Efficient Models for Correlated Data via Convolutions of Intrinsic Processes

Herbert K. H. Lee*(University of California, Santa Cruz)
Dave M. Higdon (Los Alamos National Laboratory)
Catherine A. Calder (The Ohio State University)
Christopher H. Holloman (Battelle Memorial Institute)

June 28, 2004

Abstract

Gaussian processes (GP) have proven to be useful and versatile stochastic models in a wide variety of applications including computer experiments, environmental monitoring, hydrology, and climate modeling. A GP model is determined by its mean and covariance functions. In most cases, the mean is specified to be a constant, or some other simple linear function, while the covariance function is governed by a few parameters. A Bayesian formulation is attractive since it allows for formal incorporation of uncertainty regarding the parameters governing the GP. However, estimation of these parameters can be problematic. Large datasets, posterior correlation, and inverse problems can all lead to difficulties in exploring the posterior distribution. Here we propose an alternative model which is quite tractable computationally—even with large datasets or indirectly observed data—while still maintaining the flexibility and adaptiveness of traditional GP models. This model is based on convolving simple Markov random fields with a smoothing kernel. We consider applications in hydrology and aircraft prototype testing.

Key Words: conditional autoregression, spatial correlation, moving average, inverse problem, nonstationarity

1 Introduction

Gaussian processes (GP) have proven to be effective components of stochastic representations of a wide variety of phenomena. Examples include meteorological fields (Royle et al., 1999), unknown

*Address for correspondence: School of Engineering, UCSC, 1156 High Street, Santa Cruz, CA 95064

functions (O'Hagan, 1991), complex computer model response (Sacks et al., 1989), agricultural fertility gradients (Brownie et al., 1994), and pollutant fields (Host et al., 1995), among many others.

It is common (e.g., Cressie, 1993) to define a GP $z(\mathbf{s}),\ \mathbf{s} \in \mathcal{S}$ over a space $\mathcal{S}$ such that for any finite collection $\{\mathbf{s}_1, \ldots, \mathbf{s}_n\} \in \mathcal{S}$, $(z(\mathbf{s}_1), \ldots, z(\mathbf{s}_n))$ has a multivariate normal distribution with some specified mean function $\mu(\mathbf{s}_1, \ldots, \mathbf{s}_n)$ and covariance matrix $\mathbf{C}$ with elements

$$C(z(\mathbf{s}_i), z(\mathbf{s}_j)) = \frac{1}{\lambda_z}\rho(d), \text{ where } d = \text{dist}(\mathbf{s}_i, \mathbf{s}_j).$$

The correlation function $\rho(d)$ controls the spatial dependence of $z(\mathbf{s})$. For the most part, $\mathcal{S}$ is taken to be R^p and the distance function $\text{dist}(\mathbf{s}_i, \mathbf{s}_j)$ is taken to be Euclidean distance. Other distance metrics are most often based on Euclidean distance on a transformed version of $\mathcal{S}$. Typically, transformations are rotations and dilations (Isaaks and Srivastava, 1989, ch. 16); more generally, spatial deformations (Sampson and Guttorp, 1992; Schmidt and O'Hagan, 2003) can be employed.

The choice of correlation function $\rho(d)$ plays an important role in the specification of the GP $z(\mathbf{s})$. This paper will not concentrate on this aspect of modeling, but will focus on applications for which $z(\mathbf{s})$ is expected to be reasonably smooth. In these cases, the Gaussian correlation function

$$\rho(d; \beta) = \exp\{-\beta d^2\} \tag{1.1.1}$$

is a good starting point. Alternatively, popular choices of $\rho(d)$ include the exponential and Matérn correlation functions. See Cressie (1993), Stein (1999), or Banerjee et al. (2003) for details regarding the choice of correlation function.

For simplicity, we assume the GP $z(\mathbf{s})$ has constant mean $\mu = 0$. When the observed data $\mathbf{y} = (y(\mathbf{s}_1), \ldots, y(\mathbf{s}_n))^T$ are modeled as the sum of the process $z(\mathbf{s})$ and independent white noise, the sampling model for the data is

$$L(\mathbf{y}|\mathbf{z}, \lambda_y) \propto \lambda_y^{\frac{n}{2}} \exp\{-\tfrac{1}{2}\lambda_y(\mathbf{y}-\mathbf{z})^T(\mathbf{y}-\mathbf{z})\}$$

where $\mathbf{z} = (z(\mathbf{s}_1), \ldots, z(\mathbf{s}_n))^T$ is the process $z(\mathbf{s})$ restricted to the observed data locations and λ_y governs the precision of the white noise. The process $\mathbf{z}$ has a GP prior of the form

$$\pi(\mathbf{z}|\lambda_z, \beta) \propto \lambda_z^{\frac{n}{2}} |\mathbf{R}_s(\beta)|^{-\frac{1}{2}} \exp\{-\tfrac{1}{2}\lambda_z \mathbf{z}^T \mathbf{R}_s(\beta)^{-1}\mathbf{z}\}$$

where the ijth element of the $n \times n$ correlation matrix $\mathbf{R}_s(\beta)$ is given by

$$\mathbf{R}_s(\beta)_{ij} = \exp\left\{-\beta\, \text{dist}(\mathbf{s}_i, \mathbf{s}_j)^2\right\}.$$

The formulation is completed by specifying a joint prior density for λ_y, λ_z and β given by $\pi(\lambda_y, \lambda_z, \beta)$. In this case, $\mathbf{z}$ can be integrated out of the posterior so that

$$\pi(\lambda_y, \lambda_z, \beta|\mathbf{y}) \propto |\mathbf{\Lambda}|^{\frac{1}{2}} \exp\left\{-\tfrac{1}{2}\mathbf{y}^T\mathbf{\Lambda}\mathbf{y}\right\} \times \pi(\lambda_y, \lambda_z, \beta) \tag{1.1.2}$$

where $\mathbf{\Lambda}^{-1} = \mathbf{I}_n/\lambda_y + \mathbf{R}_s(\beta)/\lambda_z$.

Functionals of this posterior can be evaluated via Markov chain Monte Carlo (MCMC) or an alternative numerical integration scheme. Any numerical approach becomes problematic as n gets large since evaluation of the posterior requires an inversion (or at least a decomposition and solution) of an $n \times n$ matrix. In addition, strong posterior dependence between λ_z and β can also make numerical integration schemes more difficult (Kern, 2000).

In the case when the data are not modeled as a Gaussian process plus Gaussian noise, a more general likelihood, or sampling density, $L(\mathbf{y}|\mathbf{z}, \lambda_y)$ is required. Here λ_y may hold parameters controlling the sampling model. Examples include spatial models for non-Gaussian data (Besag et al., 1991; Diggle et al., 1998) as well as inverse problems (Oliver et al., 1997; Higdon et al., 2003). In such cases, integrating $\mathbf{z}$ out of the posterior usually cannot be done. This places a much larger burden on the scheme being used to carry out the posterior exploration since a high-dimensional $\mathbf{z}$ must also be taken into account. Typically, sampling the vector $\mathbf{z}$ is required for posterior exploration in applications involving non-Gaussian data. For example, in inverse problems, values of the process $z(\mathbf{s})$ over a grid are usually required as inputs to a forward model which must be run to evaluate the likelihood function. The flow application of this paper is such an example.

In summary, we highlight three issues regarding estimation that can hamper the use of GP models in practice:

- strong posterior dependence regarding λ_z and β, hampering MCMC-based posterior sampling;
- difficulty in dealing with large n due to the matrix inversion required to evaluate the posterior distribution;
- inability to marginalize over the spatial field $z(\mathbf{s})$ in non-Gaussian and inverse problems, leading to high-dimensional posterior distributions which are difficult to explore effectively.

Any combination of theses problems can arise in a given application. For example, researchers studying inverse problems in hydrology have found it very difficult to obtain information regarding the spatial range parameter β due to the combined effects of the first and third points above (Oliver et al., 1997; Lee et al., 2002).

A promising approach for dealing with such problems is to use alternative representations of the underlying GP $z(\mathbf{s})$. Singular value decomposition (Schmidt and O'Hagan, 2003), Cholesky

with pivoting (Lee et al., 2002), and spatial moving average (Higdon, 2002) representations have been used successfully in applications. The lower dimensional representation of $z(\mathbf{s})$ can result in substantial gains in efficiency in the MCMC sampling. A drawback is that these representations typically require that the spatial range parameter β be specified in advance when applied to complex real-world problems, particularly inverse problems.

In this paper, we introduce a modification of the moving average representation which still gives an efficient parameterization of $z(\mathbf{s})$ while allowing inference on the range of spatial dependence. The details of this simple approach are given in the following section, including a comparison to the standard moving average and GP formulations. The usefulness of this modeling approach is then demonstrated on two applications. The first is an inverse problem from hydrology; the second is a prediction problem in aircraft prototype testing. In both applications, the modeling approach developed here allows for estimation of the spatial range in rather complicated settings.

