# Bayesian Modeling for Large Spatial Datasets

**Sudipto Banerjee [Professor]** and
Division of Biostatistics, University of Minnesota, School of Public Health, A460 Mayo Building, MMC 303, 420 Delaware Street S.E., Minneapolis MN 55455, baner009@umn.edu, Phone: 612.624.0624, Fax: 612.626.0660

**Montserrat Fuentes [Professor]**
Department Head, Statistics, North Carolina State University, SAS Hall 5105, Box 8203, 2311 Stinson Drive, Raleigh, NC 27695-8203, Fuentes@ncsu.edu, Phone: 919-515-0610, Fax: 919-515-7591

## 1 Spatial process models for point-referenced data

With the emergence of highly efficient Geographical Information Systems (GIS) databases and associated software, the modeling and analysis of spatially referenced data sets have received much attention over the last decade. Spatially-referenced data sets and their analysis using GIS arise in diverse areas of scientific and engineering investigations including geological and environmental sciences (Webster and Oliver, 2001), ecological systems (Scheiner and Gurevich, 2001), digital terrain cartography (Jones, 1997), computer experiments (Santner et al., 2003), public health (Cromley and McLafferty, 2002) and so on. A wonderful compilation of current research trends in spatial statistics is presented by Diggle et al. (2010).

Two base units of measure and mapping are commonly encountered: locations that are areas or regions with well-defined neighbors (such as pixels in a lattice, counties in a map, etc.), whence they are called *areally referenced* data; or locations that are points with coordinates (latitude-longitude, Easting-Northing etc.), in which case they are called *point referenced* or *geostatistical.* Statistical theory and methods play a crucial role in the modeling and analysis of such data by developing spatial process models, also known as stochastic process or random function models, that help in predicting and estimating physical phenomena. This proposal deals with the latter: modeling of point-referenced data. The last two decades has seen significant developments in such modeling; see, for example, the books by Cressie (1993), Chilés and Delfiner (1999), Møller (2003), Schabenberger and Gotway (2004), and Banerjee, Carlin and Gelfand (2004) for a variety of methods and applications.

Spatial process modeling envisions a *random surface y(s)* representing some variable (e.g., temperature, precipitation) that conceptually exists in continuum over the domain $s \in D$ but has been observed only at a finite set of locations and must be interpolated (or predicted) for other arbitrary locations. By interpolating at arbitrarily fine resolutions, these models *estimate* the random surface accounting for correlation between temperature and precipitation levels at locations closer to each other and produce a *response surface* for the dependent variable. Such interpolation from statistical models is often referred to as "kriging" and the response surfaces are called kriged surfaces.

## 2 Model-based kriging

The most common geostatistical setting assumes a response or dependent variable $y(s)$ observed at a generic location $s$ along with a $p \times 1$ vector of spatially referenced predictors

$x(s)$. Model-based geostatistical data analysis often proceeds from spatial regression models such as,

$$y(s) = x(s)'\beta + w(s) + \epsilon(s). \quad (1)$$

The residual from the regression is partitioned into a spatial process, $w(s)$, capturing spatial association, and an independent process, $\epsilon(s) \overset{iid}{\sim} N(0, \tau^2)$, also known as the *nugget* effect, modeling measurement error or *micro-scale* variation. Micro-scale variation is often modeled as a stationary spatial disturbance at a scale lesser than the minimum inter-site distance. Often modelers separate the micro-scale variation by inserting an $\eta(s)$ into (1), whence $\eta(s)$ captures micro-scale variation and $\epsilon(s)$ captures pure measurement error (see, e.g. Cressie, 1993, p.112).

With observations $y = (y(s_1),\ldots, y(s_n))'$ from $n$ locations, the data are treated as a partial realization of a spatial process modeled through $w(s)$. The most popular specification $w(s) \sim GP(0, C(\cdot,\cdot))$ is a zero-centered Gaussian Process determined by a valid covariance function $C(s_i, s_j)$ defined for pairs of sites $s_i$ and $s_j$ Typically the modeler specifies $C(s_1, s_2) = \sigma^2 \rho(s_1, s_2; \theta)$ where $\rho(\cdot; \theta)$ is a correlation function and $\theta$ includes parameters quantifying the rate of correlation decay and the smoothness of realizations. The data likelihood is given by $y \sim N(X\beta, \Sigma_y)$, where $X$ is an $n \times p$ matrix of regressors whose $i$-th row is $x(s_i)'$, $\Sigma_y = \sigma^2 R(\theta) + \tau^2 I_N$ and $R(\theta)$ is an $n \times n$ spatial correlation matrix with $\rho(s_i, s_j; \theta)$ as its $(i, j)$-th element. Likelihood-based inference proceeds by computing estimates from maximum likelihood (ML), restricted maximum likelihood (REML), or Generalized Estimating Equation (GEE) approaches and investigating their consistency and asymptotic properties. The books by Cressie (1993) and Schabenberger and Gotway (2004) provide excellent expositions of such methods.

