

Flexible Gaussian Processes via Convolution

Herbert K. H. Lee, Christopher H. Holloman, Catherine A. Calder, and Dave M. Higdon
Duke University (and LANL)

June 21, 2002

Abstract

Spatial and spatio-temporal processes are often described with a Gaussian process model. This model can be represented as a convolution of a white noise process and a smoothing kernel. We expand upon this model by considering convolutions of non-*iid* background processes. We highlight two particular models based on convolutions of Markov random fields and of time-varying processes. These models are illustrated using examples from hydrology and atmospheric science.

Key Words: Bayesian Dynamic Spatial Model, Forward Filtering Backward Sampling, Permeability, Ozone

1 Introduction

A common model for spatial processes is the Gaussian Process (GP) model, also referred to as kriging (see for example, Matheron 1971, Journel and Huijbregts 1978, Cressie 1991). Such models have many desirable properties, both in practice and in theory. Theoretical considerations include simplicity in the statement of the model and finding a best linear unbiased estimator, as well as asymptotic behavior (e.g., Stein 1999). Some practical reasons for their popularity include the ease of prediction under the model and the ability of the models to represent a range of different types of spatial correlation through a simple parameterization, such as with the Matern class (Matérn, 1986).

In this paper, we seek to generalize the Gaussian Process family, developing new models that expand upon the standard classes of spatial correlation that can be captured by models in the family. In particular, standard GP models are assumed to be stationary (see Section 2). In contrast, the proposed class readily incorporates models with various types of non-stationarity.

The basis for our family of models is a convolution representation of a standard Gaussian Process, which is discussed in Section 2. Flexibility is gained by considering an expanded class of latent processes. Examples in the literature of models resulting from innovative latent processes include Barry and Ver Hoef (1996), Higdon et al. (1999), Fuentes and Smith (2001), Stroud et al. (2001), and Ickstadt and Wolpert (1999). We seek to provide a general framework for such models.

2 Standard Gaussian Processes via Convolution

In a standard Gaussian process model, the process z varies over a domain S according to the distribution

$$\pi(z|\mu, \theta) \propto |\Sigma_\theta|^{-\frac{m}{2}} \exp \left\{ -\frac{1}{2}(z - \mu)^T \Sigma_\theta^{-1} (z - \mu) \right\}$$

where μ is an overall mean level and Σ is a covariance matrix of a particular form. Typical assumptions include stationarity and isotropy. Stationarity has several forms, with standard implications being a finite constant mean and variance, and a covariance structure that does not vary with location. Isotropy implies that the correlation specified by Σ for the process at two locations $z(s_i)$ and $z(s_j)$ is a function only of the distance d_{ij} between s_i and s_j . As Σ must be positive definite, one of several simple parametric forms is typically used for the correlation function, $\rho(d)$, such as the exponential correlogram $\rho(d) = e^{-d}$, the Gaussian correlogram $\rho(d) = e^{-d^2}$, or the Matérn class $\rho(d) = [(d/2)^\nu 2K_\nu(d)] / \Gamma(\nu)$. Additional parameters specify the marginal variance and the range of spatial dependence so that $\Sigma_{\theta_{ij}} = \theta_1 \rho(d_{ij} / \theta_2)$.

In most standard spatial problems, the covariance parameters can be estimated from the data. However, in some problems where no direct measurements of the spatial field are available, such as the inverse problem example in Section 4, these parameters will need to be chosen apriori, as the data do not usually contain sufficient information for the estimation of these parameters (see for example, Oliver, Cunha, and Reynolds 1997).

A convenient representation of a GP model uses process convolutions (Higdon, 2001). One may construct a Gaussian process $z(s)$ over a general spatial (and temporal) region S by convolving a continuous white noise process $x(s)$, $s \in S$ with a smoothing kernel $k(s)$ so that

$$z(s) = \int_S k(u - s)x(u)du, \text{ for } s \in S.$$

The resulting covariance function for $z(s)$ depends only on the displacement vector $d = s - s'$ and is given by

$$c(d) = \text{Var}(z(s), z(s')) = \int_S k(u - s)k(u - s')du = \int_S k(u - d)k(u)du. \quad (1)$$

As a special case, if S is R^p and $k(s)$ is isotropic, then $z(s)$ is also isotropic, with covariance function $c(d)$ that depends only on the magnitude of d . In this case there is a one to one relationship between the smoothing kernel $k(d)$ and the covariogram $c(d)$, provided either $\int_{R^p} k(s)ds < \infty$ and $\int_{R^p} k^2(s)ds < \infty$ or $c(s)$ is integrable and positive definite. The relationship is based on the convolution theorem for Fourier transforms and is discussed in more detail in Thiébaux and Pedder (1987) and Barry and Ver Hoef (1996).

