



Published in final edited form as:

Spat Stat. 2012 May 1; 1: 30–39. doi:10.1016/j.spasta.2012.02.005.

Hierarchical Modeling for Spatial Data Problems

Alan E. Gelfand¹ [Professor]

¹Department of Statistical Science, Duke University, Durham, North Carolina, 27708-0251, USA

Abstract

This short paper is centered on hierarchical modeling for problems in spatial and spatio-temporal statistics. It draws its motivation from the interdisciplinary research work of the author in terms of applications in the environmental sciences - ecological processes, environmental exposure, and weather modeling. The paper briefly reviews hierarchical modeling specification, adopting a Bayesian perspective with full inference and associated uncertainty within the specification, while achieving exact inference to avoid what may be uncomfortable asymptotics. It focuses on point-referenced (geo-statistical) and point pattern spatial settings. It looks in some detail at problems involving data fusion, species distributions, and large spatial datasets. It also briefly describes four further examples arising from the author's recent research projects.

Keywords

Data fusion; directional data; Dirichlet processes; extreme values; kernel predictors; species distributions

1 Introduction

At the outset, I am delighted to be invited to contribute this short paper profiling my research interests for this promising new journal. The time is right for a journal devoted to spatial statistics. I note that spatial statistics has experienced an unusual evolution as a field within the discipline of Statistics. The stochastic process theory that underlies much of the field was developed within the mathematical sciences by probabilists, whereas, early on, much of the statistical methodology was developed quite independently. In fact, this methodology grew primarily from the different areas of application: mining engineering leading to the development of geostatistics by Matheron and colleagues, agriculture with spatial considerations owing to the thinking of Fisher on randomization and blocking, and forestry which motivated the seminal Ph.D. thesis of Matérn. As a result, for many years, spatial statistics labored on the fringe of mainstream statistics. However, the past twenty years have seen an explosion of interest in space and space-time problems. This has been largely fueled by the increased availability of inexpensive, high speed computing (as has been the case for many other areas). Such availability has enabled the collection of large spatial and spatio-temporal datasets across many fields, has facilitated the widespread usage of sophisticated geographic information systems (GIS) software to create attractive displays, and has enabled inferential investigation of challenging, evermore appropriate and realistic models.

In the process, spatial statistics has been brought into the mainstream of statistical research with a proliferation of books, conferences and workshops, courses and short courses, and an exciting new journal! Moreover, while there has been a body of strong theoretical work

developed since the 1950's, it is safe to say that, broadly, spatial statistics has changed from a somewhat ad hoc field to a more model-driven one. Hence stochastic modeling, in particular, hierarchical modeling is the primary focus of my article, as well as a reflection of my contributions to the field. It is the opportunity to (i) frame flexible stochastic models, (ii) specify models that capture important features of complex processes, and (iii) add a full inference engine to the lovely displays that can be produced with GIS software, that has driven my research in spatial statistics. It is my objective here to give you a bit of the flavor of this work. As requested by the editor, in profiling my research activity, the associated references will primarily supply papers I have been involved with.

1.1 A paradigm shift

As we move into the second decade of the 21st century, we are witnessing a dramatic paradigm shift in the way that statisticians collaborate with researchers from other disciplines. Disappearing are the days when the statistician was called in at the end of a project to provide some routine data analysis and some summary displays. Now the statistician is an integral player in a research team, helping to formulate hypotheses, identify data needs, develop suitable stochastic models, and implement fitting of and inference from the resulting challenging models. Altogether, the statistician becomes sufficiently knowledgeable in the subject matter to “walk the walk” and “talk the talk,” adding another scientific dimension to her/his skill set.

As part of this shift, there is increasing attention paid to bigger picture science, to looking at complex processes with an integrative perspective, to bringing a range of knowledge to this effort. Increasingly, we find researchers working with observational data, less with designed experiments, recognizing that the latter can help inform about the former but the gathering of such experiments provides only one source of data for learning about the complex process. Other information sources, empirical, theoretical, physical, etc., will also be included in the synthesis.

The primary result of all of this is the development of a multi-level stochastic model. Such models are well-suited for incorporating the foregoing knowledge, allowing it to be inserted it at various levels of the modeling, as appropriate. Following the vision of Mark Berliner [3], we imagine a three stage hierarchical specification:

First stage : [*data*|*process*, *parameters*]

Second stage : [*process*|*parameters*]

Third stage : [(*hyper*)*parameters*].