We conclude with a few comments on *stationarity* of spatial processes. Stationarity is a common assumption when modeling spatial processes. A random field $w(s)$ is called *weakly stationary* (or *stationary*) if it has finite second moments, its mean function is constant and its covariance function $C(s_1, s_2) = C(h)$, where $h = s_1 - s_2$. This implies that the relationship between the values of the process at two locations only depends on the vector distance between these two locations. An *isotropic* process results when $C(s_1, s_2) = C(\|h\|)$, where $\|h\|$ is the distance between $s_1$ and $s_2$. Stationarity is a desirable property for spatial processes, but is usually a rather unrealistic assumptions for large spatial domains. There is an extensive literature on nonstationary models, e.g., Sampson and Guttorp, (1992), Nychka et al., (2002), Higdon et al., (1999), and Fuentes, (2002). Fuentes, (2005) introduces a nonparametric test for stationary.

## 3 Hierarchical models for spatial data

A different approach that has recently gaining popularity among spatial modelers follows the Bayesian inferential paradigm (Robert, 2001; Gelman et al., 2004; Carlin and Louis, 2008; Banerjee et al., 2004). Here one constructs hierarchical (or multi-level) schemes by assigning probability distributions to parameters *a priori* and inference is based upon the distribution of the parameters conditional upon the data *a posteriori*. By modeling both the observed data and any unknown regressor or covariate effects as random variables, the hierarchical Bayesian approach to statistical analysis provides a cohesive framework for combining complex data models and external knowledge or expert opinion.

Bayesian hierarchical models (e.g. Gelman et al. 2004; Banerjee et al., 2004) are widely recognized as versatile inferential tools for capturing the rich dependence structures underlying spatial data and offering full inference on spatial processes. For an $n \times 1$ vector

of observed outcomes, $\boldsymbol{y} = (y(s_1), y(s_2), \ldots, y(s_n))'$ with a first stage conditionally independent Gaussian specification and associated priors, we construct a Bayesian hierarchical model

$$p(\theta, \beta, \boldsymbol{w}|\boldsymbol{y}) \propto p(\theta) \times N(\beta|\mu_\beta, \Sigma_\beta) \times N(\boldsymbol{w}|\boldsymbol{0}, \boldsymbol{C}(\theta_1)) \times \prod_{i=1}^{n} N(y(s_i)|\boldsymbol{x}(s_i)'\beta + w(s_i), \tau^2), \quad (2)$$

where $\boldsymbol{\theta} = \{\boldsymbol{\theta}_1, \tau^2\}$;. The parameter $\tau^2$ is called the *nugget* and captures unstructured noise that may arise in the form of measurement error or micro-scale variability. Customarily, either a flat or a multivariate Gaussian prior is assigned to β. Zhang (2004) demonstrated, rather remarkably, that the process parameters $\boldsymbol{\theta}_1$ were not consistently (in the classical sense) estimable for a rather general class of covariance functions. This implies that the priors' inferential impact is not obliterated with increasing sample size; hence, weakly informative priors will be needed to identify the process parameters.

Spatial data analysis seeks to estimate the regression coefficients β, the unknown process parameters $\boldsymbol{\theta} = \{\boldsymbol{\theta}_1, \tau^2\}$, which convey the nature of spatial associations and micro-scale variability, and the spatial effects $\boldsymbol{w}$ which elicit lurking spatial patterns in the residual. Estimating (2) customarily proceeds using Markov chain Monte Carlo (MCMC) methods (e.g. Robert and Casella, 2004). With Gaussian likelihoods, often we integrate out the spatial effects $\boldsymbol{w}$. This replaces the likelihood and the prior for $\boldsymbol{w}$ by $N(\boldsymbol{y} \mid \boldsymbol{X}\beta, \boldsymbol{C}(\boldsymbol{\theta}_1) + \tau^2 \boldsymbol{I}_n)$. In any case, estimation involves $n \times n$ matrix decompositions of cubic order in the number of locations, which become exorbitant for large $n$. Evidently, multivariate and spatial-temporal settings aggravate the situation.

As modern data technologies acquire and exploit massive amounts of data, statisticians analyzing spatially referenced datasets confront settings where the number of geo-referenced locations is very large. This makes hierarchical modeling infeasible or impractical. The situation is further exacerbated in multivariate settings with several spatially dependent response variables, where the matrix dimensions increase by a factor of the number of spatially dependent variables being modeled. It is also aggravated when data is collected at frequent time points and spatiotemporal process models are used.