One can approximate the Gaussian process resulting from a convolution of white noise and a kernel by using a discrete set of points from the underlying white noise process x (Higdon, 2001). To be specific, let s_1, \dots, s_m be a specific set of locations in S . Then the GP can be approximated by

$$z(s) = \sum_{i=1}^m k(s_i - s)x(s_i) \quad (2)$$

for a given choice of smoothing kernel k , i.e., the process is computed using a finite set of basis points. Note that this still produces a continuous process, even though the underlying process is discrete. This

finite representation has many computational advantages, as only a relatively small number of sites s_i are necessary for a close approximation, saving much computing time. This reduction in dimension is also useful for a variety of theoretical reasons; for example, many inverse problems are drastically ill-conditioned, and the dimension reduction of the parameter space greatly aids in the ability to conduct inference.

The class of spatial processes can be made more flexible by either allowing the kernel to vary spatially (Barry and Ver Hoef, 1996; Higdon et al., 1999) or by using a process other than white noise for the background process. Fuentes and Smith (2001) investigates convolutions of Gaussian processes. Stroud et al. (2001) uses convolutions of mixtures of linear regressions with time-evolving coefficients to model spatio-temporal data. Alternatively, non-Gaussian latent processes can be used (e.g., Ickstadt and Wolpert 1999). Our approach is to consider other such background processes. In particular, we have found that processes which are not strictly stationary in the traditional spatial sense have been useful in several real-world applications. Also, changing the background process instead of changing the kernel tends to have fewer additional computational requirements. The next section considers some theoretical aspects of convolving alternative background processes. The sections following provide some examples of new models.

3 Convolution of non-independent processes

Rather than specifying an underlying independence process for x , one could use a more general model for x . For example, $x(s)$ could be an intrinsically stationary process with semi-variance $\gamma_x(d)$. In this case the semi-variance of the process $z(s)$ induced by equation (2) is given by $\gamma_z(d) = \gamma_z^*(d) - \gamma_z^*(0)$ where

$$\gamma_z^*(d) = \int \int k(v-d)k(u-v)\gamma_x(u) du dv.$$

Of course one could restrict $x(s)$ to a lattice to facilitate computation as well as specification of the dependence structure. By allowing more general models for $x(s)$, large scale structure can be accounted for in $x(s)$ while small scale behavior can be accounted for in the kernel specification. Examples in this paper model $x(s)$ using a Markov random field (MRF) as well as a temporally evolving specification.

As an example we compare two simple 1-d formulations with $n = 12$ observations y at locations s_1^y, \dots, s_n^y equally spaced between 0 and 10. Both formulations fix $k(s)$ to be a $N(0, .6^2)$ density. The first specifies $x(s)$ to be a fine lattice of $m = 40$ i.i.d. $N(0, 1/\beta)$ random variates. The second specifies $x(s)$ to be Brownian motion restricted to the same $m = 40$ lattice points, with increments scaled by the precision parameter β . The two formulations are summarized below:

	i.i.d. x		Brownian motion x
$L(y z, \lambda_y)$	$\propto \lambda_y^{\frac{n}{2}} \exp\{-\frac{1}{2}\lambda_y(y-z)^T(y-z)\}$	$L(y z, \lambda_y)$	$\propto \lambda_y^{\frac{n}{2}} \exp\{-\frac{1}{2}\lambda_y(y-z)^T(y-z)\}$
$\pi(x \beta)$	$\propto \beta^{\frac{m}{2}} \exp\{-\frac{1}{2}\beta \sum_{i=1}^m x_i^2\}$	$\pi(x \beta)$	$\propto \beta^{\frac{m}{2}} \exp\{-\frac{1}{2}\beta \sum_{i=1}^{m-1} (x_i - x_{i+1})^2\}$
$\pi(\beta)$	$\propto \beta^{a_x-1} e^{-b_x\beta}$	$\pi(\beta)$	$\propto \beta^{a_x-1} e^{-b_x\beta}$
$\pi(\lambda_y)$	$\propto \beta^{a_y-1} e^{-b_y\lambda_y}$	$\pi(\lambda_y)$	$\propto \beta^{a_y-1} e^{-b_y\lambda_y}$

Figure 1 shows the resulting posterior summaries for $z(s)$ under the two formulations. Because the precision parameter β in the Brownian motion formulation can modify the amount of spatial dependence in $z(s)$, its

reconstruction better matches the true process from which the data were generated, while the white noise specification overfits the data.

