Inference for Spatio-Temporal Changes of Arctic Sea Ice

Noel Cressie
(ncressie@uow.edu.au)

National Institute for Applied Statistics Research Australia (NIASRA)

University of Wollongong (UOW), Australia

This research is joint with Bohai Zhang, NIASRA, UOW, and is based on the article: Zhang, B. and Cressie, N. (2017). Estimating spatial changes over time of Arctic sea ice using hidden 2×2 tables. *NIASRA Working Paper 08-17*.

- Parkinson (2014a) considered the sea ice extent of both the Arctic and the Antarctic regions for a 35-year period; by visualizing the time series of yearly (and monthly) areas of global sea ice, a decreasing trend of the global sea ice cover was observed.

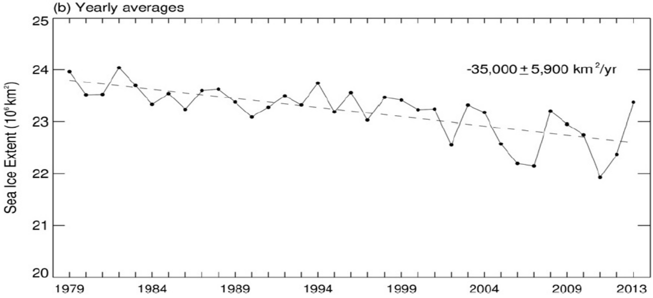


Figure: Yearly global sea ice extent (1979-2013), where the dashed line shows an ordinary-least-squares fit (from Figure 3, in Parkinson, 2014a).

Introduction, II

- Declining sea ice cover impacts the polar biogeochemical cycles (Meier et al., 2014) and can cause climate change in other regions (e.g., Mori et al., 2014; Cohen et al., 2014).
- Moreover, the albedo-ice feedback effect may lead to further retreat of the planet's ice cover (e.g., Screen et al., 2013; Pistone et al., 2014).
- An analysis of ranks of the monthly Arctic/Antarctic sea ice extents for different years can be found in Parkinson and DiGirolamo (2016).

Arctic sea ice extent has drawn considerable attention in recent years, due to the decreasing trend of ice cover in very high northern latitudes (e.g., Parkinson et al., 1999; Meier et al., 2007; Stroeve et al., 2007; Comiso et al., 2008; Parkinson, 2014a).

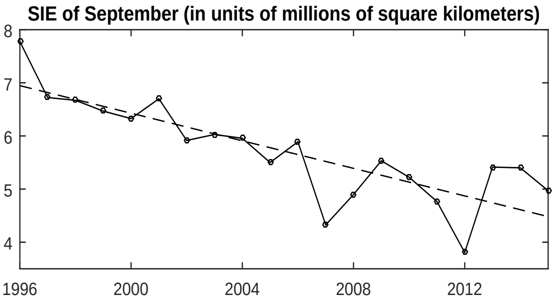


Figure: September Arctic sea ice extent for 1996 to 2015, where the dashed line shows an ordinary-least-squares fit (Zhang and Cressie, 2017).

NIASRA NATIONAL INSTITUTE FOR APPLIED STATISTICS RESEARCH AUSTRALIA

Parkinson (2014b) considered the length of the Arctic sea ice season (number of days for an area to be covered by sea ice) and created spatial maps that show the reduction of the Arctic sea ice cover.

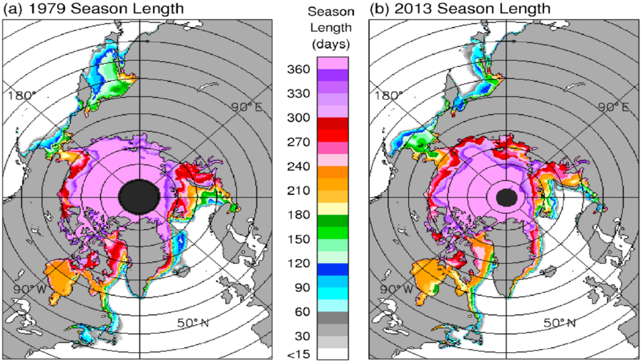


Figure: The length of the Arctic sea ice season for 1979 and 2013 (from Figure 1 in Parkinson, 2014b).

Spatio-temporal statistical models

- The geophysics literature has given results based on purely spatial or purely temporal data summaries, but generally lacking (proper) uncertainty measures. Hence, it is desirable to develop spatio-temporal statistical models for the Arctic sea ice data, from which defensible statistical inferences can be carried out.
- Descriptive spatio-temporal models: They describe spatio-temporal correlations through a valid spatio-temporal covariance function. Past and recent developments of descriptive spatio-temporal models mainly focus on Gaussian models for very large datasets (e.g., Higdon, 2002; Bevilacqua et al., 2012; Bai et al., 2012; Zhang et al., 2015; Datta et al., 2016; Zammit-Mangion and Cressie, 2017).
- Dynamic spatio-temporal models: They target the process' evolution, often discretized over time and using an autoregressive relationship (e.g., Wikle and Cressie, 1999; Wikle et al., 2001; Xu et al., 2005; Cressie et al., 2010; Cressie and Wikle, 2011; Katzfuss and Cressie, 2011; Finley et al., 2012). This is our focus here.

NIASRA NATIONAL INSTITUTE FOR APPLIED STATISTICS RESEARCH AUSTRALIA

- Arctic sea ice extent is defined as the total area of Arctic grid cells, each of whose sea ice concentration is greater than or equal to a cut-off value (say 0.15; e.g., Parkinson et al., 1999; Parkinson, 2014a).

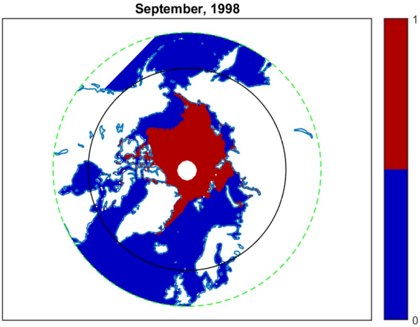


Figure: Binary sea ice cover data in September 1998; the region around the North Pole is not represented in the database (`http://nsidc.org/data/G02202`).

Spatial/spatio-temporal GLMs

- The data used to calculate sea ice extent are spatio-temporal and binary, equal to 1 if a grid cell is specified to be covered with ice, and equal to 0 otherwise.
- For spatial non-Gaussian observations, the exponential family of distributions and a spatial generalized linear model (GLM) within a hierarchical modeling framework (proposed by Diggle et al., 1998) is very flexible and has been applied to modeling large non-Gaussian spatial datasets, including empirical hierarchical models (EHM; e.g., Sengupta and Cressie, 2013; Sengupta et al., 2016; Shi and Kang, 2017), and Bayesian hierarchical models (BHM; e.g., Bradley et al., 2016; Guan and Haran, 2017; Bradley et al., 2017; Linero and Bradley, 2018, to name a few).
- Spatio-temporal GLMs in a BHM are considered by Holan and Wikle (2016); Bradley et al. (2018); Hu and Bradley (2018). This talk is about a spatio-temporal GLM in an EHM.