## 4 Hierarchical models for large spatial datasets

Modelling large spatial datasets have received much attention in the recent past. Vecchia (1988) proposed approximating the likelihood with a product of appropriate conditional distributions to obtain maximum-likelihood estimates. Stein et al. (2004) adapt this effectively for restricted maximum likelihood estimation. Another possibility is to approximate the likelihood using spectral representations of the spatial process (Fuentes, 2007). These likelihood approximations yield a joint distribution, but not a process that facilitate spatial interpolation. Another concern is the adequacy of the resultant likelihood approximation. Expertise in tailoring and tuning of a suitable spectral density estimate or a sequence of conditional distributions is required and does not easily adapt to multivariate processes.

Also, the spectral density approaches seem best suited to stationary covariance functions. Yet another approach considers compactly supported correlation functions (Furrer et al., 2006; Kaufman et al., 2008; Du et al., 2009) that yeild sparse correlation structures. More efficient sparse solvers can then be employed for kriging and variance estimation, but the tapered structures may limit modeling flexibility. Also, full likelihood-based inference still requires determinant computations that may be problematic. Another approach either replaces the process (random field) model by a *Markov* random field (Cressie, 1993) or else approximates the random field model by a Markov random field (Rue and Tjelmeland, 2002; Rue and Held 2006). This approach is best suited for points on a regular grid. With

irregular locations, realignment to a grid or a torus is required, done by an algorithm, possibly introducing unquantifiable errors in precision.

In recent work Rue, Martino and Chopin (2009) propose a promising INLA (Integrated Nested Laplace Approximation) algorithm as an alternative to MCMC that utilizes the sparser matrix structures to deliver fast and accurate posterior approximations. This uses conditional independence to achieve sparse spatial precision matrices that considerably accelerate computations, but relaxing this assumption would significantly detract from the computational benefits of the INLA and the process needs to be *approximated* by a Gaussian Markov Random Field (GMRF) (Rue and Held, 2006). Furthermore, the method involves a mode-finding exercise for hyper-parameters that may be problematic when the number of hyperparameters is more than 10. Briefly, its effectiveness is unclear for our multivariate genetic models with different structured random effects and unknown covariance matrices as hyperparameters.

Adapting the above approaches to more complex hierarchical spatial models involving multivariate processes (e.g. Wackernagel, 2006; Gelfand et al, 2004), spatiotemporal processes and spatially varying regressions (Gelfand et al., 2003) and nonstationary covariance structures (Paciorek and Schervish, 2006) is potentially problematic.

A versatile approach pursues models especially geared towards the handling of large spatial datasets. Typically, these emerge from representations of the spatial process in lower-dimensional subspaces and easily generalize to multivariate and/or spatiotemporal processes. These are often referred to as low-rank or reduced-rank spatial models and have been explored in different contexts (Wikle and Cressie, 1999; Lin et al., 2000; Higdon 2002; Kamman and Wand, 2003; Ver Hoef et al. 2004; Paciorek, 2007; Stein, 2007, 2008; Cressie and Johannesson, 2008; Banerjee et al., 2008, 2010; Crainiceaniu et al., 2008; Latmer et al. 2009). Many of these methods are variants of the so-called "subset of regressors" methods used in Gaussian process regressions for large data sets in machine learning (e.g. Rasmussen and Williams, 2006). The idea here is to assume that the spatial information available from the entire set of observed locations can be summarized in terms of a smaller, but representative, sets of locations, or "knots". When retaining richness and flexibility of hierarchical models is of primary interest, knot-based low rank models seem to be the preferred option.

Sun et al. (2011) offer a very detailed review of statistical methods for massive spatial datasets. Here, we consider one low-rank approach, referred to as Gaussian predictive process models, in some detail. Detailed descriptions of hierarchical Gaussian predictive process models are given in Banerjee et al. (2008), (2010) and Finley et al. (2009). Here, we offer a brief review.

Predictive process models use a fixed set of "knots" $\mathscr{S}^* = (s_1^*, \ldots, s_{n^*}^*)$ with $n^* << n$, which may, but need not, be a subset of $\mathscr{S}$ An optimal projection of the process $w(s)$ at a generic location $s$, based upon its realization over $\mathscr{S}$, is given by the "kriging equation" $\tilde{w}(s) = \mathrm{E}\{w(s) \mid w^*\}$, where $w^* = (w(s_1^*), w(s_2^*), \ldots, w(s_n^*))'$. We refer to $\tilde{w}(s)$ as the *predictive process* derived from the *parent process $w(s)$*. Banerjee et al. (2008) discuss several theoretical properties of the predictive process, including multivariate and spatiotemporal extensions.

