2 Space and Space-Time Modeling using Process Convolutions

Dave Higdon

Institute of Statistics and Decision Sciences, Duke University,
Durham NC 27708-0251, USA

Abstract. A continuous spatial model can be constructed by convolving a very simple, perhaps independent, process with a kernel or point spread function. This approach for constructing a spatial process offers a number of advantages over specification through a spatial covariogram. In particular, this *process convolution* specification leads to computational simplifications and easily extends beyond simple stationary models. This paper uses process convolution models to build space and space-time models that are flexible and able to accommodate large amounts of data. Data from environmental monitoring is considered.

2.1 Introduction

Modeling spatial data with Gaussian processes is the common thread of all geostatistical analyses. Some notable references in this area include Matheron (1963), Journel & Huijbregts (1978), Ripley (1981), Cressie (1991), Wackernagel (1995), and Stein (1999). A common approach is to model spatial dependence through the *covariogram* $c(\cdot)$, so that covariance between any two points depends only on the distance between them. Distance is typically Euclidean though other metrics are sometimes used.
An alternative, constructive, method for creating a Gaussian process over R^d is to take i.i.d. Gaussian random variables on a lattice in R^d and convolve them with an arbitrary kernel. Figure 2.1 shows a trivial example using a Gaussian kernel to convolve i.i.d. Gaussian noise. Successively increasing the density of the lattice by a factor of 2 in each dimension while reducing the variance of the variates by a factor of 2^d leads to a continuous Gaussian white

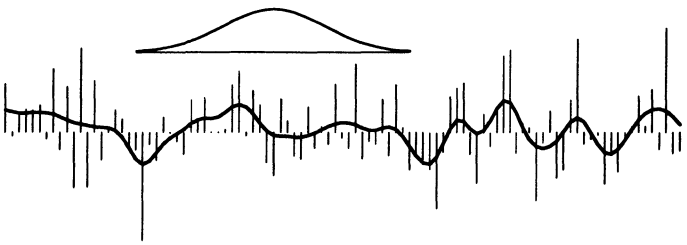

Fig. 2.1. A one-dimensional Gaussian process obtained from smoothed white noise.

C. W. Anderson et al. (eds.), *Quantitative Methods for Current Environmental Issues*

noise process over R^d. The convolution of this process can be equivalently defined using some covariogram in R^d. Though defining a process by the convolution of i.i.d. Gaussian lattice variables gives very similar results to defining a process by the covariogram, the convolution construction can be readily extended to allow for non-standard features such as non-stationarity, edge effects, dimension reduction, non-Gaussian fields, and alternative space-time models. This is the primary motivation for taking this approach.

2.2 Constructing Spatial Models via Moving Averages

One may construct a Gaussian process $z(s)$ over a general spatial (and temporal) region $\mathbf{S}$, such as the real plane, by convolving a continuous white noise process $x(s)$, $s \in \mathbf{S}$ with a smoothing kernel $k(s)$ so that

$$z(s) = \int_{\mathbf{S}} k(u-s)x(u)du, \quad \text{for } s \in \mathbf{S}. \tag{2.1}$$

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

$$c(d) = \mathrm{Cov}(z(s), z(s')) = \int_{\mathbf{S}} k(u-s)k(u-s')du = \int_{\mathbf{S}} k(u-d)k(u)du. \tag{2.2}$$

As a special case, if $\mathbf{S}$ is R^m 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 shown below,

$$\begin{array}{ccccccc} k(s) & \overset{FT}{\rightarrow} & K(\omega) & \overset{\cdot^2}{\rightarrow} & K^2(\omega) & \overset{IFT}{\rightarrow} & c(s) \\ k(s) & \overset{IFT}{\leftarrow} & C^{\frac{1}{2}}(\omega) & \overset{\sqrt{\cdot}}{\leftarrow} & C(\omega) & \overset{FT}{\leftarrow} & c(s) \end{array}$$

where FT and IFT denote the Fourier transform and its inverse, and the functions $\cdot^2$ and $\sqrt{\cdot}$ are applied pointwise. This gives the relationship between the spectrum $C(\omega)$ of a covariogram $c(s)$ and its resulting kernel $k(s)$: $C(\omega)$ is the square of the Fourier transform of $k(s)$. Note this relationship is no longer one to one if the process is not isotropic. In this case, multiple kernels can give rise to the same covariance function. The duality between the moving average process and a stationary Gaussian process determined by its variogram is explored in more detail in Thiebaux & Pedder (1985, ch. 5) and Barry & VerHoef (1996). Figure 2.2 shows kernels that give standard Gaussian, exponential and spherical covariograms for the process $z(s)$. In addition, the covariogram induced by the biweight kernel (Cleveland 1979) is also shown. Like the Gaussian kernel, it results in a covariogram that is fairly flat near the origin, and like the spherical covariogram, the dependence dies off completely after a fixed distance. This particular kernel is used in the ozone modeling examples later in this paper.

