of the Gaussian Star-Shaped Contour Model. This model can be employed for inference and prediction of contours that enclose star-shaped polygons or approximately star-shaped polygons. Approaches for fitting this model and assessment metrics for contours are also introduced.

TABLE OF CONTENTS

LIST OF FIGURES

Figure Number Page

GLOSSARY

ECMWF: European Centre for Medium-Range Weather Forecasts

FLOR: Forecast Oriented Low-Ocean Resolution

GFDL: Geophysical Fluid Dynamics Laboratory

GSCM: Gaussian Star-Shaped Contour Model

IIEE: Integrated Ice-Edge Error

MCF: Mixture Contour Forecasting

MCMC: Markov Chain Monte Carlo

NASA: National Aeronautics and Space Administration

NOAA: National Oceanic and Atmospheric Administration

RMSE: Root Mean Squared Error

S2S: Subseasonal-To-Seasonal

TAQM: Trend Adjusted Quantile Mapping

ACKNOWLEDGMENTS

First and foremost, I would like to thank my research advisors Adrian Raftery and Cecilia Bitz who have guided and encouraged me throughout graduate school. I am grateful to Adrian for teaching me how to translate statistical ideas into scientific impact and how to write clearly with the reader in mind. I am grateful to Cecilia for helping me identify valuable scientific problems and for modeling compassionate leadership in science.

Thank you also to my committee members, Donald Percival and LuAnne Thompson who have both provided insightful feedback on my research. I am grateful for the staff of the University of Washington Department of Statistics whose constant efforts enable the research and teaching in the Department of Statistics. Thank you to the many faculty in the Departments of Statistics, Biostatistics, and Atmospherics Sciences who have inspired me in their classes and provided mentorship and guidance. In particular, I would like to thank Peter Guttorp, Emilijia Perkovic, Paul Sampson, Jon Wakefield, Amy Willis, and Daniela Witten.

I am fortunate to have collaborated with a number of talented students and postdoctoral fellows during my PhD. I would particularly like to thank those who I have worked with most closely: Michael Diamond, Peter Gao, Joyce Zhang, and Yongfei Zhang. I am also grateful for the many students and postdoctoral fellows I have gotten to know during graduate school. I am happy to now count many of you as friends.

Thank you to June Morita for first encouraging me to consider graduate school and a career in research. Thanks you also to Kristin Laidre, Eliezer Gurarie, Luke Bornn, and James Gattiker for their mentorship of my research prior to graduate school.

I am also grateful for the financial support that enabled this research. The material in

this dissertation was supported by the National Oceanic and Atmospheric Administration's Climate Program Office, Climate Variability and Predictability Program through Grant NA15OAR4310161 and the National Science Foundation Graduate Research Fellowship under Grant No. DGE-1256082. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect the views of the National Science Foundation.

Finally, thank you to my parents, brothers, and partner for your constant support and encouragement. I am so very lucky to have all of you in my life.

Chapter 1

INTRODUCTION

1.1 Motivation

In the Arctic temperatures are cold enough that the top layer of the ocean freezes. Exactly where this frozen ocean water, known as sea ice, forms varies from year-to-year. The area covered by sea ice increases and decreases annually in a seasonal oscillation. Climate change has further complicated the patterns of growth and melt of sea ice. In recent decades, the total amount of ice-covered area in the Arctic has trended downward from one year to the next [Stroeve et al., 2012b, Comiso et al., 2008]. Predicting sea ice is therefore a multi-faceted challenge requiring consideration of spatial, seasonal, and temporal patterns.

Despite its difficulty to predict, forecasts for sea ice are increasingly needed. Reduced sea ice cover has made the Arctic a more appealing route for commercial shipping and tourism, resulting in increased vessel traffic [Smith and Stephenson, 2013, Melia et al., 2016]. Vessels in the Arctic often seek to avoid sea ice, since sea ice is costly and time-consuming to traverse. If reliable forecasts of sea ice were available weeks-to-months in advance, maritime routes could be planned that avoid sea ice. Of particular importance are forecasts of the location of the ice edge contour, or the boundary around the ice-covered region. Accurate assessment of the uncertainty associated with these ice edge forecasts is crucial for allowing for the assessment of risks and benefits of possible routes.

Sea ice forecasts are primarily needed in the months around the annual sea ice minimum, the time of year when the lowest proportion of the Arctic ocean's area is covered by sea ice. Vessel traffic is typically highest during these summer and early fall months, because the low sea ice cover makes navigating the Arctic easier. However, this period of time is also the most difficult to predict accurately. Large portions of the sea ice edge are far from land boundaries, resulting in minimal constraints of where the ice edge could be located. As such year-to-year variability is high and climatology, or average behavior of the sea ice in the (recent) past, does not provide a sufficiently accurate forecast of where the sea ice edge will be located in the future.

Motivated by this growing need to forecast the sea ice edge contour well, this dissertation develops two related methods to improve accuracy and calibration of sea ice forecasts. Chapters 2 and 3 introduce two methods for improved sea ice forecasts. Both methods combine statistical methods with outputs of physics-based models to issue better forecasts. In Chapter 2, I introduce Contour-Shifting, a method to anticipate and correct systematic errors in the location of the ice edge predicted by dynamic models. This material closely follows an article published in *Journal of Climate* [Director et al., 2017a]. In Chapter, 3, I develop Mixture Contour Forecasting (MCF). MCF directly models the distribution of the ice edge contour using a combination of information from historical observations and data from dynamic models. Both Contour-Shifting and MCF can be implemented with the *IceCast* R package [Director et al., 2017b, 2020]. Both methods model the sea ice edge as a connected sequence of spatial points.

The statistical challenges posed by modeling ice edge contours motivate developing more generalize techniques for modeling contours. Chapter 4 expands on the idea of direct contour modeling explored in Chapters 2 and 3. In particular, this chapter introduces a general purpose probabilistic model for contours that enclose polygons that are star-shaped or approximately star-shaped. As in the ice edge model, contours are represented as connected sequences of points. Techniques for fitting this probabilistic contour model as well as metrics to assess performance of contour models are discussed. The method proposed in this chapter provides a low-dimensional alternative for statistical modeling of boundaries compared to more typical field-based methods. Software for fitting this model and assessing credible intervals of contours is available via the *ContouR* R package [Director and Raftery, 2020].

1.2 Background

This section introduces general background on seasonal-to-subseasonal forecasting, ensemble dynamic models, statistical post-processing, and contours. More detailed background information is given as relevant within Chapters 2-4.

1.2.1 Subseasonal-To-Seasonal forecasts

In atmospheric sciences and related fields, distinctions are made between climate and weather. Climate refers to average meteorological conditions while weather refers to variability around this average. How far in advance weather can be predicted is limited by a phenomena known as chaos. In a non-linear system very slight changes in the initial state result in large changes in future states [Lorenz, 1963]. The Earth system is partially governed by non-linear behavior, so chaos limits how far in advance predictions of weather can be made. Climate predictions depend on initial conditions as well, but other factors like changes in greenhouse gas level or aerosols dominate as forecast lead times grow.

The focus of this dissertation is on forecasting sea ice weeks to months in advance. Meteorological forecasts made for these lead times are often referred to as Subseasonal-To-Seasonal (S2S) forecasts. Forecasts at these time scales are considered unusually difficult because they encompass challenges faced in both weather and climate prediction [Doblas-Reyes et al., 2013]. Like in weather forecasting, S2S forecasts must be correctly initialized with the conditions at the time the forecast is issued. Additionally, feedbacks that occur at short and moderate time scales must also be considered.

1.2.2 Ensemble dynamic models and statistical post-processing

Dynamic models are the state-of-the-art tool used for weather forecasting and climate modeling. These models numerically solve systems of differential equations to represent how the components of the Earth system evolve together. Dynamic models are typically run multiple times with different initial conditions to produce sets of forecasts, referred to as ensembles.

Because of chaos, the ensemble members produce varied forecasts. Ensembles, thus, provide some notion of the spread of possible outcomes for a given forecast. However, to sample from the full distribution can require many ensemble members, which is computationally expensive.

While ensemble dynamic models have been invaluable in understanding and predicting Earth system behavior, these models typically have biases and/or do not have variability that matches observations. These weaknesses of dynamic models arise from imperfections in obtaining or representing initial conditions [Collins, 2002, Hazeleger et al., 2013], unknown model physics, and inaccurate parameterizations (statistical approximations of processes that occur below the spatial and/or temporal scale of the model) [Meehl et al., 2014].

Statistical post-processing, or statistical methods that correct the output of dynamic models, have emerged as a way to reduce the impacts of these weaknesses of dynamic models. These methods expanded the Model Output Statistics (MOS) approach. MOS essentially used multiple linear regression to understand and anticipate differences between dynamic ensemble forecasts and observations [Glahn and Lowry, 1972]. Statistical post-processing methods have now been successfully applied to dynamic model outputs of a number of meteorological variables. Some common techniques including perturbing model output via the Geostatistical Output Perturbation (GOP) Method [Gel et al., 2004, Sloughter et al., 2007, 2013], averaging multiple ensemble predictions via Bayesian Model Averaging (BMA) [Raftery et al., 2005, Bao et al., 2010], and a local version of BMA referred to a Geostatistical Model Averaging [Kleiber et al., 2011a,b]. Quantile mapping, which transforms ensemble output based on quantiles, is also very common, particularly for climate projections [e.g., Maraun, 2016, Cannon, 2016]. Research on post-processing sea ice model output in particular is more limited [Krikken et al., 2016, Dirkson et al., 2019b,a]. Post-processing methods have typically focused on improving the sharpness and calibration of ensemble forecasts. In this context, sharpness refers to how concentrated the ensemble members are on a particular forecast. Calibration refers to how closely the distribution of the ensemble members aligns with the distribution of observations [Gneiting et al., 2007].

1.2.3 Contour models

In this dissertation, I will focus on inferring the distributions of contours based on multiple observations of contours defined by sequences of points. While this specific problem has attracted surprisingly little attention, modeling contours and boundaries more generally has been considered in statistics and related fields. Several lines of research address how to infer a boundary line from a derived spatial field. Substantial methods development has focused on inferring the boundary around the region where a spatial process exceeds a fixed level [e.g. Bolin and Lindgren, 2015, French and Sain, 2013, French and Hoeting, 2016]). Other research has employed interpolation and gradients to identify boundaries for processes on spatial fields [e.g. Womble, 1951, Jacquez et al., 2000, Banerjee and Gelfand, 2006]). Modeling defined shapes or objects, such as in medical imaging, has also been explored extensively [e.g. Dryden and Mardia, 2016, Grenander and Miller, 1998].

1.3 Data and model output

This section previews the primary observational data sources and model output employed in this dissertation.

1.3.1 Observational data

Because of its importance for the Arctic region and global climate, monitoring systems have been put in place to record sea ice's behavior. Observations of sea ice primarily focus on two variables: concentration and thickness. Concentration refers to the proportion of surface area within a defined region that contains sea ice. Thickness refers to the depth of the sea ice. Remote sensing products typically report these quantities on a grid, although some in situ data are available. This dissertation uses concentration data only, since I am focused on improving forecasts of where sea ice is located. Throughout this dissertation, I will follow the convention of categorizing a grid box as containing sea ice if the grid box has concentration of at least 15%. This thresholding is necessary because of potentially low accuracy of satellite products. Satellites cannot distinguish between open water and water melting on the sea ice's surface, referred to as melt ponds. The concentration tends to decline from about 50% to near 0% over a small region, so classifying sea ice extent depends only weakly on the precise threshold value.

The primary source of observational data considered in this dissertation is the National Aeronautics and Space Administration (NASA) Bootstrap sea ice data product [Comiso, 2000, updated 2017]. This observational data product is produced from data obtained from the Nimbus-7 SMMR and DMSP SSM/I-SSMIS passive microwave satellites. The name refers to the Bootstrap algorithm that derives sea ice concentration from sea ice emissivity using radiative transfer equations [Comiso, 1986, Comiso and Sullivan, 1986]. The bootstrap sea ice data report monthly estimates of sea ice concentration. Data are available from 1979 to present, although several months are missing. The data are reported on a Polar Stereographic grid with nominal 25 kilometer square grid boxes. We used the most recent version of the data product available when our analyses were being completed. Chapter 2 uses Version 2 [Comiso, 2000, updated 2015] and Chapter 3 uses Version 3.1 [Comiso, 2000, updated 2017]

1.3.2 Dynamic model output

I applied the post-processing methods developed in this dissertation to two dynamic sea ice models. In Chapter 2, I use the Geophysical Fluid Dynamics (GFDL) CM2.5 Forecast oriented Low-Ocean Resolution (FLOR) model produced by the National Oceanic and Atmospheric Administration (NOAA) [Vecchi et al., 2014, Msadek et al., 2014]. In Chapter 3, I use the 25-member European Centre for Medium-Range Weather Forecasts (ECMWF) model [available from Copernicus Climate Change Service, 2019]. The switch in dynamic ensembles for Chapter 3 was motivated by analyses released after the publication of Chapter 2. This research showed that the ECMWF model was the most accurate among a set of publicly-available sea ice models [Zampieri et al., 2018].

1.4 Key contributions

In this dissertation I develop statistical methods that enable more accurate and better calibrated forecasts of Arctic sea ice. I also contribute to statistical methods for modeling contours and boundaries. My key contributions can be summarized as follows:

1. I developed the Contour-Shifting method that corrects systematic errors in forecasts of the sea ice edge contour produced by dynamic ensembles [Director et al., 2017a].

2. I developed the Mixture Contour Forecasting method that produces better calibrated forecasts of where sea ice will be located weeks to months in advance. Forecasts combine statistical modeling of the ice edge with dynamic ensemble output.

3. I produced the *IceCast* R package that provides software and tutorials for implementing Contour-Shifting and Mixture Contour Forecasting [Director et al., 2017b, 2020].

4. I developed the Gaussian Star-Shaped Contour Model (GSCM), a general purpose probability model for modeling contours as a connected sequence of points. The method is designed to be applied to contours that enclose regions that are star-shaped polygons or approximately star-shaped polygons.

5. I introduced a new metric to assess coverage of credible intervals for contours that enclose regions that are star-shaped polygons or approximately star-shaped polygons.

6. I produced the *ContouR* R package for modeling and assessing contours [Director and Raftery, 2020].

1.5 Dissertation organization

The remainder of this dissertation is organized as follows. Chapters 2 and 3 focus directly on forecasting the sea ice edge. These chapters develop two related statistical methods to improve forecasts of the Arctic sea ice edge at subseasonal and seasonal time scales. Chapter 2 introduces Contour-Shifting, a method to correct biases in the location of the ice edge forecast from dynamic models. Chapter 3 develops Mixture Contour Forecasting, a method for generating distributions of sea ice edge contours that accurately represent the frequency at which different ice edges occur. Chapter 4 pivots to the development of a general statistical models for contours. In this chapter, methods for modeling contours that enclose star-shaped and approximately star-shaped polygons are developed. Metrics for assessment of probabilistic contour models are also introduced. Chapter 5 summarizes this dissertation and identifies directions for future research.

Chapter 2

IMPROVED SEA ICE FORECASTING THROUGH SPATIOTEMPORAL BIAS CORRECTION

In this chapter, we propose a new method, called Contour-Shifting, for correcting the bias in forecasts of contours such as sea ice concentration above certain thresholds. Retrospective comparisons of observations and dynamical model forecasts are used to build a statistical spatiotemporal model of how predicted contours typically differ from observed contours. Forecasted contours from a dynamical model are then adjusted to correct for expected errors in their location. The statistical model changes over time to reflect the changing error patterns that result from reducing sea ice cover in the satellite era in both models and observations. For an evaluation period from 2001-2013, these bias-corrected forecasts are on average more accurate than the unadjusted dynamical model forecasts for all forecast months in the year at four different lead times. The total area which is incorrectly categorized as containing sea ice or not is reduced by 3.3×10^5 km^2 (or 21.3%) on average. The root mean squared error of forecasts of total sea ice area is also reduced for all lead times.

This chapter is closely based on the article "Improved Sea Ice Forecasting Through Spatiotemporal Bias Correction" published in the *Journal of Climate* [Director et al., 2017a].

2.1 Introduction

Dynamical models of the climate have considerable predictive capability. However, these predictions can differ systematically from observations. Models drift from their initial conditions and toward their internal mean state due to errors in their model physics [Meehl et al., 2014]. Imperfections in obtaining or representing initial conditions also factor into these systematic differences [Collins, 2002, Hazeleger et al., 2013]. Bias correction has emerged as a way to correct for these errors. This class of methods develops statistical representations of the error patterns in climate models using retrospective comparisons of observations and model output. These statistical representations are then used to correct for the expected error in predictions obtained from dynamical model forecasts [Maraun, 2016, Meehl et al., 2014].

Arctic sea ice cover has decreased substantially in recent years, causing increased interest in predicting it [Comiso et al., 2008, Stroeve et al., 2012b]. Dynamical models forecast quantities such as sea ice concentration, thickness, and age on a spatial grid. "Perfect model" experiments, which predict results within a climate model rather than on observed variables, have been used to evaluate the predictability of sea ice. These methods, which avoid effects from errors in model physics or initial conditions, suggest that skillful forecasts could be obtained on at least seasonal time scales [Blanchard-Wrigglesworth et al., 2011b, Day et al., 2014, Tietsche et al., 2014] with significant skill for 1-2 years [Guemas et al., 2016]. However, current dynamical forecasts only show significant skill for 3-5 month lead times [Guemas et al., 2016]. In the September Sea Ice Outlook, where predictions for September sea ice are annually compiled and compared, Stroeve et al. [2014] found prediction skill to be only marginally better than the skill obtained just from estimating the linear trend. More recently, Blanchard-Wrigglesworth et al. [2015] found that dynamical models are also unskilled at predicting each other, suggesting varied errors in model physics and/or initial conditions across prediction systems. Current forecasts are notably affected by which (if any) bias correction technique is applied [Blanchard-Wrigglesworth et al., 2016, Fučkar et al., 2014]. This suggests that improved bias correction could reduce the gap between the hypothetical predictability of sea ice and the skill currently being obtained.

A range of techniques have been proposed for bias correcting forecasts from dynamical models. The simplest approach, typically referred to as mean bias adjustment, corrects forecasts based on the difference in the anomalies of predictions and observations [Meehl et al., 2014], sometimes computing different climatological means for each time period under consideration [García-Serrano and Doblas-Reyes, 2012]. Additional approaches have been introduced to account for “conditional bias”, or when the bias changes as a function of some other factor [Goddard et al., 2013]. These include techniques for correcting bias that vary with forecast time [Kharin et al., 2012], lead time [van Oldenborgh et al., 2012], and initial conditions [Fučkar et al., 2014], among others. While these methods are an improvement over mean bias adjustment, as Goddard et al. [2013] note, nonlinear relationships between the model drift and forecasted value, initialization shocks, and imperfect model responses to forcings all remain challenges for designing appropriate bias correction methods. Extended logistic regression [Wilks, 2009] and heteroskedastic extended logistic regression [Messner et al., 2014] have also been used to jointly bias correct and calibrate forecasts of sea ice extent [Krikken et al., 2016]. While adjustments are typically applied to a summary measure or to each grid box independently, explicit spatial models of bias have also recently been suggested [Arisido et al., 2017].

For sea ice in particular, bias correction has focused on correcting summary measures, such as the total sea ice area or extent [Fučkar et al., 2014, Krikken et al., 2016]. However, many current stakeholders require information about the spatial pattern of sea ice. For example, with the growth of commercial vessel traffic in some regions of the Arctic, there is increased interest in identifying safe routes for ships to follow [Huntington et al., 2015, Smith and Stephenson, 2013, Stephenson et al., 2014].

Forecasters who produce spatial maps of sea ice currently do not have adequate approaches to directly bias correct the predictions they receive from dynamical models. They must either ignore the known bias in the model output, or make ad-hoc adjustments to where the sea ice is located based only on whether bias correction methods for the total sea ice area indicate too little or too much sea ice. This latter approach is problematic, since knowing that the total sea ice area is over- or underestimated does not guarantee that any particular location will see more or less sea ice than predicted from the dynamic model. This means that it is unclear where additional sea ice should be added to the model prediction or where existing ice should be removed. Recognizing these limitations of current methods, we propose a new bias correction technique that corrects spatial error in sea ice models. This enables the production of more accurate spatial predictions than uncorrected dynamical model output.

We focus our bias-correction efforts on the marginal ice zone. This is where high concentrations of ice change over to open water [Strong, 2012] and where the highest amount of error in forecasting occurs [Tietsche et al., 2014]. Specifically, we seek to improve the positioning of the contour surrounding contiguous grid boxes with at least 15% ice concentration, a line we refer to as the sea ice contour. The general idea of our bias correction approach is to record how far and in what direction the observed and predicted contours extend from various fixed locations. Using historical data, we can then model the difference between the observation and retrospective prediction at each location and how the difference is changing over time. This provides a bias correction that can be applied to move a new predicted contour to match an observed contour more closely. We refer to this method as Contour-Shifting, since it captures this notion of moving the contour obtained from a dynamical model prediction to a new, more accurate location.

To illustrate how Contour-Shifting can make a prediction more accurate, we compare in Figure 2.1 a predicted contour obtained from a dynamical model and its bias-corrected version. Due to some irreducible forecast error, the bias-corrected contour is not expected to lie exactly on the observed region. Here, as in the remainder of this chapter, the predictions are from the CM2.5 Forecast oriented Low-Ocean Resolution (FLOR) model produced by the National Oceanic and Atmospheric Administration's Geophysical Fluid Dynamics Laboratory [Vecchi et al., 2014, Msadek et al., 2014]. The observations are from the NASA Bootstrap monthly sea ice concentrations obtained from the Nimbus-7 SMMR and DMSP SSM/I-SSMIS passive microwave satellites [Comiso, 2000, updated 2015].

The potential for increased skill from spatial bias correction becomes apparent when comparing where sea ice was predicted to where sea ice was observed in historical data. In Figure 2.2, we map the regions of disagreement between forecasts and observations for February, September, and December Arctic sea ice in three successive years. In examining the regions of disagreement between observations and predictions for particular forecast months, we see some regions and months, such as Hudson Bay in December, where sea ice is

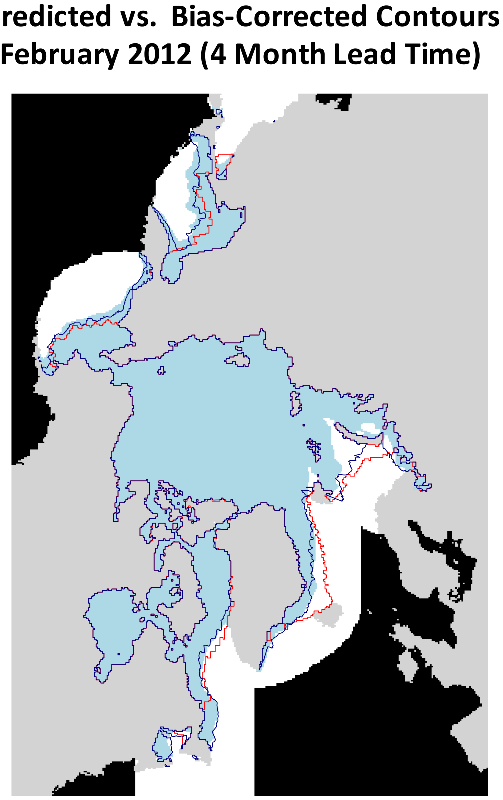


Figure 2.1: Predictions for the 15% sea ice concentration contour obtained from the GFDL model with a 4-month lead time (red) and bias-corrected using Contour-Shifting (navy blue) for February 2012. The observed sea ice region is plotted in light blue. The black regions are part of the "non-regional ocean" in the ocean mask described in Section 2.2.1 and are excluded in this analysis. The remaining white area represents grid boxes with sea ice concentration lower than 15%. Grey areas are land.

repeatedly observed, but never predicted. Similarly, we observe some regions where sea ice is repeatedly predicted, but not observed, such as the Bering Sea in February. This suggests we could obtain more accurate forecasts by correcting persistent errors of this type.

The chapter is organized as follows. Section 2.2 describes the implementation of Contour-Shifting. We discuss how to statistically model the spatial pattern of error using historical comparisons of predictions and observations and how this can be used to bias correct predictions from dynamical models. In Section 2.3 we evaluate the reduction in error this technique provides for predicting both the spatial pattern of the sea ice and the total sea ice area. Section 2.4 concludes the chapter with discussion.

2.2 Contour-Shifting method

2.2.1 Mapping ice sections

Contour-Shifting corrects the bias in predictions of the sea ice contour by relating where the sea ice edge (i.e. the 15% concentration contour) is predicted in a dynamical model to where the sea ice edge is observed. Thus to implement Contour-Shifting, we must develop a way to consistently represent the observed and predicted edges of sea ice. To do this, we compare how the predicted and observed sea ice edges relate to fixed reference locations, and how their difference changes over time.

Although sea ice is often a broken jumble of floes, most floes accumulate in contiguous sections touching land. This makes it natural to consider how far sea ice extends off land and in what direction. We develop a mapping system accordingly. Figure 2.3 is an example of how we can map the boundary of predicted or observed regions of sea ice from land. Specifically, we designate *fixed lines* generally along many of the borders between land and ocean from which we can record how far sea ice extends. (Small sections of the fixed lines may need to be on region boundaries instead as will be discussed subsequently.) We represent these connected lines as sequences of spatial points, $\boldsymbol{s}_r = (s_{r,1}, s_{r,2}, ..., s_{r,n_r})$, where r denotes the region of the Arctic in which the fixed line is located. (There are thus $n_r - 1$ line segments

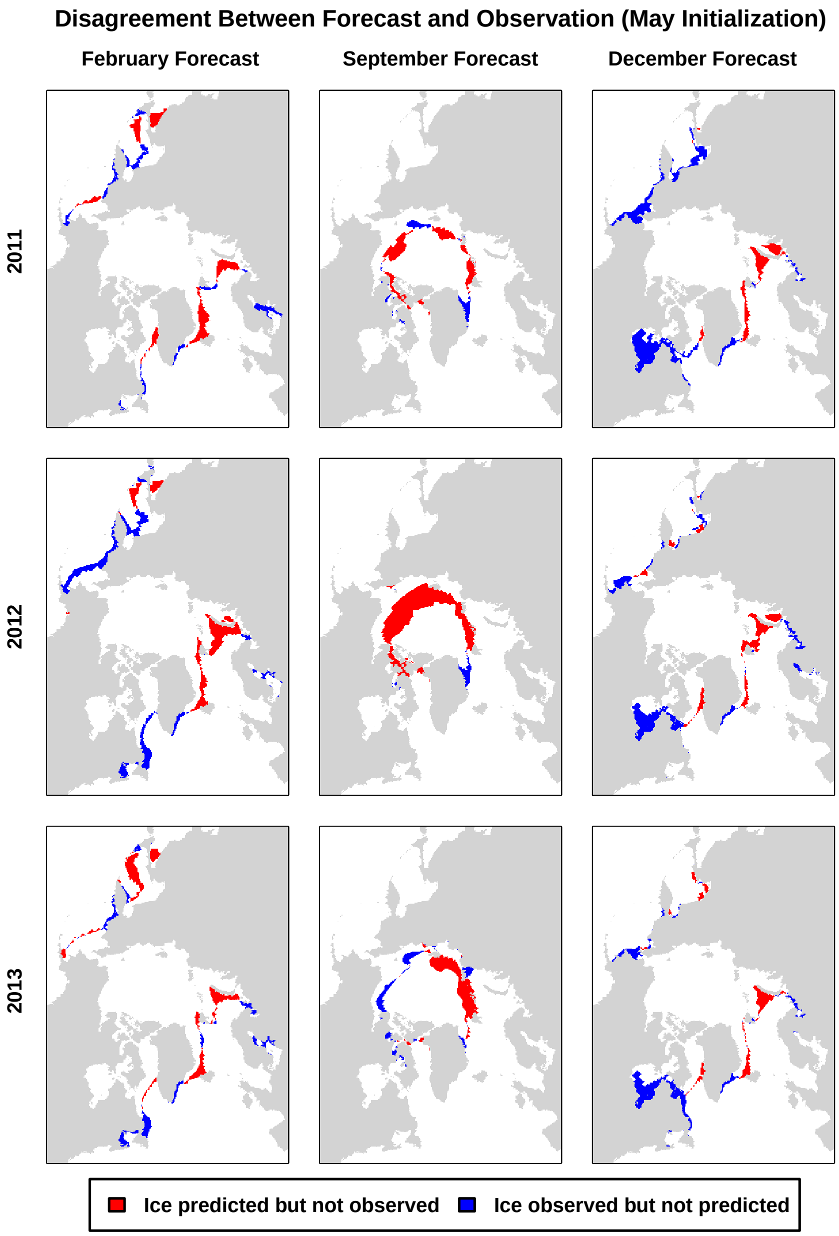


Figure 2.2: Maps indicating where the prediction of the presence of sea ice, defined as grid boxes with greater than or equal to 15% concentration, differed from observations for February, September, and December from 2011-2013. Regions where sea ice was predicted but not observed are in red and regions where sea ice was observed, but not predicted are in blue. Predictions are from May initializations. For example, the prediction in February 2011 was initialized in May 2010, while the prediction for September and December 2011 was initialized in May 2011.

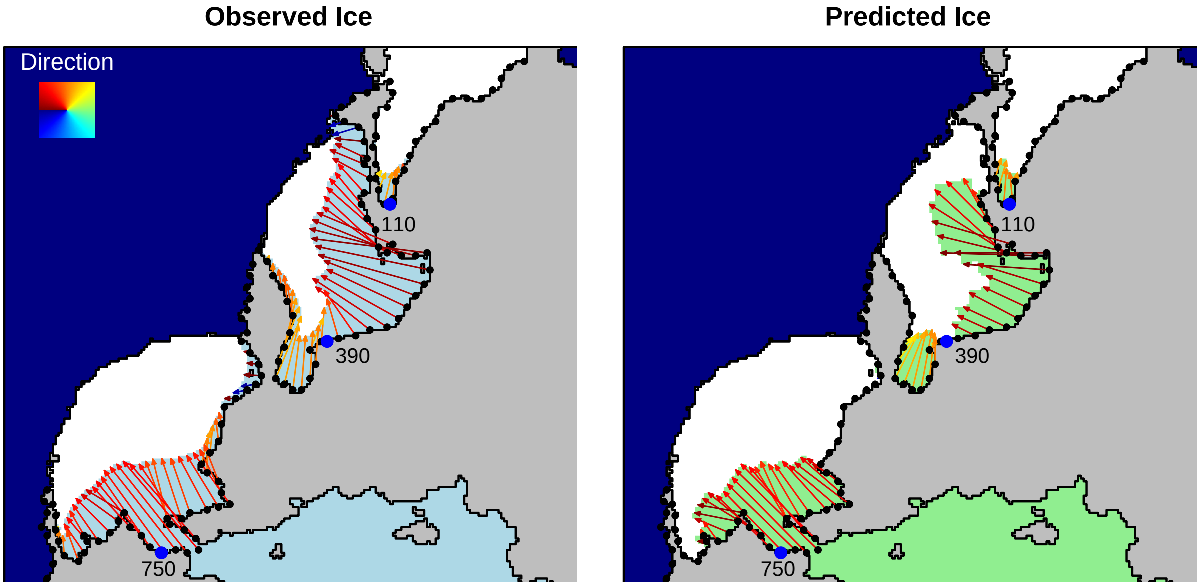


Figure 2.3: Examples of mapping vectors for the Bering Sea and Sea of Okhostk in an observation (left) and prediction (right) for February 2008. The colored areas denote grid boxes with greater than or equal to 15% sea ice concentration observed (light blue) and predicted (green). The black points indicate the fixed line on land from which the points are mapped and the color of the arrows indicate their direction. The arrowheads points to the mapped line. Only every seventh point and line is displayed for ease of visualization. The prediction was initialized in November 2007. The corresponding time series for the three blue numbered points are given in Figure 2.6.

in the r-th region, with the j-th line segment being specified by the spatial points $s_{r,j}$ and $s_{r,j+1}$.) Multiple regions and corresponding fixed lines are used in this framework to obtain more detailed representations of the location of sea ice than could be obtained with a single region. Spatial points are placed on the boundary line between the ocean and the land everywhere the line intersects a corner of a grid box. In Figure 2.3, these points are plotted as black circles. When connected in order, the points $\boldsymbol{s}_r$ trace the boundary between land and ocean. Henceforth, to simplify notation, we suppress the region subscript.

To define how sea ice extends off a fixed line in a particular region, we consider each contiguous section of sea ice that borders the fixed line. For each section, we identify all the points on the boundary of the section that intersect the fixed line. If there are k such points, we space k points evenly along the rest of the section boundary, which we denote by $\boldsymbol{s'} = (s'_1, s'_2, ..., s'_k)$ and refer to this as the *mapped line.* Specifically, we divide the section boundary line into $k-1$ equal-length increments and place a point at the k divisions between these increments. We then pair up the points on the *mapped line* with the points on the *fixed line* sequentially, and draw a vector connecting each pair. These two-dimensional vectors, which we denote by $\boldsymbol{m} = (m_1, m_2, ..., m_k)$, provide a "map" of the sea ice extending from the land border, and we refer to them as *mapping vectors.* They are plotted in Figure 2.3 colored by direction. This technique does not constrain vectors on corresponding predictions and observations to go in the same direction.

A given point i on the mapped line can be represented as

$$s'_i = s_i + m_i. \tag{2.1}$$

Sequentially connecting the points on $\boldsymbol{s}$ and then on $\boldsymbol{s'}$ provides a close approximation to the original boundaries of the sea ice sections. Regions and their associated fixed lines are selected to reflect where typical breaks in the sea ice occur and other physical boundaries. In Figure 2.4, we show the $R = 12$ regions outside the center of the Arctic and their fixed lines used in this analysis. These regions were obtained by modifying a region mask described in Parkinson and Cavalieri [2008] and available from the National Snow and Ice Data Center

[National Snow and Ice Data Center, 2017]. In regions that are nearly or completely enclosed by land, we have designated the fixed line to be some portion of the land boundary. If there is a portion of the land boundary where the sea ice clearly more commonly abuts, we use this portion. Otherwise, this selection is arbitrary. The fixed line includes a small portion of the region boundary in addition to points on land in two regions. These regions have blue and navy blue fixed lines in Figure 2.4. This is done to maintain a continuous fixed line where sea ice typically abuts while respecting region boundaries.

Compared to other regions, the sea ice edge in the center of the Arctic can follow more varied paths. We therefore need a modified definition of how to map this region. Our approach is illustrated with a sample map in Figure 2.5. We first designate an additional region for the center of the Arctic, indexed as $R + 1$. Then, we create an artificial set of reference lines and record where the sea ice intersects them. Specifically, we define an arbitrary point c near the middle of this Central Arctic region and then extend lines from c to the n_{R+1} points on the region boundary that intersect the lattice of the spatial grid. In Figure 2.5, c is indicated with a black point. We record the farthest point from c within the boundary of the sea ice in the region that intersects each line i as $s'_{R+1,i}$. We let $s_{R+1,i} = c$ for all i. As for other regions, we can then build a mapping vector, $m_{R+1,i}$ to connect $s_{R+1,i}$ and $s'_{R+1,i}$ and write

$$s'_{R+1,i} = s_{R+1,i} + m_{R+1,i}. \tag{2.2}$$

The resulting mapping vectors are shown in Figure 2.5 colored by their direction. Unlike in other regions, the angles of the vectors in the Central Arctic are fixed and the same for observations and predictions. Connecting the points of $s_{R+1,i}$ in order approximately maps out the edge of the sea ice.

Completing this process for all $R + 1$ regions for a single observation or prediction gives $R + 1$ sets of coordinates of fixed lines and associated mapping vectors. The vectors will vary for different observations and predictions, but the coordinates will be the same for any observation or prediction. This means we can develop a detailed spatial picture of how a predicted sea ice edge compares to an observed sea ice edge. We need only evaluate the

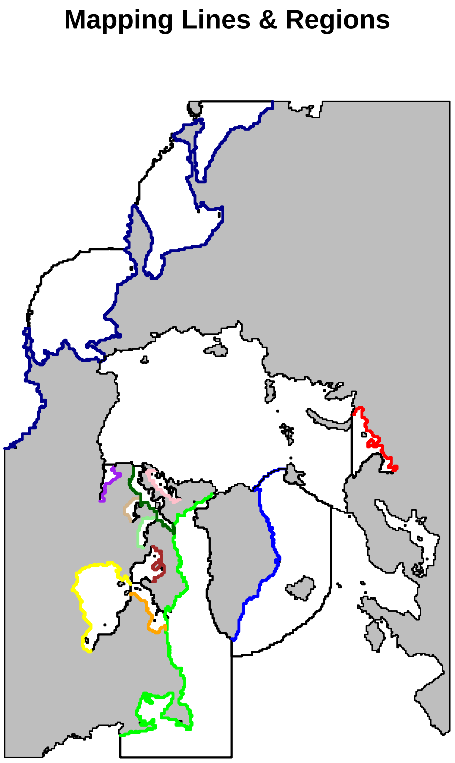


Figure 2.4: Map of region boundaries (in black) with fixed lines (colored) on top from which the ice will be mapped. The large region in the Central Arctic does not have a fixed line as is discussed in Section 2.2.1.

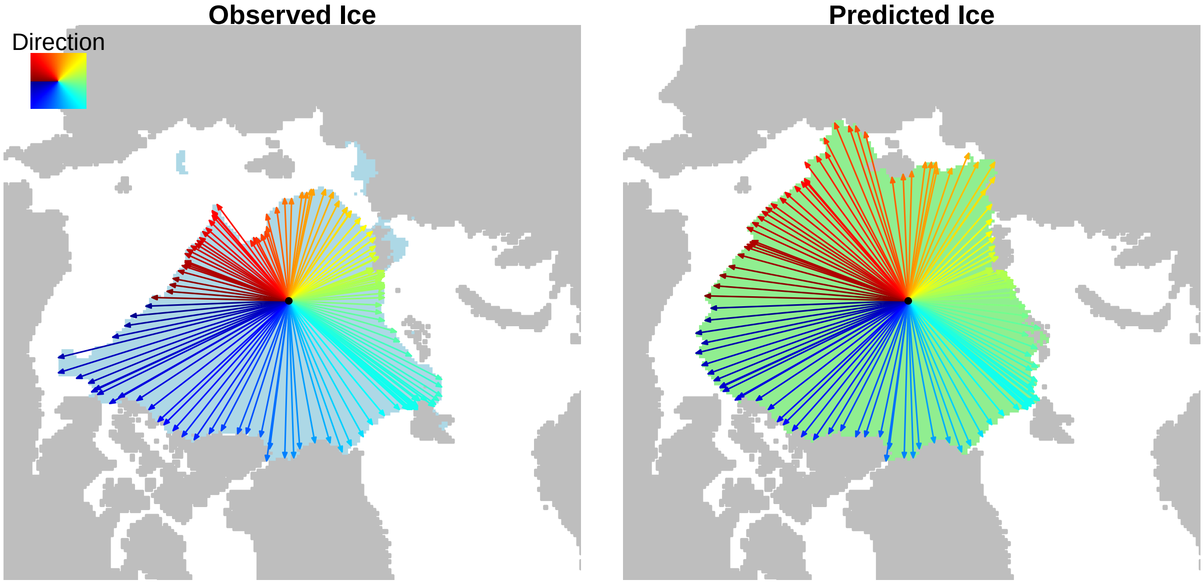


Figure 2.5: Sample of mappings in the Central Arctic region for September 2008 for the observation (left) and for the prediction with lead times of 4 months (right). The direction of the colored lines emanating from a central point in the region, indicated with a black circle, are fixed for all predictions and observations. On each line, the furthest point from the center of the region that intersects the sea ice edge is recorded. These points are those that the arrowheads touch. The line connecting these points forms an approximation to the boundary of the sea ice. Only every seventh line is plotted for ease of visualization.

differences in the mapping vectors along each point on the fixed lines. Repeating this process for a collection of observations and predictions enables relating predictions and observations to one another over time and space. For example, we can conclude that a prediction creates too much sea ice in a particular location if the vectors associated with predictions in that location tend to be longer than those associated with the observations. Similarly, we can assess if sea ice cover is decreasing over time in a region if the vectors associated with coordinates in this region typically get shorter over time.