For a zero-centered parent Gaussian process, say $w(s)$, with covariance function $C(s_1, s_2; \theta_1)$, the predictive process is $\tilde{w}(s) = E\{w(s) \mid w^*\} = c(s, \theta_1)' C^*(\theta_1)^{-1} w^*$, where $c(s; \theta_1)'$ is the $1 \times n$ vector whose $j$-th element is $C(s, s_j^*; \theta_1)$ and $C^*(\theta_1)$ is the $n^* \times n^*$ covariance

matrix with elements $C(s_i^*, s_j^*; \theta_1)$. Since $w^*$ is multivariate normal with zero mean and $n^* \times n^*$ dispersion matrix $C^*(\theta_1)$, $\tilde{w}(s)$ is itself a nonstationary Gaussian process arising from a spatially adaptive linear transformation of the parent process over the set of knots. Replacing $w(s)$ with $\tilde{w}(s)$ in (2), yields its predictive process counterpart

$$p(\theta, \beta, w^* | y) \propto p(\theta) \times N(\beta | \mu_\beta, \Sigma_\beta) \times N(w^* | \mathbf{0}, C^*(\theta_1)) \times \prod_{i=1}^{n} N(y(s_i) | x(s_i)'\beta + \tilde{w}(s_i), \tau^2). \quad (3)$$

Computational gains are achieved since matrix computations now involve the $n^* \times n^*$ matrix $C^*(\theta_1)$, where $n^*$ is chosen to be much smaller than $n$. Unlike several other knot-based approaches, the predictive process does not introduce additional parameters nor does it involve projecting data onto a grid. Thus, it avoids identifiability issues or spurious loss of uncertainty (see, e.g., Banerjee et al. 2008). Indeed, predictive process models are attractive since they are directly induced by the parent process without requiring choices of basis functions or kernels or alignment algorithms for the locations.

Being a low-rank process, the partial realizations of the predictive process produce a low-rank (degenerate) likelihood. Being smoother than the parent process, it tends to have lower variance which, in turn, inflates the residual variability often manifested as an overestimation of $\tau^2$. In fact, the following inequality holds for any fixed $\mathscr{S}$ and for any spatial process $w(s)$:

$$\text{var}\{w(s)\} = \text{var}\{E[w(s)|w^*]\} + E\{\text{var}[w(s)|w^*]\} \geq \text{var}\{E[w(s)|w^*]\} = \text{var}\{\tilde{w}(s)\}. \quad (4)$$

If the process is Gaussian then standard multivariate normal calculations yields the following closed form for this difference for site $s$,

$$\delta^2(s) = E\{\text{var}[w(s)|w^*]\} = C(s, s; \theta_1) - c(s, \theta_1)' C^*(\theta_1)^{-1} c(s, \theta_1). \quad (5)$$

One simple remedy for the bias in the predictive process model (Finley et al., 2009; Banerjee et al. 2010) is to use the process $\tilde{w}_\epsilon(s) = \tilde{w}(s) + \tilde{\epsilon}(s)$, where $\tilde{\epsilon}(s) \overset{iid}{\sim} N(0, \delta^2(s))$ and $\tilde{\epsilon}(s)$ is independent of $\tilde{w}(s)$. We call this the *modified predictive process*. Now, the variance of $\tilde{w}_\epsilon(s)$ equals that of the parent process $w(s)$ and the remedy is computationally effective – adding an independent space-varying nugget does not incur substantial computational expenses. However, the remedy is less effective in capturing small-scale spatial variation as it does not account for the spatial dependence in the residual process.

Hierarchical spatial models, including the predictive process models, are customarily estimated using Markov chain Monte Carlo (MCMC) methods (e.g. Robert and Casella, 2004). Further technical details on MCMC algorithms for hierarchical spatial models can be found in Banerjee et al. (2004) and particularly for predictive process models in Banerjee et al. (2008) and Finley et al. (2009).

## 5 Illustration

We present a brief illustration of hierarchical predictive process model using a forest biomass dataset recently analyzed by Guhaniyogi et al. (2011). Spatial modeling of forest biomass and other variables related to measurements of current carbon stocks and flux have recently attracted much attention for quantifying the current and future ecological and economic viability of forest landscapes. Interest often lies in detecting how biomass changes across the landscape (as a continuous surface) and how homogeneous it is across the region. In the United States, the Forest Inventory and Analysis (FIA) program of the USDA Forest Service collects the data needed to support these assessments. Here, we present a subset of

the analysis in Guhaniyogi et al. (2011). Our outcome variable is log metric tons of forest biomass per hectare. A July, 2003 mosaic of Landsat TM imagery, was used to calculate tasseled cap components of brightness (TC1), greenness (TC2), and wetness (TC3) to serve as explanatory variables. Figure 1(a) illustrates the georeferenced forest inventory data consisting of 6,538 forested FIA plots measured between 1999 and 2006 across the lower peninsula of Michigan.