2 Convolutions of stationary and intrinsic processes

A convenient representation of a GP model uses process convolutions (Thiébaux and Pedder, 1987; Barry and Ver Hoef, 1996; Higdon, 2002). One may construct a Gaussian process $z(\mathbf{s})$ over a region $\mathcal{S}$ by convolving a continuous, unit variance, white noise process $x(\mathbf{s})$, with a smoothing kernel $k(\mathbf{s})$:

$$z(\mathbf{s}) = \int_{\mathcal{S}} k(\mathbf{u}-\mathbf{s})x(\mathbf{u})d\mathbf{u}. \tag{2.2.1}$$

The resulting covariance function for $z(\mathbf{s})$ depends only on the displacement vector $\mathbf{d} = \mathbf{s} - \mathbf{s}'$:

$$\mathrm{Cov}(z(\mathbf{s}), z(\mathbf{s}')) = C(\mathbf{d}) = \int_{\mathcal{S}} k(\mathbf{u}-\mathbf{s})k(\mathbf{u}-\mathbf{s}')d\mathbf{u} = \int_{\mathcal{S}} k(\mathbf{u}-\mathbf{d})k(\mathbf{u})d\mathbf{u}.$$

An alternative to specifying a spatially-independent process for $x(\mathbf{s})$, is to use a more general model. If we take $x(\mathbf{s})$ to be an intrinsically stationary process with variogram $\gamma_x(\mathbf{d}) = \mathrm{Var}(x(\mathbf{s}) - x(\mathbf{s}+\mathbf{d}))$ the resulting variogram of the process $z(\mathbf{s})$ induced by Equation (2.2.1) is given by

$$\gamma_z(\mathbf{d}) = \gamma_z^*(\mathbf{d}) - \gamma_z^*(\mathbf{0}) \;\text{ where } \gamma_z^*(\mathbf{q}) = \int_{\mathcal{S}}\int_{\mathcal{S}} k(\mathbf{v}-\mathbf{q})k(\mathbf{u}-\mathbf{v})\gamma_x(\mathbf{u})\,d\mathbf{u}\,d\mathbf{v}. \tag{2.2.2}$$

With this approach, one can fix the smoothing kernel $k(\mathbf{s})$ and then modify the spatial dependence for $z(\mathbf{s})$ by controlling $\gamma_x(\mathbf{d})$. For example, if one specifies $k(\mathbf{s})$ to be a standard normal kernel and $x(\mathbf{s})$ to be 1-d Brownian motion with $\gamma_x(d) = d/\lambda_x$, then the variogram of the resulting spatial process $z(\mathbf{s})$ given by (2.2.2) would be

$$\gamma_{BM}(d) = \frac{1}{\lambda_x}\left[\frac{2}{\sqrt{\pi}}\left(\exp\{-d^2/4\} - 1\right) + d\left(\Phi(d/\sqrt{2}) - \Phi(-d/\sqrt{2})\right)\right]$$

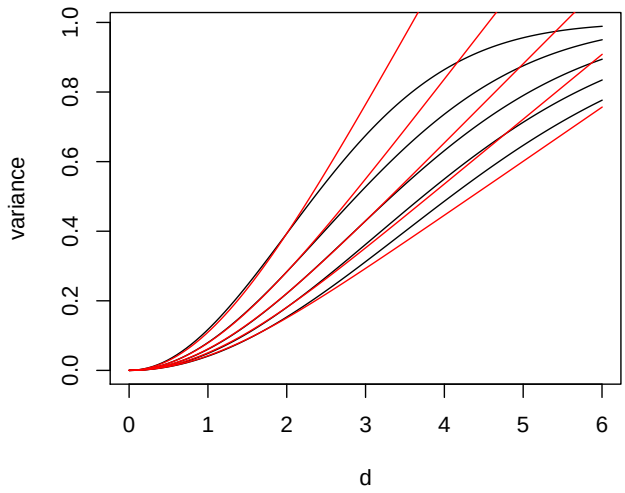

Figure 1: Gaussian variograms (black lines) and similar variograms obtained from convolving Brownian motion with a standard normal kernel (grey lines). The Gaussian variograms are obtained from smoothing white noise with normal kernels whose variances are 2, 3, 4, 5 and 6 going left to right. The variograms corresponding to the smoothed Brownian motion were obtained by choosing λ_x so that they match the Gaussian variograms at $d = 2$.

where $\Phi(\cdot)$ is the standard normal cumulative density function. An important feature is that the spatial dependence is controlled by λ_x which simply scales the variogram. By contrast, if one were to smooth white noise with a normal kernel with variance ω, the resulting variogram would be

$$\gamma_{GP}(d, \omega) \propto 1 - \exp\{-d^2/(4\omega)\} = 1 - \exp\{-\beta d^2\}.$$

This gives the equivalence between smoothing white noise with a normal kernel having variance ω and using the Gaussian correlation function (1.1.1): $\beta = (4\omega)^{-1}$.

Even though $\gamma_{BM}(d)$ gives an intrinsic process $z(\mathbf{s})$, while $\gamma_{GP}(d)$ gives a stationary one, the precision λ_x can be chosen so that $\gamma_{BM}(d)$ and $\gamma_{GP}(d)$ are very close for $d < 2.5$. Figure 1 shows how λ_x can be specified so the variograms match for various values of ω. The key point here is that by rescaling γ_{BM} one can mimic the effect of the spatial range parameter β in a GP model with a Gaussian variogram. This is an important consideration in using MCMC for posterior exploration. We can approximate the modeling capabilities of a GP while ameliorating the computational issues discussed in the previous section.

Intrinsic processes are just one choice of many for the underlying process, and a variety of others can be found in the literature. Alternatives include Gaussian processes (Fuentes and Smith, 2001), non-Gaussian processes (Ickstadt and Wolpert, 1999) and temporally evolving processes (Wikle et al., 1998; Calder et al., 2001).

2.1 Discretization of the Underlying Process

Returning to the general case, we note that in practice the theoretically continuous underlying process can be approximated by a discretized version without much loss in fidelity as long as the discretization is not too coarse relative to the smoothing kernel. To be specific we restrict the support of $x(\mathbf{s})$ to a regular lattice on $\mathcal{S}$ at locations $\mathcal{S}^x = \{\mathbf{s}_1^x, \ldots, \mathbf{s}_m^x\}$. This results in a continuous process

$$z(\mathbf{s}) = \sum_{j=1}^{m} k(\mathbf{s}_j^x - \mathbf{s}) x(\mathbf{s}_j^x) \tag{2.2.3}$$

that is controlled by the m-vector $\mathbf{x} = (x(\mathbf{s}_1^x), \ldots, x(\mathbf{s}_m^x))^T$. This discrete, low-dimensional representation can result in substantial computational gains in inverse problems or applications with large amounts of data (Calder, 2003). This discrete representation will be a very good approximation to the continuous one, provided the lattice spacings are no greater than the standard deviation of the normal kernel $k(\mathbf{s})$ (Higdon, 2002).

2.2 Simple example

As an example we compare simple 1-d formulations using a synthetic dataset. We generate $n = 12$ observations $\mathbf{y}$ at locations $s_1^y, \ldots, s_n^y$ equally spaced between 0 and 10 according the model

$$y(s_i^y) = z(s_i^y) + \epsilon_i, \; i = 1, \ldots, n$$

where $z(s)$ is a mean zero Gaussian process whose covariance function is given by $C(d) = \exp\{-(d/5)^2\}$ (this corresponds to smoothing a white noise process with a $N(0, (5/2)^2)$ density kernel) and the ϵ_is are *iid* $N(0, 0.2^2)$ random variates.

We first consider two model formulations, both of which fix $k(s)$ to be a $N(0, 0.6^2)$ density. The width of $k(s)$ is slightly more than four times narrower than the width that corresponds to the model that actually generated $z(s)$. The goal here is to see how altering the dependence structure for $x(s)$ can affect our estimate for the underlying process $z(s)$. In both formulations, the latent process $x(s)$ is specified to be non-zero over $m = 20$ regular lattice points $s_1^x, \ldots, s_m^x$ between -2 and 12. The first formulation specifies that each $x(s_j^x)$ is *iid* $N(0, 1/\lambda_x)$; the second specifies $x(s)$ to be a random walk over the same $m = 20$ lattice points, with increments scaled by the same precision parameter λ_x. We take the m-vector $\mathbf{x}$ to be $(x(s_1^x), \ldots, x(s_m^x))^T$ and note that the n-vector $\mathbf{z} = (z(s_1), \ldots, z(s_n))^T)$ is given by $\mathbf{Kx}$ where $\mathbf{K}$ is a $n \times m$ matrix whose elements are given

by $K_{ij} = k(s_j^x - s_i)$. The formulation is

$$\begin{aligned} L(\mathbf{y}|\mathbf{z}, \lambda_y) &\propto \lambda_y^{\frac{n}{2}} \exp\{-\tfrac{1}{2}\lambda_y(\mathbf{y} - \mathbf{K}\mathbf{x})^T(\mathbf{y} - \mathbf{K}\mathbf{x})\} \\ \pi(\mathbf{x}|\lambda_x) &\propto \lambda_x^{\frac{m}{2}} \exp\{-\tfrac{1}{2}\lambda_x x^T \mathbf{W} x\} \\ \pi(\lambda_x) &\propto \lambda_x^{a_x - 1} e^{-b_x \lambda_x} \\ \pi(\lambda_y) &\propto \lambda_y^{a_y - 1} e^{-b_y \lambda_y} \end{aligned}$$