The simple form of this specification belies its breadth. The process component can include multiple levels. For our interests here, it can be spatial and it can be dynamic. The data can be conditioned on whatever aspects of the process are appropriate. The stochastic forms can be multivariate, perhaps infinite dimensional with parametric and/or nonparametric specifications.

In principle, a hierarchical model can be *flattened* by suitable marginalization/integration. However, the advantage of the hierarchical form lies in convenience of specification, ease of interpretation and, often, in facilitation of model fitting. Furthermore, by recognizing the uncertainty in the model unknowns, uncertainty is properly propagated to inference arising from the model.

In view of the above, hierarchical modeling has taken over the landscape in contemporary stochastic modeling. Though analysis of such modeling can be attempted through non-

Bayesian approaches, working within the Bayesian paradigm enables exact inference and proper uncertainty assessment within the given specification.

2 Structured Random Effects and Basic Hierarchical Spatial Modeling

Arguably, the utilization of hierarchical models initially blossomed in the context of handling random effects and missing data, using the E-M algorithm [11] for likelihood analysis and Gibbs Sampling [16] for fully Bayesian analysis. With regard to random effects, both classical and frequentist modeling supply a stochastic specification for these effects, usually assumed to be a normal distribution with an associated variance component. These effects can be introduced at different levels of the modeling but, regardless, in much of the literature, they are assumed to be exchangeable, in fact i.i.d. However, with space and space-time data, random effects are specified with structured dependence.

That is, if scalar ω_i is associated with individual/measurement i , we need not insist that the vector, $\boldsymbol{\omega}$, of ω_i 's, be distributed as $\boldsymbol{\omega} \sim \mathcal{N}(\mathbf{0}, \sigma^2 I)$, as is customary. We can replace $\sigma^2 I$ with $\Sigma(\boldsymbol{\theta})$ where $\Sigma(\boldsymbol{\theta})$ has structured dependence. With a sample of size n , we could not learn about an arbitrary positive definite $n \times n$ matrix Σ but we could learn about Σ defined as a function of only a few parameters.

In the spatial setting, envisioning observations at point-referenced locations, structured dependence is frequently specified through a Gaussian process (GP). With a GP, we need only specify finite dimensional joint distributions, with the joint dependence determined by a valid covariance function. Customarily, the covariance function assigns stronger association to variables that are closer to each other in geographic space.

The standard univariate spatial model, incorporating spatial random effects [1, 9] takes the form

$$Y(s) = \mathbf{x}^T(s)\boldsymbol{\beta} + w(s) + \epsilon(s). \quad (1)$$

Here, the residual is partitioned into two pieces: the $w(s)$ are spatial random effects and the $\epsilon(s)$ are usual i.i.d. errors. $w(s)$ is from a Gaussian process, introducing say the stationary covariance function $\sigma^2(s - s'; \boldsymbol{\phi})$. $\epsilon(s)$ adds pure error (or white noise), i.e., the $\epsilon(s)$ are independent $\mathcal{N}(0, \tau^2)$.

For data $Y(s_i)$, $i = 1, \dots, n$, and $\mathbf{Y} = (Y(s_1), \dots, Y(s_n))^T$, the above model yields a marginal covariance matrix for \mathbf{Y} of the form $\Sigma = \sigma^2 R(\boldsymbol{\phi}) + \tau^2 I$, with $R_{ij} = \rho(s_i - s_j; \boldsymbol{\phi})$. The dependence incorporated into $R(\boldsymbol{\phi})$ enables us to learn about both variance components. Setting $\boldsymbol{\theta} = (\boldsymbol{\beta}, \sigma^2, \tau^2, \boldsymbol{\phi})^T$, we see that this is not a high dimensional problem (perhaps a half dozen components in $\boldsymbol{\beta}$, three or four in Σ). The likelihood is given by $\mathbf{Y}|\boldsymbol{\theta} \sim \mathcal{N}(X\boldsymbol{\beta}, \sigma^2 R(\boldsymbol{\phi}) + \tau^2 I)$. In the Bayesian setting, typically, independent priors are chosen for the parameters so $p(\boldsymbol{\theta}) = p(\boldsymbol{\beta})p(\sigma^2)p(\tau^2)p(\boldsymbol{\phi})$.

Of course, we may ask, "Where is the hierarchical modeling?" In fact, the foregoing is really a hierarchical setup by considering a first stage likelihood conditional on the spatial random effects $\mathbf{w} = (w(s_1), \dots, w(s_n))$. That is, we have:

$$\text{First stage: } Y|\boldsymbol{\theta}, \mathbf{w} \sim \mathcal{N}(X\boldsymbol{\beta} + \mathbf{w}, \tau^2 I)$$

The $Y(s_i)$ are conditionally independent given the $w(s_i)$'s.