2.2.2 Estimating the bias

Now that we have established a way to represent the sea ice edge, we can turn to the issue of estimating bias. With this mapping framework, increasing the accuracy of the sea ice edge can be achieved via reducing the error in the forecasts of the locations of the observed mapped points, $s'_{i,obs}$, for all locations i. At the time a forecast is issued any $s'_{i,obs}$ is unknown and must be predicted for each i. We denote a prediction of this quantity by $\hat{s}'_{i,obs}$. This follows the statistical convention of using the hat symbol above an unknown quantity to indicate a prediction or estimate of that quantity. We denote the mapped point obtained from the prediction from the dynamical model at location i by $s'_{i,pred}$. While this quantity is a prediction, the hat notation is not used, since $s'_{i,pred}$ is known at the time the forecast is issued. With this notation, we can now proceed to bias correction. We will use $s'_{i,pred}$ and a model for how $s'_{i,pred}$ is expected to differ systematically from $s'_{i,obs}$ to obtain a bias-corrected forecast, $\hat{s}'_{i,obs}$, for $s'_{i,obs}$.

To do this, we first need to isolate the systematic differences between observations and predictions in the mapped lines. The observed mapped point at location i, $s'_{i,obs}$, can be expressed as a function of both observations and predictions. Beginning with Equation 2.1, we can write the observed mapped point at any location i as

$$s'_{i,obs} = s_i + m_{i,obs} = s_i + m_{i,pred} + (m_{i,obs} - m_{i,pred}). \tag{2.3}$$

Further, $s'_{i,pred} = s_i + m_{i,pred}$ by definition, so we have

$$s'_{i,obs} = s'_{i,pred} + (m_{i,obs} - m_{i,pred}). \tag{2.4}$$

In other words the observed mapped vector at location i can be written as a sum of the predicted mapped vector at location i and the difference between the observed and predicted mapping vectors, $m_{i,obs} - m_{i,pred}$. This difference comes from both systematic discrepancies between the prediction and observation and irreducible error. We can decompose this error as

$$m_{i,obs} - m_{i,pred} = \mathbb{E}[m_{i,obs} - m_{i,pred}] + \epsilon_i \tag{2.5}$$

where $\mathbb{E}[m_{i,obs} - m_{i,pred}]$ denotes the expected systematic difference between the prediction and observation and ϵ_i represents the irreducible error, which we assume to be random with mean zero.

We now focus on modeling and correcting the systematic component of the error. Our goal is to obtain an unbiased estimator for $s_{i,obs}$. In general, an unbiased estimator is a function of data or other covariates, $\hat{\theta}$, that is used to predict a quantity, θ, and that on average has the same value as θ, that is $\mathbb{E}(\hat{\theta}) = \theta$. Assuming we can obtain an unbiased estimator for $\mathbb{E}[m_{i,obs} - m_{i,pred}]$, denoted by $\hat{\mathbb{E}}[m_{i,obs} - m_{i,pred}]$, we have that

$$\hat{s}'_{i,obs} = s_i + m_{i,pred} + \hat{\mathbb{E}}[m_{i,obs} - m_{i,pred}] \tag{2.6}$$

is an unbiased estimator for $s'_{i,obs}$. We can show this by taking the expectation of the estimator:

$$\mathbb{E}[\hat{s}'_{i,obs}] = \mathbb{E}[s_i + m_{i,pred} + \hat{\mathbb{E}}[m_{i,obs} - m_{i,pred}]] = s_i + m_{i,pred} + m_{i,obs} - m_{i,pred} = s_i + m_{i,obs} = s'_{i,obs}. \tag{2.7}$$

This means that on average the predicted position of the i-th mapping point will be the same as the i-th mapping point that will be observed. This is in contrast to $s'_{i,pred}$ which on average will be at a position $\mathbb{E}[m_{i,obs} - m_{i,pred}]$ units away from $s'_{i,obs}$. This error decomposition and correction has some similarities to what was proposed by Kharin et al. [2012] but provides results for many spatial locations rather than for a single summary measure.

The derivation in the previous paragraph presupposes that we have an unbiased estimator of $\mathbb{E}[m_{i,obs} - m_{i,pred}]$. Obtaining an approximately unbiased estimator is feasible with sufficient historical data and retrospective forecasts. We can map the observed and predicted sea ice regions using the technique described in Section 2.2.1. If the dynamical model bias was fixed over time, we could then estimate the bias at each location simply as the average difference between the observed and predicted mapping vectors for each year in the data set. However, this approach fails to account for the rapid change occurring in the Arctic.

Both model predictions and observations show a reduction in sea ice over time, but the rates of decline are not the same, both in total and for individual regions. For total sea ice extent, Stroeve et al. [2012a] found that the trend in most prediction systems is less than observed. To account for this, we extend the method proposed by Kharin et al. [2012] for bias-correcting total sea ice area to spatial contours. In their approach, the total sea ice area in observations and predictions are each regressed separately on time. Then, the regression models are used to estimate the predicted and observed sea ice area at the forecast time. The difference between these quantities is used as an estimate of the bias at the forecast time.

Extending this approach to our spatial setup, we estimate the bias at each point on the fixed line using regression. Outside the Central Arctic, we regress each of the x and y components of the mapping vectors for the observation and the prediction separately on time. For each point i on the mapped line, this gives four regression equations of the form,

$$m_i = \alpha_i + \beta_i t + \epsilon_i, \tag{2.8}$$

for the observed x component, observed y component, predicted x component, and predicted y component. The ϵ_i's are error terms. Consequently, the predicted bias in the mapping line at location i and time t is

$$\hat{\mathbb{E}}[m_{i,obs} - m_{i,pred}] = \hat{m}_{i,obs} - \hat{m}_{i,pred}, \tag{2.9}$$

where each m and $\hat{m}$ are vectors with an x and y component. Here $\hat{m} = \hat{\alpha}_i + \hat{\beta}_i$ where $\hat{\alpha}_i$ and $\hat{\beta}_i$ are estimates of the regression parameters. To estimate the regression parameters, we use

Huber M-estimation, a form of robust regression [Huber and Ronchetti, 2011]. Implemented in R using the *rlm* function, this technique minimizes a function of the residuals that puts reduced weight on outlying values [Venables and Ripley, 2002]. This is appropriate in this context, since we do not want a few unusual years to have undue influence on the trend lines. In Figure 2.6, we plot the x and y components of several observations and predictions along with regression lines for the sample points highlighted in Figure 2.3. Points are sometimes repeated in multiple years if the ice edge borders a land or region boundary.

In the Central Arctic, the mapped points are constrained to stay on specific lines, so the mapping vectors can be represented by just their lengths rather than by separate x and y components. Consequently, for region R + 1, we let

$$\hat{\mathbb{E}}[m_{i,obs} - m_{i,pred}] = \hat{m}_{i,obs} - \hat{m}_{i,pred} = (\hat{\alpha}_{i,obs} + \hat{\beta}_{i,obs}t) - (\hat{\alpha}_{i,pred} + \hat{\beta}_{i,pred}t)$$

where each m is now a single length rather than a vector. This reduces the number of parameters that need to be estimated, but otherwise does not affect the bias correction procedure.

2.2.3 Finalizing a prediction

To finalize a prediction, we replace the mapped points with the bias-corrected mapped points and connect them in the same way as in Section 2.2.1. Any sections of sea ice predicted that do not touch land outside the Central Arctic region are not bias-corrected, i.e. they are left as predicted by the dynamical model. In the Central Arctic region, sea ice sections that are not part of the largest contiguous region are also left unadjusted.

Some issues with the bias-corrected contour remain. In some cases the resulting contour line will intersect itself, making it an invalid boundary for a contiguous region. To remove these self-intersections, we simplify the sections of the contour surrounding self-intersections with the Douglas-Peucker algorithm [Douglas and Peucker, 1973]. This algorithm takes in a line, represented as a sequence of connected points, and outputs a new line that resembles the original line. The new line uses as few points as possible while ensuring that the new line

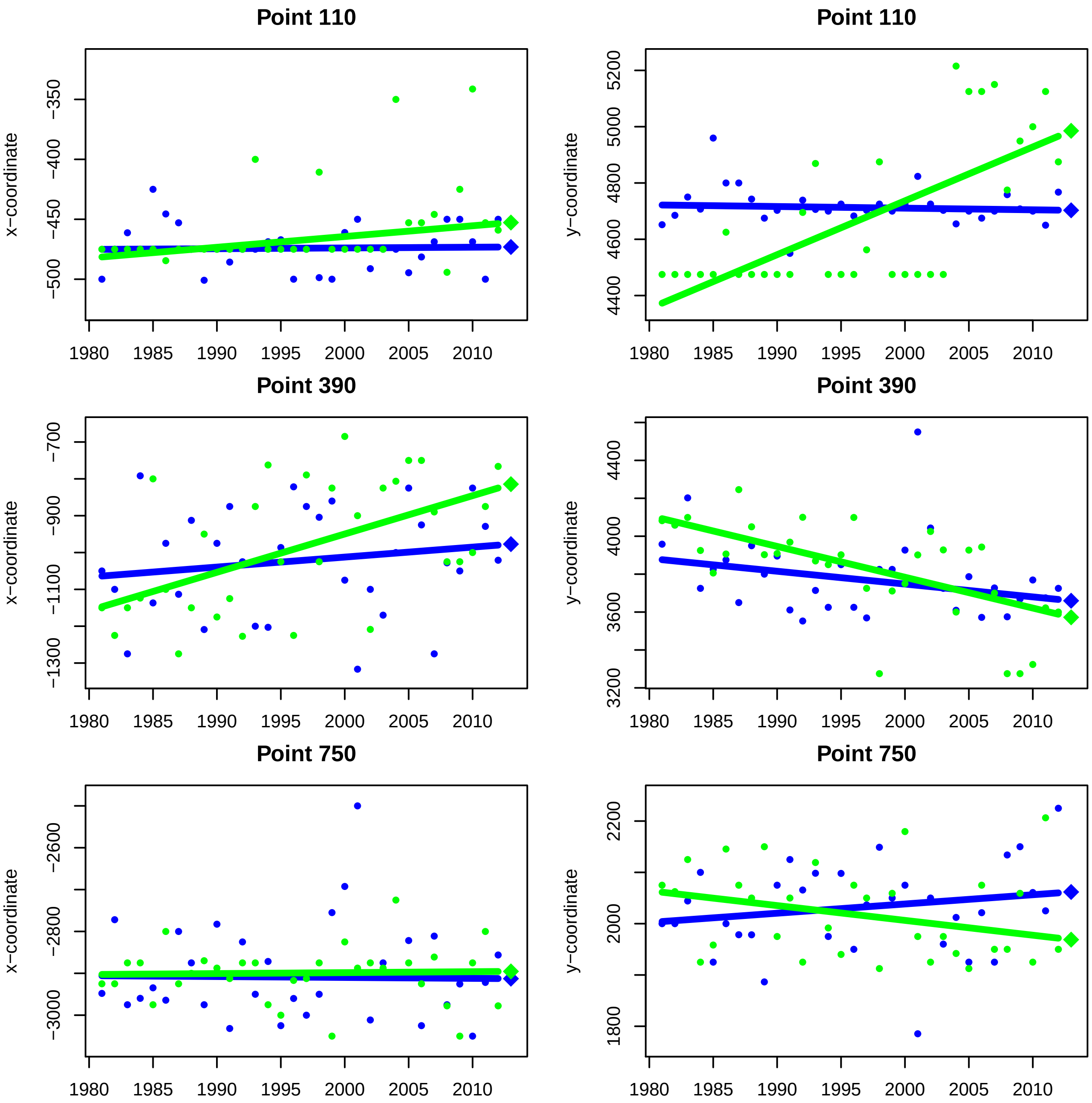


Figure 2.6: Time series for the three sample points in Figure 2.3 used to predict the bias in February 2012 at a lead time of 4 months. Panels in the left column are for the x-coordinate and panels in the right column are for the y-coordinate. Circles are recorded data and diamonds are the prediction for 2012 obtained via Huber M-estimation, a form of robust regression. Blue indicates observation and green indicates prediction. The expected bias is the difference between the two diamonds.

differs by no more than a user-specified tolerance, ϵ, from the original line. For each section of the contour that contains a self-intersection, we replace it with a similar line obtained with the Douglas-Peuker algorithm.

The tolerance ϵ is determined by an automated process. The Douglas-Peuker algorithm with a low ϵ is initially applied to the contour and a check is performed to determine if self-intersections remain. If there are self-intersections, ϵ is increased and the Douglas-Peuker algorithm is run again. This process is repeated until an ϵ is found that is large enough to give a contour with no self-intersections. In our analysis, we set the initial ϵ to 0.25 on the nominal 25 kilometer grid and increase it by 0.25 as needed.

In this way, we create a final bias-corrected contour that does not contain any self-intersections. It is also possible that a bias-corrected prediction will go through land. When this is the case, we simply remove the land contained within the region enclosed by the contour, so that the contour line goes through the interior edge of the land region. Examples of both of these adjustments to a contour are given in Figure 2.7. Finally, sections of sea ice that have been expanded such that they now overlap are merged into a single section of sea ice.

2.2.4 Implementation & Software

We have made our implementation of Contour-Shifting available in the *IceCast* R package [Director et al., 2017b]. The software package includes a one-line automated execution function that takes in observations and predictions and outputs bias-corrected predictions. Additionally, all intermediate functions and their code are accessible within the package. Interested users can use these intermediate functions to gain more flexibility in how Contour-Shifting is implemented or to extend or modify this method. A vignette, or tutorial within the package, discusses the *IceCast* package in more detail and provides examples of its use.

Within the package two main functions are used to execute Contour-Shifting. The first function maps observations and predictions and the second estimates the bias and creates the new prediction based on results from the first function. On a typical laptop, mapping

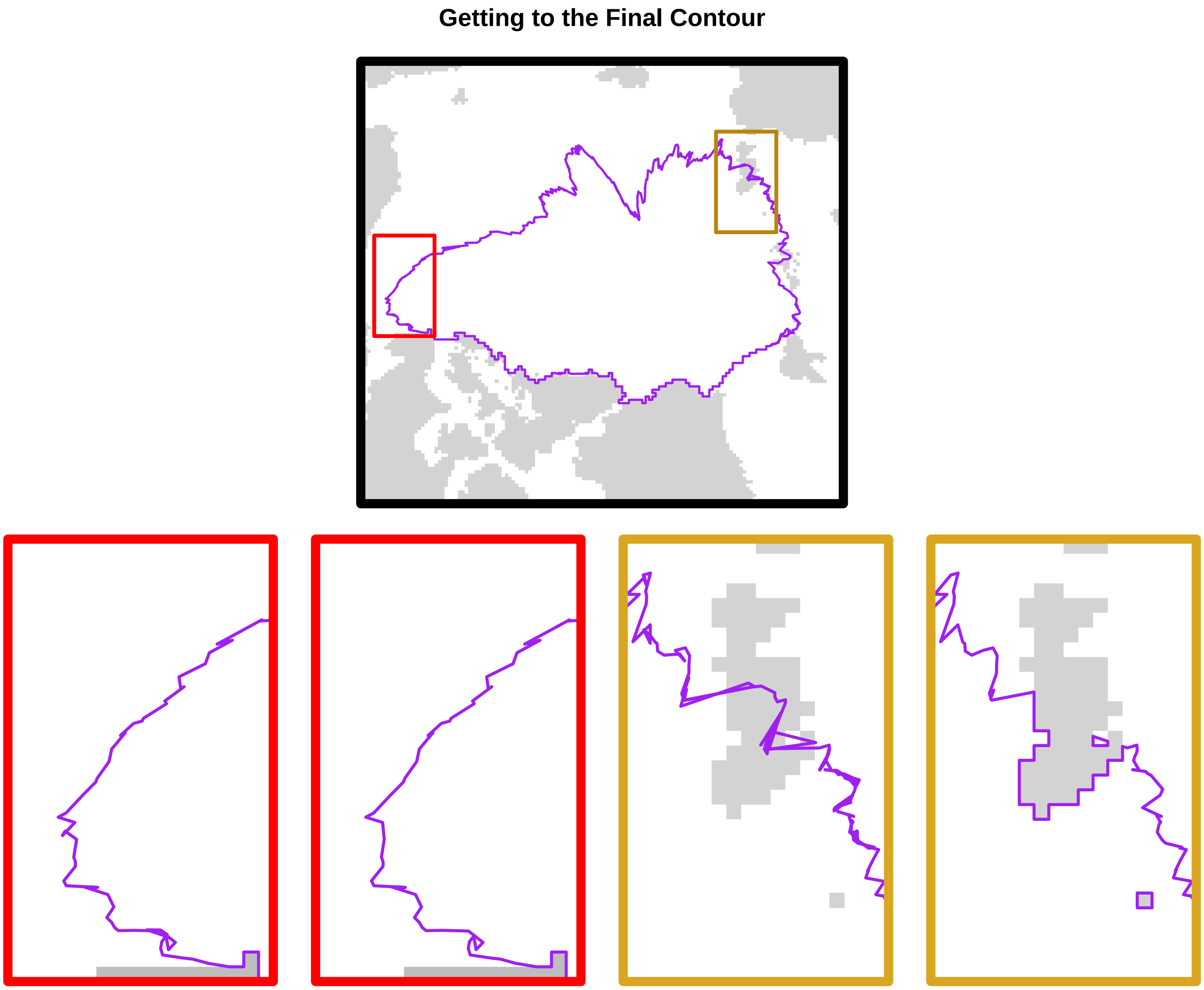


Figure 2.7: Example of modifying a bias-corrected contour when the contour line crosses over itself and/or passes through land. On the top panel is a sample contour. The red box highlights a region where the contour (in purple) crosses over itself and the gold box highlights a region where the contour goes through land. The remaining four panels zoom in on these sections with the left figure showing the original contour and the right figure showing it after correction. All issues have been corrected applying the algorithm described in Section 2.2.3. As shown in the gold box, land areas are also removed from within the contour boundary.

the prediction and observation for a single year takes around a minute. Given the mappings, estimating the bias correction and creating a new prediction for a single year takes about a minute. In other words, computation time to bias correct a single year of predictions given N years of observations and predictions is roughly N + 1 minutes. However, if multiple years are bias-corrected, the computation time per year is notably reduced since the mappings for each previous year used to learn the bias need to be calculated only once. There is some variability in computation time depending on the forecast month. Months with less sea ice tend to run faster than average, since there are less regions with sea ice to map and bias correct. The opposite is true for months with more sea ice. While Contour-Shifting does have greater computational cost than existing methods for bias correction of the total sea ice area or extent, it provides substantially more information about where sea ice is expected to be located. Further, relative to the cost of running a full dynamical forecasting system, the computation time remains trivial.

2.3 Results

2.3.1 Experimental set-up

To evaluate this bias correction approach, we consider retrospective predictions of monthly sea ice concentration from the CM2.5 Forecast-oriented Low-Ocean Resolution (FLOR) model produced by the National Oceanic and Atmospheric Administration's Geophysical Fluid Dynamics Laboratory (GFDL) from 1981-2013 [Vecchi et al., 2014, Msadek et al., 2014]. This fully-coupled global climate model has approximate atmospheric resolution of 50 km $\times$ 50 km and approximate ocean resolution of $1^\circ \times 1^\circ$. This prediction system has twelve ensemble members associated with it. The predictions were downloaded from the Earth System Grid [National Center for Atmospheric Research, 2017] and have been converted from their native grid to the nominal 25 km polar stereographic grid used in the NASA observational data product using weights obtained from the spherical coordinate remapping and interpolation package [Jones, 1997]. For total sea ice, this model is shown to be of high quality for forecasting the mean state [Msadek et al., 2014]. However, this does not mean that there is not spatial bias, since for total sea ice estimates overestimation of sea ice in one region can compensate for underestimation of sea ice in another region. This type of error, referred to as misplacement error, has been shown to form a substantial fraction of the total error in many model predictions [Goessling et al., 2016].

For our analysis, we consider monthly forecasts that have been initialized on the first day of the month and run for a year. This gives twelve initialization times per year. To obtain the prediction for the 15% sea ice edge corresponding to the ensemble forecast, we average the concentration in each grid box over the ensemble members to create a field of the average concentration. Then, we find the 15% sea ice edge for this averaged field, i.e. we find the boundary around the grid boxes with at least 15% concentration.

To assess Contour-Shifting's effect on prediction accuracy, we apply it to forecasts for all twelve months at four different lead times. We evaluate all forecast months, because sea ice has varying behavior and potential predictability during different parts of the year, such as differing persistence by season [Blanchard-Wrigglesworth et al., 2011a]. Since forecasts are for month-long periods, there can be some ambiguity in the definition of lead time. We report lead times rounding up to the nearest whole month. For example, a prediction for the monthly sea ice concentration for February made on January 1st is described as being made at a lead time of 2 months. For any prediction, we use all data prior to the forecast year to estimate the bias correction. Results are reported for 2001-2013. This corresponds to years where there was at least twenty years of data available prior to the forecast year from which to build the bias model.

Predictions from this model are compared to observations of the monthly sea ice concentration obtained from the National Aeronautics and Space Administration (NASA) satellites Nimbus-7 SMMR and DMSP SSM/I-SSMIS. The observations have been processed by the bootstrap algorithm and are distributed by the National Snow and Ice Data Center [Comiso, 2000, updated 2015]. The original land mask for the observations has been replaced with the simpler land mask (derived from approximately 1°) used in the predictions. We define a grid box to contain sea ice only if it has at least 15% sea ice concentration.

2.3.2 Spatial error reduction

In Figure 2.8, we map the initial GFDL prediction and the bias-corrected forecast of the sea ice contour on top of the observed sea ice for several example months and years. In many cases, the bias-corrected contour clearly follows the observed sea ice boundary more closely than the uncorrected dynamical model prediction does, indicating that bias-correction is improving the forecast.

To quantitively assess the improvement due to Contour-Shifting, we use the Integrated Ice-Edge Error (IIEE) measure proposed by Goessling et al. [2016]. The IIEE is an aggregate measure of the amount of error in a prediction. It is obtained by adding up the areas where ice is predicted but not observed and the areas where ice is observed but not predicted. For all months and lead times, we see a reduction in the average IIEE after bias correction has been applied. In Figure 2.9, we plot the average IIEE for the years 2001-2013 organized by forecast month for the unadjusted GFDL forecasts and for the bias-corrected forecasts. From this we can see the increased prediction skill that bias correction provides for each month. For example, for March, we see that the average IIEE for bias-corrected forecasts at all lead times is less than the lowest average IIEE for the unadjusted forecasts. This means that a bias-corrected forecast issued twelve months in advance is more accurate on average than an unadjusted forecast issued two months in advance. Similarly, the average IIEE for unadjusted September forecasts issued two months in advance is 13.8 $\times 10^5$ km^2. This is greater than the average IIEE for the bias-corrected forecast at a 2-month lead time (12.3 $\times 10^5$ km^2) and a 4-month lead time (13.4 $\times$ 10^5 km^2). This indicates that on average for September a biased-corrected forecast issued four months in advance is more accurate than an unadjusted forecast issued both two and four months in advance.

In Figure 2.10, we report the mean error reduction of this measure for all forecasts months at four different lead times. Figure 2.11 reports the mean percent change in the IIEE. For all months and lead times, we obtain an average reduction in the IIEE of 3.3 $\times 10^5$ km^2,

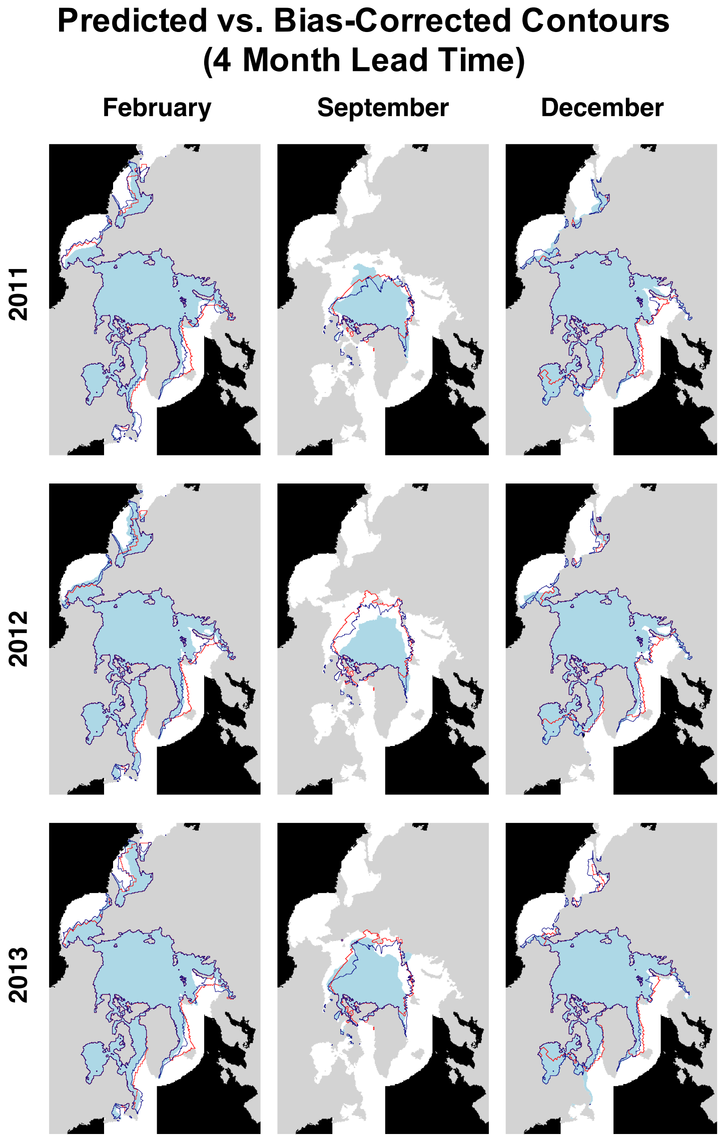


Figure 2.8: Predictions for the 15% sea ice concentration contour obtained from the GFDL model with a 4-month lead time (red) and bias-corrected using Contour-Shifting (navy) for February, September, and December from 2011-2013. The observed ice region is plotted in light blue. The black regions are part of the “non-regional ocean” in the ocean mask described in Section 2.2.1 and are excluded in this analysis. The remaining white area represents grid boxes with sea ice concentration lower than 15%.

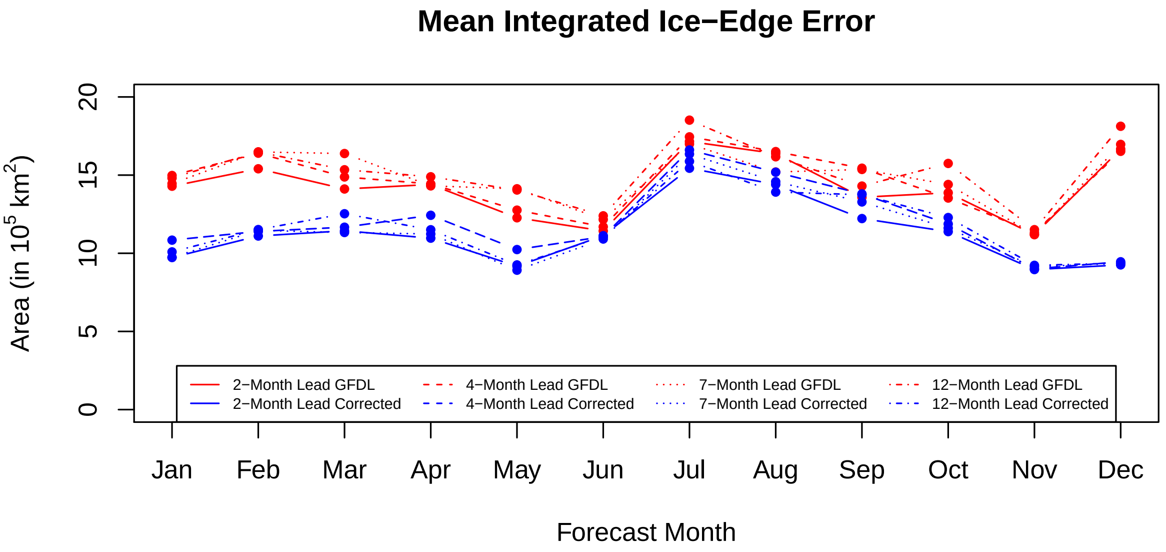
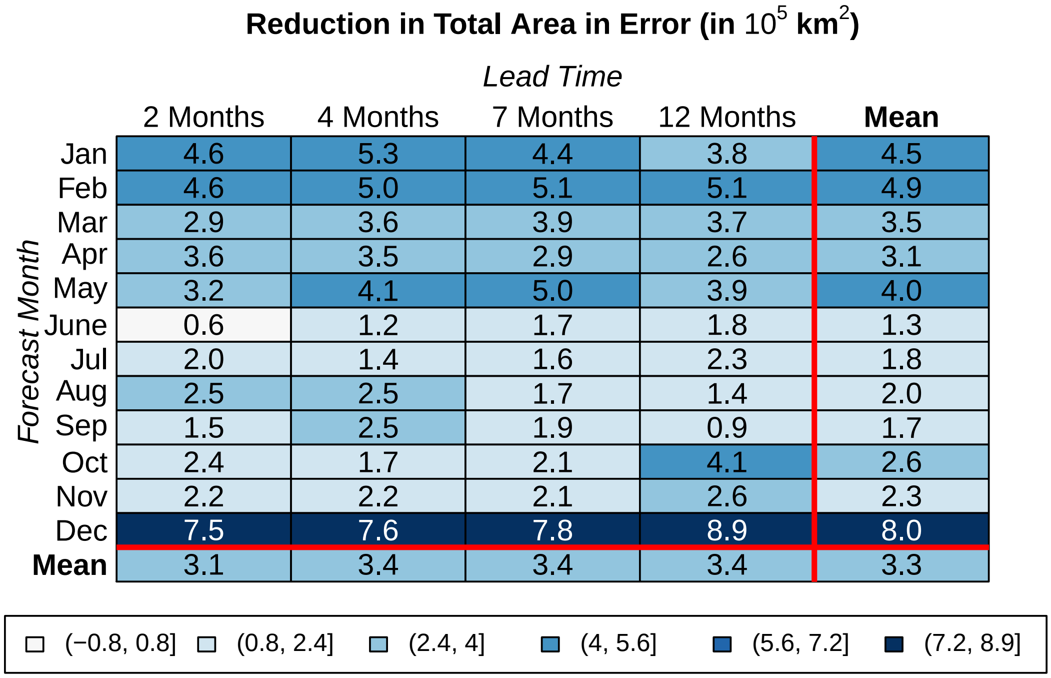


Figure 2.9: Average Integrated Ice-Edge Error for the years 2001-2013 by month for the unadjusted GFDL model output (red) and bias-corrected with Contour-Shifting (blue). The different line types indicate the lead time.

or 21.3%. To compare performance across months and lead times, we look at the percent error reduction results in Figure 2.11. We note that the amount of error reduction in a given month is not directly related to the IIEE in the original forecast. Rather, the amount of error that can be corrected is a function of how much systematic error a prediction has. Contour-Shifting will provide the most error reduction when there is a notable amount of error that follows consistent patterns. Contour-Shifting will provide less error reduction when the unadjusContour-Shiftingted forecast is quite accurate or its error is highly variable.

The largest percent error reductions occur for forecasts in the winter and in May. These months have large, consistent errors across years. For example, the month of December has a high amount of error overall and the errors are repeated across years. As can be seen in Figure 2.8, underestimation of sea ice in Hudson Bay and overestimation of sea ice off the coast of Greenland north of Iceland and in the Barents Sea occurs in most years. Thus these



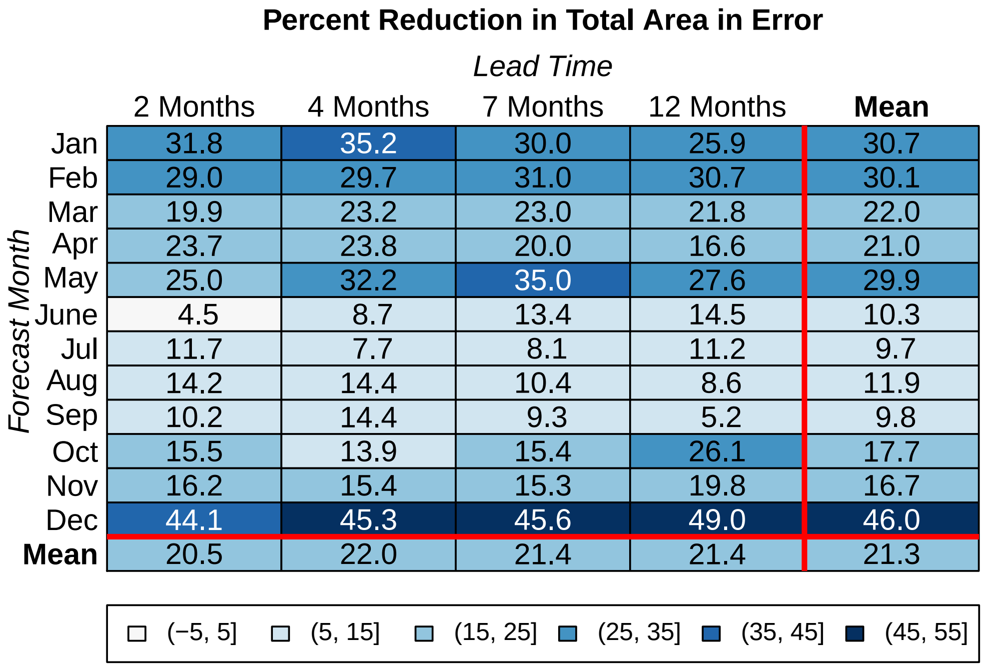
Reduction in Total Area in Error (in 10^5 km^2)

Lead Time

Forecast Month	2 Months	4 Months	7 Months	12 Months	**Mean**
Jan	4.6	5.3	4.4	3.8	4.5
Feb	4.6	5.0	5.1	5.1	4.9
Mar	2.9	3.6	3.9	3.7	3.5
Apr	3.6	3.5	2.9	2.6	3.1
May	3.2	4.1	5.0	3.9	4.0
June	0.6	1.2	1.7	1.8	1.3
Jul	2.0	1.4	1.6	2.3	1.8
Aug	2.5	2.5	1.7	1.4	2.0
Sep	1.5	2.5	1.9	0.9	1.7
Oct	2.4	1.7	2.1	4.1	2.6
Nov	2.2	2.2	2.1	2.6	2.3
Dec	7.5	7.6	7.8	8.9	8.0
Mean	3.1	3.4	3.4	3.4	3.3

□ (−0.8, 0.8] □ (0.8, 2.4] ■ (2.4, 4] ■ (4, 5.6] ■ (5.6, 7.2] ■ (7.2, 8.9]

Figure 2.10: The reduction in the mean integrated ice-edge error (IIEE) in 10^5 km^2 over the test years 2001-2013 for bias-corrected predictions compared to the original GFDL predictions at four different lead times. For each year, the IIEE is defined to be the sum of all area where sea ice is predicted but not observed or where sea ice is observed but not predicted. Results are rounded to one decimal place.



Percent Reduction in Total Area in Error

Lead Time

Forecast Month	2 Months	4 Months	7 Months	12 Months	**Mean**
Jan	31.8	35.2	30.0	25.9	30.7
Feb	29.0	29.7	31.0	30.7	30.1
Mar	19.9	23.2	23.0	21.8	22.0
Apr	23.7	23.8	20.0	16.6	21.0
May	25.0	32.2	35.0	27.6	29.9
June	4.5	8.7	13.4	14.5	10.3
Jul	11.7	7.7	8.1	11.2	9.7
Aug	14.2	14.4	10.4	8.6	11.9
Sep	10.2	14.4	9.3	5.2	9.8
Oct	15.5	13.9	15.4	26.1	17.7
Nov	16.2	15.4	15.3	19.8	16.7
Dec	44.1	45.3	45.6	49.0	46.0
Mean	20.5	22.0	21.4	21.4	21.3

□ (−5, 5] □ (5, 15] □ (15, 25] □ (25, 35] □ (35, 45] □ (45, 55]

Figure 2.11: As in Figure 2.10, except with the percent change in mean error area. Results are again rounded to one decimal place.

errors can be well corrected by Contour-Shifting.

On the other hand, the smallest percent error reductions occur from June to September. The adjusted IIEE values are only marginally better than the unadjusted IIEE values. This suggests that random, rather than systemic, errors dominate during these months and/or that our bias correction method is not identifying systematic errors. Mechanistically, this is plausible. In early summer, the heterogenous surface of the Arctic is especially affected by the ice-albedo feedback [Blanchard-Wrigglesworth et al., 2011a] and in late summer and early fall there is the highest proportion of thin ice which leads to more variability [Holland et al., 2011]. This variability makes it less likely that a consistent discrepancy between models and observations would occur and makes it so that more data would be needed to detect any bias that is there. For example, looking at consecutive September observation and predictions in the second column of Figure 2.8, the only repeated error pattern is a small underestimation of sea ice off the northeastern part of Greenland. The interior of the Arctic Ocean is variable and not substantially affected by bias correction.

No clear patterns emerge in performance across lead times, suggesting that Contour-Shifting can be applied to obtain more accurate forecasts for lead times up to at least a year. We also find little effect of lead time on IIEE even after bias correction. This suggests that initial conditions strongly affect forecast accuracy at these time scales. This result aligns with that of Msadek et al. [2014] for this dynamical model who infer that initial conditions are a primary factor in predicting sea ice extent at lead times up to a year.

2.3.3 Total sea ice area

While the key contribution of Contour-Shifting is the ability to bias correct a contour, this technique can also be employed to bias correct the total sea ice area. Sea ice area refers to the total area of grid boxes with sea ice concentration greater than 15%. To obtain a bias-corrected estimate of the total sea ice area with Contour-Shifting, one simply needs to sum the areas contained within the bias-corrected contour. This approach gives comparable performance to existing bias-correction techniques focused exclusively on correcting total sea ice area estimates.

To illustrate this, we implement the trend bias correction technique introduced in Kharin et al. [2012] and the initial condition technique introduced in Fučkar et al. [2014]. The approach of Kharin et al. [2012] requires regressing observed total sea ice area on time and regressing predicted total sea ice area on time. The difference in the regression lines at the forecast time is then used to estimate and correct for the expected difference in total sea ice area between the observation and the prediction at the forecast time. The initial condition approach is similar, except that instead of regressing on time, the observed and predicted total sea ice areas are regressed on information describing their initial conditions. In our implementation, we regress on the observed monthly sea ice areas for the month prior to initialization.

In Figure 2.12, we plot the root mean square error (RMSE) for predictions of the total sea ice area obtained versus the forecast month for four lead times. We do this for the GFDL model alone and with the various bias correction techniques. We found RMSE to be lower for the bias-corrected forecasts than for the unadjusted GFDL predictions for all the bias correction techniques, highlighting the importance of bias correction to obtain accurate sea ice forecasts. For all lead times, Contour-Shifting had lower RMSE than all the other methods considered. Contour-Shifting performed best around the time when sea ice reached its minimum. This differs from the Kharin et al. [2012] technique which performed best around June. Overall this analysis shows that Contour-Shifting not only improves the prediction of the sea ice contour, but also leads to more accurate estimates of the total sea ice area.