The computational challenge

NIASRA NATIONAL INSTITUTE FOR APPLIED STATISTICS RESEARCH AUSTRALIA

- When working in the purely spatial or the "descriptive" spatio-temporal contexts, computational challenges can be considerable, due to models that depend on large covariance matrices of spatial or spatio-temporal datasets.
- Fitting a spatial hierarchical GLM to very large spatial datasets has computational challenges, since it usually involves evaluating the likelihood of a latent high-dimensional Gaussian random vector.
- In this talk, we focus on a low-rank linear mixed effects model (Wikle et al., 2001; Cressie and Johannesson, 2006, 2008) to achieve dimension-reduction for the latent random effects in a spatio-temporal GLM, and then we model the evolution of the random effects with a multivariate dynamic model (e.g., Wikle et al., 2001; Cressie et al., 2010; Kang et al., 2010; Cressie and Wikle, 2011; Katzfuss and Cressie, 2011; Bradley et al., 2018; Hu and Bradley, 2018).
- Using a relatively small fixed number of basis functions makes computations feasible for very large spatio-temporal datasets, whether the hierarchical models are Bayesian (BHM) or empirical (EHM).

- Let $z_t(\mathbf{s})$ denote a binary ($1 \equiv$ ice; $0 \equiv$ no ice) spatio-temporal datum observed at a spatial location $\mathbf{s} \in \mathcal{D}$, where $\mathcal{D}$ is the spatial domain of interest, assumed here to be the same for all times $t \in \{1, 2, \dots, T\}$.
- As in Diggle et al. (1998), who considered spatial-only binary data, we model the spatio-temporal binary data as conditionally independent Bernoulli random variables, where the conditioning is on a latent process, $\{y_t(\mathbf{s}) : \mathbf{s} \in \mathcal{D}, t = 1, \dots, T\}$. That is, for $\mathbf{s} \in \mathcal{D}$,

$$z_t(\mathbf{s}) | y_t(\mathbf{s}) \overset{\text{ind.}}{\sim} \mathrm{Bernoulli}(p_t(\mathbf{s})), \tag{1}$$

where $y_t(\mathbf{s}) = g(p_t(\mathbf{s}))$, and $g(\cdot)$ is a link function.
- Here we choose the logit link, $g(p) = \log(p/(1-p))$, and hence

$$y_t(\mathbf{s}) = \log\left(\frac{p_t(\mathbf{s})}{1 - p_t(\mathbf{s})}\right) \text{ and } p_t(\mathbf{s}) = \frac{\exp(y_t(\mathbf{s}))}{1 + \exp(y_t(\mathbf{s}))}.$$

- The latent process $\{y_t(\mathbf{s})\}$ is further modeled as a spatio-temporal mixed effects model:

$$y_t(\mathbf{s}) = \mathbf{x}_t(\mathbf{s})'\boldsymbol{\beta}_t + \mathbf{S}_t(\mathbf{s})'\boldsymbol{\eta}_t + \xi_t(\mathbf{s}), \qquad (2)$$

where
  - $\mathbf{x}_t(\mathbf{s})$ is a p-dimensional covariate vector at location $\mathbf{s} \in \mathcal{D}$;
  - $\boldsymbol{\beta}_t$ is a p-dimensional vector of regression coefficients;
  - $\mathbf{S}_t(\mathbf{s}) \in \mathbb{R}^r$ is a basis-function vector evaluated at $\mathbf{s} \in \mathcal{D}$;
  - $\boldsymbol{\eta}_t$ is an r-dimensional mean-zero Gaussian random vector at time t;
  - $\{\xi_t(\cdot)\}$ is a Gaussian random process that is temporally independent with mean zero, and has only local or no spatial dependence that captures fine-scale variation. Here we make the white-noise assumption that $\mathrm{cov}(\xi_t(\mathbf{s}), \xi_u(\mathbf{s}')) = \sigma^2_{\xi,t} I(u = t; \mathbf{s}' = \mathbf{s})$, where $I(\cdot)$ is an indicator function;
  - $\{\boldsymbol{\eta}_t\}$ and $\{\xi_t(\cdot)\}$ are independent over both space and time.

- We use a lag-1 vector-autoregressive (VAR(1)) process to model the spatio-temporal random effects, $\{\boldsymbol{\eta}_t : t = 1, \ldots, T\}$ (e.g., Cressie and Wikle, 2011, Ch.7):

$$\boldsymbol{\eta}_1 \sim \mathcal{N}(\mathbf{0}, K), \;\; \boldsymbol{\eta}_t | \boldsymbol{\eta}_1, \ldots, \boldsymbol{\eta}_{t-1} \sim \mathcal{N}(H_t \boldsymbol{\eta}_{t-1}, U_t), \;\text{for } t = 2, \ldots, T,$$

  where we stack the random effects into $\boldsymbol{\eta} \equiv (\boldsymbol{\eta}_1', \ldots, \boldsymbol{\eta}_T')'$.
- Here, $\{H_t : t = 2, \ldots, T\}$ and $\{U_t : t = 2, \ldots, T\}$ are the $r \times r$ propagator and $r \times r$ innovation matrices at time t, respectively. The propagator matrix H_t captures the temporal cross-correlations of random effects between time points t and $t-1$.
- We shall treat K, H_t and U_t as unknown parameters to be estimated and assume that for the time period $t = 2, \ldots, T$, $H_t \equiv H$ and $U_t \equiv U$. This assumption can be weakened (as we do here) to the case where H_t and U_t are constant within shorter time periods (e.g., as discussed in Katzfuss and Cressie, 2011; Zhang and Cressie, 2017).

The data likelihood, I

- We first introduce some notation: Let $\mathcal{S}_t \equiv \{\mathbf{s}_{t,1}, \mathbf{s}_{t,2}, \ldots, \mathbf{s}_{t,N_t}\}$ be the observation locations at times $t = 1, \ldots, T$. For time t, let $\mathbf{Z}_t \equiv (z_t(\mathbf{s}_{t,1}), \ldots, z_t(\mathbf{s}_{t,N_t}))'$ be the observation vector; we stack all the space-time observations into $\mathbf{Z} \equiv (\mathbf{Z}_1', \ldots, \mathbf{Z'}_T)'$, which is a $\left(\sum_{t=1}^T N_t\right)$-dimensional vector.
- Let $\boldsymbol{\xi}_t \equiv (\xi_t(\mathbf{s}_{t,1}), \ldots, \xi_t(\mathbf{s}_{t,N_t}))'$ be the vector of the fine-scale-variation process evaluated at $\mathcal{S}_t$. Once again, we stack time-indexed vectors to yield $\boldsymbol{\xi} \equiv (\boldsymbol{\xi}_1', \ldots, \boldsymbol{\xi'}_T)'$.
- The likelihood $L(\boldsymbol{\theta}; \mathbf{Z})$ is the marginal probability,

$$\int_{\boldsymbol{\eta}} \int_{\boldsymbol{\xi}} p(\mathbf{Z}|\boldsymbol{\eta}, \boldsymbol{\xi}, \boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_T) \times p(\boldsymbol{\eta}|K, H, U) \times \prod_{t=1}^{T} p(\boldsymbol{\xi}_t|\sigma^2_{\xi,t}) \mathrm{d}\boldsymbol{\xi} \mathrm{d}\boldsymbol{\eta},$$

where $\boldsymbol{\theta} \equiv \{\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_T, \sigma^2_{\xi,1}, \ldots, \sigma^2_{\xi,T}, K, H, U\}$.

