THE AMERICAN STATISTICIAN
2021, VOL. 75, NO. 2, 185–194: Statistical Computing and Graphics
https://doi.org/10.1080/00031305.2019.1665584

Taylor & Francis
Taylor & Francis Group

STATISTICAL COMPUTING AND GRAPHICS

Making Recursive Bayesian Inference Accessible

Mevin B. Hooten[a,b,c], Devin S. Johnson[d], and Brian M. Brost[d]

[a]Colorado Cooperative Fish and Wildlife Research Unit, U.S. Geological Survey, Fort Collins, CO; [b]Department of Fish, Wildlife, and Conservation Biology, Colorado State University, Fort Collins, CO; [c]Department of Statistics, Colorado State University, Fort Collins, CO; [d]Marine Mammal Laboratory, Alaska Fisheries Science Center, NOAA Fisheries, Seattle, WA

ABSTRACT

Bayesian models provide recursive inference naturally because they can formally reconcile new data and existing scientific information. However, popular use of Bayesian methods often avoids priors that are based on exact posterior distributions resulting from former studies. Two existing Recursive Bayesian methods are: Prior- and Proposal-Recursive Bayes. Prior-Recursive Bayes uses Bayesian updating, fitting models to partitions of data sequentially, and provides a way to accommodate new data as they become available using the posterior from the previous stage as the prior in the new stage based on the latest data. Proposal-Recursive Bayes is intended for use with hierarchical Bayesian models and uses a set of transient priors in first stage independent analyses of the data partitions. The second stage of Proposal-Recursive Bayes uses the posteriors from the first stage as proposals in a Markov chain Monte Carlo algorithm to fit the full model. We combine Prior- and Proposal-Recursive concepts to fit any Bayesian model, and often with computational improvements. We demonstrate our method with two case studies. Our approach has implications for big data, streaming data, and optimal adaptive design situations.

ARTICLE HISTORY
Received September 2018
Accepted August 2019

KEYWORDS
Filtering; Hierarchical model; Iterative forecasting; Markov chain Monte Carlo; Parallel processing; Sequential inference

1. Introduction

Bayesian methods have been incredibly useful for scientific inquiry because they empower the user to customize statistical analyses for their data and desired inference as well as formally incorporate existing scientific information (Gelman et al. 2013). In particular, Bayesian hierarchical models (BHMs) also allow us to consider a complicated joint problem as a sequence of simpler conditional components. In his seminal article on BHMs, Berliner (1996) described the hierarchical model structure heuristically in terms of three quintessential components: the data model, the process model, and the parameter model. Each of these three components can be extended further, but the basic concept that statistical models should account for measurement error, process stochasticity, and parameter uncertainty, all simultaneously, is very powerful for making honest and reliable inference (Gelfand and Ghosh 2015).

Stochastic sampling approaches such as Markov chain Monte Carlo (MCMC; Gelfand and Smith 1990) have facilitated the ability to fit a wide range of Bayesian models to data (Green et al. 2015). However, as the size of datasets grows and the complexity of models increases, MCMC methods for fitting models have become limited in their applicability for big data settings (Brockwell 2006). Despite a proliferation of alternative sampling approaches (e.g., importance sampling, particle filtering, Hamiltonian Monte Carlo [HMC]; Doucet, de Freitas, and Gordon 2001; Del Moral, Doucet, and Jasra 2006; Neal 2011), MCMC is still popular, but also has fundamental weaknesses such as the inability to easily parallelize the computational procedure (beyond obtaining multiple chains; e.g., Glynn and Heidelberger 1992; Bradford and Thomas 1996; Rosenthal 2000).

Bayesian models also facilitate the formal use of preexisting information (resulting from former data analyses) in future data analyses. However, despite widespread rhetoric claiming that previous Bayesian analyses can and should be incorporated into future data analyses as prior information, it is still rarely done in practice. One potential hurdle to the formal incorporation of prior information is the inability to characterize the results of a previous data analysis as an analytically tractable prior with closed form (e.g., McCarthy and Masters 2005; Garrard et al. 2012). Thus, conventional practical guidance suggests approximating the joint posterior distribution resulting from a previous data analysis (using separate data) with an appropriate joint distribution (e.g., multivariate normal) and then use the approximate prior in the new data analysis. This practice may yield a reasonable approximation in some cases, but is unsatisfying in the sense that recursive Bayesian analyses are not coupled exactly using well-accepted stochastic sampling methods such as sequential Monte Carlo (SMC), MCMC, and HMC to fit the models.

In what follows, we discuss existing recursive Bayesian inference approaches and present a new method for performing recursive Bayesian inference using an advantageous combination of existing methods. Our methods are helpful in a variety of situations, for both hierarchical and nonhierarchical Bayesian model fitting. For ongoing data collection efforts, the procedure we describe allows us to represent previous data analyses as priors in new analyses. We show that iterative inferential pro-

CONTACT Mevin B. Hooten *hooten@rams.colostate.edu* Department of Statistics, Colorado State University, 1484 Campus Delivery, Fort Collins, CO.
Color versions of one or more of the figures in the article can be found online at *www.tandfonline.com/r/TAS*.

cedures can facilitate more rapid results using the methods we describe, especially when it is not feasible to fit the full model repetitively as new data arrive. Our approach can also be used to leverage parallel computing resources to accelerate the fitting of complicated Bayesian models such as those containing explicit dependence structure (e.g., Gaussian process models). By partitioning datasets and applying the recursive Bayesian model fitting procedure, we show that our approach can lead to more efficient algorithms for fitting Bayesian models to big datasets. Furthermore, our approach is accessible to practitioners and is compatible with SMC, MCMC, and HMC methods for fitting Bayesian models. We demonstrate our methods with two case studies: a geostatistical model fit to environmental data and a hierarchical dynamic population model fit to ecological data.

2. Methods

Also known as "sequential" inference or Bayesian filtering, recursive Bayes (RB) relies on fitting a statistical model to data in a series of steps (Särkkä 2013). Traditional RB inference has a natural appeal in studies where data are regularly collected over time and thus, it has been more commonly used in conjunction with state-space models (Chopin, Jacob, and Papaspiliopoulos 2013). While the word "sequential" also appears in SMC, and SMC methods are relevant for RB (Chopin 2002), they are not essential to the concept (as we describe in what follows).

The general concept of performing an analysis in sequence is commonplace in many fields. While many statistical methods are developed for analyzing a full dataset in a single procedure, it may be advantageous to analyze datasets in groups. For example, in addition to the situation where data arise sequentially, there may be computational advantages to analyze data in groups even when they are not indexed temporally. In what follows, we review conventional RB based on a method we refer to as "Prior-RB." We contrast Prior-RB with alternative recursive statistical procedures that rely on a sequence of stages meant to facilitate computation. We refer to these approaches as "Proposal-RB" (for reasons that will become clear). Finally, we combine these two recursive procedures to provide a framework for fitting Bayesian models more efficiently by leveraging parallel processing environments that are available in most modern computers.

2.1. Prior-Recursive Bayesian Inference

Consider a generic dataset $\mathbf{y} \equiv (y_1, \ldots, y_n)'$ and associated parametric statistical model $\mathbf{y} \sim [\mathbf{y}|\boldsymbol{\theta}]$, where $\boldsymbol{\theta}$ represents model parameters and we use the bracket notation "$[\cdot]$" to represent probability distributions (Gelfand and Smith 1990). For a specified prior $[\boldsymbol{\theta}]$, the posterior distribution is $[\boldsymbol{\theta}|\mathbf{y}] \propto [\mathbf{y}|\boldsymbol{\theta}][\boldsymbol{\theta}]$. The main concept in Prior-RB is that, for a given partition of the data $\mathbf{y} \equiv (\mathbf{y}_1', \mathbf{y}_2')'$, we can find the posterior distribution associated with the first partition $[\boldsymbol{\theta}|\mathbf{y}_1] \propto [\mathbf{y}_1|\boldsymbol{\theta}][\boldsymbol{\theta}]$ and then use it as a prior in a secondary analysis of the second partition

$$[\boldsymbol{\theta}|\mathbf{y}] \propto [\mathbf{y}_2|\boldsymbol{\theta}, \mathbf{y}_1][\boldsymbol{\theta}|\mathbf{y}_1], \quad (1)$$

$$\propto [\mathbf{y}_2|\boldsymbol{\theta}, \mathbf{y}_1][\mathbf{y}_1|\boldsymbol{\theta}][\boldsymbol{\theta}]. \quad (2)$$

The critical differences between the full model and the Prior-RB procedure are that: 1. the second stage in the Prior-RB procedure requires knowledge of the conditional data model $[\mathbf{y}_2|\boldsymbol{\theta}, \mathbf{y}_1]$ and 2. the form of the posterior resulting from the first stage in Prior-RB $[\boldsymbol{\theta}|\mathbf{y}_1]$ must be known analytically. However, if both distributions are known or at least well-approximated analytically, then we can make inference based on the full dataset, but using only the second partition of data and the output from the first stage posterior. This recursive concept is useful from a meta-analytic perspective because the same analyst does not have to compute the first stage posterior. In fact, well-reported results of a previous data analysis based on a separate dataset can serve as a sufficient statistic for reconciling inference based on both datasets.

