Bayesian Analysis (2020) **15**, Number 2, pp. 605–631

Bayesian Inference of Spatio-Temporal Changes of Arctic Sea Ice*

Bohai Zhang† and Noel Cressie‡

Abstract. Arctic sea ice extent has drawn increasing interest and alarm from geoscientists, owing to its rapid decline. In this article, we propose a Bayesian spatio-temporal hierarchical statistical model for binary Arctic sea ice data over two decades, where a latent dynamic spatio-temporal Gaussian process is used to model the data-dependence through a logit link function. Our ultimate goal is to perform inference on the dynamic spatial behavior of Arctic sea ice over a period of two decades. Physically motivated covariates are assessed using autologistic diagnostics. Our Bayesian spatio-temporal model shows how parameter uncertainty in such a complex hierarchical model can influence spatio-temporal prediction. The posterior distributions of new summary statistics are proposed to detect the changing patterns of Arctic sea ice over two decades since 1997.

Keywords: binary data, forecasting, hierarchical statistical model, latent Gaussian process, MCMC.

MSC 2010 subject classifications: 62F15, 60G15, 62J12, 62P12.

1 Introduction

Arctic sea ice extent has drawn increasing interest and alarm from geoscientists, owing to its rapid decline, particularly in the Boreal summer and early fall. This is driven by more rapid warming in the Arctic compared with other regions, a phenomenon known as the Arctic amplification (Cohen et al., 2014). The declining sea ice directly affects the biogeochemical cycle and animals in the Arctic region such as the polar bear and seabirds (Meier et al., 2014). Furthermore, the changes of Arctic sea ice can lead to changing climates in other regions of the world. For example, recent studies show that the decline of Arctic sea ice can cause extreme weather in mid-latitude regions (e.g., increasing the chance of cold Eurasian winters: Screen and Simmonds, 2013; Cohen et al., 2014; Mori et al., 2014) and influence rainfall in the state of California (e.g., Cvijanovic et al., 2017). Critically, the declining sea ice will result in an albedo–ice feedback effect (e.g., Kumar et al., 2010; Screen et al., 2013; Pistone et al., 2014), where a darker Earth surface due to less sea ice leads to additional melting.

Therefore, monitoring the spatio-temporal dynamics of Arctic sea ice is a critical component of the study of climate change. There could also be commercial interest:

*The acknowledgements section was corrected on 26 May 2020 by adding more details to Bohai Zhang's research support.

†School of Statistics and Data Science, KLMDASR, and LPMC, Nankai University, Tianjin, 300071, China, bohaizhang@nankai.edu.cn

‡National Institute for Applied Statistics Research Australia, University of Wollongong, NSW, 2522, Australia, ncressie@uow.edu.au

 https://doi.org/10.1214/20-BA1209

The loss of summer sea ice may open new shipping lanes, providing a new passage between Earth's major oceans.

In this article, we consider Arctic sea-ice-extent data obtained from a database of Arctic sea-ice concentrations (proportion of sea ice in a grid cell) produced by the National Oceanic and Atmospheric Administration (NOAA) as part of their National Snow & Ice Data Center's (NSIDC) Climate Data Record (CDR). The data are based on passive microwave remotely sensed data provided by the Nimbus 7 satellite and the F8, F11, and F13 satellites of the Defense Meteorological Satellite Program (Parkinson et al., 1999; Parkinson, 2014a), and they are projected onto approximately $25\text{km} \times 25\text{km}$ grid cells (or pixels). A 15% cut-off is the standard used to determine whether a pixel is water ($< 15\%$) or ice ($\geq 15\%$), and Arctic sea ice extent (SIE) is defined as the total area of ice pixels in the Arctic region. (e.g., Parkinson et al., 1999; Zwally et al., 2002; Meier et al., 2007; Parkinson, 2014a).

The resulting binary (0 = water and 1 = ice) data are available monthly from November, 1978; here we focus on the month of September for the two decades 1997–2016, since the Arctic sea-ice cover is typically least in that month of the year (Parkinson, 2014a). Based on the September data from 1997–2016, Figure 1 shows a time series plot of the SIE in millions of km^2 for the Arctic region.

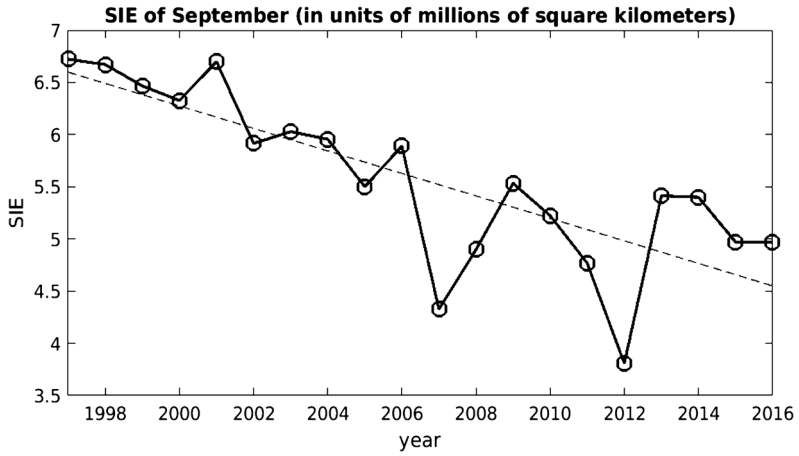


Figure 1: September Arctic sea ice extent (SIE, in units of millions of km^2) from 1997 to 2016. Arctic SIE is the total area of ice pixels in the Arctic region. The dashed line shows an ordinary-least-squares fit.

Previous analyses of Arctic sea ice by geoscientists have focussed mainly on the temporal aspect or the spatial aspect, without considering both simultaneously (e.g., Parkinson, 2014a,b; Parkinson and DiGirolamo, 2016). Notably, uncertainty measures for the summary statistics presented in their studies are either absent or not appropriate, given the dependence in the data. Related research on calibrating spatial binary outputs from computer models of Antarctic ice sheets can be found in Chang et al. (2016a,b). More recently, Olson et al. (2019) proposed a new method for assessing the dependence

of Arctic sea ice on climate variables (in the form of output from climate models), for the purpose of forecasting the Arctic sea ice. In contrast to these papers, the article by Director et al. (2017, 2019) takes a decidedly statistical approach, using a spatio-temporal statistical model to forecast selected contours of sea-ice concentration.

In this article, we use a Bayesian statistical approach for inference on Arctic sea ice based on interpretable spatial summaries for the months of September spanning two decades from 1997–2016. Specifically, we use a spatio-temporal hierarchical generalized linear mixed model (ST-GLMM) framework, where a latent Gaussian process (GP) is introduced to model the (spatio-temporal) dependence in the data, linked to the non-Gaussian (binary) observations through a logit link. Recall that the sea-ice-extent data are defined on spatial pixels of a nominal area of 25km $\times$ 25km, and hence these pixels can be treated as Basic Areal Units (BAUs) for modeling the spatial data (e.g., Nguyen et al., 2012; Zammit-Mangion and Cressie, 2020), and their centroids are used as the spatial locations of the BAU-level data. We carry out Bayesian inference on areal-based quantities for the observation period, in contrast to Director et al. (2017, 2019), who forecast selected sea-ice-concentration contours beyond their observation periods.

Our statistical model is motivated by Diggle et al. (1998), who analyzed spatial-only binary data. When the data set is very large, which is often the case for spatio-temporal geophysical data, inference on the latent GP in a ST-GLMM is very expensive and quickly becomes prohibitive. Then the modeling strategy of using low-rank or sparse-matrix approximations can be applied to facilitate the inference. This has proved to be very effective for Gaussian spatial data (e.g., Furrer et al., 2006; Banerjee et al., 2008; Cressie and Johannesson, 2008; Rue et al., 2009; Sang and Huang, 2012; Nychka et al., 2015; Datta et al., 2016; Katzfuss, 2017). The approach has been adapted to analyze large non-Gaussian spatial and spatio-temporal data by *inter alia* Sengupta and Cressie (2013), Sengupta et al. (2016), Holan and Wikle (2016), Bradley et al. (2016), Guan and Haran (2018), Bradley et al. (2018), and Linero and Bradley (2018).

In this article, we use physical knowledge supported by data to build a scientifically motivated latent process that is Gaussian on the logistic scale. Then we put prior distributions on the ST-GLMM model parameters and use a Bayesian hierarchical model (BHM) to obtain the joint posterior distribution of parameters and the latent GP given the binary SIE data. In an earlier paper, an empirical hierarchical modeling (EHM) approach was taken by Zhang and Cressie (2019) to analyze the binary SIE data with a single covariate given by distance to the North Pole.

Here, the covariates that are used to model the Arctic SIE data include the averaged Arctic surface temperature anomalies in the previous summer season and in the previous winter season, in order to forecast the presence/absence of sea ice in the following September. The effects of longitude and distance to the coastline are also found to be important covariates. While generic hierarchical modeling uses Bayes' Theorem, the model becomes fully Bayesian when all parameters have prior distributions put on them. In contrast to this BHM approach, an EHM approach involves "plugging in" the parameter estimates into the hierarchical model.

In this article, our BHM is in fact a hybrid (e.g., Wikle et al., 2019, Ch. 1), where priors were put on all but one parameter. Our simulations in Section S1 of the Sup-

plementary Material (Zhang and Cressie, 2020) demonstrated an inherent difficulty of carrying out Bayesian inference for this (fine-scale) variance parameter, so we substituted its expectation-maximization (EM, Dempster et al., 1977) estimate into the hierarchical model. We carried out a sensitivity study on inference from the proposed hierarchical statistical model under different plug-in values of σ^2_ξ, and we found that the inference was not very sensitive to its misspecification. Our simulations in Section S1 also allowed us to do some comparisons between the EHM and our (hybrid) BHM, and we found that BHM is preferred over EHM when making predictions in gaps of the spatial domain where no nearby observations are available.

It is worth emphasizing that the sea-ice-extent data are obtained from the sea-ice concentrations, which are observed areal proportions that are noisy (especially for lower concentration percentages). Here, we have proposed a Bayesian spatio-temporal model for the binary SIE data, from which the predictive distribution of the underlying process on the probability scale is less uncertain than the observed proportions. The statistical dependencies due to spatial proximity and through time are modeled explicitly, and Bayesian inferences account for them naturally and coherently. Simple summaries based on an empirical approach cannot capture the spatio-temporal variations simultaneously with proper uncertainty quantification, due to statistical dependencies in space and time.

A new aspect of our article is the choice of various functionals of the latent GP. These summary statistics feature the process dynamics and, through their predictive distributions, they illustrate the changing patterns of Arctic sea ice between the earlier decade and the recent decade. For example, we consider boxplots of the posterior means of the latent GP on the original scale (and on the probability scale), empirical temporal semivariograms, Arctic-region Hovmöller diagrams, and spatial animations, which offer different "views" of the changing spatio-temporal patterns of Arctic sea ice. Two functionals that lend themselves to animation are the ice-to-water transition probabilities and the water-to-ice transition probabilities (Zhang and Cressie, 2019), which are given in Section S5 of the Supplementary Material.

Figure 2 shows boxplots of the posterior means of the latent probability of ice, denoted as $\{p_t(\mathbf{s})\}$, where the boxplots are taken at times $t = 2001$, 2006, 2011, and 2016, for $\mathbf{s}$ ranging over the latitude bands centered at $\text{lat}_0 = 75°\text{N}$, 80°N, and 85°N. The effects of climate change in the recent decade is seen in different ways, at different latitudes: For $\text{lat}_0 = 75°\text{N}$, the boxplot averages (denoted by a dot) decrease over time, dropping well below 0.5 (the cut-off where there are more water pixels than ice pixels) during the recent decade. The high variability in the earlier decade, as captured by the size of the boxes, indicates a process in transition before collapse of the sea ice in the recent decade at that latitude. At 80°N, there is also a difference of the average posterior-mean $p_t(\mathbf{s})$-values between the earlier decade and the recent decade, and noticeably the variabilities are strikingly different. The high variability in the recent decade may precede a collapse as it did at 75°N. For the very high latitude of 85°N, all the average posterior-mean $p_t(\mathbf{s})$-values are very close to 1 with very small variability, due to still-dominant ice cover at very high latitudes. Other useful functionals based on the predictive distribution of $\{p_t(\mathbf{s})\}$ are given in Section 4, including temporal semivariograms that are able to detect a periodicity in the latent process at 85°N.

