Convolution models for nonstationary spatial data

Mark D. Risser
The Ohio State University
risser.13@osu.edu

Catherine A. Calder
The Ohio State University
calder@stat.osu.edu

Accompaniment to `Convolutions_SimData.R`

1 A Bayesian model for continuously-indexed Gaussian spatial data

A simple Bayesian model for a univariate spatial Gaussian process can be defined as follows. The quantity of interest will be denoted $\{Z(\mathbf{s}), \mathbf{s} \in G\}$, $G \subset \mathcal{R}^d$ (d is the dimension of the spatial domain, here $d = 2$), which is the observed value of $Y(\mathbf{s})$, a latent spatial process. Furthermore, suppose we have observations which are a partial realization of this random process, taken at a fixed, finite set of n spatial locations $\{\mathbf{s_1}, ..., \mathbf{s_n}\} \in G$, giving the random (observed) vector $\mathbf{Z} = (Z(\mathbf{s_1}), ..., Z(\mathbf{s_n}))'$, which will be assumed to have a multivariate Gaussian distribution, conditional on the unobserved latent process. Specifically,

$$Z(\mathbf{s_i}) = \mathbf{x_i'}\boldsymbol{\beta} + Y(\mathbf{s_i}) + \epsilon(\mathbf{s_i})$$

where $E[Z(\mathbf{s_i})] = \mathbf{x_i'}\boldsymbol{\beta}$ ($\mathbf{x_i}$ is a generic vector of observable covariate information for location $\mathbf{s_i}$), the $\epsilon(\cdot)$ are iid $\mathcal{N}(0, \tau^2)$, and $Y(\cdot)|\boldsymbol{\Xi} \sim \mathcal{GP}(\mathbf{0}, \boldsymbol{\Omega}(\boldsymbol{\Xi}))$; $\epsilon(\cdot)$ and $Y(\cdot)$ are independent. The elements of $\boldsymbol{\Omega}(\boldsymbol{\Xi})$ are $\Omega_{ij} \equiv \text{Cov}(Y(\mathbf{s_i}), Y(\mathbf{s_j}))$, defined through a parametric covariance function

$$\text{Cov}(Y(\mathbf{s_i}), Y(\mathbf{s_j})) = C(\mathbf{s_i}, \mathbf{s_j}|\boldsymbol{\Xi})$$

which depends on the generic parameter vector $\boldsymbol{\Xi}$. The full parameter vector is $\boldsymbol{\theta} = (\boldsymbol{\beta}, \tau^2, \boldsymbol{\Xi})$; integrating out the process $\mathbf{Y}$, we obtain the marginal likelihood of the observed data, given the parameters:

$$\mathbf{Z}|\boldsymbol{\theta} \sim \mathcal{N}_n(\mathbf{X}\boldsymbol{\beta}, \tau^2\mathbf{I}_n + \boldsymbol{\Omega}).$$

2 A nonstationary Matérn covariance function via kernel convolution

A univariate spatial Gaussian process $Y(\mathbf{s})$ can be defined by way of kernel convolution; specifically, define

$$Y(\mathbf{s}) = \int_{\mathbb{R}^d} K_{\mathbf{s}}(\mathbf{u}) dW(\mathbf{u}),$$

where $W(\cdot)$ is a zero-mean white noise process with unit variance and $K_{\mathbf{s}}(\cdot)$ is a spatially-varying kernel function centered at $\mathbf{s} \in G$. The kernel functions are particularly important because they define the process covariance: for $\mathbf{s_i}, \mathbf{s_j} \in G$,

$$C(\mathbf{s_i}, \mathbf{s_j}) \equiv Cov\big(Y(\mathbf{s_i}), Y(\mathbf{s_j})\big) = \int_{\mathbb{R}^d} K_{\mathbf{s_i}}(\mathbf{u}) K_{\mathbf{s_j}}(\mathbf{u}) d\mathbf{u}.$$