2.4 Discussion

We have proposed and implemented Contour-Shifting, a novel technique for bias correcting a prediction of a contour of sea ice at a particular concentration, a typical definition of the sea ice edge. Retrospective comparisons of observations and predictions are used to model how the contour obtained from a dynamical model typically differs from the contour observed. New predicted contours are then adjusted to correct for the expected discrepancy. This ap-

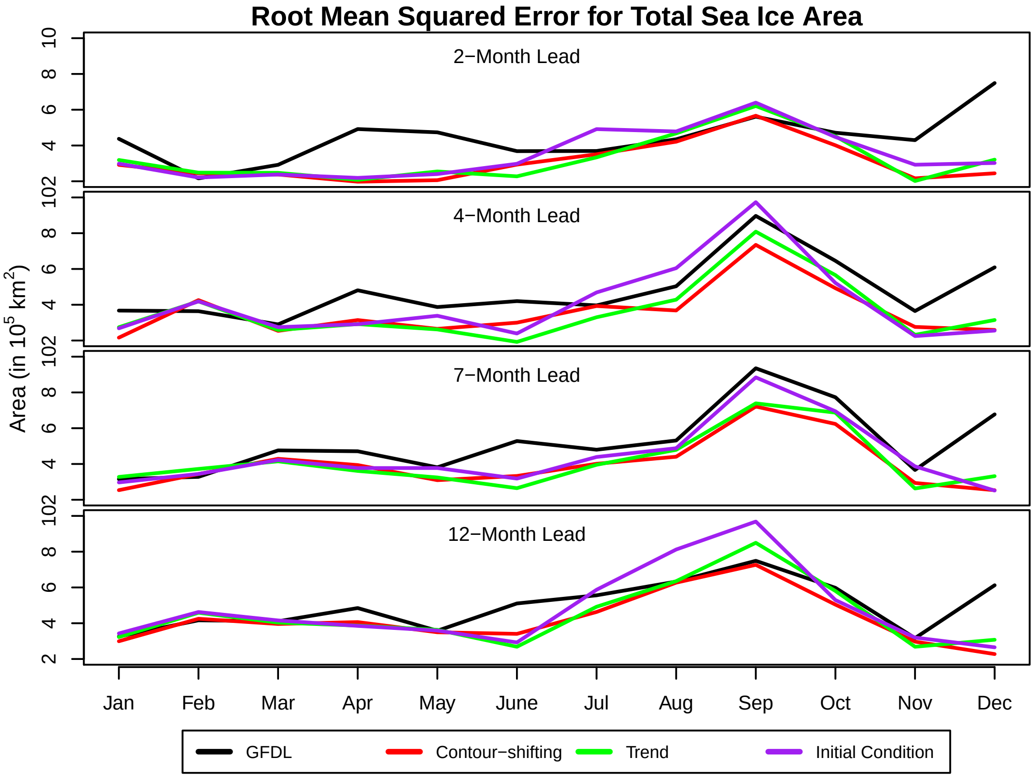


Figure 2.12: The root mean squared error for the total sea ice area prediction at four different lead times using four different methods. The black line, "GFDL" is for the unadjusted dynamical model. The other lines are for bias-corrected versions of these forecasts. The red line, "Contour-Shifting", is for the bias correction method described in this chapter. The green line, "Trend", is for the bias correction method given in Kharin et al. [2012] and the purple line, "Initial Condition", is for the bias correction method given in Fučkar et al. [2014]

proach goes beyond existing bias correction methods by addressing the spatial configuration of the sea ice rather than just summary measures of the total amount of sea ice.

In our experiments, Contour-Shifting reduced error in all forecast months and lead times considered. The total area where the prediction and the observations disagree, or IIEE, was reduced by 3.3×10^5 km^2, or 21.3%, on average. Contour-Shifting also produced error reductions in forecasts of the total sea ice area that were better than existing bias correction methods that focus only on forecasting this summary measure. The methodology of Contour-Shifting is not specific to the 15% sea ice concentration threshold used in this chapter. After categorizing grid boxes as being above or below a different threshold, a bias-corrected contour for another threshold can be obtained in the same way as the 15% threshold.

We have evaluated spatial prediction accuracy using the IIEE measure [Goessling et al., 2016]. Since Contour-Shifting leads to substantial error reduction, which evaluation criterion is used is not of great importance. Any reasonable criterion summarizing areas of discrepancy between observations and predictions would likely show improvement of forecast accuracy with bias correction. However, as Dukhovskoy et al. [2015] highlight, there are a number of properties that an ideal sea ice measure should have. As further spatial bias correction methods are developed, more nuanced measures could be useful for distinguishing among them.

2.4.1 Method comparisons

A simple alternative to Contour-Shifting for obtaining spatial predictions of sea ice would be to independently apply an existing bias correction technique to each grid box. This does notably reduce the IIEE compared to an unadjusted prediction. However, bias correcting each grid box independently occasionally leaves scattered grid boxes or groups of grid boxes that contain sea ice in an area that is otherwise open water or vice versa. We show examples of this occurring in Figure 2.13 using predictions obtained with a slightly modified version of the bias correction proposed by Kharin et al. [2012]. Specfically, we modify the Kharin et al. [2012] approach to account for the range of values which concentrations can take by replacing linear regression with logistic regression and rounding adjusted concentrations down to 1 or up to 0 as needed.

While the artifacts in Figure 2.13 are not that common, they illustrate why assuming independence of grid boxes is problematic, especially for binary data. The modeling and computation needed to ensure these artifacts never occur is substantial. A statistical spatial model over the field would be appropriate, but the presence of land and other physical features and the size of the grid complicate the design of an appropriate and computationally-tractable covariance structure. Working instead with the lower-dimensional contour avoids this limiting framework. It also opens up the possibility of extensions to this method, such as covariate-based time series models or probabilistic modeling, that would be computationally challenging to model directly on a full grid.

In Section 2.3.1, we define the 15% contour with respect to the concentration level estimates obtained from the ensemble mean for each grid box. However, this is not the only way the sea ice contour could have been defined. For example, in the Sea Ice Outlook produced by the Sea Ice Prediction Network, each grid box in each ensemble member is defined as containing sea ice or not, based on its concentration. The average of these values is used as a summary of each grid box [e.g. Sea Ice Outlook August 2014 Report]. While common, in this case this approach gives a slightly greater average IIEE for the GFDL forecasts than the technique described in Section 2.3.1. This suggests that further investigation is needed to determine how best to convert ensemble model output into a single binary prediction for each grid box. Regardless, even for these alternative definitions of the contour, Contour-Shifting still produces substantial error reductions over unadjusted forecasts.

Connecting two contours with mapping lines has some parallels to the objective curves proposed by Strong [2012] for measuring the marginal ice zone. His approach to matching up curves cannot be applied here since it requires full knowledge of both the predicted and observed curves which we do not have at the time a forecast is issued. However, the similar aspects of these methods highlight the need for techniques that relate contours systematically. We expect that other systematic approaches to matching contours could be used in the

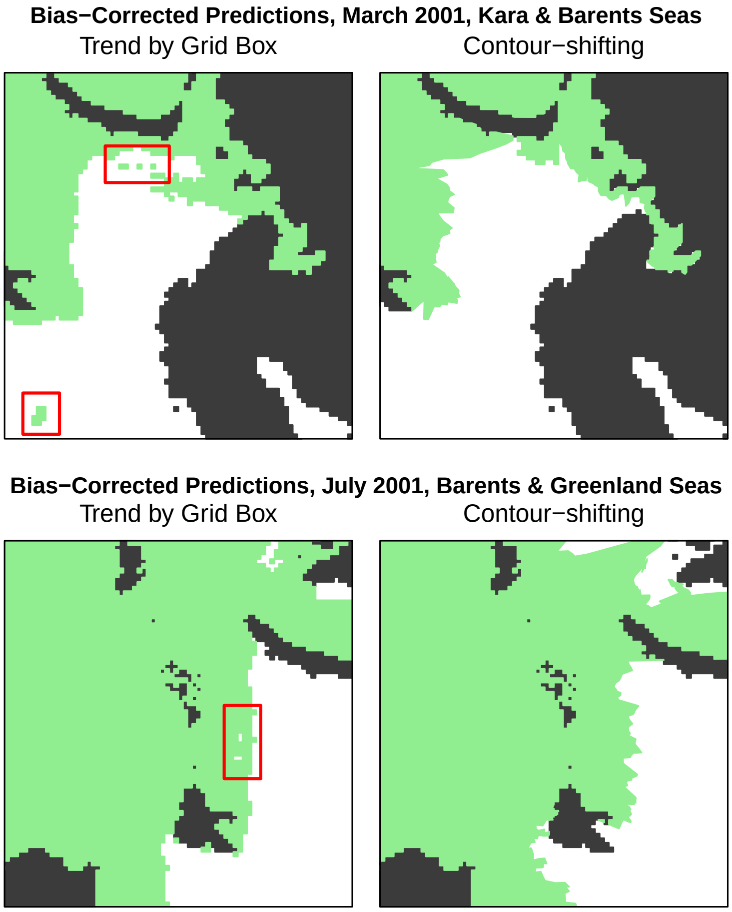


Figure 2.13: Predictions of regions with sea ice concentration of at least 15% (in green) for March 2001 and July 2001 bias corrected using a variation of the Kharin et al. [2012] approach where logistic regression is used in place of linear regression and concentrations are rounded to be within the interval $[0, 1]$ (left) and bias-corrected with Contour-Shifting (right). The red boxes in the left column highlight disconnected sets of a few grid boxes containing sea ice within open ocean and holes of a small number of grid boxes within large contiguous sections of sea ice.

general framework of Contour-Shifting with success comparable to the method we have proposed.

2.4.2 Limitations and future work

Expected changes in sea ice concentration leave some questions open about how spatial bias correction will perform in the future. There is currently limited analysis of how the predictability of sea ice will change over time. Using "perfect model" experiments (which test the ability of a model to predict its own sea ice state), Holland et al. (2011) and Tietsche et al. (2013) both infer that potential skill in predicting sea ice extent will decline. The resulting increase in irreducible error may reduce the performance of bias correction, since biases are more difficult to identify and model in a system with more variability. Additionally, perfect model experiments provide no insight into how biases between models and observations might change. If, in the future, errors are less consistent, the performance of bias correction will decrease, even if the total area in error does not increase. The general pattern of the mapping vectors in the model and observations may also change over time or could be affected by long-term variability. We have been able to obtain substantial error reductions on average using the simple assumption that the time series for the components of all mapping vectors follow a consistent linear trend. However, were the relationships between the model and the observation to change or were the relationships found to be more complicated on multi-decadal time scales, we might need more complex time series models, such as state-space models, to respond to changes in the trend and to accurately represent nonlinear structure [e.g. Brockwell and Davis, 2016]. Covariate information, particularly related to sea ice thickness and sea ice albedo, could also prove useful.

Additionally as the concentration of sea ice decreases, non-contiguous sections of ice that do not border land and holes in the sea ice may become more common. While an ideal bias correction method would provide a way to correct these features, our analysis does not address these cases. This is because there are currently insufficient examples of observations and predictions exhibiting these phenomena across space to be able to confidently develop a model for their bias. As more data become available, methodological extensions to Contour-Shifting to address biases in these features could be developed and assessed. For example, it would likely be feasible to build a model to identify and correct for regions where the size and/or frequency of holes are over- or underestimated.

Given the extensive development effort and large computational cost required to run dynamical forecasting systems, extracting all possible information from their predictions is of considerable importance. Without bias-correction methods that account for the spatial nature of climate data, predictions from these systems can only produce biased spatial information or bias-corrected summary measures. This work represents a first step to overcoming this limitation for sea ice. With Contour-Shifting, we can now use dynamical model output to correct sea ice concentration contours so they have reduced systematic differences from observations.

Chapter 3

PROBABILISTIC FORECASTING OF THE ARCTIC SEA ICE EDGE WITH CONTOUR MODELING

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 chapter introduces Mixture Contour Forecasting, a method to forecast sea ice probabilistically using a mixture of two distributions, one based on post-processed 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.

3.1 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., 2012b]. Continued reduction in sea ice is expected as the effects of climate changes increase. Reduced sea ice cover allows for increased Arctic shipping [Smith and Stephenson, 2013, Melia et al., 2016].

The importance of forecasting sea ice has increased in response, since waters without sea ice are more easily navigable than waters with sea ice. Reliable estimates of a ship's probability of encountering sea ice are needed to plan maritime routes that avoid sea ice. In this chapter, 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 then 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%. Thresholding is needed, since satellites often fail to distinguish between areas of open water and areas where water is melting on the sea ice's surface. Concentration reduces from about 50% to near 0% concentration over a small region, so the area classified as sea ice is only weakly affected by the exact threshold concentration used.

Zhang and Cressie [2019, 2020] have introduced hierarchical spatio-temporal generalized linear models for Arctic sea ice. However, many sea ice forecasts used in practice 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 typically run multiple times with slightly different initial conditions, and the outputs from the resulting runs have varying amounts of sea ice. The collection of forecasts, referred to as the ensemble, has shown skill in predicting the total area or extent of sea ice at seasonal time scales in retrospective forecasts [e.g., Sigmond et al., 2013, Msadek et al., 2014, Wang et al., 2013, Chevallier et al., 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 models at short lead times [Zampieri et al., 2018].

However, errors in ensembles are common because the underlying systems of differential equations are only approximations of the true physical processes, because initial conditions are not fully known, and because of sub-grid scale phenomena [Guemas et al., 2016, Blanchard-Wrigglesworth et al., 2015]. An ensemble can be biased, meaning that its mean behavior is systematically incorrect. It can also be poorly calibrated, meaning that the range of possible sea ice states predicted by the ensemble members does not reflect the actual uncertainty of the forecast.

Statistical post-processing, 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 chapter, 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 partially informed by the mean location of the sea ice edge contour obtained from ensemble outputs. The forecasts obtained from 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 well with a contour boundary, such as holes in the sea ice.

The MCF method provides better calibrated and more accurate probabilistic forecasts than the unadjusted ensemble and better calibrated forecasts than existing post-processing techniques. In Figure 3.1, we illustrate the extent to which MCF improves model calibration by plotting the predicted probability of sea ice presence in September against the actual proportion of times sea ice was observed for the raw ensemble and after post-processing. The predictions are from the fifth generation of the European Centre for Medium-Range Weather Forecasts (ECMWF) seasonal forecasting season (SEAS5) [Johnson et al., 2019, European Centre for Medium-Range Weather Forecasts, 2017]. We see that MCF provides much improved model calibration.

The chapter is organized as follows. In Section 3.2, we introduce a Bayesian model for generating distributions of the sea ice edge contour. This contour model is fit to observed ice edge contours from recent years and its prior is partially informed from the mean ice edge predicted from a dynamic ensemble. In Section 3.3, the contour model is combined with climatological information using a finite mixture model. In Section 3.4, we compare

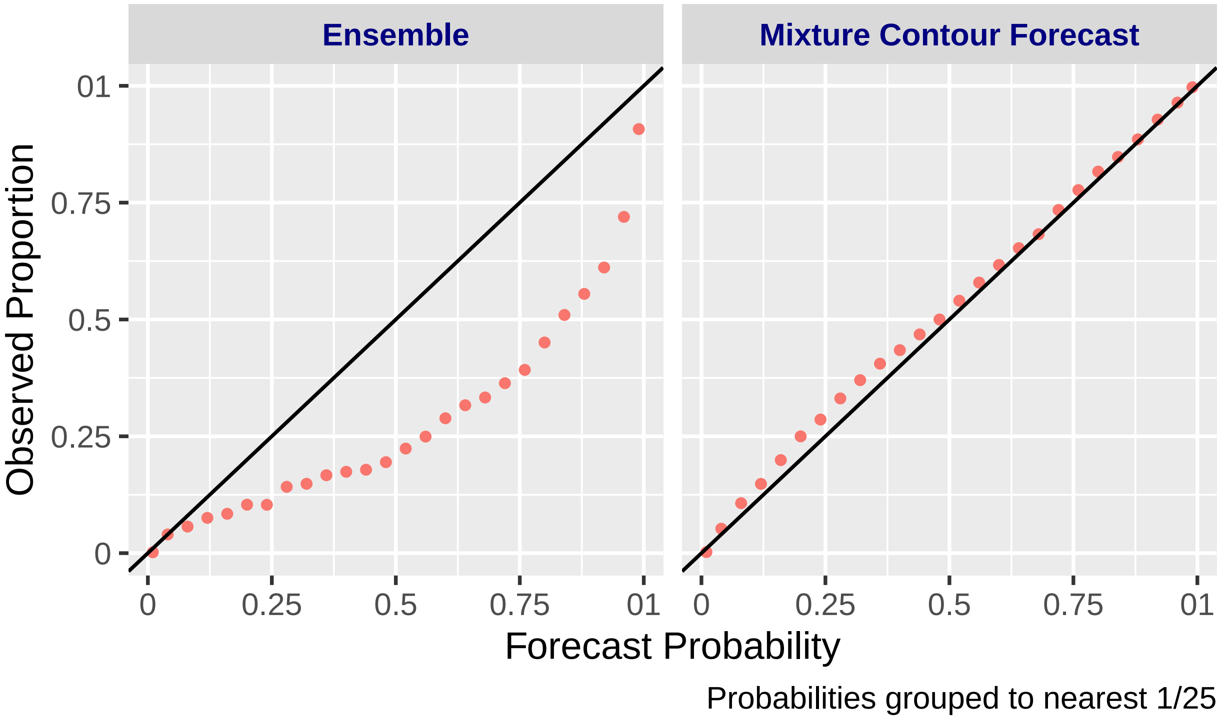


Figure 3.1: The average proportion of times sea ice was present, plotted against the predicted probability of sea ice presence for the unadjusted ECMWF ensemble forecasts (left) and for the corresponding Mixture Contour Forecasts (right) for lead times of 0.5 to 1.5 months. Results are for September 2008-2016. A perfectly calibrated forecast would have all points on the y = x line, so the MCF method forecast is better calibrated.

the performance of MCF to other post-processing and statistical forecasting techniques. In Section 3.5, we conclude with discussion.

3.2 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.

3.2.1 Notation and setup

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

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. As such, it is natural to model multiple contours separately. We focus on five regions individually. These five regions, shown in the map in Figure 3.2, exclude parts of the Arctic ocean where a contour model is not appropriate because the sea ice does not typically form one contiguous section. We selected these regions by modifying an existing region mask [Cavalieri and Parkinson, 2012] obtained from the National Snow and Ice Data Center [2017]. For notational simplicity, we do not subscript the regions and refer to the sea ice edge contour in a given region simply as $\boldsymbol{S}$.

In most regions, sea ice is formed in contiguous sections bordering land. In these regions, $\boldsymbol{S}$ is formed by a sequence of points that proceed from the coastline, into the ocean, and back to the coastline. We can reduce the number of points that need to be estimated by fixing a set of boundary points, $\boldsymbol{B} = (B_1, \ldots, B_n)$, on land and considering how far into the ocean the contour extends at each boundary location. The subset of points in $\boldsymbol{S}$ which must be fit are denoted by $\tilde{\boldsymbol{S}}$ and are indexed $\{1, 2, \ldots, n\}$. We lay out an ordered series of parallel lines, $\boldsymbol{L} = (L_1, \ldots, L_n)$, that cover the region. Each L_i extends from its corresponding point B_i to the edge of the region. We assume that one point, $\tilde{S}_i$, lies on each line, L_i. We denote the line segment from each point on the coastline, B_i, to the corresponding point $\tilde{S}_i$ as y_i. The set of all line segments is denoted by $\boldsymbol{Y} = (y_1, \ldots, y_n)$. The contour is then formed by connecting the points B_i to B_{i+1} for all $i = 1, \ldots, n-1$, B_n to $\tilde{S}_n$, $\tilde{S}_i$ to $\tilde{S}_{i-1}$ for all $i = n, \ldots, 2$, and $\tilde{S}_1$ to B_1. The left panel of Figure 3.3 illustrates these values for the Bering Sea. The angle of all lines in $\boldsymbol{L}$ is set to approximately match the direction the sea ice grows

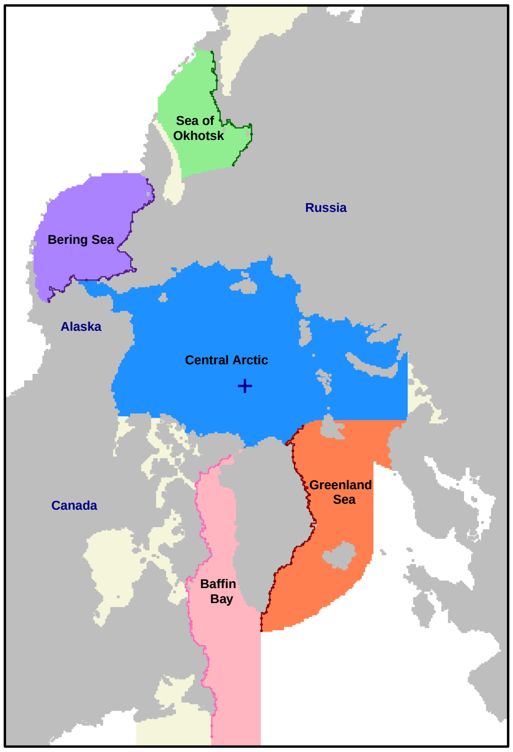


Figure 3.2: Arctic ocean regions used here. Each region in a color other than beige is fit with a contour model. The bold lines in all regions except the Central Arctic are the lines from which the fixed set of boundary points $\boldsymbol{B}$ are drawn. The boundary points themselves are plotted on top of these lines. The '+' symbol denotes the location of $\boldsymbol{B}$ in the Central Arctic Region. Areas in grey are land and areas in white are ocean regions that are not considered part of the Arctic ocean in the National Snow and Ice Data Center land mask [National Snow and Ice Data Center, 2017].

off the land in each region.

Unlike in other regions, sea ice in the Central Arctic region is not generally formed off a land boundary. To represent the Central Arctic's contour, we fix all the lines in $\boldsymbol{L}$ to originate from a single fixed central point rather than from a sequence of points. So, for this region $B_i = B_j$ for all $B_i, B_j \in \boldsymbol{B}$. The lines extend at fixed angles evenly spaced around a circle as illustrated in the right panel of Figure 3.3. In this case, $\tilde{\boldsymbol{S}} = \boldsymbol{S}$ and $n = N$.

Note that given $\boldsymbol{B}$, $\boldsymbol{Y}$, and the angles of all lines in $\boldsymbol{L}$, we have enough information to

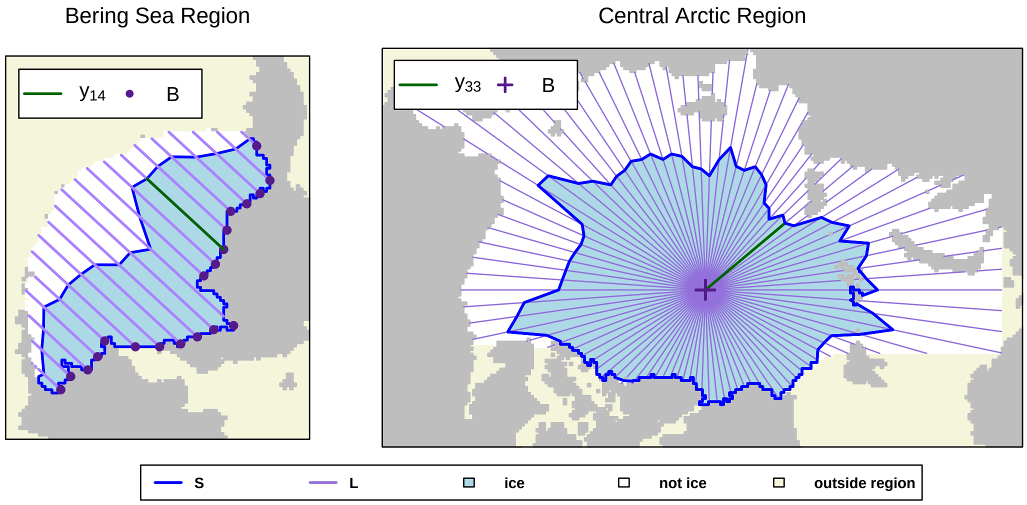


Figure 3.3: Hypothetical sea ice edge contours ($\boldsymbol{S}$), sets of fixed boundary points ($\boldsymbol{B}$), and parallel lines ($\boldsymbol{L}$) on which the points $\tilde{\boldsymbol{S}}$ will be generated for a sample typical region (left) and for the Central Arctic region (right). The green line designates the observed ice-covered line segments for the 14th (left) and 33rd (right) lines.

identify each $\tilde{S}_i$. We need only compute the length of each line segment $y_i \in \boldsymbol{Y}$. Each coordinate of the contour is then,

$$\tilde{s}_i = \boldsymbol{B}_i + (||y_i|| \cos(\theta_i), ||y_i|| \sin(\theta_i)) \tag{3.1}$$

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

3.2.2 Statistical model

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

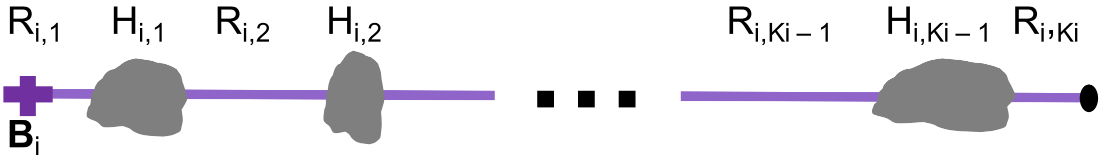


Figure 3.4: Illustration of a hypothetical line segment, L_i that crosses over multiple sections of land. Line L_i starts at point B_i, denoted by a '+' sign, and ends at the black circle. The j-th line segment crossing ocean in L_i is denoted by $R_{i,j}$ and the j-th line segment crossing land is denoted by $H_{i,j}$. There are k_i sections crossing ocean and $K_i - 1$ sections crossing land.

We now introduce notation for modeling proportions. These variables are illustrated in Figure 3.4. Let $\boldsymbol{R}_i = \{R_{i,1}, \ldots, R_{i,K_i}\}$ denote the K_i line segments that form L_i and let $\boldsymbol{H}_i = \{H_{i,1}, \ldots, H_{i,K_i-1}\}$ denote the $K_i - 1$ line segments where land is crossed. Note that

$$||L_i|| = \sum_{k=1}^{K_i} ||R_{i,k}|| + \sum_{k=1}^{K_i-1} ||H_{i,k}|| = ||\boldsymbol{R}|| + ||\boldsymbol{H}||. \tag{3.2}$$

In the common case where L_i just goes through ocean, $\boldsymbol{R} = R_{i1} = L_i$ and $H = \emptyset$.

Since the line segments forming any H_i cannot contain sea ice, we focus on modeling the proportion of the corresponding R_i that are ice-covered. More formally, let

$$\pi_i = \frac{||y_i \cap \boldsymbol{R}_i||}{||\boldsymbol{R}_i||}, \tag{3.3}$$

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

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

$$\tilde{\pi}_i = \begin{cases} \text{logit}(\pi_i) & \text{for } \epsilon \leq \pi_i \leq 1 - \epsilon \\ \text{logit}(\epsilon) & \text{for } \pi_i < \epsilon \\ \text{logit}(1-\epsilon) & \text{for } \pi_i > 1 - \epsilon, \end{cases} \tag{3.4}$$

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

$$\tilde{\boldsymbol{\pi}} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma}). \tag{3.5}$$

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 $\tilde{\boldsymbol{S}}$ 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 via

$$\pi_i = \ \text{ilogit}(\tilde{\pi}_i), \tag{3.6}$$

where $\text{ilogit}(x) = \exp(x)/(1+\exp(x))$. The length of the corresponding y_i can be computed from π_i as follows. Let D denote the maximum index of a line segment of R_i that is fully ice-covered when a proportion π_i of $\boldsymbol{R}_i$ is ice-covered:

$$D = \underset{d}{\text{argmin}} \left\{ \frac{\sum_{k=1}^{d} ||R_{i,k}||}{||\boldsymbol{R}_i||} < \pi_i \right\}. \tag{3.7}$$

Then, the length of y_i is

$$||y_i|| = \pi_i ||\boldsymbol{R}_i|| + \sum_{k=1}^{D-1} ||H_{i,k}||. \tag{3.8}$$

In other words, y_i is composed of a proportion π_i of $\boldsymbol{R}_i$ and all the line segments of $\boldsymbol{H}_i$ that must be crossed to reach the D-th segment of $\boldsymbol{R}_i$. For all i, the lengths of y_i are then used to compute $\boldsymbol{s}_i$ using Equation 3.1. Connecting the points in $\tilde{\boldsymbol{S}}$, along with the points in $\boldsymbol{B}$, if applicable, produces a generated contour.

In rare cases, the generated values in $\tilde{\boldsymbol{S}}$ 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. In such cases, a small adjustment is made with the Douglas-Peuker algorithm to the part(s) of the contours that have self-intersections. The Douglas-Peucker algorithm takes as input a line represented as a connected sequence of points and returns a new line that approximates the original line with a different connected sequence of points. The new line uses as few points as possible to approximate the old line while ensuring that the new line differs from the old line by no more than a distance of η [Douglas and Peucker, 1973]. To correct self-intersections, the Douglas-Peucker algorithm is initially applied with a small η to the part of a contour with a self-intersection. If the self-intersection is removed, the new line is used in place of the old line. If not, η is increased and the algorithm is reapplied. This process is repeated until an η is found that produces a contour with no self-intersections. In our context, these adjustments typically have minimal effect on the line itself and the area contained within the contour.

We also adjust a small percent of generated $||y_i||$ values that correspond to locations where the generated contour is very close to touching a region boundary or land (within 12.5 nominal kilometers, or half the nominal length of a grid box). In such cases, our statistical model is predicting that the contour comes very close to, but does not actually touch, a land or region boundary. This behavior is physically implausible near land, so we adjust these $||y_i||$ values to exactly align with the region or land boundary. The total area involved in this adjustments is very small. This step ensures that any individual generated contour looks physically realistic.

3.2.3 Parametric covariance

To allow for efficient fitting of Equation 3.5, 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 all regions except the Central Arctic based on the differences between the indices of the lines in $\boldsymbol{L}$.

Outside the Central Arctic, we let $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}(\boldsymbol{\sigma}, \kappa)$ where $\boldsymbol{\sigma} = (\sigma_1, \ldots, \sigma_n)$ and $\kappa > 0$. The element, $\boldsymbol{\Sigma}_{ij}$, in the i-th row and j-th column of this covariance is

$$\boldsymbol{\Sigma}_{ij} = \sigma_i \sigma_j \exp\left(-\frac{|i-j|}{\kappa}\right). \tag{3.9}$$

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. Then the difference between the indices of line i and line j does not correspond to the distance between lines L_i and L_j. So we apply an alternative covariance function based on the difference between angles θ_i and θ_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 $\boldsymbol{\sigma} = (\sigma_1, \ldots, \sigma_n)$, $\kappa > 0$, and the element in the i-th row and j-th column of the covariance matrix is

$$\boldsymbol{\Sigma}_{ij} = \sigma_i \sigma_j \exp\left(-\frac{d(\theta_i, \theta_j)}{\kappa}\right), \tag{3.10}$$

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

We find that an exponential covariance structure fits the data well. In particular, this covariance structure allows for the correlation to drop off rapidly as lines become farther apart, while maintaining some non-zero 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.

3.2.4 Number of lines in $\boldsymbol{L}$

Setting N, the number of lines in $\boldsymbol{L}$, 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. Insight into this tradeoff can be obtained by considering how well a set of observed contours can be approximated with only points on lines in $\boldsymbol{L}$. The mean difference in area between observed contours and their approximations provides an estimate of the expected area that a generated contour built with N lines cannot represent.

In this implementation, we set N based on observations of the largest region (the Central Arctic) in the forecast month with the highest variability (September). Using data from 1995-

2004, we find that approximating the observed September contours with points on $N = 90$ lines results in area differences of approximately 2.5% of the total area while maintaining feasible computation times. We set the N for other regions in proportion to how their area compares to the area of the Central Arctic. Using higher N in general leads to slightly more accurate forecasts, since this difference in area can be modestly reduced. Moderately lowering N in general will have the opposite effect. However, for small adjustments, e.g. adding or removing 5 to10 lines, this pattern may not hold due to sampling error.

3.2.5 Prior distribution of the mean sea ice edge, $\boldsymbol{\mu}$

We take a Bayesian approach to parameter estimation. We place a strong prior on $\boldsymbol{\mu}$, the mean sea ice edge, because the ensemble forecast provides considerable information about the likely location of the mean sea ice edge. We first reduce the systematic errors in the ensemble using Contour Shifting Director et al. [2017a], a bias-reduction method for ensemble forecasts that we now review.

Contour-Shifting

Director et al. [2017a] developed Contour Shifting for the Central Arctic region, and here we extend it to all regions. We summarize the method using the notation in Section 3.2.

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, $\tilde{S}^{obs}_{i,j}$. 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, $\tilde{S}^{ens}_{i,j}$. We denote these lengths by $||y^{obs}_{i,j}||$ and $||y^{ens}_{i,j}||$ respectively. Assuming linear change in these length 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,

$$||\hat{y}_{i,t}^{obs}|| = \hat{\alpha}_i^{obs} + \hat{\beta}_i^{obs} t, \tag{3.11}$$

and the ensemble mean ice edge,

$$||\hat{y}_{i,t}^{ens}|| = \hat{\alpha}_i^{ens} + \hat{\beta}_i^{ens} t. \tag{3.12}$$

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

The difference between $||\hat{y}_{i,t}^{obs}||$ and $||\hat{y}_{i,t}^{ens}||$ 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

$$||\hat{y}_{i,t}^{CS}|| = ||y_{i,t}^{ens}|| + (||\hat{y}_{i,t}^{obs}|| - ||\hat{y}_{i,t}^{ens}||), \tag{3.13}$$

where the superscript CS indicates that a Contour Shifting adjustment has been made.

Each adjusted length, $||y_{i,t}^{CS}||$, can be combined with the corresponding B_i and L_i value as in Section 3.2.1 to produce new ice edge contours. The resulting contours may have self-intersections at a small number of locations. These self-intersections can be corrected with an iterative application of the Douglas-Peuker algorithm as described in Section 3.2.1. Contours from ensembles that have been adjusted in this way are referred to as Contour-Shifted and have reduced systematic error compared to initial ensemble forecasts.

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,

$$\boldsymbol{\mu} \sim N(\boldsymbol{\mu}_0, \boldsymbol{\Lambda}_0), \tag{3.14}$$

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

$$\pi_{i,t}^{CS} = \frac{||y_{i,t}^{CS} \cap \boldsymbol{R}_i||}{||\boldsymbol{R}_i||}. \tag{3.15}$$

denote the proportion of $\boldsymbol{R}_i$ that $y_{i,t}^{CS}$ covers in the Contour-Shifted ensemble mean ice edge. Then for all i, let

$$\mu_{0,i} = \begin{cases} \text{logit}\left(||\pi_{i,t}^{CS}||\right) & \text{for } \epsilon \leq \pi_{i,t}^{CS} \leq 1-\epsilon \\ \text{logit}(\epsilon) & \text{for } \pi_{i,t}^{CS} < \epsilon \\ \text{logit}(1-\epsilon) & \text{for } \pi_{i,t}^{CS} > 1-\epsilon. \end{cases} \tag{3.16}$$

Also, the matrix $\boldsymbol{\Lambda}_0$ is a $n \times n$ diagonal covariance matrix with

$$\boldsymbol{\Lambda}_{0,ii} = \frac{\text{logit}(\max(\pi_{i,t}^{CS} - 0.125, \epsilon)) - \text{logit}(\min(\pi_{i,t}^{CS} + 0.125, 1-\epsilon)))/2}{\Phi^{-1}(.995)}. \tag{3.17}$$

This prior covariance treats all μ_i as independent. The variance for each μ_i is equivalent to the variance that would be obtained with 99% of the distribution's mass in the logit-transformed interval of $[\max(\pi_{i,t}^{CS} - 0.125, \epsilon), \min(\pi_{i,t}^{CS} + 0.125, 1-\epsilon)]$. This prior variance for $\mu_{0,i}$ typically corresponds to the mean proportion being within 0.125 of the Contour-Shifted ensemble mean proportion, $\pi_{i,t}^{CS}$. The variance is reduced if $\pi_{i,t}^{CS}$ is close to 0 or 1. (See Appendix A.1 for the derivation of the standard deviation of a normal distribution that corresponds to a particular proportion of the mass of the distribution being within a set of bounds. In this case, $M = \text{logit}(\max(\pi_{i,t}^{CS} - 0.125, \epsilon))$, $m = \text{logit}(\min(\pi_{i,t}^{CS} + 0.125, 1-\epsilon))$, and $\gamma = 0.99$.)

3.2.6 Prior for $\boldsymbol{\Sigma}$

For the prior on $\boldsymbol{\Sigma}$ we only use information about the physical constraints. While ensembles have the potential to provide information about covariance, the variability of the ensembles we have analyzed do not align with the variability seen in observations. As such, we use only physical constraints to inform our priors for the covariance parameters, $\boldsymbol{\sigma}$ and κ.

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 σ_i such that

$$\sigma_{0,i} \overset{iid}{\sim} \text{Unif}(\alpha_{\sigma,0}, \beta_{\sigma,0}), \tag{3.18}$$

where $\alpha_{\sigma,0} = 0.01$. We bound σ at $\alpha_{\sigma,0}$ rather than zero to avoid numerical issues when sampled σ values approach zero. We let

$$\beta_{\sigma_{0,i}} = \frac{(\text{logit}(\delta_2) - \text{logit}(\delta_1))/2}{\Phi^{-1}(.995)}. \tag{3.19}$$

where typically $\delta_1 = \epsilon$ and $\delta_2 = 1 - \epsilon$. This upper bound corresponds to the standard deviation of a Gaussian distribution with 99% of the distribution's mass in the interval $(\text{logit}(\delta_1), \text{logit}(\delta_2))$. This prior distribution ensures that the variance of the transformed proportion of ice-covered length does not substantially exceed the variance of a normal distribution that fully covers the interval of possible proportion values. (This bound is obtained using Appendix A.1 with $M = \text{logit}(\delta_2)$, $m = \text{logit}(\delta_1)$, and $\gamma = 0.99$.)

Exceptions to the typical δ_1 and δ_2 values are made in the Central Arctic region where $\delta_1 = 0.15$ and in the Greenland sea region where $\delta_2 = 0.73$. These exceptions reflect the fact that the ice-covered proportions in these regions have never covered the full interval $[\epsilon, 1-\epsilon]$ for any L_i in the training observations. Even at the annual minimum, lines in the Central Arctic have never had ice coverage proportions near zero. Similarly, even at the annual maximum, lines in the Greenland sea have never had ice coverage proportions exceeding 0.73.

With little information from which to anticipate how correlation decreases with distance, we use the following vague prior for κ,

$$\kappa_0 \sim \text{Unif}(\alpha_{\kappa,0}, \beta_{\kappa,0}), \tag{3.20}$$

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

3.2.7 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 from the distribution of contours from which the forecast year's contour will be drawn. With this approach we are assuming 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 $j = \{1, 2, \dots, P\}$. 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 $\boldsymbol{R}_i$ that y_{ij} covers in year j.

Combining the likelihood for the observed proportions with the prior distributions introduced in Sections 3.2.5 and 3.2.6 gives the posterior distribution

$$\begin{aligned}\prod_{j=1}^{P} \left\{ p(\tilde{\boldsymbol{\pi}}_j, \boldsymbol{\mu}, \boldsymbol{\sigma}, \kappa) \right\} p(\boldsymbol{\mu}) p(\boldsymbol{\sigma}) p(\kappa) = &\prod_{j=1}^{P} \left\{ \mathrm{N}(\tilde{\boldsymbol{\pi}}_j | \boldsymbol{\mu}, \boldsymbol{\Sigma}(\boldsymbol{\sigma}, \kappa)) \right\} \mathrm{N}(\boldsymbol{\mu} | \boldsymbol{\mu}_0, \boldsymbol{\Lambda}_0) \times \\ &\prod_{i=1}^{n} \left\{ \mathrm{Unif}(\sigma_i | \alpha_{\sigma,0}, \beta_{\sigma,0}) \right\} \mathrm{Unif}(\kappa | \alpha_{\kappa,0}, \beta_{\kappa,0}). \qquad (3.21)\end{aligned}$$

The posterior means of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ can be used with Equations 3.5, 4.3, and 3.10 to generate $\tilde{\boldsymbol{\pi}}$.

3.2.8 Model fitting

We sample from the posterior distribution in Equation 3.21 for each region independently with Markov chain Monte Carlo (MCMC), using the observed sea ice in the preceding P years. MCMC diagnostics are given in Appendix A.3. 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 complete 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 κ due to perfect correlation among the lines in these sections. In the Central Arctic sets of lines bordering the Canadian Archipelago that have proportion 1 for all P years are similarly fixed.