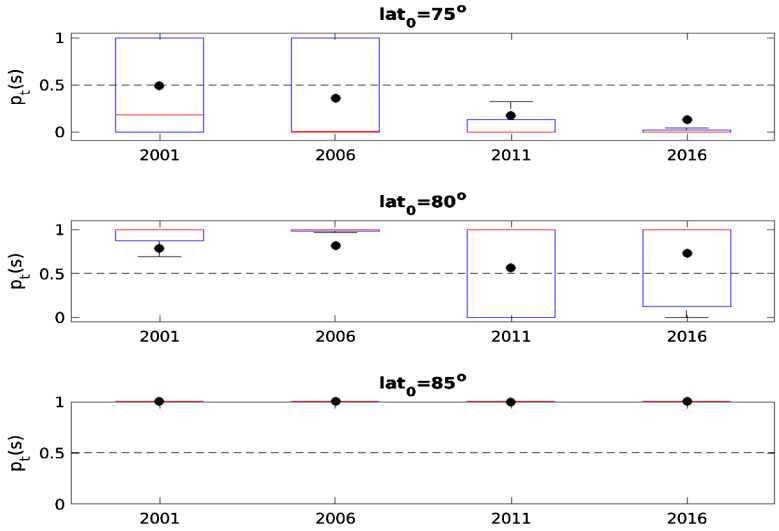


Figure 2: Boxplots of $\{E(p_t(\mathbf{s})|\text{data}) : \mathbf{s} = (\text{lon}, \text{lat}) \text{ and } \text{lat} \in (\text{lat}_0 - \Delta, \text{lat}_0 + \Delta)\}$ for three latitude bands and for years $t = 2001, 2006, 2011, 2016$, where the width of the latitude band is $2\Delta = 1°$. The dashed line indicates the value of 0.5.

The rest of the article is organized as follows. In Section 2, we introduce a spatio-temporal hierarchical model for binary data, which is used for analyzing the Arctic sea-ice-extent (SIE) data. This consists of a data model, a process model that is dynamic, and a parameter model (or prior). An important part of the process model that we build is the large-scale variation captured by covariates in the process model; Section 2 also contains a detailed description of how the covariates were selected. In Section 3, we discuss Bayesian inference of model parameters and the prediction/forecasting procedure of the BHM. A Bayesian analysis of 20 years of binary Arctic SIE data is given in Section 4, where several summary statistics and their predictive distributions are used to detect the changing patterns of Arctic sea ice from 1997–2016. In Section 5, a summary and discussion of our results are given. The article also contains five sections of Supplementary Material, to support our work presented in the main paper.

2 Spatio-Temporal Hierarchical Modeling

In this section, we introduce a spatio-temporal hierarchical statistical model for the binary Arctic SIE data introduced in Section 1. Let $z_t(\mathbf{s}) \in \{0, 1\}$ be a binary datum observed at a spatial location $\mathbf{s} \in \mathcal{D}_t$ and a time point $t \in \mathcal{T}$, where $\mathcal{D}_t$ is the spatial domain of interest and t is a discrete time point in $\mathcal{T} \equiv \{1, 2, \ldots, T\}$. We assume there is a latent process $\{y_t(\mathbf{s})\}$ that is used to define the probability that $z_t(\mathbf{s})$ equals 0 or 1, as follows. Following Diggle et al. (1998), we model $\{z_t(\mathbf{s})\}$ as conditionally independent Bernoulli random variables given the latent process $\{y_t(\mathbf{s})\}$. That is, we model

$$z_t(\mathbf{s})|y_t(\mathbf{s}) \sim \text{Ber}(p_t(\mathbf{s})), \tag{2.1}$$

independently for all $\mathbf{s} \in \mathcal{D}_t$ and $t \in \mathcal{T}$, where $\text{Ber}(p)$ is a binary random variable on $\{0, 1\}$ and equals 1 with probability p. In (2.1), $y_t(\mathbf{s}) = g(p_t(\mathbf{s}))$ and $g(\cdot)$ is a link function; here we choose the logit link such that $y_t(\mathbf{s}) = g(p_t(\mathbf{s})) = \log(p_t(\mathbf{s})/(1-p_t(\mathbf{s})))$. Equation (2.1) above defines the so-called "data model."

The "process model" is underneath the data model in the hierarchy of conditional models. Here, the model for $\{y_t(\mathbf{s})\}$, conditional on the parameters, is a spatio-temporal linear mixed model with a vector autoregressive model of order 1 (VAR(1)) for the coefficients of a relatively small set of basis functions (e.g., Wikle et al., 2001; Cressie et al., 2010; Katzfuss and Cressie, 2011; Bradley et al., 2015):

$$\begin{aligned} y_t(\mathbf{s}) &= \mathbf{x}_t(\mathbf{s})'\boldsymbol{\beta} + \mathbf{S}_t(\mathbf{s})'\boldsymbol{\eta}_t + \xi_t(\mathbf{s}), \\ \boldsymbol{\eta}_t &= \mathbf{H}_t\boldsymbol{\eta}_{t-1} + \boldsymbol{\zeta}_t, \\ \boldsymbol{\zeta}_t &\sim \text{Gau}(\mathbf{0}, \mathbf{U}_t) \text{ and } \boldsymbol{\eta}_1 \sim \text{Gau}(\mathbf{0}, \mathbf{K}), \end{aligned} \tag{2.2}$$

where at time $t \in \mathcal{T}$ and $\mathbf{s} \in \mathcal{D}_t$, $\mathbf{x}_t(\mathbf{s})$ is a p-dimensional vector of covariates at spatial location $\mathbf{s}$; $\boldsymbol{\beta}$ is the p-dimensional vector of regression coefficients associated with $\mathbf{x}_t(\mathbf{s})$; $\mathbf{S}_t(\mathbf{s})$ is an r-dimensional basis-function vector evaluated at $\mathbf{s}$; $\boldsymbol{\eta}_t$ is a vector of random coefficients of $\mathbf{S}_t(\mathbf{s})$; $\xi_t(\mathbf{s})$ is a random variable that models the fine-scale variation at $\mathbf{s}$ not captured by $\mathbf{S}_t(\mathbf{s})'\boldsymbol{\eta}_t$; and $\text{Gau}(\mu, \sigma^2)$ denotes a Gaussian distribution with mean μ and variance σ^2. The parameters in (2.2), upon which the process model is conditioned, are: $\boldsymbol{\beta}$, the p-dimensional vector of regression coefficients; and $\{\mathbf{H}_t : t = 2, \ldots, T\}$ and $\{\mathbf{U}_t : t = 2, \ldots, T\}$, the $r \times r$ propagator and $r \times r$ innovation matrices, respectively. We further assume that $\xi_t(\mathbf{s})$ in (2.2) follows $\text{Gau}(0, \sigma^2_\xi)$ distribution, $\{\xi_t(\mathbf{s})\}$ are independent of each other over both space and time, and they are also independent of the random vectors $\{\boldsymbol{\eta}_t : t = 1, \ldots, T\}$. Hence, spatio-temporal variability in the data is captured through the fixed-effects term, $\mathbf{x}_t(\mathbf{s})'\boldsymbol{\beta}$ (large-scale variation), and the random-effects term, $\mathbf{S}_t(\mathbf{s})'\boldsymbol{\eta}_t$ (small-scale variation), and (2.2) represents a spatio-temporal GP.

For a fixed time point t, the number of observations can be very large; that is, $\{y_t(\mathbf{s}) : \mathbf{s} \in \mathcal{D}_t\}$ forms a high-dimensional vector when evaluated at all the observation locations. By fixing the number of basis functions r to be a relatively small number (say a few hundred), the latent process, $\{y_t(\mathbf{s}) - \mathbf{x}_t(\mathbf{s})'\boldsymbol{\beta}\}$, is represented by a low-dimensional basis-function vector, allowing fast computations to be achieved (e.g., Wikle et al., 2001; Cressie and Johannesson, 2006, 2008; Kang et al., 2010; Zammit-Mangion and Cressie, 2020). There are many types of basis functions that could be used in this setting, such as wavelets, splines, Wendland functions, and bisquare functions. In this article, we focus on the compactly supported bisquare functions, since they have been successfully used to model very large Gaussian and non-Gaussian spatial and spatio-temporal data (e.g., Cressie and Johannesson, 2008; Sengupta and Cressie, 2013; Zhang and Cressie, 2019). In addition, specifying multi-resolution basis functions has proven to be effective in capturing spatial dependence at different scales (e.g., Wikle et al., 2001; Katzfuss and Cressie, 2011; Nychka et al., 2015; Katzfuss, 2017). Hence, we adopt a multi-resolution class.

For $j = 1, \ldots, r_i$ basis functions of the i-th resolution, we define the bisquare basis function in d-dimensional Euclidean space $\mathbb{R}^d$ as,

$$S_j^{(i)}(\mathbf{s}) \equiv \left(1 - \left(\frac{\|\mathbf{s} - \mathbf{c}_j^{(i)}\|}{\phi_i}\right)^2\right)^2 I(\|\mathbf{s} - \mathbf{c}_j^{(i)}\| < \phi_i);\ \mathbf{s} \in \mathbb{R}^d, \tag{2.3}$$

where r_i is the number of basis functions at the i-th resolution, $\mathbf{c}_j^{(i)}$ is the center of the j-th basis function $S_j^{(i)}(\cdot)$ at the i-th resolution, $\|\cdot\|$ is the Euclidean norm, ϕ_i is the radius of its spatial support (sometimes called the aperture), and $I(\cdot)$ is an indicator function. In practice, ϕ_i is specified to be 1.5 times the minimum distance between basis-function centers of the same resolution (e.g., Cressie and Johannesson, 2008). Note that in Section 4, we replace the Euclidean norm with the great-circle distance in (2.3), and it is easy to see that this type of modification could be made on any manifold equipped with a norm.

The propagator matrix $\mathbf{H}_t$ captures the temporal covariances between the elements of $\boldsymbol{\eta}_t$ and $\boldsymbol{\eta}_{t-1}$, for $t = 2, \ldots, T$. Here we treat $\mathbf{H}_t$ and $\mathbf{U}_t$ as unknown parameters, but we assume that they are constant for a fixed time period; that is, $\mathbf{H}_t \equiv \mathbf{H}$ and $\mathbf{U}_t \equiv \mathbf{U}$, for $t = 2, \ldots, T$. In practice, they will be allowed to vary from time-period to time-period (e.g., Katzfuss and Cressie, 2011; Zhang and Cressie, 2019). When modeling the Arctic SIE data, we allow the two matrices to change in successive five-year periods over the two decades of data.

Although the propagator matrix $\mathbf{H}_t$ may be considered as an $r \times r$ parameter matrix, a parsimonious representation for it can improve inference and allow for physical interpretations of its entries (Wikle et al., 2001). Alternatively, Bradley et al. (2015) proposed a class they called the Moran's I class of propagator matrices in order to avoid confounding between $\{\boldsymbol{\eta}_t\}$ and the covariates. Here we specify $\mathbf{H}_t$ for the multi-resolution basis functions by assigning a different propagator parameter for each resolution. For example, for a two-resolution design with r_1 Resolution-1 (coarse resolution) basis functions and r_2 Resolution-2 (fine resolution) basis functions, Zhang and Cressie (2019) parameterized $\mathbf{H}_t$ as follows: For $t = 2, \ldots, T$,

$$\mathbf{H}_t \equiv \mathbf{H} \equiv \begin{pmatrix} \lambda_1 \mathbf{I}_{r_1} & \mathbf{0} \\ \lambda_3 \mathbf{R} & \lambda_2 \mathbf{I}_{r_2} \end{pmatrix}, \tag{2.4}$$

where $\lambda_1, \lambda_2, \lambda_3 \in (-1, 1)$, and $\mathbf{R}$ is an $r_2 \times r_1$ matrix encoding the possible dependence from the coarse Resolution 1 to the fine Resolution 2. Specifically, for $i = 1, \ldots, r_2$ and $j = 1, \ldots, r_1$, $\mathbf{R}(i, j) = 1$ if the i-th Resolution-2 basis function is a neighbor of the j-th Resolution-1 basis function, and $\mathbf{R}(i, j) = 0$ otherwise. If the model is used for forecasting, we would need to check the eigenvalues of $\mathbf{H}$ for possible "explosive" behavior.