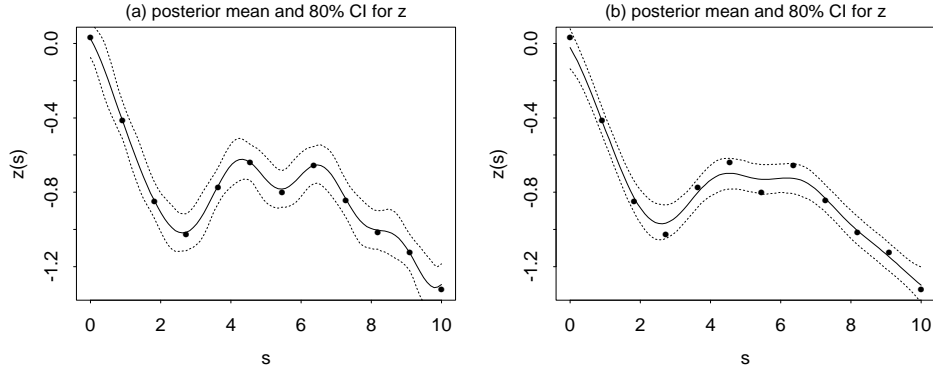


Figure 1: Comparison of posterior means and pointwise 80% credible intervals for $z(s)$ under two formulations: (a) white noise specification for $x(s)$; and (b) Brownian motion specification for $x(s)$.

4 Convolution of Markov Random Fields

As an alternative to using an underlying white noise process, if the set of basis points, s_1, \dots, s_m , is a regular grid, a Markov Random Field (MRF) can be used for the distribution of the underlying process on this grid. A Gaussian MRF for the process x is of the form

$$\pi(x) \propto \beta^{\frac{m}{2}} |W|^{\frac{1}{2}} \exp\{-\frac{1}{2}\beta x^T W x\}$$

where β is a precision parameter controlling the scale and the sparse matrix $W = \{w_{ij}\}$ controls the spatial dependence structure of x . A simple example of an MRF is a first-order model where each site only depends on its immediate neighbors, i.e.,

$$w_{ij} = \begin{cases} -1 & \text{if } i \text{ and } j \text{ are adjacent} \\ n_i & \text{if } i = j \\ 0 & \text{otherwise} \end{cases}$$

where n_i is the number of neighbors adjacent to site i . This model has the advantage of being able to capture a range of spatial dependence through a single parameter, β . This model is further discussed in Besag and Kooperberg (1995), for example. Note that an MRF is not stationary in the sense that the overall mean level is not defined. Another common representation of an MRF takes advantage of its conditional independence, as the distribution for a particular pixel depends only on the values of its neighbors:

$$x_i | x_{-i} \sim N\left(\frac{1}{w_{ii}} \sum_{j \in \partial i} -w_{ij} x_j, \frac{1}{\beta w_{ii}}\right)$$

where x_{-i} denotes all sites except x_i , and the sum runs over the neighbors of x_i .

With an MRF for the underlying process x , Equation (2) then leads to a continuous Gaussian process with a more complex correlation structure. This MRF convolution model can adapt to model different amounts of spatial dependence. For a standard GP resulting from a convolution of a white noise process, the choice of the smoothing kernel determines the spatial structure. When convolving an MRF, the choice of kernel does not fully specify the spatial structure because the structure is also affected by the smoothness of the MRF (parameterized by β above). Thus a kernel could be chosen to induce less spatial dependence than is expected, and the underlying MRF can easily increase the required smoothness and thus dependence. This MRF convolution model still retains the general characteristics of a GP model, including the ability to be well-approximated by a lower-dimensional basis representation.

In contrast to the many problems in which the data do provide information for estimating the covariance structure (e.g., for a standard GP, the parameters in a covariogram, or the choice of smoothing kernel used for the convolution), this is not generally true in inverse problems, a class of problems where there may be no direct observations of the process of interest, and information about this process must be gained via observations of a related variable. The next section provides an example from hydrology where soil structure is inferred from water flow data. In that case, the flow data do not provide enough information to estimate the correlation parameters for a standard GP. However, the data do provide enough information to draw some inferences on the precision parameter of an MRF (Lee et al., 2002). For such inverse problems, an MRF convolution model provides more spatial flexibility than a standard GP, since the precision parameter can be fit from the data, whereas none of the GP correlation parameters can be. Another common concern in inverse problems is underspecification, in that the data may be of a lower effective dimension than the parameter space. Models built using a convolution representation effectively reduce the parameter space and help with this inestimability problem. Thus the MRF convolution model brings the joint benefits of reduced dimensionality and increased flexibility.