We can extend the basic concept of Prior-RB to accommodate multiple partitions of the data. Suppose that we partition the dataset into J groups, $\mathbf{y} \equiv (\mathbf{y}_1', \ldots, \mathbf{y}_J')'$, then we can fit the first stage model as before to yield the posterior distribution $[\boldsymbol{\theta}|\mathbf{y}_1]$. For the jth data partition, we obtain the posterior

$$[\boldsymbol{\theta}|\mathbf{y}_{1:j}] \propto \left(\prod_{\nu=2}^{j} [\mathbf{y}_\nu|\boldsymbol{\theta}, \mathbf{y}_{1:(\nu-1)}] \right) [\mathbf{y}_1|\boldsymbol{\theta}][\boldsymbol{\theta}], \quad (3)$$

where, $\mathbf{y}_{1:j} \equiv (\mathbf{y}_1', \ldots, \mathbf{y}_j')'$. The J-partition Prior-RB procedure still requires analytical knowledge of each sequential posterior as well as the associated conditional data model $[\mathbf{y}_\nu|\boldsymbol{\theta}, \mathbf{y}_{1:(\nu-1)}]$.

To illustrate the Prior-RB procedure, consider the binary dataset $\mathbf{y} \equiv (0, 1, 1, 1, 0, 0, 0, 1)'$, with data model $y_i \sim \text{Bern}(\theta)$ for $i = 1, \ldots, n$ with $n = 8$. Based on a prior for θ specified as $\theta \sim \text{Beta}(1, 1)$, the posterior is a classical result in Bayesian statistics: $[\theta|\mathbf{y}] = \text{Beta}(\sum_{i=1}^{n} y_i + 1, \sum_{i=1}^{n}(1 - y_i) + 1)$, which is a beta distribution with both parameters equal to 5 in our example.

To perform the Prior-RB method for this example with binary data, we split the dataset into $J = 4$ groups resulting in $\mathbf{y}_1 \equiv (0, 1)'$, $\mathbf{y}_2 \equiv (1, 1)'$, $\mathbf{y}_3 \equiv (0, 0)'$, and $\mathbf{y}_4 \equiv (0, 1)'$. Then we analyze each dataset recursively, using the appropriate conditional data model $[\mathbf{y}_j|\theta, \mathbf{y}_{1:(j-1)}]$ for each partition of data. For this simple model, the conditional data model is Bernoulli because the original data model assumed conditional independence of the data given θ. Thus, the Prior-RB method proceeds by finding each posterior recursively: $[\theta|\mathbf{y}_1]$, $[\theta|\mathbf{y}_{1:2}]$, $[\theta|\mathbf{y}_{1:3}]$, and $[\theta|\mathbf{y}_{1:4}]$. It is easily shown that these are all beta distributions with parameter sets {2,2}, {4,2}, {4,4}, and {5,5}. Thus, the Prior-RB method results in the same posterior distribution (i.e., Beta(5, 5)) as fitting the model to all data simultaneously.

The practical application of Prior-RB in settings involving more realistic statistical models and datasets involves two challenges: 1. the ability to find the required conditional data distributions and 2. the representation of the prior for the jth stage based on the $(j-1)$th stage posterior distribution. These two challenges are exacerbated in the application of the Prior-RB method to situations where the data are not conditionally independent and/or more extensive hierarchical models are specified. We revisit these issues in the sections that follow.

2.2. Proposal-Recursive Bayesian Inference

When the data are not naturally ordered in time, it is not apparent how the Prior-RB concept may be helpful. We address

this idea in the following section, but first we set the stage for it by considering a slightly different form of recursive procedure to fit Bayesian models. Suppose the model from the previous section is expanded to accommodate latent random effects $\boldsymbol{\beta}_j$ for $j = 1, \ldots, J$ based on a natural partitioning of the dataset $\mathbf{y} = (\mathbf{y}_1', \ldots, \mathbf{y}_J')'$ (not necessarily partitioned in time). Then a generic hierarchical model structure for the data may be specified as $\mathbf{y}_j \sim [\mathbf{y}_j|\boldsymbol{\beta}_j]$, with "process" $\boldsymbol{\beta}_j \sim [\boldsymbol{\beta}_j|\boldsymbol{\theta}]$, and prior $\boldsymbol{\theta} \sim [\boldsymbol{\theta}]$, for $j = 1, \ldots, J$, and where $\boldsymbol{\beta}_j$ are $p \times 1$ vectors and the dataset partitions $\mathbf{y}_j$ are not necessarily equal-sized.

For example, consider the situation where J different datasets are collected by separate investigators and each set of coefficients $\boldsymbol{\beta}_j$ represent a subpopulation of interest. Suppose that our main goal is to make population-level inference by characterizing the parameters $\boldsymbol{\theta}$. These parameters ($\boldsymbol{\theta}$) give rise to the stochasticity associated with the subpopulation coefficients $\boldsymbol{\beta}_j$ and could represent, for example, an overall effect at the population level of a predictor variable on the response after accounting for subpopulation-level variation. When the desired sample unit is the subpopulation, the hierarchical model helps avoid pseudoreplication in the study (e.g., Hurlbert 1984).

The hierarchical model can also be thought of as a way to reconcile the results of separate data analyses in a meta-analysis framework. Lunn et al. (2013) sought to use models with similar hierarchical structure to perform meta-analysis, synthesizing results across separate studies to obtain population-level inference for $\boldsymbol{\theta}$. Assuming that each study used stochastic sampling methods (e.g., MCMC) to fit a Bayesian model to obtain a sample from the posterior distribution $[\boldsymbol{\beta}_j|\mathbf{y}_j] \propto [\mathbf{y}_j|\boldsymbol{\beta}_j][\boldsymbol{\beta}_j]$ based on the prior $[\boldsymbol{\beta}_j]$, Lunn et al. (2013) proposed a way to recursively use the results of these first stage analyses in a second stage to obtain population-level inference based on the full dataset. We refer to this approach as Proposal-RB because Lunn et al. (2013) suggested using the posterior samples from the subpopulation-level analyses as Metropolis–Hastings (M-H) proposals for $\boldsymbol{\beta}_j$ when fitting the full hierarchical model using MCMC.

The Proposal-RB approach is comprised of the following stages: 1. specify subpopulation-level priors $[\boldsymbol{\beta}_j]$ and obtain a sample from the posterior distributions $[\boldsymbol{\beta}_j|\mathbf{y}_j]$ for $j = 1, \ldots, J$ independently, then 2. fit the full model using MCMC with M-H updates for $\boldsymbol{\beta}_j$ based on the previous stage posterior as a proposal ($\boldsymbol{\beta}_j^{(*)}$).

The M-H acceptance probability for each $\boldsymbol{\beta}_j^{(*)}$ is $\min(r_j, 1)$ where r_j simplifies to

$$r_j = \frac{[\boldsymbol{\beta}_j^{(*)}|\boldsymbol{\theta}^{(k-1)}][\boldsymbol{\beta}_j^{(k-1)}]}{[\boldsymbol{\beta}_j^{(k-1)}|\boldsymbol{\theta}^{(k-1)}][\boldsymbol{\beta}_j^{(*)}]}, \tag{4}$$

with $\boldsymbol{\beta}_j^{(*)}$ arising from the first stage posterior sample and MCMC iteration k ($k = 1, \ldots, K$). The proposals $\boldsymbol{\beta}_j^{(*)}$ should be independent draws from the first stage posterior distribution for the cancellations to occur in the M-H ratio (4). Thus, in practice, we sample $\boldsymbol{\beta}_j^{(*)}$ randomly (with replacement) from the first stage Markov chains to reduce autocorrelation (Lunn et al. 2013, Appendix A). We then use a Gibbs, M-H, importance, or Hamiltonian update for the remaining model parameters $\boldsymbol{\theta}$ based on their full-conditional distribution $[\boldsymbol{\theta}|\cdot] \propto \left(\prod_{j=1}^{J}[\boldsymbol{\beta}_j|\boldsymbol{\theta}]\right)[\boldsymbol{\theta}]$ as usual (note that this full-conditional distribution does not involve $\mathbf{y}$).

Benefits of the Proposal-RB suggested by Lunn et al. (2013) are that: 1. it provides a way to use output from a first stage analysis to fit a full hierarchical model where the first stage posterior distributions are well-represented, 2. it is not limited to meta-analysis, and 3. it can dramatically simplify the M-H ratio (4) because the data model cancels in the numerator and denominator. Thus, using only output from J independent model fits and knowledge of the first stage priors $[\boldsymbol{\beta}_j]$, we can fit the full hierarchical model to obtain inference.

Aside from being a generally useful approach for fitting hierarchical models recursively, the Proposal-RB procedure is useful in data privacy situations where the original data cannot be released due to proprietary reasons, public safety, or legal restrictions (Altman et al. 2018) because the data do not appear in the second stage analysis. Proposal-RB is also trivial to implement and is naturally adapted for parallel computing environments because we can sample from each of the transient posterior distributions $[\boldsymbol{\beta}_j|\mathbf{y}_j]$ in parallel at the first stage.

To demonstrate the Proposal-RB approach, Hooten and Hefley (2019) fit a hierarchical Gaussian model to a set of biometric data taken on a sample of blue tits (*Cyanistes caeruleus*) that arose from a study of individual-level versus population-level variation in wild birds. Hooten and Hefley (2019) specified a simple hierarchical Bayesian model with random means for each individual and showed that the results are the same when fitting the full hierarchical model using traditional MCMC and when using Proposal-RB.