where the precision matrix $\mathbf{W}$ is the $m \times m$ identity matrix $\mathbf{I}_m$ if $x(s)$ is the *iid* process, while for the random walk specification $\mathbf{W}$ is a tridiagonal matrix with -1's on the off-diagonals, and 2's on the diagonals, except for the first and final diagonal entries, which are 1. The resulting full conditionals are then

$$\begin{aligned} \mathbf{x}|\lambda_y, \lambda_x, \mathbf{y} &\sim N\left((\lambda_y \mathbf{K}^T\mathbf{K} + \lambda_x \mathbf{W})^{-1}\lambda_y \mathbf{K}^T \mathbf{y}, (\lambda_y \mathbf{K}^T\mathbf{K} + \lambda_x \mathbf{W})^{-1}\right) \\ \lambda_x|\mathbf{x}, \lambda_y, \mathbf{y} &\sim \Gamma(a_x + m/2, b_x + \tfrac{1}{2}\mathbf{x}^T\mathbf{W}\mathbf{x}) \\ \lambda_y|\mathbf{x}, \lambda_x, \mathbf{y} &\sim \Gamma(a_y + n/2, b_y + \tfrac{1}{2}(\mathbf{y} - \mathbf{K}\mathbf{x})^T(\mathbf{y} - \mathbf{K}\mathbf{x})). \end{aligned}$$

For comparison, we also consider two other formulations: a white noise process for $x(s)$ with $k(s)$ specified to be a $N(0, (5/2)^2)$ density which matches the actual Gaussian process model used to generate the data, and a basic GP formulation resulting in the posterior given by (1.1.2). For the GP formulation, $\Gamma(1, 0.001)$ priors were specified for λ_y and λ_z. An exponential prior with a mean of $1/25$ was specified for the spatial range parameter β.

Figure 2 shows the resulting posterior summaries for $z(s)$ under the three formulations. The first formulation tends to overfit the data since the spatial dependence induced by the "skinny" $N(0, 0.6^2)$ kernel dies off too quickly to give a sensible representation of the $z(s)$ process that was used to produce the actual data. The random walk formulation can overcome this misspecification of $k(s)$ through the dependence structure in $x(s)$. Because the precision parameter λ_x in the random walk formulation can modify the amount of spatial dependence in the induced $z(s)$, its reconstruction better matches the reconstruction which uses a wide kernel $k(s)$ (third row of figure) which corresponds to how the data were generated. The wide kernel model gets the unrealistic advantage of using the correct kernel width (range of spatial dependence). A more honest comparison is to the GP model which treats β as an unknown parameter (bottom row of Figure 2).

Pointwise 80% posterior credible intervals for the variograms from the four spatial formulations are shown in Figure 3. Notice that the posterior for the variogram under the random walk formulation is very similar to that of the GP formulation. This suggests the simpler random walk formulation gives very similar posterior information regarding the spatial structure of $z(s)$ as compared to the conventional GP formulation which uses both λ_z and β to model the spatial dependence in $z(s)$.

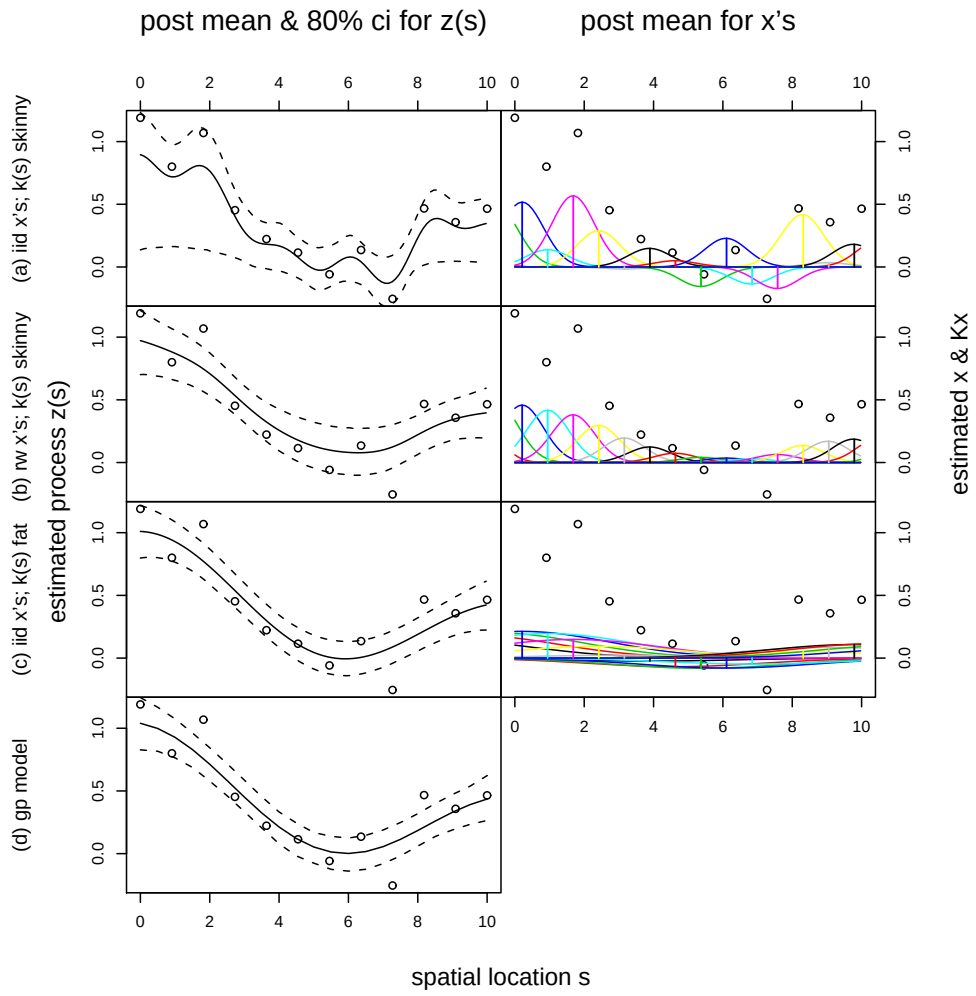

Figure 2: The left column compares posterior means and pointwise 80% credible intervals for $z(s)$ under four formulations: (a) white noise specification for $x(s)$, $k(s) \propto \exp\{-0.5(s/0.6)^2\}$; (b) random walk specification for $x(s)$, $k(s) \propto \exp\{-0.5(s/0.6)^2\}$; (c) white noise specification for $x(s)$, $k(s) \propto \exp\{-0.5(s/2.5)^2\}$; and (d) a GP model with Gaussian covariance function, treating covariance parameters as unknown. The right column shows the corresponding posterior mean estimates for each $x(s_i^x)$ (vertical lines) as well as the decomposition of the posterior mean estimate for $z(s)$ into its basis components $k(s - s_i^x)x(s_i^x)$, $i = 1, \ldots, m$. The plotting symbols denote the data $\mathbf{y}$.

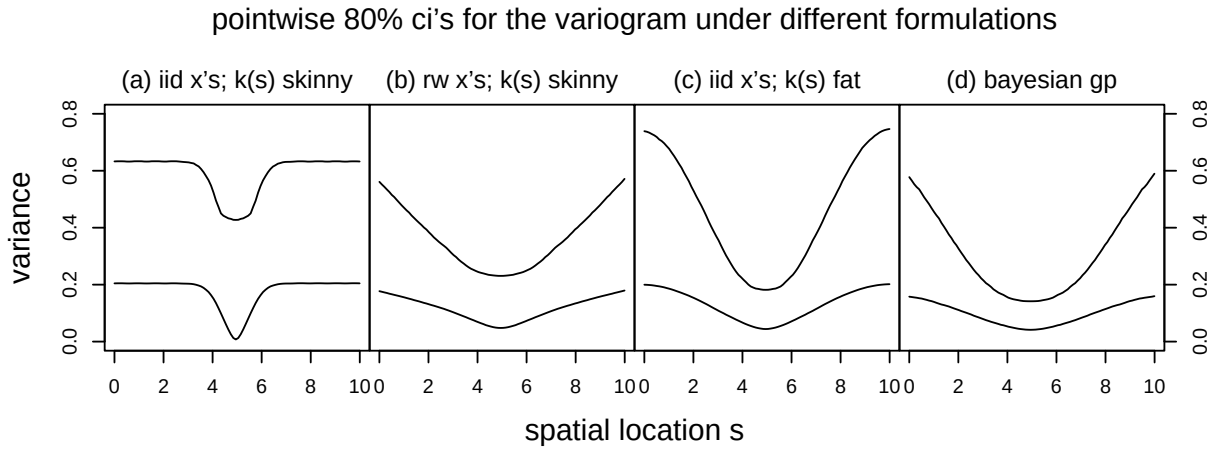

Figure 3: Pointwise 80% posterior credible intervals for the variogram – defined as $\gamma(s) = \mathrm{Var}(y(s) - y(5))$ – under the different formulations: (a) white noise specification for $x(s)$, $k(s) \propto \exp\{-0.5(s/0.6)^2\}$; (b) random walk specification for $x(s)$, $k(s) \propto \exp\{-0.5(s/0.6)^2\}$; (c) white noise specification for $x(s)$, $k(s) \propto \exp\{-0.5(s/2.5)^2\}$; and (d) conventional Gaussian process model with $C(d) \propto \exp\{-\beta d^2\}$. Note that the two models with adaptive spatial dependence ((b) and (d)) give similar estimates.