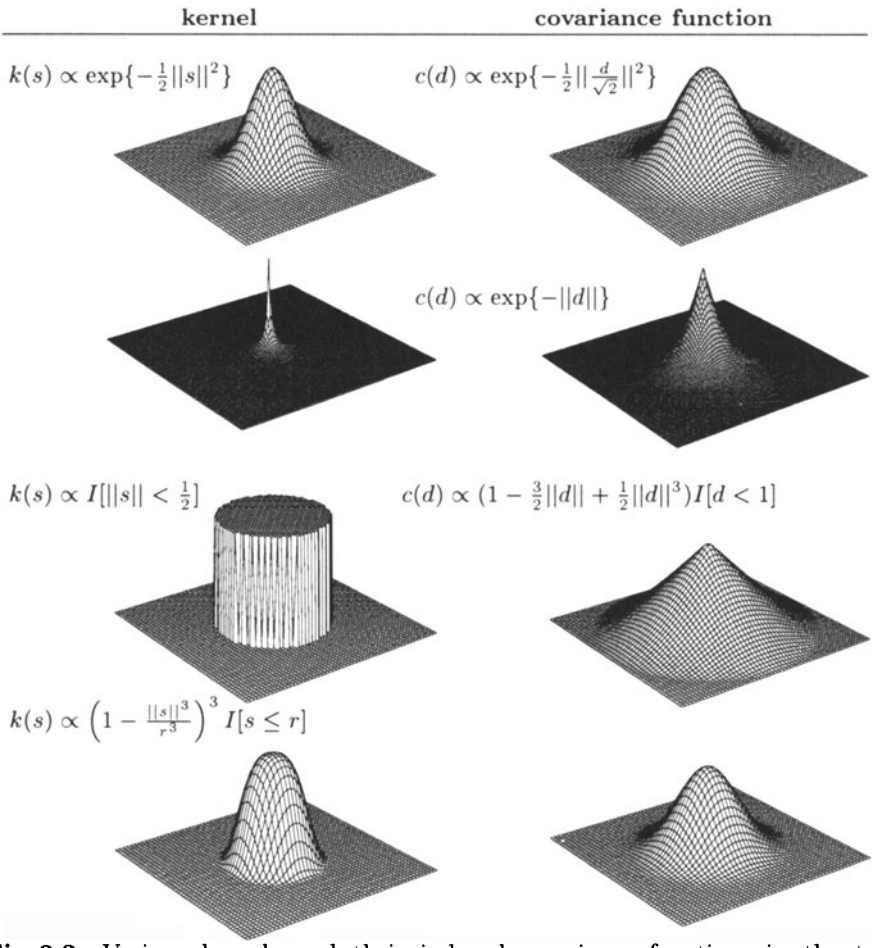

Fig. 2.2. Various kernels and their induced covariance functions in the two-dimensional plane.

2.2.1 Process Convolution Models

Under this moving average based approach, the model for the spatial process $z(s)$ is determined by the model specification for the latent process $x(s)$ and the smoothing kernel $k(s)$. One's choice of the latent process and smoothing kernel can lead to a number of interesting modeling approaches, a few of which are given below.

- **non-parametric covariance modeling**
  One appeal of the moving average representation is that one can model the smoothing kernel $k(s)$ rather than the covariogram $c(s)$ which must be positive definite. For example, Barry & Ver Hoef (1996), Kern (2000), and Ver Hoef *et al.* (2000) specify flexible models for $k(s)$ to build non-

standard covariances for Gaussian processes. Note that there are alternative approaches specifying a flexible class of positive definite covariance functions – see Gelfand & Ecker (1997) for example.

- **non-normal $x(s)$**
  A simple extension of the basic model is to modify the specification governing the latent process $x(s)$. For example Wolpert & Ickstadt (1999) specify $x(s)$ to be a Lévy process and Hartvig (2001) uses a Strauss process. Hence the resulting process $z(s)$ is not normal. A very simple example of this is to smooth out a lattice of i.i.d. Gamma(2, 1) random variables with a smoothing kernel as shown below. This essentially yields a stationary non-Gaussian field. Such a modeling approach may be appro-

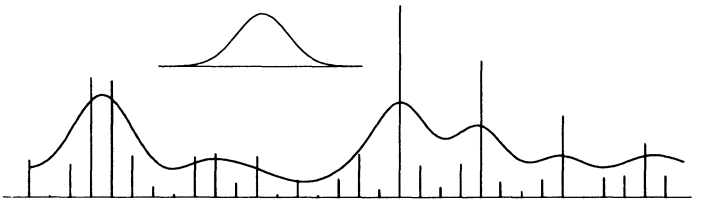

Fig. 2.3. One-dimensional smoothed gamma random field.

priate in modeling rates or concentrations which do not typically follow a normal distribution.

- **restricting the domain of $x(s)$**
  Instead of specifying $x(s)$ to be a continuous white noise process, one may choose to restrict the domain of $x(s)$. One reason may be to better account for dependencies in the system that is being modeled. For example, Kern (2000) uses such a modified process to account for irregularly shaped edges while modeling nitrogen concentrations in the Chesapeake Bay. This modification of the latent process at the edges could also be accompanied by a modification of the kernel in regions near the edges.
- **dimension reduction**
  One may choose to restrict the latent model $x(s)$ to locations $s_1, \ldots, s_m$ – over a coarse lattice for example. In this case, a small number of parameters $x(s_1), \ldots, x(s_m)$ effectively control the entire spatial process $z(s)$, even though $z(s)$ is continuous and may be required at thousands of locations. This is the main idea that I focus on for this paper. In addition to environmental monitoring problems, this approach has proven fruitful in inverse problems as well, where an economical parameterization of a spatial field can greatly facilitate computation.
- **nonstationary spatial covariance**
  The basic representation can be extended to allow the smoothing kernel

to vary with spatial location. In this case the spatial process is given by

$$z(s) = \int_{\mathbf{S}} k_s(u)x(u)du$$

where $k_s(u)$ denotes a kernel which is centered at site s and whose shape also depends on s. By allowing the family of smoothing kernels $k_s(u)$, $s \in \mathbf{S}$ to change slowly with spatial location, one can model a spatial process whose dependence structure can vary with s. Examples of using such non-stationary processes in environmental applications can be found in Higdon (1998), Higdon *et al.* (1999) and Hartvig (2001).

- **space-time models**
  By augmenting the general space $\mathbf{S}$ with time $\mathbf{T}$ so that the latent process and kernels are now defined over both space and time, $\mathbf{S} \times \mathbf{T}$, the basic formulation may be applied to space-time models. The latent process is defined over space and time $x(s,t)$ as is the smoothing kernel $k(s,t)$. See Higdon (1998) for an example. An intriguing alternative which is explored in this paper is to allow the latent process $x(s,t)$ to evolve over time. Along with a purely spatial kernel $k(s)$, a space time process is constructed by spatially smoothing the latent process.

$$z(s,t) = \int_{\mathbf{S}} k(u-s)x(u,t)du \tag{2.3}$$

  The final example of this paper combines this idea with dimension reduction so that the specified space time model can accommodate large amounts of data.