Second stage: $\mathbf{w}|\sigma^2, \phi \sim N(0, \sigma^2 R(\phi))$

The $w(s)$ provide the process model.

Third stage: priors on $(\beta, \tau^2, \sigma^2, \phi)$.

With regard to model fitting, we seek the marginal posterior $p(\boldsymbol{\theta}|\mathbf{Y})$, which is the same under the marginal and hierarchical settings. That is, we can fit the model as $f(\mathbf{Y}|\boldsymbol{\theta})p(\boldsymbol{\theta})$ or as $f(\mathbf{Y}|\boldsymbol{\theta}, \mathbf{w})p(\mathbf{w}|\boldsymbol{\theta})p(\boldsymbol{\theta})$. Fitting the marginal model using Markov chain Monte Carlo (MCMC) is usually computationally better behaved. We have a lower dimensional MCMC (no \mathbf{w} 's). Additionally, $\sigma^2 R(\phi) + \tau^2 I$ will be *diagonally dominant*, hence more stable than $\sigma^2 R(\phi)$ in terms of matrix inversion needed for sampling and likelihood evaluation.

Of course, there will be interest in the spatial random effects (and, in fact, in the entire spatial surface of the w 's) in order to see the pattern of spatial adjustment. We have not lost the w 's with the marginalized sampling. They are easily recovered, one-for-one with the posterior samples of $\boldsymbol{\theta}$, via familiar composition sampling: $p(\mathbf{w}|\mathbf{Y}) = \int p(\mathbf{w}|\boldsymbol{\theta}, \mathbf{Y})p(\boldsymbol{\theta}|\mathbf{Y})d\boldsymbol{\theta}$ [1].

In practice, we might have a non-Gaussian first stage. For instance, $Y(s)$ need not be a continuous variable. We can imagine, according to the scale of "points", presence/absence or abundance (counts) observed at a location. To build appropriate models, we replace the Gaussian likelihood with an appropriate exponential family member, resulting in spatial generalized linear models [12]. The hierarchical model above recurs. We only revise the first stage such that the $Y(s_j)$ are conditionally independent given $\boldsymbol{\beta}$ and $w(s_j)$ with $f(y(s_j)|\boldsymbol{\beta}, w(s_j), \gamma)$ an appropriate non-Gaussian likelihood such that

$$g(E(Y(s_i))) = \eta(s_i) = \mathbf{x}^T(s_i)\boldsymbol{\beta} + w(s_i), \quad (2)$$

where η is a canonical link function (such as a logit) and γ is a dispersion parameter.

Importantly, if we introduce the spatial random effects in the transformed mean, then, with continuous covariates, this encourages the means of spatial variables at proximate locations to be close to each other. Despite the conditional independence, marginal spatial dependence is induced between, say, $Y(s)$ and $Y(s')$, but the observed $Y(s)$ and $Y(s')$ need not be close to each other. In fact, the $Y(s)$ surface is everywhere discontinuous. In different terms, our second stage modeling is targeted at spatial explanation of the process, here in terms of the mean. We do not seek to achieve smoothness in the observed surface.

3 Data fusion

Data assimilation has some history in the meteorology community but has only recently received serious attention in the statistics community. We begin with the Bayesian melding model of Fuentes and Raftery [15] which has gained considerable attention. In the spatial setting, typically, we would be fusing a dataset consisting of measurements at monitoring stations, for example, exposure to ozone or particulate matter, with the output of a computer model for such exposure. The former is quite accurate but only sparsely available, often with missingness. The latter is uncalibrated but is available everywhere. The former is associated with point-referenced locations, the latter is supplied for grid cells, for example 12 km squares.

The melding or fusion model envisions a latent true exposure surface which is informed by both the station data and the computer model data. The hierarchical model arises with the two data sources providing the first stage model. The latent true model provides a process specification at the second stage, with hyperparameters at the third stage. More precisely, let the $Y(s_i)$ be the observed station data at s_i ; let $X(B_j)$ be the computer model output for grid cell B_j and let $Z(s)$ be the true exposure surface. The station data is given a measurement error model, i.e.,

$$Y(s_i) = Z(s_i) + \epsilon(s_i) \quad (3)$$

where the ϵ 's are an i.i.d. error specification. The computer output is modeled as a calibration specification, i.e., for grid cell B_j ,

$$X(B_j) = \int_{B_j} (a(s) + b(s)Z(s) + \delta(s)) ds \quad (4)$$

where $a(s)$ and $b(s)$ are Gaussian processes with the $\delta(s)$'s being pure error. Here, the challenge for the melding approach emerges. The integral for $X(B_j)$ is stochastic because, for example, $a(s)$ is not a function but a realization of a stochastic process. So, the integral can not be computed explicitly; at best we can implement a Monte Carlo integration [1]. If we have to do many of these (and, in a practical situation we would have many grid cells), we would have an enormous number of Monte Carlo integrations to do at each iteration of an MCMC fitting algorithm.