It is worth noting here that this sort of approach provides a computationally attractive alternative to direct Bayesian modeling of a Gaussian process with a correlogram that includes scaling parameters. An MCMC scheme involving scaling parameters will require a matrix inversion for each update. By comparison, the updates of λ_x in this random walk scheme (which is easily generalized to higher dimensions) require no inversions, just a scan through the latent $\mathbf{x}$ vector.

2.3 Convolving intrinsic Gaussian MRF's

The above random walk convolution formulation can easily be generalized to higher dimensions by specifying an appropriate lattice $\mathcal{S}_x$ for the support of $x(\mathbf{s})$. One can then specify a Gaussian Markov random field (MRF) prior for the m-vector $\mathbf{x}$ over lattice locations so that

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

where the precision λ_x scales $\mathbf{x}$, and $\mathbf{W}$ is a sparse matrix that specifies the conditional dependence structure in $\mathbf{x}$. This model is also called an intrinsic autoregressive model. Our applications use a first order MRF (Besag and Kooperberg, 1995) so that $\mathbf{W}$ has elements

$$W_{jk} = \begin{cases} -1 & \text{if } \mathbf{s}_j^x \text{ and } \mathbf{s}_k^x \text{ are adjacent} \\ n_j & \text{if } \mathbf{s}_j^x = \mathbf{s}_k^x \\ 0 & \text{otherwise} \end{cases}$$

where n_j is the number of lattice sites $\mathbf{s}_k^x$ directly adjacent to site $\mathbf{s}_j^x$. This first order $\mathbf{W}$ reduces to the random walk formulation in 1-d. Alternatives such as locally quadratic MRF's (Besag and Kooperberg, 1995) could also be used here. The resulting full conditionals are exactly those given in the example in the previous section. In cases were m is large, the multivariate update for $\mathbf{x}$ may have to be replaced with a modified updating scheme, such as single site updating.

3 Application in Subsurface Flow

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 wells 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

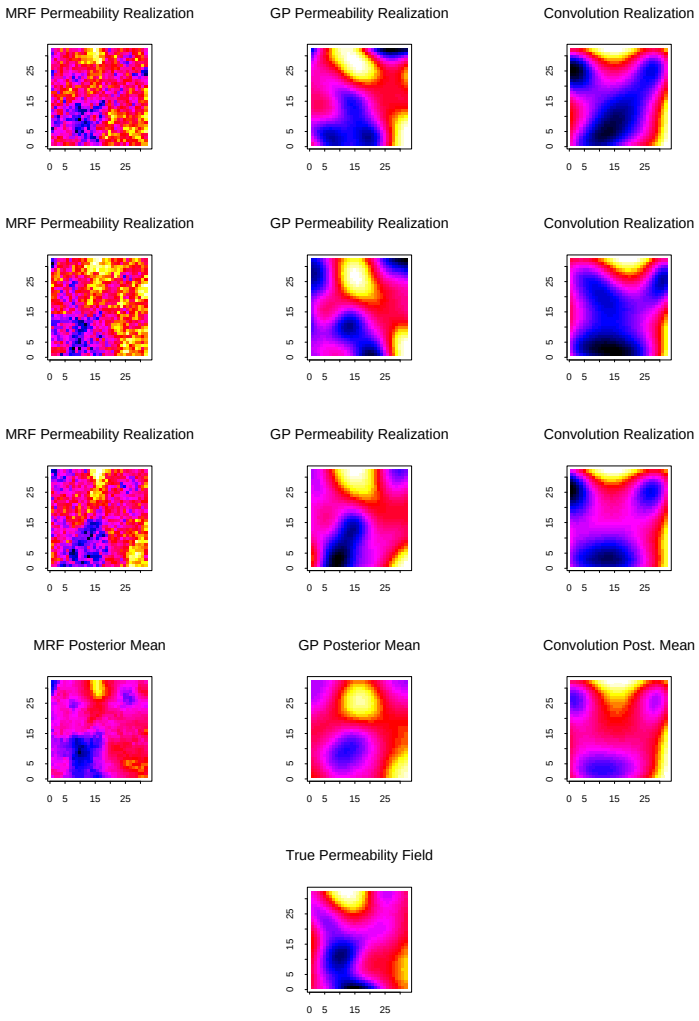

Figure 4: Comparison of three models: MRF, GP, and Convolution of MRF. The top three rows are posterior realizations and the fourth row is posterior means. The truth is in the bottom center.

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, $\mathbf{b}$, 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 for $\mathbf{z} = \log(\boldsymbol{\psi})$: a Markov random field, a Gaussian process, and the convolution of an MRF model of the previous section.

The plots in Figure 4 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 variogram, 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 λ_x 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 Gaussian process model, using the correct values for its spatial dependence parameters (since these cannot be effectively estimated from flow data, see for example, the discussion in Oliver, Cunha, and Reynolds, 1997). The last full column is for the new convolution of an MRF model described in the previous section, where λ_x 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. All three models are able to capture the main features of the true permeability field, although the MRF produces realizations which are much less smooth.

The GP model was fit under the very unrealistic supposition that the spatial dependence range parameter β is known. In practice this is never the case. Both the MRF and convolution of MRF models allow the spatial dependence to be treated as unknown and its uncertainty is accounted for in the posterior. Of these two models, the MRF model leads to more difficulties in MCMC im-

plementation because of its high-dimensional representation of $\mathbf{z}$. The parsimonious representation of $\mathbf{z}$ in the convolution formulation leads to a simpler, and far more efficient MCMC scheme. It uses only basic Metropolis updates for components of $\mathbf{z}$, and it samples its posterior an order of magnitude more quickly relative to the MRF formulation.

4 Store separation for a prototype aircraft

This section develops a space-time model that relies heavily on the process convolution approach to make tractable the computations necessary for estimation.

The U. S. Air Force needs to ensure that store launches are safe for prototype aircraft under development. This application looks at launches of the AIM-9M missile from the side bay of the F-22 prototype. Upon launch, the missile is propelled along a guiding rail and then enters into free flight. The guidance system of the missile kicks in shortly after the launch and steers the missile/store on a straight path in the direction of the flying aircraft.

Prior to any launches on the actual aircraft, both computational fluid dynamics (CFD) simulations and wind tunnel simulations are carried out for various flight conditions of the aircraft. The flight conditions are indexed by dynamic pressure, altitude, G-force on the aircraft, and angle of attack. These flight conditions are denoted by the 4-variate parameter $\mathbf{s}$.

The missile trajectory can be described by a time series of three center of mass coordinates and three orientation coordinates ($\mathsf{x}, \mathsf{y}, \mathsf{z}$,$\mathsf{yaw}$,$\mathsf{pitch}$,$\mathsf{roll}$). Only the first 0.3 seconds of the trajectory are considered here since any danger to the aircraft from the store launch is most likely to occur during this interval.

The solid lines in Figure 6 show the yaw angle trajectories from flight tests, CFD simulations, and wind tunnel simulations as a function of the flight condition $\mathbf{s}$. In the figure, dynamic pressure and altitude are varying, while the aircraft is flying at a constant load of 1 g. The angle of attack is fairly steep ($\sim$ 26 degrees above vertical) for flight conditions in the low pressure-high altitude region, and becomes nearly level ($\sim$ 0 degrees) for the high pressure-low altitude conditions.

In this example we follow the conceptual framework of using a GP or related model for analyzing computer experiments (Sacks et al., 1989; Currin et al., 1991). Information from physical experiments can then be directly incorporated into the model (Kennedy and O'Hagan, 2001; Reese et al., 2004).

4.1 Data

Data for this analysis are $n_{\text{FT}} = 3$ flight test trajectories $y_{\text{FT}}(\mathbf{s}, t)$, $n_{\text{WT}} = 56$ wind tunnel trajectory simulations $y_{\text{WT}}(\mathbf{s}, t)$, and $n_{\text{CFD}} = 3$ CFD trajectory simulations $y_{\text{CFD}}(\mathbf{s}, t)$. Each of these three data types are modeled as an "ideal" trajectory $\eta(\mathbf{s}, t)$ plus a discrepancy term that is specific to each data type.

$$\begin{aligned} y_{\text{FT}}(\mathbf{s}, t) &= \eta(\mathbf{s}, t) + \delta_{\text{FT}}(\mathbf{s}, t) + e_{\text{FT}} \\ y_{\text{WT}}(\mathbf{s}, t) &= \eta(\mathbf{s}, t) + \delta_{\text{WT}}(\mathbf{s}, t) + e_{\text{WT}} \\ y_{\text{CFD}}(\mathbf{s}, t) &= \eta(\mathbf{s}, t) + \delta_{\text{CFD}}(\mathbf{s}, t) + e_{\text{CFD}} \end{aligned}$$

The discrepancy terms $\delta_{\text{WT}}(\mathbf{s}, t)$ and $\delta_{\text{CFD}}(\mathbf{s}, t)$ account for systematic discrepancy between the simulation and the ideal trajectory $\eta(\mathbf{s}, t)$. It is expected that these discrepancies are continuous over flight condition $\mathbf{s}$ and time t. The term $\delta_{\text{FT}}(\mathbf{s}, t)$ allows correlation over time (but not flight condition) and accounts for replicate variability and correlated observation error due to the process of estimating displacement and orientation of the store from video of the flight test launch. Independent observation error is accounted for with the term e_{FT} which is modeled as independent over $\mathbf{s}$ and t and is attributable mainly to the resolution of the photography used to determine the flight test trajectories. Similarly the wind tunnel and CFD trajectories are also allowed an *iid* error term to account for slight mismatch due to the basis representation of the trajectories described below.