The parameters for the data model given by (2.1) and the parameters for the process model given by (2.2) are $\boldsymbol{\theta} \equiv \{\boldsymbol{\beta}, \sigma_\xi^2, \mathbf{K}, \mathbf{H}, \mathbf{U}\}$. For Bayesian inference, we need to assign a prior to $\boldsymbol{\theta}$: For variance-covariance parameters $\mathbf{K}$, $\mathbf{U}$, and σ_ξ^2, we used Inverse-Wishart

distributions for conditional conjugacy; for the regression coefficients $\boldsymbol{\beta}$, we specified an improper non-informative prior $\pi(\boldsymbol{\beta}) \propto 1$; and for each of the $\{\lambda_i\}$ of $\mathbf{H}$ given by (2.4), we specified independent Unif$(-1, 1)$ distributions. In the simulation setting given in Section S1 of the Supplementary Material, we found that a single parameter, σ_ξ^2, that is better handled through EM estimation than through assigning it a prior, and hence we do likewise for analyzing the Arctic SIE data.

2.1 Specification of Covariates for the Arctic Sea-Ice-Extent Data

In this subsection, we discuss how to select covariates for modeling the spatio-temporal Arctic SIE data in the month of September for each year spanning the two decades from 1997–2016. September is chosen as a time when Arctic sea ice is least, and it avoids modeling the within-year seasonal variation (Parkinson et al., 1999); consequently, we are not able to detect changes in seasonal variation over the years. Henceforth $\mathcal{D}_t \equiv \mathcal{D}$, a region defined by pixels whose latitudes are greater than or equal to 60°N that we call the Arctic region.

Recall that in Zhang and Cressie (2019), the only covariate used was the distance of any spatial location in $\mathcal{D}$ to the North Pole, which is a proxy for surface temperatures in the Arctic region; its coefficient was allowed to vary with $t \in \mathcal{T}$ in order to detect climate change. In this article, we consider the Arctic surface temperatures directly and use the GISTEMP surface-temperature-anomaly data to obtain covariates. Specifically, we use the previous year's summer-average and winter-average surface-temperature anomalies to define two covariates that allow for temporal variation as well as spatial variation. In addition, for some spatial pixels that are close to coastlines but far from the North Pole, we observe the presence of sea ice (e.g., see the left panel of Figure 3). Hence, we use the distance of spatial location to the nearest coastline to define a covariate that models local spatial effects. Finally, an exploratory data analysis based on a simple logistic-regression model showed that longitude played a role in modeling these data, probably due to the different dynamics in different regions of the Arctic (Section S2 of the Supplementary Material).

Based on preliminary data analyses that included use of standard logistic regressions, we fitted the following mean function for the hidden spatio-temporal GP:

$$\begin{aligned}\mathbf{x}_t(\mathbf{s})'\boldsymbol{\beta} \;=\; & \beta_0 + \bar{x}^{su}_{t-1}\beta_1 + \bar{x}^{wi}_{t-1}\beta_2 + (x^{su}_{t-1}(\mathbf{s}) - \bar{x}^{su}_{t-1})\beta_3 + (x^{wi}_{t-1}(\mathbf{s}) - \bar{x}^{wi}_{t-1})\beta_4 \\ & + \cos(\pi s_1/180)\beta_5 + \sin(\pi s_1/180)\beta_6 + x^{pl}(\mathbf{s})\beta_7 + x^{cs}(\mathbf{s})\beta_8, \quad (2.5)\end{aligned}$$

where $\mathbf{s} \equiv (s_1, s_2)' \in \mathcal{D}$, and s_1 and s_2 are the longitude and latitude of $\mathbf{s}$; $\bar{x}^{su}_{t-1}$ and $\bar{x}^{wi}_{t-1}$ are the spatially averaged (i.e., over $\mathcal{D}$) Boreal summer (averaged over Jun, Jul, Aug) and winter (averaged over Dec, Jan, Feb) surface-temperature anomalies each indexed by the year $(t-1)$ in which the first month falls. The components of the spatial averages $x^{su}_{t-1}(\mathbf{s})$ and $x^{wi}_{t-1}(\mathbf{s})$ at the spatial location $\mathbf{s}$ and the previous year $(t-1)$ are used to capture the regional effects of changes of surface temperatures. That is, summer (winter) in year $(t-1)$ corresponds to the three-month season starting in Jun (Dec) in year $(t-1)$. These first four covariates aim to characterize the effects of the previous

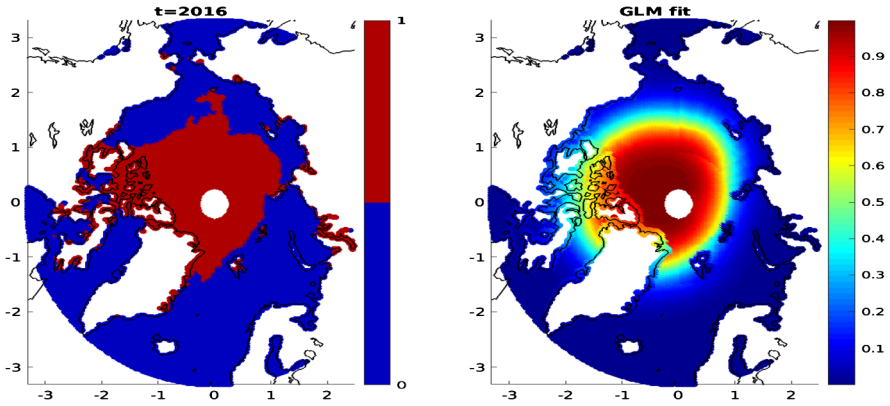


Figure 3: September Arctic SIE data at $t = 2016$ (binary, left panel) and the GLM-fitted logistic regression fitted to the 2016 data (using the covariates in (2.5)) on the probability scale (right panel). The white circle around the North Pole in both panels represents a region not covered by the sensor due to remote sensing limitations.

year's surface-temperature anomalies on forming the sea ice in September in the current year.

The remaining four covariates are purely spatial: $x^{pl}(\mathbf{s})$ is the distance between $\mathbf{s}$ and the North Pole and was the only covariate included in the hidden spatio-temporal GP of the EHM fitted by Zhang and Cressie (2019); the covariates based on longitude represent periodic spatial heterogeneity, analogous to modeling a linear trend on the real line; and $x^{cs}(\mathbf{s})$ is a more local spatial effect accounting for short distances to the coast (less than 50km). Let $d^{cs}(\mathbf{s})$ denote the great-circle distance from $\mathbf{s}$ to the nearest coastline. Our exploratory data analysis based on logistic regression showed that the coefficients of $x^{pl}(\mathbf{s})$ are larger for data close to the coastline than for data far away from the coastline. Then we define $x^{cs}(\mathbf{s}) \equiv x^{pl}(\mathbf{s}) \times I(d^{cs}(\mathbf{s}) < 50\text{km})$ to provide adjustments for the effect of distance to the coast. The right panel of Figure 3 shows the logistic-regression fit of (2.5) for the Arctic sea-ice-extent data for $t = 2016$.

In summary, we shall use the eight covariates in (2.5) to model the GP's fixed effects in Section 4, and the remaining spatio-temporal variability in the GP is modeled statistically with random effects given in (2.2).

3 Fully Bayesian Inference

In this section, we discuss the Bayesian inference of model parameters for the hierarchical statistical model given by (2.1) and (2.2). Let $\mathcal{D}_t^o \equiv \{\mathbf{s}_{t,1}^o, \mathbf{s}_{t,2}^o, \ldots, \mathbf{s}_{t,N_t}^o\}$ be the spatial locations with observed data at time $t \in \{1, \ldots, T\}$. For each time t, we stack the data into a column vector, $\mathbf{Z}_t^o \equiv (z_t(\mathbf{s}_{t,1}^o), \ldots, z_t(\mathbf{s}_{t,N_t}^o))'$, and we define $\boldsymbol{\xi}_t^o \equiv (\xi_t(\mathbf{s}_{t,1}^o), \ldots, \xi_t(\mathbf{s}_{t,N_t}^o))'$ to be the fine-scale-variation vector evaluated at $\mathcal{D}_t^o$. Let $\mathbf{Z}^o \equiv (\mathbf{Z}_1^{o'}, \ldots, \mathbf{Z}_T^{o'})'$, $\boldsymbol{\xi}^o \equiv (\boldsymbol{\xi}_1^{o'}, \ldots, \boldsymbol{\xi}_T^{o'})'$, and $\boldsymbol{\eta} \equiv (\boldsymbol{\eta}_1', \ldots, \boldsymbol{\eta}_T')'$; then the data likelihood

is:

$$
\begin{aligned}
L(\boldsymbol{\theta}; \mathbf{Z}^o) &= \int_{\boldsymbol{\eta}} \int_{\boldsymbol{\xi}^o} p(\mathbf{Z}^o|\boldsymbol{\eta}, \boldsymbol{\xi}^o, \boldsymbol{\beta}) \times p(\boldsymbol{\eta}|\mathbf{K}, \mathbf{H}, \mathbf{U}) \times p(\boldsymbol{\xi}^o|\sigma_\xi^2) d\boldsymbol{\xi}^o d\boldsymbol{\eta} \\
&= \int_{\boldsymbol{\eta}} \int_{\boldsymbol{\xi}^o} \prod_{t=1}^{T} \prod_{i=1}^{N_t} (1 + \exp(-(2z_{t,i}^o - 1) y_{t,i}^o))^{-1} \\
&\quad \times (2\pi)^{-r/2} |\mathbf{K}|^{-1/2} \exp(-\boldsymbol{\eta_1}' \mathbf{K}^{-1} \boldsymbol{\eta}_1 / 2) \\
&\quad \times (2\pi)^{-(T-1)r/2} |\mathbf{U}|^{-(T-1)/2} \prod_{t=2}^{T} \exp(-(\boldsymbol{\eta}_t - \mathbf{H}\boldsymbol{\eta}_{t-1})' \mathbf{U}^{-1} (\boldsymbol{\eta}_t - \mathbf{H}\boldsymbol{\eta}_{t-1}) / 2) \\
&\quad \times \prod_{t=1}^{T} (2\pi\sigma_\xi^2)^{-N_t/2} \exp(-\boldsymbol{\xi}_t^{o'} \boldsymbol{\xi}_t^o / (2\sigma_\xi^2)) \mathrm{d}\boldsymbol{\xi}^o \mathrm{d}\boldsymbol{\eta},
\end{aligned}
$$

where $p(\cdot|\cdot)$ represents the appropriate conditional-probability density; under this notation, $p(\mathbf{Z}^o|\boldsymbol{\eta}, \boldsymbol{\xi}^o, \boldsymbol{\beta})$ is a product of Bernoulli densities, $p(\boldsymbol{\eta}|\mathbf{K}, \mathbf{H}, \mathbf{U})$ is a multivariate Gaussian density, and $p(\boldsymbol{\xi}^o|\sigma_\xi^2)$ is a product of univariate Gaussian densities. For notational simplicity, we define $z_{t,i}^o \equiv z_t(\mathbf{s}_{t,i}^o)$, and $y_{t,i}^o \equiv y_t(\mathbf{s}_{t,i}^o)$. Since the integration above does not have a closed form due to the nonlinearity of the data model, Zhang and Cressie (2019) modified the spatial-only methodology given in Sengupta and Cressie (2013) and developed an EM algorithm to estimate model parameters, $\boldsymbol{\theta}$, for spatio-temporal data $\mathbf{Z}^o$. Then they substituted the resulting estimate $\hat{\boldsymbol{\theta}}$ into the hierarchical model, resulting in an EHM. In this article, our approach is Bayesian, and (with the exception of one parameter, σ_ξ^2) we put prior distributions on the elements of $\boldsymbol{\theta}$, resulting in a BHM and a joint posterior distribution of $\boldsymbol{\theta}$, $\boldsymbol{\eta}$, and $\boldsymbol{\xi}^o$.

In real-world applications, some of the parameters may be fixed by the modeler at physically meaningful values or at estimates, and the remaining parameters could have priors put on them. This results in what might be called a hybrid hierarchical model that is not an EHM but not quite a BHM either (e.g., Wikle et al., 2019, Ch.1). When analyzing the Arctic SIE data, σ_ξ^2 was very difficult to infer from a fully Bayesian analysis. This may be because it is hard to identify separately the small-scale variation and the fine-scale variation from the total variation observed on binary data, particularly when $\mathbf{K}$ is an arbitrary $r \times r$ positive-definite matrix. However, we found the situation unimproved by making parametric assumptions about $\mathbf{K}$ and, more fundamentally, this may be because not all the covariance parameters are identifiable under fixed-domain asymptotics (Zhang, 2004).

In Section 4, we plug in the EM estimate of σ_ξ^2, resulting in a hybrid BHM, although we continue to call the model a BHM. However, in the rest of Section 3, we present a fully Bayesian methodology that includes a prior on σ_ξ^2, noting that the algorithms simplify a little when we fix σ_ξ^2.