Finally, we have the second stage process model, say,

$$Z(s) = \mu(s) + \eta(s). \quad (5)$$

Here, the mean, $\mu(s)$, captures the large scale structure, perhaps through covariates, perhaps through a trend surface, while $\eta(s)$ capture the small scale structure or second order dependence through a GP. Again, model fitting is challenging; Fuentes and Raftery [15] observe that they were only able to successfully fit the model in the case that $b(s) = b$.

So, again, Bayesian melding has two important limitations. First, it is computationally intensive. Since computer model outputs usually cover large spatial domains, thereby introducing a very large number of grid cells, a very large number of stochastic integrals need to be computed. Secondly, as proposed, it does not incorporate a temporal dimension and, computationally, a dynamic extension is, practically, infeasible.

Again, accommodating the spatial misalignment between the two types of data is of fundamental importance for both improved predictions of exposure as well as for evaluation and calibration of the numerical model. Fully model-based alternatives, so-called downscalers, can address these foregoing limitations [4, 5]. As its name implies, the downscaler scales the output from numerical models to point level. The static spatial version, specified within a Bayesian framework, regresses the observed data on the numerical model output using spatially-varying coefficients which are, in turn, specified through a correlated spatial Gaussian process. In particular, the downscaler replaces (3), (4), and (5) with

$$Y(s_i) = \beta_0(s_i) + \beta_1(s_i)X(B_j) + \epsilon(s_i) \quad (6)$$

where $s_i \in B_j$. In (6), $(\beta_0(s), \beta_1(s))$ is modeled as a bivariate GP.

The benefit of (6) is evident. We model the monitoring stations (which are relatively few) rather than the grid cells associated with the computer model (which are usually very many). In different terms, we do not offer a latent process model for the data; we assume the process modeling has been incorporated into the computer model. Rather, the goal of the fusion is to provide a process for local calibration between the two sources.

As an example, we have applied the static downscaler to ozone concentration data for the Eastern US, comparing it to Bayesian melding ([15]) and ordinary kriging ([9]). The downscaler outperforms Bayesian melding in terms of computing speed and is superior to both Bayesian melding and ordinary kriging in terms of predictive performance; predictions obtained are better calibrated and predictive intervals have empirical coverage closer to the nominal values.

The space-time downscaler offers a natural extension of the form

$$Y_t(s_i) = \beta_{0t}(s_i) + \beta_{1t}(s_i) X_t(B_j) + \epsilon_t(s_i). \quad (7)$$

The space-time downscaler is immediately recognized as a dynamic space-time model [17]. Bayesian model fitting using the forward filter, backward sample (ffbs) algorithm [6, 14] becomes the customary approach.

A more sophisticated example of a space-time data fusion appears in [28], motivated by a problem of estimation of annual wet chemical deposition in the eastern United States. We note that precipitation is required in order to have wet deposition. Hence, our modeling for monitoring stations must allow for point masses at 0 for both precipitation and wet deposition. So, we model precipitation and then, deposition given precipitation, with spatial misalignment. The computer model output also supplies values of 0 for wet deposition in some grid boxes at some time points so, here as well, we need to allow point masses at 0. For the latter, we introduce a latent environmental process at the grid scale to enable these point masses. It is modeled through a conditionally autoregressive (CAR) specification. Similarly, for the former, we condition both precipitation and deposition on a point-referenced latent environmental process. The downscaling connects this latter process to the grid-scale process using a measurement error model. We fitted this model to weekly wet chemical deposition data both for the sulfate and nitrate compounds covering the eastern United States. The model was validated with set-aside data from a number of monitoring sites. Predictive Bayesian methods enable inference on aggregated summaries such as quarterly and annual deposition maps.