In all, the data comprise $n = n_{\text{FT}} + n_{\text{WT}} + n_{\text{CFD}}$ trajectories of the $p_s = 6$ degrees of freedom (x,y,z,roll,pitch,yaw) describing the store position and orientation over time. It is convenient to take $\mathbf{s}$ to denote these n flight conditions where data have been recorded or simulated. We define

$$\mathbf{s} = \begin{pmatrix} \mathbf{s}_1 \\ \vdots \\ \mathbf{s}_n \end{pmatrix} = \begin{pmatrix} \mathbf{s}_{\text{FT}} \\ \mathbf{s}_{\text{WT}} \\ \mathbf{s}_{\text{CFD}} \end{pmatrix} = \begin{pmatrix} s_{11} & \cdots & s_{1p_s} \\ \vdots & \ddots & \vdots \\ s_{n1} & \cdots & s_{np_s} \end{pmatrix}.$$

4.2 Trajectory model for $\eta(\mathbf{s}, t)$

We define an ideal trajectory $\eta(\mathbf{s}, t)$ which would be the average of repeated flight tests carried out at the flight configuration $\mathbf{s}$. This ideal trajectory can be decomposed into a mean trajectory $\eta_0(t)$—which is the same for any flight condition $\mathbf{s}$—and a component $\eta_1(\mathbf{s}, t)$ that also varies with $\mathbf{s}$:

$$\eta(\mathbf{s}, t) = \eta_0(t) + \eta_1(\mathbf{s}, t).$$

We model $\eta_0(t)$ as a convolution of an MRF . The smoothing kernel $k(\mathbf{s}) = \left(1 - \frac{||\mathbf{s}||^3}{\lambda_k^3}\right)^3 I[|s| < \lambda_k]$ is a tricube kernel (Cleveland, 1979) whose width λ_k is specified separately for each trajectory

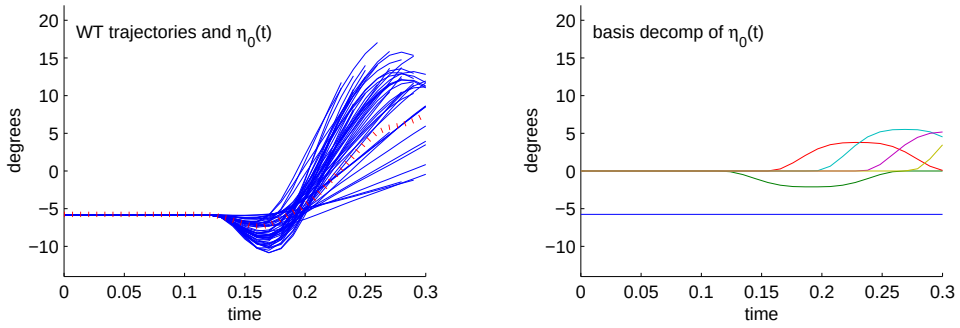

Figure 5: Left frame: An estimate for the mean yaw angle trajectory $\eta_0(t)$ (dotted line) is shown with the yaw angle trajectories for the wind tunnel simulations (solid lines). Right frame: The basis decomposition for $\eta_0(t) = \sum_j k_j(t) x_{\eta_0 j}$, with the lines showing the basis terms $k_j(t) x_{\eta_0 j}$.

type. A 1-d MRF is specified for $x(t)$ with support at time locations $t_1^x, \ldots, t_{p_\eta}^x$. Typically p_η is between 5 and 9, depending on which of the six trajectory components are being modeled. The mean trajectory component $\eta_0(t)$ is modeled by

$$\eta_0(t) = \mu + \sum_{j=1}^{p_\eta} k(t_j^x - t) x_{\eta_0 j} = \sum_{j=0}^{p_\eta} k_j(t) x_{\eta_0 j} \tag{4.4.1}$$

where $k_0(t) = 1$, $k_j(t) = k(t_j^x - t)$ and $x_{\eta_0 0} = \mu$. We restrict $\eta_0(t)$ to 30 time points $t_1, \ldots, t_{30}$ evenly spaced between 0 and 0.3 seconds. The 30-vector $\boldsymbol{\eta}_0$ can then be written as

$$\boldsymbol{\eta}_0 = \mathbf{K}_{\eta_0} \mathbf{x}_{\eta_0}$$

where $\mathbf{K}_{\eta_0}$ is a $30 \times (p_\eta + 1)$ matrix with ij elements $k_j(t_i)$. Figure 5 shows the decomposition of $\boldsymbol{\eta}_0$ for the yaw angle trajectory.

The process $\eta_1(\mathbf{s}, t)$ is constructed similarly

$$\eta_1(\mathbf{s}, t) = \sum_{j=1}^{p_\eta} k_j(t) x_{\eta_1 j}(\mathbf{s}) \tag{4.4.2}$$

but now the basis scaling is controlled by the p_η-dimensional process $x_{\eta_1}(\mathbf{s})$ which varies with flight condition $\mathbf{s}$; the constant term is not required. As with $\eta_0(t)$, we represent $\eta_1(\mathbf{s}, t)$ at any given flight condition $\mathbf{s}$ by the 30-vector $\boldsymbol{\eta}_1(\mathbf{s})$ given by

$$\boldsymbol{\eta}_1(\mathbf{s}) = \mathbf{K}_{\eta_1} \mathbf{x}_{\eta_1}(s)$$

where $\mathbf{K}_{\eta_1}$ is now the $30 \times p_\eta$ matrix with ij elements $k_j(t_i)$. The MRF coefficient vector $\mathbf{x}_{\eta_1}(\mathbf{s})$ varies with flight condition. For a collection of n flight conditions $\mathbf{s} = (\mathbf{s}_1, \ldots, \mathbf{s}_n)^T$, we take $\boldsymbol{\eta}_1(\mathbf{s})$ to be the $30n$-vector which holds n trajectories. In this case $\boldsymbol{\eta}_1(\mathbf{s})$ can be written as

$$\boldsymbol{\eta}_1(\mathbf{s}) = (\mathbf{I}_n \otimes \mathbf{K}_{\eta_1}) \mathbf{x}_{\eta_1}(\mathbf{s})$$

where $\mathbf{x}_{\eta_1}(\mathbf{s})$ is now a np_η-vector determining the n trajectories.

The models for $\boldsymbol{\eta}_0$ and $\boldsymbol{\eta}_1(\mathbf{s})$ are completed by specifying prior distributions for $\mathbf{x}_{\eta_0}$ and $\mathbf{x}_{\eta_1}(\mathbf{s})$. We specify an MRF prior for $\mathbf{x}_{\eta_0}$. For $\mathbf{x}_{\eta_1}(\mathbf{s})$, we use a combination process prior for which the correlation function depends on the flight conditions at which simulations or flight tests have been carried out. Restricting to the data consisting of n trajectories, $\eta(\mathbf{s}, t)$ is determined by the $p_\eta + 1$-vector $\mathbf{x}_{\eta_0}$ and the $p_\eta n$-vector $\mathbf{x}_{\eta_1}$. The implied model for these vectors is then given by

$$\begin{aligned} \mathbf{x}_{\eta_0} &\sim N(\boldsymbol{\mu}, \text{diag}(\sigma_\mu^2, \boldsymbol{\Sigma}_t(\boldsymbol{\lambda}_{\eta_0}))) \\ \mathbf{x}_{\eta_1} &\sim N(\mathbf{0}, \boldsymbol{\Sigma}_t(\boldsymbol{\lambda}_{\eta_1}) \otimes \mathbf{R}_s(\boldsymbol{\beta}_\eta)) \end{aligned}$$

where

- $\boldsymbol{\Sigma}_t(\boldsymbol{\lambda})$ is a $p_\eta \times p_\eta$ covariance matrix for a random walk $\{x_1, \ldots, x_{p_\eta}\}$ for which $x_1 \sim N(0, 1/\lambda_1)$, and the following independent increments $x_j - x_{j-1} \sim N(0, 1/\lambda_j)$, $j = 2, \ldots, p_\eta$. Hence the covariance matrix is defined by $\boldsymbol{\Sigma}_t(\boldsymbol{\lambda}) = \mathbf{W}_t^{-1}(\boldsymbol{\lambda})$ where the ijth element of tridiagonal precision matrix $\mathbf{W}_t(\boldsymbol{\lambda})$ is given by

$$\mathbf{W}_t(\boldsymbol{\lambda})_{ij} = \begin{cases} -\lambda_{\max(i,j)} & \text{if } |i-j| = 1 \\ \lambda_i + \lambda_{i+1} & \text{if } i = j \text{ and } i < p_\eta \\ -\lambda_i & \text{if } i = j = p_\eta \\ 0 & \text{otherwise} \end{cases}$$