Candidate models included a simple non-spatial regression and a predictive process model. Knot locations were chosen by applying the `k-means` clustering algorithm to the observed locations. We assumed that the parent covariance function is exponential: $C(s_1, s_2; \theta_1) = \sigma^2 \exp(-\phi\|s_1 - s_2\|)$; therefore, $\theta_1 = \{\sigma^2, \phi\}$. Based on results from an initial variogram analysis of the non-spatial model's residuals, the priors for $\tau^2$ and $\sigma^2$ for the predictive process models followed *IG*(2 0.5) distributions (mean equaling 0.5), while the prior for the spatial decay parameter $\phi$ followed a *U*(0.006 3), which corresponds to support from 1–500 km. This is a broad range of support, given the maximum distance between any two plots is 460 km. For all models the regression coefficients each received a *flat* prior, i.e. $\Sigma_\beta^{-1} = O$ a matrix of zeroes.

Posterior inference was based on three chains of 25, 000 iterations (the first 5,000 iterations were discarded as burn-in). The samplers were coded in C++ and Fortran and leveraged Intel's Math Kernel Library threaded BLAS and LAPACK routines for matrix computations. All analyses were conducted on a Linux workstation using two Intel Nehalem quad-Xeon processors. Table 1 offers parameter estimates for the predictive process candidate models. Over the range of knot intensities, both the non-spatial and predictive process models produce comparable estimates of the regression coefficients – several of which explain a significant amount of variability in log biomass. The predictive process model seems to estimate a larger nugget, $\tau^2$, and a smaller partial sill, $\sigma^2$. Figure 1(b) is an interpolated surface of the non-spatial model residuals. We would expect the fitted spatial random effects of the candidate models to look somewhat similar to this residual surface, which is evident from Figure 1(c).

## 6 Summary and discussion

We have offered a brief overview of certain recent developments in the problem of fitting desired Bayesian hierarchical spatial modeling specifications to large datasets. To do so, we propose simply replacing the parent spatial process specification by its induced predictive process specification. One need not digress from the modelling objectives to think about choices of basis functions or kernels or alignment algorithms for the locations. The resulting class of models essentially falls under the setup of hierarchical generalized linear mixed models.

As in existing low-rank kriging approaches, knot selection is required and some sensitivity to the number of knots is expected (see Finley, Sang, Banerjee and Gelfand, 2009, for details). With a fairly even distribution of data locations, one possibility is to select knots on a uniform grid overlaid on the domain. Alternatively, selection can be achieved through a formal design-based approach based upon minimization of a spatially averaged predictive variance criterion (e.g., Diggle and Lophaven, 2006). However, in general the locations are highly irregular, generating substantial areas of sparse observations where we wish to avoid placing knots, since they would be "wasted" and possibly lead to inflated predictive process variances and slower convergence. Here, more practical space-covering designs (e.g., Royle and Nychka, 1998) can yield a representative collection of knots that better cover the domain. Another alternative is to apply popular clustering algorithms such as `k-means` or more robust median-based *partitioning around medoids* algorithms (e.g., Kaufman and

Rousseeuw, 1990). User-friendly implementations of these algorithms are available in `R` packages such as `fields` and `cluster` and have been used in spline-based low-rank kriging models (Ruppert et al. 2003). While for most applications a reasonable grid of knots should lead to robust inference, with fewer knots the separation between them increases and estimating random fields with fine-scale spatial dependence become difficult. Indeed, learning about fine scale spatial dependence is always a challenge (see, e.g., Cressie, 1993, p.114).

It is also not uncommon to find space-time datasets with a very large number of distinct time points, possibly with different time points observed at different locations (e.g., real estate transactions). Predictive processes can be used to improve the applicability of a class of dynamic space-time models proposed by Gelfand et al. (2005) by alleviating a computational bottleneck without sacrificing model flexibility and with minimal loss of information. Recently, Finley et al. (2011) focused on the common setting where space is considered continuous but time is taken to be discrete. Here, data is viewed as arising from a time series of spatial processes. Some examples of data that fit this description include: US Environmental Protection Agency's Air Quality System which reports pollutants' mean, minimum, and maximum at 8 and 24 hour intervals; climate model outputs of weather variables generated on hourly or daily intervals, and; remotely sensed landuse/landcover change recorded at annual or decadal time steps.

Finally, we conclude with some remarks on computing. With multiple processors, substantial gains in computing efficiency can be realized through parallel processing of matrix operations. We intend continued migration of our lower-level C++ code to the existing `spBayes` (http://cran.r-project.org; Finley et. al. 2007) package in the `R` environment to facilitate accessibility to predictive process models.