Specifically, $L(\boldsymbol{\theta}; \mathbf{Z})$

$$= \int_{\boldsymbol{\eta}} \int_{\boldsymbol{\xi}} \prod_{t=1}^{T} \prod_{i=1}^{N_t} (1 + \exp(-(2z_{t,i} - 1)y_{t,i}))^{-1} \times |K|^{-1/2} \exp(-\boldsymbol{\eta}_{\mathbf{1}}{}' K^{-1} \boldsymbol{\eta}_1 / 2)$$

$$\times (2\pi)^{-T \cdot r/2} |U|^{-(T-1)/2} \prod_{t=2}^{T} \exp(-(\boldsymbol{\eta}_t - H\boldsymbol{\eta}_{t-1})' U^{-1} (\boldsymbol{\eta}_t - H\boldsymbol{\eta}_{t-1})/2)$$

$$\times \prod_{t=1}^{T} (2\pi\sigma^2_{\xi,t})^{-N_t/2} \exp(-\boldsymbol{\xi}_t' \boldsymbol{\xi}_t / (2\sigma^2_{\xi,t})) \mathrm{d}\boldsymbol{\xi} \mathrm{d}\boldsymbol{\eta}, \qquad (3)$$

- In (3), we abbreviate the notation: $z_{t,i} \equiv z_t(\mathbf{s}_{t,i})$ and $y_{t,i} \equiv y_t(\mathbf{s}_{t,i})$, where recall that $y_t(\mathbf{s}_{t,i})$ is given by (2).
- The likelihood (3) does not have an analytical form, and hence we use the EM algorithm (Dempster et al., 1977), iterated to convergence, to obtain maximum likelihood estimates of the parameters $\boldsymbol{\theta}$.

The complete log-likelihood

- We treat the latent-random-effects vector $\boldsymbol{\eta}$ and the fine-scale-variation vector $\boldsymbol{\xi}$ as unobserved random variables.
- The complete log-likelihood, $\ell_c(\boldsymbol{\theta}) \equiv \log p(\mathbf{Z}, \boldsymbol{\eta}, \boldsymbol{\xi} | \boldsymbol{\theta})$, is:

$$
\begin{aligned}
&-\sum_{t=1}^{T}\sum_{i=1}^{N_t} \log(1+\exp(-(2z_{t,i}-1)y_{t,i})) - \frac{1}{2}\log|K| - \frac{1}{2}\boldsymbol{\eta}_1' K^{-1}\boldsymbol{\eta}_1 \\
&-\frac{1}{2}\sum_{t=2}^{T}(\boldsymbol{\eta}_t - H\boldsymbol{\eta}_{t-1})' U^{-1}(\boldsymbol{\eta}_t - H\boldsymbol{\eta}_{t-1}) - \frac{(T-1)}{2}\log|U| \\
&-\frac{N_t}{2}\sum_{t=1}^{T}\log\sigma_{\xi,t}^2 - \frac{1}{2}\sum_{t=1}^{T}\frac{\boldsymbol{\xi}_t'\boldsymbol{\xi}_t}{\sigma_{\xi,t}^2} + c_1,
\end{aligned} \tag{4}
$$

where c_1 is a constant that does not depend on $\boldsymbol{\theta}$.

NIASRA
NATIONAL INSTITUTE FOR APPLIED STATISTICS RESEARCH AUSTRALIA

- The expectation step (E-step) of the EM algorithm is with respect to $p(\boldsymbol{\eta}, \boldsymbol{\xi}|\mathbf{Z}, \boldsymbol{\theta})$. Suppose that we have completed the ℓ-th iteration, resulting in $\boldsymbol{\theta}^{(\ell)}$; then at the $(\ell+1)$-th iteration, the E-step is:

$$
\begin{aligned}
Q(\boldsymbol{\theta}; \boldsymbol{\theta}^{(\ell)}) = & -\sum_{t=1}^{T}\sum_{i=1}^{N_t} E(\log(1+\exp(-(2z_{t,i}-1)y_{t,i}))|\mathbf{Z}, \boldsymbol{\theta}^{(\ell)}) \\
& -\frac{1}{2}\mathrm{tr}\{E(\boldsymbol{\eta}_1\boldsymbol{\eta}_1'|\mathbf{Z}, \boldsymbol{\theta}^{(\ell)})K^{-1}\} - \frac{1}{2}\log|K| - \frac{(T-1)}{2}\log|U| \\
& -\frac{1}{2}\sum_{t=2}^{T}\mathrm{tr}\Big\{E(\boldsymbol{\eta}_t\boldsymbol{\eta}_t'|\mathbf{Z}, \boldsymbol{\theta}^{(\ell)})U^{-1} - E(\boldsymbol{\eta}_t\boldsymbol{\eta}_{t-1}'|\mathbf{Z}, \boldsymbol{\theta}^{(\ell)})H'U^{-1} \\
& -HE(\boldsymbol{\eta}_{t-1}\boldsymbol{\eta}_t'|\mathbf{Z}, \boldsymbol{\theta}^{(\ell)})U^{-1} + HE(\boldsymbol{\eta}_{t-1}\boldsymbol{\eta}_{t-1}'|\mathbf{Z}, \boldsymbol{\theta}^{(\ell)})H'U^{-1}\Big\} \\
& -\frac{1}{2}\sum_{t=1}^{T}\mathrm{tr}\{E(\boldsymbol{\xi}_t\boldsymbol{\xi}_t'|\mathbf{Z}, \boldsymbol{\theta}^{(\ell)})\}/\sigma^2_{\xi,t} - \frac{N_t}{2}\sum_{t=1}^{T}\log\sigma^2_{\xi,t} + c_2,
\end{aligned}
$$

where c_2 does not depend on $\boldsymbol{\theta}$.

NIASRA NATIONAL INSTITUTE FOR APPLIED STATISTICS RESEARCH AUSTRALIA

- Problem: $p(\boldsymbol{\eta}, \boldsymbol{\xi} | \mathbf{Z}, \boldsymbol{\theta})$ cannot be expressed in closed form.
- We resolve this by using a Laplace approximation (e.g., Sengupta and Cressie, 2013; Sengupta et al., 2016), which requires the posterior mode of $p(\boldsymbol{\eta}, \boldsymbol{\xi} | \mathbf{Z}, \boldsymbol{\theta})$.
- Suppose at the ℓ-th iteration, we obtain $\boldsymbol{\theta}^{(\ell)}$; then the maximum of $p(\boldsymbol{\eta}, \boldsymbol{\xi} | \mathbf{Z}, \boldsymbol{\theta}^{(\ell)})$ can be obtained equivalently by maximizing the complete likelihood $p(\mathbf{Z}, \boldsymbol{\eta}, \boldsymbol{\xi} | \boldsymbol{\theta}^{(\ell)})$, with respect to $\boldsymbol{\eta}$ and $\boldsymbol{\xi}$.
- The Laplace approximation replaces the posterior distribution, $p(\boldsymbol{\eta}, \boldsymbol{\xi} | \mathbf{Z}, \boldsymbol{\theta}^{(\ell)})$, with a multivariate Gaussian distribution whose mean is given by the posterior mode of $(\boldsymbol{\eta}, \boldsymbol{\xi})$ and whose covariance matrix is given by the inverse of the negative Hessian matrix (He) of the posterior distribution evaluated at the mode.

