GEOSTATISTICAL MODELING

Richard L. Smith
University of North Carolina, Chapel Hill

rls@email.unc.edu
www.stat.unc.edu/faculty/rsmith.html

SAMSI/GSP WORKSHOP ON SPATIAL-TEMPORAL STATISTICS

JUNE 2 2003

1. **Spatial covariances**
2. **Model identification and estimation**
3. **Prediction and interpolation**
4. **Spatial-temporal models**
5. **Application to data on fine particulate matter**

Major references

Cressie (1993)

Stein (1999)

Chilès and Delfiner (1999)

Matérn (1986)

My own course notes:

http://www.stat.unc.edu/postscript/rs/envstat/env.html

Software

S-PLUS Spatial Statistics module

SAS PROC MIXED (for ML or REML estimation of variogram models; there are also variogram plotting and kriging procedures within SAS)

The GeoR package (includes instructions for downloading R):

http://www.est.ufpr.br/geoR/

GSP's "Fields" package:

http://www.cgd.ucar.edu/stats/Software/Fields/

My own programs and data sets:

http://www.unc.edu/depts/statistics/postscript/rs/envstat/env2.html

1. Spatial covariances

Basic structure: A stochastic process $\{Z(s),\ s \in D\}$, $D \subseteq \mathcal{R}^d$, usually though not necessarily $d = 2$.

Mean function

$$\mu(s) = \mathrm{E}\{Z(s)\}, \quad s \in D.$$

Covariance function

$$C(s_1, s_2) = \mathrm{Cov}\{Z(s_1), Z(s_2)\}.$$

Z is *Gaussian* if all joint distributions are multivariate normal.

Z is *second-order stationary* if $\mu(s) \equiv \mu$ and

$$\mathrm{Cov}\{Z(s_1), Z(s_2)\} = C(s_1 - s_2),$$

for all $s_1 \in D$, $s_2 \in D$, where $C(s)$ is $\mathrm{Cov}\{Z(s), Z(0)\}$.

The Variogram. Assume $\mu(s)$ is a constant, which we may without loss of generality take to be 0, and then define

$$\mathrm{Var}\{Z(s_1) - Z(s_2)\} = 2\gamma(s_1 - s_2).$$

This makes sense only if the left hand side depends on s_1 and s_2 only through their difference $s_1 - s_2$. Such a process is called *intrinsically stationary*. The function $2\gamma(\cdot)$ is called the *variogram* and $\gamma(\cdot)$ the *semivariogram*.

Intrinsic stationarity is weaker than second-order stationarity. However, if the latter holds we have

$$\gamma(h) = C(0) - C(h).$$

We shall usually assume second-order stationarity though many of the results hold just assuming intrinsic stationarity.

Isotropy. Suppose the process is intrinsically stationary with semivariogram $\gamma(h)$, $h \in \mathcal{R}^d$. If $\gamma(h) = \gamma_0(||h||)$ for some function γ_0, i.e. if the semivariogram depends on its vector argument h only through its length $||h||$, then the process is *isotropic*.

Examples:

1. *Exponential-power form:*

$$\gamma_0(t) = \begin{cases} 0 & \text{if } t = 0, \\ c_0 + c_1(1 - e^{-|t/R|^p}) & \text{if } t > 0. \end{cases}$$

Here $0 < p \leq 2$. $p = 1$ is called *exponential*, $p = 2$ is *Gaussian*.

2. *Spherical:* (for d=1,2,3,)

$$\gamma_0(t) = \begin{cases} 0 & \text{if } t = 0, \\ c_0 + c_1\left\{\frac{3}{2}\frac{t}{R} - \frac{1}{2}(\frac{t}{R})^3\right\} & \text{if } 0 < t \leq R, \\ c_0 + c_1 & \text{if } t \geq R. \end{cases}$$

3. *Power law:*

$$\gamma_0(t) = \begin{cases} 0 & \text{if } t = 0, \\ c_0 + c_1 t^\lambda & \text{if } t > 0. \end{cases}$$

Valid if $0 \leq \lambda < 2$. $\lambda = 1$ is *linear variogram*. This case is *not* second-order stationary.

4. *Matérn:*

$$C_0(t) = \frac{1}{2^{\theta_2 - 1}\Gamma(\theta_2)} \left(\frac{2\sqrt{\theta_2}t}{\theta_1}\right)^{\theta_2} \mathcal{K}_{\theta_2}\left(\frac{2\sqrt{\theta_2}t}{\theta_1}\right).$$

$\theta_1 > 0$ is the spatial scale parameter and $\theta_2 > 0$ is a shape parameter. $\Gamma(\cdot)$ is the usual gamma function while $\mathcal{K}_{\theta_2}$ is the modified Bessel function of the third kind of order θ_2. $\theta_2 = \frac{1}{2}$ corresponds to the exponential form of semivariogram, and the limit $\theta_2 \to \infty$ results in the Gaussian form.

