Nonstationary Spatial Covariance Functions *

Michael L. Stein

ABSTRACT: A method is described for producing explicit nonstationary spatial covariance functions that, for example, allows both the local geometric anisotropy and the degree of differentiability to vary spatially.

Keywords: Spatial anisotropy, Spectral density, Matérn model, Long-range dependence

* Michael L. Stein is Professor, Department of Statistics, University of Chicago, Chicago, IL 60637 (email: stein@galton.uchicago.edu). Although the research described in this article has been funded wholly or in part by the United States Environmental Protection Agency through STAR cooperative agreement R-82940201-0 to the University of Chicago, it has not been subjected to the Agency's required peer and policy review and therefore does not necessarily reflect the views of the Agency and no official endorsement should be inferred.

1. Introduction

There has been a growing recognition of the need for nonstationary spatial covariance functions in a variety of disciplines. The atmospheric and environmental sciences are two obvious areas of application; for example, Holland, et al. (2003) review the use of nonstationary spatial covariance functions for the statistical analysis of air quality data. In machine learning, Gaussian process models provide a popular approach to nonparametric regression (see Seeger (2004) and the references therein), and specification of the covariance structure is obviously critical in this setting. Paciorek and Schervish (2004a) discuss the use of nonstationary models for Gaussian process regression. Since regression models can have any number of regressors, there is a need for models for valid nonstationary spatial covariance functions in arbitrary dimensions. For both ease of computation and interpretation, it is helpful to have explicit expressions for these covariance functions.

Paciorek (2003) (see also Paciorek and Schervish (2004a,b)) describe a method for producing explicit expressions for valid spatial covariance functions with locally varying geometric anisotropies. However, this approach does not allow one to vary other aspects of the covariance structure spatially, such as the differentiability of the process or the index of long range dependence. By using spatially varying spectra, Pintore and Holmes (2004) demonstrate that one can get explicit spatial covariance functions whose degree of differentiability varies in space. Section 2 shows how these two ideas can be combined to produce very flexible classes of explicit nonstationary spatial covariance functions.

Higdon (1998) and Fuentes (2002) describe methods for generating nonstationary covariance functions as integrals, but these integrals generally have to be carried out numerically. Indeed, the key insight in Paciorek (2003) is that there is a class of models for which the integrations required by Higdon (1998) can be carried out analytically. Nychka, Wikle and Royle (2002) describe a wavelet approach to producing nonstationary spatial covariance functions, but although the wavelet approach leads to fast computations, it produces covariance functions that are sums over a large number of terms. Another approach to producing explicit covariance functions with locally varying geometric anisotropies is to use spatial deformations (Sampson and Guttorp 1992, Perrin and Senoussi 2000, Sampson, Damian and Guttorp 2001, Schmidt and O'Hagan 2003, Clerc and Mallat 2003). As with Paciorek's (2003) method, spatial deformations could be combined with spatially varying spectra to obtain a broad class of explicit nonstationary covariance functions.

Section 3 briefly compares these approaches and discusses some potential areas of application and limitations of the models given in Section 2.

2. Main Result

The following theorem gives a representation for a broad class of valid covariance functions for random fields on $\mathbb{R}^p$. We will call a real function R on $\mathbb{R}^p \times \mathbb{R}^p$ a covariance function if $R(\mathbf{x}, \mathbf{x}) < \infty$ for all $\mathbf{x} \in \mathbb{R}^p$ and if it assigns nonnegative values to all quadratic forms: for all finite n, all $\mathbf{x}_1, \dots \mathbf{x}_n$ in $\mathbb{R}^p$ and all real scalars $a_1, \dots, a_n$, $\sum_{i,j=1}^n a_i a_j R(\mathbf{x}_i, \mathbf{x}_j) \geq 0$.