We use Metropolis steps for updating each μ_i, σ_i, and κ. Normal proposals are used for each parameter at each iteration centered at their current value. For each element i and iteration t, the log acceptance ratio for $\mu_i^{(t)}$ is

$$\begin{aligned}&-\frac{1}{2}\sum_{j=1}^{P}\left(\tilde{\boldsymbol{\pi}}_j-\boldsymbol{\mu}^{(t)}\right)^T\boldsymbol{\Sigma}^{-1}(\tilde{\boldsymbol{\pi}}_j-\boldsymbol{\mu}^{(t)})-\frac{1}{2}(\boldsymbol{\mu}^{(t)}-\boldsymbol{\mu}_0)^T\boldsymbol{\Lambda}^{-1}(\boldsymbol{\mu}^{(t)}-\boldsymbol{\mu}_0)\\&+\frac{1}{2}\sum_{j=1}^{P}(\tilde{\boldsymbol{\pi}}_j-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\tilde{\boldsymbol{\pi}}_j-\boldsymbol{\mu})+\frac{1}{2}(\boldsymbol{\mu}-\boldsymbol{\mu}_0)^T\boldsymbol{\Lambda}^{-1}(\boldsymbol{\mu}-\boldsymbol{\mu}_0),\end{aligned}\tag{3.22}$$

where $\boldsymbol{\mu}^{(t)}$ denotes the $\boldsymbol{\mu}$ vector on the t-th iteration with the i-th element proposed. For each element i and iteration t, the log acceptance ratio for $\sigma_i^{(t)}$ is

$$\begin{aligned}&-\frac{n}{2}\log|(\boldsymbol{\Sigma}^{(t)})|-\frac{1}{2}\sum_{j=1}^{P}\left(\tilde{\boldsymbol{\pi}}_j-\boldsymbol{\mu}\right)^T(\boldsymbol{\Sigma}^{(t)})^{-1}(\tilde{\boldsymbol{\pi}}_j-\boldsymbol{\mu})+\mathbb{1}[\sigma_i^{(t)}\in(\alpha_{\sigma,0},\beta_{\sigma,0})]\\&+\frac{n}{2}\log|\boldsymbol{\Sigma}|+\frac{1}{2}\sum_{j=1}^{P}\left(\tilde{\boldsymbol{\pi}}_j-\boldsymbol{\mu}\right)^T(\boldsymbol{\Sigma}^{(t)})^{-1}(\tilde{\boldsymbol{\pi}}_j-\boldsymbol{\mu}),\end{aligned}\tag{3.23}$$

where $\sigma_i^{(t)}$ denotes the proposal for the i-th element of $\boldsymbol{\sigma}$ on the t-iteration and $\Sigma^{(t)}$ denotes the corresponding covariance matrix with $\sigma_i^{(t)}$ proposed. On the t-th iteration, the log acceptance ratio for $\kappa^{(t)}$ is

$$\begin{aligned}&-\frac{n}{2}\log|(\boldsymbol{\Sigma}^{(t)})|-\frac{1}{2}\sum_{j=1}^{P}\left(\tilde{\boldsymbol{\pi}}_j-\boldsymbol{\mu}\right)^T(\boldsymbol{\Sigma}^{(t)})^{-1}(\tilde{\boldsymbol{\pi}}_j-\boldsymbol{\mu})+\mathbb{1}[\kappa^{(t)}\in(\alpha_{\kappa,0},\beta_{\kappa,0})]\\&+\frac{n}{2}\log|\boldsymbol{\Sigma}|+\frac{1}{2}\sum_{j=1}^{P}\left(\tilde{\boldsymbol{\pi}}_j-\boldsymbol{\mu}\right)^T(\boldsymbol{\Sigma}^{(t)})^{-1}(\tilde{\boldsymbol{\pi}}_j-\boldsymbol{\mu}),\end{aligned}\tag{3.24}$$

where $\kappa^{(t)}$ denotes the proposal for κ on the t-th iteration and $\boldsymbol{\Sigma}^{(t)}$ now denotes the corresponding covariance matrix with $\kappa^{(t)}$ proposed.

3.3 Mixture Contour Forecasting

The contour model generally provides reasonable forecasts of the sea ice edge contour, but does have some weaknesses. The first is that 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 still form away from this main area. Areas of open water, called polynyas, also sometimes form as holes within the main sea ice area. The contour model proposed in the previous section cannot represent these features. Secondly, 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 of ice contours 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 advantages and disadvantages. The climatology forecast can represent features such as holes in the sea ice or 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 edge configurations. 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 whether sea ice was present in observations for some grid box s and year t in the training period. Let $g_p(\gamma_{s,t})$ and $g_c(\gamma_{s,t})$ be the estimated 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 estimated probability is the proportion of the time that grid box s is within the area enclosed by the generated contours for time t. The predicted probability of sea ice presence at grid box s at time t is then

$$p(\gamma_{s,t}) = wg_p(\gamma_{s,t}) + (1-w)g_c(\gamma_{s,t}) \tag{3.25}$$

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

$$l(w) = \sum_t \sum_s \log\{w\, a_s g_p(\gamma_{s,t}) + (1-w)a_s g_c(\gamma_{s,t})\}. \tag{3.26}$$

The variable a_s is the proportion of the entire area that is in grid box s, i.e, $\sum_s a_s = 1$. The use of a_s accounts for the fact that the grid boxes do not all have the same area. Assuming spatial and temporal independence is almost certainly inaccurate; however, 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 et al., 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 straightforward 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

$$\hat{z}^{(j)}_{p,s,t} = \frac{w^{(j-1)} a_s g_p(\gamma_{s,t})}{w^{(j-1)} a_s g_p(\gamma_{s,t}) + (1-w^{(j-1)}) a_s g_c(\gamma_{s,t})}, \tag{3.27}$$

and the M-step is

$$w^{(j)} = \frac{\sum_t \sum_s a_s \hat{z}^{(j)}_{p,s,t}}{\sum_t \sum_s a_s} \tag{3.28}$$

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 3.28 may be unequal to the number of years in the training period.

3.4 Method evaluation

3.4.1 Model outputs and observations

All post-processing 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 sea ice concentration 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 post-processing, ECMWF has been shown to be generally the most skillful [Zampieri et al., 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 25km by 25km grid [National Snow and 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 sea ice concentration produced from the National Aeronautics and Space Administration satellites Nimbus-7 SMMR and DMSP SSM/I-SSMIS. This data can be downloaded from the National Snow and Ice Data Center [Comiso, 2000, updated 2017]. Grid boxes with sea ice concentrations of at least 15% are treated as having sea ice present. Otherwise grid boxes are treated as not containing sea ice.

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 and Raftery, 2020] are treated as land.

The forecasts previously described are summarized in lines 1-4 of Table 3.1. Beginning in 1993, all years preceding the forecast year are used in fitting Contour Shifting. A ten-year rolling window is used to fit the statistical model for generating contours and in the climatology forecast weighted in MCF. The time length of ten years is used, since recent analyses have shown that climatology computed over a ten-year period provides reasonably accurate sea ice forecasts. Such forecasts become nearly as accurate as ECMWF ensemble forecasts at lead times of 1.5 months [Zampieri et al., 2018]. Slight changes in the number of years used for this purpose is unlikely to affect the results. However, using many fewer years would not provide enough samples to fit the parameters accurately. Using a much bigger number of years would also degrade performance. Because of the rapid reduction in Arctic sea ice area, contours from past decades will differ notably from recent ice edge contours. A three-year rolling window is used to determine the weights in MCF. Performance accuracy is generally insensitive to this choice (see Appendix A.4). One hundred contours are generated for each forecast.

3.4.2 Reference forecasts

We compare our results to two additional reference forecasts summarized in lines 5-6 of Table 3.1. Trend Adjusted Quantile Mapping (TAQM) is another recently developed statistical post-processing method for sea ice [Dirkson et al., 2019b]. TAQM fits a parametric probability distribution to ensemble model output and applies a specialized version of quantile mapping to produce probabilistic forecasts of sea ice concentration. TAQM does not predict the probability of sea ice presence directly, but Dirkson et al. [2019b] do use the resulting distribution of the sea ice concentration to predict the probability of sea ice presence (concentration of at least 15%).

We also compute a damped persistence forecast in a manner similar to Wayand et al. [2019]. Damped persistence forecasts estimate the sea ice concentration in forecast month m using linear regression and the observed sea ice concentration in the initialization month i. For each grid box, the concentration for month m in year t, denoted by $C_{m,t}$, is estimated as

$$\hat{C}_{m,t} = \hat{\beta}_m t + (C_{i,t_i} - \hat{\beta}_i t_i)\hat{\rho}, \tag{3.29}$$

Table 3.1: 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 ten years preceding the forecast year that contain sea ice	indicator of whether at least five of the ten 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 post-processed using technique in Dirkson et al. [2019b]	NA
Damped Persistence	NA	indicator of whether predicted sea ice concentration from a damped persistence forecast is at least 0.15 (modified from Wayand et al. [2019])

where C_{i,t_i} is the observed concentration in the i-th initialization month, $\hat{\beta}_m$ and $\hat{\beta}_i$ are the coefficients for linear regressions of C_m and C_i on year, and $\hat{\rho}$ is the empirical correlation of C_m and C_i estimated from past observations. If all values of C_m and/or C_i in the training period are the same, the empirical correlation is undefined. In these cases, we set $\hat{\rho}$ to 0, which makes Equation 3.29 equivalent to linear regression. When $i \leq m$, $t_i = t$, otherwise $t_i = t - 1$. Grid boxes with predicted concentration of at least 0.15 are forecasted to contain sea ice. Observations beginning in 1981 and extending up to the initialization time are used in fitting.

3.4.3 Visualizing forecasts

Uncertainty information is needed for maritime planning to adequately evaluate risks and benefits. Like Gneiting et al. [2007], we consider accurate model calibration to be vital for probabilistic forecasts. We illustrate the importance of calibration in this context with Figure 3.5, which shows samples of four probabilistic forecasts for September 2008. The corresponding observed sea ice edge for September 2008 is also plotted for reference. Figure 3.5 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 more substantial. For the climatology and ensemble forecasts, the observed contour goes through some regions with zero probability, suggesting that these forecasts are not sufficiently variable. For these forecasts, there are areas where sea ice is predicted with probability 0, but sea ice is observed. Discrepancies like these

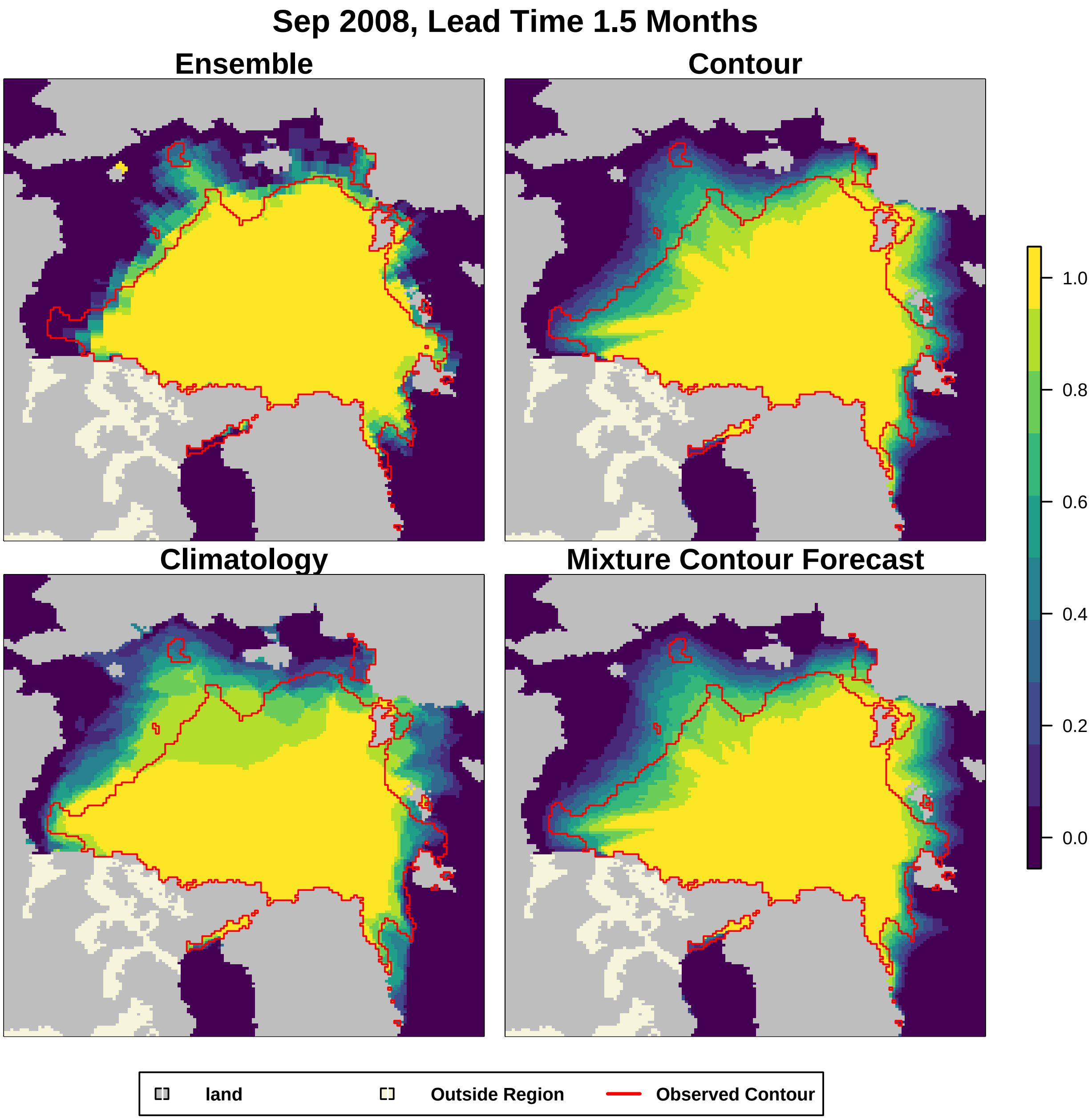


Figure 3.5: Forecasts of the probability of sea ice presence for September 2008 using different methods. The forecasts are described in Table 3.1. The red line is the observed sea ice edge contour and grey areas are land. For MCF, the observed ice edge is almost completely within areas with positive probability and has little area where sea ice is was predicted with probability 1 but sea ice was not present in observations. In contrast, the observed sea ice edge more often goes through regions with zero probability in the ensemble and climatology forecasts. In the climatology forecast, there is also the most area where sea ice is predicted with probability 1 but is not present in observations.

between the forecasted probability and what will likely occur makes maritime planning and risk mitigation with these latter types of forecasts difficult.

3.4.4 Assessing Calibration

We now evaluate model calibration for the probabilistic forecasts. We evaluate calibration with reliability diagrams that 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, i.e. grid boxes forecasted to contain sea ice with a given probability actually 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], so we emphasize performance in these peak shipping months. In Figure 3.6 we show the reliability diagrams for the peak-shipping months for the probabilistic forecasts. Predictions from MCF are substantially better calibrated than the ensemble and better calibrated than TAQM during these months, especially at short lead times. Figures A.4 and A.5 in Appendix A.2 show that MCF always improves calibration over the unadjusted ensemble and generally improves calibration compared to TAQM. We note that MCF has been designed specifically for predicting sea ice presence, while TAQM addresses the more general goal of forecasting sea ice 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.

3.4.5 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

$$\frac{\sum_t \sum_s a_s (f_{s,t} - o_{s,t})^2}{T}, \tag{3.30}$$

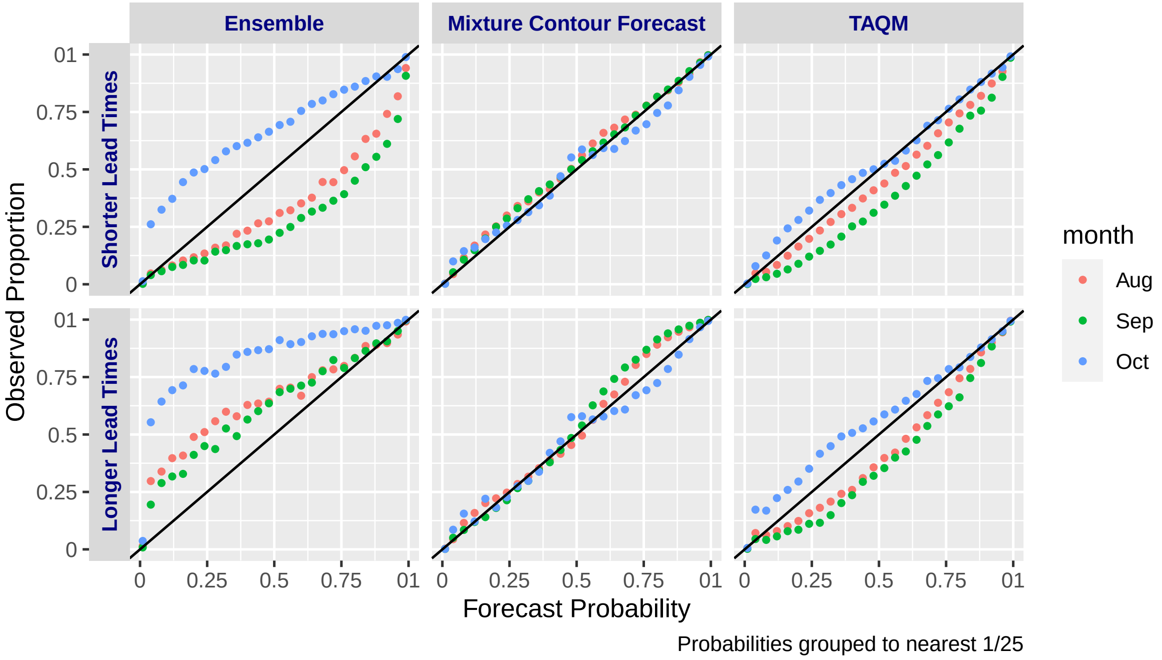


Figure 3.6: The average proportion of the time sea ice was present plotted against the predicted probability of sea ice presence for the raw ECMWF forecasts (left), after post-processing with MCF (middle), and TAQM (right). Forecasts are grouped into lead times of 0.5 and 1.5 months (top) and 2.5-6.5 months (bottom). A perfectly calibrated forecast would have all points on the diagonal $y = x$ line.

where $f_{s,t}$ and $o_{s,t}$ denote the forecast and observation in grid box s in year t respectively. The value a_s is the proportion of the total area that is in grid box s. The observed value is 1 when the sea ice 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 3.7, 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. As lead time increases, MCF's performance converges to equal or better performance than climatology. TAQM also generally improves accuracy of forecasts.

Figures A.1 and A.2 in Appendix A.2 show that TAQM and MCF have similar overall

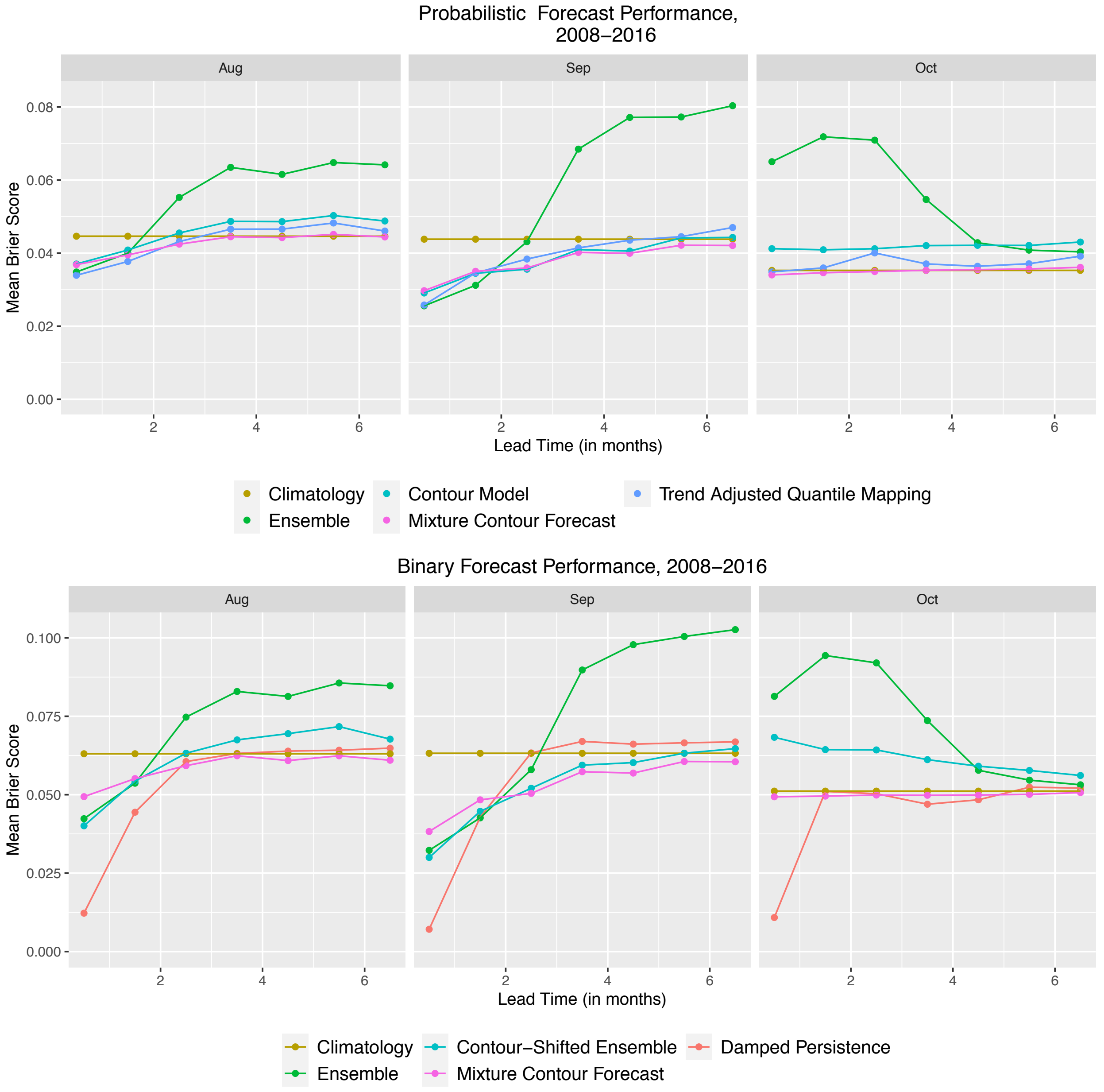


Figure 3.7: Top: Average Brier scores by month for the test years 2008-2016 for the probabilistic forecasts. The Brier Score for each grid box is weighted based on its area. Forecasts are described in Table 3.1. Bottom: As above, but for the binary forecasts.

accuracy, but that the pattern of their performance by lead time and month varies. For peak shipping months, MCF outperforms TAQM, suggesting that our specialized modeling of the sea ice edge has benefits for maritime planning use. For other applications, more general techniques like TAQM may be more appropriate. 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.

3.4.6 Binary Forecasts

We also briefly assess binary forecasts with the bottom panel in Figure 3.7. Binary forecasts are inherently poorly calibrated, and so are not optimal, but can be useful in method assessment. The Contour-Shifted forecast clearly improves accuracy compared to the ensemble. In other words, Contour Shifting does reduce some systematic bias that affects typical ensembles. Binary MCF performs similarly to the Contour-Shifted ensemble in general, but 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. Brier scores for binary forecasts for all seasons are in Figure A.3 in Appendix A.2.

3.4.7 Understanding mixture weights

Since MCF is fitted separately for each forecast month and lead time, we can examine how the weights on the contour model and climatology forecast vary between months and lead times. Figure 3.8 shows the average weight placed on the contour model for the years in the test set. High weights typically occur at short lead times, reflecting the fact that the ensemble typically has the most skill soon 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

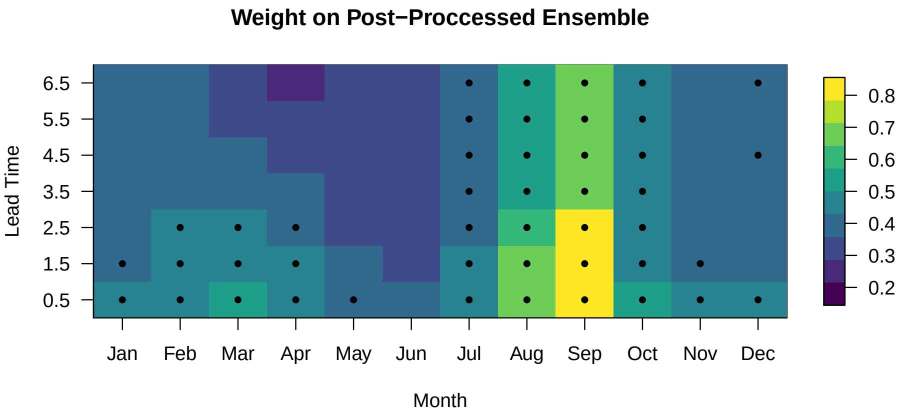


Figure 3.8: Weight on the contour model by month and lead time. A black dot indicates that the weight is at least 0.4. The contour gets more weight at short lead times and in months near the sea ice minimum. The weight on climatology is equal to one minus the weight on the contour model.

variability, so climatology tends to perform poorly and the ensemble's ability to simulate evolving physical conditions becomes more important.

3.5 Discussion

We have introduced the Mixture Contour Forecasting method for issuing probabilistic sea ice forecasts. MCF forecasts are probabilistic and well calibrated, meaning that 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. At most lead times and forecast months, probabilistic MCF forecasts are also as or more accurate than the raw ECMWF ensemble and the other post-processed and statistical forecasts.

Because MCF provides well calibrated and relatively accurate forecasts, MCF's use has the potential to 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 that says where they can legally and safely travel. For vessels that are easily damaged in sea ice, encountering 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 face danger if they encounter it.

Our model evaluation weights both types of misclassification errors equally (predicting the presence of sea ice when it was not observed and predicting the absence of sea ice when it was observed.) However, the probabilistic forecasts provided by MCF do allow us to account for different costs of the two types of error. In particular, MCF opens up the possibility of 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 advantageous. MCF provides a framework for modeling that could be extended to other situations where the boundary is of interest.

As implemented in this chapter, 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 et al., 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 align with observed variability and the relationship between variability in observations and variability in the ensemble is inconsistent both spatially and temporally. Thus 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 3.4 is not the only ensemble prediction system. The post-processing techniques developed in this chapter could be directly applied to other ensembles, since they do not use any specific features of the ECMWF ensemble. However, model biases and calibration issues vary, so exact performance would need to be assessed. Different ensembles also vary by which forecast months they perform well in, and vary more by the extent to which skill declines with lead time [Zampieri et al., 2018]. Thus extending MCF to use multiple ensemble members as has been done for other meteorological variables could provide further skill [e.g., Raftery et al., 2005, Dirkson et al., 2019a].

Our analysis of sea ice forecasting highlights situations where statistical post-processing can provide value. Many aspects of physical processes are known to evolve following well established equations. Such information can only be crudely approximated with a purely 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.

Chapter 4

CONTOUR MODELS FOR BOUNDARIES ENCLOSING STAR-SHAPED AND APPROXIMATELY STAR-SHAPED POLYGONS

Boundaries on spatial fields divide regions with particular features from surrounding background areas. These boundaries are often described with contour lines. To measure and record these boundaries, contours are often represented as ordered sequences of spatial points that connect to form a line. While methods to identify boundary lines from interpolated spatial fields are well-established, less focus has been placed on modeling data consisting of multiple sequences of spatial points. For data of the latter form, we introduce the Gaussian Star-Shaped Contour Model (GSCM). Sequences of connected spatial points are generated via a model for the distance to these points from a fixed starting point. The GSCM is designed for modeling contours that enclose regions that are star-shaped polygons or approximately star-shaped polygons. Metrics are introduced to assess this assumption. Simulation studies illustrate the performance of GSCM under various scenarios. An analysis of Arctic sea ice edge contour data highlights how GSCMs can be applied to observational data.

4.1 Introduction

Boundaries that enclose regions are often subjects of scientific interest. Contour lines typically divide a contiguous region with some defining feature from surrounding background areas. Many contour models have focused on inferring a boundary for a spatial process observed at various spatial points. A less studied problem is how to infer the distribution of contours from multiple fully-observed contours. Specifically, we focus on data comprised of multiple sets of spatial points in the 2-D plane that connect in order to form contours. In this context, uncertainty in the contours can be inferred from observing multiple instances of the boundary line, such as contours observed at multiple time points.

This type of data can also be represented by grids of binary values that indicate if each grid box is inside or outside the boundary. The associated contours are then the ordered sequences of points that form the boundary between the grid boxes inside and outside the region. In this chapter, we introduce the Gaussian Star-Shaped Contour Model (GSCM) for inferring distributions of contours from data sets of this form. GSCMs are designed for modeling contours that enclose regions that are star-shaped (Definition 4) or approximately star-shaped polygons

As a motivating example, we consider the Arctic sea ice edge contour. Figure 4.1 shows a sample sea ice edge contour. The sea ice edge contour forms a boundary between the area covered by sea ice and the surrounding open water. For sea ice edge contours, questions of both inference and prediction are relevant. Predictions of the ice edge contour are needed weeks to months in advance for maritime planning, since vessels seek to avoid sea ice-covered areas that are costly to navigate. Polar scientists are also interested in where ice edge contours are more and less variable and the extent to which ice edge contours change over time.

Previous research has developed contour and boundary models for a number of other data types. Analysis has often focused on inferring a single boundary from observations on a spatial field. Research on exceedance levels has developed methods to infer contours that describe where a property goes above some level [Bolin and Lindgren, 2015, French and Sain, 2013, French and Hoeting, 2016]. Wombling methods find contours by identifying curvilinear gradients [Womble, 1951, Banerjee and Gelfand, 2006]. Statistical shape analysis [Dryden and Mardia, 2016, Srivastava and Klassen, 2016] provides tools to model boundaries corresponding to particular objects with discernible features. While important statistical techniques, none of these methods infer distributions of contours from multiple observed contours. The GSCM seeks to fill this methodological gap.

In the remainder of the chapter, we develop GSCMs and assess their performance. Section 4.2 more precisely defines contours and how to represent them. Section 4.3 introduces

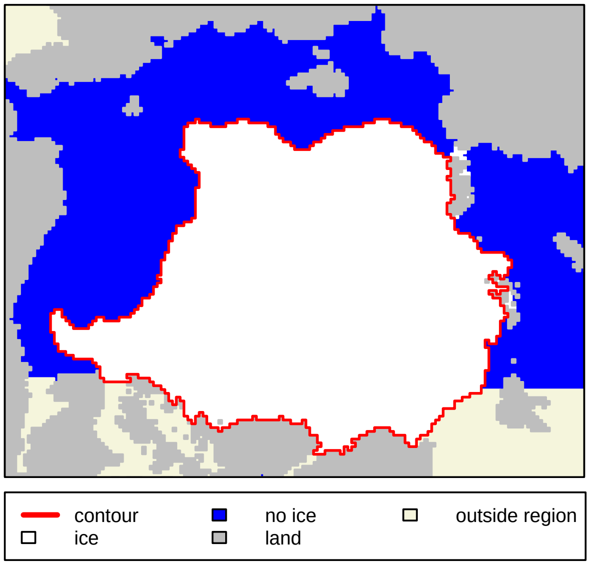


Figure 4.1: Contour forming the boundary around the main contiguous area covered by sea ice in a central region of the Arctic in September 2017. Methods for modeling contours like this one that enclose approximately star-shaped polygons are introduced in Section 4.6.

the GSCM for modeling contours enclosing star-shaped polygons and discusses details of fitting this probability model. Section 4.4 introduces a metric for model assessment and Section 4.5 presents simulation studies. Section 4.6 extends GSCMs to contours enclosing approximately star-shaped polygons and Section 4.7 considers an example with sea ice edge contours. Section 4.8 concludes the chapter with discussion, including further comparison of GSCMs to other contour and boundary methods.

4.2 Contour definitions

Our focus in this chapter is on modeling contours that act as the boundary between a region that has some particular feature and the surrounding background region. There are multiple feasible ways such contours could be defined. In this section, we give two representations for these contours that will be used as a basis for subsequent modeling and assessment.

4.2.1 Point-sequence representation

Contours and the regions they enclose can be described using connected sequences of points. We refer to this description of a contour as a *point-sequence representation.* We define the following concepts.

Definition 1. *Contour point sequence,* $\boldsymbol{S}$*: An ordered set of spatial points* $\boldsymbol{S} = (\boldsymbol{s}_1, \ldots, \boldsymbol{s}_n)$*, with* $n > 2$*, where each* $\boldsymbol{s}_i$ *consists of the x-y coordinates of a spatial location.*

Definition 2. *Contour line,* $\overline{\boldsymbol{S}}$*: The connected line formed by connecting* $\boldsymbol{s}_i$ *to* $\boldsymbol{s}_{i+1}$ *for* $i = 1, \ldots, n-1$ *and connecting* $\boldsymbol{s}_n$ *to* $\boldsymbol{s}_1$*.*

Definition 3. *Enclosed polygon,* $\underline{\boldsymbol{S}}$*: The polygon formed by the interior of the contour line* $\overline{\boldsymbol{S}}$*.*

The left panel of Figure 4.2 illustrates these definitions for a contour described by a point-sequence representation. The main advantage of the point-sequence representation is its flexibility. Any contour enclosing a polygon can be represented exactly with a sequence of spatial points, $\boldsymbol{S}$. Also, the level of detail in the contour can be increased simply by increasing the density of points.

Gridded data describing a contiguous region can also be converted to the point-sequence representation. Consider a binary grid that indicates whether each grid box is inside or outside the region of interest. The corresponding contour $\boldsymbol{S}$ is composed of corner points of grid boxes that touch the outside of the region on one side and its inside on the other. The points of $\boldsymbol{S}$ are ordered to align with the order in which they would be touched if one were to trace around the boundary. Where and in what direction to start tracing is arbitrary, since these choices only determine the indexing of the points in $\boldsymbol{S}$, not the line, $\overline{\boldsymbol{S}}$, or polygon, $\underline{\boldsymbol{S}}$, formed by connecting these points. The point-sequence representation can be used with any grid box resolution without reduced accuracy, though finer grids will require more points in $\boldsymbol{S}$.

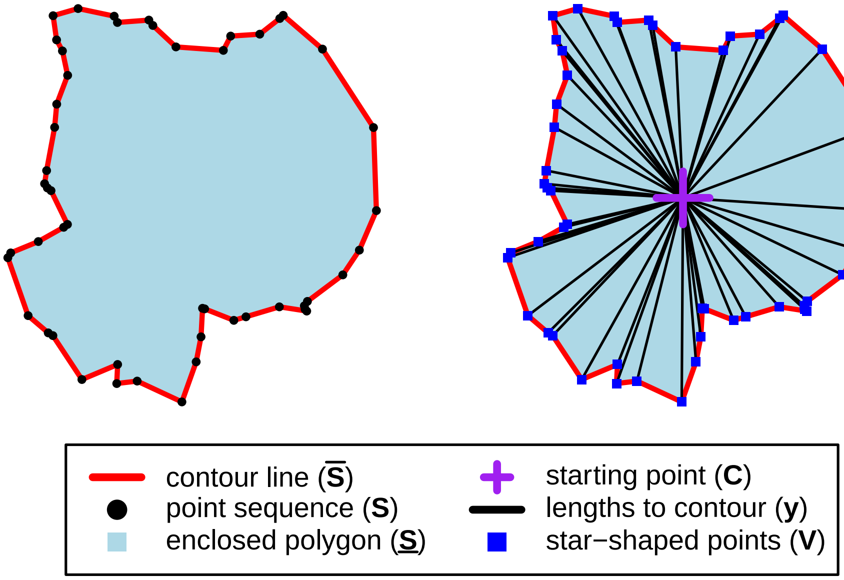


Figure 4.2: Components of a contour represented with a point-sequence representation (left) and a star-shaped representation (right).

Notation

We need to distinguish between points, lines and polygons. An ordered sequence of spatial points will be denoted by a boldface letter, such as $\boldsymbol{S}$. The full line formed by connecting all points in order, $\overline{\boldsymbol{S}}$, will be denoted by an overline. A line segments will be denoted by an overline over two letters that represent the start and end points of the segment, such as $\overline{\boldsymbol{CD}}$. The polygon enclosed by a line, $\underline{\boldsymbol{S}}$, will be denoted by an underline.

Contours with a fractal nature

We acknowledge that the point-sequence representation does not directly account for contours whose true nature is fractal. As such, representing fractal contours as a connected sequence of points may be an approximation. In contours of sea ice and other physical world examples, as the spatial scale of observations increases, the level of detail of the contour also increases [Mandelbrot, 1967]. With each increase in spatial resolution additional line segments are needed to describe the increased detail. In other words, the length of the contour increases each time the spatial scale considered increases. This fractal nature of some contours means that these contour can never be fully expressed with a finite, ordered set of spatial points, since these contours' true lengths are infinite.

For typical statistical applications, however, limits exist on the precision of measurements and the relevant spatial scale of scientific interest. So, while the fractal or Hausdorff dimension can be estimated [Gneiting et al., 2012], the level of detail of the boundary that will be measured and is typically of interest will rarely be a fractal. Therefore for many applications, making the simplifying assumption that the contour at the spatial scale of scientific interest can be defined by a sequence of spatial points $\boldsymbol{S}$ is reasonable. Additional discussion of contours as fractals is given in Section 4.8.2.

4.2.2 Star-shaped representation

Point-sequence representations are natural and describe contours with the highest precision typically available in observations. However, point-sequence representations have weaknesses for describing multiple contours and their distributional behavior. Since contours differ in length, two spatial points on different contours with the same index on their point-sequence representations can refer to very different physical locations. This can make it hard to describe or compare spatially-dependent features, and thus to infer distributions of contours. Instead we build an alternative representation of contours that we will refer to as the *star-shaped representation.* This will be appropriate for contours that enclose star-shaped polygons or approximately star-shaped polygons.

Before defining the star-shaped representation, we review the standard definitions of a star-shaped polygon and its kernel [Preparata and Shamos, 1985, pp. 18].

Definition 4. *Star-shaped polygon: A polygon* $\underline{\boldsymbol{P}}$ *is star-shaped if there exists a point* $\boldsymbol{D}$ *within* $\underline{\boldsymbol{P}}$ *such that the line segment* $\overline{\boldsymbol{Dp}}$ *is fully contained within* $\underline{\boldsymbol{P}}$ *for all points* $\boldsymbol{p}$ *on line* $\overline{\boldsymbol{P}}$*.*

All convex polygons are star-shaped, but the set of star-shaped polygons is substantially larger and encompasses a wide range of shapes. Figure 4.3 shows nine example polygons that are star-shaped.

Definition 5. *Kernel of a star-shaped polygon,* $\mathcal{K}(\underline{\boldsymbol{P}})$*: The set of point(s) that satisfy the criterion for* $\boldsymbol{D}$ *in Definition 4 is referred to as the kernel of the polygon,* $\mathcal{K}(\underline{\boldsymbol{P}})$*.*

The set of convex polygons is the subset of star-shaped polygons such that $\mathcal{K}(\underline{\boldsymbol{P}}) = \underline{\boldsymbol{P}}$.

We now turn to defining the star-shaped representation. For any star-shaped polygon, $\underline{\boldsymbol{S}}$, lines can be drawn from some point, $\boldsymbol{C}$, in the kernel of $\underline{\boldsymbol{S}}$ to all points on a contour, $\boldsymbol{S}$, by the definition of star-shaped. Assume for the moment that the contour point sequence $\boldsymbol{S}$ is unknown, but that the location of $\boldsymbol{C}$ is known along the lengths and directions of the lines from $\boldsymbol{C}$ to $\boldsymbol{S}$. Then $\boldsymbol{S}$ could be derived from this information with trigonometry. Taking inspiration from this fact, we develop the star-shaped representation. First we define a line set:

Definition 6. *Line set,* $\mathcal{L}(\boldsymbol{C}, \boldsymbol{\theta})$*: A set of* p *lines,* $\mathcal{L} = (\ell_1, \ldots, \ell_p)$*, extending infinitely outward from some starting point* $\boldsymbol{C} = (C_x, C_y)$ *at* $p > 2$ *unique angles,* $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_p)$*, where* C_x *and* C_y *are x- and y-coordinates respectively.*

We also define a set of spatial points that produce a star-shaped polygon when connected in order:

Definition 7. *Star-shaped point set,* $\boldsymbol{V}(\boldsymbol{C}, \boldsymbol{\theta}, \boldsymbol{y})$*: A set of* p *spatial points* $\boldsymbol{V} = (\boldsymbol{v}_1, \ldots, \boldsymbol{v}_p)$ *such that*

$$\boldsymbol{v}_i = (C_x + y_i \cos(\theta_i), C_y + y_i \sin(\theta_i)) \tag{4.1}$$

where $\boldsymbol{y} = (y_1, \ldots, y_p)$ *is a set of* p *distances,* $\boldsymbol{C} = (C_x, C_y)$ *is a spatial point, and* $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_p)$ *is a set of* p *unique angles.*

A star-shaped point set can be used to approximate a contour when the distances are selected systematically:

Definition 8. *Star-shaped representation,* $\tilde{\boldsymbol{V}}(\boldsymbol{C}, \boldsymbol{\theta}, \boldsymbol{S})$*: Let* $\underline{\boldsymbol{S}}$ *be a star-shaped polygon and* $\boldsymbol{C} \in \mathcal{K}(\underline{\boldsymbol{S}})$ *be a starting point. Then, the star-shaped representation of the contour* $\boldsymbol{S}$*, denoted by* $\tilde{\boldsymbol{V}}(\boldsymbol{C}, \boldsymbol{\theta}, \boldsymbol{S})$*, is the star-shaped point set,* $\boldsymbol{V}(\boldsymbol{C}, \boldsymbol{\theta}, \boldsymbol{y})$*, where* $\boldsymbol{y} = (y_1, \ldots, y_p)$ *is the set of distances from* $\boldsymbol{C}$ *to the intersection point of the contour line* $\overline{\boldsymbol{S}}$ *and each line* ℓ_i *in the line set* $\mathcal{L}(\boldsymbol{C}, \boldsymbol{\theta})$*.*

The right panel of Figure 4.2 shows the components of the star-shaped representation for a sample contour. Let $\overline{\boldsymbol{V}}$ refer to the contour line formed by connecting $\boldsymbol{v}_i$ to $\boldsymbol{v}_{i+1}$ for $i = 1, \ldots, p-1$ and $\boldsymbol{v}_p$ to $\boldsymbol{v}_1$ for $p > 2$. Let $\underline{\boldsymbol{V}}$ refer to the polygon contained within $\overline{\boldsymbol{V}}$.