- **building dependent spatial processes**
  Finally, the process convolution approach gives an approach to build dependent spatial processes (see Ver Hoef & Barry 1998, and Ver Hoef *et al.* 2000 for example). The basic idea is to build processes $z_j(s)$ that share part of a common latent process in their construction. For example, two processes $z_1(s)$ and $z_2(s)$ could be constructed as:

$$z_1(s) = \int_{\mathbf{S}_0 \cup \mathbf{S}_1} k_1(u-s)x(u)du$$
$$z_2(s) = \int_{\mathbf{S}_0 \cup \mathbf{S}_2} k_2(u-s)x(u)du$$

  where the underlying latent process $x(s)$ resides on the union of the disjoint spaces $\mathbf{S} = \mathbf{S}_0 \cup \mathbf{S}_1 \cup \mathbf{S}_2$ and is independent on these separate subspaces. The dependence of $z_1(s)$ and $z_2(s)$ arises from their shared dependence upon $x(s)$ for $s \in \mathbf{S}_0$. A schematic of this approach is given in Figure 2.4 below.

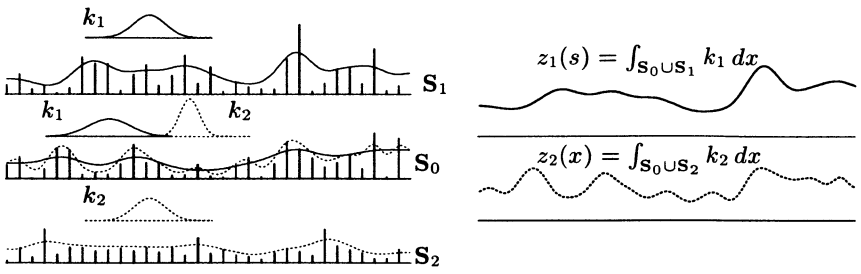

Fig. 2.4. A correlation is induced between processes $z_1(s)$ and $z_2(s)$ through common dependence upon $x(s)$ within $\mathbf{S}_0$. This spatial example is readily extendible to multiple fields over space and time.

2.3 Basic Spatial Model

The goal here is to present methodology for constructing spatial models that are flexible and sufficiently tractable so that inference can be carried out with fairly large datasets. Because of this, a sparse support set for the latent process $x(s)$ is required. Hence this paper will not consider models which specify $x(s)$ to be a continuous white noise process. For such an example see Higdon *et al.*(1999).

Here we develop the formulation of the basic model. Some modifications are considered afterwards. Let $y_1, \ldots, y_n$ be data recorded over the spatial locations $s_1, \ldots, s_n$ in $\mathbf{S}$. Perhaps the simplest spatial model represents the data as the sum of an overall mean μ, a spatial process $z = (z_1, \ldots, z_n)^T$, and Gaussian white noise $\epsilon = (\epsilon_1, \ldots, \epsilon_n)^T$ with variance σ_ϵ^2,

$$y = \mu + z + \epsilon$$

where the elements of z are the restriction of the spatial process $z(s)$ to the data locations $s_1, \ldots, s_n$.

We define $z(s)$ to be a mean zero Gaussian process. But rather than specify $z(s)$ through its covariance function, it is determined by the latent process $x(s)$ and the smoothing kernel $k(s)$. We restrict the latent process $x(s)$ to be nonzero at the spatial sites $\omega_1, \ldots, \omega_m$, also in $\mathbf{S}$ and define $x = (x_1, \ldots, x_m)^T$ where $x_j = x(\omega_j)$, $j = 1, \ldots, m$. Each x_j is then modeled as independent draws from a $N(0, \sigma_x^2)$ distribution. The resulting continuous Gaussian process is then

$$z(s) = \sum_{j=1}^{m} x_j k(s - \omega_j) \tag{2.4}$$

where $k(\cdot - \omega_j)$ is a kernel centered at ω_j. For the applications considered in this paper the smoothing kernel $k(\cdot)$ will be a radially symmetric kernel, such as a bivariate Gaussian density or any of the other kernels in Figure 2.2.

This gives the linear model

$$y = \mu \mathbf{1}_n + Kx + \epsilon \tag{2.5}$$

where $\mathbf{1}_n$ is the n-vector of 1's, the elements of K are given by

$$\begin{aligned} K_{ij} &= k(s_i - \omega_j)x_j, \\ x &\sim N(0, \sigma_x^2 I_m), \text{ and} \\ \epsilon &\sim N(0, \sigma_\epsilon^2 I_n). \end{aligned}$$

This is a basic mixed effects model. Inference can be carried out using a statistical package that uses likelihood based approaches for general mixed models such as SAS's `proc mixed` (Wolfinger *et al.* 1996) or `lme` in S-Plus (Pinheiro & Bates 2000). See the appendix for S-Plus 5.0 code for fitting the 1-d models here.

An alternative to straight likelihood based methods is a Bayesian approach. A fully Bayesian approach requires a prior specification for the remaining parameters μ, σ_x^2, and σ_ϵ^2. A 'default' formulation would give an improper uniform prior to μ ($\pi(\mu) \propto 1$) and rather flat gamma priors to σ_x^{-2} and σ_ϵ^{-2}. No closed form solution is available under this formulation, hence MCMC or some other approach for exploring the resulting posterior distribution will be required. If one fixes the ratio σ_x/σ_ϵ, the posterior distribution for x can be obtained in closed form (a multivariate t distribution). For applications that are data rich, it may be quite reasonable to estimate σ_x/σ_ϵ up front and then treat it as fixed.

2.3.1 Simple 1-d Example

Consider the data pairs (y_i, s_i), $i = 1, \ldots, n = 30$ shown in Figure 2.5. A process convolution model (2.5) is constructed by defining $k(s)$ to be a univariate Gaussian density with a standard deviation of 2, and defining the latent process support so that the ω_js are $m = 7$ equally spaced points ranging from -1 to 12 (as shown in the lower portion of Figure 2.5).

Note that this combination of kernel width and spacings of the ω_js yields a spatial process $z(s)$ via (2.4) that is nearly stationary. If the spacings become much larger, or if the kernel width is reduced, the covariance structure for $z(s)$ becomes unduly influenced by sparseness artifacts.