When the models and/or datasets are more complex, the Proposal-RB method can lead to computational improvements. For example, Hooten et al. (2016) and Gerber et al. (2018) applied the Proposal-RB method to make population-level inference about resource selection by animals. In particular, Hooten et al. (2016) showed that the Proposal-RB approach suggested by Lunn et al. (2013) may be particularly useful for cases where the data model is numerically challenging to evaluate. Specifically, Hooten et al. (2016) considered a hierarchical point process model for animal telemetry data where the data model was specified as a product of weighted distributions (Patil and Rao 1977) that contained integrals of a function that included model parameters. These integrals were a crux in implementing the spatial point process model because an optimization or stochastic sampling algorithm such as MCMC must numerically calculate the integral repeatedly when fitting the model to data. The Proposal-RB approach used by Hooten et al. (2016) simplified the sampling procedure in the second stage analysis substantially because the integrals in the data model did not need to be calculated again after the first stage, resulting in a procedure that required less computational time than fitting the full model jointly.

Overall, the Proposal-RB method is useful for fitting certain classes of hierarchical models to data that are naturally partitioned. However, Proposal-RB does not directly translate to nonhierarchical models and cases where the data are not conditionally independent. When the data are not conditionally independent, we can still fit independent models for each

partition in the first stage, but the data model will not cancel in the second stage M-H updates for $\boldsymbol{\beta}_j$. If the data model is trivial to calculate, it is possible that the Proposal-RB approach may still be useful, but data models with dependence (e.g., Gaussian process models) can be numerically challenging to calculate repetitively.

Similarly, for nonhierarchical models, natural partitions of the data may not exist and it becomes more difficult to envision useful partition-specific first stage models. While it may be possible to contrive an auxiliary variable approach that augments a nonhierarchical model with an artificial latent process (e.g., Albert and Chib 1993), we propose a simpler alternative in what follows.

2.3. Prior-Proposal-Recursive Bayesian Inference

Proposal-RB is useful for meta-analysis and fitting hierarchical models, but the concepts in Proposal-RB do not automatically transfer to nonhierarchical models, or if so, may not be helpful computationally. Therefore, we propose a combination of Prior- and Proposal-RB (hereafter, PP-RB) concepts that makes RB more accessible to practitioners and facilitates inference for model parameters for a wide class of Bayesian models.

Our PP-RB approach assumes the data can be partitioned as described earlier such that $\mathbf{y} \equiv (\mathbf{y}_1', \ldots, \mathbf{y}_J')'$ and we can implement the Prior-RB procedure for recursively fitting the full model in stages. To implement the PP-RB approach, we first obtain a sample from $[\boldsymbol{\theta}|\mathbf{y}_1]$ as before, then, for the next $J-1$ stages, we recursively obtain samples from

$$[\boldsymbol{\theta}|\mathbf{y}_{1:j}] \propto [\mathbf{y}_j|\boldsymbol{\theta}, \mathbf{y}_{1:(j-1)}][\boldsymbol{\theta}|\mathbf{y}_{1:(j-1)}], \tag{5}$$

for $j = 2, \ldots, J$. Borrowing the technique from Proposal-RB where we use the transient posterior from the previous stage as the proposal (in addition to the prior, as in Prior-RB), our M-H acceptance probability for the jth stage and kth MCMC iteration can be written as $\min(r_j, 1)$ with

$$r_j = \frac{[\mathbf{y}_j|\boldsymbol{\theta}^{(*)}, \mathbf{y}_{1:(j-1)}][\boldsymbol{\theta}^{(*)}|\mathbf{y}_{1:(j-1)}][\boldsymbol{\theta}^{(k-1)}|\mathbf{y}_{1:(j-1)}]}{[\mathbf{y}_j|\boldsymbol{\theta}^{(k-1)}, \mathbf{y}_{1:(j-1)}][\boldsymbol{\theta}^{(k-1)}|\mathbf{y}_{1:(j-1)}][\boldsymbol{\theta}^{(*)}|\mathbf{y}_{1:(j-1)}]}, \tag{6}$$

$$= \frac{[\mathbf{y}_j|\boldsymbol{\theta}^{(*)}, \mathbf{y}_{1:(j-1)}]}{[\mathbf{y}_j|\boldsymbol{\theta}^{(k-1)}, \mathbf{y}_{1:(j-1)}]}, \tag{7}$$

where $\boldsymbol{\theta}^{(*)}$ is the kth realization from the transient posterior sample from the previous stage. Notice that the M-H ratio in (7) consists only of a ratio of the conditional data models. Thus, the PP-RB approach still requires the knowledge and calculation of the conditional data model at each MCMC iteration and stage. However, because the set of posterior realizations we use as proposals throughout the procedure are acquired as a result of the first stage analysis, we can pre-calculate the log density (or mass function) of the conditional data model for each proposal $\boldsymbol{\theta}^{(*)}$ in parallel between stages 1 and 2 in the procedure. With values for the numerator in the M-H ratio resulting from our quasi-prefetching technique (i.e., the pre-calculation of log densities for all possible proposals of $\boldsymbol{\theta}$; Brockwell 2006), performing the updates for $\boldsymbol{\theta}$ is less computationally intensive. Furthermore, because we need only save the values for $\log[\mathbf{y}_j|\boldsymbol{\theta}^{(*)}, \mathbf{y}_{1:(j-1)}]$ after the first stage, the PP-RB approach has low memory requirements between stages.

3. PP-RB Application to Geostatistics

Our PP-RB approach can be applied to fit a wide range of Bayesian models recursively. As a first demonstration of the PP-RB method, we apply it to fit the standard geostatistical model (Cressie 1993), which is very commonly used in environmental and ecological applications. The data used for this illustration are measurements of sea surface temperature (SST) on the eastern and northern Bering Sea shelf near the Pribilof Islands, Alaska. The data were obtained as part of the 2017 NOAA Fisheries bottom trawl surveys used to assess the condition of groundfish stocks in the Bering Sea. The SST measurements are collected in the same locations as the fishing trawls (Figure 1, "FULL"). There are $n = 520$ observations in this dataset spaced on a 20 km grid with additional locations surveyed near the Pribilof Islands and St. Matthew Island.

Parametric geostatistical modeling involves the use of Gaussian processes that are ubiquitous throughout many different fields. The use of Gaussian processes in spatially explicit models has a long history in statistics, but has experienced a resurgence lately due to the need to flexibly and efficiently model large datasets and provide optimal predictions in space and time (e.g., Gelfand and Schliep 2016; Hooten and Johnson 2017; Heaton et al. 2019).

For this example, we specify a version of the full Bayesian geostatistical model (Arab, Hooten, and Wikle 2017) as

$$\mathbf{y} \sim \text{N}(\mathbf{X}\boldsymbol{\beta}, \boldsymbol{\Sigma}(\sigma^2, \phi, \tau^2)), \tag{8}$$

$$\boldsymbol{\beta} \sim \text{N}(\boldsymbol{\mu}_\beta, \boldsymbol{\Sigma}_\beta), \tag{9}$$

$$\sigma^2 \sim \text{Inv-}\chi^2(\alpha_1, \alpha_2), \tag{10}$$

$$\phi \sim \text{half-N}(0, \gamma), \tag{11}$$

$$\tau^2 \sim \text{Unif}(0, 1), \tag{12}$$

where the full dataset is denoted as $\mathbf{y} \equiv (y_1, \ldots, y_n)'$ and comprises observations of SST (°C) at locations $\mathbf{s}_1, \ldots, \mathbf{s}_n$ in continuous space $\mathcal{S}$. The spatial covariance of $\mathbf{y}$ is modeled with sill, range, and nugget parameters as

$$\boldsymbol{\Sigma}(\sigma^2, \phi, \tau^2) \equiv \sigma^2((1-\tau^2)\mathbf{R}(\phi) + \tau^2\mathbf{I}). \tag{13}$$

We used a Matérn (Matérn 1986; Guttorp and Gneiting 2006) covariance function with smoothness parameter set to 3/2 to model the latent spatial structure and parameterize the correlation matrix $\mathbf{R}(\phi)$. The entries of the correlation matrix are $R_{ij} = (1 + d_{ij}/\phi)\exp(-d_{ij}/\phi)$, where ϕ is a parameter that controls the range of spatial structure and the Euclidean distance between locations $\mathbf{s}_i$ and $\mathbf{s}_j$ is $d_{ij} = ||\mathbf{s}_i - \mathbf{s}_j||_2$. For simplicity, the spatial process is assumed to be second-order stationary and isotropic (although our PP-RB approach can be applied in cases with more general assumptions as well). For covariates, we used the easting and northing associated with each spatial location.

The full posterior distribution associated with our geostatistical model is $[\boldsymbol{\beta}, \sigma^2, \phi, \tau^2|\mathbf{y}] \propto [\mathbf{y}|\boldsymbol{\beta}, \sigma^2, \phi, \tau^2][\boldsymbol{\beta}][\sigma^2][\phi][\tau^2]$. To fit the full geostatistical model in (8)–(12), we constructed a MCMC algorithm based on conjugate updates for $\boldsymbol{\beta}$ and σ^2, and used a M-H update for ϕ and τ^2 (see Appendix B for details on the implementation). For our example, we used conjugate Jeffreys' specifications (Jeffreys 1946) for the priors $[\boldsymbol{\beta}] \propto 1$ and $[\sigma^2] = 1/\sigma^2$.