4 Species distributions

Understanding spatial patterns of species diversity and the distributions of individual species is a consuming problem in biogeography and conservation. That is, ecologists try to explain where species are and why. In a series of papers [20, 21, 25], data from the Cape Floristic Region (CFR) in South Africa, one of six floral kingdoms in the world and a global hotspot of diversity and endemism, was used to study species distributions. This work considered data collected at a fixed set of locations, assessing presence or absence of a species at the location. Then, the question of presence/absence becomes a matter of building an explanatory model for Bernoulli trials. In fact, in the context of hierarchical spatial models, we developed “probability-of-presence” surfaces in response to environmental features and species characteristics. Furthermore, we introduced notions such as *suitability* and *availability* to develop *potential* and *adjusted* presence/absence surfaces in response to transformation of the land.

However, often, available data is collected only as a set of presence locations (e.g., museum collections and other non-systematic sampling settings), thus precluding the possibility of a presence-absence analysis. In [8], we propose that it is natural to view presence-only data as a point pattern over a region. We use a hierarchical model to treat the presence data as a realization of a spatial point process, whose intensity is driven by local environmental features. Spatial dependence in the intensity levels is modeled with random effects involving a zero mean Gaussian process. We augment the model to capture highly variable and typically sparse sampling effort as well as land transformation, both of which degrade the point pattern. Again, the Cape Floristic Region (CFR) in South Africa provides an extensive body of such species data. The potential (i.e., nondegraded) intensity surfaces over the entire area are of interest from a conservation and policy perspective. The region is divided into ~ 37, 000 minute x minute grid cells. (In this region a minute by minute grid cell is roughly 1.5 by 1.8 kilometers.) To work with a Gaussian process over a very large number of cells we use predictive spatial process approximation (see Section 5). Bias correction was implemented by adding a heteroscedastic error component. We illustrated with six different species, some prevalent, some sparse. Importantly, we made comparison with the now popular Maxent approach [26], though the latter is limited with regard to inference and can only provide intensities normalized to densities. Using simulation examples with a known intensity and different loss functions, we showed an order of magnitude improvement with our point pattern analysis. An additional feature of our modeling is the opportunity to infer about species richness. Species richness is the number of distinct species in an areal unit and thus is a concept that depends upon spatial scale. It is the primary determinant of biodiversity at a given scale.

Continuing work with the CFR data turned to the question of modeling abundance, i.e., quantifying presence. Again, the CFR provides a rich class of species abundance data for such modeling. We developed a multi-stage Bayesian hierarchical model for explaining species abundance over the foregoing 37, 000 grid cells [7]. Species abundance is observed at some locations within some cells. The abundance values are ordinally categorized. Environmental and soil-type factors, likely to influence the abundance pattern, were included in the model. Similar to the presence/absence and presence-only problems, we formulated the empirical abundance pattern as a degraded version of the potential pattern, with the degradation effect accomplished in two stages. First, we adjusted for land use transformation and then we adjusted for misclassification error in yielding the observed abundance classifications. Notably, only 28% of the grid cells have been sampled and, for sampled grid cells, the number of sampled locations ranges from one to more than one hundred. Still, we are able to develop potential and transformed abundance surfaces over the entire region.

In the hierarchical framework, categorical abundance classifications are induced by continuous latent surfaces. The degradation model above is built on the latent scale. On this scale, an areal level spatial regression model was used for modeling the dependence of species abundance on the environmental factors, incorporating spatial random effects to capture anticipated similarity in abundance pattern among neighboring regions. The model was fitted for several different species. With categorical data, display of the resultant abundance patterns is a challenge; we offered several different views. Again, the patterns are of importance with implications for species competition and, more generally, for planning and conservation.

5 Larger spatial datasets

With scientific data available at geocoded locations, investigators increasingly employ spatial process models for carrying out statistical inference. Within the Bayesian framework,

hierarchical models are routinely fitted using Markov chain Monte Carlo. However, fitting such spatial models often involves expensive matrix decompositions whose computational complexity increases in cubic order with the number of spatial locations, rendering such models infeasible for large spatial data sets. This computational burden is exacerbated in multivariate settings with several spatially dependent response variables. It is also aggravated when data is collected at numerous time points and spatiotemporal process models are used.

Approaches to tackle this problem primarily adopt one of three paths. The first seeks approximations for the spatial process using kernel convolutions, moving averages, or basis functions. Essentially, the spatial process $w(s)$ is replaced by an approximation $\tilde{w}(s)$ that represents the realizations in a lower-dimensional subspace. The second approach works in the spectral domain of the spatial process and approximates the likelihood in terms of spectral densities, avoiding the matrix computations. The third approach either replaces the process (random field) model by a Markov random field model or else approximates it by a Markov random field. See [2] for a full literature review.