Theorem 1. *Let* $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_p)$ *with* $\theta_i < \theta_{i+1}$ *and* $\theta_i \in (0, 2\pi)$ *for all* i*. For a star-shaped polygon* $\underline{\boldsymbol{S}}$ *there exists* $\boldsymbol{\theta}$ *and* $\boldsymbol{y}$ *such that* $\tilde{\boldsymbol{V}}(\boldsymbol{C}, \boldsymbol{\theta}, \boldsymbol{y}) = \boldsymbol{S}$ *for any* $\boldsymbol{C} \in \mathcal{K}(\underline{\boldsymbol{S}})$*. (Proof in Appendix B.1.)*

Corollary 1. *Let* ℓ_θ *denote the line that extends infinitely outward from* $\boldsymbol{C}$ *at angle* $\theta \in (0, 2\pi)$ *and that intersects* $\overline{\boldsymbol{S}}$*. For any* $\theta \in (0, 2\pi)$*, the line* ℓ_θ *is distinct, i.e.* $\ell_\theta \neq \ell_{\theta'}$ *for any* θ, θ' *such that* $\theta \neq \theta'$*. (Proof in Appendix B.1.)*

When considering multiple contours, the star-shaped representation's value becomes clear. With the star-shaped representation, representing how contours differ and what variability is like in different spatial areas is straightforward. For example, consider two contours $\boldsymbol{S}_k$ and $\boldsymbol{S}_\ell$ described with a star-shaped representation using some line set $\mathcal{L}(\boldsymbol{C}, \boldsymbol{\theta})$. With the star-shape representations of both contours, $\tilde{\boldsymbol{V}}(\boldsymbol{C}, \boldsymbol{\theta}, \boldsymbol{S}_k)$ and $\tilde{\boldsymbol{V}}(\boldsymbol{C}, \boldsymbol{\theta}, \boldsymbol{S}_\ell)$, determining how much further one contour extends in a particular direction is straightforward. Simply find the differences between $y_{i,k}$ and $y_{i,\ell}$ where $y_{i,k}$ and $y_{i,\ell}$ are the distances from $\boldsymbol{C}$ to contour $\boldsymbol{S}_k$ and $\boldsymbol{S}_\ell$ along a line extending at angle θ_i. Similarly, with a large set of contours, the variability of the contours along any line $\ell_i \in \mathcal{L}(\boldsymbol{C}, \boldsymbol{\theta})$ can be computed directly from the variability of the corresponding values of $\boldsymbol{y}_i$ in a set of star-shaped representations.

Finally, note that for a contour enclosing a star-shaped polygon, the star-shaped representation is identical to the point-sequence representation when $p = n$, $\boldsymbol{C} \in \mathcal{K}(\underline{\boldsymbol{P}})$, and θ_i

aligns with the direction of the line segment $\overline{\boldsymbol{C}\boldsymbol{s}_i}$ for all i. Note that when these conditions are met, the points $\boldsymbol{V}$ are the same for any choice of starting point $\boldsymbol{C}$ within the kernel of $\underline{\boldsymbol{S}}$. However, the angles $\boldsymbol{\theta}$ and lengths $\boldsymbol{y}$ will differ.

4.3 Star-shaped contour model

4.3.1 General model

We now propose a probability model for generating contours that enclose star-shaped polygons. We call this model the *star-shaped contour model.*

Definition 9. *Star-shaped contour model,* $\boldsymbol{V}(\boldsymbol{C}, \boldsymbol{\theta}, \pi)$*: Let* $\boldsymbol{C} = (C_x, C_y)$ *be a fixed starting point,* $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_p)$ *be a fixed set of* p *unique angles, and* π *be a probability distribution from which a set of values* $\boldsymbol{y} = (y_1, \ldots, y_p) > 0$ *can be drawn. These parameters form a star-shaped probability model if drawn sets* $\boldsymbol{y}$ *are used to form corresponding star-shaped points sets,* $\boldsymbol{V}(\boldsymbol{C}, \boldsymbol{\theta}, \boldsymbol{y})$*, as given in Definition 7.*

We now consider a distribution for π appropriate in many circumstances. We assume that $\boldsymbol{y}$ follows a Gaussian distribution,

$$\boldsymbol{y} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma}), \tag{4.2}$$

where $\boldsymbol{\mu}$ is a mean vector and the parameter $\boldsymbol{\Sigma}$ is a positive-definite covariance function. We further assume that $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are such that mass on non-positive $\boldsymbol{y}$ is negligible. (In practice, if in a small proportion of cases, a generated y_i is non-positive, its value can be set to some small $\eta > 0$.) We call this the *Gaussian Star-Shaped Contour Model* (GSCM). The GSCM can be seen as a finite approximation to the planar version of Gaussian Random Particles proposed in Hansen et al. [2015]. GSCMs can produce a fairly flexible set of contours. The first three rows in Figure 4.3 illustrates the types of contours GSCMs can produce.

Finding $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ that generally result in negligible non-positive values is reasonable because of how $\boldsymbol{y}$ is constructed. The values $\boldsymbol{y}$ represent distances from a starting point $\boldsymbol{C}$ and so are automatically non-negative. Additionally, lengths close to zero can be avoided,

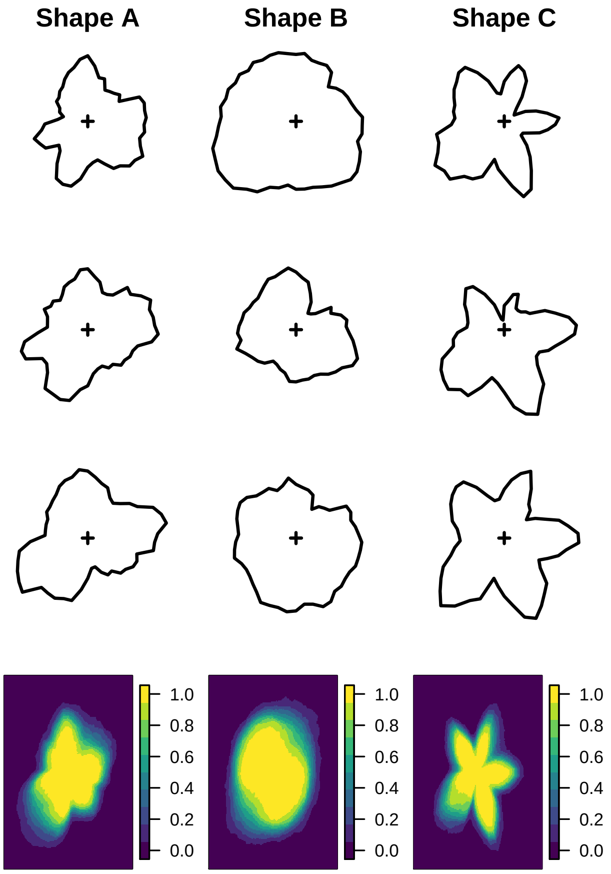


Figure 4.3: Rows 1-3: Nine examples of contours enclosing star-shaped polygons generated from GSCMs with three different parameter settings, organized by column. The cross sign denotes the starting point, $\boldsymbol{C}$. Row 4: Estimated probability of a grid box being contained within contours generated by GSCMs with the column's parameters settings. Probabilities estimated from 100 generated contours. The GSCM parameter settings are referred to as *Shape A* (left), *Shape B* (middle), and *Shape C* (right). For all shapes $p = 50$ and $\kappa = 2$. Values for $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}$ for all shapes can be found in Appendix B.2.

since $\boldsymbol{C}$ can be selected among points in the kernel of the polygon that will typically include some centrally located points. So reasonable $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ that align with typical observable $\boldsymbol{y}$ values will not yield substantial non-positive $\boldsymbol{y}$.

A realistic covariance matrix $\boldsymbol{\Sigma}$ for Equation 4.2 will take account of the fact that the lines in the line set $\mathcal{L}$ are organized via angles $\boldsymbol{\theta}$ extending outward from the starting point $\boldsymbol{C}$. As such, the correlation structure for the set of distances, $\boldsymbol{y}$, should also be a function of the angles. Covariances based on angles are complicated by the fact that 0 and 2π represent the same angle. So, the difference between two angles does not necessarily correspond to how far apart the angles actually are. Specialized covariance functions have been derived that remain valid when distances are indexed by angle [Gneiting, 2013]. Denote the angle between θ_i and θ_j by $d(\theta_i, \theta_j) \in [0, \pi]$.

Typically, the correlation between y_i and y_j will decrease as $d(\theta_i, \theta_j)$ increases. For the simulation examples in this chapter, we focus on an exponential covariance structure, $\boldsymbol{\Sigma}(\boldsymbol{\sigma}, \kappa)$ where $\boldsymbol{\sigma} = (\sigma_1, \ldots, \sigma_p)$ and $\kappa > 0$. The element, $\boldsymbol{\Sigma}_{ij}$, in the i-th row and j-th column of this covariance is

$$\boldsymbol{\Sigma}_{ij} = \sigma_i \sigma_j \exp\left(-\frac{d(\theta_i, \theta_j)}{\kappa}\right). \tag{4.3}$$

4.3.2 Fitting GSCMs

We now turn to how to estimate the parameters of a GSCM given a dataset of observed contours. We assume that the data are comprised of a set of N observed contours, $\mathbf{S} = (\boldsymbol{S}_1, \ldots, \boldsymbol{S}_N)$, that enclose regions that are star-shaped polygons. We also assume that the contours are generated from a common $\boldsymbol{C}$ and $\boldsymbol{\theta}$, i.e., $\boldsymbol{C} = \boldsymbol{C}_1 = \cdots \boldsymbol{C}_N$ and $\boldsymbol{\theta} = \boldsymbol{\theta}_1, \cdots \boldsymbol{\theta}_N$. We fix $\hat{\boldsymbol{C}}$ and $\hat{\boldsymbol{\theta}}$, and then estimate the parameters $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ that describe the distribution of random distances, $\boldsymbol{y}$, from $\hat{\boldsymbol{C}}$ at angles $\hat{\boldsymbol{\theta}}$.

Fixing the starting point $\hat{\boldsymbol{C}}$ and the set of angles $\hat{\boldsymbol{\theta}}$

Our goal in finding $\hat{\boldsymbol{C}}$ and $\hat{\boldsymbol{\theta}}$ is to identify a starting point and set of angles that can be used to describe the details of the observed contours. We first discuss finding $\hat{\boldsymbol{C}}$ and $\hat{\boldsymbol{\theta}}$ individually, each conditional on the other. Then, we describe an iterative algorithm to fix both values.

Finding $\hat{\boldsymbol{C}}$ conditional on $\boldsymbol{\theta}$: Conditional on $\boldsymbol{\theta}$, the fixed value of the starting point, $\hat{\boldsymbol{C}}$, is selected to minimize the differences between the enclosed polygons of the observed contours and enclosed polygons formed by star-shaped representations of the contours. We define the set that differs between an observed polygon, $\underline{\boldsymbol{S}_i}$, and a star-shaped approximation to that contour with starting point, $\hat{\boldsymbol{C}}$, as

$$A(\hat{\boldsymbol{C}}, \boldsymbol{\theta}, \boldsymbol{S}_i) := \{(\underline{\boldsymbol{S}_i}^c \cap \underline{\tilde{\boldsymbol{V}}}(\hat{\boldsymbol{C}}, \boldsymbol{\theta}, \boldsymbol{S}_i) \cup (\underline{\boldsymbol{S}}_i \cap \underline{\tilde{\boldsymbol{V}}}^c(\hat{\boldsymbol{C}}, \boldsymbol{\theta}, \boldsymbol{S}_i))\}. \tag{4.4}$$

where the superscript c denotes the complement of the set. The area contained within this set is denoted by $|A(\hat{\boldsymbol{C}}, \boldsymbol{\theta}, \boldsymbol{S}_i)|$. Assuming the true $\boldsymbol{\theta}$ is known, Theorem 1 guarantees that at least one point $\hat{\boldsymbol{C}}$ exists where $|A(\boldsymbol{C}, \boldsymbol{\theta}, \boldsymbol{S}_i)| = 0$ for all i. So, the best selection of $\hat{\boldsymbol{C}}$ would be the point that gives $|A(\hat{\boldsymbol{C}}, \boldsymbol{\theta}, \boldsymbol{S}_i)| = 0$ for all i,

$$\hat{\boldsymbol{C}} = \{\boldsymbol{C} : |A(\boldsymbol{C}, \boldsymbol{\theta}, \boldsymbol{S}_1)| = 0, \ldots, |A(\boldsymbol{C}, \boldsymbol{\theta}, \boldsymbol{S}_N)| = 0)\} \tag{4.5}$$

Solved exactly, equation 4.5 gives $\hat{\boldsymbol{C}} = \boldsymbol{C}$. In practice, $\hat{\boldsymbol{C}}$ is typically numerically approximated as

$$\hat{\boldsymbol{C}} = \underset{\mathbf{C} \in \boldsymbol{\mathcal{D}}}{\operatorname{argmin}} \left\{ f(|A(\boldsymbol{C}, \boldsymbol{\theta}, \boldsymbol{S}_1)|, \ldots, |A(\boldsymbol{C}, \boldsymbol{\theta}, \boldsymbol{S}_N)|) \right\}, \tag{4.6}$$

where f is some function, typically the mean, and $\boldsymbol{\mathcal{D}}$ is a grid of possible locations of the starting point. The finer the grid of $\boldsymbol{\mathcal{D}}$ the closer $\hat{\boldsymbol{C}}$ will be to $\boldsymbol{C}$.

Computation can generally be substantially reduced by narrowing the grid of possible locations, $\boldsymbol{\mathcal{D}}$, using ideas related to the kernels of star-shaped polygons. We note that the true starting point will be contained with an area we call the *intersection kernel* defined as follows:

Definition 10. *Intersection kernel,* $\boldsymbol{\mathcal{K}}(\boldsymbol{V}(\boldsymbol{C}, \boldsymbol{\theta}, \pi))$*:* *The intersection of the kernels of all possible star-shaped polygons that can be generated from a particular star-shaped contour* *model,* $\boldsymbol{V}(\boldsymbol{C}, \boldsymbol{\theta}, \pi)$*, or* $\cap_{i=1}^{\infty}\mathcal{K}(\underline{\boldsymbol{S}}_i)$ *where* $\underline{\boldsymbol{S}}_i$ *denotes the i-th of the infinitely many star-shaped contours that could be generated from* $\boldsymbol{V}(\boldsymbol{C}, \boldsymbol{\theta}, \pi)$*.*

We can estimate the intersection kernel with observed contours drawn from a star-shaped probability model.

Definition 11. *Estimated intersection kernel,* $\hat{\boldsymbol{\mathcal{K}}}(\underline{\mathbf{S}})$*: Let* $\underline{\mathbf{S}}$ *be a set of* N *star-shaped polygons, then the intersection of the kernels of the polygons is*

$$\hat{\boldsymbol{\mathcal{K}}}(\underline{\mathbf{S}}) = \mathcal{K}(\underline{\boldsymbol{S}}_1) \cap \ldots \cap \mathcal{K}(\underline{\boldsymbol{S}}_N) \tag{4.7}$$

The estimated intersection kernel is a subset of the true intersection kernel. Note further that

Theorem 2. *The true starting point* $\boldsymbol{C} \in \hat{\boldsymbol{\mathcal{K}}}(\underline{\mathbf{S}})$ *where* $\underline{\mathbf{S}}$ *is a set of* N *star-shaped polygons. (Proof in Appendix B.1.)*

Therefore, we need only perform the optimization in Equation 4.6 for the reduced set of points such that $\boldsymbol{\mathcal{D}} \in \hat{\boldsymbol{\mathcal{K}}}(\underline{\mathbf{S}})$.

Computation of $\hat{\boldsymbol{\mathcal{K}}}(\underline{\mathbf{S}})$ in practice is straightforward. Each $\mathcal{K}(\underline{\boldsymbol{S}}_i)$ is found via the intersection of a set of interior half-planes defined by each edge of $\overline{\boldsymbol{S}}_i$. Specifically, the interior half-planes intersected are those that have dividing lines overlapping the edges' line segments and that are on the same side of the dividing lines as the interior of the polygon. The kernel of $\boldsymbol{S}_i$, $\mathcal{K}(\underline{\boldsymbol{S}}_i)$, is the intersection of all these interior half-planes [Shamos, 1975]. Intersecting all the individual polygons, $\mathcal{K}(\underline{\boldsymbol{S}}_i)$, produces $\hat{\boldsymbol{\mathcal{K}}}(\underline{\mathbf{S}})$. For any $\underline{\boldsymbol{S}}_i$ with n edges, $\mathcal{K}(\underline{\boldsymbol{S}}_i)$ can be found as described in $O(n \log n)$ time [Shamos, 1975]. An alternative algorithm can find $\mathcal{K}(\underline{\boldsymbol{S}}_i)$ in $O(n)$ time [Lee and Preparata, 1979].

Finding $\hat{\boldsymbol{\theta}}$ conditional on $\boldsymbol{C}$:

Conditional on $\boldsymbol{C}$, the set of angles $\hat{\boldsymbol{\theta}}$ used in modeling the GSCM should also be selected to minimize the differences between the enclosed polygons of the observed contours and

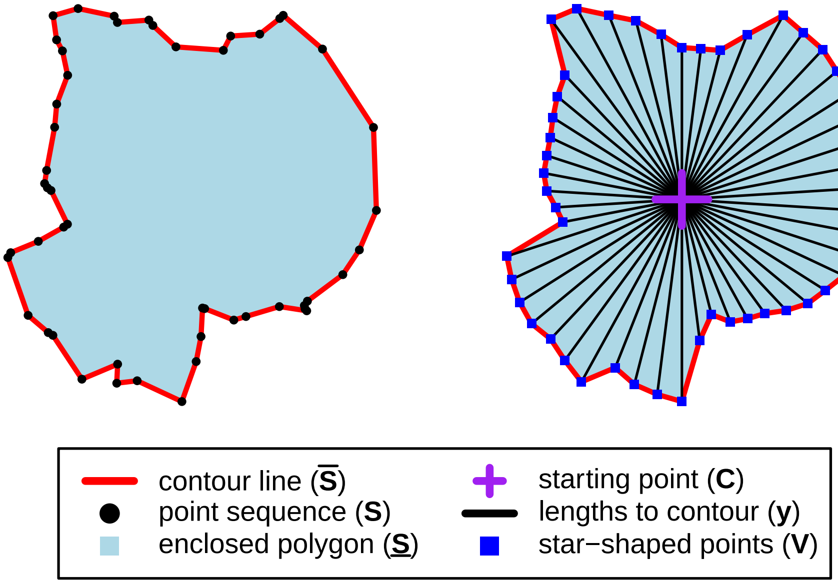


Figure 4.4: Components of a contour represented with a point-sequence representation (left) and a star-shaped representation approximating the point sequence with evenly-spaced angles (right).

enclosed polygons formed by star-shaped representations of contours. Following Corollary 1 any set of distinct angles can be used to form a star-shaped representation of a contour. To help ensure smoothness in generated contours $\hat{\boldsymbol{\theta}}$ and for simplicity, we recommend use evenly spaced angles. Setting $\hat{\boldsymbol{\theta}}$ then reduces to finding $\hat{p}$, the number of elements in $\hat{\boldsymbol{\theta}}$. Let $\boldsymbol{\theta}_p$ denote a set of p evenly angles with $\theta_1 = (2\pi/p)/2 = \pi/p$. Figure 4.4 illustrates how a star-shaped contour is approximated with a star-shaped representation with evenly-spaced angles.

We want to find some $\hat{p}$ that maintains model accuracy while reducing the computation needed to fit the GSCM. Since $\hat{p}$ controls the dimensions of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$, larger $\hat{p}$ results in more computation. We set some allowable amount of error dependent on the application and identify the approximately lowest $\hat{p}$ that does not exceed this allowable error on average. To allow comparisons of polygons of different sizes, we often express this allowable error as a proportion, δ, of the average area of the polygons. This constraint on the allowable error can be expressed as

$$f(|A(\boldsymbol{C}, \hat{\boldsymbol{\theta}}, \boldsymbol{S}_1)|, \ldots, |A(\boldsymbol{C}, \hat{\boldsymbol{\theta}}, \boldsymbol{S}_N)|) < \frac{\delta}{N} \sum_{i=1}^{N} |\underline{\boldsymbol{S}}_{\boldsymbol{i}}|, \tag{4.8}$$

where f is some function, typically the mean, and $|\underline{\boldsymbol{S}}_{\boldsymbol{i}}|$ denotes the area of the polygon $\underline{\boldsymbol{S}}_i$ in the set of N observed polygons. Using lower values of δ will generally result in higher $\hat{p}$ and lower area differences. The balance between computation time and allowable error depends on the context, so the appropriate value of δ will be informed by the application of interest. Section 4.5.3 uses simulation to explore the implications of selecting different δ for coverage accuracy.

We use Algorithm 1 to find an approximately minimal $\hat{p}$ that satisfies the constraint in Equation 4.8. To avoid selecting a larger $\hat{p}$ than necessary, $p^{(0)}$ should generally be initialized such that Equation 4.8 is not likely to be satisfied. Smaller values of a will make $\hat{p}$ more precise, but will require more computation than larger a.

Finding $\hat{\boldsymbol{C}}$ and $\hat{\boldsymbol{\theta}}$: In practice, neither $\hat{\boldsymbol{C}}$ nor $\hat{\boldsymbol{\theta}}$ will be known. So, we Algorithm 2 to iteratively set both values together. As in Algorithm 1, the initial value, $p^{(0)}$, should typically selected to be low enough that we are confident that Equation 4.8 is not satisfied. Otherwise, we may select a larger $\hat{p}$ than is needed. Smaller values of a will result in more precise determination of $\hat{p}$, but computation will increase correspondingly.

Computing a posterior

Once $\hat{\boldsymbol{C}}$ and $\hat{\boldsymbol{\theta}}$ have been determined model fitting is straightforward. For a set $\boldsymbol{S}$ of N observed contours, the values $\boldsymbol{y}_j$ that form the appropriate star-shaped approximation should be computed directly. Then from Equation 4.2, the corresponding likelihood for these $\mathbf{y}$ is just that of a multivariate normal distribution, namely

$$\prod_{j=1}^{N} (2\pi)^{-\hat{p}/2} \det(\boldsymbol{\Sigma}(\cdot)^{-1/2} \exp\left\{-\frac{1}{2}(\boldsymbol{y}_j - \boldsymbol{\mu})^T \boldsymbol{\Sigma}(\cdot)^{-1}(\boldsymbol{y}_j - \boldsymbol{\mu})\right\}. \tag{4.9}$$

Algorithm 1: Finding $\hat{\boldsymbol{\theta}}$ conditional on $\boldsymbol{C}$

```
Initialize $p^{(0)}$ and set $t \leftarrow 0$
Compute RHS of Equation 4.8
while Equation 4.8 does not hold do
    Compute $\boldsymbol{\theta}^{(t)}$ for $p^{(t)}$
    Compute LHS of Equation 4.8 with $\boldsymbol{\theta} = \boldsymbol{\theta}^{(t)}$
    if Equation 4.8 does not hold then
        Set $t \leftarrow t + 1$
        Set $p^{(t+1)} \leftarrow ap^{(t)}$ where is $a > 1$
    else
        Set $\hat{\boldsymbol{\theta}} \leftarrow \hat{\boldsymbol{\theta}}^{(t)}$
    end
end
```

Estimates of the mean vector $\boldsymbol{\mu}$ and the parameter controlling $\boldsymbol{\Sigma}(\cdot)$ can be estimated using any standard method.

In the simulations and example that follow we take a Bayesian approach and assume an exponential covariance as in Equation 4.3 with parameters $\boldsymbol{\sigma}$ and κ. Priors should be selected to reflect the understanding of the scientific problem. For demonstration purposes, we set the following simple prior distributions for $\boldsymbol{\mu}$, $\boldsymbol{\sigma}$ and κ. We assume a multivariate normal prior distribution for $\boldsymbol{\mu}$, namely

$$\boldsymbol{\mu} \sim \text{MVN}(\boldsymbol{\mu}_{\mathbf{0}}, \boldsymbol{\Lambda}_0). \tag{4.10}$$

The hyperparameters μ_0 and $\boldsymbol{\Lambda}_0$ give the mean and covariance of the prior distribution. For the example exponential covariance defined in Equation 4.3 we assume simple uniform priors

Algorithm 2: Finding $\hat{\boldsymbol{C}}$ and $\hat{\boldsymbol{\theta}}$

```
Initialize $p^{(0)}$ and set $t \leftarrow 0$
Compute RHS of Equation 4.8
while Equation 4.8 does not hold do
    Compute $\boldsymbol{\theta}^{(t)}$ for $p^{(t)}$
    Find $\hat{\boldsymbol{C}}^{(t)}$ using Equation 4.6 with $\boldsymbol{\theta} = \boldsymbol{\theta}^{(t)}$
    Compute LHS of Equation 4.8 with $\boldsymbol{C} = \hat{\boldsymbol{C}}^{(t)}$ and $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}^{(t)}$.
    if Equation 4.8 does not hold then
        Set $t \leftarrow t + 1$
        Set $p^{(t+1)} \leftarrow ap^{(t)}$ where is $a > 1$
    else
        Set $\hat{\boldsymbol{C}} \leftarrow \hat{\boldsymbol{C}}^{(t)}$
        Set $\hat{\boldsymbol{\theta}} \leftarrow \hat{\boldsymbol{\theta}}^{(t)}$
    end
end
```

on κ and $\boldsymbol{\sigma}$, namely

$$\kappa \sim \text{Unif}(0, \beta_{\kappa,0}) \tag{4.11}$$

$$\sigma_j \sim \text{Unif}(0, \beta_{\sigma,0}) \tag{4.12}$$

where the hyperparameters $\beta_{\kappa,0}$ and $\beta_{\sigma,0}$ are upper bounds. Samples from the posterior distributions of the parameter can be found via standard Markov chain Monte Carlo (MCMC). In our implementation, Gibbs steps are used for μ and Metropolis steps for $\boldsymbol{\sigma}$ and κ.

Estimating gridded probabilities and credible intervals

The mean parameter estimates from the sampled MCMC define a GSCM from which K contours, denoted by $\hat{\mathbf{S}} = (\hat{\boldsymbol{S}}_1, \dots \hat{\boldsymbol{S}}_K)$, can be generated. If K is large enough, we can compute the probability of a given area being contained within a contour. Each generated contour can be approximated by a binary grid, $\mathbf{G}$, of dimension $r \times v$. Let $g_{i,j}$ indicate the grid box in the i-th row and j-th column of $\mathbf{G}$. Let $\mathbb{1}g_{i,j,k}$ indicate whether the majority of the area in grid box $g_{i,j}$ is inside or outside contour k. Most grid boxes will be entirely inside or outside the contour; however, grid boxes that intersect the generated contour will contain area both inside and outside the contour. Ideally, the grid selected should be fine enough to ensure that little area is contained within these transitional grid boxes. Averaging the binary grids produces an $r \times v$ matrix, $\hat{\mathbf{P}}$, with elements $\hat{p}_{i,j} = \sum_{i=1}^{K} \mathbb{1}g_{i,j,k}/K$, that indicate the probability of grid box $g_{i,j}$ being contained within a contour. The last row of Figure 4.3 shows estimated gridded probabilities obtained from 100 generated contours from the corresponding GSCMs.

Credible regions for the location of the contour can be computed from $\hat{\mathbf{P}}$. The $(1-\alpha)$ credible region, $\boldsymbol{I}_{1-\alpha}$, is formed from a union of grid boxes that satisfy the condition

$$\boldsymbol{I}_{1-\alpha} = \left\{ g_{i,j} : \frac{\alpha}{2} < \hat{p}_{i,j} < 1 - \frac{\alpha}{2} \right\}. \tag{4.13}$$

Rescaling data

For numerical convenience in fitting and generating contours, it is often desirable for all contours to be contained within the $[0,1] \times [0,1]$ unit square. Observed data will typically need to be rescaled to be within these bounds. Data should be rescaled such that generated contours do not extend outside the unit square. A good rescaling can ensure that contours that extend outside the unit square are rare. Specifically, observed contours $\mathbf{S} = (\boldsymbol{S}_1, \ldots \boldsymbol{S}_N)$ are rescaled to be within an $[\epsilon, 1-\epsilon] \times [\epsilon, 1-\epsilon]$ square. This rescaling provides a buffer region of width ϵ on the outside of the unit square in which no contours have been observed. Generated contours that extend farther than the observed contours will then be in this buffer region rather than outside the unit square. The higher the variability of contours, the larger the value of ϵ needed to avoid generating contours that go beyond the unit square.

To translate a set of observed coordinates, $\mathbf{S} = (\boldsymbol{S}_1, \ldots, \boldsymbol{S}_n)$, to the square of dimension $[\epsilon, 1-\epsilon] \times [\epsilon, 1-\epsilon]$, let $\min(\mathbf{S}_x)$ and $\max(\mathbf{S}_x)$ denote the minimum and maximum observed x-coordinates from all spatial points in all contours in $\mathbf{S}$. Define $\min(\mathbf{S}_y)$ and $\max(\mathbf{S}_x)$ analogously for the y-coordinates. Let $\boldsymbol{s}_{i,j} = (s^x_{i,j}, s^y_{i,j})$ denote the i-th point in the j-th contour. Then for all i and all j, the equivalent rescaled coordinates for observed points $\boldsymbol{s}_{i,j}$ are

$$\tilde{s}^x_{i,j} = \epsilon + (1-2\epsilon)(s^x_{i,j} - \min(\mathbf{S}_x))/(\max(\mathbf{S}_x) - \min(\boldsymbol{S}_x)), \tag{4.14}$$

$$\tilde{s}^y_{i,j} = \epsilon + (1-2\epsilon)(s^y_{i,j} - \min(\mathbf{S}_y))/(\max(\mathbf{S}_y) - \min(\mathbf{S}_y)). \tag{4.15}$$

4.4 Coverage Metric

To assess the appropriateness of a probabilistic contour model a metric is needed. A good model would correctly identify the region where the contour could plausibly be located. We focus on the coverage of prediction intervals for star-shaped contours. With an accurate contour model, the variability of the generated contour would be correctly represented along all parts of the contour. In designing an appropriate metric, we leverage the star-shaped structure of the data. The general idea is to consider coverage along each line in a line set. Each marginal assessment determines if variability is correctly represented on the line extending outward from some starting point, $\boldsymbol{C}^*$, at a particular angle.

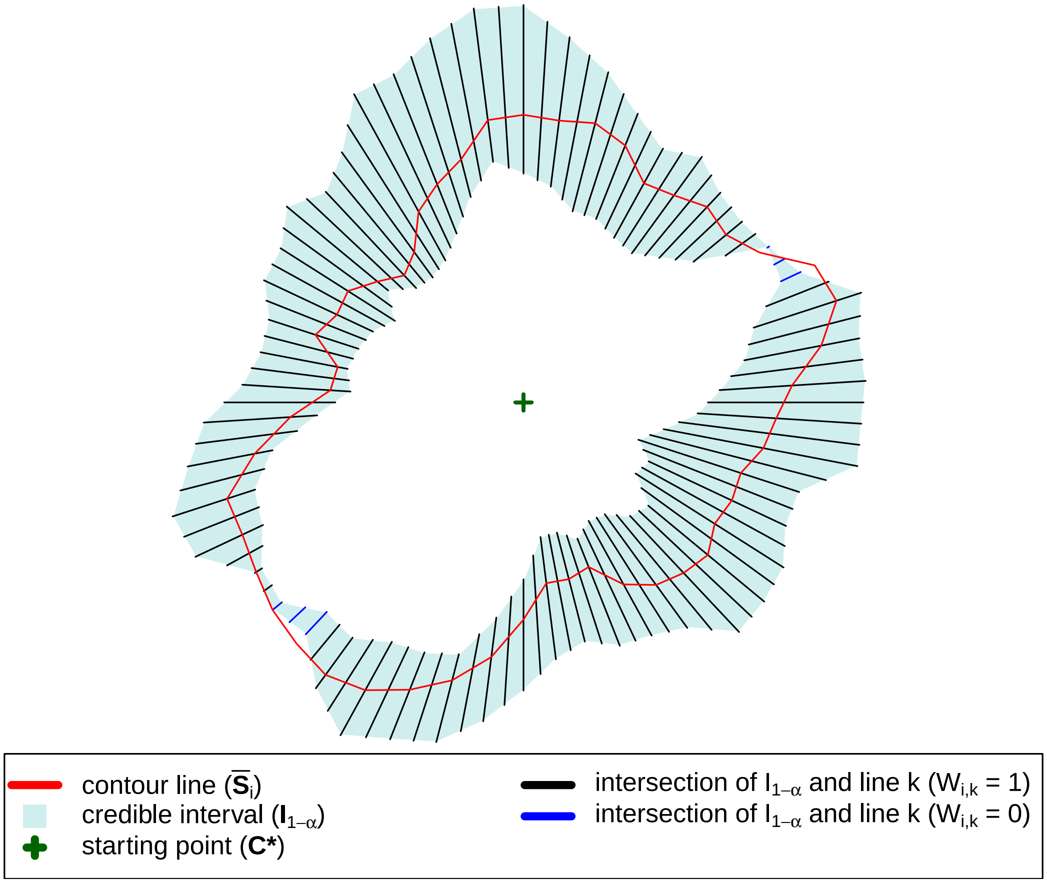


Figure 4.5: Illustration of coverage assessment for a contour line, $\overline{\boldsymbol{S}}_i$ (red), and a $1-\alpha$ credible region, $\boldsymbol{I}_{1-\alpha}$ (light blue). The line segments, $I_{1-\alpha,k}$, corresponding to the intersection of the $\boldsymbol{I}_{1-\alpha}$ credible region and line ℓ_k are colored black when they cover $S_{i,k}$ and blue otherwise. The center of the star-shaped polygon from which the contour is generated is denoted by a green cross sign.

To make this general idea precise, we define several quantities that are illustrated in Figure 4.5. As in Equation 4.13, let $\boldsymbol{I}_{1-\alpha}$ be the $1-\alpha$ credible region obtained from some contour distribution. Define $I_{1-\alpha,k}$ as the line segment formed from the intersection of the

$\boldsymbol{I}_{1-\alpha}$ credible region and the line ℓ_k in some test line set $\mathcal{L}^*(\boldsymbol{C}^*, \boldsymbol{\theta}^*)$ with M lines. We refer to $I_{1-\alpha,k}$ as a test line. Also, we define $R_{i,k}$ as the intersection of some observed contour $\boldsymbol{S}_i$ and line ℓ_k. Note that $R_{i,k}$ will always be a single point when the polygon $\underline{\boldsymbol{S}}$ is exactly star-shaped and $\boldsymbol{C}^* \in \mathcal{K}(\underline{\boldsymbol{S}_i})$.

Let $W_{i,k} = \mathbb{1}[R_{i,k} \in I_{1-\alpha,k}]$ indicate whether the intersection points of the observed contour and the line $\ell_k \in \mathcal{L}^*(\boldsymbol{C}^*, \boldsymbol{\theta}^*)$ are contained within the intersection of the credible region and line ℓ_k. Then, for credible intervals with perfect coverage, for any i, k,

$$Pr(W_{i,k}) = \mathbb{E}[W_{i,k}] = 1 - \alpha. \tag{4.16}$$

In other words, the part of the true contour that intersects line ℓ_k is contained within the credible contour region along line ℓ_k with probability $1 - \alpha$.

To better understand coverage results with this metric, we consider coverage behavior for a set of N observed contours, $\mathbf{S} = (\boldsymbol{S}_1, \dots, \boldsymbol{S}_N)$. We assume the contours being assessed are drawn from some star-shaped contour model as in Definition 9 with fixed but unknown $\boldsymbol{C}$, $\boldsymbol{\theta}$, and π. Conditional on $\boldsymbol{C}$, $\boldsymbol{\theta}$ and π, each $\boldsymbol{S}_i$ is assumed to be independent of $\boldsymbol{S}_j$ for all i, j.

For a credible interval with perfect coverage for any i, k,

$$\sum_{i=1}^{N} W_{i,k} \approx N(1 - \alpha). \tag{4.17}$$

for sufficiently large sample size, N. This equation is used to assess coverage in practice. Equation 4.17 not holding for any k indicates the variability of the contour around line ℓ_k is not correctly represented by the credible region.

With this set-up, $W_{i,k}$ and $W_{j,k}$ are independent for all i, j. So, for a given line ℓ_k,

$$\sum_{i=1}^{N} W_{i,k} \sim \text{Binomial}(N, 1 - \alpha). \tag{4.18}$$

Since the distributional behavior of the quantity $\sum_{i=1}^{N} W_{i,k}$ is known, the expected variability of the mean coverage given a particular sample size is known and can be accounted for in a simulation study or cross-validation experiment.

We also consider how coverage on one test lines in $\mathcal{L}^*(\boldsymbol{C}^*, \boldsymbol{\theta}^*)$ relates to coverage on another test line. The location of points on a contour are correlated. So, $W_{i,k}$ and $W_{i,\ell}$ are correlated. Typically $W_{i,k}$ and $W_{i,\ell}$ will be more correlated the smaller the angle distance from θ_k and θ_ℓ. What this means is that while

$$\mathbb{E}\left[\sum_{k=1}^{M} W_{i,k}\right] = M(1-\alpha), \tag{4.19}$$

is fixed and known, the distribution of $\sum_{k=1}^{M} W_{i,k}$ is not a good metric to assess to coverage since its variability depends on the correlation structure of the points on the contour. For a contour with high correlation among points, the quantity $\sum_{k=1}^{M} W_{i,k}$ will be affected substantially by what contour happens to be sampled. For intuition, consider a contour with high correlation among all i, j. In this case, the contour is likely to be entirely within the credible interval or entirely outside of it, so $\sum_{k=1}^{M} W_{i,k}$ will be either 0 or M.

Overall, these relationships among coverage for different numbers of sampled contours and different numbers of angles show that sample sizes need to be considered in terms of the number of contours observed, N. The metric $\sum_{i=1}^{N} W_{i,k}$ should be considered for all elements in $\boldsymbol{\theta}$.

In selecting the number of test lines, M, the goal should be to space lines such that accuracy is assessed with detail appropriate for the application. Since the true exact value of p, is unknown, we cannot simply set $M = p$. However, based on the observed data, we should have a general idea of p. So, we set $M >> p$, to ensure that $M > p$. We also evenly space $\boldsymbol{\theta}^*$ with $\theta_1^* = (2\pi/M)/2 = \pi/M$. By assessing coverage on a substantially greater number of test lines than number of true lines, we ensure that coverage is assessed near every true line.

To carry out this assessment, a fixed starting point $\boldsymbol{C}^*$ and $\boldsymbol{\theta}^*$ should be selected. In simulation studies, the true value of $\boldsymbol{C}$ will be known and we can let $\boldsymbol{C}^* = \boldsymbol{C}$. For assessment of real data such as in a cross-validation study, a starting point for $\boldsymbol{C}^*$ will be unknown and must be determined. We recommend using a $\boldsymbol{C}^*$ that minimizes area differences between the enclosed polygons of observations and enclosed polygons formed by star-shaped representations as in Section 4.3.2, i.e. let $\boldsymbol{C}^* = \hat{\boldsymbol{C}}$.

This assessment approach differs from how contours have been assessed in the context of level exceedances. There, a credible region or confidence region has often been defined as the region that covers the true contour in its entirety $(1-\alpha)$-proportion of the time [e.g. Bolin and Lindgren, 2015, French, 2014]. With credible (confidence) regions constructed to satisfy this definition, coverage can be assessed by determining what proportion of the time the true contour is fully contained within the region. We opt not to use this metric since our goal is to develop a method to generate contours directly. Correlation along the contour makes assessing the probability of capturing the entire contour difficult. Our metric reflects that we are most concerned with getting the right variability in all parts of the contour. We are less concerned with identifying a larger area that contains that entirety of the contour with high probability. Our intervals are therefore narrower than would be required for these global intervals.