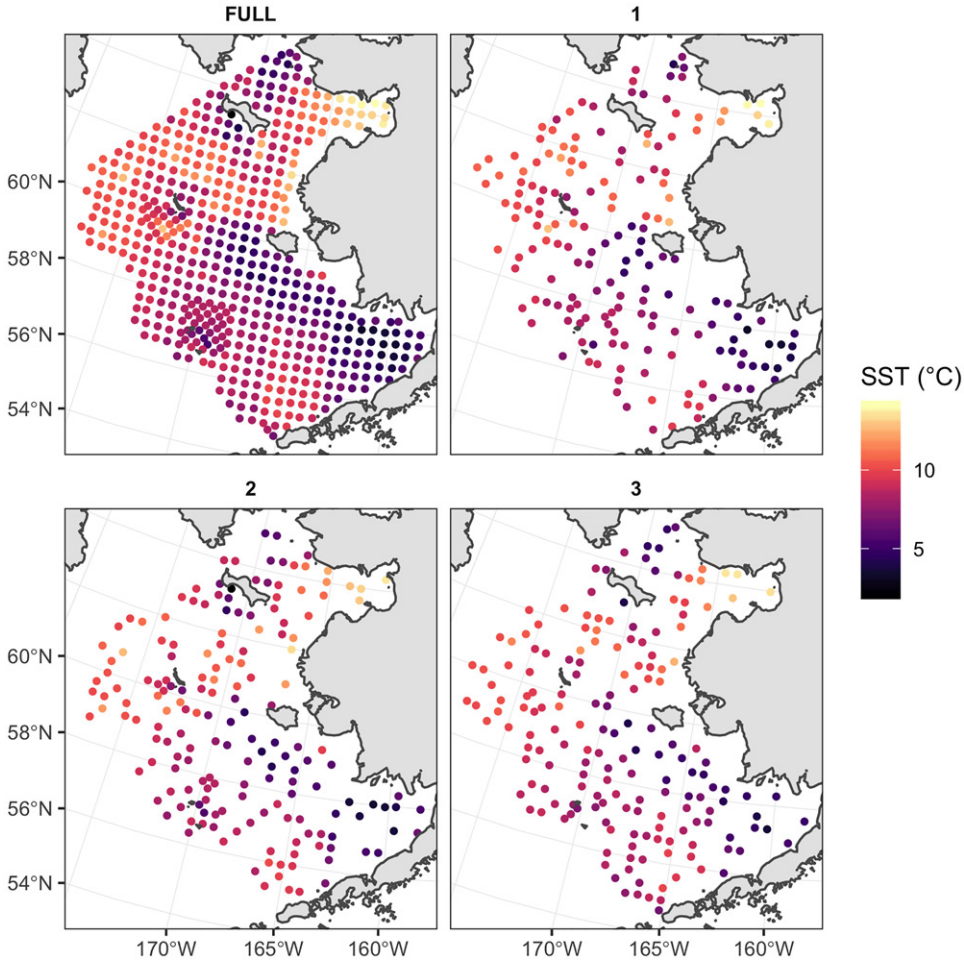

Figure 1. Sea surface temperature measurements (°C) from NOAA Fisheries 2017 bottom trawl survey on the eastern and northern Bering Sea shelf off the coast of Alaska. The "FULL" plot depicts all data together, while the remaining plots show the locations used in each partition of the data for the PP-RB approach.

The general PP-RB procedure to fit the Bayesian geostatistical model for $J > 3$ partitions of the data involves the following steps:

1. Partition the data into $\mathbf{y} \equiv (\mathbf{y}_1', \ldots, \mathbf{y}_J')'$ subsets.
2. Stage 1: Fit the Bayesian geostatistical model in (8)–(12) to the first partition of data to yield a MCMC sample from $[\boldsymbol{\beta}, \sigma^2, \phi, \tau^2 | \mathbf{y}_1]$ resulting in realizations $\boldsymbol{\beta}^{(k)}$, $\sigma^{2(k)}$, $\phi^{(k)}$, $\tau^{2(k)}$ for MCMC iteration $k = 1, \ldots, K$.
3. Calculate $\log[\mathbf{y}_j | \boldsymbol{\beta}^{(k)}, \sigma^{2(k)}, \phi^{(k)}, \tau^{2(k)}, \mathbf{y}_{1:(j-1)}]$ for realizations $k = 1, \ldots, K$ and partitions $j = 2, \ldots, J$, in parallel.
4. Stage 2: Perform block M-H updates for model parameters using $\boldsymbol{\beta}^{(k)}$, $\sigma^{2(k)}$, $\phi^{(k)}$, and $\tau^{2(k)}$ randomly from the first stage transient posterior as proposals in (7) according to Appendix A.
5. Stage 3: Sampling randomly from the resulting MCMC sample from the second stage as proposals (Appendix A), perform the third stage M-H updates based on the ratio (7).
6. Stages 4–J: Repeat for all stages, conditioning on the posterior from the previous stage each time.

The precalculation step between stages 1 and 2 in our PP-RB procedure is the computational crux because we must evaluate the log density of the conditional Gaussian distribution repetitively. Based on well-known multivariate Gaussian properties (e.g., Gentle 2007), the jth conditional data distribution for our geostatistical model is $[\mathbf{y}_j | \boldsymbol{\beta}^{(k)}, \sigma^{2(k)}, \phi^{(k)}, \tau^{2(k)}, \mathbf{y}_{1:(j-1)}] = \mathrm{N}(\tilde{\boldsymbol{\mu}}_j, \tilde{\boldsymbol{\Sigma}}_j)$, with conditional mean and covariance

$$\tilde{\boldsymbol{\mu}}_j \equiv \mathbf{X}_j\boldsymbol{\beta} + \boldsymbol{\Sigma}_{j,1:(j-1)}\boldsymbol{\Sigma}^{-1}_{1:(j-1),1:(j-1)}(\mathbf{y}_{1:(j-1)} - \mathbf{X}_{1:(j-1)}\boldsymbol{\beta}), \tag{14}$$

$$\tilde{\boldsymbol{\Sigma}}_j \equiv \boldsymbol{\Sigma}_{j,j} - \boldsymbol{\Sigma}_{j,1:(j-1)}\boldsymbol{\Sigma}^{-1}_{1:(j-1),1:(j-1)}\boldsymbol{\Sigma}_{1:(j-1),j}. \tag{15}$$

Thus, to evaluate the conditional data model, we must calculate two matrix inverses as well as several matrix products and a determinant. The floating point operations (FLOPS) associated with inverting $\boldsymbol{\Sigma}_{1:(j-1),1:(j-1)}$ are the most numerically intensive, on the order of $O(n^3_{1:(j-1)})$ (where $n_{1:(j-1)}$ is the dimension of $\mathbf{y}_{1:(j-1)}$), which is less than $O(n^3)$ required for the full dataset. In the case where we have two equal sized partitions (i.e., $J = 2$), the FLOPS associated with matrix inverses are $O(2(\frac{n}{2})^3) = O(\frac{n^3}{4})$, four times less than for the full dataset. Additionally, after the log conditional data model is evaluated for a given set of parameters, we do not need to retain its mean and covariance matrix, which reduces our storage requirements substantially.

We applied the PP-RB approach to fit the Bayesian geostatistical model to the SST data using $J = 3$ partitions and $K = 200{,}000$ MCMC iterations. Figure 1 shows the full spatial dataset and the $J = 3$ partitions of data, subsampled randomly from the full dataset. The computational time required to perform the

entire PP-RB procedure based on a first stage model fit with $K = 200{,}000$ MCMC iterations was 1.7 hr whereas the time required to fit the full model with the same number of MCMC iterations was 6.9 hr. Thus, our PP-RB approach based on $J = 3$ random partitions of the spatial data resulted in an algorithm that was approximately 4 times faster to obtain the same inference from the exact model without approximating the covariance function.

We summarized the inference resulting from the two model fits in Figure 2, where the 95% credible intervals and posterior means for each parameter are shown for the full model (in black) and for each stage of the PP-RB procedure in colors ranging from red (stage 1) to green (stage 3). It is clear that our inference concerning all geostatistical model parameters improves as we fit the models in each stage of the PP-RB procedure recursively (the green credible intervals match the black ones in Figure 2). In particular, for the β_3 regression coefficient (associated with the northing covariate), our inference changes from nonsignificant to significantly different than zero (based on the 95% credible interval) between the second and third stages as the credible interval is shrunk toward the full-data posterior when we incorporate additional partitions of data.

4. PP-RB Application to Online Updating

To illustrate the PP-RB approach for "online" updating (i.e., a strategy for efficiently assimilating new data as they become available; Schifano et al. 2016; Wang et al. 2018), we analyzed temporal data resulting from surveys of Steller sea lion populations. Steller sea lions are listed as endangered under the U.S. Endangered Species Act over much of their geographic range. The National Marine Fisheries Service of the U.S. Federal Government monitors the status of this species in Alaska by conducting aerial surveys to count the number of sea lion pups born in the Aleutian Islands and Gulf of Alaska each year. We focused our analysis on counts at two different rookery sites (Marmot and Sugarloaf) monitored during 1978–2013; although both sites were not surveyed every year and survey effort was generally sparse early in the monitoring program. The Steller sea lion pup count data are available in the R package "agTrend."

