Models and Computations for Nonstationary Spatial Processes

Montserrat Fuentes
Department of Statistics
North Carolina State University
Raleigh
N.C. 27695-8203, U.S.A.
fuentes@stat.ncsu.edu

Richard L. Smith and Stanislav Kolenikov
Department of Statistics
University of North Carolina
Chapel Hill
N.C. 27599-3260, U.S.A.
rls@email.unc.edu, skolenik@email.unc.edu

1. Introduction

Modern environmental science typically involves large data sets from monitors distributed in an irregular spatial network. Sometimes the problem is to determine mean levels or trends of a pollutant at the network sites, and to interpolate between the sites. More sophisticated applications involve computer models that simulate meteorology and atmospheric chemistry, but monitoring data are required either to initiate or to validate the computer model. Statistical solutions of these problems require models for spatial processes. To date, the most commonly applied models in these contexts are random field models with stationary, isotropic covariance structure. However, it is increasingly recognized that stationary models are inadequate for practical applications.

In this paper, we briefly survey the currently available classes of nonstationary spatial models, and then propose a new class of models, including a method of fitting the models. Applications discussed involve networks of SO_2 and particular matter monitors over the United States.

2. Models for nonstationary spatial processes

Let $Z(\mathbf{x})$ denote a Gaussian spatial process indexed by a spatial variable $\mathbf{x} \in D$, where D is some domain in $\mathbb{R}^d$ for some $d > 1$. Typically $d = 2$ in applications, and $Z(\mathbf{x})$ denotes the level of some environmental variable at a spatial location $\mathbf{x}$. The process is stationary if the covariance function between two locations

$$C(\mathbf{x}_1, \mathbf{x}_2) = \mathrm{cov}\{Z(\mathbf{x}_1), Z(\mathbf{x}_2)\}$$

depends only on the vector distance $\mathbf{x}_1 - \mathbf{x}_2$ between the two locations. If the covariance depends only on the scalar Euclidean distance $||\mathbf{x}_1 - \mathbf{x}_2||$, then it is said to be *isotropic*. A number of parametric models for stationary, isotropic processes are known, and geostatistical analysis typically consists of selecting one of these models, fitting it to the data and using the resulting model for interpolation and prediction (Ripley 1981, Cressie 1993).

When the process is nonstationary, many possible models are known. The most widely studied classes of models in recent years stem from a seminar paper by Sampson and Guttorp (1992), which assumed that the observed process is stationary and isotropic after applying some smooth nonlinear deformation to the space, i.e.

$$C(\mathbf{x}_1, \mathbf{x}_2) = C_0(||\mathbf{g}(\mathbf{x}_1) - \mathbf{g}(\mathbf{x}_2)||)$$

where $\mathbf{g} : \mathbb{R}^d \to \mathbb{R}^d$ is the deformation and C_0 is a stationary isotropic covariance function. However, there are limitations on the classes of models that can be described in this way, and this has prompted us to look for alternative models.

Haas (1995) has proposed a *moving windows* approach in which prediction at a site is based on a stationary model fitted to observations within a window centered at that site. This approach allows one to approximate quite general classes of spatial models, but it is not easy to patch together the models in different windows to create an overall nonstationary model. Holland *et al.* (1999) proposed an approach combining stationary geostatistical models with the *empirical orthogonal functions* approach widely used in oceanography and atmospheric science. Higdon *et al.* (1999) proposed a model using kernel functions to represent the nonstationary component of the model. Our approach also uses kernel functions but in a different way. Fuentes and Smith (2001) have developed the approach in more detail.

We represent Z as a convolution of local stationary processes,

$$Z(\mathbf{x}) = \int_D K(\mathbf{x} - \mathbf{s}) Z_{\boldsymbol{\theta}(\mathbf{s})}(\mathbf{x}) d\mathbf{s}, \tag{1}$$

where K is a kernel function and $Z_{\boldsymbol{\theta}}(\mathbf{x}),\ \mathbf{x} \in D$ is a family of (independent) stationary Gaussian processes indexed by $\boldsymbol{\theta}$. The parameter $\boldsymbol{\theta}$ is allowed to vary across space to reflect the lack of stationary of the process.