4.5 Simulation studies

4.5.1 Simulation details

In the following simulation studies, we consider how the star-shaped model performs in inferring distributions of contours from datasets of observed contours. We consider performance with varying numbers of observations, different constraints for determining the number of lines fit (δ as defined in Section 4.3.2), and varied GSCM parameters.

In many of our simulations, we focus on a particular GSCM with $p = 50$ that we will refer to as *Shape A*. The correlation structure of $\boldsymbol{y}$ follows the exponential form given in Equation 4.3. The vector of mean distances, $\boldsymbol{\mu}$, and variance parameter vector, $\boldsymbol{\sigma}$, change gradually. The exact values of $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}$ can be found in Appendix B.2. Unless otherwise noted, κ is set to 2. Example generated contours and gridded probability estimates for *Shape A* are given in the left panel of Figure 4.3.

The values $\hat{\boldsymbol{C}}$ and $\hat{\boldsymbol{\theta}}$ are found as described in Section 4.3.2. The parameter values are fit from observations using MCMC. Chains are run for 50,000 iterations with the first

15,000 iterations discarded as burn-in. The prior parameters are $\boldsymbol{\mu_0} = (0.2, \ldots, 0.2)$, $\beta_{\kappa_0} = (8, \ldots, 8)$, $\beta_{\sigma_0} = (0.15, \ldots, 0.15)$, and $\boldsymbol{\Lambda}_0 = .05\mathbf{I}_{\hat{p}}$ where $\mathbf{I}_{\hat{p}}$ is a diagonal matrix of dimension $\hat{p}$ by $\hat{p}$. The value $\hat{p}$ refers to the number of angles in $\hat{\boldsymbol{\theta}}$.

For all simulations, we use 40 evaluation runs. On each evaluation run, we estimate the model parameters using some number N of sample contours generated with the true parameter values. We then generate 100 contours with the mean estimated parameter values and find the corresponding credible intervals as defined in Section 4.3.2. We generate a final contour using the true parameter values to act as the "observed" contour in the simulation. For all evaluation runs, we record the coverage for a set of $M = 100$ evenly spaced test lines with $\theta_1^* = \pi/M$. We report the mean coverage over the 40 evaluations as defined in Section 4.4. We also report the standard deviation of the covered proportion across the 100 test lines.

4.5.2 Number of observations

Our first simulation varies the number of simulated "observed" contours used to fit the *Shape A* GSCM. We consider coverage performance for 10, 20, and 50 simulated observed contours with $\delta = 0.02$. Table 4.5.2 displays the results of this simulation.

We find that coverage improves for $N = 20$ compared to $N = 10$. We find slightly worse performance for $N = 50$ than $N = 20$, although the performances are not substantially different given that we had only 40 evaluation runs. These results indicate that obtaining some minimum sample size is important for coverage performance. For small sample sizes, on the order of $N = 10$, the data alone may not be enough to produce accurate coverage, particularly for the 80% credible interval. However, data sets of this size can potentially still be modeled correctly if informative priors can supplement the observations.

4.5.3 Value of δ

In these simulations, we evaluate how coverage accuracy is affected by the parameter δ introduced in Section 4.3.2. This parameter controls the number of lines used in fitting, $\hat{p}$,

Table 4.1: Mean coverage values for 40 simulations of fitting the contour distribution for *Shape A* with different number of observed contours. In each simulation, $M = 100$ evenly-spaced test lines were evaluated with $\theta_1^* = \pi/M$. Standard deviations across the test lines are given in parentheses. Priors and MCMC details are given in Section 4.5.1

Nominal	**N = 10**	**N= 20**	**N= 50**
0.80	0.87 (0.04)	0.79 (0.07)	0.76 (0.06)
0.90	0.94 (0.04)	0.89 (0.05)	0.86 (0.05)
0.95	0.98 (0.02)	0.95 (0.03)	0.93 (0.04)

and how accurate $\hat{\boldsymbol{C}}$ must be. We evaluate coverage for 40 evaluations runs of *Shape A* with δ set to $0.03, 0.02$, and 0.01. These δ selections set the allowable maximal area difference to 3%, 2%, and 1% of the mean area contained within the observed contours. We also consider how correlation in $\boldsymbol{y}$ affects the need for different p by evaluating each δ for three different smoothness parameter, κ, values: 1, 2, and 4. Table 4.5.3 displays the mean coverage across test lines for three α-levels along with the mean $\hat{p}$ found. On each evaluation the number of sampled contours is set to $N = 20$.

We find that the mean coverage accuracy is only modestly affected by the value of δ. These results support the idea that using a lower δ in many cases will reduce computation while not reducing model performance. We also find that, for a given δ, an increase in κ corresponds to a decrease in $\hat{p}$. In other words, for a contour with higher correlation among $\boldsymbol{y}$, a smaller set of lines can adequately represent the contour distribution.

4.5.4 GSCM Parameters

We also evaluate contour models that have mean values, $\boldsymbol{\mu}$, that vary more slowly (*Shape B*) and more quickly (*Shape C*) around the contour, compared to *Shape A*. Figure 4.3 shows sample contours and probability distributions for these shapes. Both models are defined to have $p = 50$ and $\kappa = 2$. Exact values for $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}$ for *Shape B* and *Shape C* can be found in Appendix B.2. We report coverage results in Table 4.5.4 for three α-levels using $N = 20$

Table 4.2: Mean coverage values for 40 simulations fitting the contour distribution for *Shape A* with different values of δ. In each simulation, 20 observed contours were sampled and $M = 100$ evenly-spaced test lines with $\theta^* = \pi/M$ were evaluated. Standard deviations across the test lines are given in parentheses. The mean $\hat{p}$ is given for each δ along with the standard deviation across the evaluation runs in parentheses. Apostrophes indicate that the entry is the same as the line above it. Priors and MCMC details are given in Section 4.5.1

$\boldsymbol{\kappa}$	**Nominal**	$\boldsymbol{\delta = 0.03}$		$\boldsymbol{\delta = 0.02}$		$\boldsymbol{\delta = 0.01}$	
		Coverage	**Mean** $\hat{p}$	**Coverage**	**Mean** $\hat{p}$	**Coverage**	**Mean** $\hat{p}$
1	0.8	0.86 (0.05)	38.48 (0.8)	0.87 (0.06)	45.65 (1.8)	0.86 (0.05)	55.20 (9.2)
	0.9	0.94 (0.04)	”	0.94 (0.04)	”	0.94 (0.03)	”
	0.95	0.98 (0.02)	”	0.98 (0.02)	”	0.98 (0.02)	”
2	0.8	0.82 (0.05)	32.65 (0.9)	0.80 (0.06)	41.27 (1.2)	0.84 (0.06)	50.50 (1.9)
	0.9	0.92 (0.04)	”	0.90 (0.05)	”	0.91 (0.04)	”
	0.95	0.96 (0.03)	”	0.95 (0.04)	”	0.95 (0.03)	”
4	0.8	0.86 (0.06)	28.38 (0.67)	0.79 (0.05)	36.67 (0.9)	0.80 (0.07)	48.45 (1.0)
	0.90	0.93 (0.04)	”	0.89 (0.04)	”	0.91 (0.04)	”
	0.95	0.97 (0.03)	”	0.94 (0.03)	”	0.97 (0.02)	”

Table 4.3: Mean coverage values for 40 simulations fitting the contour distribution for shapes A, B and C. In each simulation, $N = 20$ simulated observed contours were sampled as data and $M = 100$ evenly-spaced test lines were evaluated with $\theta_1^* = \pi/M$. Standard deviations across the test lines are given in parentheses. Priors and MCMC details are given in Section 4.5.1.

Nominal	**Shape A**	**Shape B**	**Shape C**
0.8	0.79 (0.07)	0.79 (0.06)	0.87 (0.05)
0.9	0.89 (0.05)	0.89 (0.05)	0.95 (0.03)
0.95	0.95 (0.03)	0.95 (0.04)	0.98 (0.02)

simulated observed samples and $\delta = 0.02$.

We find reasonably accurate coverage performance for all three different shapes. The method performs slightly worse for *Shape C* than for *Shape A* and *Shape B*, likely due to the difficulty of getting the pointed sections of *Shape A* in the correct location if $\hat{p}$ is even slightly underestimated. This result indicates that for contours that look like *Shape A*, lower δ values may be appropriate. Overall, the good performance across parameter settings indicates that a range of contours can be well approximated by a GSCM.

4.6 Modeling contours enclosing approximately star-shaped polygons

A GSCM can also be applied to data where the polygons enclosed by the contours are not exactly star-shaped, as long as they are approximately so. Fitting proceeds similarly to Section 4.3.2. This section describes how to assess whether the the GSCM is appropriate given the observed data, and how the fitting procedure is altered if the observed contours enclose polygons that are not exactly star-shaped.

4.6.1 Assessing appropriateness of GSCM

Two main assumptions must be met to apply the GSCM: the polygons enclosed by the contours must be approximately star-shaped and all contours should have at least one common point to allow for the definition of a starting point. The latter assumption can be assessed by determining whether the observed set of contours have overlapping areas. The former assumption can be assessed using metrics that describe how close an observed contour is to enclosing a polygon that is star-shaped. The metrics focus on the difference in the area between a star-shaped representation of the contour and the true contour. If these differences are small for a set of observed contours, $\mathbf{S}$, then the GSCM can be applied.

We relax Definition 8 to make precise how to approximate an arbitrary polygon with the star-shaped representation. Two main differences between star-shaped polygons and arbitrary polygons are addressed in these new definitions: 1) an arbitrary polygon may not have a kernel and 2) the contour line enclosing an arbitrary polygon may intersect with some of the lines in the line set multiple times. The new star-shaped representation definitions are:

Definition 12. *Underestimated star-shaped representation,* $\tilde{\boldsymbol{V}}_u(\boldsymbol{C}, \boldsymbol{\theta}, \boldsymbol{S})$*: Let* $\underline{\boldsymbol{S}}$ *be a polygon described by ordered spatial points* $\boldsymbol{S} = (s_1, \dots, s_n)$*, let* $\boldsymbol{C} \in \underline{\boldsymbol{S}}$ *be a starting point, let* $\boldsymbol{\theta}$ *be an arbitrary set of* p *unique angles, and let* $\boldsymbol{y} = (y_1, \dots, y_p)$ *be a set of distances from* $\boldsymbol{C}$ *to the* ***closest*** *intersection point of the contour line* $\overline{\boldsymbol{S}}$ *and each line* ℓ_i *in the line set* $\mathcal{L}(\boldsymbol{C}, \boldsymbol{\theta})$*. Then, the star-shaped representation of the contour,* $\boldsymbol{V}(\boldsymbol{C}, \boldsymbol{\theta}, \boldsymbol{y})$*, is the underestimated star-shaped representation,* $\tilde{\boldsymbol{V}}_u(\boldsymbol{C}, \boldsymbol{\theta}, \boldsymbol{S})$*.*

Definition 13. *Overestimated star-shaped representation,* $\tilde{\boldsymbol{V}}_o(\boldsymbol{C}, \boldsymbol{\theta}, \boldsymbol{S})$*: Let* $\underline{\boldsymbol{S}}$ *be a polygon described by ordered spatial points* $\boldsymbol{S} = (s_1, \dots, s_n)$*, let* $\boldsymbol{C} \in \underline{\boldsymbol{S}}$ *be a starting point, let* $\boldsymbol{\theta}$ *be an arbitrary set of* p *unique angles, and let* $\boldsymbol{y} = (y_1, \dots, y_p)$ *be the set of distances from* $\boldsymbol{C}$ *to the* ***farthest*** *intersection point of the contour line* $\overline{\boldsymbol{S}}$ *and each line* ℓ_i *in the line set* $\mathcal{L}(\boldsymbol{C}, \boldsymbol{\theta})$*. Then, the star-shaped representation of the contour,* $\boldsymbol{V}(\boldsymbol{C}, \boldsymbol{\theta}, \boldsymbol{y})$*, is the overestimated star-shaped representation,* $\tilde{\boldsymbol{V}}_o(\boldsymbol{C}, \boldsymbol{\theta}, \boldsymbol{S})$*.*

The names of these representations highlight that these polygons generally under- or overestimate the area contained within the true polygon. For contours that enclose star-shaped polygons, if $\boldsymbol{C} \in \mathcal{K}(\underline{\boldsymbol{S}})$, only one intersection is found between the contour line and all lines in the line set. So, $\tilde{\boldsymbol{V}}_o(\boldsymbol{C}, \boldsymbol{\theta}, \boldsymbol{S}) = \tilde{\boldsymbol{V}}_u(\boldsymbol{C}, \boldsymbol{\theta}, \boldsymbol{S}) = \tilde{\boldsymbol{V}}(\boldsymbol{C}, \boldsymbol{\theta}, \boldsymbol{S})$.

We now formalize these differences in area between the original polygon and the under- or overestimated star-shaped representation. For notational convenience, let $\tilde{\boldsymbol{V}}_u = \tilde{\boldsymbol{V}}_u(\boldsymbol{C}, \boldsymbol{\theta}, \boldsymbol{S})$ and $\tilde{\boldsymbol{V}}_o = \tilde{\boldsymbol{V}}_o(\boldsymbol{C}, \boldsymbol{\theta}, \boldsymbol{S})$. Then the sets that differ between the true contour and the under- and overestimated star-shaped representations are

$$A_u(\boldsymbol{C}, \boldsymbol{\theta}, \boldsymbol{S}) := \{(\underline{\boldsymbol{S}}^c \cap \underline{\tilde{\boldsymbol{V}}_u}) \cup (\underline{\boldsymbol{S}} \cap \underline{\tilde{\boldsymbol{V}}_u}^c)\}, \tag{4.20}$$

$$A_o(\boldsymbol{C}, \boldsymbol{\theta}, \boldsymbol{S}) := \{(\underline{\boldsymbol{S}}^c \cap \underline{\tilde{\boldsymbol{V}}_o}) \cup (\underline{\boldsymbol{S}} \cap \underline{\tilde{\boldsymbol{V}}_o}^c)\}, \tag{4.21}$$

where the superscript c denotes the complement of the set. Let $|A_u(\boldsymbol{C}, \boldsymbol{\theta}, \boldsymbol{S})|$ and $|A_o(\boldsymbol{C}, \boldsymbol{\theta}, \boldsymbol{S})|$ be the area contained within these sets. Figure 4.6 illustrates the quantities described in this section for four contours.

The area differences are zero only if the polygons are star-shaped. More precisely:

Theorem 3. *For any polygon* $\underline{\boldsymbol{S}}$ *that is not star-shaped* $|A_u(\boldsymbol{C}, \boldsymbol{\theta}, \boldsymbol{S})| > 0$ *and* $|A_o(\boldsymbol{C}, \boldsymbol{\theta}, \boldsymbol{S})| > 0$ *for any* $\boldsymbol{C}$ *and* $\boldsymbol{\theta}$. *(Proof in Appendix B.1.)*

For easier comparison of these area differences across sets of polygons of different sizes, these areas can be expressed as percentages of the mean total area of the polygons. Table 1 reports the difference areas for the contours in Figure 4.6, and their star-shaped representations as a percentage of the total area of the polygon.

4.6.2 Fitting GSCMs to approximately star-shaped polygons

The approach to fitting in Section 4.3.2 needs to be altered slightly for contours that enclose regions that are only approximately star-shaped contours. The values of $\boldsymbol{y}$ need to be computed using the under- or overestimated star-shaped approximation as given in Definitions 12 and 13. Whether to use the under- or overestimated star-shaped representation depends on the application. In some cases, asymmetric risks may motivate selecting a model that generally over- or underestimates the area within the polygon. Otherwise, both $|A_u(\boldsymbol{C}, \boldsymbol{\theta}, \boldsymbol{S})|$

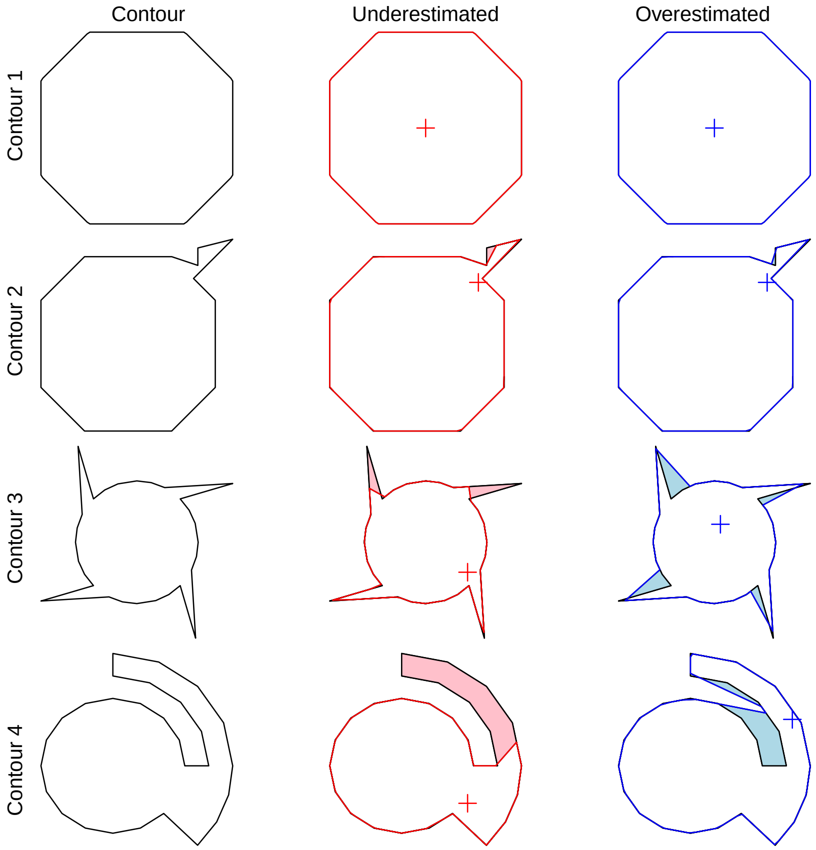


Figure 4.6: Four contours (left) with their underestimated star-shaped approximations, $\tilde{\boldsymbol{V}}_u(\hat{\boldsymbol{C}}_u, \boldsymbol{\theta}, \boldsymbol{S})$, (red, center), and their overestimated star-shaped approximations, $\tilde{\boldsymbol{V}}_o(\hat{\boldsymbol{C}}_o, \boldsymbol{\theta}, \boldsymbol{S})$ (blue, right). The polygon in the top row is star-shaped, and the other three polygons are not. Pink and blue sections in the center and right panel show the respective area differences, $A_u(\hat{\boldsymbol{C}}_u, \boldsymbol{\theta}, \boldsymbol{S})$ and $A_o(\hat{\boldsymbol{C}}_o, \boldsymbol{\theta}, \boldsymbol{S})$. The red crosses in the central panel and the blue crosses in the right panel denote the estimated starting points, $\hat{\boldsymbol{C}}_u$ and $\hat{\boldsymbol{C}}_o$. The vector $\boldsymbol{\theta}$ contains 200 elements spaced evenly in the interval $[0, 2\pi]$.

and $|A_o(\boldsymbol{C}, \boldsymbol{\theta}, \boldsymbol{S})|$ can be computed for the set of observed contours $\boldsymbol{S}$ and whichever representation results in less area differences can be selected. Fitting the posterior proceeds as in Section 4.3.2. Finding probabilities and credible intervals proceeds as in Section 4.3.2.

We find $\hat{\boldsymbol{C}}$ and $\hat{\boldsymbol{\theta}}$ using nearly the same algorithm as in Section 4.3.2, except that we update the star-shaped representation in Equation 4.6 and Equation 4.8 to be the under- or

Table 4.4: Values of the area differences for the under- and overestimated star-shaped approximations, $|A_u(\hat{\boldsymbol{C}}_u, \boldsymbol{\theta}, \boldsymbol{S})|$ and $|A_o(\hat{\boldsymbol{C}}_o, \boldsymbol{\theta}, \boldsymbol{S})|$, for the contours in Figure 4.6. Area differences are computed numerically and expressed as a percentage of the total area of the polygon. The vector $\boldsymbol{\theta}$ contains 200 elements spaced evenly in the interval $[0, 2\pi]$.

	Underestimated	**Overestimated**
Contour 1	0.00	0.00
Contour 2	0.43	0.24
Contour 3	4.44	9.78
Contour 4	15.65	8.46

overestimated star-shaped representation. Specifically, we replace Equation 4.6 with

$$\hat{\boldsymbol{C}}_u = \underset{\boldsymbol{C} \in \boldsymbol{\mathcal{D}}}{\operatorname{argmin}} \{f(|A_u(\boldsymbol{C}, \boldsymbol{\theta}, \boldsymbol{S}_1)|, \dots, |A_u(\boldsymbol{C}, \boldsymbol{\theta}, \boldsymbol{S}_N)|)\} \tag{4.22}$$

for the underestimated representation and

$$\hat{\boldsymbol{C}}_o = \underset{\boldsymbol{C} \in \boldsymbol{\mathcal{D}}}{\operatorname{argmin}} \{f(|A_u(\boldsymbol{C}, \boldsymbol{\theta}, \boldsymbol{S}_1)|, \dots, |A_u(\boldsymbol{C}, \boldsymbol{\theta}, \boldsymbol{S}_N)|)\} \tag{4.23}$$

for the overestimated representation. The values $\hat{\boldsymbol{C}}_u$ and $\hat{\boldsymbol{C}}_{\boldsymbol{o}}$ denote the estimated starting points for the two approximations respectively. As before, f is some function, typically the mean. Like in Section 4.3.2, we evaluate these functions of the area difference at a grid of possible locations, $\boldsymbol{\mathcal{D}}$. Unlike in Section 4.3.2 we cannot use the estimated intersection kernel to reduce the size of $\boldsymbol{\mathcal{D}}$, since the polygons are not exactly star-shaped. However, we note that the starting point must be common to all observed polygons, so we constrain $\boldsymbol{\mathcal{D}}$ to only cover areas in the intersection of all observed polygons, $\underline{\boldsymbol{S}}_1 \cap \dots \cap \underline{\boldsymbol{S}}_N$. Equation 4.8 is also slightly changed so that

$$f(|A_u(\boldsymbol{C}, \hat{\boldsymbol{\theta}}, \boldsymbol{S}_1)|, \dots, |A_u(\boldsymbol{C}, \hat{\boldsymbol{\theta}}, \boldsymbol{S}_N)|) < \frac{\delta}{N} \sum_{i=1}^{N} |\underline{\boldsymbol{S}}_{\boldsymbol{i}}| \tag{4.24}$$

for the underestimated representation and

$$f(|A_o(\boldsymbol{C}, \hat{\boldsymbol{\theta}}, \boldsymbol{S}_1)|, \dots, |A_o(\boldsymbol{C}, \hat{\boldsymbol{\theta}}, \boldsymbol{S}_N)|) < \frac{\delta}{N} \sum_{i=1}^{N} |\underline{\boldsymbol{S}}_{\boldsymbol{i}}| \tag{4.25}$$

for the overestimated representation. The value f still represents some function and δ is still a proportion.

4.6.3 Coverage metric for approximately star-shaped polygons

This metric introduced in Section 4.6.1 is essentially the same when applied to approximately star-shaped polygons. However, $R_{i,k}$, the intersection of an observed contour $\boldsymbol{S}_i$ and line ℓ_k, may now contain multiple points since polygon $\underline{\boldsymbol{S}}_i$ is only approximately star-shaped. Assuming the polygons are approximately star-shaped, a single point of intersection will still be more common. Aside from this change, the definition of $W_{i,k} = \mathbb{1}[R_{i,k} \in I_{1-\alpha,k}]$ remains the same when $R_{i,k}$ is composed of multiple points. So, all subsequent definitions and properties in Section 4.6.1 hold as well.

4.6.4 Simulation study of contours enclosing approximately star-shaped polygons

In these simulations, we explore the performance of our method for estimating a GSCM for contours that enclose polygons that are only approximately star-shaped. We simulate contours that vary systematically in how much the polygons they enclose differ from being star-shaped polygons. Figure 4.7 shows examples of the types of contours we will evaluate.

To obtain these contours, we first generate polygons that are star-shaped. We then append sections to these polygons, causing the polygons to often no longer be star-shaped. The appended sections loop around the outside of the initial polygon for some number of points on the initial contour. The number of initial points looped over are selected randomly from some uniform distribution. Appended sections that loop around a larger number of points are longer than appended sections that loop over a smaller number of points. Longer appended sections result in more area that cannot be described with a star-shaped representation than shorter appended sections. How close the appended section is to the initial star-shaped polygon and how wide the appended section are also selected randomly from uniform distributions.

We consider 40 evaluation runs for three different distributions of the number of initial points that the appended section loops over. The initial star-shaped contour is set to be *Shape A* with $\kappa = 2$ and $p = 50$. In each evaluation run, the GSCM is fit to $N = 20$ simulated contours. Rather than fixing δ to determine $\hat{p}$ and $\hat{\boldsymbol{C}}$, we set $\hat{p} = p = 50$ in these simulations and estimate $\hat{\boldsymbol{C}}$ conditional on $\hat{p}$. This choice simplifies the interpretation of our results. (Larger δ's would be needed as the area differing from a star-shaped polygon increases. So, it would be difficult to distinguish whether performance differences were due to changes in $\hat{p}$ or changes in how close to star-shaped the polygons are.) We initially run these simulations using a random location for the appended section and then repeat these simulations with a fixed location for the appended section. Results are reported in Table 4.6.4.

With a random location for the appended section, we find that applying GSCMs still results in reasonable coverage for the 90% and 95% credible intervals. For the 80% credible interval, performance is moderately degraded. Interestingly, average performance does not seem to be correlated with how large the appended section to the contour is. When an appended section is added to a fixed location, we find that the mean coverage across the test lines is relatively accurate; however, the standard deviation is quite high, suggesting that coverage is actually poor in some parts of the contours. Figure 4.8 illustrates this variability in performance across the test lines by plotting the proportion of evaluation runs covered by the 90% credible intervals for each test line individually for the case with the number of initial points looped over distributed Uniform(4, 5).

The location of the fixed appended section is under-covered. In contrast, no obvious patterns are seen when the location of the appended section is random. These results indicate that GSCMs can be applied to contours that enclose polygons that modestly differ from star-shaped. However, if areas differing from the star-shaped representation occur in the same location repeatedly, additional modeling of these areas may be needed to avoid systematic errors in coverage.

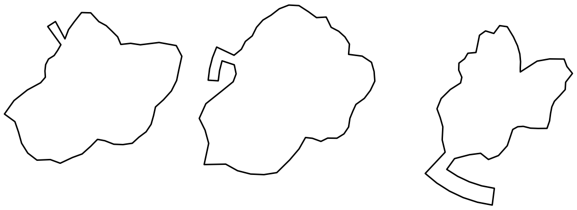


Figure 4.7: Three generated contours that enclose polygons that are only approximately star-shaped. From left to right, contours have increasing area that cannot be described with a star-shaped representation.

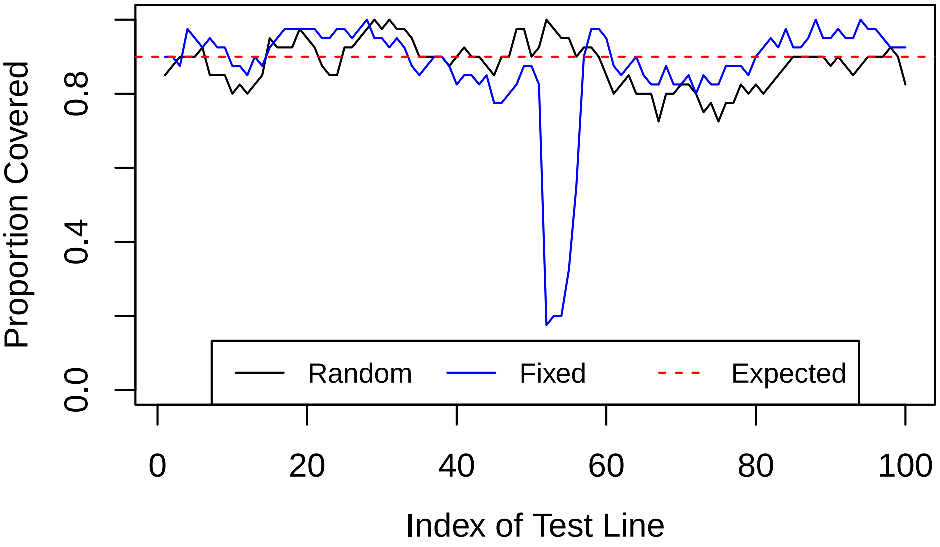


Figure 4.8: Proportion of lines covered out of 40 evaluation runs plotted against the index of each of $M = 100$ evenly-spaced test lines for the 90% credible interval for the approximately star-shaped contours. The black line corresponds to when the appended sections are added to a random location and the blue lines corresponds to when the appended sections are added to a fixed location. The number of initial points looped over are selected randomly from a Uniform (4, 5) distribution. Nominal coverage is in red. Priors and MCMC details are given in Section 4.5.1.

Table 4.5: Mean coverage values for 40 simulations of fitting the contour distribution for approximately star-shaped data. The number of initial points for the appended section to wrap around are selected randomly from a Uniform(a, b) distribution. In each simulation, $N = 20$ simulated observed contours were sampled as data and $M = 100$ evenly-spaced test lines with $\theta_1^* = \pi/M$ were evaluated. All simulations use $p = 50$. The appended sections are located in a random position in the first three cases and a fixed location in the second three cases. Standard deviations across the test lines are given in parentheses. Priors and MCMC details are given in Section 4.5.1

	Nominal	**Unif$(0,1)$**	**Unif$(2,3)$**	**Unif$(4,5)$**
Random Location:	0.8	0.90 (0.04)	0.88 (0.07)	0.90 (0.05)
	0.9	0.96 (0.02)	0.94 (0.04)	0.93 (0.04)
	0.95	0.98 (0.02)	0.97 (0.02)	0.95 (0.03)
Fixed Location:	0.8	0.79 (0.08)	0.77 (0.12)	0.77 (0.16)
	0.9	0.88 (0.06)	0.87 (0.10)	0.87 (0.15)
	0.95	0.94 (0.04)	0.93 (0.07)	0.93 (0.13)

4.7 Example: September Arctic sea ice edge contour

We now return to the motivating example of the Arctic sea ice edge. Modeling the distribution of the sea ice edge is useful for understanding how the range of plausible sea ice edges changes over time. Also, predicting the sea ice edge is helpful for maritime planning, since going through sea ice is expensive in time and fuel. We focus on modeling the sea ice edge in September, the month when the ice-covered area in the Arctic is at its annual minimum. Maritime vessel traffic is typically highest in September because of the large area without sea ice.

Arctic sea ice edge contours observed in previous Septembers provide some information about what types of sea ice edges might be expected in the near future. However, ice edges from past decades are unlikely to be seen again, since, with climate change, the Arctic has seen a rapid reduction in the area covered by sea ice [Comiso et al., 2008, Stroeve et al., 2012b]. September sea ice varies substantially from year to year, so recent observations cover only a small fraction of the physically plausible ice edges that could occur. Therefore, to make statements about how probable sea ice edges will be in the near future and/or to describe plausible sea ice edges that could have been observed in past years' conditions, we need a model for the sea ice edge. The GSCM can be applied to generate plausible sea ice edge contours and corresponding credible intervals for the ice edge contour.

To assess the sea ice edges generated by the GSCM, we perform a leave-one-out cross validation experiment on the September sea ice edge contour in a region in the central Arctic for ten recent years. Data on sea ice are obtained from a monthly-average observational product produced by the National Aeronautics and Space Administration satellites Nimbus-7 SMMR and DMSP SSM/I-SSMIS and downloaded from the National Snow and Ice Data Center [Comiso, 2000, updated 2017]. Each grid box in this data product has an associated concentration, or percentage of ice-covered area. To form the sea ice edge, we categorize all grid boxes with at least 15% concentration as containing sea ice and all grid boxes with less than 15% concentration as open water. This thresholding is often used by sea ice researchers,

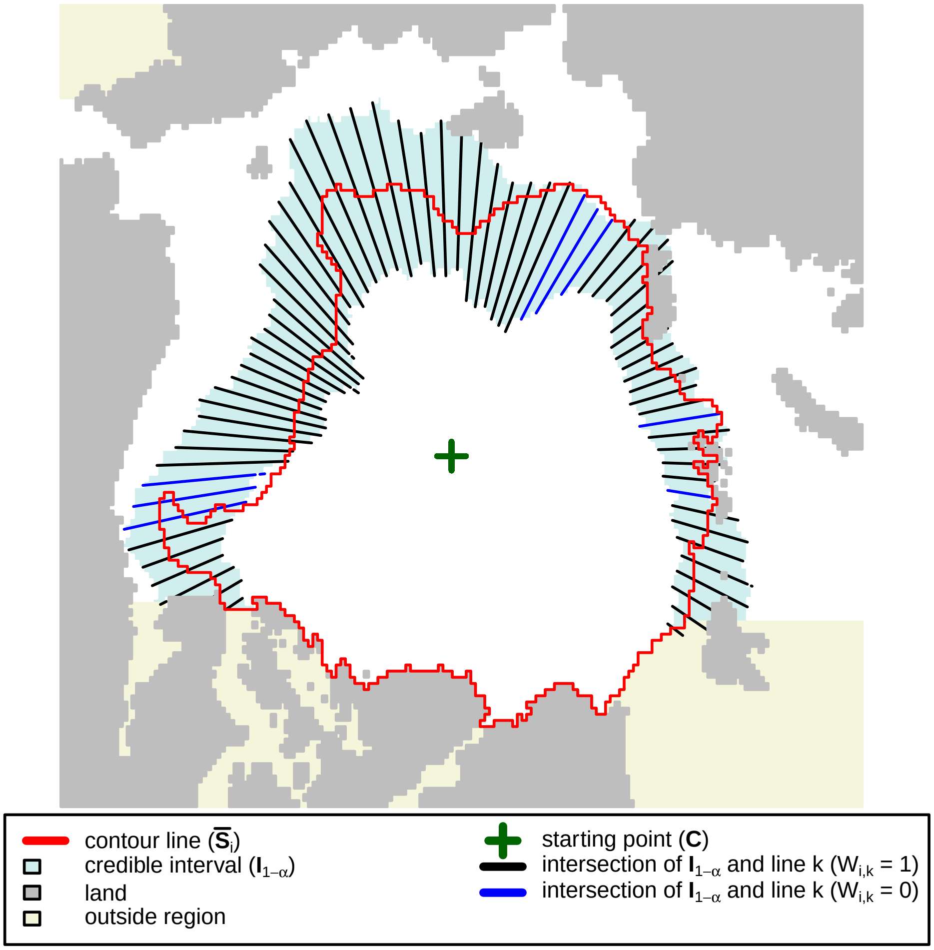


Figure 4.9: The September 2017 sea ice edge contour for the central Arctic region (red) with an 80% credible region fitted from the GSCM with data from 2008-2016. Line segments, $I_{0.8,k}$, corresponding to the intersection of the $\boldsymbol{I}_{0.8}$ credible region and line ℓ_k are colored black when they cover the contour and blue otherwise. The starting point of the evaluation is denoted by a green cross sign.

because the transition from complete sea ice cover to open water occurs over a narrow spatial range and satellite estimates for the low concentrations are not considered reliable.

We assume for this analysis that the ice edge contours observed over the different years are independent and that the contour distribution of the ice edge is stationary over time. While this stationarity assumption is not strictly true given climate change, for a time period of ten years the effects of the climate change trend are small relative to year-to-year variability. For each year j from 2008 to 2017, we fit the GSCM using the contours observed in the other nine years. We then try to "predict" the distribution of possible ice edges that would have been plausible in year j. Data have been rescaled as described in Section 4.3.2 with $\epsilon = 0.1$. We exclude from fitting a section of the ice edge contour that always borders land, since there is no variability in the associated $\boldsymbol{y}$ values. We use the overestimated approximation, since for some maritime planning applications, predicting too much sea ice may be less dangerous than predicting too little sea ice.

As shown in Section 4.5.2, nine years of data may not be sufficient for fitting a GSCM well. In this context, though, we have considerable information that we can incorporate into the prior to augment the observations. In particular, the contour in the area we are trying to predict is bounded primarily by land and also by the boundary around the region. So, the lines in the line set $\boldsymbol{\mathcal{L}} = (\ell_1, \dots, \ell_p)$ are also bounded. We leverage their fixed maximal lengths in setting the priors. Specifically, we let $\mu_{0,i} = 0.5\ell_i$ and $\beta_{0,\sigma_i} = (\ell_i/2)/\Phi^{-1}(.995)$. The latter corresponds to the standard deviation of a normal distribution with 99% of its mass falling in the interval $(0, ||\ell_i||)$ where $||\ell_i||$ is the length of ℓ_i. We also set Λ_0 be a diagonal matrix with 0.5^2 on the diagonal and $\beta_{\kappa_0} = 10$. We use MCMC for fitting with 50,000 iterations, of which 25,000 are omitted as burn-in.

We evaluate coverage as in Section 4.4 using the approximately optimal starting point selected using the method in Section 4.6.2 with $\delta = 0.05$ and $p = 100$. We exclude from evaluation test lines that always border land and have no variability. This leaves 71 test lines. To illustrate this coverage evaluation, we plot in Figure 4.9 the 80% credible interval estimated for the 2017 sea ice edge fit with data from the other nine years. As in Section

Table 4.6: Mean coverage for a leave-one-out cross-validation study for the 2008-2017 September sea ice edge. $M = 71$ evenly-spaced test lines were evaluated with $\theta_1^* = \pi/M$. Standard deviations across the test lines are given in parentheses.

Nominal	**Mean**
0.80	0.75 (0.12)
0.90	0.87 (0.08)
0.95	0.92 (0.07)

4.4 lines extending outward from the estimated starting point are colored blue if they do not cover the left-out contour and black otherwise. Table 4.7 reports the mean coverage of the credible intervals averaged over the ten years and 71 testing lines. Three nominal α-levels are considered. The observed and nominal coverage values are similar, suggesting that the GSCM has appropriate coverage. More broadly, we have shown that GSCM has potential for generating and describing real contours of scientific interest.

4.8 Discussion

We have introduced the GSCM for modeling contours that enclose polygons that are star-shaped or approximately star-shaped. Unlike many existing contour methods, GSCM is designed to be applied to data that are composed of multiple observations of connected sequences of points. Simulation studies illustrated how GSCMs provide accurate coverage under a number of different scenarios. Analysis of September Arctic sea ice also showed how GSCMs can be applied to sea ice edges. We conclude this chapter with a discussion of how GSCMs relate to other contour models and directions for future research.

4.8.1 Other approaches

A large body of research addresses contours in other contexts. The star-shaped model provides tools for modeling contours in the case where multiple contour boundaries are directly observed and the distribution of possible contours is the primary object of inference. Other methods for contours and boundaries may be appropriate for other applications.

Much of the existing research on boundaries relates to level exceedances, also called excursions. The classic level exceedance problem refers to inferring the contour enclosing regions where some latent process exceeds a certain level u. Inference is based on measurements taken at various spatial points on a random spatial field. Early work by Polfeldt [1999] considers how to make statements about the accuracy of contour maps in this context. Lindgren and Rychlik [1995] first define contour uncertainty regions using unions of crossing intervals, or line sections where transitions from below and above u occur.

More recently, Bolin and Lindgren [2015] introduce a method for inferring exceedance levels with irregularly-spatial measurements when the latent spatial field is Gaussian. Their approach provides a way to make global statements about the uncertainty of the full contour. Bolin and Lindgren [2017] then extend this method to estimate the uncertainty of multiple contours to produce contour maps with appropriate uncertainty estimates. Both methods leverage Integrated Nested Laplace Approximations for efficient computation and can be implemented with the `excursions` R package [Bolin and Lindgren, 2018].

French [2014] provides an alternative simulation-based method for making global statements about the location of the contour. Methods for identifying the exceedance region are also explored from both Bayesian and frequentist perspectives [French and Sain, 2013, French and Hoeting, 2016]. Level exceedance methods and GSCMs both focus on inferring contours and their uncertainty. However, in the former boundaries and their uncertainty are inferred with measurements of a continuous process made at spatially-referenced points while in the latter distributions of plausible contours are inferred from direct observations of the contour boundaries themselves.