In recent work [2, 13] we offered and refined a dimension reduction approach that we call a predictive process model for spatial and spatiotemporal data. Every spatial (or spatiotemporal) process induces a predictive process model (in fact, arbitrarily many of them). The latter models project process realizations of the former to a lower-dimensional subspace thereby reducing the computational burden. Hence, we can accommodate nonstationary, non-Gaussian, possibly multivariate, possibly spatiotemporal processes in the context of larger datasets.

We consider a set of ‘‘knots’’ $\mathcal{S}^* = \{s_1^*, \dots, s_m^*\}$ which may or may not form a subset of the entire collection of locations. A Gaussian process would yield

$\mathbf{w}^* = [w(s_i^*)]_{i=1}^m \sim MVN(0, C^*(\theta))$ as its realizations over \mathcal{S}^* , where $C^*(\theta) = [C(s_i^*, s_j^*; \theta)]_{i,j=1}^m$ is the corresponding $m \times m$ covariance matrix. The predictive process approximation at site s_0 becomes

$$\tilde{w}(s_0) = E[w(s_0) | \mathbf{w}^*] = \mathbf{c}^T(s_0; \theta) C^{*-1}(\theta) \mathbf{w}^*, \quad (8)$$

where $\mathbf{c}(s_0; \theta) = [C(s_0, s_j^*; \theta)]_{j=1}^m$. This single site interpolator, in fact, defines a spatial process $\tilde{w}(s) \sim GP(0, \tilde{C}(\cdot))$ with nonstationary covariance function $\tilde{C}(s, s'; \theta) = \mathbf{c}^*(s; \theta) C^{*-1}(\theta) \mathbf{c}^*(s', \theta)$ where $\mathbf{c}^*(s; \theta) = [C(s, s_j^*; \theta)]_{j=1}^m$. We refer to $\tilde{w}(s)$ as the *predictive process* derived from $w(s)$ (which we call the *parent process*). The process is completely specified given the covariance function of the parent process and \mathcal{S}^* . The joint distribution associated with the realizations at any set of locations is singular if the set has at most m locations. Further properties of realizations of this process are discussed in [2]. Replacing $w(s)$ in (1) with $\tilde{w}(s)$, we modify the spatial regression model in (1) to:

$$Y(s) = \mathbf{x}^T(s) \beta + \tilde{w}(s) + \epsilon(s). \quad (9)$$

The predictive process offers the advantage of *off-the-shelf* approximation since it is induced by the process we seek to work with. As such, it is readily extended to spatiotemporal models. Like other dimension reduction techniques, it requires knot selection, raising potentially interesting design problems [13]. Lastly, it is not a panacea for very large space-time problems. The amount of required matrix multiplication and associated storage limit its employment to order 10^4 locations.

Nowadays, really large space and space-time datasets would involve order 10^6 locations or more. These settings will exceed the capability of fitting hierarchical models within a fully Bayesian framework using Markov chain Monte Carlo. Alternatives that may be considered include fixed rank kriging ([10]) and integrated nested Laplace approximation (INLA) ([27]).

6 Further examples

I conclude with a few paragraphs on problems which I have explored recently but do not have space to describe in detail.

Example 1

Customary modeling for continuous point-referenced data assumes a Gaussian process which is often taken to be stationary. When such models are fitted within a Bayesian framework, the unknown parameters of the process are assumed to be random so a random Gaussian process results, i.e., the stochastic process of random variables is random itself. A richer way to model random stochastic processes is to place a non-parametric distribution on a class of stochastic processes. We have proposed a novel spatial Dirichlet process mixture model to produce a random spatial process which is neither Gaussian nor stationary. (See [19] with further extension in [18]). We first specified a spatial Dirichlet process (SDP) model for spatial data. This process plays the role of $w(s)$ in (1). Due to the almost sure discreteness associated with direct use of Dirichlet process models, we introduce mixing of this process against a pure error process, $\epsilon(s)$, to create residuals of the form in (1). Again, these residuals are non-Gaussian and can have skewed and multimodal distributions. The SDP and the mixed SDP have interesting properties. In the Bayesian framework, posterior model fitting is implemented using Gibbs sampling and spatial prediction can be handled. In practice, the SDP requires replicates of the process.