# REFERENCES

Banerjee, S.; Carlin, BP.; Gelfand, AE. Hierarchical Modeling and Analysis for Spatial Data. Boca Raton: FL: Chapman and Hall/CRC Press; 2004.

Banerjee S, Finley AO, Waldmann P, Ericsson T. Hierarchical spatial process models for multiple traits in large genetic trials. Journal of the American Statistical Association. 2010; 105:506–521. [PubMed: 20676229]

Banerjee S, Gelfand AE, Finley AO, Sang H. Gaussian predictive process models for large spatial datasets. Journal of the Royal Statistical Society Series B. 2008; 70:825–848.

Carlin, BP.; Louis, TA. Bayesian Methods for Data Analysis. Third edition. Boca Raton: FL: Chapman and Hall/CRC Press; 2008.

Chilés, JP.; Delfiner, P. Geostatistics: Modelling Spatial Uncertainty. New York: John Wiley and Sons; 1999.

Crainiceanu CM, Diggle PJ, Rowlingson B. Bivariate binomial spatial modeling of Loa loa Prevalence in tropical Africa (with discussion). Journal of the American Statistical Association. 2008; 103:21–37.

Cressie, NAC. Statistics for Spatial Data. 2nd edition. New York: Wiley; 1993.

Cressie NAC, Johannesson G. Spatial prediction for massive datasets. Journal of the Royal Statistical Society Series B. 2008; 70:209–226.

Du J, Zhang H, Mandrekar VS. Fixed-domain asymptotic properties of tapered maximum likelihood estimators. Annals of Statistics. 2009; 37:3330–3361.

Finley AO, Banerjee S, Carlin BP. spBayes: an R package for univariate and multivariate hierarchical point-referenced spatial models. Journal of Statistical Software. 2007; 19:4.

Finley AO, Banerjee S, Gelfand AE. Bayesian dynamic modeling for large spacetime datasets using Gaussian predictive processes. Journal of Geographical Information Systems. 2011 (submitted).

Finley AO, Sang H, Banerjee S, Gelfand AE. Improving the performance of predictive process modeling for large datasets. Computational Statistics and Data Analysis. 2009; 53:2873–2884. [PubMed: 20016667]

Fuentes M. Periodogram and other spectral methods for nonstationary spatial processes. Biometrika. 2002; 89:197–210.

Fuentes M. A formal test for nonstationarity of spatial stochastic processes. Journal of Multivariate Analysis. 2005; 96:30–55.

Fuentes M. Approximate likelihood for large irregularly spaced spatial data. Journal of the American Statistical Association. 2007; 102:321–331. [PubMed: 19079638]

Furrer R, Genton MG, Nychka D. Covariance tapering for interpolation of large spatial datasets. Journal of Computational and Graphical Statistics. 2006; 15:502–523.

Gelfand AE, Banerjee S, Gamerman D. Univariate and multivariate dynamic spatial modelling. Environmetrics. 2005; 16:465–479.

Gelfand AE, Kim H, Sirmans CF, Banerjee S. Spatial modeling with spatially varying coefficient processes. Journal of the American Statistical Association. 2003; 98:387–396.

Gelfand AE, Schmidt A, Banerjee S, Sirmans CF. Nonstationary multivariate process modelling through spatially varying coregionalization. Test. 2004; 13:263–312.

Gelman, A.; Carlin, JB.; Stern, HS.; Rubin, DB. Bayesian Data Analysis. 2nd edition. Boca Raton: FL: Chapman and Hall/CRC Press; 2004.

Harville, DA. Matrix Algebra from a Statistician's Perspective. New York: Springer; 1997.

Henderson HV, Searle SR. On deriving the inverse of a sum of matrices. SIAM Review. 1981; 23:53–60.

Higdon, D. Space and space-time modeling using process convolutions. In: Anderson, C.; Barnett, V.; Chatwin, PC.; El-Shaarawi, AH., editors. In Quantitative methods for current environmental issues. Springer-Verlag; 2002. p. 37-56.

Higdon, D.; Swall, J.; Kern, J. Bayesian Statistics. Vol. 6. Oxford University Press; 1999. Non-Stationary Spatial Modeling.

Jones, CB. Geographical Information Systems and Computer Cartography. Harlow, Essex, UK: Addison Wesley Longman; 1997.

Kamman EE, Wand MP. Geoadditive models. Journal of the Royal Statistical Society C (Applied Statistics). 2003; 52:1–18.

Kaufman CG, Schervish MJ, Nychka DW. Covariance tapering for likelihood-based estimation in large spatial data sets. Journal of the American Statistical Association. 2009; 103:1545–1555.