THEOREM 1. *Suppose* Σ *is a mapping from* $\mathbb{R}^p$ *to positive definite* $p \times p$ *matrices,* μ *is a nonnegative measure on* $[0, \infty)$*, and for each* $\mathbf{x} \in \mathbb{R}^p$, $g(\cdot; \mathbf{x}) \in L^2(\mu)$. *Defining* $\Sigma(\mathbf{x}, \mathbf{y}) = \frac{1}{2}\Sigma(\mathbf{x}) + \frac{1}{2}\Sigma(\mathbf{y})$ *and* $Q(\mathbf{x}, \mathbf{y}) = (\mathbf{x} - \mathbf{y})'\Sigma(\mathbf{x}, \mathbf{y})^{-1}(\mathbf{x} - \mathbf{y})$,

$$R(\mathbf{x}, \mathbf{y}) = \frac{|\Sigma(\mathbf{x})|^{1/4}|\Sigma(\mathbf{y})|^{1/4}}{|\Sigma(\mathbf{x}, \mathbf{y})|^{1/2}} \int_0^\infty e^{-wQ(\mathbf{x},\mathbf{y})} g(w; \mathbf{x}) g(w; \mathbf{y}) \mu(dw) \tag{1}$$

is a covariance function. □

Paciorek (2003) gives this result when g does not depend on $\mathbf{x}$. Note that if c is a real-valued function on $\mathbb{R}^p$, then if $R(\mathbf{x}, \mathbf{y})$ is a covariance function on $\mathbb{R}^p \times \mathbb{R}^p$, so is $c(\mathbf{x})c(\mathbf{y})R(\mathbf{x}, \mathbf{y})$.

PROOF OF THEOREM 1. The proof follows closely that in Paciorek (2003, pp. 26–27). First, note that $R(\mathbf{x}, \mathbf{x}) = \int_0^\infty g(w; \mathbf{x})^2 \mu(dw) < \infty$ for all $\mathbf{x}$ since $g(\cdot; \mathbf{x}) \in L^2(\mu)$. Next, define $\Sigma_i = \Sigma(\mathbf{x}_i)$, $\Sigma_{ij} = \Sigma(\mathbf{x}_i, \mathbf{x}_j)$ $Q_{ij} = Q(\mathbf{x}_i, \mathbf{x}_j)$ and let K_i^w be the density of a multivariate normal random vector with mean $\mathbf{x}_i$ and covariance matrix $(4w)^{-1}\Sigma_i$. Then, using a straightforward convolution argument (Paciorek 2003, p. 27) for the second equality,

$$\begin{aligned}
&\sum_{i,j=1}^n a_i a_j R(\mathbf{x}_i, \mathbf{x}_j) \\
&\quad = \sum_{i,j=1}^n a_i a_j \frac{|\Sigma_i|^{1/4}|\Sigma_j|^{1/4}}{|\Sigma_{ij}|^{1/2}} \int_0^\infty e^{-wQ_{ij}} g(w; \mathbf{x}_i) g(w; \mathbf{x}_j) \mu(dw) \\
&\quad = \pi^{p/2} \sum_{i,j=1}^n a_i a_j |\Sigma_i|^{1/4}|\Sigma_j|^{1/4} \int_0^\infty \left\{ \int_{\mathbb{R}^p} K_i^w(\mathbf{u}) K_j^w(\mathbf{u}) \mathbf{du} \right\} w^{-p/2} g(w; \mathbf{x}_i) g(w; \mathbf{x}_j) \mu(dw) \\
&\quad = \pi^{p/2} \int_0^\infty \int_{\mathbb{R}^p} \left\{ \sum_{i=1}^n a_i |\Sigma_i|^{1/4} K_i^w(\mathbf{u}) g(w; \mathbf{x}_i) \right\}^2 \mathbf{du}\, w^{-p/2} \mu(dw),
\end{aligned}$$

which is nonnegative, proving that R assigns nonnegative values to all quadratic forms. □