Example 2

In most spatio-temporal applications, covariates are measured in addition to a spatio-temporal response. In such cases, the purpose of the statistical model is often to quantify the change in expectation of the response due to a change in the predictors, while accounting for the spatio-temporal structure of the response, the predictors, or both. The most common approach for building such a model is to express the response-predictors relationship at a single spatio-temporal coordinate. For spatio-temporal problems, however, the relationship between the response and predictors may be more complex. In other words, the response at a single spatio-temporal coordinate may be affected by levels of the predictors in spatio-temporal proximity to the response location. We have proposed [22, 23] a flexible modeling framework to capture spatial and temporal lagged effects between a predictor and a response. Specifically, kernel functions are used to weight a spatio-temporal covariate surface in a regression model for the response. The kernels are assumed to be parametric and nonstationary with the data informing the parameter values of the kernel. We have illustrated the methodology on simulated data as well as a physical data set of ozone concentrations to be explained by temperature. We find substantial improvement in the regression model using these kernel-based predictors. They also help to illuminate the nature of ozone response to temperature.

Example 3

Directional data arise in oceanography (wave directions) and meteorology (wind directions), and, more generally, with periodic measurements recorded in degrees or angles on a circle. We have introduced a fully model-based approach to handle angular data in the case of measurements taken at spatial locations, anticipating structured dependence between these

measurements [24]. Starting with a wrapped model for directional data, we formulate a wrapped Gaussian spatial process model for this setting, induced from a customary *inline* Gaussian process. We have examined the properties of this process, including the induced correlation structure. We build a hierarchical model to handle this situation and show how to fit this model straightforwardly using Markov chain Monte Carlo methods. Our approach enables spatial interpolation and can accommodate measurement error. We provide illustration with a set of angular wave direction data from the Adriatic coast of Italy, generated through a complex computer model. An alternative approach starts with the projected normal model for directional data and enables the projected normal spatial process, built from a bivariate Gaussian process model. Such models are more flexible than usual wrapped or von Mises models and easily handle regression. However, they can be more challenging to fit.

Example 4

We propose a hierarchical modeling approach for explaining a collection of point-referenced extreme values. A typical setting would be a set of annual maximum daily temperatures at a collection of monitoring stations. We again build a hierarchical model with a first stage specification for the annual maxima that are Generalized Extreme Value (GEV) distributions. At the second stage we have the GEV parameters μ (location), σ (scale), and ξ (shape). The first two can be specified to reflect underlying spatio-temporal structure. In [29], the first stage assumed conditional independence of the maxima. As at the end of Section 2, this assumption can provide process explanation but implies that realizations of the surface of spatial maxima will be everywhere discontinuous.

For phenomena such as temperature, such behavior is unrealistic. In [30], we relaxed this assumption, offering a spatial process model for extreme values which provides mean square continuous realizations, using a Gaussian copula specification. The behavior of the surface is driven by the spatial dependence which is unexplained under the second stage spatio-temporal specification for the GEV parameters. In this sense, the smoothing is viewed as fine scale or short range smoothing while the larger scale smoothing will be captured in the second stage of the modeling. We are able to implement spatial interpolation for extreme values based on this model. A simulation study and a study on actual annual maximum rainfalls for a region in South Africa was used to illustrate the performance of the model.

7 Summary

I foresee a promising future for the field of spatial statistics. Increasingly, data layers with spatial referencing will become available. Increasingly, researchers will appreciate the benefit of incorporating spatial information in helping to explain the complex processes they study. As in other areas of Statistics, the result will be methodology driven by application, creating theoretical, modeling, and computational challenges. Though spatial work will continue to find a home in many publication forums, the amount of spatial work currently being conducted and anticipated into the future argues for the need for a targeted forum in the field. This bodes a bright future for this new journal, *Spatial Statistics*.

Acknowledgments

The work of the author was supported in part by NSF DMS 0914906 and NSF CDI 0940671.

References

1. Banerjee, S.; Carlin, BP.; Gelfand, AE. Hierarchical modeling and analysis for spatial data. Chapman & Hall/CRC; Boca Raton: 2004.