As in the original paper by Higdon et al. (1998), if d-variate Gaussian densities are used for the kernel functions, a closed form can be obtained for $C(\mathbf{s_i}, \mathbf{s_j})$, namely

$$C(\mathbf{s_i}, \mathbf{s_j}) = (2\sqrt{\pi})^{-d} \left| \frac{\mathbf{\Sigma_i} + \mathbf{\Sigma_j}}{2} \right|^{-1/2} \exp\{-Q_{ij}\} = (2\sqrt{\pi})^{-d} \left| \frac{\mathbf{\Sigma_i} + \mathbf{\Sigma_j}}{2} \right|^{-1/2} \rho\left(\sqrt{Q_{ij}}\right), \quad (1)$$

where $\mathbf{\Sigma_i} \equiv \mathbf{\Sigma}(\mathbf{s_i})$ is the $(d \times d)$ covariance matrix for the Gaussian kernel function centered at location $\mathbf{s_i}$ (henceforth called the **kernel matrix**), $Q_{ij} = (\mathbf{s_i} - \mathbf{s_j})' \left(\frac{\mathbf{\Sigma_i} + \mathbf{\Sigma_j}}{2}\right)^{-1} (\mathbf{s_i} - \mathbf{s_j})$, and $\rho(\cdot)$ is the standard Gaussian correlation function. The kernel matrices $\mathbf{\Sigma}(\mathbf{s_i})$ can be interpreted as a locally varying geometric anisotropy matrix; i.e., $\mathbf{\Sigma}(\mathbf{s_i})$ controls the anisotropic behavior of the process in a small neighborhood of each point $\mathbf{s_i}$.

However, using a Gaussian correlation function as in (1) has the undesirable property of giving process realizations which are infinitely differentiable. Building off the ideas in Paciorek (2003), Stein (2005) proves that a generalization of (1) still gives a valid covariance function. Specifically,

$$C(\mathbf{s_i}, \mathbf{s_j}) = \sigma(\mathbf{s_i})\sigma(\mathbf{s_j}) \left| \frac{\mathbf{\Sigma_i} + \mathbf{\Sigma_j}}{2} \right|^{-1/2} \mathcal{M}_{\frac{\nu(\mathbf{s_i}) + \nu(\mathbf{s_j})}{2}} \left(\sqrt{Q_{ij}}\right) \quad (2)$$

is a valid covariance function, where $\sigma(\mathbf{s_i})$ is a locally varying "variance" parameter ($\sigma(\mathbf{s_i})$ is proportional to the actual process variance at $\mathbf{s_i}$), $\nu(\mathbf{s_i})$ is a locally varying smoothness parameter, and $\mathcal{M}_\nu(\cdot)$ is a Matérn correlation function with smoothness ν. The model in (2) is extremely flexible, as it allows the process variance, smoothness, and geometric anisotropy to vary over space; furthermore, using the Matérn class of correlation functions avoids the undesirable smoothness properties of (1).

The model to be used in the simulated dataset is a simplified version of (2) (due to Paciorek and Schervish (2006)) which focuses on the spatially varying geometric anisotropy of a spatial process. Specifically, we will set $\sigma(\mathbf{s_i}) \equiv \sigma|\mathbf{\Sigma_i}|^{1/4}$ and $\nu(\mathbf{s_i}) \equiv \nu$ for all $\mathbf{s_i} \in G$, and allow the $\mathbf{\Sigma}(\mathbf{s_i})$ to vary in space. Formally, the covariance function we will use is

$$C(\mathbf{s_i}, \mathbf{s_j}) = \sigma^2 |\mathbf{\Sigma_i}|^{1/4} |\mathbf{\Sigma_j}|^{1/4} \left| \frac{\mathbf{\Sigma_i} + \mathbf{\Sigma_j}}{2} \right|^{-1/2} \mathcal{M}_\nu \left(\sqrt{Q_{ij}}\right), \quad (3)$$

which is a valid covariance function on $\mathcal{R}^d$, $d \geq 1$. This choice of $\sigma(\mathbf{s_i})$ gives a constant process variance:

$$\text{Var}\big(Y(\mathbf{s_i})\big) \equiv C(\mathbf{s_i}, \mathbf{s_i}) = \sigma^2 |\mathbf{\Sigma_i}|^{1/4} |\mathbf{\Sigma_i}|^{1/4} \left| \frac{\mathbf{\Sigma_i} + \mathbf{\Sigma_i}}{2} \right|^{-1/2} \mathcal{M}_\nu(0) = \sigma^2.$$