The resulting mixed effects model is fit using REML (Patterson & Thompson 1971) and the fitted values (best linear unbiased predictors) are given by the solid line in the top part of the figure. The bottom part of the figure shows the locations of the ω_js (black dots) and also shows the estimated values for each x_j. From (2.5) the fitted values can be represented as

$$\hat{y} = \hat{\mu} \mathbf{1}_n + \sum_{j=1}^{m} K^j \hat{x}_j$$

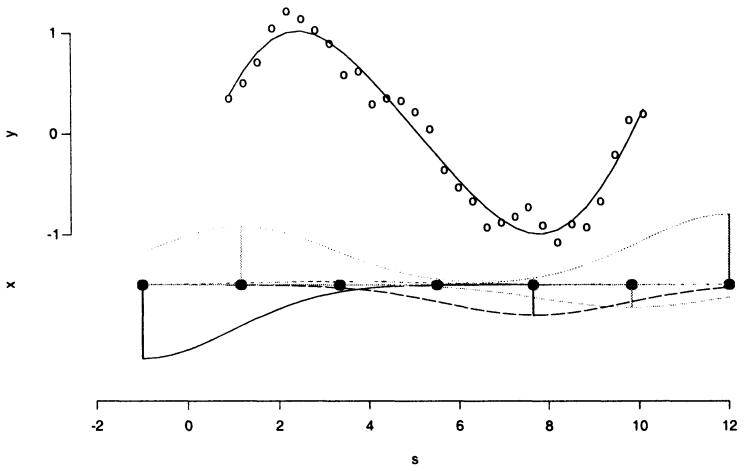

Fig. 2.5. A simulated spatial dataset with the resulting fitted values using a mixed model fit with REML (top figure). The grid locations $\omega_1, \ldots, \omega_m$ are marked by the solid dots below. The estimated values for the x_js are shown by the vertical segments at each ω_j (reduced by a factor of 3). The kernels centered at each ω_j show the 'bases' $k(s - \omega_j)x_j$ that sum to the fitted surface.

where K^j is the j^{th} column of the matrix K of equation (2.5). The lines in the bottom part of Figure 2.5 show the terms $K^j x_j$, $j = 1, \ldots, 7$.

2.3.2 A 2-d Example

As a more serious example we consider the maximum 8 hour averaged ozone concentration over the eastern United States on a late spring day in 1999 (Figure 2.6). After looking at some empirical variograms of the data, I settled on specifying $k(s)$ to be an isotropic two dimensional tricube kernel with a range of 9 degrees. Hence the induced covariogram dies off at about 15 degrees. More detailed approaches for choosing the smoothing kernel can be found in Barry & Ver Hoef (1996), Ver Hoef & Barry (1998), and Ver Hoef *et al.* (2000). The selected ω_js are shown by the black dots in the right hand frame of Figure 2.6. They are arranged on an isotropic hexagonal grid so that each interior ω_j is equidistant to three other $\omega_{j'}$s. Again the fitted surface was obtained via REML. Note that when the support of $x(s)$ is rather sparsely spaced as it is here, the straight least squares fit is nearly identical to the REML fit. The second order properties of these fitted values have not yet been compared.

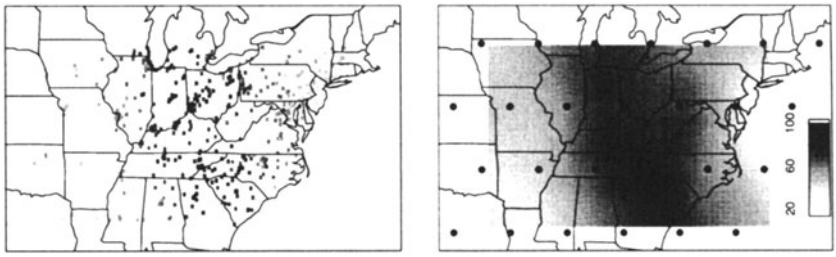

Fig. 2.6. Measured ozone concentrations and fitted concentration surface. Left: Daily maximum of the eight hour running average ozone concentration for a single day. Right: Fitted concentration surface; the dots represent the support points $\omega_1, \ldots, \omega_{27}$ of the latent process $x(s)$.

2.4 A Multiresolution Model

One can formulate a multiresolution model by representing the spatial process $z(s)$ as a sum of component processes

$$z(s) = \sum_{\ell=1}^{p} z_\ell(s)$$

where each $z_\ell(s)$ is represented through its own process convolution model determined by the pair $\{x_\ell(s), k_\ell(s)\}$, $\ell = 1, \ldots, p$. Each latent process $x_\ell(s)$ has support over spatial sites $\omega_{\ell 1}, \ldots, \omega_{\ell m_\ell}$ and we use $x_{\ell j}$ as shorthand for $x(\omega_{\ell j})$. Hence the separate processes are given by

$$z_\ell(s) = \sum_{j=1}^{m_\ell} k_\ell(s - \omega_{\ell j}) x_{\ell j}.$$

The multiresolution is captured in the kernels which become narrower as ℓ increases. Hence each additional $z_\ell(s)$ accounts for additional small scale detail.
As with the basic formulation, the model can be represented as a mixed model

$$y = \mu \mathbf{1} + \sum_{\ell=1}^{p} K^\ell x_\ell + \epsilon$$

where x_ℓ is the m_ℓ vector $(x_{\ell 1}, \ldots, x_{\ell m_\ell})^T$, the elements of K^ℓ are given by

$$\begin{aligned} K^\ell_{ij} &= k_\ell(s_i - \omega_{\ell j}) x_{\ell j}, \\ x_\ell &\sim N(0, \sigma^2_{x_\ell} I_{m_\ell}), \text{ and} \\ \epsilon &\sim N(0, \sigma^2_\epsilon I_n). \end{aligned}$$

2.4.1 Simple 1-d Example

As an example, Figure 2.7 shows the $n = 30$ data points from the previous 1-d example. The spatial locations s_i vary between 1 and 10 and the actual data points were created according to the formula

$$y(s_i) = \sin(2\pi[s_i/10]) + 0.2\sin(2\pi[s_i/2.5]) + e_i$$

where the e_is are i.i.d. $N(0, .1^2)$. So the first term gives large scale variation while the second term gives a fifth of the variation at 4 times the frequency. The multiresolution model is constructed as follows:

scale (ℓ)	$k_\ell(s)$	support of $x_\ell(s)$
coarse (1)	normal; sd=2	7 sites equally spaced between -1 and 12
medium (2)	normal; sd=1	14 sites equally spaced between -1 and 12
fine (3)	normal; sd=$\frac{1}{2}$	28 sites equally spaced between -1 and 12