The Matérn class of covariance functions is widely used to model stationary and isotropic spatial processes (Stein, 1999) and, for interpoint distance d and positive parameters ϕ, ν and α, is given by $\phi\mathcal{M}_\nu(\alpha d) = \phi(\alpha d)^\nu \mathcal{K}_\nu(\alpha d)$, where $\mathcal{K}_\nu$ is a modified Bessel function of order ν. A process with this covariance function is m times mean square differentiable in any direction if and only if $\nu > m$. Take $\mu(dw) = w^{-1}e^{-1/(4w)}dw$, $g(w;\mathbf{x}) = w^{-\nu(\mathbf{x})/2}$ and assume $\nu(\mathbf{x}) > 0$ for all $\mathbf{x}$, so that g satisfies the conditions of Theorem 1. Then for Σ positive definite-valued and c real-valued, using (3.471.9) in Gradshteyn and Ryzhik and absorbing $|\Sigma(\mathbf{x})|^{1/4}$ into $c(\mathbf{x})$,

$$R(\mathbf{x},\mathbf{y}) = \frac{c(\mathbf{x})c(\mathbf{y})}{|\Sigma(\mathbf{x},\mathbf{y})|^{1/2}} \mathcal{M}_{\{\nu(\mathbf{x})+\nu(\mathbf{y})\}/2}\big(Q(\mathbf{x},\mathbf{y})^{1/2}\big) \tag{2}$$

is a covariance function on $\mathbb{R}^p \times \mathbb{R}^p$ for all p. Paciorek and Schervish (2004a,b) give the special case when ν is constant in $\mathbf{x}$ and Pintore and Holmes (2004) essentially give the special case when Σ is constant in $\mathbf{x}$. Equation (2) gives a single model that allows for spatially varying geometric anisotropies and ranges (through Σ) and spatially varying degrees of differentiability (through ν). Thus, in terms of the local behavior of the process, (2) is an extremely flexible model.

In terms of long-range behavior, (2) is less flexible because all Matérn covariance functions decay exponentially at large distances. Thus, for example, if c is bounded and ν and the eigenvalues of Σ are bounded away from 0 and ∞ in $\mathbf{x}$, then there exist positive constants a and b such that $R(\mathbf{x},\mathbf{y}) \le ae^{-b|\mathbf{x}-\mathbf{y}|}$ for all $\mathbf{x}$ and $\mathbf{y}$. To generate models whose covariance functions decay more slowly as $|\mathbf{x}-\mathbf{y}|$ increases, take $\mu(dw) = w^{-1}e^{-w}dw$ and $g(w;\mathbf{x}) = w^{\delta(\mathbf{x})/2}$, where $\delta(\mathbf{x}) > 0$ guarantees $g(\cdot;\mathbf{x}) \in L^2(\mu)$. For Σ positive definite-valued and c real-valued, we have

$$R(\mathbf{x},\mathbf{y}) = \frac{c(\mathbf{x})c(\mathbf{y})}{|\Sigma(\mathbf{x},\mathbf{y})|^{1/2}\{1+Q(\mathbf{x},\mathbf{y})\}^{\{\delta(\mathbf{x})+\delta(\mathbf{y})\}/2}} \tag{3}$$

is a covariance function on $\mathbb{R}^p \times \mathbb{R}^p$ for all p. Paciorek and Schervish (2004b) give (3) when δ is constant. In the stationary case (c, δ and Σ constant), the process is long-range dependent if and only if its autocovariance function is not absolutely integrable, which will hold here if and only if $\delta \le p/2$. It is not so obvious what one should mean by long-range dependence for nonstationary processes, although Heyde and Yang (1997) give a definition for processes on the integers that can be generalized to processes on $\mathbb{R}^p$. Certainly, if in (3), the eigenvalues of Σ were bounded away from 0 and ∞ and $\delta(\mathbf{x}) \le p/2$ for all $\mathbf{x}$, one would want to call the resulting process long-range dependent.

Models of the forms (2) and (3) can be summed to obtain models with spatially varying degrees of local and long-range behavior. Because models of the form (2) do not have long-range dependence, one can change the functions ν and Σ in (2) without affecting the long-range dependence of the process. Similarly, because all models of the form (3) correspond to processes that are infinitely mean-square differentiable, one can change the functions δ and Σ in (3) without affecting the differentiability of the process. However, it may be desirable to have a single model that allows variation in both the long-term and local behavior of the process. Denoting interpoint distance by d, Gneiting and Schlather (2004) note that the model $(1 + d^\alpha)^{-\beta}$ is a valid isotropic autocovariance function in any number of dimensions for $0 < \alpha \leq 2$ and $\beta > 0$, and that one can achieve different degrees of both local behavior and long-range dependence through the choice of α and β. However, the degrees of local behavior are more restricted than for the Matérn class in that only nondifferentiable ($\alpha < 2$) and infinitely differentiable ($\alpha = 2$) processes are attainable. Furthermore, since an explicit spectral representation for this class of covariance functions does not appear to be available for arbitrary α, it is not clear how one would apply Theorem 1 to this class, nor is it clear such an application would lead to explicit expressions for some class of nonstationary processes in which, in some sense, α and β could vary spatially.