- $\mathbf{R}_s(\boldsymbol{\beta})$ specifies a $n \times n$ "spatial" correlation that is a function of the $n \times p_s$ flight condition matrix s. The p_s-dimensional parameter vector $\boldsymbol{\beta}$ specifies the strength of the spatial dependence in the coordinate directions of the flight condition space. The ijth element of $\mathbf{R}_s(\boldsymbol{\beta})$ is given by

$$\mathbf{R}_s(\boldsymbol{\beta})_{ij} = \exp\left\{ -\sum_{k=1}^{p_s} \beta_k (s_{ik} - s_{jk})^2 \right\}. \tag{4.4.3}$$

We set all elements of $\boldsymbol{\lambda}_{\eta_0}$ to be equal to λ_0, and then independent $\Gamma(a_{\eta j}, b_{\eta j})$ priors are specified for λ_0 and for the components of $\boldsymbol{\lambda}_{\eta_1}$. Rather diffuse priors are used here ($a_{\eta j} = 1, b_{\eta j} = 0.001$) since the total number of trajectories in the data will give good information regarding $\boldsymbol{\lambda}_\eta$. We specify a uniform prior over five preselected values for each dimension of the p_s-vector $\boldsymbol{\beta}$ which controls the strength of dependence along each of the flight condition coordinates. Note that this Gaussian process prior for $\mathbf{x}_{\eta_1}(\mathbf{s})$ is particularly useful because it gives predictions for new trajectories at untried flight conditions $\mathbf{s}^*$.

4.3 Discrepancy models

Each of the discrepancy terms $\delta_{\text{FT}}(\mathbf{s},t)$, $\delta_{\text{WT}}(\mathbf{s},t)$, and $\delta_{\text{CFD}}(\mathbf{s},t)$ are modeled with a MRF convolution representation which is identical to that of $\eta_1(\mathbf{s},t)$ given in (4.4.2). The difference lies in the prior specifications for the terms $\mathbf{x}_{\delta_{\text{FT}}}(\mathbf{s}_{\text{FT}})$, $\mathbf{x}_{\delta_{\text{WT}}}(\mathbf{s}_{\text{WT}})$, $\mathbf{x}_{\delta_{\text{CFD}}}(\mathbf{s}_{\text{CFD}})$, which are vectors of length $p_\eta n_{\text{FT}}$, $p_\eta n_{\text{WT}}$, and $p_\eta n_{\text{CFD}}$ respectively. Recall that p_η is the number of MRF elements used in (4.4.2). The prior specification for the discrepancies is summarized below:

basis representation	size	prior for x	size
$\boldsymbol{\delta}_{\text{FT}} = (\mathbf{I}_{n_{\text{FT}}} \otimes \mathbf{K}_{\delta_{\text{FT}}})\mathbf{x}_{\delta_{\text{FT}}}$	$30n_{\text{FT}}$	$\mathbf{x}_{\delta_{\text{FT}}} \sim N(\mathbf{0}, \boldsymbol{\Sigma}_t(\boldsymbol{\lambda}_{\text{FT}}) \otimes \mathbf{I}_{n_{\text{FT}}})$	$p_\eta n_{\text{FT}}$
$\boldsymbol{\delta}_{\text{WT}} = (\mathbf{I}_{n_{\text{WT}}} \otimes \mathbf{K}_{\delta_{\text{WT}}})\mathbf{x}_{\delta_{\text{WT}}}$	$30n_{\text{WT}}$	$\mathbf{x}_{\delta_{\text{WT}}} \sim N(\mathbf{0}, \boldsymbol{\Sigma}_t(\boldsymbol{\lambda}_{\text{WT}}) \otimes \mathbf{R}_{s_{\text{WT}}}(\beta))$	$p_\eta n_{\text{WT}}$
$\boldsymbol{\delta}_{\text{CFD}} = (\mathbf{I}_{n_{\text{CFD}}} \otimes \mathbf{K}_{\delta_{\text{CFD}}})\mathbf{x}_{\delta_{\text{CFD}}}$	$30n_{\text{CFD}}$	$\mathbf{x}_{\delta_{\text{CFD}}} \sim N(\mathbf{0}, \boldsymbol{\Sigma}_t(\boldsymbol{\lambda}_{\text{CFD}}) \otimes \mathbf{R}_{s_{\text{CFD}}}(\beta))$	$p_\eta n_{\text{CFD}}$

Here $\mathbf{R}_{s_\ell}(\boldsymbol{\beta})$ is the correlation matrix constructed using (4.4.3) and restricted to the flight conditions $\mathbf{s}_\ell$ for $\ell \in \{\mathsf{FT}, \mathsf{WT}, \mathsf{CFD}\}$. Recall the covariance matrix $\boldsymbol{\Sigma}_t(\boldsymbol{\lambda})$ specifies a random walk (beginning at 0) whose successive increments are $N(0, 1/\lambda_j)$, $j = 1, \ldots, p_\eta$. Each component of the p_η-vector $\boldsymbol{\lambda}_\ell$ is given an independent $\Gamma(a_{\ell j}, b_{\ell j})$ prior which was elicited with the help of flight engineers and experience from previous tests involving other aircraft. Note we use ℓ as a placeholder for the datatype: FT, WT, or CFD.

4.4 Error models

The error terms $\mathbf{e}_{\text{FT}}$, $\mathbf{e}_{\text{WT}}$, and $\mathbf{e}_{\text{CFD}}$ are modeled as white noise vectors, each with its own precision parameter κ_{FT}, κ_{WT}, and κ_{CFD}. The prior formulation is then

$$\begin{aligned}
\mathbf{e}_{\text{FT}} &\sim N(\mathbf{0}, \mathbf{I}_{30n_{\text{FT}}}/\kappa_{\text{FT}}) & \kappa_{\text{FT}} &\sim \Gamma(a_{e_{\text{FT}}}, b_{e_{\text{FT}}}) \\
\mathbf{e}_{\text{WT}} &\sim N(\mathbf{0}, \mathbf{I}_{30n_{\text{WT}}}/\kappa_{\text{WT}}) & \kappa_{\text{WT}} &\sim \Gamma(a_{e_{\text{WT}}}, b_{e_{\text{WT}}}) \\
\mathbf{e}_{\text{CFD}} &\sim N(\mathbf{0}, \mathbf{I}_{30n_{\text{CFD}}}/\kappa_{\text{CFD}}) & \kappa_{\text{CFD}} &\sim \Gamma(a_{e_{\text{CFD}}}, b_{e_{\text{CFD}}}).
\end{aligned}$$

We specify the a_{e_ℓ}'s to be 1 and the b_{e_ℓ}'s to be 0.0001.

4.5 Posterior distribution and estimation

The resulting posterior distribution has the form:

$$\begin{aligned}
\pi(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\beta}, \kappa | \mathbf{y}) \;\propto\; & |\mathbf{W}_e|^{\frac{1}{2}} \exp\left\{-\tfrac{1}{2}(\mathbf{y} - \mathbf{K}\mathbf{x})^T \mathbf{W}_e (\mathbf{y} - \mathbf{K}\mathbf{x})\right\} \times |\mathbf{W}_x|^{\frac{1}{2}} \exp\left\{-\tfrac{1}{2}\mathbf{x}^T \mathbf{W}_x \mathbf{x}\right\} \times \\
& \prod_{\ell \in \{\text{FT,WT,CFD}\}} \prod_{j=1}^{p_\eta} \lambda_{\ell j}^{a_{\ell j}-1} \exp\{-b_{\ell j}\}\lambda_{\ell j} \times \\
& \prod_{\ell \in \{\text{FT,WT,CFD}\}} \kappa_\ell^{a_{e_\ell}-1} \exp\{-b_{e_\ell}\}\kappa_\ell \times \pi(\boldsymbol{\beta})
\end{aligned}$$

where

$$\mathbf{K} = [\mathbf{1}_n \otimes [\mathbf{K}_{\eta_0} | \mathbf{K}_{\eta_1}] | \mathrm{diag}(\mathbf{I}_{n_{\mathrm{FT}}} \otimes \mathbf{K}_{\delta_{\mathrm{FT}}}, \mathbf{I}_{n_{\mathrm{WT}}} \otimes \mathbf{K}_{\delta_{\mathrm{WT}}}, \mathbf{I}_{n_{\mathrm{CFD}}} \otimes \mathbf{K}_{\delta_{\mathrm{CFD}}})]$$