- By introducing some notation to form the quadratic term in $\boldsymbol{\eta}$, we can obtain closed-form expressions for the first-order and second-order derivatives of $p(\mathbf{Z}, \boldsymbol{\eta}, \boldsymbol{\xi} | \boldsymbol{\theta}^{(\ell)})$ with respect to $\boldsymbol{\eta}$ and $\boldsymbol{\xi}$.
- The posterior mode of $(\boldsymbol{\eta}, \boldsymbol{\xi})$ can be obtained iteratively by running the Fisher-scoring algorithm (e.g., Jennrich and Sampson, 1976) until convergence.
- Let $\mathrm{He}(\boldsymbol{\eta}, \boldsymbol{\xi})$ denote the Hessian matrix for $\ell_c \equiv \log p(\mathbf{Z}, \boldsymbol{\eta}, \boldsymbol{\xi} | \boldsymbol{\theta}^{(\ell)})$; hence,

$$\mathrm{He}(\boldsymbol{\eta}, \boldsymbol{\xi}) = \begin{pmatrix} \frac{\partial^2 \ell_c}{\partial \boldsymbol{\eta} \boldsymbol{\eta}'} & \frac{\partial^2 \ell_c}{\partial \boldsymbol{\eta} \partial \boldsymbol{\xi}'} \\ \frac{\partial^2 \ell_c}{\partial \boldsymbol{\xi} \partial \boldsymbol{\eta}'} & \frac{\partial^2 \ell_c}{\partial \boldsymbol{\xi} \boldsymbol{\xi}'} \end{pmatrix}.$$

- The conditional posterior (co)variances of $(\boldsymbol{\eta}', \boldsymbol{\xi}')'$ can be approximated by $-\mathrm{He}(\hat{\boldsymbol{\eta}}, \hat{\boldsymbol{\xi}})^{-1}$, where $(\hat{\boldsymbol{\eta}}, \hat{\boldsymbol{\xi}})$ is the posterior mode of $p(\boldsymbol{\eta}, \boldsymbol{\xi} | \mathbf{Z}, \boldsymbol{\theta}^{(\ell)})$.

The maximization step (M-step)

- The maximization step (M-step) yields $\boldsymbol{\theta}^{(\ell+1)} = \arg\max_{\boldsymbol{\theta}} Q(\boldsymbol{\theta}; \boldsymbol{\theta}^{(\ell)})$.
- The closed-form solutions of K, $\{\sigma^2_{\xi,t}\}$, H, and U at this step are:

$$K^{(\ell+1)} = E(\boldsymbol{\eta}_1 \boldsymbol{\eta}_1' | \mathbf{Z}, \boldsymbol{\theta}^{(\ell)}),$$

$$(\sigma^2_{\xi,t})^{(\ell+1)} = \frac{1}{N} \mathrm{tr}(E(\boldsymbol{\xi}_t \boldsymbol{\xi}_t' | \mathbf{Z}, \boldsymbol{\theta}^{(\ell)})),$$

$$H^{(\ell+1)} = \left(\sum_{t=2}^{T} E(\boldsymbol{\eta}_t \boldsymbol{\eta}_{t-1}' | \mathbf{Z}, \boldsymbol{\theta}^{(\ell)}) \right) \left(\sum_{t=2}^{T} E(\boldsymbol{\eta}_{t-1} \boldsymbol{\eta}_{t-1}' | \mathbf{Z}, \boldsymbol{\theta}^{(\ell)}) \right)^{-1},$$

$$U^{(\ell+1)} = \frac{1}{T-1} \sum_{t=2}^{T} E\left((\boldsymbol{\eta}_t - H^{(\ell+1)} \boldsymbol{\eta}_{t-1})(\boldsymbol{\eta}_t - H^{(\ell+1)} \boldsymbol{\eta}_{t-1})' | \mathbf{Z}, \boldsymbol{\theta}^{(\ell)} \right).$$

- The regression coefficients, $\{\boldsymbol{\beta}_t,\ t = 1, \ldots, T\}$, cannot be expressed in closed form, but they can be estimated using a one-step Newton-Raphson update within the EM algorithm (Sengupta and Cressie, 2013).

- First iterate to convergence to obtain the EM estimates of the model parameters, denoted by $\hat{\boldsymbol{\theta}}$. Then substitute them into the predictive distribution to yield the empirical predictive distribution, $p(\boldsymbol{\eta}, \boldsymbol{\xi} | \mathbf{Z}, \hat{\boldsymbol{\theta}})$.
- Our approach is then to simulate from $p(\boldsymbol{\eta}, \boldsymbol{\xi} | \mathbf{Z}, \hat{\boldsymbol{\theta}})$ using Markov chain Monte Carlo (MCMC), which in turn yields a predictive distribution of $\{y_t(\mathbf{s}) : \mathbf{s} \in \mathcal{D}, t = 1, \ldots, T\}$; see below.
- MCMC: Since the full conditional distributions, $p(\boldsymbol{\eta} | \boldsymbol{\xi}, \mathbf{Z}, \hat{\boldsymbol{\theta}})$ and $p(\boldsymbol{\xi} | \boldsymbol{\eta}, \mathbf{Z}, \hat{\boldsymbol{\theta}})$, do not have a closed form, the "Metropolis-Hastings within Gibbs sampler" MCMC (e.g., Gelfand and Smith, 1990; Gelman et al., 2014) is used to obtain predictive samples from $(\boldsymbol{\eta}, \boldsymbol{\xi})$.
- Predictive samples of $\{y_t(\mathbf{s})\}$ follow from (2), which is:

$$y_t(\mathbf{s}) = \mathbf{x}_t(\mathbf{s})'\boldsymbol{\beta}_t + \mathbf{S}_t(\mathbf{s})'\boldsymbol{\eta}_t + \xi_t(\mathbf{s}).$$

- Temporal homogeneity during $\{1, \ldots, T\}$ implies that the basis-function vector $\mathbf{S}(\cdot) \equiv (S_1(\cdot), \ldots, S_r(\cdot))'$ does not depend on t.
- Here we focus on the compactly supported multi-resolution bisquare functions: For $j = 1, \ldots, r$, define

$$S_j(\mathbf{s}) \equiv \left(1 - \left(\frac{\|\mathbf{s} - \mathbf{c}_j\|}{\phi_j}\right)^2\right)^2 I(\|\mathbf{s} - \mathbf{c}_j\| < \phi_j); \ \mathbf{s} \in \mathbb{R}^d, \tag{5}$$

  where $\mathbf{c}_j$ is the center of the j-th basis function $S_j(\cdot)$, ϕ_j is the radius of its spatial support, and $I(\cdot)$ is an indicator function.
- The choice of $\{\phi_j\}$ determines the multiple resolutions, which are used to capture different dependence scales (e.g., Cressie and Johannesson, 2008; Nychka et al., 2015; Katzfuss, 2017); some basis functions with centers outside the study domain are included to accommodate boundary effects (Cressie and Kang, 2010).