Latimer AM, Banerjee S, Sang H, Mosher E Jr, Silander JA. Hierarchical models for spatial analysis of large data sets: A case study on invasive plant species in the northeastern United States. Ecology Letters. 2009; 12:144–154. [PubMed: 19143826]

Lin X, Wahba G, Xiang D, Gao F, Klein R, Klein B. Smoothing spline ANOVA models for large data sets with Bernoulli observations and the randomized GACV. Annals of Statistics. 2000; 28:1570–1600.

Møller, J. Spatial Statistics and Computational Method. New York: Springer; 2003.

Nychka D, Wikle CK, Royle JA. Multiresolution models for nonstationary spatial covariance functions. Statistical Modelling: An International Journal. 2002; 2:315–331.

Paciorek CJ. Computational techniques for spatial logistic regression with large datasets. Computational Statistics and Data Analysis. 2007; 51:3631–3653. [PubMed: 18443645]

Paciorek CJ, Schervish MJ. Spatial modelling using a new class of nonstationary covariance functions. Environmetrics. 2006; 17:483–506. [PubMed: 18163157]

Rasmussen, CE.; Williams, CKI. Gaussian Processes for Machine Learning. Cambridge: MA: MIT Press; 2006.

Robert, C. The Bayesian Choice. 2nd edition. New York: Springer; 2001.

Robert, CP.; Casella, G. Monte Carlo Statistical Methods. New York: Springer-Verlag;

Rue, H.; Held, L. Gaussian Markov Random Fields: Theory and Applications. Boca Raton: FL: Chapman and Hall/CRC Press; 2006.

Rue H, Tjelmeland H. Fitting Gaussian Markov random fields to Gaussian fields. Scandinavian
   Journal of Statistics. 2002; 29:31–49.

Rue H, Martino S, Chopin N. Approximate Bayesian inference for latent Gaussian models by using
   integrated nested Laplace approximations (with discussion). Journal of the Royal Statistical
   Society, Series B. 2009; 71:1–35.

Sampson PD, Guttorp P. Nonparametric estimation of nonstationary spatial covariance structure.
   Journal of the American Statistical Association. 1992; 87:108–119.

Santner, TJ.; Williams, BJ.; Notz, WI. The Design and Analysis of Computer Experiments. Springer;
   2003.

Schabenberger, O.; Gotway, CA. Statistical Methods for Spatial Data Analysis (Texts in Statistical
   Science Series). Boca Raton: Chapman and Hall/CRC; 2004.

Scheiner, SM.; Gurevitch, J. Design and Analysis of Ecological Experiments. 2nd edition. Oxford,
   UK: Oxford University Press; 2001.

Stein, ML. Interpolation of Spatial Data: some theory for kriging. New York: Springer; 1999.

Stein ML, Chi Z, Welty LJ. Approximating likelihoods for large spatial datasets. Journal of the Royal
   Statistical Society, Series B. 2004; 66:275–296.

Stein ML. Spatial variation of total column ozone on a global scale. Annals of Applied Statistics.
   2007; 1:191–210.

Stein ML. A modeling approach for large spatial datasets. Journal of the Korean Statistical Society.
   2008; 37:3–10.

Sun, Y.; Li, B.; Genton, MG. Geostatistics for large datasets. In: Montero, JM.; Porcu, E.; Schlather,
   M., editors. Advances And Challenges In Space-time Modelling Of Natural Events. Springer:
   2011. (to appear).

Vecchia AV. Estimation and model identification for continuous spatial processes. Journal of the
   Royal Statistical Society Series B. 1988; 50:297–312.

Ver Hoef JM, Cressie NAC, Barry RP. Flexible spatial models based on the fast Fourier transform
   (FFT) for cokriging. Journal of Computational and Graphical Statistics. 2004; 13:265–282.

Wackernagel, H. Multivariate Geostatistics: An introduction with Applications. 3rd edition. New
   York: Springer-Verlag; 2006.

Wikle C, Cressie N. A dimension-reduced approach to space-time Kalman filtering. Biometrika. 1999;
   86:815–829.