3.1 Parameter Inference

Let $\pi(\cdot)$ denote a generic prior distribution. The joint posterior distribution, $p(\boldsymbol{\eta}, \boldsymbol{\xi}^o, \boldsymbol{\theta}|\mathbf{Z}^o)$, is proportional to the product of the complete likelihood, $p(\mathbf{Z}^o, \boldsymbol{\eta}, \boldsymbol{\xi}^o|\boldsymbol{\theta})$, and the

prior distribution, $\pi(\mathbf{K}, \mathbf{H}, \mathbf{U}, \boldsymbol{\beta}, \sigma_\xi^2)$. For $\mathbf{H}$ parameterized as in (2.4) and $\text{IW}(\nu, \boldsymbol{\Phi})$ denoting an Inverse Wishart distribution with parameters ν and $\boldsymbol{\Phi}$, we put $\pi(\mathbf{K}) \sim \text{IW}(\nu_K, \boldsymbol{\Phi}_K)$, $\pi(\mathbf{U}) \sim \text{IW}(\nu_U, \boldsymbol{\Phi}_U)$, $\pi(\lambda_i) \sim \text{Unif}(-1, 1)$ independently for $i = 1, 2, 3$, and $\pi(\boldsymbol{\beta}, \sigma_\xi^2) \propto (\sigma_\xi^2)^{-1/2}$. Then we can integrate out $\mathbf{K}$ and $\mathbf{U}$ and sample $\{\boldsymbol{\eta}, \boldsymbol{\xi}^o, \boldsymbol{\beta}, \{\lambda_i\}, \sigma_\xi^2\}$ from

$$p(\boldsymbol{\eta}, \boldsymbol{\xi}^o, \boldsymbol{\beta}, \{\lambda_i\}, \sigma_\xi^2 | \mathbf{Z}^o) \propto p(\mathbf{Z}^o | \boldsymbol{\beta}, \boldsymbol{\eta}, \boldsymbol{\xi}^o) \times |\boldsymbol{\eta}_1 \boldsymbol{\eta}_1' + \boldsymbol{\Phi}_{\boldsymbol{K}}|^{-\frac{\nu_K + 1}{2}} \times (\sigma_\xi^2)^{-(N+1)/2} \times e^{-\frac{\boldsymbol{\xi}^{o\prime} \boldsymbol{\xi}^o}{2\sigma_\xi^2}} \times |\sum_{t=2}^{T} (\boldsymbol{\eta}_t - \mathbf{H}\boldsymbol{\eta}_{t-1})(\boldsymbol{\eta}_t - \mathbf{H}\boldsymbol{\eta}_{t-1})' + \boldsymbol{\Phi}_{\boldsymbol{U}}|^{-\frac{\nu_U + T - 1}{2}}, \tag{3.1}$$

where $N = \sum_{t=1}^{T} N_t$. Note that we do not integrate out σ_ξ^2, because we would like to sample each $\xi_{t,i}^o$ individually such that the computations are parallelizable (and we note that ultimately we fix σ_ξ^2 at its EM estimate in Section 4). The posterior samples of $\mathbf{K}$ and $\mathbf{U}$ can be sampled from the following full conditionals:

$$p(\mathbf{K}|\cdot) \sim \text{IW}(\nu_K + 1, \boldsymbol{\eta}_1 \boldsymbol{\eta}_1' + \boldsymbol{\Phi}_{\boldsymbol{K}}), \tag{3.2}$$

$$p(\mathbf{U}|\cdot) \sim \text{IW}\left(\nu_U + T - 1, \sum_{t=2}^{T} (\boldsymbol{\eta}_t - \mathbf{H}\boldsymbol{\eta}_{t-1})(\boldsymbol{\eta}_t - \mathbf{H}\boldsymbol{\eta}_{t-1})' + \boldsymbol{\Phi}_{\boldsymbol{U}}\right), \tag{3.3}$$

where "$\cdot$" denotes conditioning on all other parameters (and of course on the data). In order to make $\{\boldsymbol{\eta}_t : t = 1, \ldots, T\}$ well constrained in the prior when modeling the Arctic SIE data, we chose the Inverse Wishart parameter $\nu_K = 2r$. This makes the prior distribution more informative; then $\boldsymbol{\Phi}_K = (3r + 1)\hat{\mathbf{K}}$ and $\boldsymbol{\Phi}_U = (3r + 1)\hat{\mathbf{U}}$, to make the priors of $\mathbf{K}$ and $\mathbf{U}$ concentrate around their respective EM estimates, denoted as $\hat{\mathbf{K}}$ and $\hat{\mathbf{U}}$.

For $\boldsymbol{\beta}$, $\{\lambda_i\}$, $\boldsymbol{\eta}$, and $\boldsymbol{\xi}^o$, whose full conditionals do not have a closed form, we used a Metropolis-Hastings algorithm within the Gibbs sampler (Gelfand and Smith, 1990; Gelman et al., 2014) to draw their posterior samples. Following Zhang and Cressie (2019), we generated the posterior samples of $\{\boldsymbol{\eta}_t\}$ successively; for $\boldsymbol{\xi}^o$, we generated the posterior samples of $\{\xi_{t,i}^o\}$ individually. Details of the Markov chain Monte Carlo (MCMC) sampling algorithm and convergence diagnostics for the Arctic SIE data are given in Section S3 of the Supplementary Material.

3.2 Prediction and Forecasting

Let $\mathcal{D}_t^u \equiv \{\mathbf{s}_{t,1}^u, \mathbf{s}_{t,2}^u, \ldots, \mathbf{s}_{t,m_t}^u\}$ be the spatial locations specified for prediction at time $t \in \{1, \ldots, T\}$, and let $\mathbf{Y}_t^u \equiv (y_t(\mathbf{s}_{t,1}^u), \ldots, y_t(\mathbf{s}_{t,m_t}^u))'$ denote the vector at those prediction locations. Note that the prediction-location set $\mathcal{D}_t^u$ and the observation-location set $\mathcal{D}_t^o$ may have a non-empty intersection. If $\mathcal{D}_t^u$ and $\mathcal{D}_t^o$ overlap, which is our interest here when modeling the Arctic SIE data, then for location $\mathbf{s}_{t,i}^o \in \mathcal{D}_t^o$, we plug in the posterior samples of $\boldsymbol{\eta}_t$, $\xi_t(\mathbf{s}_{t,i}^o)$, and $\boldsymbol{\theta}$ into the process-model equation, $y_t(\mathbf{s}_{t,i}^o) = \mathbf{x}_t(\mathbf{s}_{t,i}^o)'\boldsymbol{\beta} + \mathbf{S}_t(\mathbf{s}_{t,i}^o)'\boldsymbol{\eta}_t + \xi_t(\mathbf{s}_{t,i}^o)$, to obtain the samples from the predictive distribution of $y_t(\mathbf{s}_{t,i}^o)$,

from which summaries (e.g., mean, variance, and quantiles) of $y_t(\mathbf{s}^o_{t,i})$ can be readily obtained.

Now consider the scenario that the prediction-location set $\mathcal{D}^u_t$ and the observation-location set $\mathcal{D}^o_t$ do not overlap, which is the case when predicting in the region around the North Pole where no observations are available. Now samples from the predictive distribution can be obtained based on the posterior samples of $\boldsymbol{\eta}$ and $\boldsymbol{\theta}$. Since

$$\begin{aligned} p(\mathbf{Y}^u_t|\mathbf{Z}^o) &= \int p(\mathbf{Y}^u_t|\boldsymbol{\eta}_t, \boldsymbol{\theta}, \mathbf{Z}^o) \times p(\boldsymbol{\eta}_t, \boldsymbol{\theta}|\mathbf{Z}^o) \mathrm{d}\boldsymbol{\eta}_t \mathrm{d}\boldsymbol{\theta} \\ &= \int p(\mathbf{Y}^u_t|\boldsymbol{\eta}_t, \boldsymbol{\theta}) \times p(\boldsymbol{\eta}_t, \boldsymbol{\theta}|\mathbf{Z}^o) \mathrm{d}\boldsymbol{\eta}_t \mathrm{d}\boldsymbol{\theta}, \end{aligned} \tag{3.4}$$

the predictive samples of $\mathbf{Y}^u_t$ can be drawn by the method of composition: Having drawn the posterior samples of $\boldsymbol{\eta}_t$ and $\boldsymbol{\theta}$, $\mathbf{Y}^u_t$ is drawn from $p(\mathbf{Y}^u_t|\boldsymbol{\eta}_t, \boldsymbol{\theta})$, which is a Gaussian distribution. The predictive mean and predictive variance of $\mathbf{Y}^u_t$ are then obtained from:

$$\begin{aligned} E(\mathbf{Y}^u_t|\mathbf{Z}^o) &= E(E(\mathbf{Y}^u_t|\boldsymbol{\eta}_t, \boldsymbol{\theta})|\mathbf{Z}^o) = \mathbf{X}^u_t E(\boldsymbol{\beta}|\mathbf{Z}^o) + \mathbf{S}^u_t E(\boldsymbol{\eta}_t|\mathbf{Z}^o), \\ \mathrm{var}(\mathbf{Y}^u_t|\mathbf{Z}^o) &= E(\mathrm{var}(\mathbf{Y}^u_t|\boldsymbol{\eta}_t, \boldsymbol{\theta})|\mathbf{Z}^o) + \mathrm{var}(E(\mathbf{Y}^u_t|\boldsymbol{\eta}_t, \boldsymbol{\theta})|\mathbf{Z}^o) \\ &= E(\sigma^2_\xi|\mathbf{Z}^o)\mathbf{I}_{m_t} + \mathrm{var}(\mathbf{X}^u_t\boldsymbol{\beta} + \mathbf{S}^u_t\boldsymbol{\eta}_t|\mathbf{Z}^o), \end{aligned}$$

where $\mathbf{I}_{m_t}$ is the $m_t \times m_t$ identity matrix, $\mathbf{X}^u_t \equiv (\mathbf{x}_t(\mathbf{s}^u_{t,1}), \ldots, \mathbf{x}_t(\mathbf{s}^u_{t,m_t}))'$, and $\mathbf{S}^u_t \equiv (\mathbf{S}_t(\mathbf{s}^u_{t,1}), \ldots, \mathbf{S}_t(\mathbf{s}^u_{t,m_t}))'$.

Given the predictive samples of $\mathbf{Y}^u$, then for $\mathbf{Z}^u_t \equiv (z_t(\mathbf{s}^u_{t,1}), \ldots, z_t(\mathbf{s}^u_{t,m_t}))'$, its predictive samples are drawn from

$$p(\mathbf{Z}^u_t|\mathbf{Z}^o) = \int p(\mathbf{Z}^u_t|\mathbf{Y}^u_t, \mathbf{Z}^o) \times p(\mathbf{Y}^u_t|\mathbf{Z}^o) \mathrm{d}\mathbf{Y}^{\mathbf{u}}_{\mathbf{t}} = \int p(\mathbf{Z}^u_t|\mathbf{Y}^u_t) \times p(\mathbf{Y}^u_t|\mathbf{Z}^o) \mathrm{d}\mathbf{Y}^{\mathbf{u}}_{\mathbf{t}}, \tag{3.5}$$

again by the method of composition, using the posterior samples of $\mathbf{Y}^u_t$. The predictive mean and predictive variance of $\mathbf{Z}^u_t$ are

$$\begin{aligned} E(\mathbf{Z}^u_t|\mathbf{Z}^o) &= E(E(\mathbf{Z}^u_t|\mathbf{Y}^u_t)|\mathbf{Z}^o) = E(g^{-1}(\mathbf{Y}^u_t)|\mathbf{Z}^o), \\ \mathrm{var}(\mathbf{Z}^u_t|\mathbf{Z}^o) &= E(\mathrm{var}(\mathbf{Z}^u_t|\mathbf{Y}^u_t)|\mathbf{Z}^o) + \mathrm{var}(E(\mathbf{Z}^u_t|\mathbf{Y}^u_t)|\mathbf{Z}^o) \\ &= E(\mathrm{diag}\{g^{-1}(\mathbf{Y}^u_t)(1 - g^{-1}(\mathbf{Y}^u_t))\}|\mathbf{Z}^o) + \mathrm{var}(g^{-1}(\mathbf{Y}^u_t)|\mathbf{Z}^o), \end{aligned}$$

where $\mathrm{diag}\{\mathbf{a}\}$ denotes a diagonal matrix with its diagonal entries given by the vector $\mathbf{a}$, and recall that $g(\cdot)$ is the logit link function.