4.1 Flow Example

The study of the flow of liquids, particularly groundwater, through porous media, such as soil, is an important problem in engineering. Reliable solutions exist for the forward problem of determining how water flows when the physical characteristics (e.g., permeability) of the aquifer are known. A topic of interest to both statisticians and engineers is the inverse problem of using flow data to infer the permeability structure of the aquifer. An overview of this inverse problem can be found in Yeh (1986). Important applications include contaminant cleanup (James et al., 1997; Jin et al., 1995) and oil production (Xue and Datta-Gupta, 1996).

The estimation of the permeabilities turns out to be a difficult problem. Core samples analyzed in a lab can provide direct estimates at points, but there is significant measurement error involved, and such samples are expensive, so at best only a few samples will be available. Instead, collecting indirect data, such as flow data, is often more cost-effective. In a flow experiment, one or more injection wells force water underground into the aquifer while multiple producer wells extract this water. Water is pumped through this system until equilibrium is reached, and then a tracer (such as a fluorescent or radioactive dye) is injected at the injection well(s) and the concentration of the tracer is measured over time at each of the producer wells. The

resulting concentration curves, and especially the breakthrough times (the times of first arrival of the tracer at a production well), provide information about the underlying permeability field, as it takes longer for the tracer to move through areas of lower permeability. In many cases, the breakthrough time is essentially a sufficient statistic for the whole concentration curve, with the times providing almost as much information as using the entire curves (Vasco et al., 1998; Yoon et al., 1999). Thus we focus on only the breakthrough times here.

The resulting data lead to the inverse problem of inferring the spatial permeability field (treated here as having a scalar value at each point in space) from the flow data, possibly without any direct permeability measurements. For the forward problem of predicting flow from a given configuration of permeabilities, various versions of computer code exist which solve differential equations given by physical laws (i.e., conservation of mass, Darcy’s Law, and Fick’s Law) to determine the flows. Here we use the S3D streamtube code of Datta-Gupta (King and Datta-Gupta, 1998). The likelihood for the permeabilities, $\boldsymbol{\psi}$, compares the true breakthrough times, b , to the fitted times, $\hat{\mathbf{b}}$, computed by the flow simulator for a given value of the permeabilities (i.e., the fitted breakthrough times are a complex, non-analytical but deterministic transformation of the permeabilities). Conditional on the data, we use an *iid* Gaussian error structure:

$$L(\boldsymbol{\psi} \mid \mathbf{b}) \propto \exp \left\{ -\frac{1}{2\sigma^2} \sum_h (b_h - \hat{b}_h(\boldsymbol{\psi}))^2 \right\}.$$

This likelihood is usually highly ill-conditioned, in that different configurations of the permeability field can lead to similar concentration curves. Thus some regularity conditions must be imposed on the problem, and this is typically done by putting restrictions on the permeability field, specifying it to be of some parametric form. Gaussian processes are the standard approach in the literature. A natural approach is to induce regularity by using a spatial model as a prior for the permeabilities. We compare three different spatial prior models here, a standard Markov Random Field (MRF), a standard Gaussian process, and the new MRF convolution model of this section.

The plots in Figure 2 are from a simulated two-dimensional example. The setup is a typical inverted nine-spot configuration, with a central injection well and eight production wells in the corners and at the midpoints of the edges. We use a grid size of 32 by 32. The true field is a realization from a Gaussian process with a specified covariogram, and this is run through the simulator to generate breakthrough time data at the eight production wells. Next the three different models are fit to these data. In the first column of the figure is the model using a first-order symmetric Markov Random Field spatial prior, which estimates its spatial dependence parameter β from the data (this model and the regular GP were fit using the methods of Lee et al. (2002)). In the second column is a standard Gaussian process model, using the correct values for its spatial dependence parameters (since these cannot be effectively estimated from flow data). The last full column is for the new MRF convolution model described in this section, where β is also fit from the data. The top three rows show realizations from the posterior, and the last full row shows posterior means. The true field is shown in the bottom center. Darker pixels represent higher permeability. Note that despite having the advantage of being given the correct correlation structure, the posterior mean of the GP model does not do as effective a job of capturing the full spatial structure of the true field, in particular the two “arms” of