Wombling methods also focus on spatial boundaries. First considered by Womble [Womble, 1951], these methods typically apply bilinear interpolation to spatially-referenced data points. The gradients of the interpolated functions are used to infer boundaries. Jacquez et al. [2000] summarizes early research on primarily deterministic methods for identifying these boundaries. Principled Bayesian statistical methods have since been developed [Banerjee and Gelfand, 2006, Gelfand and Banerjee, 2015] and recent research has introduced Wombling methods for areal data [Lu and Carlin, 2005, Li et al., 2015] and point processes [Liang et al., 2009]. Wombling boundaries are inferred from spatially-referenced or areal data. The proposed GSCM differs from Wombling techniques, since it is targeted to be applied to repeated direct observations of contour boundaries.

Statistical shape analysis [e.g. Dryden and Mardia, 2016, Srivastava and Klassen, 2016] describes features and variation around boundaries. Shapes typically have consistent and definable features. In these types of applications, location and rotational effects are often ignored in describing the distribution around the shape. As an example, shape analysis is often appropriate for biological imaging research. Deformable templates were developed to describe distributions around shapes with definable features such as the parts of a hand [Amit et al., 1991, Grenander and Keenan, 1993, Grenander and Miller, 1998]. The applications motivating GSCMs differ from the applications motivating shape analysis in that the physical location of the boundary is of interest and no features are present.

Many image analysis methods have been developed to segment images or identify edges. Techniques such as mathematical morphology [Haralick et al., 1987, Lee et al., 1987], watershed segmentation [e.g. Gauch, 1999], and more recently deep learning are applied to identify or sharpen the uncertainty of a single observed boundary in an image. The goals of these methods again differ from the goal of GSCM to define variability over multiple observations of boundaries.

Another alternative to GSCMs would be to model directly whether the points on a lattice are inside or outside a contour boundary. However, methods for modeling binary data on a lattice such as the autologistic [Besag, 1974], centered autologistic [Caragea and Kaiser, 2009], and the Spatial Generalized Linear Mixed Model [Besag et al., 1991, Diggle et al., 1998, Hughes and Haran, 2013] are not structured to guarantee that all the grid boxes inside the contour form in a contiguous section. Hence, these methods are not designed directly for modeling contour boundaries.

4.8.2 Fractal contours and GSCMs

We have treated contours as connected sequences of points, but many contours have fractal-like properties. We now discuss how a fractal contour could be converted to a connected sequence of points for modeling with a GSCM. A true fractal contour, represented by a set F, could be approximately represented by a smaller set of points $\boldsymbol{S}$. In a 2-dimensional Euclidean space, $\mathbb{R}^2$, consider a countable or finite set $\{U\}$ of circles of radius δ. We say $\{U\}$ covers F if $F \subset \cup_{i=1}^{\infty} U_i$ and we refer to the set $\{U\}$ as a δ-cover of F. The value $N_\delta(F)$ is the smallest number of circles of radius δ that could be used to cover F [Falconer, 2004, pp. 27-28]. Let $\{U^*\}$ denote one covering that contains $N_\delta(F)$ circles. Since the contour F is assumed to be finite, the number of circles in $\{U^*\}$ will also be finite. In Figure 4.10, we plot a δ-cover over a visualization of a fractal contour F with several finite self-similar layers. The δ-cover plotted is for visualization and may not contain exactly $N_\delta(F)$ circles.

We can define the elements in the sequence of points forming the contour, $\boldsymbol{S}$, to be the starting points, $\{M^*\}$, of the circles in $\{U^*\}$. The starting points should be arranged in the order in which they would be touched if one were to trace over the fractal contour line, F. Since the contour is a closed loop, where and in what direction to start tracing the fractal contour, F, only affects the indexing of the starting points and not the contour line formed by connecting these points. With this procedure the distance from any point F to a point in $\boldsymbol{S}$ is no more than δ. In general there are multiple δ-cover that contains $N_\delta(F)$ circles; therefore, a criterion needs to be specified as to which set $\{U^*\}$ is used to define $\boldsymbol{S}$. For example, the set $\{U^*\}$ could be selected to be the δ-cover with $N_\delta(F)$ circles that has the circle center closest to the highest $x-$ and highest $y-$ coordinate in the domain, i.e. the top right corner of the domain.

Once $\boldsymbol{S}$ has been obtained, GSCMs can be used. In particular, the under- or overestimated star-shaped representation to any fractal contour be found for $\boldsymbol{S}$ as described in Definitions 12 and 13. A star-shaped representation can represent the area contained within some fractal contours fairly well. For example, the visualized F and the overestimated star-shaped representation for $\boldsymbol{S}$ obtained using the visualized δ-cover are similar. The differing area represents only 7.4% of the total area of the visualized fractal contour when $p = 200$ lines are used in the line set.

4.8.3 Limitations and extensions

A limitation of GSCM is that it can be applied only to contours that enclose polygons that are star-shaped or approximately star-shaped. Using multiple starting points would be one way to relax this assumption. From each starting point a different star-shaped contour could be generated. These individual contours could be combined to produce an overall contour that encloses a polygon that is not star-shaped. Another promising direction for relaxing the star-shaped assumption is to develop a method for appending sections to an initial contour that encloses star-shaped polygons. Extending the way contours were generated in the simulation in Section 4.6.4 might be a natural place to start to do this.

Another limitation is that GSCM assumes all generated line lengths $\boldsymbol{y}$ to be drawn from a Gaussian distribution. While GSCMs can cover a range of shapes, in some applications the observed y may be skewed, bounded above, or otherwise differ from being distributed approximately normally. Therefore, exploring alternatives to GSCM that allow for more flexible distributions of y would be valuable.

Overall, though, GSCMs provide a promising avenue for modeling data composed of multiple observed contours. Representing contours with sequences of points, which are of lower dimension than spatial fields, could enable more detailed and efficient modeling of contour boundaries.

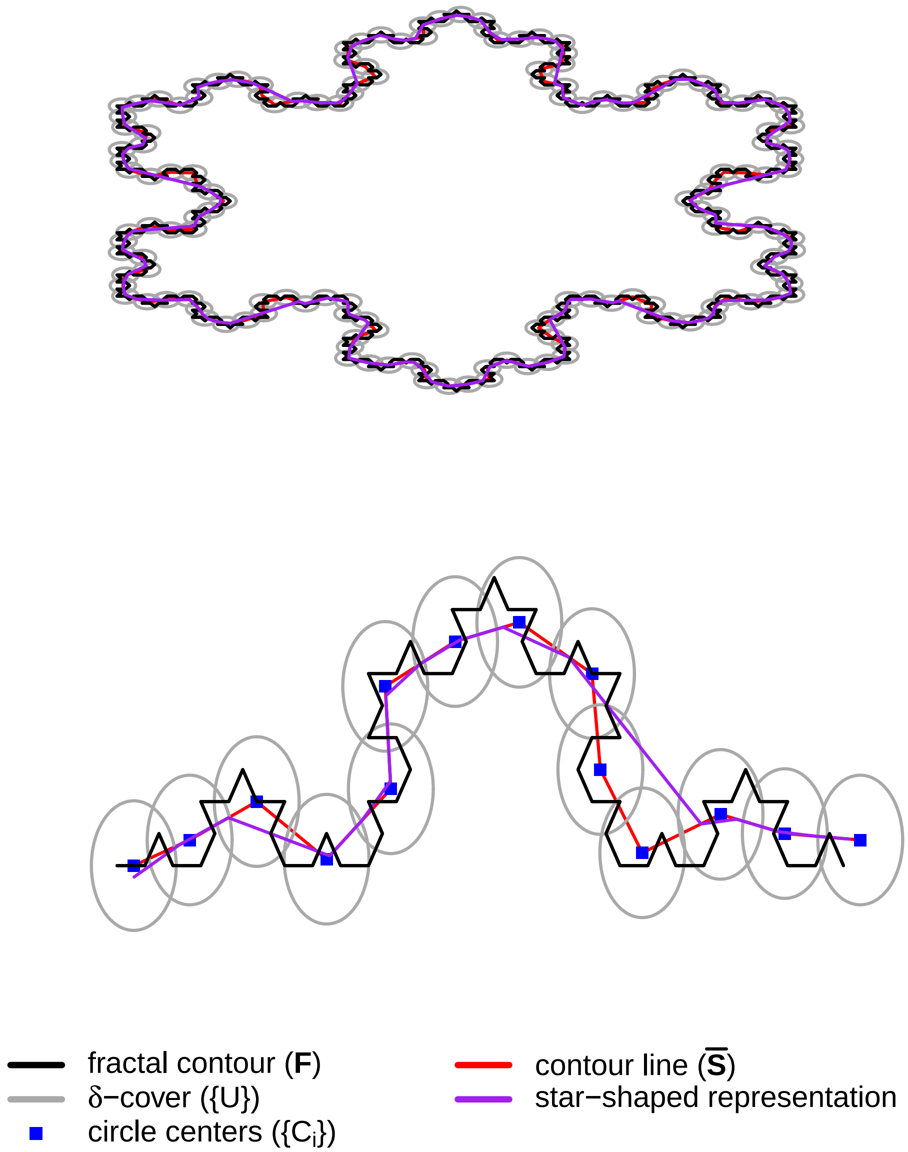


Figure 4.10: Top: A visualization of a fractal contour, $\boldsymbol{F}$, with several self-similar layers shown (black line), a δ-cover of $\boldsymbol{F}$ (grey circles), and a line connecting the starting points of the δ-cover (red line). Bottom: Magnified section of the top figure with circle centers, $\boldsymbol{S}$, (blue squares), added. Note that the δ-cover shown is just a visualization and may not contain $N_\delta(F)$ elements.

Chapter 5

DISCUSSION

This chapter reviews the key advancements made in this dissertation and discusses future research that could stem from this work. More specific summary and discussion are found within each chapter (see Sections 2.4, 3.5, and 4.8).

5.1 Summary of research

Motivated by the need for better forecasts of the Arctic sea ice edge weeks to months in advance, I have developed models for inference and prediction of contours. Chapters 2 and 3 introduced methods for forecasting sea ice using a combination of statistical modeling and dynamic ensemble model output. Chapter 2 focused on bias correction of dynamic ensemble model output and Chapter 3 focused on statistical methods for issuing calibrated forecasts. General methods for modeling contours as connected sequences of points were considered in Chapter 4.

The main contribution of Chapter 2 is the development of Contour-Shifting, a technique for correcting systematic biases in the mean sea ice edge forecast from dynamic ensembles. Differences between observations and retrospective forecasts of the sea ice edge are modeled to predict expected errors in future forecasts. Contour-Shifting accounts for location, time-dependence, and seasonality of expected errors. Retrospective forecasts from the GFDL CM2.5 FLOR model [Vecchi et al., 2014, Msadek et al., 2014] for 2001-2013 are used to evaluate the method. Across four lead times, Contour-Shifting is shown to substantially improve dynamic ensemble forecast accuracy of where sea ice is located. Contour-Shifting also increases the accuracy of dynamic ensemble forecasts of the total sea ice extent.

Chapter 3 develops the Mixture Contour Forecasting (MCF) method. MCF enables issuing probabilistic forecasts of where sea ice will be located weeks to months in advance. Forecasts are based on a new contour probability model that represents the ice edge contour as a connected sequence of points. This model is fit with recent observational data and model output from the Contour-Shifted dynamic ensemble. These contour-based forecasts are weighted with climatological data for each grid box in a two-component mixture model. For an evaluation period from 2008-2016 using retrospective ECMWF forecasts [available from Copernicus Climate Change Service, 2019], the MCF method produces forecasts that maintain accuracy and improve calibration compared to existing dynamic and statistical forecasting methods.

In Chapter 4 the ideas for modeling ice edge contours developed in Chapters 2 and 3 are clarified and generalized for statistical use more broadly. A probabilistic model for contours is introduced that represents contours as ordered, connected sequences of points. Points on the contours are approximated with sets of lengths defining how far the contour extends in various directions from a fixed point. To sample from this contour distribution, sets of lengths are sampled from a probability distribution that accounts for spatial correlation. This model is referred to as the Gaussian Star-Shaped Contour Model (GSCM), since it is most appropriate for contours that enclose star-shaped or approximately star-shaped polygons. The performance of GSCMs is demonstrated with simulation studies and a cross-validation experiment with sea ice edge contours. A metric for assessing the coverage of contours enclosing star-shaped region is also introduced.

5.2 Future research

The models and ideas in this dissertation can certainly be improved and extended. This section describes several directions for possible future research.

5.2.1 Sea ice forecasting

The Contour-Shifting and MCF methods developed in this dissertation could be further refined. The current Contour-Shifting method considers time series of the lengths the sea ice extends from fixed points in dynamic models and observations. The lengths the sea ice extend are assumed to be changing linearly over time. However, some time series of lengths show non-linear behavior or step changes such as when the ice moves away from a land boundary. More flexible time series models could account for this behavior, although avoiding overfitting may be a challenge. For both observations and retrospective forecasts only short time series are available for fitting.

Additionally, the current Contour-Shifting method identifies the bias in the dynamic model output for each length along the contour independently. Correlation among these systematic errors is expected with particularly high correlation for lengths that are close together in space. More sophisticated models for estimating the bias could be developed that fit all biases in lengths jointly as one model. This approach would likely result in more accurate forecasts, since information among locations would be shared.

In this dissertation, I have implemented Contour-Shifting and MCF with a single dynamic ensemble forecast. However, different dynamic ensembles capture different features of the sea ice more accurately. Research with other post-processing methods has shown that multi-model forecasts can produce more skillful forecasts than single model forecasts [Dirkson et al., 2019a]. Extending Contour-Shifting and MCF to use output from multiple dynamic ensembles could result in more accurate forecasts.

The calibration of forecasts obtained with MCF could also likely be improved by increasing the information from the dynamic ensemble outputs used in forecasts. Currently, dynamic ensemble output is used to inform the mean sea ice edge forecast, but dynamic ensemble output is not used to inform the forecasts' variability. The variability of ensemble members could provide additional information about the range of possible ice edges that are feasible for a particular forecast time. However, how well the variability in the dynamic ensemble compares to observations varies spatially and seasonally. So, methods will need to be developed that carefully determine the relationship among the dynamic ensemble variability and the observational variability for different spatial locations and forecast months. Development of these techniques could be particularly valuable for future forecasts if rates of sea thinning increase. Since thinner sea ice is more variable [Holland et al., 2011], if sea ice thickness begins to change rapidly from one year to the next, the variability from observations in the recent past may not match variability at the forecast time. Information from the dynamic ensemble that accounts for this thinning could be particularly valuable for adjusting MCF to this potential reduction in consistency of variability.

In addition to where sea ice is located, operational forecast users are often interested in sea ice thickness. So, development of methods that jointly model sea ice concentration and sea ice thickness are a natural extension of this research. One approach would be to develop a two-step process for generating draws from a sea ice thickness distribution. First, a sea ice edge contour could be generated using the GSCM developed in this dissertation. Then non-zero sea ice thickness values could be sampled from some spatial field model within the area enclosed by the generated ice edge. A methodological challenge will be to ensure that the sea ice thickness decreases smoothly to zero at the ice edge, rather than abruptly shifting to zero at the ice edge.

5.2.2 Contour modeling

Because modeling of connected sequences of points has been little-studied, numerous extensions of GSCMs are feasible. I have only presented a model for generating a single contour in this dissertation. In various contexts, modeling multiple related contours could also be useful. For example, joint models for contours could be used to infer or predict pairs of inner or outer perimeters. Similarly, two contours of correlated quantities could be modeled jointly. In the former case, the key methodological challenge will be how to ensure that a generated inner contour never extends further than its corresponding generated outer contour. In the latter case, a challenge will be defining a covariance structure that accurately represents correlation both among the individual points on each contours and between the two contours.

GSCMs could also plausibly be extended to provide approximations of processes that change slowly over spatial fields. Consider a spatial process whose values increase, or decrease, monotonically moving outward from a particular point. Such a process might correspond to the concentration of a pollutant moving outward from its source. A small set of jointly-modeled contours could represent where the process exceeds certain levels. The covariance among the contours would need to be structured to ensure monotonicity. This research would likely parallel that of Bolin and Lindgren [2017] who consider the uncertainty of multiple contours on a map in the context of level sets. However, the low-dimensional structure of the GSCM could reduce computation in cases where the process changes slowly enough over space that modeling the process at every spatial location on the grid is unnecessary to infer the boundaries of interest.

A final extension of GSCMs would be to consider how the model can be generalized to use with contours that are not approximately star-shaped. One path to relaxing the star-shaped assumption is to consider how to build contours that extend out from two or more starting points. Such an approach might work well in cases where a contour enclosing a polygon that is not star-shaped contains several sub-areas that are star-shaped. Another possibility would be to initially generate a contour enclosing a star-shaped polygon and then use a secondary probability model to append additional areas to this first area.

BIBLIOGRAPHY

Y. Amit, U. Grenander, and M. Piccioni. Structural image restoration through deformable templates. *Journal of the American Statistical Association*, 86(414):376–387, 1991.

M. W. Arisido, C. Gaetan, D. Zanchettin, and A. Rubino. A Bayesian hierarchical approach for spatial analysis of climate model bias in multi-model ensembles. *Stochastic Environmental Research and Risk Assessment*, 31:2645–2657, 2017.

S. Banerjee and A. E. Gelfand. Bayesian Wombling: Curvilinear gradient assessment under spatial process models. *Journal of the American Statistical Association*, 101(476):1487–1501, 2006.

L. Bao, T. Gneiting, E. P. Grimit, P. Guttorp, and A. E. Raftery. Bias correction and Bayesian model averaging for ensemble forecasts of surface wind direction. *Monthly Weather Review*, 138(5):1811–1821, 2010.

J. Besag. Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society: Series B (Methodological)*, 36(2):192–225, 1974.

J. Besag, J. York, and A. Mollié. Bayesian image restoration, with two applications in spatial statistics. *Annals of the Institute of Statistical Mathematics*, 43(1):1–20, 1991.

E. Blanchard-Wrigglesworth, K. C. Armour, C. M. Bitz, and E. DeWeaver. Persistence and inherent predictability of Arctic sea ice in a GCM ensemble and observations. *Journal of Climate*, 24(1):231–250, 2011a.

E. Blanchard-Wrigglesworth, C. M. Bitz, and M. M. Holland. Influence of initial conditions and climate forcing on predicting Arctic sea ice. *Geophysical Research Letters*, 38(18), 2011b.

E. Blanchard-Wrigglesworth, R. I. Cullather, W. Wang, J. Zhang, and C. M. Bitz. Model forecast skill and sensitivity to initial conditions in the seasonal sea ice outlook. *Geophysical Research Letters*, 42(19):8042–8048, 2015.

E. Blanchard-Wrigglesworth, A. Barthélemy, M. Chevallier, R. Cullather, N. Fučkar, F. Massonnet, P. Posey, W. Wang, J. Zhang, C. Ardilouze, et al. Multi-model seasonal forecast of Arctic sea-ice: forecast uncertainty at pan-Arctic and regional scales. *Climate Dynamics*, pages 1–12, 2016.

D. Bolin and F. Lindgren. Excursion and contour uncertainty regions for latent Gaussian models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 77(1): 85–106, 2015.

D. Bolin and F. Lindgren. Quantifying the uncertainty of contour maps. *Journal of Computational and Graphical Statistics*, 26(3):513–524, 2017.

D. Bolin and F. Lindgren. Calculating probabilistic excursion sets and related quantities using excursions. *Journal of Statistical Software*, 86(5):1–20, 2018.

G.W. Brier. Verification of forecasts expressed in terms of probability. *Monthly Weather Review*, 78(1):1–3, 1950.

P. J. Brockwell and R. A. Davis. *Introduction to Time Series and Forecasting*. Springer, third edition, 2016.

M. Bushuk, R. Msadek, M. Winton, G. A. Vecchi, R. Gudgel, A. Rosati, and X. Yang. Skillful regional prediction of Arctic sea ice on seasonal timescales. *Geophysical Research Letters*, 44(10):4953–4964, 2017.

A. J. Cannon. Multivariate bias correction of climate model output: Matching marginal distributions and intervariable dependence structure. *Journal of Climate*, 29(19):7045–7064, 2016.

P. C. Caragea and M. S. Kaiser. Autologistic models with interpretable parameters. *Journal of Agricultural, Biological, and Environmental Statistics*, 14(3):281, 2009.

D. J. Cavalieri and C. L. Parkinson. Arctic sea ice variability and trends, 1979-2010. *The Cryosphere*, 6(4):881, 2012.

M. Chevallier, D. Salas y Mélia, A. Voldoire, M. Déqué, and G. Garric. Seasonal forecasts of the pan-Arctic sea ice extent using a GCM-based seasonal prediction system. *Journal of Climate*, 26(16):6092–6104, 2013.

M. Collins. Climate predictability on interannual to decadal time scales: the initial value problem. *Climate Dynamics*, 19(8):671–692, 2002.

J. Comiso. Bootstrap sea ice concentrations from Nimbus-7 SMMR and DMSP SSM/I-SSMIS. version 2, 2000, updated 2015.

J. Comiso. Bootstrap sea ice concentrations from Nimbus-7 SMMR and DMSP SSM/I-SSMIS. version 3, 2000, updated 2017.

J. C. Comiso. Characteristics of Arctic winter sea ice from satellite multispectral microwave observations. *Journal of Geophysical Research: Oceans*, 91(C1):975–994, 1986.

J. C. Comiso and C. W. Sullivan. Satellite microwave and in situ observations of the Weddell Sea ice cover and its marginal ice zone. *Journal of Geophysical Research: Oceans*, 91(C8): 9663–9681, 1986.

J. C. Comiso, C. L. Parkinson, R. Gersten, and L. Stock. Accelerated decline in the Arctic sea ice cover. *Geophysical Research Letters*, 35(1), 2008.

Copernicus Climate Change Service. Copernicus climate change service climate data store. `https://cds.climate.copernicus.eu`, 2019.

J .J. Day, S. Tietsche, and E. Hawkins. Pan-Arctic and regional sea ice predictability: Initialization month dependence. *Journal of Climate*, 27(12):4371–4390, 2014.

A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society: Series B (Methodological)*, 39(1):1–22, 1977.

P. J. Diggle, J. A. Tawn, and R. A. Moyeed. Model-based geostatistics. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 47(3):299–350, 1998.

H. M. Director and A. E. Raftery. *ContouR: Implementing Gaussian Star-Shaped Contour Models (GSCMs)*, 2020. R package version 1.0.0, `https://github.com/hdirector/ContouR`.

H. M. Director, A. E. Raftery, and C. M. Bitz. Improved sea ice forecasting through spatiotemporal bias correction. *Journal of Climate*, 30(23):9493–9510, 2017a.

H. M. Director, A. E. Raftery, and C. M. Bitz. *IceCast: Apply Statistical Post-Processing to Improve Sea Ice Predictions*, 2017b. R package version 1.1.0, `https://CRAN.R-project.org/package=IceCast`.

H. M. Director, A. E. Raftery, and C. M. Bitz. *IceCast: Apply Statistical Post-Processing to Improve Sea Ice Predictions*, 2020. R package version 3.0.0, `https://github.com/hdirector/IceCastV3`.

A. Dirkson, B. Denis, and W. J. Merryfield. A multimodel approach for improving seasonal probabilistic forecasts of regional Arctic sea ice. *Geophysical Research Letters*, 2019a.

A. Dirkson, W. J. Merryfield, and A. H. Monahan. Calibrated probabilistic forecasts of Arctic sea ice concentration. *Journal of Climate*, 32(4):1251–1271, 2019b.

F. J. Doblas-Reyes, J. García-Serrano, F. Lienert, A. P. Biescas, and L. R. L. Rodrigues. Seasonal climate predictability and forecasting: status and prospects. *Wiley Interdisciplinary Reviews: Climate Change*, 4(4):245–268, 2013.

D. H. Douglas and T. K. Peucker. Algorithms for the reduction of the number of points required to represent a digitized line or its caricature. *Cartographica: The International Journal for Geographic Information and Geovisualization*, 10(2):112–122, 1973.

I. L. Dryden and K. V. Mardia. *Statistical shape analysis: with applications in R.* John Wiley & Sons, second edition, 2016.

D. S. Dukhovskoy, J. Ubnoske, E. Blanchard-Wrigglesworth, H. R. Hiester, and A. Proshutinsky. Skill metrics for evaluation and comparison of sea ice models. *Journal of Geophysical Research: Oceans*, 120(9):5910–5931, 2015.

B. Ellis and L. Brigham. Arctic marine shipping assessment 2009 report. 2009.

European Centre for Medium-Range Weather Forecasts. ECMWF SEAS5 user guide. `https://www.ecmwf.int/sites/default/files/medialibrary/2017-10/System5_guide.pdf`, 2017.

K. Falconer. *Fractal geometry: mathematical foundations and applications.* John Wiley & Sons, 2004.

J. P. French. Confidence regions for the level curves of spatial data. *Environmetrics*, 25(7): 498–512, 2014.

J. P. French and J. A. Hoeting. Credible regions for exceedance sets of geostatistical data. *Environmetrics*, 27(1):4–14, 2016.

J. P. French and S. R. Sain. Spatio-temporal exceedance locations and confidence regions. *The Annals of Applied Statistics*, 7(3):1421–1449, 2013.

N. S. Fučkar, D. Volpi, V. Guemas, and F. J. Doblas-Reyes. A posteriori adjustment of near-term climate predictions: Accounting for the drift dependence on the initial conditions. *Geophysical Research Letters*, 41(14):5200–5207, 2014.

J. García-Serrano and F. J. Doblas-Reyes. On the assessment of near-surface global temperature and North Atlantic multi-decadal variability in the ENSEMBLES decadal hindcast. *Climate Dynamics*, 39(7-8):2025–2040, 2012.

J. M. Gauch. Image segmentation and analysis via multiscale gradient watershed hierarchies. *IEEE Transactions on Image Processing*, 8(1):69–79, 1999.

Y. Gel, A. E. Raftery, and T. Gneiting. Calibrated probabilistic mesoscale weather field forecasting: The geostatistical output perturbation method. *Journal of the American Statistical Association*, 99(467):575–583, 2004.

A. E. Gelfand and S. Banerjee. Bayesian Wombling: finding rapid change in spatial maps. *Wiley Interdisciplinary Reviews: Computational Statistics*, 7(5):307–315, 2015.

H. R. Glahn and D. A. Lowry. The use of model output statistics (MOS) in objective weather forecasting. *Journal of Applied Meteorology*, 11(8):1203–1211, 1972.

T. Gneiting. Strictly and non-strictly positive definite functions on spheres. *Bernoulli*, 19 (4):1327–1349, 2013.

T. Gneiting, F. Balabdaoui, and A. E. Raftery. Probabilistic forecasts, calibration and sharpness. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 69 (2):243–268, 2007.

T. Gneiting, H. Ševčíková, and D. B. Percival. Estimators of fractal dimension: Assessing the roughness of time series and spatial data. *Statistical Science*, pages 247–277, 2012.

L. Goddard et al. A verification framework for interannual-to-decadal predictions experiments. *Climate Dynamics*, 40(1-2):245–272, 2013.

H. F. Goessling, S. Tietsche, J. J. Day, E. Hawkins, and T. Jung. Predictability of the Arctic sea-ice edge. *Geophysical Research Letters*, 2016.

U. Grenander and D. M. Keenan. On the shape of plane images. *SIAM Journal on Applied Mathematics*, 53(4):1072–1094, 1993.

U. Grenander and M. I. Miller. Computational anatomy: An emerging discipline. *Quarterly of Applied Mathematics*, 56(4):617–694, 1998.

V. Guemas, E. Blanchard-Wrigglesworth, M. Chevallier, J. J. Day, M. Déqué, F. J. Doblas-Reyes, N. S. Fučkar, A. Germe, E. Hawkins, S. Keeley, Koenigk T., Mélia D. S., and Tietsche S. A review on Arctic sea-ice predictability and prediction on seasonal to decadal time-scales. *Quarterly Journal of the Royal Meteorological Society*, 142(695):546–561, 2016.

L. V. Hansen, T. L. Thorarinsdottir, E. Ovcharov, T. Gneiting, and D. Richards. Gaussian random particles with flexible Hausdorff dimension. *Advances in Applied Probability*, 47 (2):307–327, 2015.

R. M. Haralick, S. R. Sternberg, and X. Zhuang. Image analysis using mathematical morphology. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-9(4): 532–550, 1987.

W. Hazeleger, V. Guemas, B. Wouters, S. Corti, I. Andreu-Burillo, F. J. Doblas-Reyes, K. Wyser, and M. Caian. Multiyear climate predictions using two initialization strategies. *Geophysical Research Letters*, 40(9):1794–1798, 2013.

M. M. Holland, D. A. Bailey, and S. Vavrus. Inherent sea ice predictability in the rapidly changing Arctic environment of the Community Climate System Model, version 3. *Climate Dynamics*, 36(7-8):1239–1253, 2011.

P. J. Huber and E. M. Ronchetti. *Robust Statistics*. Springer, second edition, 2011.

J. Hughes and M. Haran. Dimension reduction and alleviation of confounding for spatial generalized linear mixed models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 75(1):139–159, 2013.

H. P. Huntington, R. Daniel, A. Hartsig, K. Harun, M. Heiman, R. Meehan, G. Noongwook, L. Pearson, M. Prior-Parks, M. Robards, and Stetston G. Vessels, risks, and rules: Planning for safe shipping in Bering Strait. *Marine Policy*, 51:119–127, 2015.

G. M. Jacquez, S. Maruca, and M. J. Fortin. From fields to objects: a review of geographic boundary analysis. *Journal of Geographical Systems*, 2(3):221–241, 2000.

S. J. Johnson, T. N. Stockdale, L. Ferranti, M. A. Balmaseda, F. Molteni, L. Magnusson, S. Tietsche, D. Decremer, A. Weisheimer, G. Balsamo, S. P. E. Keeley, K. Mogensen, H. Zuo, and B. M. Monge-Sanz. SEAS5: the new ECMWF seasonal forecast system. *Geoscientific Model Development*, 12(3), 2019.

P. W. Jones. A user's guide for SCRIP: a spherical coordinate remapping and interpolation package. *Los Alamos National Laboratory, Los Alamos, USA*, 1997.

V. V. Kharin, G. J. Boer, W. J. Merryfield, J. F. Scinocca, and W. S. Lee. Statistical adjustment of decadal predictions in a changing climate. *Geophysical Research Letters*, 39 (19), 2012.

W. Kleiber, A. E. Raftery, J. Baars, T. Gneiting, C. F. Mass, and E. Grimit. Locally calibrated probabilistic temperature forecasting using geostatistical model averaging and local Bayesian model averaging. *Monthly Weather Review*, 139(8):2630–2649, 2011a.

W. Kleiber, A. E. Raftery, and T. Gneiting. Geostatistical model averaging for locally calibrated probabilistic quantitative precipitation forecasting. *Journal of the American Statistical Association*, 106(496):1291–1303, 2011b.

F. Krikken, M. Schmeits, W. Vlot, V. Guemas, and W. Hazeleger. Skill improvement of dynamical seasonal Arctic sea ice forecasts. *Geophysical Research Letters*, 2016.

D. Lee and F. P. Preparata. An optimal algorithm for finding the kernel of a polygon. *Journal of the Association for Computing Machinery*, 26(3):415–421, 1979.

J. Lee, R. Haralick, and L. Shapiro. Morphologic edge detection. *IEEE Journal on Robotics and Automation*, 3(2):142–156, 1987.

P. Li, S. Banerjee, T. A. Hanson, and A. M. McBean. Bayesian models for detecting difference boundaries in areal data. *Statistica Sinica*, 25(1):385, 2015.

S. Liang, S. Banerjee, and B. P. Carlin. Bayesian Wombling for spatial point processes. *Biometrics*, 65(4):1243–1253, 2009.

G. Lindgren and I. Rychlik. How reliable are contour curves? Confidence sets for level contours. *Bernoulli*, 1(4):301–319, 1995.

E. N. Lorenz. Deterministic nonperiodic flow. *Journal of the Atmospheric Sciences*, 20(2): 130–141, 1963.

H. Lu and B. P. Carlin. Bayesian areal Wombling for geographical boundary analysis. *Geographical Analysis*, 37(3):265–285, 2005.

B. Mandelbrot. How long is the coast of Britain? Statistical self-similarity and fractional dimension. *Science*, 156(3775):636–638, 1967.

D. Maraun. Bias correcting climate change simulations—a critical review. *Current Climate Change Reports*, 2(4):211–220, 2016.

G. A. Meehl et al. Decadal climate prediction: an update from the trenches. *Bulletin of the American Meteorological Society*, 95(2):243–267, 2014.

N. Melia, K. Haines, and E. Hawkins. Sea ice decline and 21st century trans-Arctic shipping routes. *Geophysical Research Letters*, 43(18):9720–9728, 2016.

J. W. Messner, G. J. Mayr, A. Zeileis, and D. S. Wilks. Heteroscedastic extended logistic regression for postprocessing of ensemble guidance. *Monthly Weather Review*, 142(1): 448–456, 2014.

R. Msadek, G. A. Vecchi, M. Winton, and R. G. Gudgel. Importance of initial conditions in seasonal predictions of Arctic sea ice extent. *Geophysical Research Letters*, 41(14): 5208–5215, 2014.

National Center for Atmospheric Research. Earth system grid at NCAR. `https://www.earthsystemgrid.org/home.html`, 2017.

National Snow and Ice Data Center. Region mask for the northern hemisphere. `http://nsidc.org/data/polar-stereo/tools_masks.html`, 2017.

C. L. Parkinson and D. J. Cavalieri. Arctic sea ice variability and trends, 1979–2006. *Journal of Geophysical Research: Oceans*, 113(C7), 2008.

M. Plummer, N Best, K. Cowles, and K. Vines. Coda: Convergence diagnosis and output analysis for MCMC. *R News*, 6(1):7–11, 2006. URL `https://journal.r-project.org/archive/`.

T. Polfeldt. On the quality of contour maps. *Environmetrics*, 10(6):785–790, 1999.

F. P. Preparata and M. I. Shamos. *Computational geometry: an introduction.* Springer Science & Business Media, 1985.

A. E. Raftery and S. M. Lewis. Practical Markov chain Monte Carlo: comment: one long run with diagnostics: implementation strategies for Markov chain Monte Carlo. *Statistical Science*, 7(4):493–497, 1992.

A. E. Raftery and S. M. Lewis. The number of iterations, convergence diagnostics and generic metropolis algorithms. *Practical Markov Chain Monte Carlo*, 7(98):763–773, 1995.

A. E. Raftery, T. Gneiting, F. Balabdaoui, and M. Polakowski. Using Bayesian model averaging to calibrate forecast ensembles. *Monthly Weather Review*, 133(5):1155–1174, 2005.

Sea Ice Outlook August 2014 Report. `https://www.arcus.org/sipn/sea-ice-outlook/2014/august`, 2014.

M. I. Shamos. Geometric complexity. In *Proceedings of the Seventh Annual ACM Symposium on Theory of Computing*, pages 224–233, 1975.

M. Sigmond, J. C. Fyfe, G. M. Flato, V. V. Kharin, and W. J. Merryfield. Seasonal forecast skill of Arctic sea ice area in a dynamical forecast system. *Geophysical Research Letters*, 40(3):529–534, 2013.

J. M. Sloughter, A. E. Raftery, T. Gneiting, and C. Fraley. Probabilistic quantitative precipitation forecasting using Bayesian model averaging. *Monthly Weather Review*, 135(9): 3209–3220, 2007.

J. M. Sloughter, T. Gneiting, and A. E. Raftery. Probabilistic wind vector forecasting using ensembles and Bayesian model averaging. *Monthly Weather Review*, 141(6):2107–2119, 2013.

L. C. Smith and S. R. Stephenson. New trans-Arctic shipping routes navigable by midcentury. *Proceedings of the National Academy of Sciences*, 110(13):E1191–E1195, 2013.

A. Srivastava and E. P. Klassen. *Functional and shape data analysis*. Springer, 2016.

S. R. Stephenson, L. W. Brigham, and L. C. Smith. Marine accessibility along Russia's northern sea route. *Polar Geography*, 37(2):111–133, 2014.

J. C. Stroeve, V. Kattsov, A. Barrett, M. Serreze, T. Pavlova, M. Holland, and W. N. Meier. Trends in Arctic sea ice extent from CMIP5, CMIP3 and observations. *Geophysical Research Letters*, 39(16), 2012a.

J. C. Stroeve, M. C. Serreze, M. M. Holland, J. E. Kay, J. Malanik, and A. P. Barrett. The Arctic's rapidly shrinking sea ice cover: a research synthesis. *Climatic Change*, 110(3): 1005–1027, 2012b.

J. C. Stroeve, L. C. Hamilton, C. M. Bitz, and E. Blanchard-Wrigglesworth. Predicting September sea ice: Ensemble skill of the SEARCH sea ice outlook 2008–2013. *Geophysical Research Letters*, 41(7):2411–2418, 2014.

C. Strong. Atmospheric influence on Arctic marginal ice zone position and width in the Atlantic sector, February–April 1979–2010. *Climate Dynamics*, 39(12):3091–3102, 2012.

S. Tietsche, J. J. Day, V. Guemas, W. J. Hurlin, S. P. E. Keeley, D. Matei, R. Msadek, M. Collins, and E. Hawkins. Seasonal to interannual Arctic sea ice predictability in current global climate models. *Geophysical Research Letters*, 41(3):1035–1043, 2014.

G. J. van Oldenborgh, F. J. Doblas-Reyes, B. Wouters, and W. Hazeleger. Decadal prediction skill in a multi-model ensemble. *Climate Dynamics*, 38(7-8):1263–1280, 2012.

G.A. Vecchi et al. On the seasonal forecasting of regional tropical cyclone activity. *Journal of Climate*, 27(21):7994–8016, 2014.

W. N. Venables and B. D. Ripley. *Modern Applied Statistics with S*. Springer, New York, fourth edition, 2002.

W. Wang, M. Chen, and A. Kumar. Seasonal prediction of Arctic sea ice extent from a coupled dynamical forecast system. *Monthly Weather Review*, 141(4):1375–1394, 2013.

N. E. Wayand, C. M. Bitz, and E. Blanchard-Wrigglesworth. A year-round subseasonal-to-seasonal sea ice prediction portal. *Geophysical Research Letters*, 46(6):3298–3307, 2019.

D. S. Wilks. Extending logistic regression to provide full-probability-distribution MOS forecasts. *Meteorological Applications*, 16(3):361–368, 2009.

W. H. Womble. Differential systematics. *Science*, 114(2961):315–322, 1951.

L. Zampieri, H. F. Goessling, and T. Jung. Bright prospects for Arctic sea ice prediction on subseasonal time scales. *Geophysical Research Letters*, 45(18):9731–9738, 2018.

B. Zhang and N. Cressie. Estimating spatial changes over time of Arctic Sea ice using hidden 2× 2 tables. *Journal of Time Series Analysis*, 40(3):288–311, 2019.

B. Zhang and N. Cressie. Bayesian inference of spatio-temporal changes of Arctic sea ice. *Bayesian Analysis*, 15(2):605–631, 2020.

J. Zhuang. xESMF: Universal regridder for geospatial data. https://github.com/JiaweiZhuang/xESMF, 2018.

D. L. Zimmerman and M. L. Stein. Classical geostatistical methods. In A. E. Gelfand, P. Diggle, P. Guttorp, and M. Fuentes, editors, *Handbook of Spatial Statistics*, pages 29–44. CRC Press: Boca Raton, FL, 2010.

Appendix A

APPENDIX TO CHAPTER 3

A.1 Standard deviation corresponding to γ proportion of mass of a Gaussian within symmetric bounds

Consider a Gaussian distribution with known mean $\mu = (m + M)/2$, where $M > m$. The standard deviation, $\sigma > 0$, such that $100 \times \gamma$ percent of the mass of the distribution is within m and M is

$$\sigma = \frac{(M - m)/2}{\Phi^{-1}((1 + \gamma)/2)} \tag{A.1}$$

where $\gamma \in (0, 1)$ and $\Phi^{-1}(\cdot)$ is the standard normal inverse cumulative distribution function.