We analyzed Steller sea lion pup counts using the hierarchical model

$$y_{s,t} \sim \text{Pois}(\lambda_{s,t}), \tag{16}$$
$$\log(\lambda_{s,1}) \sim \text{N}(\mu_1, \sigma_1^2), \tag{17}$$
$$\log\left(\lambda_{s,t}\right) \sim \text{N}(\phi_s + \log\left(\lambda_{s,t-1}\right), \sigma_s^2), \tag{18}$$
$$\phi_s \sim \text{N}(0, \sigma_\phi^2), \tag{19}$$
$$\sigma_s^2 \sim \text{IG}(\alpha, \beta), \tag{20}$$

where $y_{s,t}$ is the observed pup count at sites $s = 1, 2$ (i.e., Marmot and Sugarloaf sites) in year $t = 1, \ldots, T$ (see Appendix C for implementation details). These sites were not monitored in year $T+1 = 2014$, thus we sought to use the available data from 1978 through 2013 ($\mathbf{y}_{s,1:T}$) to predict sea lion pup count intensity in the year 2015 ($\lambda_{s,T+2}$) as rapidly as possible after obtaining the subsequent observations $y_{s,T+2}$. We used the PP-RB approach to accomplish this task without the burden of fitting the model to the full dataset.

In the context of online updating, we assume a first-stage analysis has been conducted based on previous pup counts $\mathbf{y}_{s,1:T}$ resulting in a MCMC sample comprised of $\boldsymbol{\theta}_s^{(k)} \equiv (\phi_s^{(k)}, \sigma_s^{2(k)}, \boldsymbol{\lambda}_s^{(k)})'$ (where $k = 1, \ldots, K$ indexes MCMC iterations from the first-stage analysis). When new data $y_{s,T+2}$ arrive and we wish to update inference using the second-stage algorithm, we first sample the new intensity parameters $\lambda_{s,T+1}$ and $\lambda_{s,T+2}$ from their predictive distributions

$$\log\left(\lambda_{s,T+1}^{(*)}\right) \sim \text{N}\left(\phi_s^{(k)} + \log\left(\lambda_{s,T}^{(k)}\right), \sigma_s^{2(k)}\right), \tag{21}$$
$$\log\left(\lambda_{s,T+2}^{(*)}\right) \sim \text{N}\left(\phi_s^{(k)} + \log\left(\lambda_{s,T+1}^{(*)}\right), \sigma_s^{2(k)}\right). \tag{22}$$

Then, the kth M-H acceptance ratio to update all parameters, including $\lambda_{s,T+2}$, in the second-stage analysis is

$$r = \frac{[y_{s,T+2}|\lambda_{s,T+2}^{(*)}][\lambda_{s,T+2}^{(*)}, \lambda_{s,T+1}^{(*)}, \boldsymbol{\theta}_s^{(*)}|\mathbf{y}_{s,1:T}] \times [\lambda_{s,T+2}^{(k-1)}, \lambda_{s,T+1}^{(k-1)}, \boldsymbol{\theta}_s^{(k-1)}|\mathbf{y}_{s,1:T}]}{[y_{s,T+2}|\lambda_{s,T+2}^{(k-1)}][\lambda_{s,T+2}^{(k-1)}, \lambda_{s,T+1}^{(k-1)}, \boldsymbol{\theta}_s^{(k-1)}|\mathbf{y}_{s,1:T}] \times [\lambda_{s,T+2}^{(*)}, \lambda_{s,T+1}^{(*)}, \boldsymbol{\theta}_s^{(*)}|\mathbf{y}_{s,1:T}]}, \tag{23}$$

$$= \frac{[y_{s,T+2}|\lambda_{s,T+2}^{(*)}]}{[y_{s,T+2}|\lambda_{s,T+2}^{(k-1)}]}, \tag{24}$$

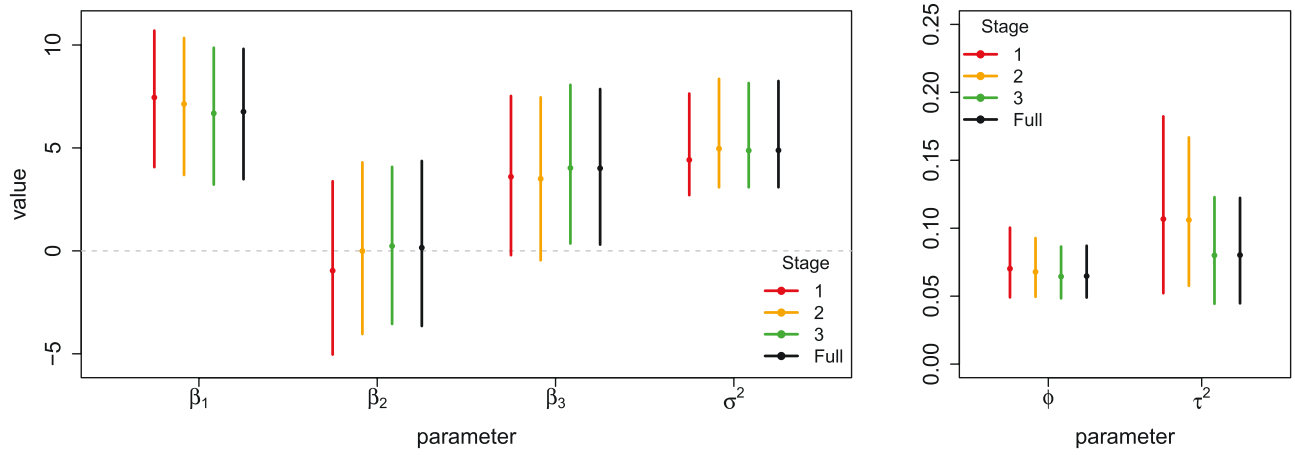

Figure 2. Posterior means (points) and 95% credible intervals resulting from fitting the Bayesian geostatisical model to the full dataset (black) and for each partition $j = 1, \ldots, 3$ using the PP-RB approach.

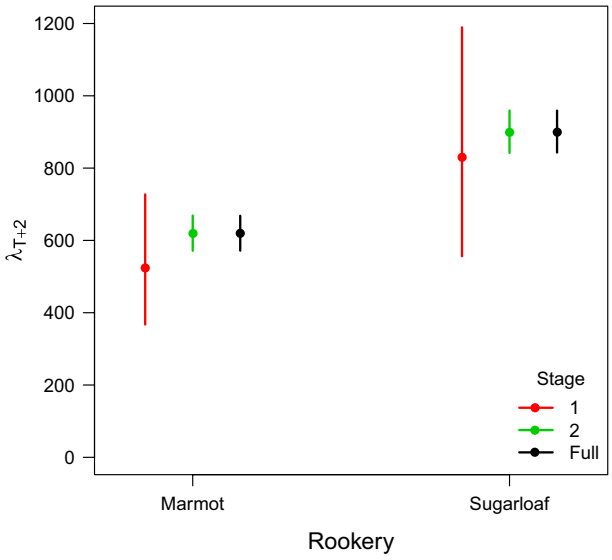

Figure 3. Posterior means (points) and 95% credible intervals (vertical lines) for Steller sea lion pup count intensities $\lambda_{s,T+2}$ at two sites in the Aleutian Islands, AK for year $T + 2 = 2015$. The black intervals correspond to the results from the full model, whereas the red and green intervals correspond to the inference obtained from the first- and second-stages of the PP-RB analysis.

which simplifies to a function containing only Poisson probability mass functions resulting from the fact that the proposal distribution for $(\lambda_{s,T+2}, \lambda_{s,T+1}, \boldsymbol{\theta}_s)$ is $[\lambda_{s,T+2}, \lambda_{s,T+1}, \boldsymbol{\theta}_s | \mathbf{y}_{s,1:T}]$, which is sampled during the first-stage analysis. As before, we draw proposals $\lambda^{(*)}_{s,T+2}$ randomly with replacement from the stage one MCMC posterior predictive sample (Appendix A) and accept the proposal with probability $\min(r, 1)$.

To compare the PP-RB method for online updating with the model fit to the entire dataset simultaneously, we used $K = 100{,}000$ MCMC iterations for both the full dataset ($\mathbf{y}_{s,1:T}$ and $y_{s,T+2}$) and the dataset without the last year of data ($\mathbf{y}_{s,1:T}$). We then relied on the PP-RB method to assimilate the final year of data $y_{s,T+2}$ in a second algorithm using the M-H updates described in (24). Although predictions from the first-stage analysis were highly uncertain (i.e., wide red credible intervals in Figure 3), upon incorporation of the new data ($y_{s,T+2}$), inference was virtually identical to the full-data posterior (i.e., green and black credible intervals match in Figure 3). For this analysis, updating our inference using the second-stage algorithm and the final year of data was 59 times faster than fitting the model to the full dataset simultaneously.

5. Discussion

In this era where new data are constantly streaming in and both sensing and storage technology are improving, online statistical models have become more challenging to fit efficiently. Dietze et al. (2018) made a strong case for the need to fit statistical models to incoming data operationally and regularly provide iterative forecasts based on important ecological and environmental data streams. Statistically rigorous recursive inference and forecasting is clearly useful in all fields, but existing methods for assimilating new data recursively or performing meta-analysis may be inaccessible to practitioners or computationally infeasible. Our PP-RB approach relies on well-known Bayesian updating principles and commonly used MCMC methods for fitting models. The PP-RB approach combines two existing RB concepts (i.e., Prior- and Proposal-RB) to result in a broadly applicable multistage technique for fitting Bayesian models sequentially to partitioned datasets.