When $t = T + 1$, that is, when forecasting one-year ahead, we use the fact that $p(\boldsymbol{\eta}_{T+1}, \boldsymbol{\theta}|\mathbf{Z}^o) = \int p(\boldsymbol{\eta}_{T+1}|\boldsymbol{\eta}_T, \boldsymbol{\theta}) \times p(\boldsymbol{\eta}_T, \boldsymbol{\theta}|\mathbf{Z}^o) \mathrm{d}\boldsymbol{\eta}_T$, and equation (3.4), to obtain the one-step-ahead forecast of $\mathbf{Y}^u_{T+1}$. Using equation (3.5), $\mathbf{Z}^u_{T+1}$ can be similarly obtained.

A simulation study comparing parameter estimation and prediction performance of the EHM and the BHM is given in Section S1 of the Supplementary Material. From the simulation results, we found that when there are a limited number of observations, the EM estimates of the regression coefficients can have a large bias, but these parameters were estimated reasonably well by the BHM. We also found that the BHM yielded a

considerably smaller prediction error than that of the EHM when predicting in spatial gaps where no nearby observations are available. Also in Section S1, a sensitivity study on the inference of the proposed BHM was conducted for different plug-in values of σ_ξ^2. We observed that the inference results were very similar and, for the different plug-in values, the 95% credible intervals of the model parameters covered their respective true values.

4 Bayesian Analysis of the Arctic Sea-Ice-Extent Data

In this section, we fitted the spatio-temporal hierarchical statistical model developed in Section 2 and Section 3.1 to Arctic sea-ice-extent data obtained from remote sensing from 1997 to 2016 (20 years in total). The temporal domain is $t = 1997, \ldots, 2016$, which indexes the sea-ice data in the month of September for each of the 20 years. The original data are on a 304×448 longitude-latitude grid, involving $136,192$ monthly observations. Since the grid cells (pixels) with latitude smaller than 60°N are always water pixels in the month of September, we focus on the spatial domain that covers the observation locations with latitude 60° and above, resulting in the Arctic region from the south end of Greenland to the North Pole; see the left panel of Figure 3. In total, there are $26,342$ observation locations, and these spatial locations stay the same over time (i.e., $\mathcal{D}_t \equiv \mathcal{D}$).

We partitioned the entire 20 years into four time periods: 1997–2001, 2002–2006, 2007–2011, and 2012–2016, in order to allow the dimension-reduced propagator matrix $\mathbf{H}$ and the corresponding innovation matrix $\mathbf{U}$ to vary over different time periods. That is, we fitted the BHM (2.1)–(2.5) with priors given in Section 3.1, but σ_ξ^2 was estimated (EM estimation) from the data in each time period separately and substituted into the BHM. Note that the end year of the previous time period was used to initialize the analysis of the current time period: Specifically, $\mathbf{K}_t = \mathbf{H}\mathbf{K}_{t-1}\mathbf{H}' + \sigma_\xi^2\mathbf{I}_{N_t}$, for $t = 2, \ldots, T$, where $\mathbf{K}_1 \equiv \mathbf{K}, \mathbf{H}$, and σ_ξ^2 were obtained from the previous time period. This helps to avoid artificial abrupt transitions when inferring $\{y_t(\mathbf{s})\}$.

The surface-temperature-anomaly data that were used to define covariates $\mathbf{x}_t(\mathbf{s})$ in (2.5) come from the Goddard Institute for Space Studies Surface Temperature Analysis (GISTEMP) project (e.g., Hansen et al., 2010; GISTEMP-Team, 2018). This data set is on a $2° \times 2°$ longitude-latitude grid, and hence its resolution is coarser than the spatial grid of the Arctic SIE data, which has a resolution of approximately 25km × 25km. To each sea-ice-observation location (i.e., center of the grid cell), we assigned its surface-air-temperature-anomaly value to be the closest surface-air-temperature-anomaly datum in the GISTEMP data set. Recall that we use eight covariates in (2.5) for modeling the fixed-effects term for the latent process $\{y_t(\mathbf{s})\}$, all of which are defined at all observation locations by this interpolation method. Further discussion of the covariates is given in the next subsection.

4.1 BHM Fitting

Following Zhang and Cressie (2019), we used a two-resolution design for the basis functions. We defined $r_1 = 45$ Resolution-1 basis-function centers and $r_2 = 172$ Resolution-2

basis-function centers on regularly spaced geodesic grids on the polar cap. Some basis-function centers were placed outside the study domain to account for boundary effects (Cressie and Kang, 2010). Specifically, the Matlab function GridSphere (Laven, 2015) was used to generate basis-function centers of two resolutions, where the great-circle radii of the Resolution-1 and Resolution-2 basis functions are 881.71km and 440.86km, respectively. The bisquare basis functions in (2.3) were used to form $\{\mathbf{S}_t(\mathbf{s})'\boldsymbol{\eta}_t\}$, the small-scale-variation component of $\{y_t(\mathbf{s})\}$, and the basis-function matrices are $\mathbf{S}_t \equiv \mathbf{S}$ for each time period. We then fitted the proposed dimension-reduced Bayesian ST-GLMM to the Arctic SIE data for Periods 1–4, with the fine-scale variance σ^2_ξ in each given period fixed at its EM estimate.

Parameter	Period 1	Period 2	Period 3	Period 4
β_0	-1.298 $(-1.392, -1.205)$	-2.284 $(-2.339, -2.226)$	-3.416 $(-3.483, -3.349)$	-2.748 $(-2.811, -2.686)$
β_1	-0.026 $(-0.114, 0.059)$	-1.572 $(-1.714, -1.430)$	0.735 $(0.595, 0.879)$	-1.545 $(-1.745, -1.339)$
β_2	0.892 $(0.764, 1.018)$	-0.225 $(-0.326, -0.116)$	0.159 $(-0.032, 0.322)$	-0.949 $(-1.077, -0.821)$
β_3	0.543 $(0.493, 0.589)$	-0.011 $(-0.058, 0.040)$	0.209 $(0.162, 0.255)$	0.195 $(0.150, 0.238)$
β_4	-0.461 $(-0.511, -0.410)$	0.089 $(0.043, 0.134)$	0.503 $(0.452, 0.557)$	0.052 $(0.009, 0.095)$
β_5	-3.593 $(-3.661, -3.525)$	-2.056 $(-2.125, -1.982)$	-1.192 $(-1.254, -1.131)$	-0.925 $(-0.988, -0.863)$
β_6	-2.626 $(-2.686, -2.566)$	-2.453 $(-2.530, -2.378)$	-2.545 $(-2.609, -2.481)$	-2.920 $(-2.989, -2.857)$
β_7	-7.286 $(-7.404, -7.170)$	-5.893 $(-6.010, -5.778)$	-4.970 $(-5.072, -4.875)$	-4.605 $(-4.709, -4.507)$
β_8	1.393 $(1.248, 1.534)$	1.050 $(0.901, 1.187)$	0.723 $(0.598, 0.856)$	0.645 $(0.525, 0.759)$
λ_1	0.590 $(0.587, 0.592)$	0.492 $(0.489, 0.495)$	0.691 $(0.690, 0.694)$	0.558 $(0.555, 0.560)$
λ_2	0.400 $(0.398, 0.403)$	0.473 $(0.470, 0.476)$	0.525 $(0.523, 0.527)$	0.441 $(0.438, 0.443)$
λ_3	0.010 $(0.010, 0.011)$	0.019 $(0.019, 0.020)$	-0.028 $(-0.028, -0.027)$	-0.013 $(-0.014, -0.013)$
σ^2_ξ	0.393 (fixed)	0.430 (fixed)	0.341 (fixed)	0.352 (fixed)

Table 1: Posterior means and 95% credible intervals (in parentheses) for scalar parameters from fitting a BHM to the Arctic SIE data. The fine-scale variance σ^2_ξ is fixed at its EM estimate in each time period.

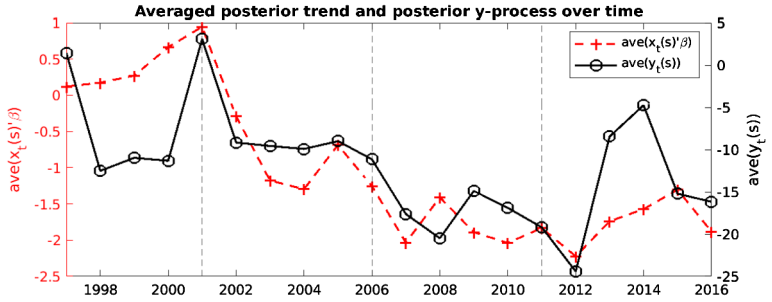


Figure 4: The posterior means of $\text{ave}\{\mathbf{x}_t(\mathbf{s})'\boldsymbol{\beta} : \mathbf{s} = (\text{lon}, \text{lat}), \text{and lat} \in (\text{lat}_0 - \Delta, \text{lat}_0 + \Delta)\}$ and $\text{ave}\{y_t(\mathbf{s}) : \mathbf{s} = (\text{lon}, \text{lat}), \text{and lat} \in (\text{lat}_0 - \Delta, \text{lat}_0 + \Delta)\}$, where the target latitude value is $\text{lat}_0 = 75°\text{N}$ and $2\Delta = 1°$ is the bandwidth. The ends of Period 1, 2, 3, and 4 are 2001, 2006, 2011, and 2016, respectively, and are shown with a vertical dashed line.

Table 1 gives the parameter-inference results for our proposed BHM, where both posterior means and corresponding 95% credible intervals of parameters are reported. The covariates contribute in different ways, depending on the time period. To visualize the spatio-temporal variability that they capture, Figure 4 shows a time series plot of

the posterior mean of ave$\{\mathbf{x}_t(\mathbf{s})'\boldsymbol{\beta}\}$ as well as the corresponding time series plot of the posterior mean of ave$\{y_t(\mathbf{s})\}$, where in each case $\mathbf{s}$ is averaged over a 1° latitude band centered on $\text{lat}_0 = 75°\text{N}$. From the scales on the vertical axis in Figure 4, the covariates are less important than the spatio-temporal variability in the random effects.

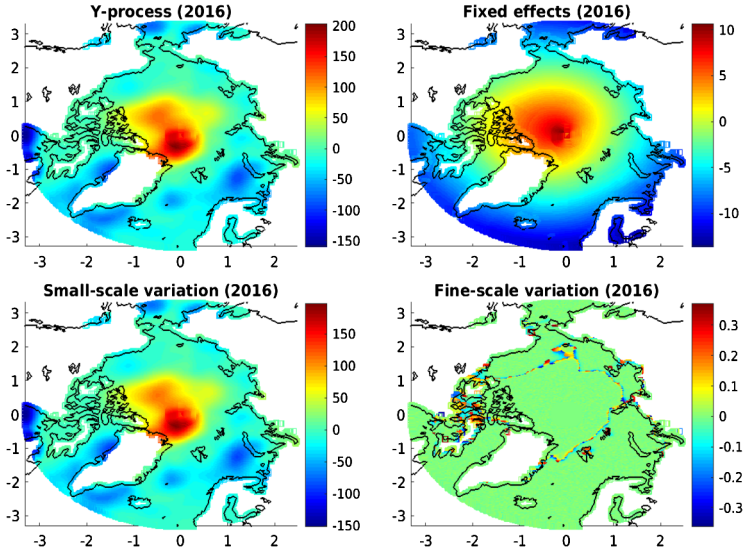
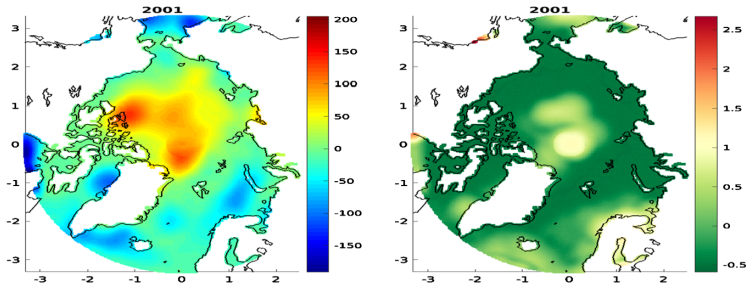


Figure 5: Year $t = 2016$. Top left: Posterior mean of the y-process. Top right: Posterior mean of the fixed-effects component of the y-process. Bottom left: Posterior mean of the small-scale component of the y-process. Bottom right: Posterior mean of the fine-scale component of the y-process. Adding these three components results in the posterior mean of the y-process, shown top left. (Note that the scales are different and the small-scale variation's contribution to the y-process dominates.)