3. Discussion

Although a sum of models of the forms (2) and (3) provides an extensive range of possible models for nonstationary covariance functions, it does have some limitations. Assuming Σ and g are continuous in $\mathbf{x}$, at least roughly, all models of the form given by Theorem 1 are locally geometrically anisotropic; that is, are locally isotropic in a neighborhood of $\mathbf{x}$ after some affine transformation of the coordinates depending on $\mathbf{x}$. Such a restriction excludes, for example, models with different degrees of differentiability in different directions, or models with local anisotropies that are other than geometric. Another limitation of models obtained through Theorem 1 is the ability to produce negative correlations. Indeed, if g does not depend on $\mathbf{x}$, then R in (1) is trivially nonnegative. When g does depend on $\mathbf{x}$, one can get $R(\mathbf{x}, \mathbf{y}) < 0$ by, for example, taking $g(w; \mathbf{x}) = -g(w; \mathbf{y})$. However, the fact that the ability to obtain negative correlations is tied to the presence of a spatially varying spectrum may not always be desirable.

As noted in the introduction, spatial deformations provide an alternative to the approach in Paciorek (2003) for producing spatially varying geometric anisotropies. For example, if D is an invertible mapping from $\mathbb{R}^p$ to $\mathbb{R}^p$ and ν is a positive function on $\mathbb{R}^p$, then $\mathcal{M}_{\{\nu(\mathbf{x})+\nu(\mathbf{y})\}/2}(D(\mathbf{x}), D(\mathbf{y}))$ is a covariance function on $\mathbb{R}^p \times \mathbb{R}^p$ that allows spatial variation in both the local smoothness and the local anisotropy of the process. It would be interesting to study the relationship between these covariance functions and those in (2), although either approach is likely to be sufficiently flexible for many applications. However, it may be easier to think about how Σ, the positive definite matrix valued function in (2), controls the strength of correlations, than the deformation D, especially for the larger values of p that can occur in Gaussian process regression. For example, even if one has a way to ensure D is invertible on the domain of interest, it may be difficult to keep track of all interpoint distances induced by D and make sure that no such distances have been made unintentionally too large or too small. Furthermore, Paciorek's approach may be easier to apply to parametric models for nonstationarity, such as when the strength of local dependence is related to some known topographical feature. On the other hand, deformations can be applied to any valid isotropic covariance function on $\mathbb{R}^p$, whereas (1) with g constant implies that the underlying isotropic covariance function (i.e., with Σ constant) be valid in any number of dimensions (Yaglom, 1987, p. 354). Nott and Dunsmuir (2002) describe an approach to extending a covariance structure for a finite number of locations to a valid (generally) nonstationary covariance function on all of $\mathbb{R}^p$ that could also be combined with spatially varying spectra. However, their approach seems more geared towards fitting covariance structures to spatial data at a modest number of monitoring sites with replicates in time rather than to producing models for covariance functions.

For spatial datasets of modest size, a model of the form (2) may provide too much flexibility unless the forms of c, ν and Σ are severely constrained in some manner. For example, even in the simple case where the random field is Gaussian with known mean, all three functions are constant and $\Sigma = \theta I$, then the scalar parameters c and θ cannot be consistently estimated based on observations in a bounded subset of $\mathbb{R}^p$ if $p < 4$ (Zhang 2004). Thus, if one allows c and Σ to vary smoothly in space, even if ν is known and constant, one cannot hope to estimate both c and Σ consistently based on observations from a single realization of the corresponding Gaussian random field. Furthermore, in many applications, especially when the random field is observed with nonnegligible measurement error, it may be difficult to estimate ν locally and one may then want to simplify the model by assuming it is constant. However, in some cases, the local smoothness of a spatial process may exhibit some simple spatial pattern. For example, Fang and Stein (1998) find that the smoothness of longitudinal variations in total column ozone in the Earth's atmosphere shows a clear dependence on latitude. Thus, although we may not often want to use models such as (2) and (3) or sums of them in their full generality, it is useful to have models with such flexibility that we can then specialize to meet the needs of any particular application.