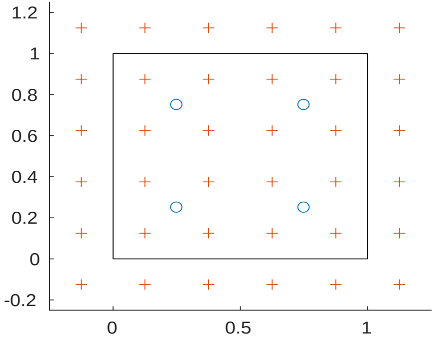


Figure: A simple example showing centers of bisquare basis functions, where circles and pluses are for Resolution-1 and Resolution-2 basis functions, respectively. The spatial domain of interest $\mathcal{D} = [0, 1] \times [0, 1]$ is outlined.

- We further parameterize the propagator matrix H to model dependence for both within-resolution basis functions and between-resolution basis functions.
- For basis functions with r_1 Resolution-1 basis functions and r_2 Resolution-2 basis functions, we specify H as follows:

$$H \equiv \begin{pmatrix} \rho_1 I_{r_1} & 0 \\ \rho_3 R & \rho_2 I_{r_2} \end{pmatrix} \equiv \rho_1 \begin{pmatrix} I_{r_1} & 0 \\ \tilde{\rho}_3 R & \tilde{\rho}_2 I_{r_2} \end{pmatrix}, \tag{6}$$

where ρ_1 and $\rho_2 \equiv \rho_1 \tilde{\rho}_2 \in (0, 1)$ model the within-resolution autocorrelations for Resolution-1 and Resolution-2 basis functions, respectively; $\rho_3 \equiv \rho_1 \tilde{\rho}_3 \in (0, 1)$ models the between-resolution autocorrelations of basis functions multiplied by an $r_2 \times r_1$ matrix R; and R is sparse with non-zero entries equal to 1 if a finer-resolution basis function is a (spatial) neighbor of a coarser-resolution basis function.

Application to Arctic sea ice cover data

- Recall that the Arctic sea ice extent (SIE) is obtained as the sum of the areas of grid cells whose sea ice concentration is greater than or equal to 15% (e.g., Parkinson et al., 1999; Zwally et al., 2002; Meier et al., 2007; Parkinson, 2014a).
- Arctic sea ice cover datasets come from remote sensing of Arctic sea ice concentrations, which are areal proportions of sea ice over spatial grid cells in the Arctic.
- Here we considered the National Oceanic and Atmospheric Administration (NOAA)/National Snow & Ice Data Center's (NSIDC) Climate Data Record (CDR) of passive microwave sea ice concentrations (e.g., Peng et al., 2013; Meier et al., 2017).
- There are $136,192$ observations (stored as a 304×448 matrix) for each daily or monthly dataset with the possibility of missing values (e.g., around the North Pole), and each spatial grid cell has a nominal area of $25\text{km} \times 25\text{km}$.

Data description, I

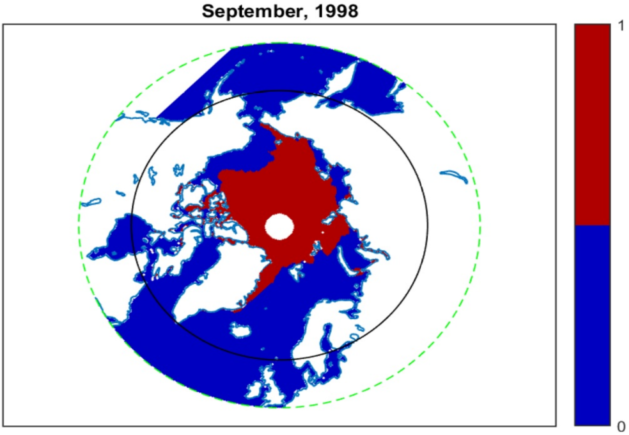
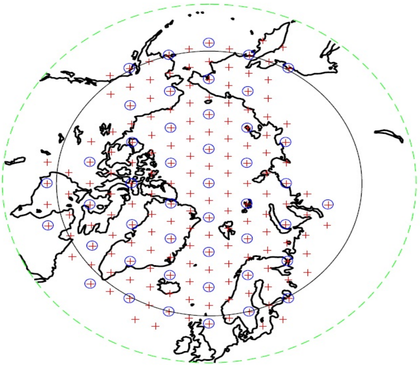


Figure: Binary sea ice cover data in September 1998; the region around the North Pole has no data. (The same figure was shown earlier.)

NIASRA NATIONAL INSTITUTE FOR APPLIED STATISTICS RESEARCH AUSTRALIA

- We focus on the spatio-temporal binary data used to obtain the SIE for the month of September over the 20 years from 1996 to 2015, inclusive. The month of September typically has the minimum Arctic sea ice extent for the year (e.g., Parkinson, 2014a).
- Our study domain $\mathcal{D}$ is defined by locations with latitudes greater than or equal to 60°N, which covers the Arctic region, ranging from the south end of Greenland to the North Pole.
- The spatial locations of the September data are the same for different years, resulting in a spatial binary dataset of $26,342$ observations for each of the 20 years, for latitudes $\geq 60^\circ$N.
- We split the 20 years into four time periods: $1996 - 2001$, $2001 - 2006$, $2006 - 2011$, and $2011 - 2015$; then we applied the proposed spatio-temporal model to data in each of these four time periods, assuming that $H_t \equiv H$ and $U_t \equiv U$ in a given period, but allowing them to be different from one period to the next.

- We used the multi-resolution bisquare basis functions given in (5), where great-circle distance replaced Euclidean distance. We used two resolutions with $r_1 = 45$ Resolution-1 basis functions (centers are shown as a blue circle) and $r_2 = 172$ Resolution-2 basis functions (centers are shown as a red plus).



EM-estimation results, III

Table: EM estimates of the propagator-matrix parameters.

Period	ρ_1	ρ_2	ρ_3
1	0.53	0.40	0.06
2	0.36	0.43	0.03
3	0.59	0.48	0 (fixed)
4	0.48	0.52	0.05

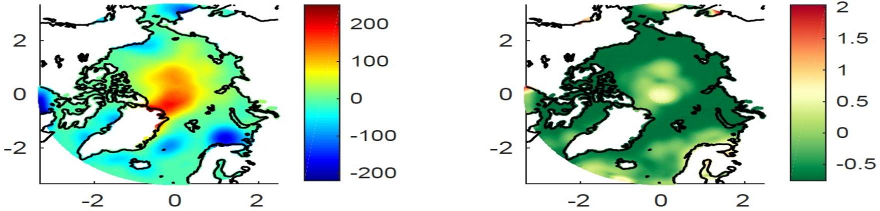
We conclude that there is some variability in the within-resolution correlations from period to period, and that the between-resolution correlations of basis functions are negligible.