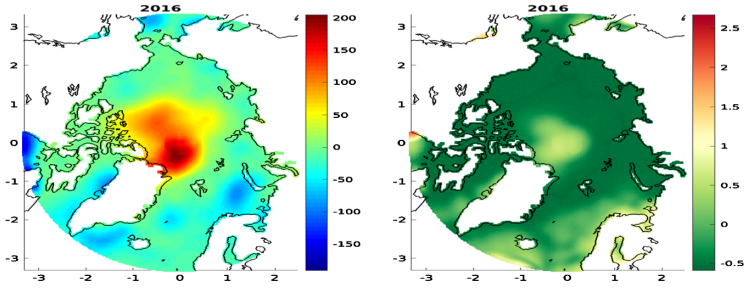
For the propagator matrix $\mathbf{H}$, the within-resolution covariance parameters λ_1 and λ_2 have a dominant magnitude, and the correlations for Resolution-1 and Resolution-2 basis functions between t and $t-1$ are always positive. The between-resolution covariance parameter λ_3 has a very small magnitude, where it is positive for the first two periods and then becomes negative for the latter two periods. The changing sign of λ_3 implies that the autoregressive structure of $\{\boldsymbol{\eta}_t\}$ changes from the earlier decade to the recent decade, something we see in a striking way when we compare temporal semivariograms between the two decades (Section 4.2).

Figure 5 shows the posterior mean of the latent process $\{y_{2016}(\mathbf{s})\}$, along with its fixed-effects and its random-effects components. The detailed spatial variations are mainly captured by the small-scale-variation component, $\{\mathbf{S}(\mathbf{s})'\boldsymbol{\eta}_{2016}\}$, while the amplitude of the fine-scale-variation component is tiny and mostly located around the ice–water boundaries.

Figure 6 and Figure 7 show the posterior means and posterior standard deviations of $\{y_t(\mathbf{s}) : t = 2001, 2006\}$ and $\{p_t(\mathbf{s}) : t = 2001, 2006\}$, respectively; the 0.05 and 0.95



(a) $t = 2001$, y-scale



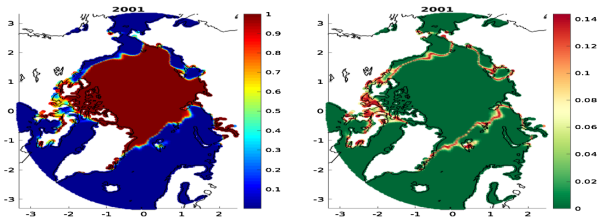
(b) $t = 2016$, y-scale

Figure 6: The posterior means (left panels) and posterior standard deviations (on the log scale; right panels) of the logit-transformed process $\{y_t(\mathbf{s})\}$. (a) $t = 2001$; (b) $t = 2016$.

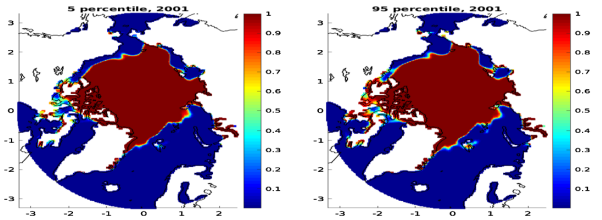
posterior quantiles of $\{p_t(\mathbf{s})\}$ are also displayed. We observe that on the logit scale (y-scale), large uncertainties typically appear at locations where process values have large magnitudes. On the probability scale (p-scale), large uncertainties occur at the ice–water boundary locations, as expected. It is clear that the Arctic sea ice has decreased substantially from the end year of Period 1 (2001) to the end year of Period 4 (2016).

4.2 Evolution of Arctic Sea Ice

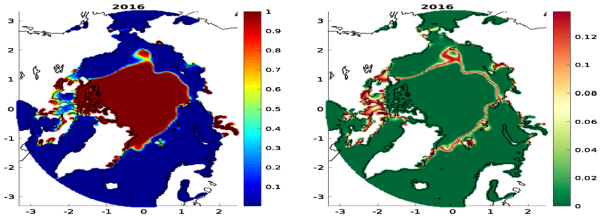
In this subsection, we first consider the spatial variability of the latent process (on the y-scale) in different latitude bands for the end years of Periods 1–4. Specifically, at $t = 2001, 2006, 2011, 2016$, we compare the five-number boxplots of $\{E(y_t(\mathbf{s})|\mathbf{Z}^o)\}$ within the three latitude bands around 75°N, 80°N, and 85°N. Figure 8 shows the boxplots augmented with the average (denoted by a dot); by varying $\mathbf{s} = (\text{lon}, \text{lat})$ within the given latitude band $(\text{lat}_0 - \Delta, \text{lat}_0 + \Delta)$ for $2\Delta = 1°$, we obtain the boxplot. Because on the y-scale, the value 0 corresponds to 0.5 on the probability scale, which is a natural cut-off value that says the potential for ice and water is the same, we feature the value 0 in the boxplots on the y-scale. By looking across the columns, we see the evolution of Arctic sea ice over the two decades of data, at time points five years apart. The earlier



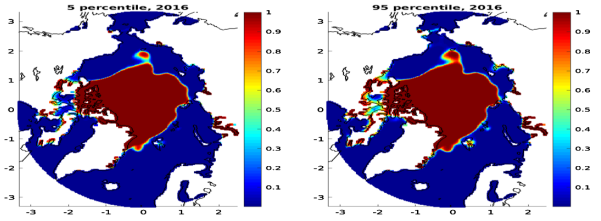
(a) $t = 2001$, posterior mean and s.d. (p-scale)



(b) $t = 2001$, the 0.05 and 0.95 quantiles (p-scale)



(c) $t = 2016$, posterior mean and s.d. (p-scale)



(d) $t = 2016$, the 0.05 and 0.95 quantiles (p-scale)

Figure 7: (a) and (c): The posterior means (left panel) and posterior standard deviations (right panel) of the probability process $\{p_t(\mathbf{s})\}$. (b) and (d): Posterior 0.05 quantile (left panel), and posterior 0.95 quantile (right panel) of $p_t(\mathbf{s})$. The year $t = 2001$ is shown in (a) and (b), and the year $t = 2016$ is shown in (c) and (d).

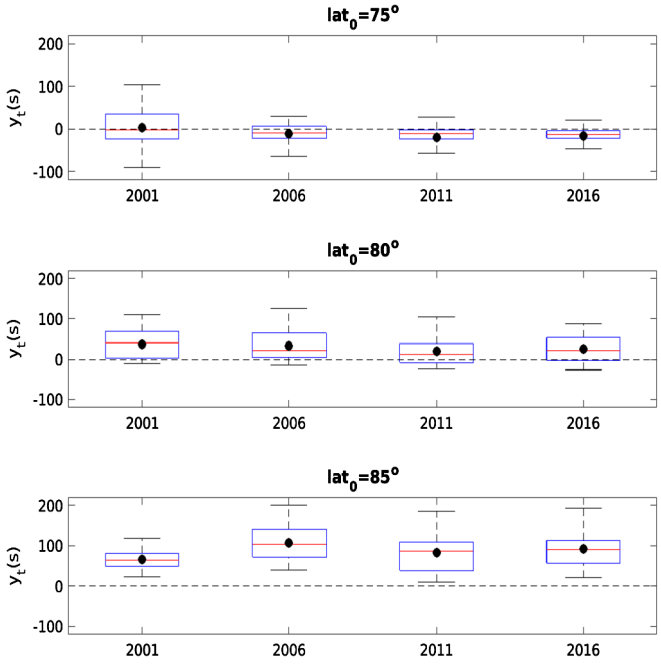


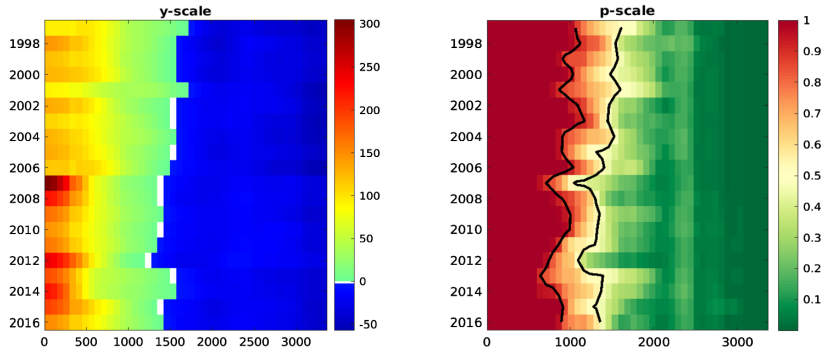
Figure 8: Boxplots of $\{E(y_t(\mathbf{s})|\mathbf{Z}^o) : \mathbf{s} = (\text{lon}, \text{lat}), \text{and lat} \in (\text{lat}_0 - \Delta, \text{lat}_0 + \Delta)\}$ for three latitude bands and for years $t = 2001, 2006, 2011, 2016$, where the width of the latitude band is $2\Delta = 1°$. The dashed line indicates the value of zero for $y_t(\mathbf{s})$ (which corresponds to 0.5 on the probability scale).

decade is characterized by the behavior at $t = 2001$ and 2006, and the recent decade by the behavior at $t = 2011$ and 2016.

The evolution is clear at $\text{lat}_0 = 75°$N and 80°N: The central tendencies decrease over time (more water and less ice). At 85°N, the sea ice appears to be in equilibrium over the period of observation, but a more refined (temporal semivariogram) analysis below, shows oscillatory behavior in the recent decade.

Figure 2 shows similar boxplots on the p-scale in different latitude bands, which provide a clearer contrast between the earlier and the latest decade. A detailed discussion is given in Section 1, where attention is directed towards a possible future collapse of the sea ice at latitude 80°N.

The yearly temporal dynamics of the latent process $\{y_t(\mathbf{s})\}$ (as well as that of $\{p_t(\mathbf{s})\}$) can be visualized through Hovmöller diagrams (e.g., Cressie and Wikle, 2011, Section 5.1.2). Figure 9 shows the Hovmöller diagrams of the posterior means of both $\{y_t(\mathbf{s})\}$ and $\{p_t(\mathbf{s})\}$, where the horizontal axis is the great-circle distance of $\mathbf{s}$ to the North Pole in units of km, and the vertical axis is time in units of year. Specifically, for each distance x and year t, we obtained the spatially averaged posterior mean of $y_t(\mathbf{s})$ as $\bar{y}_t(x) = \text{ave}\{E(y_t(\mathbf{s})|\mathbf{Z}^o) : \|\mathbf{s} - \mathbf{s}^*\| \in (x - h, x + h)\}$, where $\mathbf{s}^*$ is the North Pole,



(a) y-scale (b) p-scale

Figure 9: Hovmöller diagrams showing the posterior means of the spatial averages, $\{\bar{y}_t(x)\}$ (left panel) and $\{\bar{p}_t(x)\}$ (right panel), for each year, where the average is taken over all pixels at distance $x \pm h$ from the North Pole. The horizontal axis is in units of km, and the vertical axis is in units of year. The right panel also shows the 0.9 and 0.5 contours from left to right.

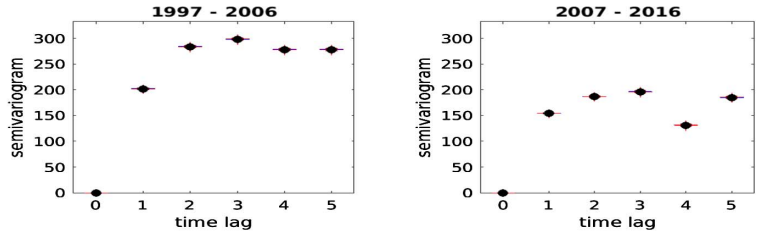
$2h = 75$km is a *distance bandwidth*, and $\|\mathbf{s} - \mathbf{s}^*\|$ computes the great-circle distance between two spatial locations. An analogous quantity on the p-scale, denoted by $\bar{p}_t(x)$, is obtained and can be interpreted as the potential for sea ice at (x, t). These are plotted as a function of (x, t) in Figure 9.

The 0.9 and 0.5 contours plotted on the p-scale show that in the second decade the potential for sea ice at distances 1000 km or more has decreased substantially. Section S2 of the Supplementary Material shows, through Hovmöller diagrams in different regions of the Arctic, that this decrease is not uniform. The regional Hovmöller diagrams in Section S2 provide a very effective way of assessing where the sea ice is most sensitive to a warming climate.