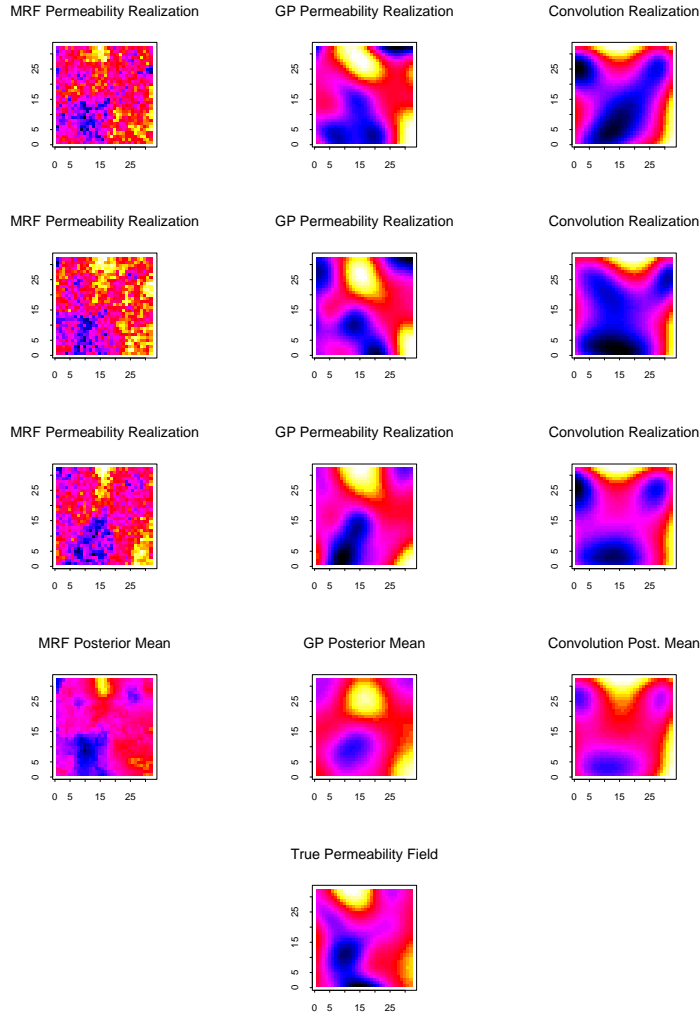


Figure 2: Comparison of three models, MRF, GP, and MRF Convolution. The top three rows are posterior realizations and the fourth row is posterior means. The truth is in the bottom center.

higher permeability, as the MRF and MRF convolution models. Note that these arms are important, since they correlate to paths of water flow, so that merely putting blobs of high permeability in several locations, as the GP model does, is suboptimal. In a sense, the ordinary GP appears to be over-smoothing, in that it favors an oversimplified permeability configuration. On the other hand, the individual posterior realizations from the MRF model are much rougher than the truth, so estimates of uncertainty derived from these may be too large. The MRF convolution model does a good job of capturing a reasonable amount of smoothness in each MCMC realization, and also producing a posterior mean that adequately reflects the true spatial structure. Thus it benefits from the flexibility of the MRF while still retaining the smoothness of a GP.

5 Convolution of Temporally Evolving Processes

The process convolution approach is easily extended to a space-time model by allowing the white noise process to evolve over time, now denoted $x(s, t)$. A simple example would be an autoregressive or random walk process:

$$x(s_i, t) = \beta x(s_i, t-1) + \nu_{i,t},$$

where $\nu_{i,t} \stackrel{iid}{\sim} N(0, \tau^2)$. Thus for $\beta = 1$, this is an independent Gaussian random walk over time at each location. For $0 < \beta < 1$, this is a stationary AR(1) process, and for $\beta = 0$ this reduces to the white noise model from before.

With this model, temporal dependence is captured through the latent process $x(s, t)$, while spatial dependence comes from smoothing this latent process in space using the convolution kernel, k . This model can be extended to a more general Bayesian dynamic spatial model (BDSM) which allows more complicated dependence on the previous time step, e.g.:

$$\mathbf{x}_t = \mathbf{G}(\boldsymbol{\beta})\mathbf{x}_{t-1} + \boldsymbol{\nu}_t \quad \boldsymbol{\nu}_t \sim N(0, \tau^2 \mathbf{I}).$$

Here $\mathbf{G}(\boldsymbol{\beta})$ is a matrix so that each $x(s_i, t)$ depends on a linear combination of any or all of the values of the latent process at the previous time step, \mathbf{x}_{t-1} , and $\boldsymbol{\beta}$ is a vector of parameters for the weights in this linear combination. Under an assumption of Gaussian error, the linear structure of the BDSM allows the complete conditional distributions of the $\mathbf{x}(t)$'s to be computed in closed form given the values of the other model parameters. Implementation is straightforward using the Forward Filtering Backward Sampling (FFBS) algorithm (Carter and Kohn, 1994; Frühwirth-Schnatter, 1994; West and Harrison, 1997). The complete conditional distributions for the other parameters are also available in closed form, allowing the use of Gibbs sampling for inference.