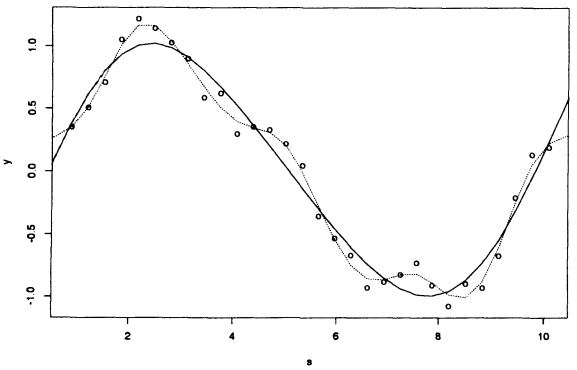

Fig. 2.7. Simulated spatial dataset with the resulting fitted values under the basic model (solid line) and the multiresolution model (dotted line). Here the multiresolution model finds the small scale structure in the data.

The fitted process is shown with the fit from the basic coarse model. Note that the multiscale formulation picks up the signal from the high frequency sine term. The estimated variance components under the two models are given below. Note that the multiresolution model picks up the high frequency sin() term at the finest resolution ($\ell = 3$) for which $k_3(s)$ nearly matches a half cycle of the sine wave.

model output for 1-d dataset

model	$\hat{\mu}$	$\hat{\sigma}_y$	$\hat{\sigma}_{x_1}$	$\hat{\sigma}_{x_2}$	$\hat{\sigma}_{x_3}$	reml logLik
basic	-0.01	0.15	8.25	–	–	0.38
multires	0.11	0.08	4.01	0.00	0.25	9.74

2.4.2 A 2-d Example

Similarly, this multiresolution model can be applied to the ozone example of the previous section. Here the multiresolution model is specified to have $p = 2$ levels. The first, coarse component ($\ell = 1$) is exactly the specification from the original ozone model so that the pair $\{x_1(s), k_1(s)\}$ is given in Section 2.3.2. The fine component refines the coarse specification by a factor of two in both spatial coordinates. Hence the process $x_2(s)$ is restricted to a denser set of points $\omega_{2,1}, \ldots, \omega_{2,87}$ shown in the bottom left frame of Figure 2.8. Along with this, the kernel $k_2(s)$ has half the scale of $k_1(s)$ so that the kernel vanishes 4.5 degrees from its center.

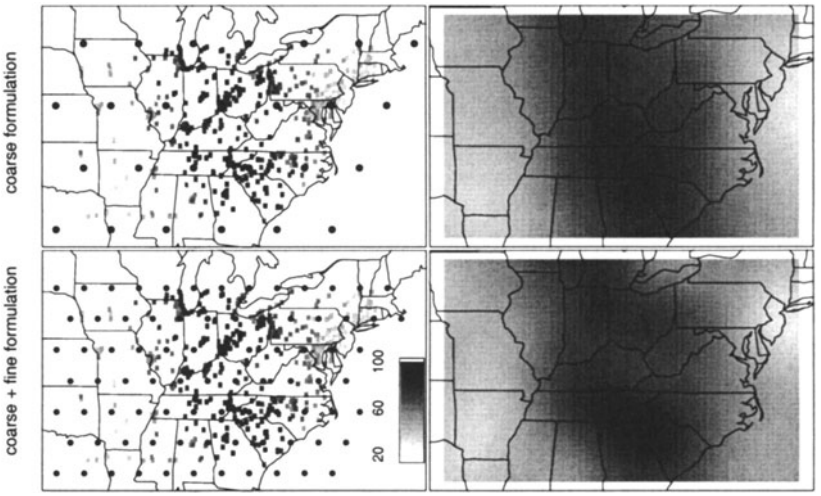

Fig. 2.8. Basic (coarse) and multiresolution models applied to the ozone data. Top left: data and location of ω_{1j}s for the latent process; Top right: fit under the basic formulation; Bottom left: data and location of the ω_{2j}s corresponding to the fine scale process – the coarse ω_{1j}s are shown in the top left figure; Bottom right: fit under the multiresolution formulation.

The data set is fit using REML and the results are summarized below. The predicted values are shown in the right hand frames of Figure 2.8. As with the 1-d example, the multiresolution fit differs from that of the basic model in the finer scale details. Such a decomposition of the ozone field is attractive since it is expected that the smaller scale fluctuations are fairly unpredictable. However the larger scale variability in the ozone concentrations does show some persistence from day to day.

model output for ozone dataset

model	$\hat{\mu}$	$\hat{\sigma}_y$	$\hat{\sigma}_{x_1}$	$\hat{\sigma}_{x_2}$	reml logLik
basic	54.33	8.04	80.35	–	−1742
multires	55.10	7.37	55.88	24.12	−1721

2.5 Building Space-Time Models

Perhaps the biggest attraction to these process convolution models is that they give a framework for developing new classes of space and space-time models that allow for more realistic space-time dependence while maintaining some analytic tractability. Generally, one can construct a space-time process by first defining a simple, possibly discrete, process over space and time, and then smoothing it out with one or more kernels, giving a smooth process over space and time.

This constructive approach is appealing since the resulting models can be extended to allow for generalizations such as non-stationarity, non-Gaussian models, and non-separable space-time dependence structures. See Wolpert & Ickstadt (1998), Ickstadt & Wolpert (1999), and Higdon *et al.* (1999) for some purely spatial applications, and Higdon (1998) for a space-time model. In addition, models can be constructed in such a way to facilitate computation – such as restricting the underlying process to reside on a lattice so that fast Fourier transforms can be employed.

For example, a simple but non-trivial example of a non-separable, non-Gaussian space-time process can be constructed as follows:

1. distribute random variables over space s and time t according to some marked point process.
2. smooth out this process according to some kernel defined over space s and time t.