The covariance of $Z_{\boldsymbol{\theta}(\mathbf{s})}$ is stationary with parameter $\boldsymbol{\theta}(\mathbf{s})$,

$$\text{cov}\{Z_{\boldsymbol{\theta}(\mathbf{s})}(\mathbf{s_1}), Z_{\boldsymbol{\theta}(\mathbf{s})}(\mathbf{s_2})\} = C_{\boldsymbol{\theta}(\mathbf{s})}(\mathbf{s_1} - \mathbf{s_2}).$$

A possible covariance model for $Z_{\boldsymbol{\theta}(\mathbf{s})}$ is the Matérn model

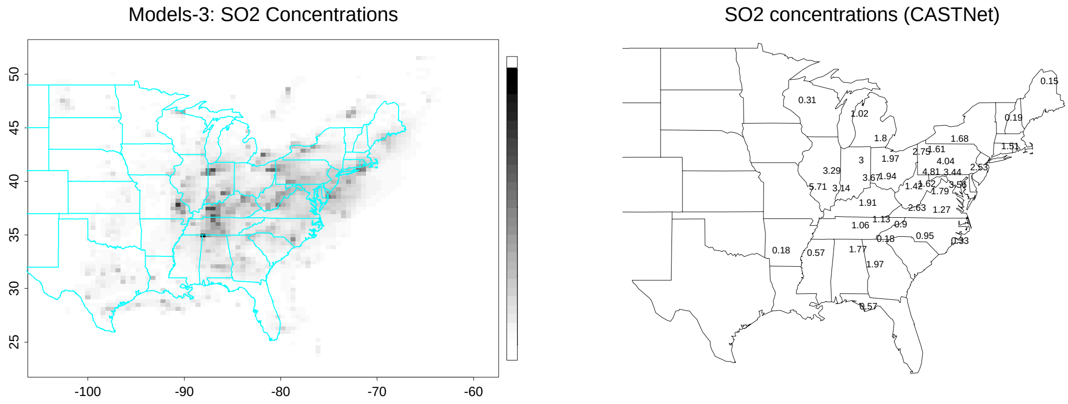
$$C_{\boldsymbol{\theta}(\mathbf{s})}(\mathbf{x}) = \frac{\sigma_{\mathbf{s}}}{2^{\nu_{\mathbf{s}}-1}\Gamma(\nu_{\mathbf{s}})\alpha_{\mathbf{s}}^{2\nu_{\mathbf{s}}}} (2\nu_{\mathbf{s}}^{1/2}|\mathbf{x}|/\rho_{\mathbf{s}})^{\nu_{\mathbf{s}}} \mathcal{K}_{\nu_{\mathbf{s}}}(2\nu_{\mathbf{s}}^{1/2}|\mathbf{x}|/\rho_{\mathbf{s}}), \tag{2}$$

where $\mathcal{K}_{\nu_{\mathbf{s}}}$ is a modified Bessel function and $\boldsymbol{\theta}(\mathbf{s}) = (\nu_{\mathbf{s}}, \sigma_{\mathbf{s}}, \rho_{\mathbf{s}})$ represent the three parameters that define the Matérn process.

The covariance $C(\mathbf{x_1}, \mathbf{x_2}; \theta)$ of Z may be represented as

$$C(\mathbf{x_1}, \mathbf{x_2}; \boldsymbol{\theta}) = \int_D K(\mathbf{x_1} - \mathbf{s}) K(\mathbf{x_2} - \mathbf{s}) C_{\boldsymbol{\theta}(\mathbf{s})}(\mathbf{x_1} - \mathbf{x_2}) d\mathbf{s}. \tag{3}$$

In practice, we fit the model (1) where K is the Epanechnikov kernel $K(\mathbf{u}) \propto (1 - ||\mathbf{u}||^2/h^2)$ with h an arbitrary bandwidth, and the integral over D is replaced by a sum over a grid of cells covering the observation region.



(a) (b)

Figure 1: **(a):** Output of Models-3, weekly average of SO_2 concentrations (ppb), for the week of July 11, 1995. The resolution is 36 km^2 **(b):** Weekly average of SO_2 concentrations (ppb) at the Clean Air Status and Trends Network (CASTNet) sites, for the week of July 11, 1995.