Zhang H. Inconsistent estimation and asymptotically equal interpolations in model-based geostatistics.
   Journal of the American Statistical Association. 2004; 99:250–261.
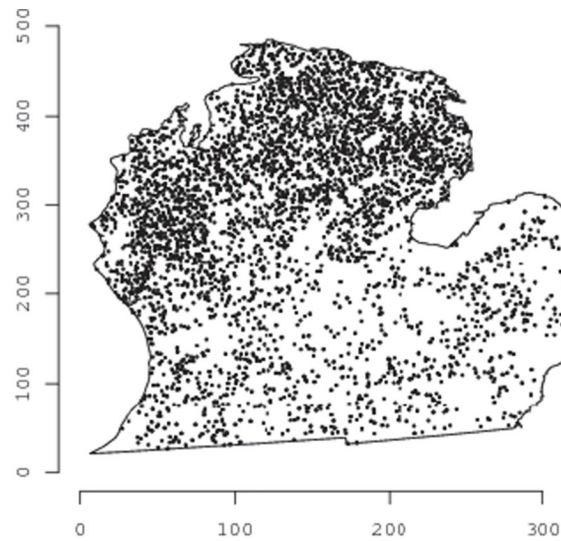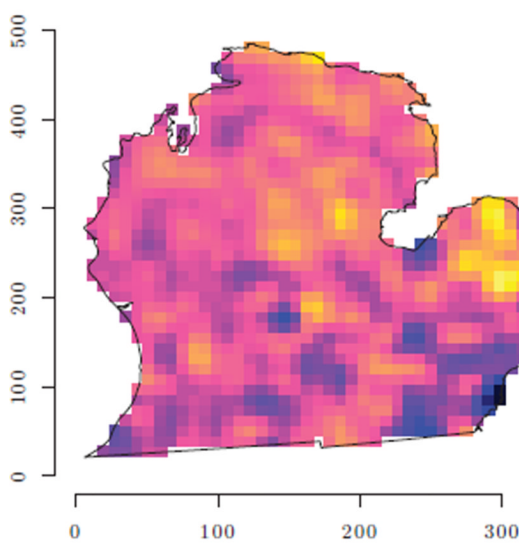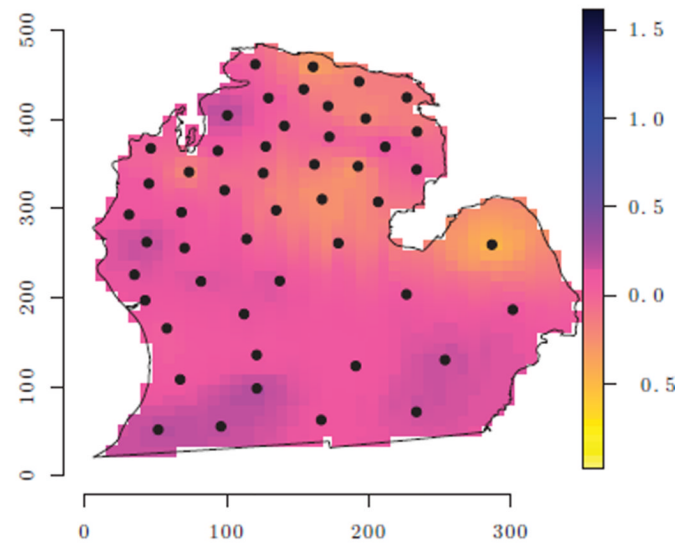
**Figure 1.**
Forest biomass dataset and associated estimates for the 50 knot predictive process models:
(a) location of forest inventory plots; (b) interpolated surface of the non-spatial model
residuals; and (c) predictive process model estimated spatial random effects with knot
locations.

## Table 1

Nonspaial and predictive process candidate models' parameter posterior credible intervals (50 (2.5 97.5)) and model fit criterion for the forest biomass dataset. Run time is in hours for a single chain of 25,000 interations on a single non-hyperthreaded processor.

| | Non-spatial | Predictive process | | |
|---|---|---|---|---|
| | | **50** | **100** | **200** |
| $\beta_0$ | 10.98 (10.95, 11.01) | 10.99 (10.71, 11.26) | 11.00 (10.68, 11.30) | 11.01 (10.97, 11.06) |
| $\beta_{TC1}$ | 0.07 (0.02, 0.12) | 0.03 (−0.02, 0.09) | 0.03 (−0.02, 0.09) | 0.05 (0.00, 0.11) |
| $\beta_{TC2}$ | −0.03 (−0.09, 0.02) | −0.00 (−0.07, 0.06) | −0.01 (−0.07, 0.06) | −0.02 (−0.08, 0.04) |
| $\beta_{TC3}$ | 0.45 (0.41, 0.49) | 0.43 (0.39, 0.48) | 0.43 (0.39, 0.48) | 0.44 (0.39, 0.48) |
| $\sigma^2$ | – | 0.17 (0.09, 0.33) | 0.14 (0.08, 0.24) | 0.28 (0.16, 0.48) |
| $\tau^2$ | 1.01 (0.97, 1.04) | 0.93 (0.84, 0.98) | 0.96 (0.87, 1.00) | 0.74 (0.53, 0.86) |
| $\phi$ | | 0.016 (0.009, 0.051) | 0.011 (0.007, 0.062) | 0.016 (0.011, 0.023) |
| G | 5935 | 5804 | 5759 | 5640 |
| P | 5939 | 5842 | 5830 | 5810 |
| D | 11874 | 11646 | 11590 | 11450 |
| Run time (hours) | – | 9.44 | 19.31 | 37.08 |