This scheme for generating a realization from such a process is shown in Figure 2.9. In this particular example, i.i.d. exponential random variables are associated with each location, resulting in a strictly positive, non-Gaussian field. Note that space is only one-dimensional here to make the figure easy to understand – a 2- or 3-dimensional spatial component could also be considered.

This fairly simple example might be satisfactory for a pollutant concentration under a constant, prevailing wind. However, it may not very well account for a changing wind pattern. In such cases it might be preferable to have the underlying point process evolve over time and smooth only over the spatial component. Figure 2.10 shows a latent process defined over a spatial lattice that is slowly changing in magnitude and spatial location over time. Such a process is appealing since the spatial 'shifting' of this latent process could be linked to meteorological data for the region.

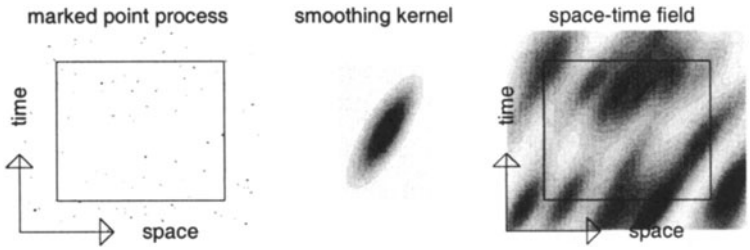

Fig. 2.9. Smoothing a marked point process defined over space and time yields a space time surface.

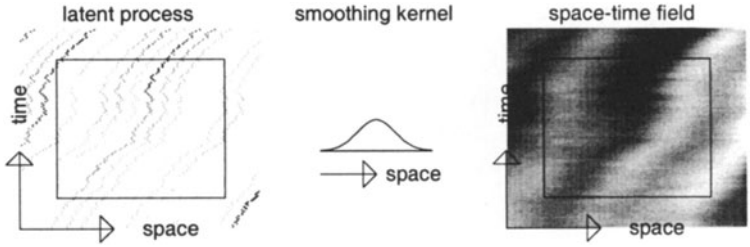

Fig. 2.10. A space time process constructed by spatially smoothing a process $x(s,t)$ whose values and support are both evolving over time.

2.5.1 A Space-Time Model for Ozone Concentrations

Using the ideas above, a space-time model for 30 consecutive days of ozone concentrations is constructed. We let $t \in \mathbf{T} = \{1, \ldots, 30\}$ index time (in days) and define the spatial support for the latent process over $\mathbf{W} = \{\omega_1, \ldots, \omega_{27}\}$ exactly as in the basic model of Section 2.3.2 so that $x(s,t)$ is nonzero over the set $\mathbf{W} \times \mathbf{T}$. Rather than define the random variables $x_{jt} = x(\omega_j, t)$ to be i.i.d., each sequence $\{x_{jt}\}$, $t = 1, \ldots, 30$ is specified to follow a Gaussian random walk. By accounting for the temporal dependence within the latent process $x(s,t)$, the induced space-time process is obtained by smoothing out $x(s,t)$ spatially

$$\begin{aligned} z(s,t) &= \int_{\mathbf{S}} k(u-s)x(u,t)du \\ &= \sum_j k(\omega_j - s)x_{jt} \end{aligned}$$

The data consist of around 500 measurements recorded each of the 30 days. The first 8 days of measurements are shown in Figure 2.11. With so many stations, each day there are a small number that cannot report data. Hence the actual number of datapoints recorded each time step n_t varies. Conditional

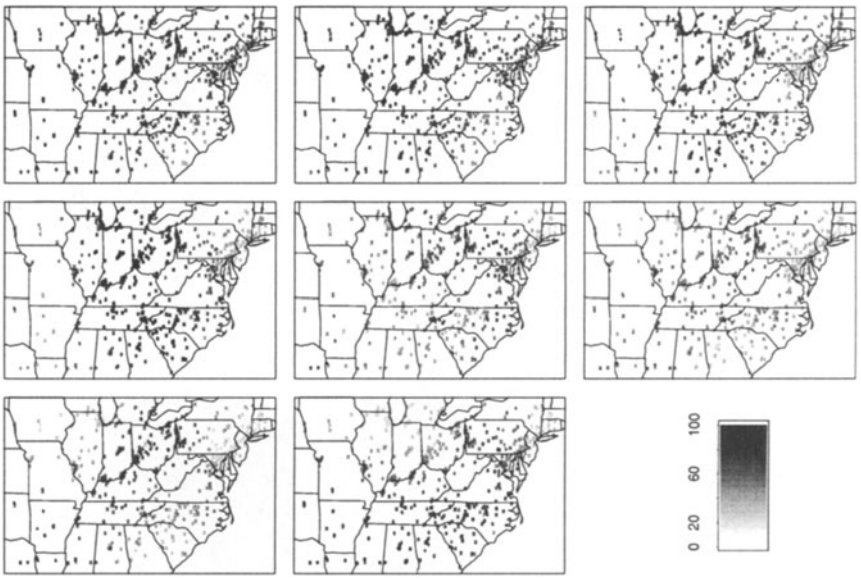

Fig. 2.11. Daily maximum for the eight hour running average of ozone concentration for eight consecutive days.

on the latent process values $x_t = (x_{1,t}, \ldots, x_{27,t})^T$, $t = 1, \ldots, 30$, a model for the data $y_t = (y_{1t}, \ldots, y_{n_t t})^T$ which is recorded at sites $s_{1t}, \ldots, s_{n_t t}$ can be expressed by the two evolution equations

$$y_t = K^t x_t + \epsilon_t \tag{2.6}$$

$$x_t = x_{t-1} + \nu_t \tag{2.7}$$

where K^t is the $n_t \times 27$ matrix given by

$$\begin{aligned} K^t_{ij} &= k(s_{it} - \omega_j) x_{jt}, \; t = 1, \ldots, 30, \\ \epsilon_t &\overset{\text{i.i.d.}}{\sim} N(0, \sigma^2_\epsilon), \; t = 1, \ldots, 30, \\ \nu_t &\overset{\text{i.i.d.}}{\sim} N(0, \sigma^2_\nu), \; t = 1, \ldots, 30, \text{ and} \\ x_1 &\sim N(0, \sigma^2_x I_{27}). \end{aligned} \tag{2.8}$$

This model is readily amenable to the dynamic linear model (DLM) machinery of West & Harrison (1997). Other alternatives are a fully Bayesian analysis via MCMC or a REML based approach. See Stroud *et al.* (1999) for a very similar DLM based modeling approach.