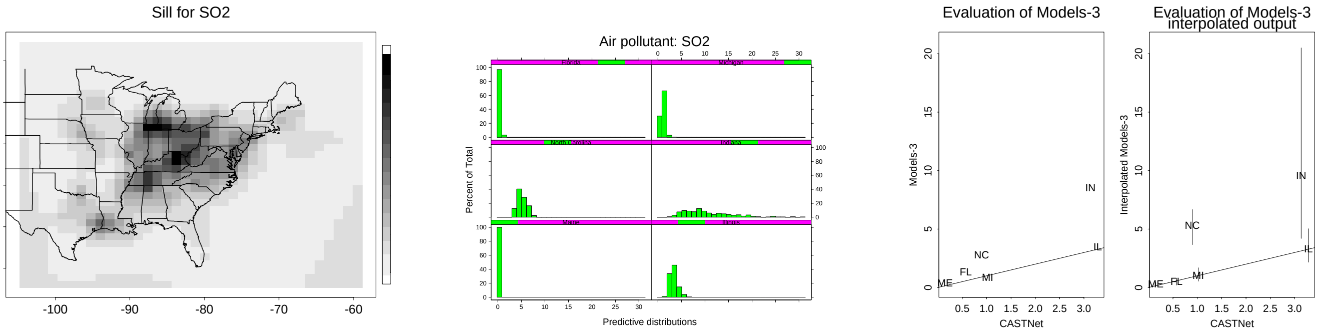
3. Evaluation of a numerical model for sulphur dioxide

As a first application, we consider data from a U.S. monitoring network for SO_2 (known as *CASTNet*), alongside the projections of a computer model for the same conditions ("Models-3" of the U.S. Environmental Protection Agency). As an example, Fig. 1(a) shows the Models-3 data for average SO_2 concentrations (ppb) over the week of July 11, 1995, aggregated in 36 km^2 grid cells, and Fig. 1(b) shows monitoring data at the CASTNet sites for the same week. We would like to be able to evaluate how well Models-3 has succeeded in representing the true conditions as measured at the monitors. The nonstationary model of section 2 was fitted to the Models-3 output data, taking ito account that these are aggregated over grid cells rather than point data, using a Bayesian scheme which was then used to obtain predictive distributions of observations at the CASTNet monitoring sites.

Fig. 2(a) is a map of the local sill parameter (σ_s in the Matérn model), and makes clear the nonconstancy of this parameter. Fig. 2(b) illustrates predictive distributions at CASTNet sites in six states — note, in particular, the high posterior variances for the sites in NC and IN. Fig. 2(c) plots CASTNet values against both raw Models-3 output and the results of the Bayesian predictive analysis for the six sites. The latter plot is much more informative about the fit Models-3 to the real data.

4. Monitoring fine particulate matter

A second application is introduced here and will be presented in more detail during the meeting. In 1997, the U.S. Environmental Protection Agency proposed a new standard for particulate matter, based on fine particles or $PM_{2.5}$. Following this proposal, a new network was set up, comprising some 800 stations, to assess the current state of $PM_{2.5}$ concentrations.



(a) (b) (c)

Figure 2: **(a):** Spatial estimates of the local sill parameter. **(b):** Posterior predictive distributions at CASTNet sites in FL, MI (top), NC, IN (middle), ME, IL (bottom). **(c):** Left, CASTNet measurements for the week starting July 11, 1995, versus the values of Models-3 for the pixels that are the closest to each CASTNet site. Right, CASTNet measurements versus the modes and 90% credible intervals of the posterior predictive distributions derived from Models-3 at the CASTNet locations.

In ongoing work, we propose nonstationary spatial-temporal models for the data and use them to reconstruct a surface of mean $PM_{2.5}$ levels across the United States.

REFERENCES

Cressie, N. (1993), *Statistics for Spatial Data.* Second edition, John Wiley, New York.

Fuentes, M. and Smith, R.L. (2001), A new class of nonstationary spatial models. Manuscript in preparation.

Haas, T.C. (1995), Local prediction of a spatio-temporal process with an application to wet sulfate deposition. *J. Amer. Statist. Assoc.* **90**, 1189–1199.

Higdon, D., Swall, J. and Kern, J. (1999), Non-stationary spatial modeling. In *Bayesian Statistics 6*, eds. J.M. Bernardo *et al.*, Oxford University Press, pp. 761–768.

Holland, D., Saltzman, N., Cox, L.H. and Nychka, D. (1999), Spatial prediction of sulfur dioxide in the eastern United States. In *geoENV II — Geostatistics for Environmental Applications*, eds. Gómez-Hernández, J., Soares, A. and Froidevaux, R., Kluwer, Dordrecht, 65–76.

Ripley, B.D. (1981), *Spatial Statistics.* Wiley, New York.

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

RESUME

We propose a new class of nonstationary spatial processes and associated estimation methods. Applications are given to the spatial distributions of sulphur dioxide and particulate matter over the United States.