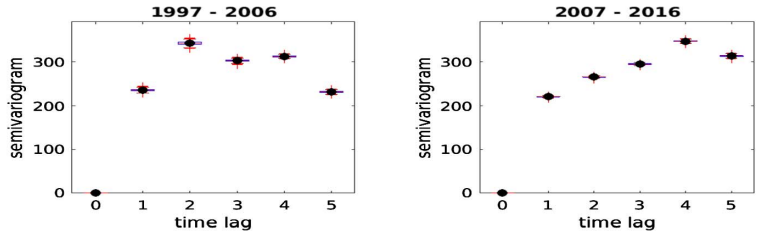
The sea-ice extent plotted in Figure 1 and the potential for sea ice (p-scale) plotted in Figure 9(b) indicate a slower and deeper oscillatory behavior in the recent decade. We investigate the temporal-dependence structure for the two decades by using the empirical temporal semivariogram computed for the detrended latent process for different latitude bands: For $h = 1, \ldots, M$,

$$\bar{\gamma}(h; \text{lat}_0) \equiv \frac{1}{2}\text{ave}\{(\delta_t(\mathbf{s}) - \delta_{t+h}(\mathbf{s}))^2 : \mathbf{s} = (\text{lon}, \text{lat}), \text{and lat} \in (\text{lat}_0 - \Delta, \text{lat}_0 + \Delta)\},$$

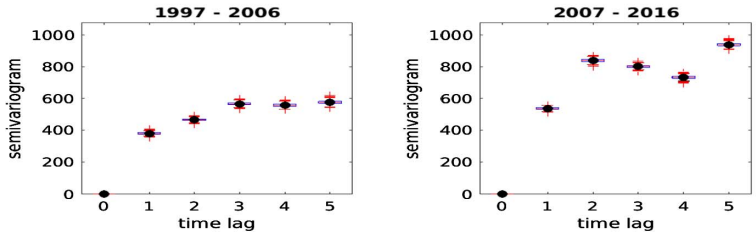
where $\delta_t(\mathbf{s}) \equiv \mathbf{S}_t(\mathbf{s})'\boldsymbol{\eta}_t + \xi_t(\mathbf{s})$ is the random-effect component obtained from a sample from the posterior distribution, $\mathbf{s} \equiv (\text{lon}, \text{lat})'$, lat_0 is the target latitude, $2\Delta = 1^\circ$ is the width of the latitude band, and the upper temporal lag, M, is half the range of possible lags rounded up (e.g., Cressie, 1988). Then we obtained the five-number boxplots of the posterior distribution of $\bar{\gamma}(h; \text{lat}_0)$ as well as $E(\bar{\gamma}(h; \text{lat}_0)|\mathbf{Z}^o)$ (denoted by a dot) from the MCMC samples from the posterior distribution of $\{\delta_t(\mathbf{s})\}$ for the earlier decade, 1997–2006, and for the recent decade, 2007–2016, where in both cases $M = \lfloor 9/2 \rfloor = 5$.



(a) $\text{lat}_0 = 75°\text{N}$



(b) $\text{lat}_0 = 80°\text{N}$



(c) $\text{lat}_0 = 85°\text{N}$

Figure 10: The boxplots and the average (indicated by the solid black dot) of samples from the posterior distribution of temporal semivariograms, $\{\bar{\gamma}(h; \text{lat}_0) : h = 1, \ldots, 5\}$, for the three latitude bands, around 75°N, 80°N, and 85°N, where the width of the latitude band is $2\Delta = 1°$.

Figure 10 shows these for $\text{lat}_0 = 75°\text{N}, 80°\text{N}$, and 85°N, and it should be noted from the boxplots that our posterior inference is very precise. There are several remarkable features that indicate differences between the two decades. At $\text{lat}_0 = 75°\text{N}$, the sill is much smaller for the recent decade, yet the sill is much larger for the recent decade at $\text{lat}_0 = 85°\text{N}$ (and about the same at $\text{lat}_0 = 80°\text{N}$). Thus, the contrast in the evolution of the sea ice has flipped when comparing higher latitudes to lower latitudes. In the recent decade, the consistent "dip" at lag 4 of the semivariograms for $\text{lat}_0 = 75°\text{N}$ and $\text{lat}_0 = 85°\text{N}$ indicates a periodicity that is reflected in the substantial oscillations around the trend line seen in Figure 1. That it appears at high latitudes indicates the need to monitor the sea ice for evolutionary changes both away from and close to the North Pole.

The empirical semivariograms have added to our understanding of the temporal behavior of Arctic sea ice, already obtained from boxplots and Hovmöller diagrams. Of

concern is a possible scenario in the future where the polar ice cap collapses in a matter of years not decades.

5 Conclusion

In this article, we develop a Bayesian hierarchical model (BHM) for the binary Arctic sea-ice-extent (SIE) data, where the 0–1 observations are modeled using a logistic link function in the data model and a spatio-temporal linear mixed-effects model in the process model. Compared with the empirical hierarchical model (EHM) in Zhang and Cressie (2019), the BHM accounts for parameter uncertainties and, in Section S1 of the Supplementary Material, we show in a simulation study that a BHM can yield better prediction accuracy than an EHM in spatial gaps where there are no nearby data. Furthermore, the BHM allows the uncertainties of parameters such as regression coefficients and covariance parameters to be directly obtained from MCMC samples, which allows proper statistical interpretation of the effects of covariates and the data-dependence structure. In contrast, parameter uncertainties from the EM algorithm require numerical differentiations (e.g., Meng and Rubin, 1991; Jamshidian and Jennrich, 2000). When constructing the ST-GLMM for the latent process, several physically motivated covariates were used for the fixed-effects component. The BHM yields credible intervals for the coefficients of these covariates, and they are almost all substantially different from zero.

When modeling the Arctic SIE data, we assume $\mathbf{H}_t$ stays the same in each five–year time period. This can be extended to the case where $\mathbf{H}_t$ varies over time by putting a prior on the distribution of $\mathbf{H}_t$ (e.g., Berliner et al., 2000). Further parameterization of $\mathbf{K}$ increases the degrees of the freedom and would lead to better estimation of both $\mathbf{K}$ and σ^2_ξ. For example, $\mathbf{K}$ may be parameterized using a parametric covariance function, where the basis-function centers are the spatial locations (e.g., Katzfuss and Cressie, 2009). Alternatively, the neighborhood information of the basis-function centers may be used to form a known baseline covariance matrix $\mathbf{K}_0$ (Bradley et al., 2015) or a known baseline precision matrix $\mathbf{Q}_0$ (e.g., Reich et al., 2006; Hughes and Haran, 2013). Then let $\mathbf{K} = \sigma^2_K \mathbf{K}_0$ or $\mathbf{K}^{-1} = \tau^2_K \mathbf{Q}_0$, which involves only one unknown parameter σ^2_K (or τ^2_K) to be estimated.

The focus of our paper is to propose useful functionals that characterize the space-time changes of Arctic sea ice, particularly between an earlier decade and a recent decade. The boxplots of the posterior means of $\{p_t(\mathbf{s})\}$ in Figure 2 in intermediate latitude bands show an obvious declining tendency over two decades. We also used Hovmöller diagrams to visualize the temporal dynamics of the posterior means of the latent processes, $\{y_t(\mathbf{s})\}$ and $\{p_t(\mathbf{s})\}$, both for the whole Arctic region (Section 4.2) and regionally (Section S2 of the Supplementary Material). We found that for the entire Arctic, the 0.5-probability contour of the posterior mean of the latent probability process, $\{p_t(\mathbf{s})\}$, which represents an important boundary of the sea–ice potential, has moved northwards over time. Not surprisingly, this behavior is more obvious for years when substantially smaller Arctic SIE was observed (e.g., 2007 and 2012). Furthermore, we made a comparison of the empirical temporal semivariograms between the earlier

decade before 2007 and the recent decade from 2007 at different latitudes, and we found that their temporal-dependence structures were very different. Our results demonstrate the importance of monitoring sea ice at high latitudes as well as at latitudes where the sea–ice boundary occurs, and they propose summaries of the data that detect changes in the sea ice. We have also observed and summarized highly oscillatory behavior in the recent decade, noting the possibility of substantial changes in a matter of years rather than decades.

A change in Arctic sea ice can impact the weather in other regions, which indicates that future research should focus on teleconncections of the effects of Arctic forcing on other environmental variables (e.g., surface temperature and precipitation) in other regions (Cohen et al., 2014). Another important research project would be the albedo–ice feedback of the Arctic region, which could be studied using a bivariate spatio-temporal statistical model of sea-ice concentrations and albedo.

Supplementary Material

Bayesian spatio-temporal modeling of Arctic sea ice extent. Supplementary Material (DOI: 10.1214/20-BA1209SUPP; .pdf). The Supplementary Material contains a simulation study that compares the inference performance of the empirical hierarchical model (EHM) and the Bayesian hierarchical model (BHM) in Section S1; a detailed discussion of the dynamics of sea ice in different regions of the Arctic in Section S2; details of the MCMC sampling algorithm and convergence diagnostics for the Arctic sea-ice-extent data in Section S3; the classification accuracy of the Bayesian inference in Section S4; and visualization of the ice-to-water and water-to-ice transition probabilities in Section S5.

References

Banerjee, S., Gelfand, A., Finley, A., and Sang, H. (2008). "Gaussian predictive process models for large spatial data sets." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 70: 825–848. MR2523906. doi: https://doi.org/10.1111/j.1467-9868.2008.00663.x. 607

Berliner, L. M., Wikle, C. K., and Cressie, N. (2000). "Long-lead prediction of Pacific SSTs via Bayesian dynamic modeling." *Journal of climate*, 13: 3953–3968. 625

Bradley, J. R., Holan, S. H., and Wikle, C. K. (2015). "Multivariate spatio-temporal models for high-dimensional areal data with application to longitudinal employer-household dynamics." *Annals of Applied Statistics*, 9: 1761–1791. MR3456353. doi: https://doi.org/10.1214/15-AOAS862. 610, 611, 625

Bradley, J. R., Holan, S. H., and Wikle, C. K. (2018). "Computationally efficient distribution theory for Bayesian inference of high-dimensional dependent count-valued data (with discussion)." *Bayesian Analysis*, 13: 253–310. MR3773410. doi: https://doi.org/10.1214/17-BA1069. 607

Bradley, J. R., Wikle, C. K., and Holan, S. H. (2016). "Bayesian spatial change of support for count-valued survey data with application to the American Community Survey." *Journal of the American Statistical Association*, 111: 472–487. MR3538680. doi: https://doi.org/10.1080/01621459.2015.1117471. 607

Chang, W., Haran, M., Applegate, P., and Pollard, D. (2016a). "Calibrating an ice sheet model using high-dimensional binary spatial data." *Journal of the American Statistical Association*, 111: 57–72. MR3494638. doi: https://doi.org/10.1080/01621459.2015.1108199. 606

Chang, W., Haran, M., Applegate, P., and Pollard, D. (2016b). "Improving ice sheet model calibration using paleoclimate and modern data." *Annals of Applied Statistics*, 10: 2274–2302. 606

Cohen, J., Screen, J. A., Furtado, J. C., Barlow, M., Whittleston, D., Coumou, D., Francis, J., Dethloff, K., Entekhabi, D., Overland, J., et al. (2014). "Recent Arctic amplification and extreme mid-latitude weather." *Nature Geoscience*, 7: 627–637. 605, 626

Cressie, N. (1988). "A graphical procedure for determining nonstationarity in time series." *Journal of the American Statistical Association*, 83: 1108–1116. MR0997589. 623

Cressie, N. and Johannesson, G. (2006). "Spatial prediction for massive datasets." In *Mastering the Data Explosion in the Earth and Environmental Sciences: Proceedings of the Australian Academy of Science Elizabeth and Frederick White Conference*, 1–11. Australian Academy of Science, Canberra, Australia. 610

Cressie, N. and Johannesson, G. (2008). "Fixed rank kriging for very large spatial data sets." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 70: 209–226. MR2412639. doi: https://doi.org/10.1111/j.1467-9868.2007.00633.x. 607, 610, 611

Cressie, N. and Kang, E. L. (2010). "High-resolution digital soil mapping: Kriging for very large datasets." In Viscarra Rossel, R. A., McBratney, A. B., and Minasny, B. (eds.), *Proximal Soil Sensing*, 49–63. Springer, Dordrecht, NL. 618

Cressie, N., Shi, T., and Kang, E. L. (2010). "Fixed rank filtering for spatio-temporal data." *Journal of Computational and Graphical Statistics*, 19: 724–745. MR2732500. doi: https://doi.org/10.1198/jcgs.2010.09051. 610

Cressie, N. and Wikle, C. K. (2011). *Statistics for Spatio-Temporal Data*. John Wiley & Sons, Hoboken, NJ, USA. MR2848400. 622