Overall, the PP-RB method we presented is very accessible to practitioners because it relies on a first-stage posterior sample (that can be acquired using automated software) followed by a sequence of simple M-H updates. The multicore architecture of modern computers can be leveraged to accelerate the PP-RB by precomputing the conditional log-likelihoods for each first-stage MCMC sample in parallel, but parallel computing is not necessary to use PP-RB in general. For example, when the results of a previous analysis are available and we only seek to obtain inference based on a single new incoming data partition (i.e., partition $\mathbf{y}_J$), no recursion is necessary. In that case, we simply condition on the most recent model output (i.e., based on partitions $\mathbf{y}_{1:(J-1)}$).

The main challenge associated with the PP-RB method is that we need to evaluate conditional log-likelihoods. The PP-RB procedure is straightforward when the conditional log-likelihood function is analytically tractable (such as for the multivariate normal distribution), but approximation of the conditional log-likelihood must be used when it is not tractable. We have had good success approximating the conditional log-likelihood using SMC and, in some cases, pseudo-likelihood approaches in a variety of other applications.

The data partitions required in PP-RB need not correspond to a meaningful aspect of space, time, or model structure, but in some cases, natural partitions may be available (i.e., spatio-temporal data) and can be used. In fact, partition design is an important area of future research related to PP-RB because it could lead to optimal recursive strategies and even faster inference. In fact, Gramacy and Apley (2015) and Guinness (2018) explored similar concepts related to the design of partitions for fitting approximate Gaussian process models more efficiently. Those partitioning concepts can be used in conjunction with our PP-RB approach and may extend to a broader set of Bayesian statistical models. However, poorly selected partitions may result in suboptimal inference because the early stages could result in MCMC samples that do not adequately explore the correct posterior in practice. For example, Zimmerman (2006) found that designs with clusters comprised of sampling locations that are near each other in space facilitate the estimation of covariance parameters in geostatistical models.

Many other approaches to Gaussian process approximations have been developed over the past several decades and are appearing with greater regularity recently. For example, Vecchia (1988) presented a Gaussian process approximation based on the same type of recursive expression of the data model we used in our geostatistical example from the previous section (also see Stein, Chi, and Welty 2004; Huang and Sun 2018; Katzfuss and Guinness 2018). This concept led to several recursive approaches to developing approximate Gaussian process models that have been proposed recently, including predictive processes (Banerjee et al. 2008) and nearest neighbor processes (Datta et al. 2016; Finley et al. 2018), both of which are compatible with our PP-RB method. Furthermore, any of the alternative approaches for approximating Gaussian process covariance

matrices using reduced-rank or sparse parameterizations (e.g., Higdon 2002; Furrer, Genton, and Nychka 2006; Cressie and Johannesson 2008; Wikle 2010; Lindgren, Rue, and Lindstrom 2011; Gramacy and Apley 2015; Nychka et al. 2015; Katzfuss 2017) are also compatible with our PP-RB method, as long as they are applied in a Bayesian context (also see Heaton et al. (2019) for an excellent review). Finally, there may be value in pairing subsampling methods (e.g., Liang et al. 2013; Kleiner et al. 2014; MacLaurin and Adams 2015; Barbian and Assuncao 2017) with PP-RB to reduce computational requirements further.

The natural recursive nature of the PP-RB method is not limited to use for improving computational efficiency, it also reconciles well with optimal design and monitoring strategies. Optimal adaptive design, especially in the spatio-temporal context, is becoming more popular in environmental (e.g., Wikle and Royle 1999, 2005) and ecological statistics (e.g., Hooten et al. 2009; Hooten, Ross, and Wikle 2012; Williams et al. 2018). Our PP-RB method can be used to rapidly assimilate new data and characterize posterior forecast distributions that can be optimized to reduce the uncertainty associated with ongoing monitoring efforts without requiring a reanalysis of the entire cumulative dataset.

In terms of alternative methods for efficient Bayesian computing, a variety of computing strategies have become popular because of increasing computational demands due to larger datasets and more complex models. For example, related classes of computing strategies are: Consensus MC (Scott et al. 2016), Weierstrass samplers (Wang and Dunson 2013), embarrassingly parallel MCMC (Neiswanger, Wang, and Xing 2013), and Modular Bayes (Jacob et al. 2017), among others.

Appendix A

The cancellations in the M-H ratios (i.e., Equations (4), (7), and (24)) occur when our proposed parameter values independently arise from the proposal distribution. Lunn et al. (2013) suggested sampling the proposals randomly with replacement from the stage one Markov chains to reduce dependence. To illustrate this concept, we denote the stage one Markov chain values for the parameters in (4) as $\boldsymbol{\beta}_j^{(*,k)}$ for $k = 1, \ldots, K$ MCMC iterations. Then if we sample our proposal at random from those values, the implied proposal distribution is a categorical distribution on the set of $\boldsymbol{\beta}_j^{(*,k)}$ with probabilities $[\boldsymbol{\beta}_j^{(*,k)}|\mathbf{y}_j] / \sum_{v=1}^{K}[\boldsymbol{\beta}_j^{(*,v)}|\mathbf{y}_j]$ for $k = 1, \ldots, K$. Because the denominator sums over the space of $\boldsymbol{\beta}$ in our proposal, it is a function of the data $f(\mathbf{y}_j)$ only. Thus, the proposal distribution has the form

$$\frac{[\boldsymbol{\beta}_j^{(*,k)}|\mathbf{y}_j]}{\sum_{v=1}^{K}[\boldsymbol{\beta}_j^{(*,v)}|\mathbf{y}_j]} \propto f(\mathbf{y}_j)[\boldsymbol{\beta}_j^{(*,k)}|\mathbf{y}_j], \tag{A.1}$$

$$\propto [\boldsymbol{\beta}_j^{(*,k)}|\mathbf{y}_j], \tag{A.2}$$

$$\propto [\mathbf{y}_j|\boldsymbol{\beta}_j^{(*,k)}][\boldsymbol{\beta}_j^{(*,k)}], \tag{A.3}$$

as required in (4). Other options could involve thinning the stage one MCMC sample to reduce dependence in the proposed values or randomly permute the stage one MCMC sample (Hooten and Hefley 2019). Thinning is the most common way to reduce dependence in the MCMC sample before making inference using Monte Carlo integration, but it may reduce the number of possible proposal values in Proposal-RB and PP-RB substantially. By contrast, permuting the stage one MCMC sample will not remove the dependence in the Markov chains completely, but does allow us to use the entire set of potential proposals from the first stage (unlike the sampling with replacement approach).

Appendix B

The Bayesian geostatistical model for the full dataset $\mathbf{y}$ was specified as

$$\mathbf{y} \sim \text{N}(\mathbf{X}\boldsymbol{\beta}, \boldsymbol{\Sigma}(\sigma^2, \phi, \tau^2)), \tag{B.1}$$

$$\boldsymbol{\beta} \sim \text{N}(\boldsymbol{\mu}_\beta, \boldsymbol{\Sigma}_\beta), \tag{B.2}$$

$$\sigma^2 \sim \text{Inv-}\chi^2(\alpha_1, \alpha_2), \tag{B.3}$$

$$\phi \sim \text{half-N}(\gamma^2), \tag{B.4}$$

$$\tau^2 \sim \text{Unif}(0, 1). \tag{B.5}$$

We fit the Bayesian geostatistical model to the full dataset using $K = 20{,}000$ MCMC iterations and hyperparameters $\boldsymbol{\mu}_\beta = (0, 0, 0)'$, $\boldsymbol{\Sigma}_\beta^{-1} = \mathbf{0}$ (i.e., flat prior), $\alpha_1 = 0, \alpha_2 = 0$, and $\gamma = 0.05$. The coordinates for the spatial locations were scaled to be within $[0, 1] \times [0, 1]$ and the $\gamma = 0.05$ choice implies that ≈95% of the posterior mass for the effective range of spatial correlation lies between 0 and 1/3 the maximum distance between spatial locations.

The full-conditional distributions for this geostatistical model are conjugate for $\boldsymbol{\beta}$ and σ^2. For $\boldsymbol{\beta}$ the full-conditional distribution is $[\boldsymbol{\beta}|\cdot] = \text{N}(\mathbf{A}^{-1}\mathbf{b}, \mathbf{A}^{-1})$ with

$$\mathbf{A} \equiv \mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X} + \boldsymbol{\Sigma}_\beta^{-1}, \tag{B.6}$$

$$\mathbf{b} \equiv \mathbf{y}'\boldsymbol{\Sigma}^{-1}\mathbf{X} + \boldsymbol{\mu}_\beta'\boldsymbol{\Sigma}_\beta^{-1}. \tag{B.7}$$

For σ^2 the full-conditional distribution is $[\sigma^2|\cdot] = \text{Inv-}\chi^2(\tilde{\alpha}_1, \tilde{\alpha}_2)$ with

$$\tilde{\alpha}_1 \equiv n + \alpha_1, \tag{B.8}$$

$$\tilde{\alpha}_2 \equiv \frac{\alpha_1\alpha_2 + nS^2}{\alpha_1 + n}, \tag{B.9}$$

where $S^2 = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'((1 - \tau^2)\mathbf{R}(\phi) + \tau^2\mathbf{I})^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})/n$.