The empirical predictive distributions

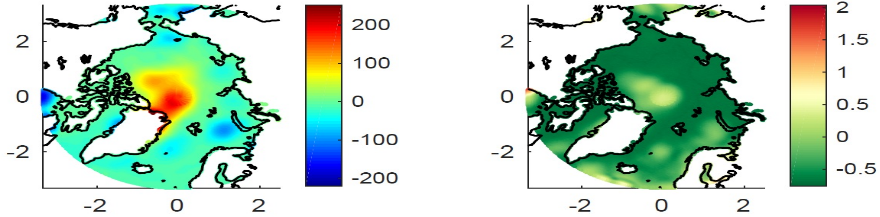
- After obtaining the EM estimates of model parameters and MCMC samples of $(\boldsymbol{\eta}, \boldsymbol{\xi})$, we can readily infer the empirical predictive distribution of $\{y_t(\mathbf{s})\}$ from (2) (and equivalently of $\{p_t(\mathbf{s})\}$ using the expit transformation, $p_t(\mathbf{s}) = \text{expit}(y_t(\mathbf{s})) \equiv e^{y_t(\mathbf{s})}/(1 + e^{y_t(\mathbf{s})})$).

- Recall that $p_t(\mathbf{s})$ is the probability that spatial pixel $\mathbf{s}$ is ice at time t, and $y_t(\mathbf{s})$ is that probability on the logit scale.

- Suppose that we have generated samples, $\{y_t^{(\ell)}(\mathbf{s}) : \mathbf{s} \in \mathcal{D}, t = 1, \ldots, T\}$, for $\ell = 1, \ldots, L$, from the predictive distribution of $\{y_t(\cdot)\}$, using an MCMC algorithm. Then based on those MCMC samples, the empirical predictive means and empirical predictive standard errors of $\{y_t(\mathbf{s})\}$ (and $\{p_t(\mathbf{s})\}$) can be readily obtained.



(a) 1998 (Period 1): $E(y_t(\mathbf{s})|\mathbf{Z}, \hat{\boldsymbol{\theta}})$ and $(\mathrm{var}(y_t(\mathbf{s})|\mathbf{Z}, \hat{\boldsymbol{\theta}}))^{\frac{1}{2}}$ (on the log scale).



(b) 2011 (Period 3): $E(y_t(\mathbf{s})|\mathbf{Z}, \hat{\boldsymbol{\theta}})$ and $(\mathrm{var}(y_t(\mathbf{s})|\mathbf{Z}, \hat{\boldsymbol{\theta}}))^{\frac{1}{2}}$ (on the log scale).

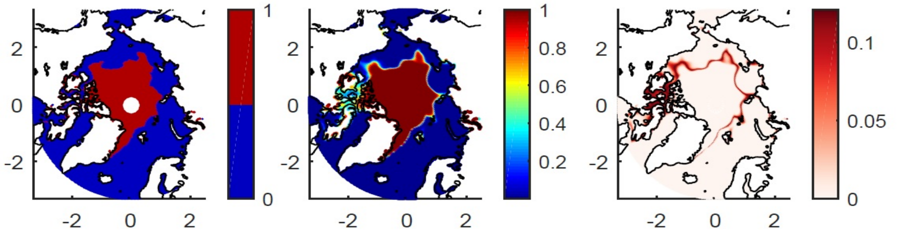


Figure: Plots of the data and the predictive distribution of $\{p_t(\mathbf{s})\}$ for year 2011. From left to right, the sea ice cover data, the predictive mean of $p_{2011}(\cdot)$, and the predictive standard error of $p_{2011}(\cdot)$.

- The latent process, $p_t(\mathbf{s}) = \exp(y_t(\mathbf{s}))/(1 + \exp(y_t(\mathbf{s})))$, contracts the scale of spatial variability into an almost dichotomous spatial process.
- Prediction uncertainties are particularly large for spatial locations around the boundaries of the Arctic sea ice cover.

Posterior summaries based on $y_t(\mathbf{s})$

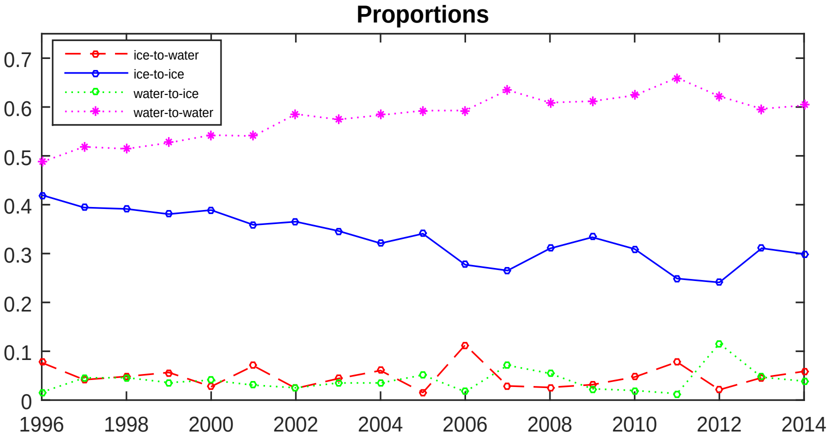
NIASRA
NATIONAL INSTITUTE FOR APPLIED STATISTICS RESEARCH AUSTRALIA

- Recall that $g(\cdot)$ is the logit function and 0.15 is an often-used sea ice concentration cut-off value used to classify whether or not a spatial grid cell is covered by ice. The following table can be used to infer the joint ice-to-water, ice-to-ice, water-to-ice, and water-to-water probabilities:

Table: Joint probabilities at t and $t+1$ for fixed spatial locations.

t \ t+1	Water	Ice
Ice	$\pi_{11}(\mathbf{s}; t, t+1) \equiv$ $Pr(y_t(\mathbf{s}) \geq g(0.15), y_{t+1}(\mathbf{s}) < g(0.15) \vert \mathbf{Z}, \boldsymbol{\theta})$	$\pi_{12}(\mathbf{s}; t, t+1) \equiv$ $Pr(y_t(\mathbf{s}) \geq g(0.15), y_{t+1}(\mathbf{s}) \geq g(0.15) \vert \mathbf{Z}, \boldsymbol{\theta})$
Water	$\pi_{21}(\mathbf{s}; t, t+1) \equiv$ $Pr(y_t(\mathbf{s}) < g(0.15), y_{t+1}(\mathbf{s}) < g(0.15) \vert \mathbf{Z}, \boldsymbol{\theta})$	$\pi_{22}(\mathbf{s}; t, t+1) \equiv$ $Pr(y_t(\mathbf{s}) < g(0.15), y_{t+1}(\mathbf{s}) \geq g(0.15) \vert \mathbf{Z}, \boldsymbol{\theta})$

- For each t, consider averaging the joint probabilities in the ice-to-water joint-probability table, over all the spatial locations in $\mathcal{D}$.
- This is equal to the predictive mean of the proportion of grid cells in the ice-to-water state.