5.1 Ozone Example

Ozone is a common pollutant in urban areas. Concentration of ozone varies over space and time, and may relate to many other factors, including meteorological variables, automobile usage, and industrial activity. While some of these potential covariates, such as temperature, may be easily measured, others such as industrial activity are nearly impossible to measure accurately. Here we propose a space-time model for ozone concentration based only on ozone measurements, trying to learn directly from the data about the spatio-temporal structure of ozone pollution. Such an analysis can be considered an exploratory analysis that may be useful in directing further research. Calder et al. (2001) provide additional discussion of this approach. A similar approach (with a different model) was taken by Meiring et al. (1998).

This dataset consists of ozone measurements for 30 consecutive days of 1999 at 512 (irregularly spaced) sites in the eastern United States. For each day at each site, the maximum eight hour average ozone concentration is recorded. We denote the sites by r_j , $j \in \{1, \dots, 512\}$. We work with the natural logarithm transformation of the ozone measurements, denoted $y(r_j, t)$ for $t \in \{1, \dots, 30\}$.

We model the log ozone concentrations as a convolution of a temporally-evolving latent process. The latent process, \mathbf{x} , is defined over a coarse regular grid of sites s_i . Using a fairly simple spatio-temporal

structure, the value of each point of the latent process depends on the value at that point at the previous timestep as well as the values of the point’s first-order neighbors at the previous timestep:

$$x_{s_i,t} = \beta_i^0 x_{(s_i,t-1)} + \beta_i^N x_{(s_{i_N},t-1)} + \beta_i^E x_{(s_{i_E},t-1)} + \beta_i^S x_{(s_{i_S},t-1)} + \beta_i^W x_{(s_{i_W},t-1)} + \nu_{(i,t)} \quad (3)$$

where s_{i_N} is the site immediately to the north of site s_i , s_{i_E} is the site immediately to the east of site s_i , and so forth. Note that in this model, the β^k are constant over time, but vary over space. The latent process is then convolved with a circular Gaussian kernel, k , and we assume *iid* Gaussian error on the observations:

$$y(r_j, t|\mathbf{x}) = \sum_i k(d(r_j, s_i))x(s_i, t) + \mu + \varepsilon_{r_j,t} \quad \varepsilon_{r_j,t} \stackrel{iid}{\sim} N(0, \sigma^2)$$

where $d(r, s)$ is the distance between points r and s , and μ is an overall mean level. We use conjugate priors when possible, i.e., Gaussian for μ , independent inverse-gammas for σ^2 and τ^2 , and independent Gaussian processes for each β^k . The model is fit via Gibbs sampling using FFBS. Further details are available in Calder et al. (2001).

Figure 3 shows the posterior means for the spatio-temporal coefficients, the β ’s of Equation (3). At each location s_i in the lattice, a circle represents β^0 , the effect of the level at that site from the previous time step, and an arrow is shown for each of the directional β ’s for that point. Those points along the edges of the lattice have fewer than those in the interior. The arrows are plotted with their direction and color indicating the sign of their posterior means. For example, the influence on the latent process at s_i from the process at s_{i_W} (its neighbor to the west) is shown as a black arrow to the east if the value is positive or a grey arrow back toward the west if the value is negative. The sizes of the arrows are scaled such that one degree of latitude or longitude corresponds to .2 on the β scale, e.g., an arrow of length 2° in Figure 3 corresponds to $\beta = .4$. The circles are plotted with their radii on the same scale as the arrow lengths, and in black for positive values and grey for negative values.

Figure 3 displays some obvious patterns. Both the Northeast and the Midwest show strong evidence of spatio-temporal correlation, with ozone moving to the north and east in the Northeast, and to the east and south in the Midwest. These trends mirror standard weather patterns in these regions. In contrast, in the plains and in the South, the spatio-temporal relationship is quite small.

6 Conclusions

The standard Gaussian Process model has seen much success in spatial modeling. Yet many problems present complications, such as non-stationarity, which cannot be accommodated by the standard model. By using a convolution representation and then expanding the class of background processes beyond the usual white noise process, we can achieve a wide variety of spatial processes. This expanded family is sufficiently flexible to deal with many of the common problems with standard GP models.

Large gains in computational efficiency can be made from the discretized version of the convolution. Such discretizations may also be amenable to existing fast algorithms, such as the Forward Filtering Backward Sampling algorithm. The expanded family of processes of this paper are thus both flexible, and computationally tractable.