The full-conditional distribution for the spatial parameters ϕ and τ^2 will not be conjugate, but we can sample it using an M-H update in the first-stage algorithm. We write the full-conditional distribution for ϕ and τ^2 as

$$[\phi, \tau^2|\cdot] \propto [\mathbf{y}|\boldsymbol{\beta}, \sigma^2, \phi, \tau^2][\phi][\tau^2]. \tag{B.10}$$

We use the random walk method with rejection sampling for proposing values of $(\phi^{(*)}, \tau^{2(*)})$ (where we reject the update when $\phi^{(*)}$ or $\tau^2 \leq 0$ and $\tau^2 > 1$), the resulting M-H ratio is

$$r = \frac{\text{N}(\mathbf{y}|\mathbf{X}\boldsymbol{\beta}^{(k)}, \boldsymbol{\Sigma}(\sigma^{2(k)}, \phi^{(*)}, \tau^{2(*)}))\text{half-N}(\phi^{(*)}|\gamma)}{\text{N}(\mathbf{y}|\mathbf{X}\boldsymbol{\beta}^{(k)}, \boldsymbol{\Sigma}(\sigma^{2(k)}, \phi^{(k-1)}, \tau^{2(k-1)}))\text{half-N}(\phi^{(k-1)}|\gamma)}. \tag{B.11}$$

The random walk proposal distribution is adaptively tuned to reach an acceptance rate of ≈0.3.

Appendix C

The full BHM for the Steller sea lion count data $(y_{s,t})$ is specified as

$$y_{s,t} \sim \text{Pois}(\lambda_{s,t}), \tag{C.1}$$

$$\log(\lambda_{s,1}) \sim \text{N}(\mu_1, \sigma_1^2), \tag{C.2}$$

$$\log(\lambda_{s,t}) \sim \text{N}(\phi_s + \log(\lambda_{s,t-1}), \sigma_s^2), \tag{C.3}$$

$$\phi_s \sim \text{N}(0, \sigma_\phi^2), \tag{C.4}$$

$$\sigma_s^2 \sim \text{IG}(\alpha, \beta), \tag{C.5}$$

for $s = 1, \ldots, 2$ and $t = 2, \ldots, T$. For hyperparameters, we specified $\mu_1 = 8.7, \sigma_1^2 = 1.69, \sigma_\phi^2 = 1, \alpha = 1$, and $\beta = 20$.

We fit this full model using a MCMC algorithm and $K = 100{,}000$ iterations with conjugate updates for model parameters ϕ_s and σ_s^2, and M-H updates for $\lambda_{s,t}$ for all s and t. The full-conditional distributions for model parameters are as follows. For ϕ_s, the full-conditional is $[\phi_s|\cdot] = \mathrm{N}(a^{-1}b, a^{-1})$, with

$$a \equiv \frac{T-1}{\sigma_s^2} + \frac{1}{\sigma_\phi^2}, \tag{C.6}$$

$$b \equiv \frac{\sum_{t=2}^{T}(\log(\lambda_{s,t}) - \log(\lambda_{s,t-1}))}{\sigma_s^2}. \tag{C.7}$$

For σ_s^2 the full-conditional is $[\sigma_s^2|\cdot] = \mathrm{IG}(\tilde{\alpha}, \tilde{\beta})$, with

$$\tilde{\alpha} \equiv \frac{T-1}{2} + \alpha, \tag{C.8}$$

$$\tilde{\beta} \equiv \left(\frac{\sum_{t=2}^{T}(\lambda_{s,t} - \phi_s - \lambda_{s,t-1})^2}{2} + \frac{1}{\beta} \right)^{-1}. \tag{C.9}$$

For $\log(\lambda_{s,t})$ the full-conditionals are as follows. For $t = 1$, the full-conditional is

$$[\log(\lambda_{s,1})|\cdot] \propto [y_{s,1}|\lambda_{s,1}][\log(\lambda_{s,2})|\phi_s, \sigma_s^2, \log(\lambda_{s,1})][\log(\lambda_{s,1})], \tag{C.10}$$

for $t = 2, \ldots, T-1$, the full-conditional is

$$[\log(\lambda_{s,t})|\cdot] \propto [y_{s,t}|\lambda_{s,t}][\log(\lambda_{s,t+1})|\phi_s, \sigma_s^2, \log(\lambda_{s,t})] \tag{C.11}$$

$$[\log(\lambda_{s,t})|\phi_s, \sigma_s^2, \log(\lambda_{s,t-1})], \tag{C.12}$$

and, for $t = T$, the full-conditional is

$$[\log(\lambda_{s,T})|\cdot] \propto [y_{s,T}|\lambda_{s,T}][\log(\lambda_{s,T})|\phi_s, \sigma_s^2, \log(\lambda_{s,T-1})]. \tag{C.13}$$

We used random walk Metropolis proposals for the log intensity state variables such that $\log(\lambda_{s,t})^{(*)} \sim \mathrm{N}(\log(\lambda_{s,t})^{(k-1)}, \sigma_{\text{tune}}^2)$.

Funding

This research was funded by NSF EF 1241856, NSF DMS 1614392, and NSF DEB 1927177. Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

References

Albert, J. H., and Chib, S. (1993), "Bayesian Analysis of Binary and Polychotomous Response Data," *Journal of the American Statistical Association*, 88, 669–679. [188]

Altman, M., Wood, A., O'Brien, D. R., and Gasser, U. (2018), "Practical Approaches to Big Data Privacy Over Time," *International Data Privacy Law*, 8, 29–51. [187]

Arab, A., Hooten, M. B., and Wikle, C. K. (2017), "Hierarchical Spatial Models," in *Encyclopedia of Geographical Information Science* (2nd ed.), Boston, MA: Springer. [188]

Banerjee, S., Gelfand, A. E., Finley, A. O., and Sang, H. (2008), "Gaussian Predictive Process Models for Large Spatial Data Sets," *Journal of the Royal Statistical Society*, Series B, 70, 825–848. [191]

Barbian, M. H., and Assuncao, R. M. (2017), "Spatial Subsemble Estimator for Large Geostatistical Data," *Spatial Statistics*, 22, 68–88. [192]

Berliner, L. M. (1996), "Hierarchical Bayesian Time Series Models," in *Maximum Entropy and Bayesian Methods*, eds. K. Hanson and R. Silver, Dordrecht, The Netherlands: Kluwer Academic, pp. 15–22. [185]

Bradford, R., and Thomas, A. (1996), "Markov Chain Monte Carlo Methods for Family Trees Using a Parallel Processor," *Statistics and Computing*, 6, 67–75. [185]

Brockwell, A. E. (2006), "Parallel Markov Chain Monte Carlo Simulation by Pre-Fetching," *Journal of Computational and Graphical Statistics*, 15, 246–261. [185,188]

Chopin, N. (2002), "A Sequential Particle Filter Method for Static Models," *Biometrika*, 89, 539–552. [186]

Chopin, N., Jacob, P. E., and Papaspiliopoulos, O. (2013), "SMC^2: An Efficient Algorithm for Sequential Analysis of State Space Models," *Journal of the Royal Statistical Society*, Series B, 75, 397–426. [186]

Cressie, N. A. C. (1993), *Statistics for Spatial Data* (revised edition), New York: Wiley. [188]

Cressie, N. A. C., and Johannesson, G. (2008), "Fixed Rank Kriging for Very Large Spatial Data Sets," *Journal of the Royal Statistical Society*, Series B, 70, 209–226. [192]

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. [191]

Dietze, M., Fox, A., Beck-Johnson, L., Betancourt, J. L., Hooten, M. B., Jarnevitch, C., Kiett, T., Kenney, M., Laney, C., Larsen, L., Loescher, H., Lunch, C., Pijanowski, B., Randerson, J., Reid, E., Tredennick, A., Vargas, R., Weathers, K., and White, E. (2018), "Iterative Near-Term Ecological Forecasting: Needs, Opportunities, and Challenges," *Proceedings of the National Academy of Sciences of the United States of America*, 115, 1424–1432. [191]

Del Moral, P., Doucet, A., and Jasra, A. (2006), "Sequential Monte Carlo Samplers," *Journal of the Royal Statistical Society*, Series B, 68, 411–436. [185]

Doucet, A., de Freitas, J. F. G., and Gordon, N. J. (2001), *Sequential Monte Carlo Methods in Practice*, New York: Springer. [185]

Finley, A., Datta, A., Cook, B. C., Morton, D. C., Andersen, H. E., and Banerjee, S. (2018), "Efficient Algorithms for Bayesian Nearest Neighbor Gaussian Processes," arXiv no. 1702.00434v3. [191]

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. [192]

Garrard, G. E., McCarthy, M. A., Vesk, P. A., Radford, J. Q., and Bennett, A. F. (2012), "A Predictive Model of Avian Natal Dispersal Distance Provides Prior Information for Investigating Response to Landscape Change," *Journal of Animal Ecology*, 81, 14–23. [185]

Gelfand, A. E., and Ghosh, S. (2015), "Hierarchical Modeling," in *Bayesian Theory and Applications*, eds. P. Damien, P. Dellaportas, N. G. Polson, and D. A. Stephens, Oxford: Oxford University Press. [185]

Gelfand, A. E., and Schliep, E. M. (2016), "Spatial Statistics and Gaussian Processes: A Beautiful Marriage," *Spatial Statistics*, 18, 86–104. [188]

Gelfand, A. E., and Smith, A. F. (1990), "Sampling-Based Approaches to Calculating Marginal Densities," *Journal of the American Statistical Association*, 85, 398–409. [185,186]