2. 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; 70:825–48.
3. Berliner, LM. Hierarchical Bayesian time series models.. In: Hanson, K.; Silver, R., editors. *Maximum entropy and Bayesian methods.* Kluwer Academic Publishers; 1996. p. 15-22.
4. Berrocal VJ, Gelfand AE, Holland DM. A spatio-temporal downscaler for output from numerical models. *Journal of Agricultural, Biological and Environmental Statistics.* 2010; 15:176–97.
5. Berrocal VJ, Gelfand AE, Holland DM. A bivariate space-time downscaler under space and time misalignment. *Annals of Applied Statistics.* 2010; 4:1942–75. [PubMed: 21853015]
6. Carter CK, Kohn R. On Gibbs sampling for state space models. *Biometrika.* 1994; 81:541–53.
7. Chakraborty A, Gelfand AE, Silander JA Jr. Latimer AM, Wilson A. Modeling large scale species abundance through latent spatial processes. *Annals of Applied Statistics.* 2010; 4:1403–29.
8. Chakraborty A, Gelfand AE, Silander JA Jr. Wilson A, Latimer A. Point pattern modeling for degraded presence-only data over large regions. *Journal of the Royal Statistical Society, Series C.* 2011; 60:757–76.
9. Cressie, NAC. *Statistics for spatial data.* Wiley; New York: 1993.
10. Cressie NAC, Johannesson G. Spatial prediction for massive datasets. *Journal of the Royal Statistical Society Series B.* 2008; 70:209–226.
11. Dempster AP, Laird NM, Rubin DB. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B.* 1977; 39:1–38.
12. Diggle PJ, Moyeed RA, Tawn JA. Model-based geostatistics (with discussion). *Applied Statistics.* 1998; 47:299–350.
13. Finley AO, Sang H, Banerjee S, Gelfand AE. Improving the performance of predictive process modeling for large datasets. *Computational Statistics and Data Analysis.* 2009; 53:2873–84. [PubMed: 20016667]
14. Frühwirth-Schnatter S. Data augmentation and dynamic linear models. *Journal of Time Series Analysis.* 1994; 15:183–202.
15. Fuentes M, Raftery AE. Model evaluation and spatial interpolation by Bayesian combination of observations with outputs from numerical models. *Biometrics.* 2005; 61:36–45. [PubMed: 15737076]
16. Gelfand AE, Smith AFM. Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association.* 1990; 85:398–409.
17. Gelfand AE, Banerjee S, Gamerman D. Spatial process modelling for univariate and multivariate dynamic spatial data. *Environmetrics.* 2005; 16:465–79.
18. Gelfand, AE.; Guindani, M.; Petrone, S. Bayesian nonparametric modeling for spatial data analysis using Dirichlet processes (with discussion).. In: Bernardo, JM.; Bayarri, MJ.; Berger, JO.; Heckerman, D.; Smith, AFM.; West, M., editors. *Bayesian Statistics 8.* Oxford University Press; London: 2007. p. 175-200.
19. Gelfand AE, Kottas A, MacEachern SM. Bayesian nonparametric spatial modeling with Dirichlet process mixing. *Journal of the American Statistical Association.* 2005; 100:1021–35.
20. Gelfand AE, Silander JA Jr. Wu S, Latimer A, Lewis PO, Rebelo AG, Holder M. Explaining species distribution patterns through hierarchical modeling. *Bayesian Analysis.* 2005; 1:41–92.
21. Gelfand AE, Schmidt AM, Wu S, Silander JA Jr. Latimer A, Rebelo AG. Explaining species diversity through species level hierarchical modeling. *Journal of Royal Statistical Society, Series C.* 2005; 54:1–20.
22. Heaton MJ, Gelfand AE. Spatial prediction using kernel averaged predictors. *Journal of Agricultural, Biological, and Environmental Statistics.* 2011; 10:233–52.
23. Heaton, MJ.; Gelfand, AE. Kernel averaged predictors for spatio-temporal regression models.. 2011. In review
24. Jona Lasinio, G.; Gelfand, AE.; Jona Lasinio, M. Analyzing spatial directional data using wrapped Gaussian processes.. 2011. In revision
25. Latimer A, Wu S, Gelfand AE, Silander JA Jr. Building statistical models to analyze species distributions. *Ecological Applications.* 2006; 16:33–50. [PubMed: 16705959]

26. Phillips SJ, Dudyk M. Modeling of species distributions with Maxent: New extensions and a comprehensive evaluation. *Ecography*. 2008; 31:161–75.
27. Rue H, Martino S, Chopin N. Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximation. *Journal of the Royal Statistical Society, Series B*. 2009; 71:1–35.
28. Sahu SK, Gelfand AE, Holland DM. Fusing point and areal level space-time data with application to wet deposition. *Journal of the Royal Statistical Society - Series C*. 2010; 59:77–103.
29. Sang H, Gelfand AE. Hierarchical modeling for extreme values observed over space and time. *Environmental and Ecological Statistics*. 2009; 16:407–26.
30. Sang H, Gelfand AE. Continuous spatial process models for spatial extreme values. *Journal of Agricultural, Biological, and Environmental Statistics*. 2010; 15:49–65.