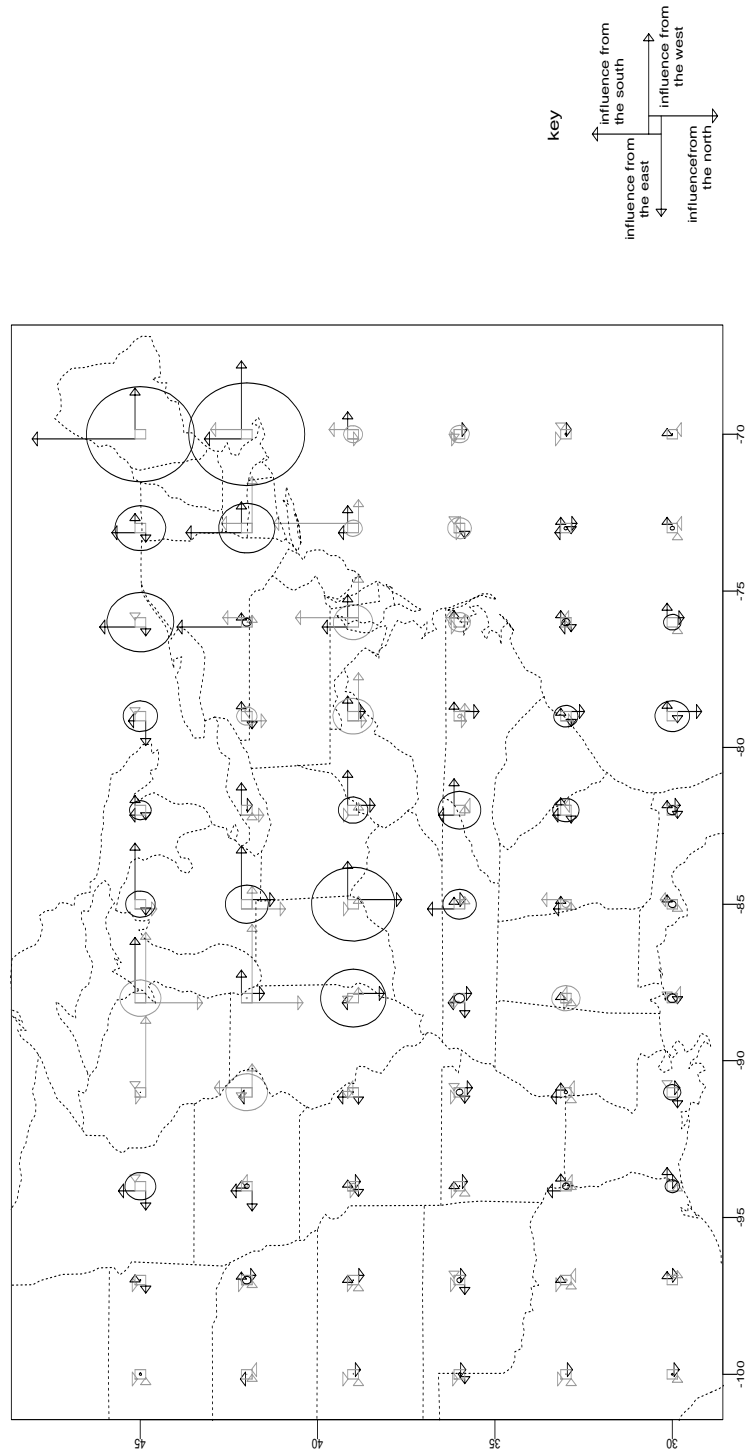


Figure 3: Posterior means of the β 's in arrow form. Black represents a positive posterior mean of a β , and grey represents a negative posterior mean of a β . The arrows indicate strength and direction of influence of neighboring ozone concentration levels from the previous day, and the circle the influence of the level at the same point from the previous day.

Acknowledgments

This work was partially supported by National Science Foundation grant DMS 9873275.