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., and Rubin, D. B. (2013), *Bayesian Data Analysis*, Boca Raton, FL: Chapman & Hall/CRC. [185]

Gentle, J. E. (2007), *Matrix Algebra: Theory, Computations, and Applications in Statistics*, New York: Springer. [189]

Gerber, B. D., Hooten, M. B., Peck, C. P., Rice, M. B., Gammonley, J. H., Apa, A. D., and Davis, A. J. (2018), "Accounting for Location Uncertainty in Azimuthal Telemetry Data Improves Ecological Inference," *Movement Ecology*, 6, 14. [187]

Glynn, P. W., and Heidelberger, P. (1992), "Analysis of Initial Transient Deletion for Parallel Steady-State Simulations," *SIAM Journal on Scientific and Statistical Computing*, 13, 904–922. [185]

Gramacy, R. B., and Apley, D. W. (2015), "Local Gaussian Process Approximation for Large Computer Experiments," *Journal of Computational and Graphical Statistics*, 24, 561–578. [191,192]

Green, P., Latuszynski, K., Pereyra, M., and Robert, C. (2015), "Bayesian Computation: A Summary of the Current State, and Samples Backwards and Forwards," *Statistical Computation*, 25, 835–862. [185]

Guinness, J. (2018), "Permutation and Grouping Methods for Sharpening Gaussian Process Approximations," *Technometrics*, 60, 415–429. [191]

Guttorp, P., and Gneiting, T. (2006), "Studies in the History of Probability and Statistics XLIX: On the Matérn Correlation Family," *Biometrika*, 93, 989–995. [188]

Heaton, M. J., Datta, A., Finley, A. O., Furrer, R., Guhaniyogi, R., Gerber, F., Gramacy, R. B., Hammerling, D., Katzfuss, M., Lindgren, F., and Nychka, D. W. (2019), "A Case Study Competition Among Methods for Analyzing Large Spatial Data," *Journal Agricultural, Biological, and Environmental Statistics*, 24, 398–425. [188,192]

Higdon, D. (2002), "Space and Space-Time Modeling Using Process Convolutions," in *Quantitative Methods for Current Environmental Issues*, New York: Springer. [192]

Hooten, M. B., Buderman, F. E., Brost, B. M., Hanks, E. M., and Ivan, J. S. (2016), "Hierarchical Animal Movement Models for Population-Level Inference," *Environmetrics*, 27, 322–333. [187]

Hooten, M. B., and Hefley, T. J. (2019), *Bringing Bayesian Models to Life*, Boca Raton, FL: Chapman & Hall/CRC. [187,192]

Hooten, M. B., and Johnson, D. S. (2017), "Basis Function Models for Animal Movement," *Journal of the American Statistical Association*, 112, 578–589. [188]

Hooten, M. B., Ross, B. E., and Wikle, C. K. (2012), "Optimal Spatio-Temporal Monitoring Designs for Characterizing Population Trends," in *Design and Analysis of Long-Term Ecological Monitoring Studies*, eds. R. A. Gitzen, J. J. Millspaugh, A. B. Cooper, and D. S. Licht, Cambridge, MA: Cambridge University Press. [192]

Hooten, M. B., Wikle, C. K., Sheriff, S. L., and Rushin, J. W. (2009), "Optimal Spatio-Temporal Hybrid Sampling Designs for Ecological Monitoring," *Journal of Vegetation Science*, 20, 639–649. [192]

Huang, H., and Sun, Y. (2018), "Hierarchical Low Rank Approximation of Likelihoods for Large Spatial Datasets," *Journal of Computational and Graphical Statistics*, 27, 110–118. [191]

Hurlbert, S. H. (1984), "Pseudoreplication and the Design of Ecological Field Experiments," *Ecological Monographs*, 54, 187–211. [187]

Jacob, P. E., Murray, L. M., Holmes, C. C., and Robert, C. P. (2017), "Better Together? Statistical Learning in Models Made of Modules," arXiv no. 1708.08719. [192]

Jeffreys, H. (1946), "An Invariant Form for the Prior Probability in Estimation Problems," *Proceedings of the Royal Society of London,* Series A, *Mathematical and Physical Sciences*, 186, 453–461. [188]

Katzfuss, M. (2017), "A Multi-Resolution Approximation for Massive Spatial Datasets," *Journal of the American Statistical Association*, 112, 201–214. [192]

Katzfuss, M., and Guinness, J. (2018), "A General Framework for Vecchia Approximations of Gaussian Processes," arXiv no. 1708.06302v3. [191]

Kleiner A., Talwalkar, A., Sarkar, P., and Jordan, M. I. (2014), "A Scalable Bootstrap for Massive Data," *Journal of the Royal Statistical Society*, Series B, 76, 795–816. [192]

Liang F., Cheng, Y., Song, Q., Park, J., and Yang, P. A. (2013), "Resampling-Based Stochastic Approximation Method for Analysis of Large Geostatistical Data," *Journal of the American Statistical Association*, 108, 325–339. [192]

Lindgren, F., Rue, H., and Lindstrom, J. (2011), "An Explicit Link Between Gaussian Fields and Gaussian Markov Random Fields: The Stochastic Partial Differential Equation Approach," *Journal of the Royal Statistical Society*, Series B, 73, 423–498. [192]

Lunn D., Barrett, J., Sweeting, M., and Thompson, S. (2013), "Fully Bayesian Hierarchical Modelling in Two Stages, With Application to Meta-Analysis," *Journal of the Royal Statistical Society*, Series C, 62, 551–572. [187,192]

MacLaurin, D., and Adams, R. P. (2015), "Firefly Monte Carlo: Exact MCMC With Subsets of Data," in *Proceedings of the Conference on Uncertainty in Artificial Intelligence*. [192]

Matérn, B. (1986), *Spatial Variation* (2nd ed.), Berlin: Springer-Verlag. [188]

McCarthy, M. A., and Masters, P. I. P. (2005), "Profiting From Prior Information in Bayesian Analyses of Ecological Data," *Journal of Applied Ecology*, 42, 1012–1019. [185]

Neal, R. M. (2011), "MCMC Using Hamiltonian Dynamics," in *Handbook of Markov Chain Monte Carlo*, eds. S. Brooks, A. Gelman, G. L. Jones, and X.-L. Meng, New York: CRC Press. [185]

Neiswanger, W., Wang, C., and Xing, E. (2013), "Asymptotically Exact, Embarrassingly Parallel MCMC," arXiv no. 1311.4780. [192]

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. [192]

Patil G., and Rao, C. (1977), "The Weighted Distributions: A Survey of Their Applications," in *Applications of Statistics*, ed. P Krishnaiah, Amsterdam: North Holland Publishing Company. [187]

Rosenthal, J. S. (2000), "Parallel Computing and Monte Carlo Algorithms," *Far East Journal of Theoretical Statistics*, 4, 207–236. [185]

Särkkä, S. (2013), *Bayesian Filtering and Smoothing*, Cambridge, MA: Cambridge University Press. [186]

Schifano E. D., Wu, J., Wang, C., Yan, J., and Chen, M. H. (2016), "Online Updating of Statistical Inference in the Big Data Setting," *Technometrics*, 58, 393–403. [190]

Scott, S. L., Blocker, A. W., Bonassi, F. V., Chipman, H. A., George, E. I., and McCulloch, R. E. (2016), "Bayes and Big Data: The Consensus Monte Carlo Algorithm," *International Journal of Management Science and Engineering Management*, 11, 78–88. [192]

Stein, M. L., Chi, Z., and Welty, L. J. (2004), "Approximating Likelihoods for Large Spatial Data Sets," *Journal of the Royal Statistical Society*, Series B, 66, 275–296. [191]

Vecchia, A. (1988), "Estimation and Model Identification for Continuous Spatial Processes," *Journal of the Royal Statistical Society*, Series B, 50, 297–312. [191]

Wang, C., Chen, M. H., Wu, J., Yan, J., Zhang, Y., and Schifano, E. (2018), "Online Updating Method With New Variables for Big Data Streams," *Canadian Journal of Statistics*, 46, 123–146. [190]

Wang, X., and Dunson, D. B. (2013), "Parallelizing MCMC via Weierstrass Sampler," arXiv no. 1312.4605. [192]

Wikle, C. K. (2010), "Low-Rank Representations for Spatial Processes," in *Handbook of Spatial Statistics*, eds. A. Gelfand, P. Diggle, M. Fuentes, and P. Guttorp, Boca Raton, FL: Chapman & Hall/CRC. [192]

Wikle, C. K., and Royle, J. A. (1999), "Space-Time Dynamic Design of Environmental Monitoring Networks," *Journal of Agricultural, Biological and Environmental Statistics*, 4, 489—507. [192]

——— (2005), "Dynamic Design of Ecological Monitoring Networks for Non-Gaussian Spatio-Temporal Data," *Environmetrics*, 16, 507–522. [192]

Williams, P. J., Hooten, M. B., Womble, J. N., Esslinger, G. G., and Bower, M. R. (2018), "Monitoring Dynamic Spatio-Temporal Ecological Processes Optimally," *Ecology*, 99, 524–535. [192]

Zimmerman, D. L. (2006), "Optimal Network Design for Spatial Prediction, Covariance Parameter Estimation, and Empirical Prediction," *Environmetrics*, 17, 635–652. [191]