For this example flat, but proper, Gamma priors were specified for the precisions $1/\sigma^2_\epsilon$, $1/\sigma^2_\nu$, and $1/\sigma^2_x$. Estimation was then carried out using MCMC.

The resulting posterior mean estimates for the fitted values are shown in Figure 2.12. In addition, Figure 2.13 shows the posterior means for $x(s,t)$ at three interior spatial support points as a function of time.

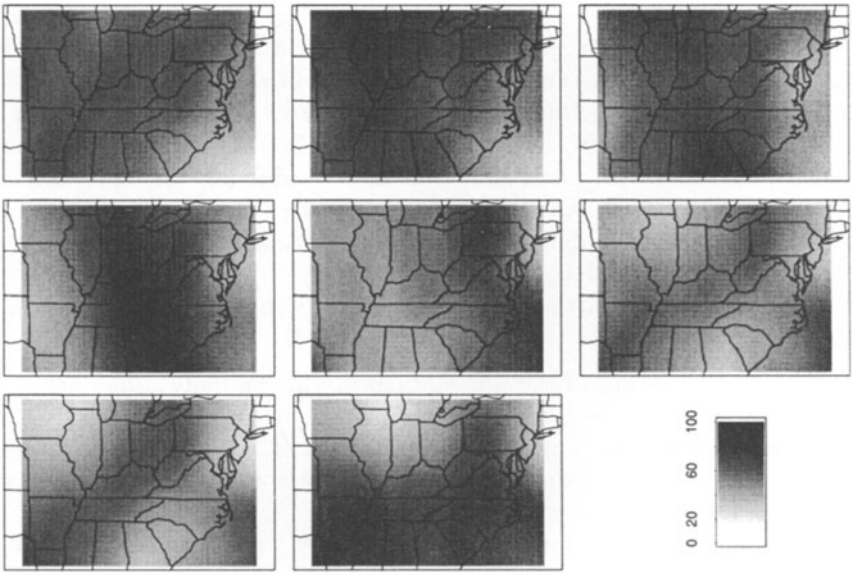

Fig. 2.12. Posterior mean surfaces of the 8 hour daily maximum ozone concentrations for eight consecutive days.

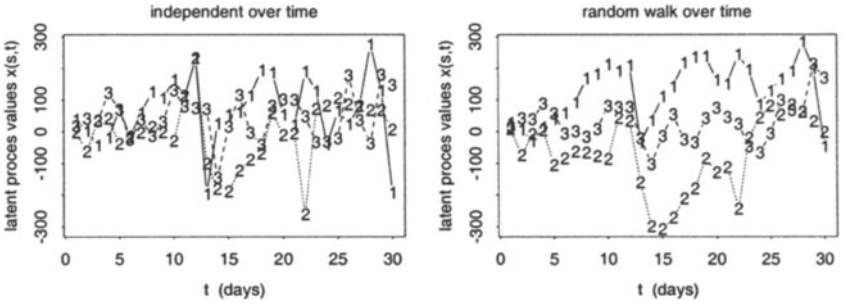

Fig. 2.13. Posterior mean estimates of the latent process $x(\omega_j, t)$, $t = 1, \ldots, 30$ at three interior locations ω_j. The left hand plot shows estimates under a model with i.i.d. x_{jt}; the right hand plot shows the estimates under the random walk formulation.

2.6 Discussion

This paper has focused on tools for building space and space-time models, rather than focusing on actual data analyses. The hope is to give readers an idea of how such models may be used in applications.

The approach is similar in spirit to using empirical orthogonal functions to reduce dimensionality in spatial fields (see von Storch & Zwiers 1999, for example), but the process convolution approach offers some advantages. Like wavelets, process convolutions allow local control over the spatial field. Interpolation is 'built-in' to process convolutions; it is less obvious how one should interpolate an EOF representation of the data. And finally, the approach can be understood through standard Gaussian process theory. Hence the vast literature regarding building Gaussian process models is also applicable to process convolution models.

Finally, I note that research in regard to modeling the ozone data is ongoing. Some issues of interest are:

- What is the nature of the non-stationarity in the daily ozone fields? If dependence does vary with spatial location, is the nature of this non-stationarity similar from day to day? Nychka *et al.* (1999) give a wavelet based approch for estimating non-stationarity from replicate observations over time.
- In the multiresolution model, how many resolution levels are appropriate? Presumably, the answer to this question depends on the actual inference problem, as well as the spatial process to be modeled. Also, the data are likely to inform only to a limited resolution. Any information about smaller scales will have to be obtained from some other source and built into the prior model. How can such fine scale information be obtained?
- It may be advantageous to combine the multiresolution model with the space-time model. It's possible that the coarse resolution component of the spatial model shows temporal dependence, while finer resolution components do not. How should one determine which resolutions should be incorporated into the temporal component of the model?
- By conditioning on a couple of estimated parameters, the posterior distribution can be obtained in closed form. Such a posterior form will be amenable to powerful utility based design methodology (Müller 1999). It will be of interest to understand how this 'conditional' posterior differs from the more appropriate marginal posterior.

We expect this line of research to play a role in furthering our understanding the nature and variability of ozone and other pollutants.