References

- Barry, R. P. and Ver Hoef, J. M. (1996). “Blackbox Kriging: Spatial Prediction Without Specifying Variogram Models.” *Journal of Agricultural, Biological, and Environmental Statistics*, 1, 297–322.
- Besag, J. and Kooperberg, C. (1995). “On Conditional and Intrinsic Autoregressions.” *Biometrika*, 82, 733–746.
- Calder, C. A., Holloman, C., and Higdon, D. (2001). “Exploring Space-Time Structure in Ozone Concentration Using a Dynamic Process Convolution Model.” In *Case Studies in Bayesian Statistics 6*. To appear.
- Carter, C. and Kohn, R. (1994). “On Gibbs Sampling for State Space Models.” *Biometrika*, 81, 541–553.
- Cressie, N. A. C. (1991). *Statistics for Spatial Data*. New York: Wiley-Interscience.
- Frühwirth-Schnatter, S. (1994). “Data Augmentation and Dynamic Linear Models.” *Journal of Time Series Analysis*, 15, 183–202.
- Fuentes, M. and Smith, R. L. (2001). “A new class of nonstationary spatial models.” Tech. Rep. 2534, North Carolina State University, Department of Statistics.
- Higdon, D. (2001). “Space and Space-Time Modeling using Process Convolutions.” Tech. Rep. 01-03, Duke University, Institute of Statistics and Decision Sciences.
- Higdon, D., Swall, J., and Kern, J. (1999). “Non-Stationary Spatial Modeling.” In *Bayesian Statistics 6*, eds. J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith, 761–768. Oxford University Press.
- Ickstadt, K. and Wolpert, R. L. (1999). “Spatial Regression for Marked Point Processes.” In *Bayesian Statistics 6*, eds. J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith, 323–341. Oxford University Press.
- James, A. I., Graham, W. D., Hatfield, K., Rao, P. S. C., and Annable, M. D. (1997). “Optimal Estimation of Residual Non-aqueous Phase Liquid Saturation Using Partitioning Tracer Concentration Data.” *Water Resources Research*, 33, 2621–2636.
- Jin, M., Delshad, M., Dwarakanath, V., McKinney, D. C., Pope, G. A., Sepehrnoori, K., Tilburg, C. E., and Jackson, R. E. (1995). “Partitioning Tracer Test for Detection, Estimation, and Remediation Performance Assessment of Subsurface Non-aqueous Phase Liquids.” *Water Resources Research*, 31, 1201–1211.
- Journel, A. G. and Huijbregts, C. J. (1978). *Mining Geostatistics*. New York: Academic Press.

- King, M. J. and Datta-Gupta, A. (1998). “Streamline Simulation: A Current Perspective.” *In Situ*, 22, 1, 91–140.
- Lee, H., Higdon, D., Bi, Z., Ferreira, M., and West, M. (2002). “Markov Random Field Models for High-Dimensional Parameters in Simulations of Fluid Flow in Porous Media.” *Technometrics*. To appear.
- Matérn, B. (1986). *Spatial Variation*. 2nd ed. New York: Springer-Verlag.
- Matheron, G. (1971). *The Theory of Regionalized Variables and its Applications*. Fontainebleau: Ecole des Mines.
- Meiring, W., Guttorp, P., and Sampson, P. D. (1998). “Space-time Estimation of Grid-cell Hourly Ozone Levels for Assessment of a Deterministic Model.” *Environmental and Ecological Statistics*, 5, 197–222.
- Oliver, D. S., Cunha, L. B., and Reynolds, A. C. (1997). “Markov Chain Monte Carlo Methods for Conditioning a Permeability Field to Pressure Data.” *Mathematical Geology*, 29, 1, 61–91.
- Stein, M. L. (1999). *Interpolation of Spatial Data: Some Theory for Kriging*. New York: Springer-Verlag.
- Stroud, J., Müller, P., and Sansó, B. (2001). “Dynamic Models for Spatio-Temporal Data.” *Journal of the Royal Statistical Society, Series B*, 63, 673–689.
- Thiébaux, H. J. and Pedder, M. A. (1987). *Spatial Objective Analysis with Applications in Atmospheric Science*. London: Academic Press.
- Vasco, D. W., Yoon, S., and Datta-Gupta, A. (1998). “Integrating Dynamic Data Into High-Resolution Reservoir Models Using Streamline-Based Analytic Sensitivity Coefficients.” Society of Petroleum Engineers 1998 Annual Technical Conference, SPE 49002.
- West, M. and Harrison, J. (1997). *Bayesian Forecasting and Dynamic Models*. 2nd ed. New York: Springer-Verlag.
- Xue, G. and Datta-Gupta, A. (1996). “A New Approach to Seismic Data Integration Using Optimal Non-parametric Transformations.” Society of Petroleum Engineers 1996 Annual Technical Conference, SPE 36500.
- Yeh, W. W. (1986). “Review of Parameter Identification in Groundwater Hydrology: the Inverse Problem.” *Water Resources Research*, 22, 95–108.
- Yoon, S., Barman, I., Datta-Gupta, A., and Pope, G. A. (1999). “In-Situ Characterization of Residual NAPL Distribution Using Streamline-Based Inversion of Partitioning Tracer Tests.” SPE/EPA Exploration and Production Environmental Conference, SPE 52729.