$$\mathbf{y} = \begin{pmatrix} \mathbf{y}_{\mathrm{FT}} \\ \mathbf{y}_{\mathrm{WT}} \\ \mathbf{y}_{\mathrm{CFD}} \end{pmatrix}, \quad \mathbf{x} = \begin{pmatrix} \mathbf{x}_{\eta_0} \\ \mathbf{x}_{\eta_1} \\ \mathbf{x}_{\delta_{\mathrm{FT}}} \\ \mathbf{x}_{\delta_{\mathrm{WT}}} \\ \mathbf{x}_{\delta_{\mathrm{CFD}}} \end{pmatrix},$$

$$\mathbf{W}_e = \mathrm{diag}(\kappa_{\mathrm{FT}} \mathbf{I}_{n_{\mathrm{FT}}}, \kappa_{\mathrm{WT}} \mathbf{I}_{n_{\mathrm{WT}}}, \kappa_{\mathrm{CFD}} \mathbf{I}_{n_{\mathrm{CFD}}}), \text{ and}$$

$$\mathbf{W}_x = \mathrm{diag}\left(0, \mathbf{W}_t(\boldsymbol{\lambda}_{\eta_0}), \mathbf{W}_t(\boldsymbol{\lambda}_{\eta_1}) \otimes \mathbf{R}_s^{-1}(\boldsymbol{\beta}), \mathbf{W}_t(\boldsymbol{\lambda}_{\delta_{\mathrm{FT}}}) \otimes \mathbf{I}_{n_{\mathrm{FT}}}, \right.$$
$$\left. \mathbf{W}_t(\boldsymbol{\lambda}_{\delta_{\mathrm{WT}}}) \otimes \mathbf{R}_{s_{\mathrm{WT}}}^{-1}(\boldsymbol{\beta}), \mathbf{W}_t(\boldsymbol{\lambda}_{\delta_{\mathrm{CFD}}}) \otimes \mathbf{R}_{s_{\mathrm{CFD}}}^{-1}(\boldsymbol{\beta}) \right).$$

This formulation leads to standard forms for the full conditional distributions for the parameters $\mathbf{x}$, κ, and $\boldsymbol{\lambda}$; that of $\boldsymbol{\beta}$ is not standard:

$$\begin{aligned}
\mathbf{x}|\cdots &\sim N(\boldsymbol{\mu}_x, \mathbf{V}_x) \\
&\quad \boldsymbol{\mu}_x = (\mathbf{W}_\epsilon + \mathbf{W}_x)^{-1} \mathbf{W}_\epsilon \mathbf{K}^T \mathbf{y} \\
&\quad \mathbf{V}_x = (\mathbf{W}_\epsilon + \mathbf{W}_x)^{-1} \\
\kappa_\ell|\cdots &\sim \Gamma(a_{e_\ell} + \tfrac{1}{2} 30 n_\ell, b_{e_\ell} + \tfrac{1}{2}(\mathbf{y}_\ell - \hat{\mathbf{y}}_\ell)^T (\mathbf{y}_\ell - \hat{\mathbf{y}}_\ell)), \ \ell = \mathrm{FT, WT, CFD} \\
\lambda_{\ell j}|\cdots &\sim \Gamma(a_{\ell j} + \tfrac{1}{2} n_\ell, b_{\ell j} + \tfrac{1}{2} \mathbf{d}_{\ell j}^T \mathbf{R}_{s_\ell}^{-1}(\boldsymbol{\beta}) \mathbf{d}_{\ell j}), \ \ell = \mathrm{FT, WT, CFD}, \ k = j, \ldots, p_\eta \\
\pi(\boldsymbol{\beta}|\cdots) &\propto |\mathbf{W}_x|^{\frac{1}{2}} \exp\{-\frac{1}{2} \mathbf{x}^T \mathbf{W}_x \mathbf{x}\} \times \pi(\boldsymbol{\beta}),
\end{aligned}$$

where $\hat{\mathbf{y}} = \mathbf{K}\mathbf{x}$. The full conditional for $\lambda_{\ell j}$ involves the vector $\mathbf{d}_{\ell j}$ which is constructed from the current values of the $n_\ell p_\eta$-vector $\mathbf{x}_{\delta_\ell}$ as follows. Construct the $n_\ell \times p_\eta$ matrix $\mathbf{d}_\ell$ from first differences of $\mathbf{x}_{\delta_\ell}$:

$$\mathbf{d}_\ell = \begin{pmatrix} x_{11} & x_{12} - x_{11} & \cdots & x_{1p\eta} - x_{1(p_\eta - 1)} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n_\ell 1} & x_{n_\ell 2} - x_{n_\ell 1} & \cdots & x_{n_\ell p\eta} - x_{n_\ell (p_\eta - 1)} \end{pmatrix}$$

The n_ℓ-vector $\mathbf{d}_{\ell j}$ is defined to be the jth column of the matrix $\mathbf{d}_\ell$.

These simple forms for the full conditionals make the posterior distribution amenable to sampling via Markov chain Monte Carlo (Gilks et al., 1996) or maximization with an E-M algorithm for a restricted maximization (Dempster et al., 1984). In either case, dealing with this high-dimensional parameter space and dataset can be handled with relative ease, whereas a traditional GP approach would be substantially more difficult.

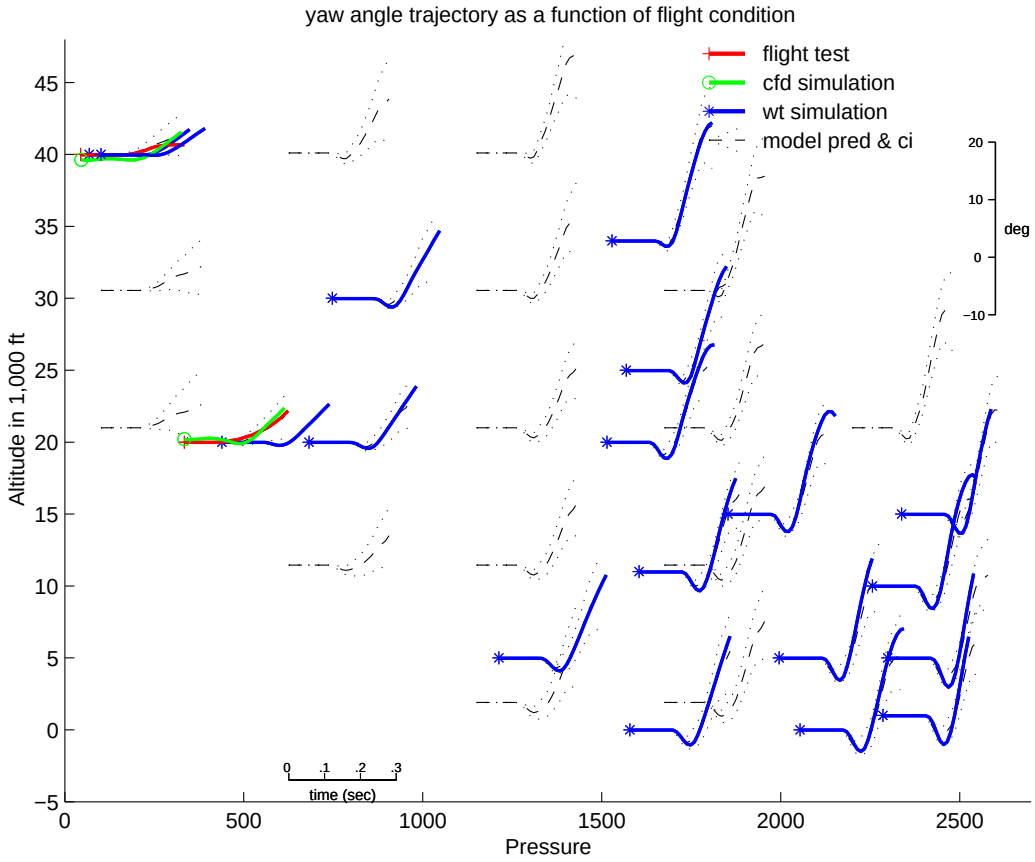

Figure 6: Observed and simulated trajectories along with the predicted trajectories (with a pointwise 2sd envelope) as a function of flight condition. The beginning of each trajectory marks the flight conditions which corresponds to it. The large vertical and horizontal axes give the dynamic pressure and altitude corresponding to the flight condition. The figure shows only launches for which the load was at 1g. The angle of attack varies from 26 degrees for trajectories in the upper left corner, to 0 degrees for most all of the trajectories near the lower right corner. Here the yaw-angle trajectory is shown. The small axes within the plot correspond to the yaw angle by time trajectories.