Appendix

Below is the code that produces the REML solutions for the one-dimensional examples. This was run using S-Plus 2000 for Windows.

```
# a fake dataset to make the bumps with
n_30    # of data points
m_7     # number of support sites for x(s)
# create sites s
s_seq(1,10,length=n)
# create the data y
e1_rnorm(n,sd=.1)
e2_cos(s/10*2*pi*4)*.2
y_sin(s/10*2*pi)+e2+e1
plot(s,y)

# locations of support points
w_seq(1-2,10+2,length=m)
# width of kernel
sdkern_2

# create the matrix K
K_matrix(NA,ncol=m,nrow=n)
for(i in 1:m){
  K[,i]_dnorm(s,mean=w[i],sd=sdkern)
}

# create a dataframe to hold the data
df1_data.frame(y=y,K=K,sub=1)
df1$sub_as.factor(df1$sub)

# now a fit a mixed model using lme
a1_lme(fixed= y ~ 1,
       random= pdIdent(~K-1),
       data=df1,na.action=na.omit)
# obtain and plot the fitted values
a1p_as.vector(predict(a1,df1))
lines(s,a1p,lty=1)

# now a multiscale version....
m1_c(7,14,28) # number of support points at each scale
w_list(NULL)  # list to hold the support locations
# make support locations w
for(i in 1:3) w[[i]]_seq(1-2,10+2,length=m1[i])
sdkern_c(2,1,.5) # kernel width by scale

# generate K matrices for each scale
K1_matrix(NA,ncol=m1[1],nrow=n)
for(i in 1:m1[1]) K1[,i]_dnorm(s,mean=w[[1]][i],sd=sdkern[1])
```

```
K2_matrix(NA,ncol=m1[2],nrow=n)
for(i in 1:m1[2]) K2[,i]_dnorm(s,mean=w[[2]][i],sd=sdkern[2])
K3_matrix(NA,ncol=m1[3],nrow=n)
for(i in 1:m1[3]) K3[,i]_dnorm(s,mean=w[[3]][i],sd=sdkern[3])

# create dataframe df2
df2_data.frame(y=y,K1=K1,K2=K2,K3=K3,sub=1)

# fit mixed effects model
a2_lme(fixed= y ~ 1,
       random= list(sub = pdIdent(~K1-1),sub=pdIdent(~K2-1),
                sub=pdIdent(~K3-1)),
       data=df2,na.action=na.omit)
# get predictions
a2p_as.vector(predict(a2,df2))
# plot it
plot(s,y)
lines(s,a1p,lty=1)
lines(s,a2p,lty=2)
```

References

Barry, R. P. & Ver Hoef, J. M. (1996). Blackbox kriging: spatial prediction without specifying variogram models. *Journal of Agricultural, Biological, and Environmental Statistics*, **1**, 297–322.

Cleveland, W. S. (1979). Robust locally weighted regression and smoothing scatterplots. *Journal of the American Statistical Association*, **74**, 829–836.

Cressie, N. A. C. (1991). *Statistics for Spatial Data*. Wiley-Interscience.

Gelfand, A. E. & Ecker, M. D. (1997). Bayesian variogram modeling for an isotropic spatial process. *Journal of Agricultural, Biological, and Environmental Statistics*, **2**, 347–369.

Guttorp, P., Meiring, W. & Sampson, P. D. (1994). A space-time analysis of ground-level ozone data. *Environmetrics*, **5**, 241–254.

Hartvig, N. V. (2001). A stochastic geometry model for fMRI data. *Scandinavian Journal of Statistics*, to appear.

Higdon, D. M. (1998). A process-convolution approach to modeling temperatures in the north Atlantic Ocean. *Journal of Environmental and Ecological Statistics*, **5**, 173–190.

Higdon, D. M., Swall, J. & Kern, J. C. (1999). Non-stationary spatial modeling. In *Bayesian Statistics 6. Proceedings of the Sixth Valencia International Meeting*, 761–768. Oxford University Press.

Kern, J. C. (2000). *Bayesian Process-Convolution Approaches to Specifying Spatial Dependence Structure,* unpublished PhD dissertation, Institute of Statistics & Decision Sciences, Duke University, Durham, USA.

Ickstadt, K. & Wolpert, R. L. (1999). Spatial regression for marked point processes. In *Bayesian Statistics 6. Proceedings of the Sixth Valencia International Meeting*, 323–341.

Littell, R. C., Milliken, G. A., Stroup, W. W. & Wolfinger, R. D. (1996). *SAS System for Mixed Models*. SAS Institute.

Matérn, B. (1986). *Spatial Variation (Second Edition)*. Springer-Verlag.

Meiring, W., Guttorp, P. & 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.

Müller, P. (1999). Simulation based optimal design. In *Bayesian Statistics 6. Proceedings of the Sixth Valencia International Meeting*, 323–341.

Nychka, D., Wikle, C. & Royle, J. A. (1999). Large spatial prediction problems and nonstationary random fields. Technical report, National Center for Atmospheric Research.

Patterson, H. D. & Thompson, R. (1971). Recovery of inter-block information when block sizes are unequal. *Biometrika*, **58**, 545–554.

Pinheiro, J. & Bates, D. (2000). *Mixed Effects models in S and S-plus*. New York: Springer.

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

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

Stroud, J., Müller, P. & Sanso, B. (1999). Dynamic models for spatio-temporal data. Technical Report 99-20, Institute of Statistics and Decision Sciences, Duke University.

Thiébaux, H. J. (1997). The power of the duality in spatial-temporal estimation. *Journal of Climatology*, **10**, 567–573.

Thiébaux, H. J. & Pedder, M. A. (1987). *Spatial Objective Analysis: with applications in atmospheric science*. San Diego: Academic Press.

Ver Hoef, J. & Barry, R. P. (1998). Constructing and fitting models for cokriging and multivariable spatial prediction. *Journal of Statistical Planning and Inference*, **69**, 275–294.

Ver Hoef, J., Cressie, N. & Barry, R. (2000). Flexible spatial models based on the fast Fourier transform (FFT) for cokriging. Technical report, Department of Statistics, The Ohio State University.

von Storch, H. & Zwiers, F. W. (1999). *Statistical Analysis in Climate Research*. New York: Cambridge University Press.

Wackernagel, H. (1995). *Multivariate Geostatistics. An Introduction With Applications*. Springer-Verlag.

West, M. & Harrison, J. (1997). *Bayesian Forecasting and Dynamic Models (Second Edition)*. New York: Springer-Verlag.

Wolpert, R. L. & Ickstadt, K. (1998). Poisson/gamma random field models for spatial statistics. *Biometrika*, **85**, 251–267.

Yaglom, A. M. (1987). *Correlation Theory of Stationary and Related Random Functions. Volume I: Basic Results*. Springer-Verlag, New York.

Yaglom, A. M. (1987). *Correlation Theory of Stationary and Related Random Functions. Volume II: Supplementary Notes and References*. Springer-Verlag, New York.