Proof. Let X be a random variable with $X \sim N(\mu, \sigma)$. Note that the value of σ that produces $\Pr(m \leq X \leq M) = \gamma$ is equivalent to the value of σ that produces $Pr(x \leq M) = (1 - \gamma)/2 + \gamma = (1 + \gamma)/2$. Then,

$$\Pr\left(Z \leq \frac{M - (M + m)/2}{\sigma}\right) = \Pr\left(Z \leq \frac{(M - m)/2}{\sigma}\right) = \frac{(1 + \gamma)}{2}$$

Hence,

$$\frac{(M - m)/2}{\sigma} = \Phi^{-1}\left(\frac{1 + \gamma}{2}\right). \tag{A.2}$$

$\square$

A.2 Additional figures

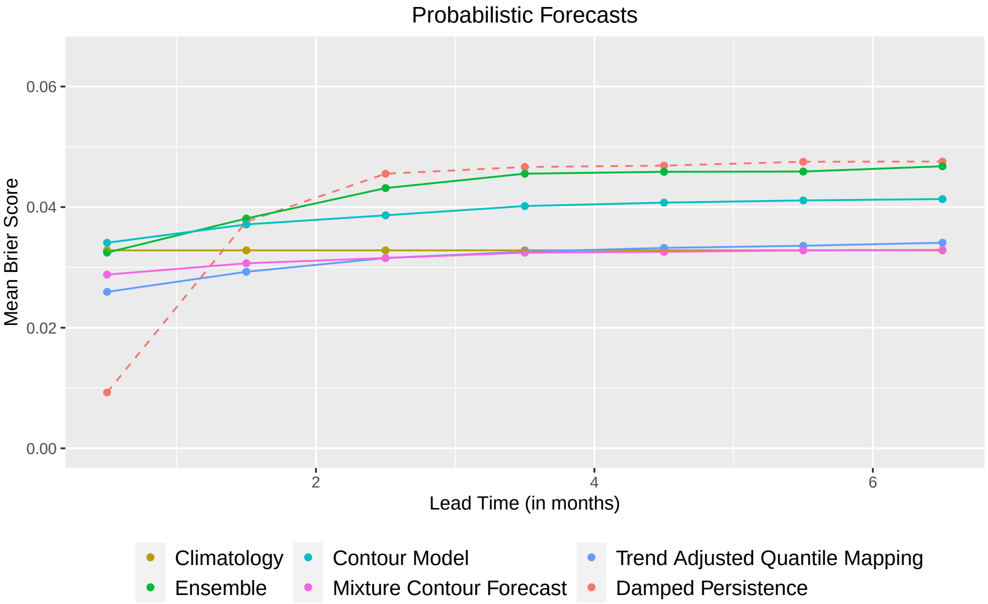


Figure A.1: Overall Brier scores for the test years 2008-2016 for the probabilistic forecasts and a damped persistence reference binary forecast. The Brier Score for each grid box is weighted based on its area. Forecasts are described in Table 1 in the main text.

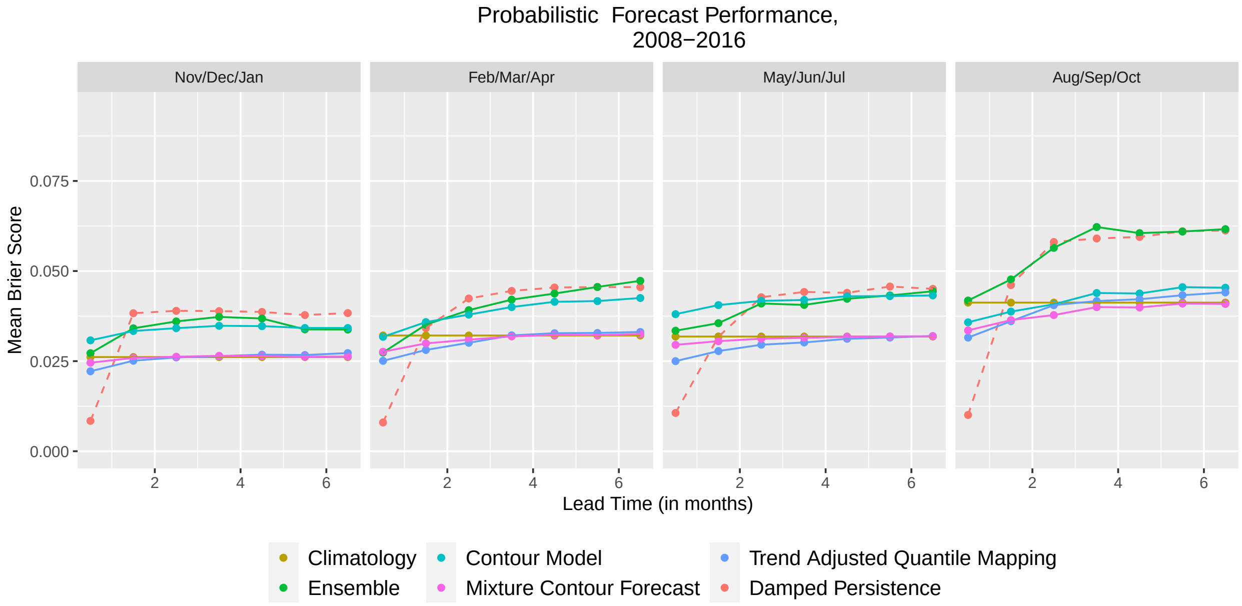


Figure A.2: Average Brier scores grouped into three-month sets for the test years 2008-2016 for the probabilistic forecasts and a damped persistence reference binary forecast. The Brier Score for each grid box is weighted based on its area. Forecasts are described in Table 1 in the main text.

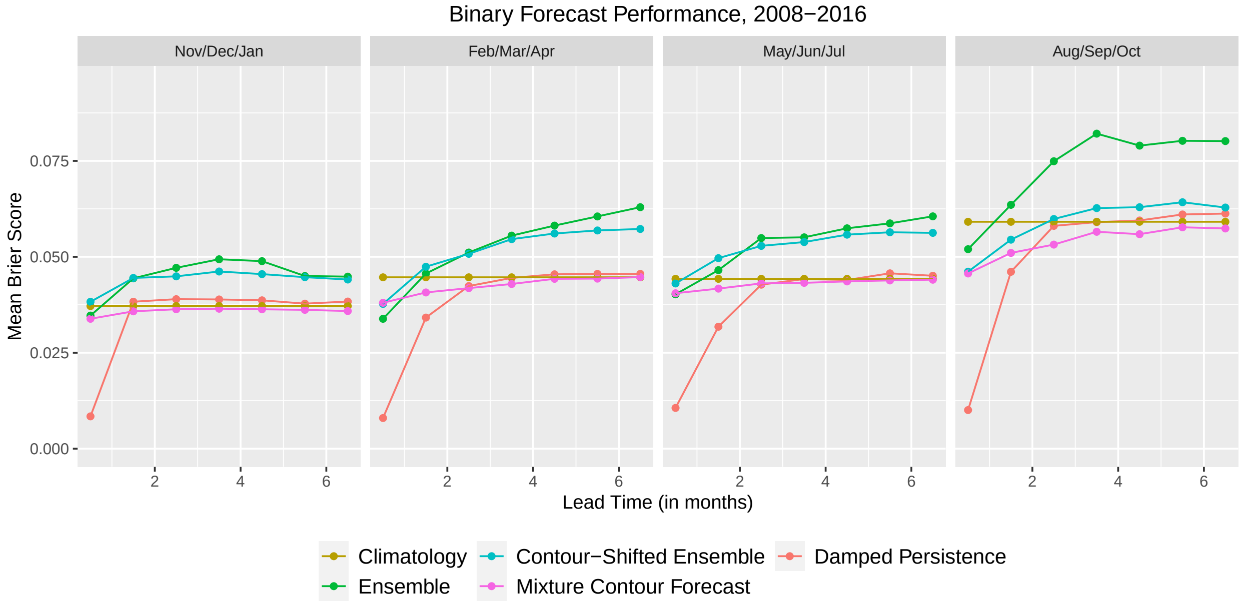


Figure A.3: As in Figure A.2, except for binary forecasts.

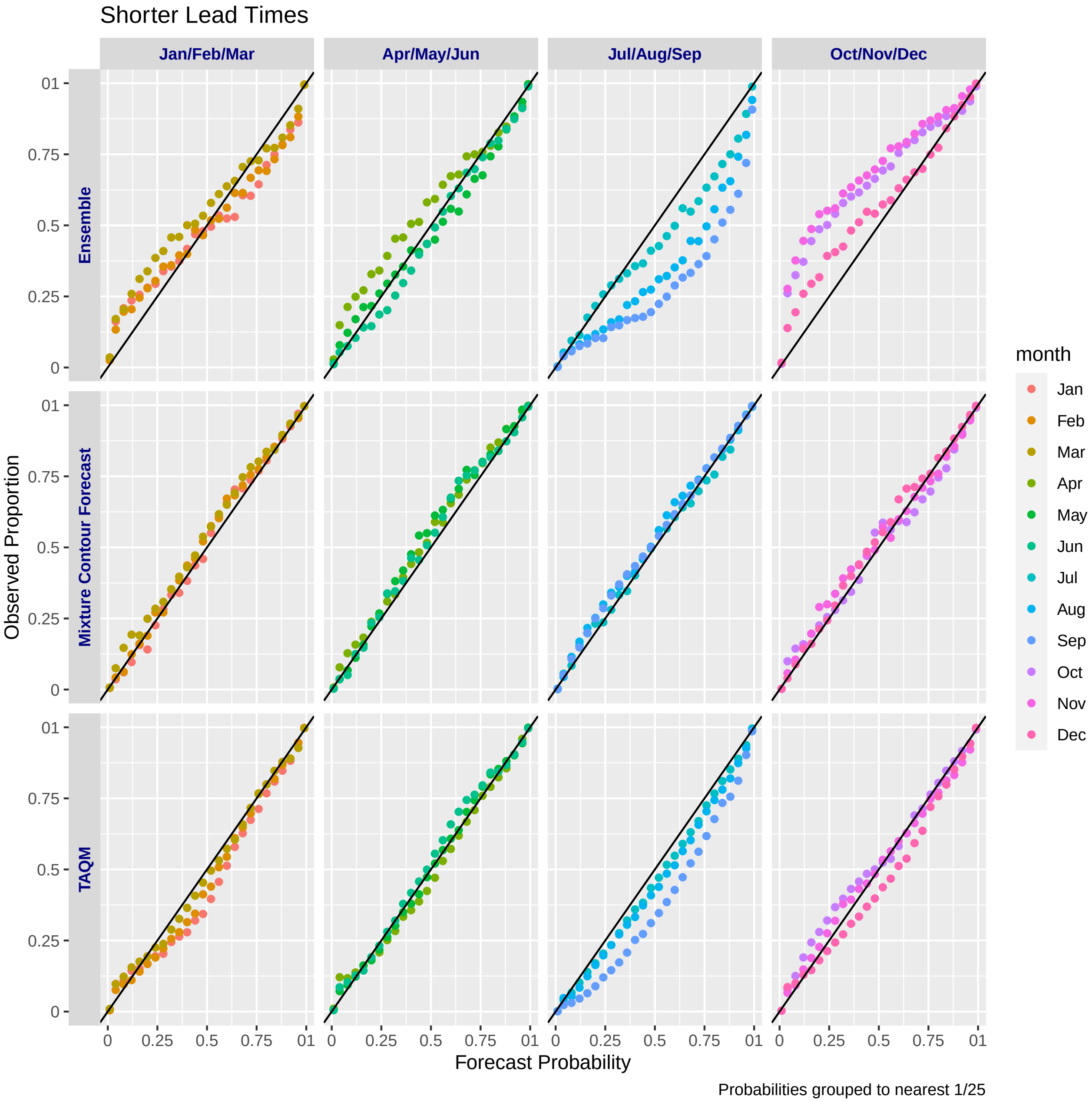


Figure A.4: Plots of the average proportion of times sea ice was observed against the predicted probability of sea ice presence for the raw ECMWF (top), MCF (middle), and TAQM (bottom) forecasts for lead times of 0.5 - 1.5 months. Results are grouped into three-month sets and all grid boxes are equally weighted.

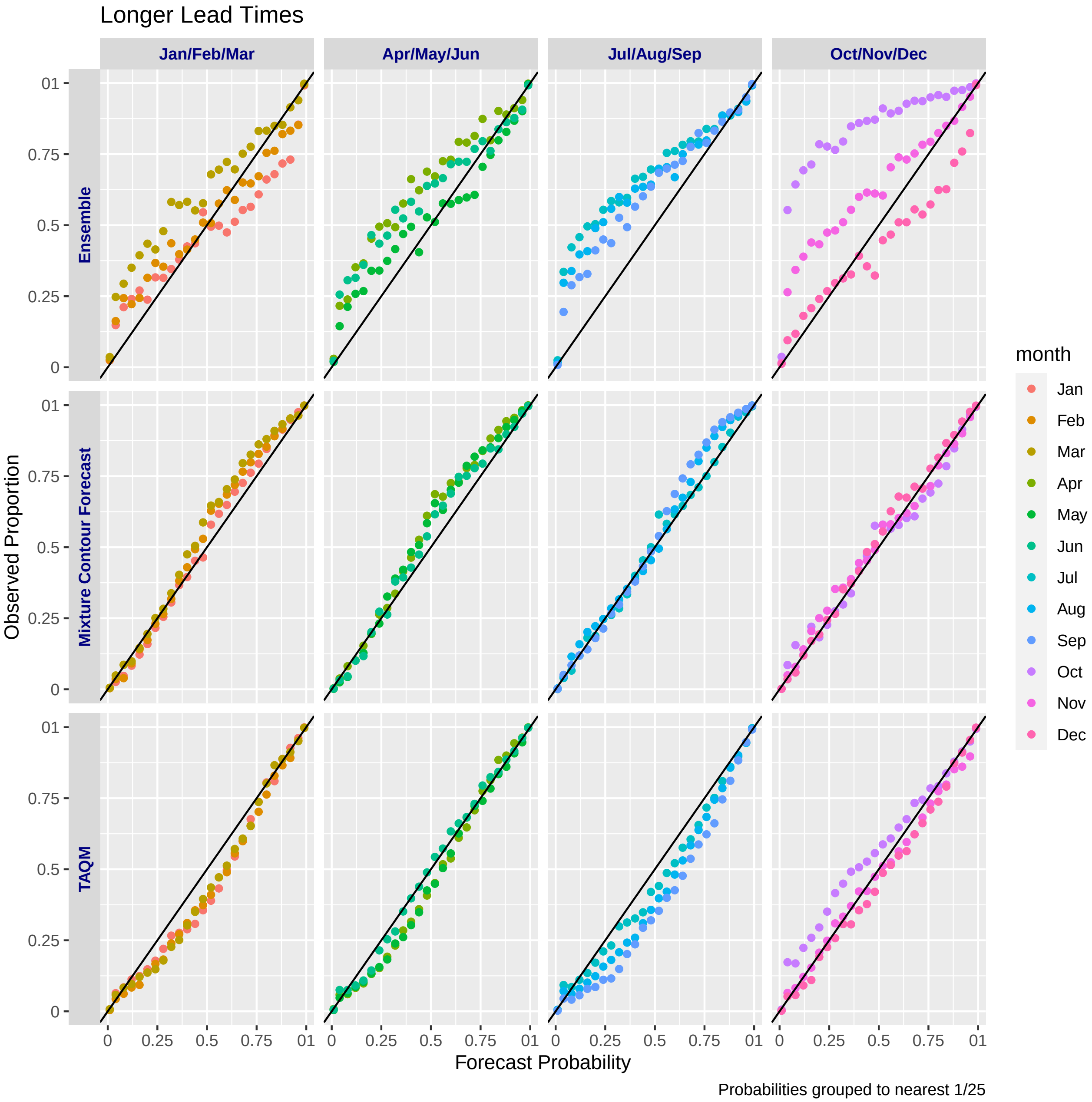


Figure A.5: As in Figure A.4 but for lead times of 2.5 - 6.5 months.

A.3 MCMC diagnostics

As in all MCMC, diagnostic analysis is needed to determine the appropriate chain length. We selected the number iterations to run the sampler by considering traceplots and model diagnostics for sample forecast months. In this Section, we evaluate one month in detail. Our analysis serves a dual purpose. It demonstrates that the number of iterations used in this chapter's analysis is reasonable and serves as a model for how MCMC diagnostics could be applied for future results obtained with this method.

We note that our primary goal in this chapter is prediction and not inference and that only the mean of each parameter distribution is used. So, only reasonable sampling from the posterior distribution is needed for good performance. Therefore, we have not repeated this analysis on every forecast month and year. However, if our goals were to shift to inference on the parameter distributions, we would recommend doing a more thorough evaluation that would involve all parameters in each forecast month and year. We would also recommend repeating this initial diagnostic analysis if major changes are made to the method, such as changing the ensemble prediction system used in setting the hyperparameter $\boldsymbol{\mu}_0$.

A.3.1 Example Evaluation: September 2005, 1.5-month lead

We evaluate the chains for September 2005 at a 1.5-month lead time using the training years of 1995-2004. We selected this month as an example, since the location of the sea ice edge is highly variable at this time of year. The model parameters are consequently likely to have high variability and need more iterations for fitting. Other months could potentially be fit adequately with less iterations.

Three regions contain sea ice in September 2005 and have a contour model fit for them: the Central Arctic, Baffin Bay, and Greenland Sea. Using the *coda* R package [Plummer et al., 2006], we compute the Raftery and Lewis Diagnostic for κ and most values of μ_i and σ_i [Raftery and Lewis, 1992, 1995]. We use r = 0.0125 and report the maximum number of iterations needed after assessing both q = 0.025 and q = 0.975. In Table A.1 and Table A.2, we report the 50-th, 95-th, and 100-th percentiles of the estimated number of iterations needed from all μ_i and σ_i for each region. We omit from this analysis some chains for σ_i and μ_i when more than 95% of the samples are within .05 of one of its boundaries (i.e., the upper or lower bound of the corresponding uniform prior as defined in Section 2.5 and Section 2.6 of the main text.) This omission, or something similar, is needed because in some cases the parameter value that maximizes the posterior is on the boundary of the range. The chain will correctly not move or move little in such cases and the Raftery and Lewis Diagnostic does not make sense. In Table A.3, we report the estimated number of iterations needed for κ for the three regions.

Table A.1: The 50-th, 95-th, and 100-th percentile for the estimated chain lengths for μ_i obtained from the Raftery and Lewis Diagnostic for the three regions evaluated. Values rounded to the nearest whole number.

Region	$n_{est,50}$	$n_{est,95}$	$n_{est,100}$
Central Arctic	5820	10972	12715
Baffin Bay	3840	5414	5489
Greenland Sea	3829	8704	8884

Table A.2: As in Table A.1, except for σ_i

Region	$n_{est,50}$	$n_{est,95}$	$n_{est,100}$
Central Arctic	5380	10745	11520
Baffin Bay	5397	6182	6216
Greenland Sea	7345	15540	16533

In Figure A.6, we show the traceplot for a typical chain for μ_i for each of the three regions. Figure A.7 is an analogous figure for a typical σ_i. We selected the index i in each case by

Table A.3: Estimated chain lengths from the Raftery and Lewis Diagnostic for κ

Region	n_κ
Central Arctic	3856
Baffin Bay	46544
Greenland Sea	36060

finding the chain with the estimated sample size closest to the median estimated number of iterations needed for all indices. Finally, Figure A.8 shows the traceplots for κ for the three regions.

The traceplots illustrate that the chains converge quickly. The κ_i parameter tends to sample the space most slowly, so κ controls the number of iterations needed. In particular, the maximal value from the Raftery and Lewis diagnostic is approximately 45,000. The traceplots show that a burn-in of 5000 is sufficient for the chains to have reached their posterior density region. These results motivate using 55,000 iterations for all chains with the initial 5000 iterations removed as burn-in.

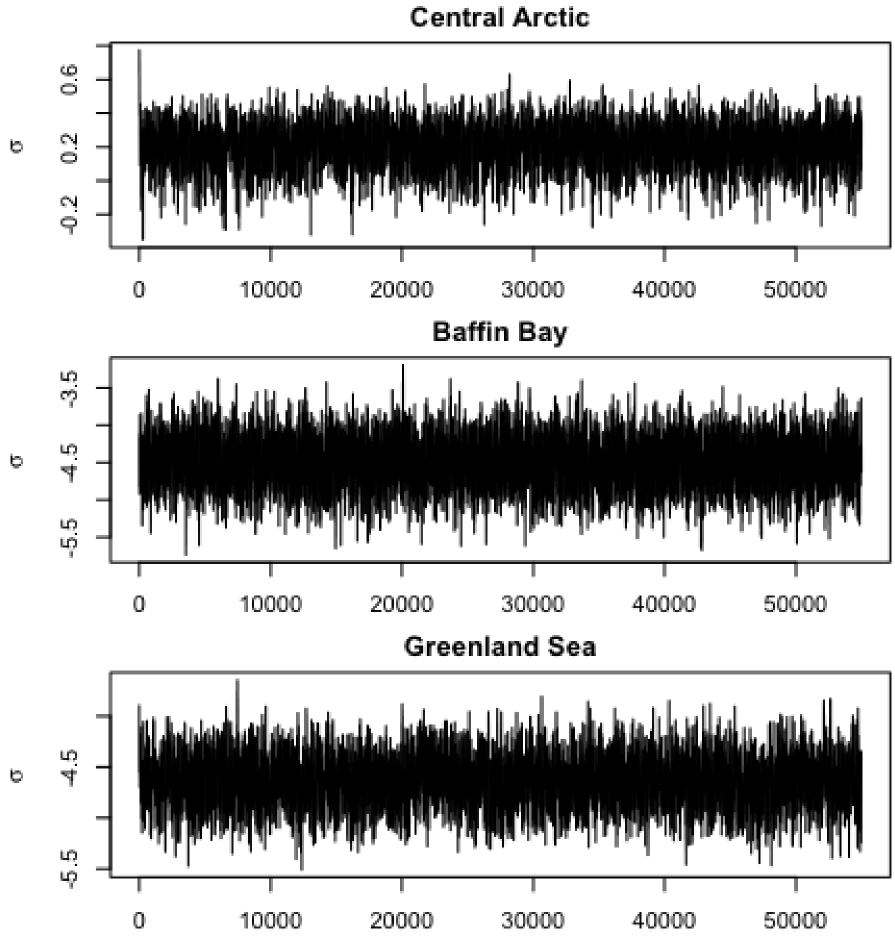


Figure A.6: Traceplots for the chains in each of the three evaluated regions for a typical μ_i.

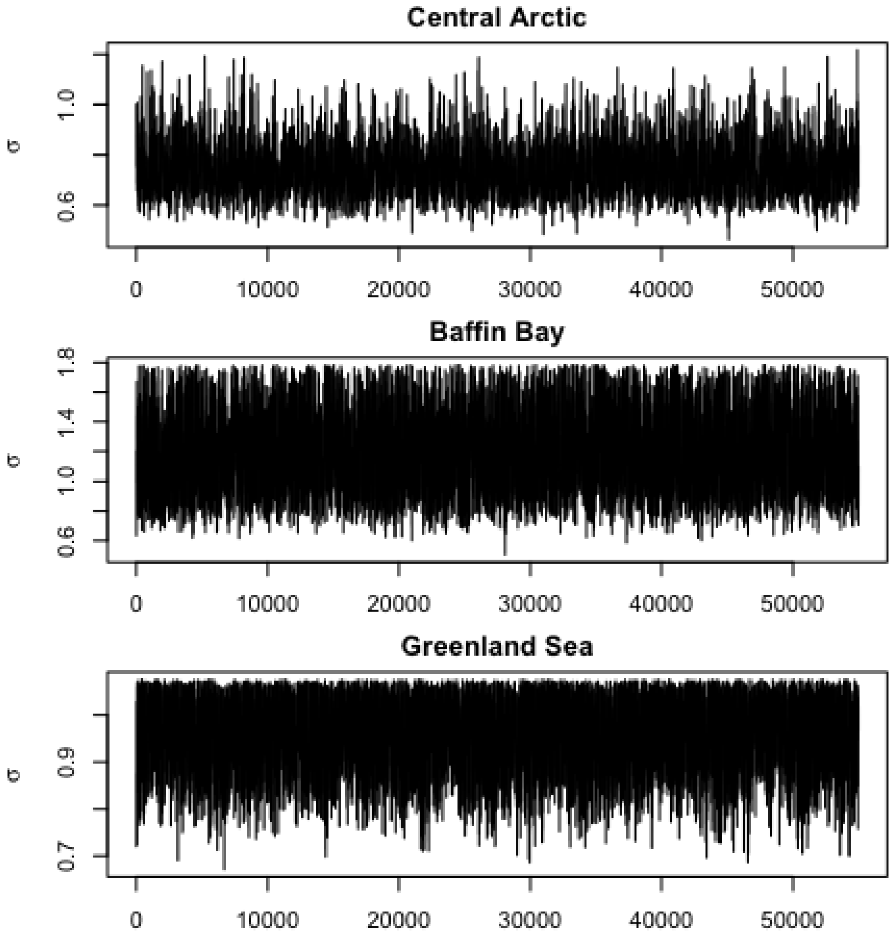


Figure A.7: Traceplots for the chains in each of the three evaluated regions for a typical σ_i.

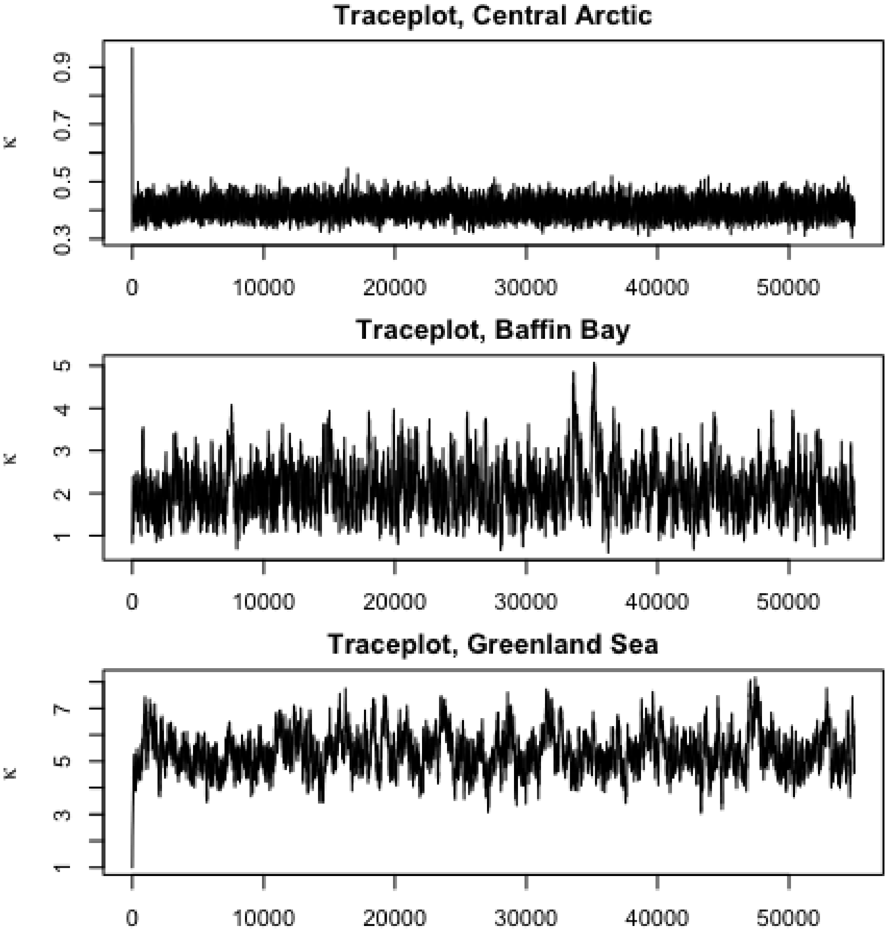


Figure A.8: Traceplots for the parameter κ in the three evaluated regions.

A.4 Length of Training Periods

Several aspects of the modeling in this chapter rely on fitting parameters using multiple previous years of data as a training period. Since the Arctic is changing, more recent data is likely to be more relevant, but using only a small number of years of data may lead to parameter estimates that are too variable. Computational cost and limited amounts of data further constrain the training lengths possible. So, determining the appropriate training window length is not obvious. In this section, we discuss our rationale for the values used in the chapter.

Because Contour-Shifting explicitly models the time trend, we follow Director et al. [2017a] and use all available data prior to the forecast year to fit the bias correction. The earliest ensemble output ECMWF available on the Copernicus Climate Change Service Climate Data Store [Copernicus Climate Change Service, 2019] is for 1993, so that is the earliest year used in fitting Contour-Shifting.

The damped persistence forecast also relies on a trend. So, we use all available training data in fitting. Since this method only requires observations and they are available for earlier years than ensemble output, we fit this model with data beginning in 1981.

In the chapter, we use only three years of training data in a rolling window to fit the weights in the mixture. Since the weights are dependent on results after post-processing, there is not enough data to do a proper cross validation while leaving aside a large enough test set. We find three years works well in practice, but it is not clear this is optimal. For future use, we evaluate the performance of different training lengths for the rolling window. We fit the weights and corresponding MCF forecasts for the years 2012-2016 using rolling windows of training lengths from 1-7 years. We report the mean Brier score over months and years. As in the main chapter, we weight the grid boxes by their area. We find that a five-year training period performs best and would recommend this training length for operational use. We do note however that forecast performance does not appear to be particularly sensitive to this choice.

Table A.4: Mean area-weighted Brier scores for MCF on the test set of 2012-2016 for different numbers of years of training data used to determine the weight on the climatology versus the contour model forecast.

Years	Mean Brier Score
1	0.03313
2	0.03317
3	0.03315
4	0.03293
5	0.03281
6	0.03286
7	0.03293

Appendix B

APPENDIX TO CHAPTER 4

B.1 Proofs

B.1.1 Proof of Theorem 1

Let $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_p)$ with $\theta_i < \theta_{i+1}$ and $\theta_i \in (0, 2\pi)$ for all i. For a star-shaped polygon $\underline{\boldsymbol{S}}$ there exists $\boldsymbol{\theta}$ and $\boldsymbol{y}$ such that $\tilde{\boldsymbol{V}}(\boldsymbol{C}, \boldsymbol{\theta}, \boldsymbol{y}) = \boldsymbol{S}$ for any $\boldsymbol{C} \in \mathcal{K}(\underline{\boldsymbol{S}})$.

Proof. Consider the point $\boldsymbol{s}_i \in \boldsymbol{S}$. Since $\boldsymbol{C} \in \mathcal{K}(\underline{\boldsymbol{S}})$, there exists a line segment, $\overline{\boldsymbol{C}\boldsymbol{s}_i}$, from $\boldsymbol{C}$ to $\boldsymbol{s}_i$ that is entirely contained within the polygon $\underline{\boldsymbol{S}}$. Let $\theta_i = \text{atan2}(\boldsymbol{s}_{i,y} - C_y, \boldsymbol{s}_{i,x} - C_x)$ where atan2(b, a) is the two-quadrant arctangent representing the angle between the positive x-axis and the line segment from the origin to point (b, a). Then the corresponding ℓ_i in the line set covers the line segment $\overline{\boldsymbol{C}\boldsymbol{s}_i}$. By construction, the line ℓ_i intersects $\boldsymbol{s}_i$. The line ℓ_i also cannot intersect any other points on $\overline{\boldsymbol{S}}$, since the existence of such points would violate the assumption that $\boldsymbol{S}$ is star-shaped. Define $y_i = \sqrt{(\boldsymbol{s}_{i,y} - C_y)^2 + (\boldsymbol{s}_{i,x} - C_x)^2}$, then $\boldsymbol{s}_i = \hat{\boldsymbol{v}}_i$. Repeat this construction of θ_i, ℓ_i and y_i for all $\boldsymbol{s}_i$. Then $\boldsymbol{s}_i = \hat{\boldsymbol{v}}_i$ for all i, hence $\tilde{\boldsymbol{V}}(\boldsymbol{C}, \boldsymbol{\theta}, \boldsymbol{y}) = \boldsymbol{S}$. ☐

B.1.2 Proof of Corollary 1

Let ℓ_θ denote the line that extends infinitely outward from $\boldsymbol{C}$ at angle $\theta \in (0, 2\pi)$ and that intersects $\overline{\boldsymbol{S}}$. For any $\theta \in (0, 2\pi)$, the line ℓ_θ is distinct, i.e. $\ell_\theta \neq \ell_{\theta'}$ for any θ, θ' such that $\theta \neq \theta'$.

Proof. Let $\theta_i < \theta < \theta_{i+1}$. The line ℓ_θ intersects $\overline{\boldsymbol{S}}$ in the line segment $\overline{\boldsymbol{s}_i\boldsymbol{s}_{i+1}}$, but not at either $\boldsymbol{s}_i, \boldsymbol{s}_{i+1}$. Since this fact holds for any θ_i, θ_{i+1}, each line ℓ_θ must be distinct. ☐

B.1.3 Proof of Theorem 2

The true starting point $\boldsymbol{C} \in \hat{\boldsymbol{\mathcal{K}}}(\underline{\mathbf{S}})$ where $\underline{\mathbf{S}}$ is a set of N star-shaped polygons.

Proof. The starting point $\boldsymbol{C}$ will be in every $\mathcal{K}(\underline{\boldsymbol{S}}_i)$ by the definition of a kernel. So, any point p that is in $\mathcal{K}(\underline{\boldsymbol{S}}_i)$ but not $\mathcal{K}(\underline{\boldsymbol{S}}_j)$ for any i, j is not $\boldsymbol{C}$. Hence, $\boldsymbol{C}$ must be in $\hat{\boldsymbol{\mathcal{K}}}(\underline{\mathbf{S}})$. $\square$

B.1.4 Proof of Theorem 3

For any polygon $\underline{\boldsymbol{S}}$ that is not star-shaped $|A_u(\boldsymbol{C}, \boldsymbol{\theta}, \boldsymbol{S})| > 0$ and $|A_o(\boldsymbol{C}, \boldsymbol{\theta}, \boldsymbol{S})| > 0$ for any $\boldsymbol{C}$ and $\boldsymbol{\theta}$.

Proof. We show that $|A_u(\boldsymbol{C}, \boldsymbol{\theta}, \boldsymbol{S})| > 0$. The proof for $|A_o(\boldsymbol{C}, \boldsymbol{\theta}, \boldsymbol{S})| > 0$ is analogous. The quantity $|A_u(\boldsymbol{C}, \boldsymbol{\theta}, \boldsymbol{S})| = 0$ only if $(\underline{\boldsymbol{S}}^c \cap \underline{\tilde{\boldsymbol{V}}_u}) = \emptyset$ and $(\underline{\boldsymbol{S}} \cap \underline{\tilde{\boldsymbol{V}}_u}^c) = \emptyset$. These sets are both empty only if $\tilde{\boldsymbol{V}}_u = \boldsymbol{S}$. Hence, we need show only that no $\boldsymbol{\theta}$, $\boldsymbol{C}$, $\boldsymbol{y}$ combination exists that allows $\tilde{\boldsymbol{V}}_u = \boldsymbol{S}$. Polygon $\underline{\boldsymbol{S}}$ is not star-shaped. So, there exists at least one point $s^* \in \overline{\boldsymbol{S}}$ such that for any point $\boldsymbol{C} \in \underline{\boldsymbol{S}}$, the line $\overline{\boldsymbol{C}s^*}$ goes outside polygon $\underline{\boldsymbol{S}}$. Since the line $\overline{\boldsymbol{C}s^*}$ exits $\underline{\boldsymbol{S}}$ before reaching s^*, there is at least one additional intersection point between line $\overline{\boldsymbol{C}s^*}$ and contour line $\overline{\boldsymbol{S}}$. Let s^{**} denote this intersection. The lines $\overline{\boldsymbol{C}s^*}$ and $\overline{\boldsymbol{C}s^{**}}$ are at the same angle, denoted θ^*. For either s^* or s^{**} to be $\tilde{\boldsymbol{V}}_u$, we must put a a line ℓ^* in the line set that extends at angle $\theta^* = \text{atan2}(s^*_y - C_y, s^*_x - C_x) = \text{atan2}(s^{**}_y - C_y, s^{**}_x - C_x)$ where atan2(b, a) is the two-quadrant arctangent representing the angle between the positive x-axis and the line segment from the origin to point (a, b). $\tan^{-1}(b/a) \in (0, 2\pi)$ defines the angle between the x-axis and the line from $(0, 0)$ to (a, b). However, any corresponding selection of y^* that results in s^* in $\tilde{\boldsymbol{V}}_u$ ensures that s^{**} is not in $\tilde{\boldsymbol{V}}_u$, since only one point in $\tilde{\boldsymbol{V}}_u$ is created for each line in the line set. Hence, there is no $\boldsymbol{y}$ that would allow $\tilde{\boldsymbol{V}}_u \neq \boldsymbol{S}$, so $|A_u(\boldsymbol{C}, \boldsymbol{\theta}, \boldsymbol{S})| > 0$. $\square$

B.2 Simulation shape parameters

This section gives a list of the parameter values used in simulations. All values are rounded to the nearest thousandth. For *Shape A*, *Shape B*, and *Shape C*, all the following parameters are shared.

$$\kappa = 2 \tag{B.1}$$

$$\boldsymbol{C} = (0.5, 0.5) \tag{B.2}$$

$$\begin{aligned}\boldsymbol{\theta} = (&0.063, 0.188, 0.314, 0.440, 0.565, 0.691, 0.817, 0.942, 1.068, 1.194, 1.319, 1.445, 1.571, 1.696, 1.822,\\ &1.948, 2.073, 2.199, 2.325, 2.450, 2.576, 2.702, 2.827, 2.953, 3.079, 3.204, 3.330, 3.456, 3.581, 3.707,\\ &3.833, 3.958, 4.084, 4.210, 4.335, 4.461, 4.587, 4.712, 4.838, 4.964, 5.089, 5.215, 5.341, 5.466, 5.592,\\ &5.718, 5.843, 5.969, 6.095, 6.220)\end{aligned} \tag{B.3}$$

$$\begin{aligned}\boldsymbol{\sigma} = (&0.035, 0.044, 0.053, 0.062, 0.071, 0.080, 0.080, 0.073, 0.065, 0.058, 0.050, 0.042, 0.035, 0.035, 0.035,\\ &0.035, 0.035, 0.035, 0.035, 0.035, 0.035, 0.035, 0.035, 0.035, 0.035, 0.035, 0.044, 0.053, 0.062, 0.071,\\ &0.080, 0.080, 0.073, 0.065, 0.058, 0.050, 0.042, 0.035, 0.035, 0.035, 0.035, 0.035, 0.035, 0.035, 0.035,\\ &0.035, 0.035, 0.035, 0.035, 0.035)\end{aligned} \tag{B.4}$$

The mean parameters for *Shape A*, *Shape B*, and *Shape C* are

$$\begin{aligned}\boldsymbol{\mu}_A = (&0.294, 0.304, 0.306, 0.290, 0.264, 0.241, 0.220, 0.213, 0.219, 0.239, 0.259, 0.282, 0.298, 0.300, 0.276,\\ &0.254, 0.236, 0.216, 0.211, 0.217, 0.212, 0.206, 0.200, 0.192, 0.195, 0.239, 0.287, 0.313, 0.318, 0.321,\\ &0.322, 0.321, 0.319, 0.316, 0.312, 0.308, 0.280, 0.240, 0.197, 0.197, 0.214, 0.231, 0.247, 0.264, 0.263,\\ &0.256, 0.249, 0.244, 0.270, 0.283)\end{aligned} \tag{B.5}$$

$$\begin{aligned}\boldsymbol{\mu}_B = (&0.3, 0.3, 0.3, 0.3, 0.3, 0.3, 0.3, 0.3, 0.3, 0.3, 0.3, 0.3, 0.3, 0.3, 0.3, 0.3, 0.3, 0.3, 0.3, 0.3, 0.3, 0.3, 0.3, 0.3,\\ &0.3, 0.3, 0.3, 0.3, 0.3, 0.3, 0.3, 0.3, 0.3, 0.3, 0.3, 0.3, 0.3, 0.3, 0.3, 0.3, 0.3, 0.3, 0.3, 0.3, 0.3, 0.3, 0.3, 0.3,\\ &0.3, 0.3)\end{aligned} \tag{B.6}$$

$$\begin{aligned}\boldsymbol{\mu}_C = (&0.300, 0.263, 0.225, 0.188, 0.150, 0.150, 0.187, 0.225, 0.262, 0.300, 0.290, 0.243, 0.197, 0.150, 0.150,\\ &0.200, 0.250, 0.300, 0.300, 0.274, 0.248, 0.222, 0.196, 0.170, 0.170, 0.200, 0.230, 0.260, 0.290, 0.320,\\ &0.320, 0.298, 0.275, 0.253, 0.230, 0.150, 0.200, 0.250, 0.300, 0.350, 0.360, 0.312, 0.265, 0.218, 0.170,\\ &0.170, 0.203, 0.235, 0.268, 0.300).\end{aligned} \tag{B.7}$$