Following the notation above, the parameter vector is now

$$\boldsymbol{\theta} = (\boldsymbol{\beta}, \tau^2, \sigma^2, \nu, \mathbf{\Sigma}(\cdot)).$$

3 Kernel parameterization, specific values used, and data dictionary

Given that the focus of the dataset is on the spatial dependence structure, the process mean will be fixed to zero; i.e., $\boldsymbol{\beta} \equiv \mu = 0$. The default values used for other parameters in the simulated data are $\tau^2 = 0.1$ (nugget effect), $\sigma^2 = 1$ (σ^2 is proportional to the process variance), and $\nu = 0.5$ (process smoothness; $\nu = 0.5$ implies an exponential covariance structure). The dataset contains $n = 676$ observations on a spatial region of $G = [0, 5] \times [0, 5]$; the gridded locations have a spacing of 0.2.

The kernel matrices are parameterized as follows. Each 2×2 kernel matrix $\mathbf{\Sigma_i}$ is

$$\mathbf{\Sigma_i} = \begin{bmatrix} \sigma^2_{1,i} & \sigma_{3,i} \\ \sigma_{3,i} & \sigma^2_{2,i} \end{bmatrix},$$

although we could equivalently use the representation

$$\mathbf{\Psi_i} = \begin{bmatrix} \log(\sigma^2_{1,i}) \\ \log(\sigma^2_{2,i}) \\ \tan(\frac{\pi}{2}\rho_i) \end{bmatrix} \equiv \begin{bmatrix} \psi^i_1 \\ \psi^i_2 \\ \psi^i_3 \end{bmatrix}$$

(where $\rho_i = \sigma_{3,i}/(\sigma_{1,i} \cdot \sigma_{2,i}) \in (-1, 1)$), so that each element of $\mathbf{\Psi_i}$ has support $\mathcal{R}$. Independent Gaussian processes can be defined for each of these three components across the n locations, with constant mean and exponential covariance structure. For example, define $\boldsymbol{\psi_1} = (\psi^1_1, \psi^2_1, ..., \psi^n_1)'$; then the values used for the dataset can be a draw from

$$\boldsymbol{\psi_1} \sim \mathcal{N}_n(m_1 \mathbf{1_n}, \mathbf{C_1}),$$

where $\mathbf{C_1}$ has ij element $v_1^2 \cdot \exp\left\{-\frac{||\mathbf{s_i} - \mathbf{s_j}||}{r_1}\right\}$, $i, j = 1, ..., n$. Thus, we require parameters for the GP of the kernel matrices, namely three mean parameters $\mathbf{m} = (m_1, m_2, m_3)'$, three variance parameters $\mathbf{v} = (v_1, v_2, v_3)'$, and three range parameters $\mathbf{r} = (r_1, r_2, r_3)'$. These values were chosen in an attempt to give something that resembles a reasonable kernel matrix process; the specific values are $\mathbf{m} = (-2.3, -1.9, 1.0)$, $\mathbf{v} = (1, 1, 1)$, and $\mathbf{r} = (4, 4, 4)$.

3.1 Data dictionary

Two datasets are included: one with data on a grid, and one with data spaced irregularly over G. Each dataset contains four variables: x, y, observed data, and a holdout indicator. Ten percent of the data has been selected as a hold-out sample, and this final variable indicates if the observation is to be held out (1) or not (0).

References

Higdon, D., Swall, J., and Kern, J. (1998). Non-stationary spatial modeling. *Bayesian Statistics*, 6.

Paciorek, C. J. (2003). *Nonstationary Gaussian processes for regression and spatial modelling*. PhD thesis, Carnegie Mellon University.

Paciorek, C. J. and Schervish, M. J. (2006). Spatial modeling using a new class of nonstationary covariance functions. *Environmetrics*, 17:483–506.

Stein, M. L. (2005). Nonstationary spatial covariance functions. *Unpublished technical report*.