Ice-to-water joint probabilities

NIASRA
NATIONAL INSTITUTE FOR APPLIED STATISTICS RESEARCH AUSTRALIA

- Recall that $\{y_t^{(\ell)}(\mathbf{s}) : \mathbf{s} \in \mathcal{D}, t = 1, \ldots, T\}$, for $\ell = 1, \ldots, L$ are samples from the empirical predictive distribution of $y_t(\mathbf{s})$, using an MCMC algorithm.
- Let $h(\{y_t(\cdot)\})$ be a functional to be predicted. Then its predictive mean, $E(h(\{y_t(\cdot)\})|\mathbf{Z}, \hat{\boldsymbol{\theta}})$, can be obtained empirically, through averaging the samples $\{h(\{y_t^{(\ell)}(\cdot)\}) : \ \ell = 1, \ldots, L\}$. That is, the predictive mean is approximately $\frac{1}{L}\sum_{\ell=1}^{L} h(\{y_t^{(\ell)}(\cdot)\})$.
- Put $h(\{y_t(\cdot)\}) = I(y_t(\mathbf{s}) \geq g(0.15), y_{t+1}(\mathbf{s}) < g(0.15))$; then from the MCMC samples, π_{11} can be obtained (up to MCMC error) by

$$\pi_{11}(\mathbf{s}; t, t+1) = \frac{1}{L}\sum_{\ell=1}^{L} I(y_t^{(\ell)}(\mathbf{s}) \geq g(0.15)) I(y_{t+1}^{(\ell)}(\mathbf{s}) < g(0.15));$$

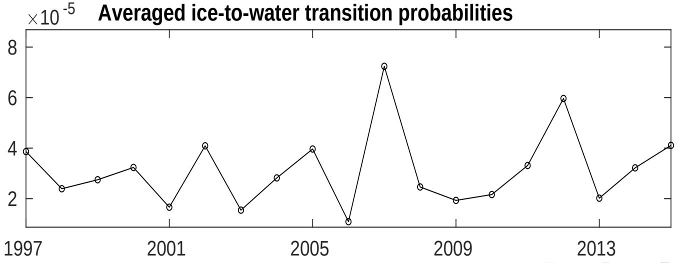
  other predictive probabilities can be estimated analogously.
- Each pixel $\mathbf{s}$ has a time series of predictive π_{11}'s, which we averaged over $\mathbf{s}$ on the previous slide. This is different from the time series of the empirical proportions of ice-to-water pixels.

The risk of transitioning to water from ice

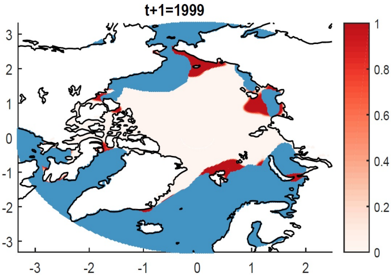
- The ice-to-water transition (IWT) probability at pixel $\mathbf{s}$ and time $t+1$ is the conditional probability, $\pi_{t+1|t}(\mathbf{s}) \equiv \pi_{11}(\mathbf{s}; t, t+1)/\pi_{1\cdot}(\mathbf{s}; t)$, given pixel $\mathbf{s}$ is ice. Here, $\pi_{1\cdot}(\mathbf{s}; t) \equiv \pi_{11}(\mathbf{s}; t, t+1) + \pi_{12}(\mathbf{s}; t, t+1) > 0$.
- Now take a weighted average to obtain the quantity, $\mathrm{IWT}_{t+1} \equiv$

$$\frac{\sum_i \pi_{t+1|t}(\mathbf{s}_i)\pi_{1\cdot}(\mathbf{s}_i; t)}{\sum_i \pi_{1\cdot}(\mathbf{s}_i; t)} = \frac{\sum_i E(I(y_t(\mathbf{s}_i) \geq g(0.15), y_{t+1}(\mathbf{s}_i) < g(0.15))|\mathbf{Z}, \hat{\boldsymbol{\theta}})}{\sum_i E(I(y_t(\mathbf{s}_i) \geq g(0.15))|\mathbf{Z}, \hat{\boldsymbol{\theta}})},$$

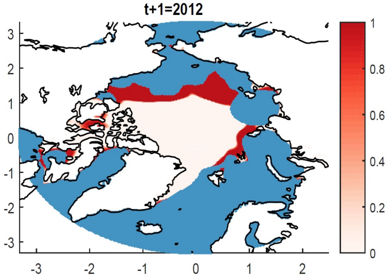
which can be obtained from MCMC samples.



Map of ice-to-water transition probabilities



(a) $t = 1998$; $t + 1 = 1999$



(b) $t = 2011$; $t + 1 = 2012$

Figure: The conditional probability map of $\pi_{t+1|t}(\mathbf{s})$ for $t + 1 = 1999$ and 2012. The blue color indicates the water regions with $\pi_{1\cdot}(\mathbf{s}; t) = 0$. High-risk areas are indicated by a darker red. A more complete picture is obtained through a time series of these maps, which we give as an animation.

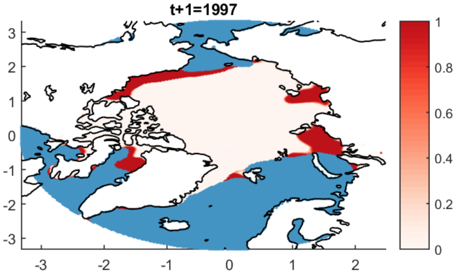


Figure: The conditional probability map of $\pi_{t+1|t}(\mathbf{s})$, from $t+1 = 1997$ to $t+1 = 2015$. The blue color indicates the water regions with $\pi_{1\cdot}(\mathbf{s}; t) = 0$. High-risk areas are indicated by a darker red.

Conclusions

- We have proposed a hierarchical spatio-temporal generalized linear model for analyzing binary sea ice cover datasets over time.
- The spatio-temporal dependencies are modeled by a latent spatio-temporal linear mixed-effects model, which achieves both dimension-reduction for computational efficiency and a flexible nonstationary spatial field at different time points.
- We “smoothed" but did not “forecast" here: Based on the predictive samples of $\{y_t(\cdot)\}$, several summaries are given that provide different perspectives on the changes over time of Arctic sea ice cover.
- In particular, we considered a latent 2×2 table based on the joint empirical predictive distribution of $y_t(\mathbf{s})$ and $y_{t+1}(\mathbf{s})$ at two consecutive time points, from which we visualized changes in the ice-to-water state across years.
- Knowing where changes occur over time is critical to understanding changes in polar biogeochemical cycles and albedo-ice feedback.

THE R SERIES

SPATIO-TEMPORAL STATISTICS WITH R

CHRISTOPHER K. WIKLE
ANDREW ZAMMIT-MANGION
NOEL CRESSIE

CRC Press
Taylor & Francis Group
A CHAPMAN & HALL BOOK

NIASRA
NATIONAL INSTITUTE FOR APPLIED STATISTICS RESEARCH AUSTRALIA

Bai, Y., P. Song, and T. E. Raghunathan (2012). Joint composite estimating functions in spatiotemporal models. *Journal of the Royal Statistical Society: Series B 74*, 799–824.

Bevilacqua, M., C. Gaetan, J. Mateu, et al. (2012). Estimating space and space-time covariance functions for large data sets: A weighted composite likelihood approach. *Journal of the American Statistical Association 107*, 268–280.