References

Clerc, M. and Mallat, S. (2003), Estimating deformations of stationary processes, *Ann. Statist.* **31**, 1772–1821.

Fang, D. and Stein, M. L. (1998), Some statistical methods for analyzing the TOMS data, *J. Geophys. Res.* **103**, 26,165–26,182.

Fuentes, M. (2002), Spectral methods for nonstationary spatial processes, *Biometrika* **89**, 197–210.

Gneiting, T. and Schlather, M. (2004), Stochastic models that separate fractal dimension and the Hurst effect, *SIAM Rev.* **46**, 269–282.

Gradshteyn, I. S. and Ryzhik, I. M. (2000), *Table of Integrals, Series, and Products* (sixth ed.), (Academic Press, San Diego).

Heyde, C. C. and Yang, Y. (1997), On defining long-range dependence, *J. Appl. Probab.* **34**, 939–944.

Higdon, D. (1998), A process-convolution approach to modeling temperatures in the North-Atlantic Ocean, *Envir. Ecol. Statist.* **5**, 173–190.

Holland, D. M., Cox, W. M., Scheffe, R., Cimorelli, A. J., Nychka, D. and Hopke, P. K. (2003), Spatial prediction of air quality data, *Environmental Manager* (August 2003), 31–35.

Nott, D. J. and Dunsmuir, W. T. M. (2002), Estimation of nonstationary spatial covariance structure, *Biometrika* **89**, 819–829.

Nychka, D., Wikle, C. and Royle, J. A. (2002), Multiresolution models for nonstationary spatial covariance functions, *Statist. Model.* **2**, 315–332.

Paciorek, C. (2003), *Nonstationary Gaussian Processes for Regression and Spatial Modelling*, unpublished Ph.D. dissertation, Carnegie Mellon University, Department of Statistics.

Paciorek, C. and Schervish, M. (2004a), Nonstationary covariance functions for Gaussian process regression, in S. Thrun, L. Saul and B. Schölkopf, eds. *Advances in Neural Information Processing Systems 16* (MIT Press, Cambridge, MA) pp. 273–280.

Paciorek, C. and Schervish, M. (2004b), Spatial smoothing using a new class of nonstationary covariance functions, available at http://www.biostat.harvard.edu/~paciorek/

Perrin, O. and Senoussi, R. (2000), Reducing non-stationary random fields to stationarity and isotropy using a space deformation, *Statist. Probab. Lett.* **48**, 23–32.

Pintore, A. and Holmes, C. (2004), Spatially adaptive non-stationary covariance functions via spatially adaptive spectra, available at http://www.stats.ox.ac.uk/~cholmes/Reports/spectral_tempering.pdf

Sampson, P. D., Damian, D. and Guttorp, P. (2001), Advances in modeling and inference for environmental processes with nonstationary spatial covariance, in P. Monestiez, D. Allard and R. Froidvaux, eds. *GeoENV 2000: Geostatistics for Environmental Applications* (Kluwer, Dordrecht) pp. 17–32.

Sampson, P. D. and Guttorp, P. (1992), Nonparametric estimation of nonstationary spatial covariance structure, *J. Amer. Statist. Assoc.* **87**, 108–119.

Schmidt, A. M. and O'Hagan, A. (2003), Bayesian inference for non-stationary spatial covariance structure via spatial deformations, *J. Roy. Statist. Soc.* B **65**, 743–758.

Seeger, M. (2004), Gaussian processes for machine learning, *Int. J. Neural Syst.* **14**, 69–106.

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

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

Zhang, H. (2004), Inconsistent estimation and asymptotical equal interpolations in model-based geostatistics, *J. Amer. Statist. Assoc.* **99**, 250–261.