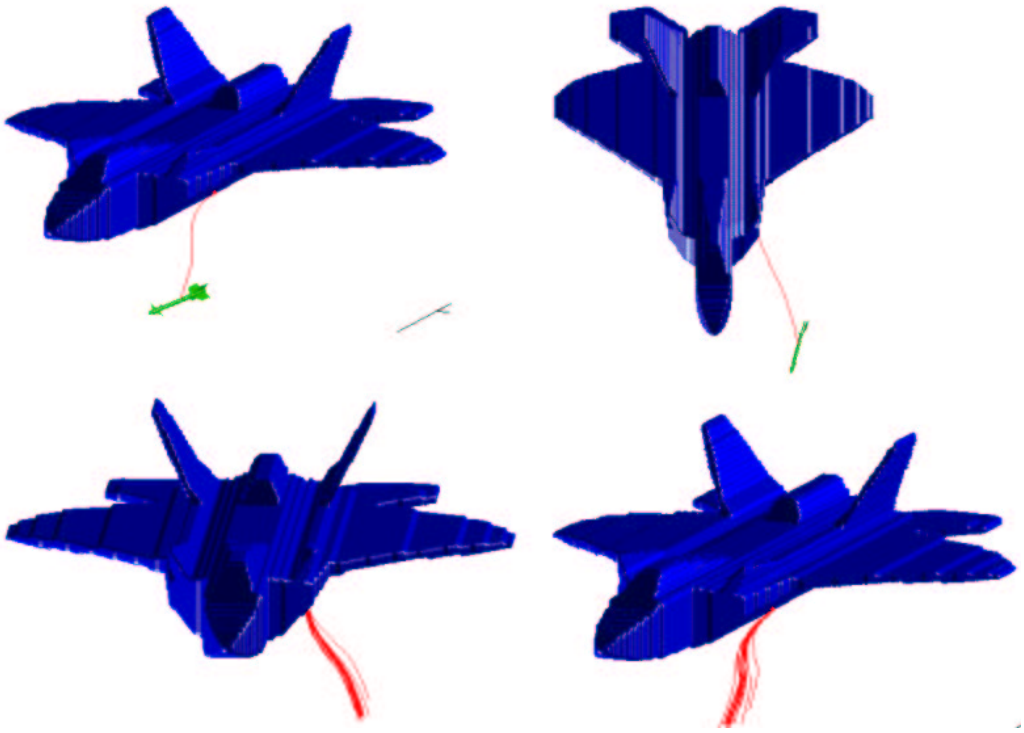

Figure 7: Simulated missile trajectories from the posterior distribution. The lines show center of mass trajectories. The orientation of the missile at one point along the trajectory is shown in the top figures.

4.6 Results

The predictive distribution for $\mathbf{x}$ can be obtained using straightforward Gaussian process theory (Cressie, 1993). Given the predictive distribution for $\mathbf{x}$, the resulting trajectory distribution is determined by the convolution/basis representation. The resulting uncertainty in the predicted outcome of a flight test at untried flight conditions is considerable. Due to the limited number of flight tests and CFD runs available, the data are not greatly informative about the discrepancy terms and the variability in flight tests. Hence the predictions rely a fair bit on the prior distribution specified for these quantities. As more flight tests become available, this reliance on the prior will be overcome by information from the new data.

Figure 6 shows the predicted trajectories for the yaw angle and their uncertainty for various flight conditions. Figure 7 shows posterior realizations of missile trajectories for a flight condition of $\mathbf{s} = (500, 20000, 1, 15)$. The posterior distribution for the trajectories at this configuration clearly indicates that there is no appreciable danger from a launch at this flight condition. Of course this analysis is based on simulations which have limited failure modes built into them. The analysis cannot account for failure modes omitted from the simulations without substantial flight test data.

5 Discussion

This paper introduced a convolution model which uses a fixed smoothing kernel along with a parameter λ_x to scale an underlying MRF $x(\mathbf{s})$. This leads to a number of computational advantages over a traditional GP model, while still capturing much of its spatial dependence structure. The convolution of an MRF formulation, with its simplified dependence structure and MCMC efficiency, can be used to extract spatial information in low signal to noise problems for which ordinary GP implementations have great difficulty. The flow application of this paper is one such example.

This convolution can extend its spatial dependence by increasing the value of the precision parameter λ_x. As the spatial distance becomes large, the variogram of the spatial process reverts to the variogram of $x(\mathbf{s})$. Though increasing the precision λ_x of the MRF extends the spatial dependence, how small the range of spatial dependence can be made is limited by the smoothing kernel $k(\mathbf{s})$. Hence, if $k(\mathbf{s})$ is chosen too large, the spatial dependence will have too large a range and the MRF cannot correct for this.

In this paper we have focused on a Gaussian MRF for modeling $x(\mathbf{s})$, our intrinsic underlying process. This framework clearly generalizes and fits into the context of convolutions of non-white processes, as discussed in Section 2. The convolution approach brings a number of computational and implementational advantages, as we have illustrated above and as discussed in Higdon (2002).

Acknowledgments

This work was partially supported by National Science Foundation grants DMS 9873275 and DMS 0233710.

References

Banerjee, S., Carlin, B. P., and Gelfand, A. E. (2003). *Hierarchical Modeling and Analysis for Spatial Data*. Boca Raton, FL: Chapman & Hall.

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.

Besag, J., York, J., and Mollié, A. (1991). "Bayesian image restoration, with two applications in spatial statistics (with discussion)." *Annals of the Institute of Statistical Mathematics*, 43, 1–59.

Brownie, C., Bowman, D. T., and Burton, J. W. (1994). "Estimating spatial variation in analysis of data from yield trials: A comparison of methods." *Agronomy Journal*, 85, 1244–1253.

Calder, C. A. (2003). "Exploring Latent Structure in Spatial Temporal Processes Using Process Convolutions." Ph.D. thesis, Duke University, Institute of Statistics and Decision Sciences.

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.

Cleveland, W. S. (1979). "Robust Locally-Weighted Regression and Smoothing Scatterplots." *Journal of the American Statistical Association*, 74, 829–836.

Cressie, N. A. C. (1993). *Statistics for Spatial Data*. revised ed. Somerset, NJ: Wiley-Interscience.

Currin, C., Mitchell, T., Morris, M., and Ylvisaker, D. (1991). "Bayesian Prediction of Deterministic Functions, With Applicatins to the Design and Analysis of Computer Experiments." *Journal of the American Statistical Association*, 86, 953–963.

Dempster, A. P., Selwyn, M. R., Patel, C. M., and Roth, A. J. (1984). "Statistical and computational aspects of mixed model analysis." *Applied Statistics*, 33, 203–214.

Diggle, P. J., Tawn, J. A., and Moyeed, R. A. (1998). "Model-based geostatistics (with discussion)." *Applied Statistics*, 47, 299–326.

Fuentes, M. and Smith, R. L. (2001). "A new class of nonstationary spatial models." Tech. Rep. 2534, North Carolina State University, Department of Statistics.

Gilks, W. R., Richardson, S., and Spiegelhalter, D. J. (1996). *Markov Chain Monte Carlo in Practice*. London: Chapman and Hall.

Higdon, D. (2002). "Space and Space-time Modeling Using Process Convolutions." In *Quantitative Methods for Current Environmental Issues*, eds. C. Anderson, V. Barnett, P. C. Chatwin, and A. H. El-Shaarawi, 37–56. London: Springer-Verlag.

Higdon, D., Lee, H., and Holloman, C. (2003). "Markov chain Monte Carlo-based approaches for inference in computationally intensive inverse problems." In *Bayesian Statistics 7*, eds. J. M. Bernardo, M. J. Bayarri, J. O. Berger, A. P. Dawid, D. Heckerman, A. F. M. Smith, and M. West, 181–197. Oxford: Oxford University Press.

Host, G., Omre, H., and Switzer, P. (1995). "Spatial Interpolation Errors for Monitoring Data." *Journal of the American Statistical Association*, 90, 853–861.

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.

Isaaks, E. H. and Srivastava, R. M. (1989). *Applied geostatistics*. Oxford: 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.

Kennedy, M. C. and O'Hagan, A. (2001). "Bayesian Calibration of Computer Models." *Journal of the Royal Statistical Society, Series B*, 63, 425–464.

Kern, J. C. (2000). "Bayesian Process-Convolution Approaches to Specifying Spatial Dependence Structure." Ph.D. thesis, Duke University, Institute of Statistics and Decision Sciences.

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*, 44, 230–241.

O'Hagan, A. (1991). "Bayes-Hermite quadrature." *Journal of Statistical Planning and Inference*, 29, 145–260.

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.

Reese, C. S., Wilson, A. G., Hamada, M., Martz, H. F., and Ryan, K. J. (2004). "Integrated Analysis of Computer and Physical Experiments." *Technometrics*, 46, 153–164.

Royle, J. A., Berliner, L. M., Wikle, C. K., and Milliff, R. (1999). "A hierarchical spatial model for constructing wind fields from scatterometer data in the Labrador Sea." In *Case Studies in Bayesian Statistics*, vol. IV, 367–382. Springer-Verlag.

Sacks, J., Welch, W. J., Mitchell, T. J., and Wynn, H. P. (1989). "Design and Analysis of Computer Experiments." *Statistical Science*, 4, 409–435.

Sampson, P. D. and Guttorp, P. (1992). "Nonparametric estimation of nonstationary spatial covariance structure." *Journal of the American Statistical Association*, 87, 108–119.

Schmidt, A. M. and O'Hagan, A. (2003). "Bayesian inference for nonstationary spatial covariance structure via spatial deformations." *Journal of the Royal Statistical Society, Series B*, 65, 745–758.

Stein, M. L. (1999). *Interpolation of Spatial Data: Some Theory for Kriging*. New York: Springer-Verlag.

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.

Wikle, C. K., Berliner, L. M., and Cressie, N. (1998). "Hierarchical Bayesian Space-time Models." *Environmental and Ecological Statistics*, 5, 117–154.

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.