Bradley, J. R., S. H. Holan, and C. K. Wikle (2017). Bayesian hierarchical models with conjugate full-conditional distributions for dependent data from the natural exponential family. arXiv preprint arXiv:1701.07506.

Bradley, J. R., S. H. Holan, and C. K. Wikle (2018). Computationally efficient distribution theory for Bayesian inference of high-dimensional dependent count-valued data (with discussion). *Bayesian Analysis*. DOI: 10.1214/17-BA1069.

Bradley, J. R., C. K. Wikle, and S. H. Holan (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.

Cohen, J., J. A. Screen, J. C. Furtado, et al. (2014). Recent Arctic amplification and extreme mid-latitude weather. *Nature Geoscience 7*, 627–637.

Comiso, J. C., C. L. Parkinson, R. Gersten, et al. (2008). Accelerated decline in the Arctic sea ice cover. *Geophysical Research Letters 35*, L01703.

NIASRA
NATIONAL INSTITUTE FOR APPLIED STATISTICS RESEARCH AUSTRALIA

Cressie, N. and G. Johannesson (2006). Spatial prediction for massive datasets. In *Proc. Australian Academy of Science Elizabeth and Frederick White Conf*, pp. 1–11. Canberra: Australian Academy of Science.

Cressie, N. and G. Johannesson (2008). Fixed rank kriging for very large spatial data sets. *Journal of the Royal Statistical Society: Series B 70*, 209–226.

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

Cressie, N., T. Shi, and E. L. Kang (2010). Fixed rank filtering for spatio-temporal data. *Journal of Computational and Graphical Statistics 19*, 724–745.

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

Datta, A., S. Banerjee, A. O. Finley, et al. (2016). Nonseparable dynamic nearest neighbor Gaussian process models for large spatio-temporal data with an application to particulate matter analysis. *Annals of Applied Statistics 10*, 1286–1316.

Dempster, A. P., N. M. Laird, and D. B. Rubin (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society: Series B 39*, 1–38.

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

NIASRA
NATIONAL INSTITUTE FOR APPLIED STATISTICS RESEARCH AUSTRALIA

Finley, A. O., S. Banerjee, and A. E. Gelfand (2012). Bayesian dynamic modeling for large space-time datasets using Gaussian predictive processes. *Journal of Geographical Systems 14*, 29–47.

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

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

Guan, Y. and M. Haran (2017). A computationally efficient projection-based approach for spatial generalized linear mixed models. Journal of Computational and Graphical Statistics, forthcoming. arXiv preprint: arXiv:1609.02501.

Higdon, D. (2002). Space and space-time modeling using process convolutions. In C. W. Anderson, V. Barnett, P. C. Chatwin, and A. H. El-Shaarawi (Eds.), *Quantitative Methods for Current Environmental Issues*, pp. 37–56. Springer, London, UK.

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

Hu, G. and J. Bradley (2018). A Bayesian spatio-temporal model for analyzing earthquake magnitudes. STAT, forthcoming.

References IV

NIASRA
NATIONAL INSTITUTE FOR APPLIED STATISTICS RESEARCH AUSTRALIA

Jennrich, R. I. and P. Sampson (1976). Newton-Raphson and related algorithms for maximum likelihood variance component estimation. *Technometrics 18*, 11–17.

Kang, E. L., N. Cressie, and T. Shi (2010). Using temporal variability to improve spatial mapping with application to satellite data. *Canadian Journal of Statistics 38*, 271–289.

Katzfuss, M. (2017). A multi-resolution approximation for massive spatial datasets. *Journal of the American Statistical Association 112*, 201–214.

Katzfuss, M. and N. Cressie (2011). Spatio-temporal smoothing and EM estimation for massive remote-sensing data sets. *Journal of Time Series Analysis 32*, 430–446.

Linero, A. R. and J. R. Bradley (2018). Multi-rubric models for ordinal spatial data with application to online ratings from Yelp. Annals of Applied Statistics, forthcoming.

Meier, W., F. Fetterer, M. Savoie, et al. (2017). NOAA/NSIDC Climate Data Record of Passive Microwave Sea Ice Concentration, Version 3. [09/1996-09/2015]. Boulder, Colorado USA. NSIDC: National Snow and Ice Data Center. doi: `http://dx.doi.org/10.7265/N59P2ZTG` [Accessed: 09 October 2017].

Meier, W. N., G. K. Hovelsrud, B. E. Oort, 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.

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

Mori, M., M. Watanabe, H. Shiogama, et al. (2014). Robust Arctic sea-ice influence on the frequent Eurasian cold winters in past decades. *Nature Geoscience 7*, 869–873.

Nychka, D., S. Bandyopadhyay, D. Hammerling, et al. (2015). A multiresolution Gaussian process model for the analysis of large spatial datasets. *Journal of Computational and Graphical Statistics 24*, 579–599.

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

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

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

Parkinson, C. L. and N. E. DiGirolamo (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.

Peng, G., W. Meier, D. Scott, et al. (2013). A long-term and reproducible passive microwave sea ice concentration data record for climate studies and monitoring. *Earth System Science Data 5*, 311–318.

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

Screen, J. A., I. Simmonds, C. Deser, et al. (2013). The atmospheric response to three decades of observed Arctic sea ice loss. *Journal of Climate 26*, 1230–1248.

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

Sengupta, A., N. Cressie, B. H. Kahn, et al. (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.

Shi, H. and E. L. Kang (2017). Spatial data fusion for large non-Gaussian remote sensing datasets. *Stat 6*, 390–404.

Stroeve, J., M. M. Holland, W. Meier, et al. (2007). Arctic sea ice decline: Faster than forecast. *Geophysical Research Letters 34*, L09501, doi: 10.1029/2007GL029703.

Wikle, C. K. and N. Cressie (1999). A dimension-reduced approach to space-time Kalman filtering. *Biometrika 86*, 815–829.

Wikle, C. K., R. F. Milliff, D. Nychka, et al. (2001). Spatiotemporal hierarchical Bayesian modeling: Tropical ocean surface winds. *Journal of the American Statistical Association 96*, 382–397.

Xu, K., C. K. Wikle, and N. I. Fox (2005). A kernel-based spatio-temporal dynamical model for nowcasting weather radar reflectivities. *Journal of the American Statistical Association 100*, 1133–1144.

References VII

Zammit-Mangion, A. and N. Cressie (2017). FRK: An R package for spatial and spatio-temporal prediction with large datasets. arXiv preprint: arXiv:1705.08105.

Zhang, B. and N. Cressie (2017). Estimating spatial changes over time of Arctic sea ice using hidden 2×2 tables. https://niasra.uow.edu.au/content/groups/public/@web/@inf/@math/documents/mm/uow240617.pdf.

Zhang, B., H. Sang, and J. Z. Huang (2015). Full-scale approximations of spatio-temporal covariance models for large datasets. *Statistica Sinica*, 99–114.

Zwally, H. J., J. C. Comiso, C. L. Parkinson, et al. (2002). Variability of Antarctic sea ice 1979–1998. *Journal of Geophysical Research: Oceans 107*, C5, 3041.