Cvijanovic, I., Santer, B. D., Bonfils, C., Lucas, D. D., Jch, C., and Zimmerman, S. (2017). "Future loss of Arctic sea-ice cover could drive a substantial decrease in California's rainfall." *Nature Communications*, 8: 1947. 605

Datta, A., Banerjee, S., Finley, A. O., and Gelfand, A. E. (2016). "Hierarchical Nearest-Neighbor Gaussian Process Models for Large Geostatistical Datasets." *Journal of the American Statistical Association*, 111: 800–812. MR3538706. doi: https://doi.org/10.1080/01621459.2015.1044091. 607

Dempster, Arthur P., Laird, Nan M. and Rubin, Donald B. (1977). "Maximum likelihood from incomplete data via the EM algorithm." *Journal of the Royal Statistical Society: Series B (Methodological)*, 39: 1–22. 608

Diggle, P. J., Tawn, J., and Moyeed, R. (1998). "Model-based geostatistics (with discussion)." *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 47: 299–350. MR1626544. doi: https://doi.org/10.1111/1467-9876.00113. 607, 609

Director, H. M., Raftery, A. E., and Bitz, C. M. (2017). "Improved sea ice forecasting through spatiotemporal bias correction." *Journal of Climate*, 30: 9493–9510. 607

Director, H. M., Raftery, A. E., and Bitz, C. M. (2019). "Probabilistic forecasting of the Arctic sea ice edge with contour modeling." https://arxiv.org/abs/1908.09377. 607

Furrer, R., Genton, M. G., and Nychka, D. (2006). "Covariance tapering for interpolation of large spatial datasets." *Journal of Computational and Graphical Statistics*, 15: 502–523. MR2291261. doi: https://doi.org/10.1198/106186006X132178. 607

Gelfand, A. E. and Smith, A. F. M. (1990). "Sampling-based approaches to calculating marginal densities." *Journal of the American Statistical Association*, 85: 398–409. MR1141740. 615

Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. (2014). *Bayesian Data Analysis*, 3rd edn. Chapman & Hall/CRC, Boca Raton, FL, USA. MR2027492. 615

GISTEMP-Team (2018). "GISS Surface Temperature Analysis (GISTEMP). NASA Goddard Institute for Space Studies." Dataset accessed 2019-01-08 at https://data.giss.nasa.gov/gistemp/. 617

Guan, Y. and Haran, M. (2018). "A computationally efficient projection-based approach for spatial generalized linear mixed models." *Journal of Computational and Graphical Statistics*, 27: 701–714. MR3890863. doi: https://doi.org/10.1080/10618600.2018.1425625. 607

Hansen, J. R., Ruedy, R., Sato, M., and Lo, K. (2010). "Global Surface Temperature Change." *Reviews of Geophysics*, 48: RG4004. doi: 10.1029/2010RG000345. 617

Holan, S. H. and Wikle, C. K. (2016). "Hierarchical dynamic generalized linear mixed models for discrete-valued spatio-temporal data." In Davis, R. A., Holan, S. H., Lund, R., and Ravishanker, N. (eds.), *Handbook of Discrete-Valued Time Series*, 327–348. Chapman & Hall/CRC, Boca Raton, FL, USA. MR3699412. 607

Hughes, J. and Haran, M. (2013). "Dimension reduction and alleviation of confounding for spatial generalized linear mixed models." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 75: 139–159. MR3008275. doi: https://doi.org/10.1111/j.1467-9868.2012.01041.x. 625

Jamshidian, M. and Jennrich, R. I. (2000). "Standard errors for EM estimation." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 62: 257–270. MR1749538. doi: https://doi.org/10.1111/1467-9868.00230. 625

Kang, E. L., Cressie, N., and Shi, T. (2010). "Using temporal variability to improve spatial mapping with application to satellite data." *Canadian Journal of Statistics*, 38: 271–289. MR2682762. doi: https://doi.org/10.1002/cjs.10063. 610

Katzfuss, M. (2017). "A multi-resolution approximation for massive spatial datasets." *Journal of the American Statistical Association*, 112: 201–214. MR3646566. doi: https://doi.org/10.1080/01621459.2015.1123632. 607, 610

Katzfuss, M. and Cressie, N. (2009). "Maximum likelihood estimation of covariance parameters in the spatial-random-effects model." In *Proceedings of the Joint Statistical Meetings, Alexandria, VA: American Statistical Association*, 3378–3390. MR2894757. doi: https://doi.org/10.1198/jasa.2011.tm09680. 625

Katzfuss, M. and Cressie, N. (2011). "Spatio-temporal smoothing and EM estimation for massive remote-sensing data sets." *Journal of Time Series Analysis*, 32: 430–446. MR2841794. doi: https://doi.org/10.1111/j.1467-9892.2011.00732.x. 610, 611

Kumar, A., Perlwitz, J., Eischeid, J., Quan, X., Xu, T., Tao, Z., Hoerling, M., Jha, B., and Wang, W. (2010). "Contribution of sea ice loss to Arctic amplification." *Geophysical Research Letters*, 37: 389–400. 605

Laven, K. V. (2015). "Grid Sphere (version 1.15)." doi: https://au.mathworks.com/matlabcentral/fileexchange/28842-grid-sphere [Accessed: 17 February 2020]. 618

Linero, A. R. and Bradley, J. R. (2018). "Multi-rubric models for ordinal spatial data with application to online ratings from Yelp." *Annals of Applied Statistics*, 12: 2054–2074. MR3875692. doi: https://doi.org/10.1214/18-AOAS1143. 607

Meier, W. N., Hovelsrud, G. K., Oort, B. E., Key, J. R., Kovacs, K. M., Michel, C., Haas, C., Granskog, M. A., Gerland, S., Perovich, D. K., et al. (2014). "Arctic sea ice in transformation: A review of recent observed changes and impacts on biology and human activity." *Reviews of Geophysics*, 52: 185–217. 605

Meier, W. N., Stroeve, J., and Fetterer, F. (2007). "Whither Arctic sea ice? A clear signal of decline regionally, seasonally and extending beyond the satellite record." *Annals of Glaciology*, 46: 428–434. 606

Meng, X.-L. and Rubin, D. B. (1991). "Using EM to obtain asymptotic variance-covariance matrices: The SEM algorithm." *Journal of the American Statistical Association*, 86: 899–909. 625

Mori, M., Watanabe, M., Shiogama, H., Inoue, J., and Kimoto, M. (2014). "Robust Arctic sea-ice influence on the frequent Eurasian cold winters in past decades." *Nature Geoscience*, 7: 869–873. 605

Nguyen, H., Cressie, N., and Braverman, A. (2012). "Spatial statistical data fusion for remote sensing applications." *Journal of the American Statistical Association*, 107: 1004–1018. MR3010886. doi: https://doi.org/10.1080/01621459.2012.694717. 607

Nychka, D., Bandyopadhyay, S., Hammerling, D., Lindgren, F., and Sain, S. (2015). "A multiresolution Gaussian process model for the analysis of large spatial datasets." *Journal of Computational and Graphical Statistics*, 24: 579–599. MR3357396. doi: https://doi.org/10.1080/10618600.2014.914946. 607, 610

Olson, R., An, S.-I., Fan, Y., Chang, W., Evans, J., and Lee, J.-Y. (2019). "A novel method to test non-exclusive hypotheses applied to Arctic ice projections from dependent models." *Nature Communications*, 10: 3016. 606

Parkinson, C. L. (2014a). "Global sea ice coverage from satellite data: Annual cycle and 35-yr trends." *Journal of Climate*, 27: 9377–9382. 606

Parkinson, C. L. (2014b). "Spatially mapped reductions in the length of the Arctic sea ice season." *Geophysical Research Letters*, 41: 4316–4322. 606

Parkinson, C. L., Cavalieri, D. J., Gloersen, P., Zwally, H. J., and Comiso, J. C. (1999). "Arctic sea ice extents, areas, and trends, 1978–1996." *Journal of Geophysical Research: Oceans*, 104: 20837–20856. 606, 612

Parkinson, C. L. and DiGirolamo, N. E. (2016). "New visualizations highlight new information on the contrasting Arctic and Antarctic sea-ice trends since the late 1970s." *Remote Sensing of Environment*, 183: 198–204. 606

Pistone, K., Eisenman, I., and Ramanathan, V. (2014). "Observational determination of albedo decrease caused by vanishing Arctic sea ice." *Proceedings of the National Academy of Sciences*, 111: 3322–3326. 605

Reich, B. J., Hodges, J. S., and Zadnik, V. (2006). "Effects of residual smoothing on the posterior of the fixed effects in disease-mapping models." *Biometrics*, 62: 1197–1206. MR2307445. doi: https://doi.org/10.1111/j.1541-0420.2006.00617.x. 625

Rue, H., Martino, S., and Chopin, N. (2009). "Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 71: 319–392. MR2649602. doi: https://doi.org/10.1111/j.1467-9868.2008.00700.x. 607

Sang, H. and Huang, J. Z. (2012). "A full scale approximation of covariance functions for large spatial data sets." *Journal of the Royal Statistical Society, Series B (Statistical Methodology)*, 74: 111–132. MR2885842. doi: https://doi.org/10.1111/j.1467-9868.2011.01007.x. 607

Screen, J. A. and Simmonds, I. (2013). "Exploring links between Arctic amplification and mid-latitude weather." *Geophysical Research Letters*, 40: 959–964. 605

Screen, J. A., Simmonds, I., Deser, C., and Tomas, R. (2013). "The atmospheric response to three decades of observed Arctic sea ice loss." *Journal of Climate*, 26: 1230–1248. 605

Sengupta, A. and Cressie, N. (2013). "Hierarchical statistical modeling of big spatial datasets using the exponential family of distributions." *Spatial Statistics*, 4: 14–44. MR3130833. 607, 610, 614

Sengupta, A., Cressie, N., Kahn, B. H., and Frey, R. (2016). "Predictive inference for big, spatial, non-Gaussian data: MODIS cloud data and its change-of-support." *Australian & New Zealand Journal of Statistics*, 58: 15–45. MR3499149. doi: https://doi.org/10.1111/anzs.12148. 607

Wikle, C. K., Milliff, R. F., Nychka, D., and Berliner, L. M. (2001). "Spatiotemporal hierarchical Bayesian modeling: Tropical ocean surface winds." *Journal of the American Statistical Association*, 96: 382–397. MR1939342. doi: https://doi.org/10.1198/016214501753168109. 610, 611

Wikle, C. K., Zammit-Mangion, A., and Cressie, N. (2019). *Spatio-Temporal Statistics with R*. Chapman & Hall/CRC, Boca Raton, FL, USA. 607, 614

Zammit-Mangion, A. and Cressie, N. (2020). "FRK: An R package for spatial and spatio-temporal prediction with large datasets." *Journal of Statistical Software*. doi: https://arxiv.org/abs/1705.08105, to appear. 607, 610

Zhang, B. and Cressie, N. (2019). "Estimating spatial changes over time of Arctic sea ice using hidden 2×2 tables." *Journal of Time Series Analysis*, 40: 288–311. MR3946154. doi: https://doi.org/10.1111/jtsa.12425. 607, 608, 610, 611, 612, 613, 614, 615, 617, 625

Zhang, B. and Cressie, N. (2020). "Bayesian spatio-temporal modeling of Arctic sea ice extent. Supplementary Material." *Bayesian Analysis*. doi: https://doi.org/10.1214/20-BA1209SUPP. 608

Zhang, H. (2004). "Inconsistent estimation and asymptotically equal interpolations in model-based geostatistics." *Journal of the American Statistical Association*, 99: 250–261. MR2054303. doi: https://doi.org/10.1198/016214504000000241. 614

Zwally, H. J., Comiso, J. C., Parkinson, C. L., Cavalieri, D. J., and Gloersen, P. (2002). "Variability of Antarctic sea ice 1979–1998." *Journal of Geophysical Research: Oceans*, 107: C5, 3041. 606

Acknowledgments

Cressie's research is supported by an Australian Research Council Discovery Project, number DP190100180. Zhang's research is supported by the Fundamental Research Funds for the Central Universities, Nankai University (63191406), and National Science Foundation China Project, number 11901316. It is also partially supported by the Key Laboratory of Pure Mathematics and Combinatorics (LPMC) of Ministry of Education, China and the Key Laboratory for Medical Data Analysis and Statistical Research (KLMDASR) of Tianjin, China. We would like to thank Bruno Sansó for encouraging us to write up our research into this article, and we are very grateful for the many comments and suggestions from the Associate Editor and four reviewers.