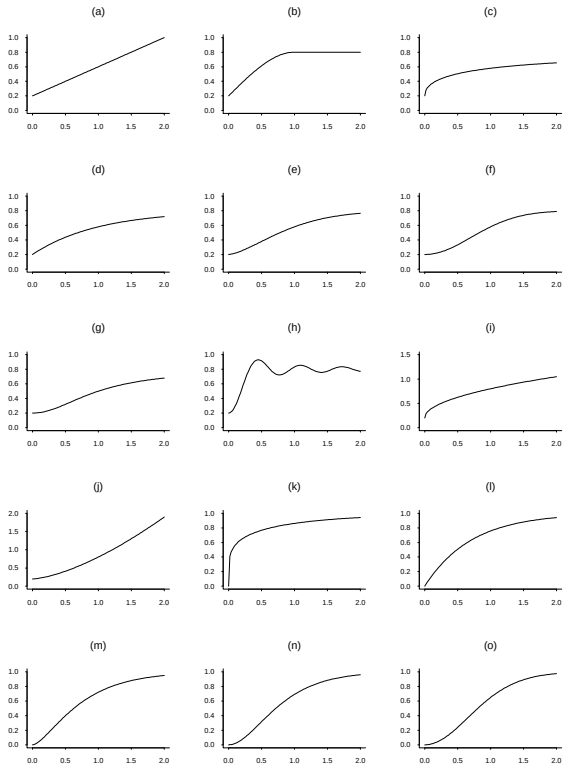

Fig. 1. Examples of isotropic variograms.

(a) Linear

(b) Spherical

(c) Exponential power, $p = 0.5$

(d) Exponential

(e) Exponential power, $p = 1.5$

(f) Gaussian

(g) Rational quadratic

(h) Wave

(i) Power, $\lambda = 0.5$.

(j) Power, $\lambda = 1.5$.

(k)–(o) Matérn with $\theta_2 = 0.1$, 0.5, 1, 2, 10.

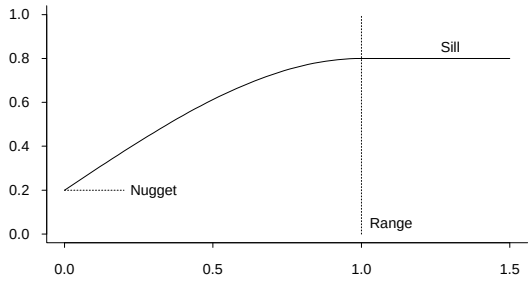

Fig. 2. Idealized form of variogram function, illustrating the nugget, sill and range.

Positive definiteness

Any covariance function must satisfy

$$\sum_i \sum_j a_i a_j C(s_i, s_j) \geq 0$$

for any finite set of points $s_1, ..., s_n$ and arbitrary real coefficients $a_1, ..., a_n$. For the variogram, if $\sum a_i = 0$,

$$\sum_i \sum_j a_i a_j \gamma(s_i - s_j) \leq 0.$$

These conditions are obviously necessary. That they are also sufficient is a consequence of *Bochner's theorem.*

More complete characterizations follow through spectral representations (see lectures by Fuentes and Stein at this worksop)

2. Model identification and estimation

Assume a process $\{Z(s),\ s \in D\}$ observed at a finite number of points $s_1, ..., s_N$.

The *sample variogram* is often used as an initial guide to the form of spatial model. It can be drawn as either a *variogram cloud*, or a *binned variogram*.

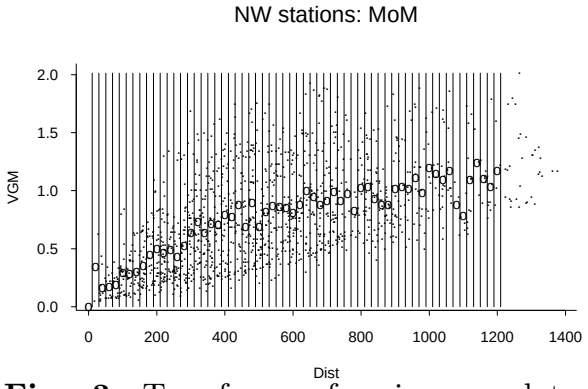

Fig. 3. Two forms of variogram plot superimposed.

Fitting parametric models

Sample variogram not negative definite: therefore, not acceptable as an estimate of population variogram

Solution: fit a parametric model

- curve fitting to the variogram,
- maximum likelihood (ML),
- restricted maximum likelihood (REML),
- Bayesian estimators.

We concentrate here on the ML and REML procedures. They are generally practical for numerical fitting, and can be combined with various model selection measures (e.g. AIC, BIC, likelihood ratio testing) to choose among parametric models.

Maximum likelihood estimation
(Mardia and Marshall 1984)

Assume Gaussian process. General model (includes regression terms):

$$Z \sim N_n(X\beta, \Sigma),$$
$$\Sigma = \alpha V(\theta),$$

X a $n \times q$ matrix of covariates, α a scale parameter and $V(\theta)$ determined by θ, parameters of spatial model.

Minimize the negative log likelihood function:

$$\ell(\beta, \alpha, \theta) =$$
$$\frac{n}{2}\log(2\pi) + \frac{n}{2}\log\alpha + \frac{1}{2}\log|V(\theta)|$$
$$+ \frac{1}{2\alpha}(Z - X\beta)^T V(\theta)^{-1}(Z - X\beta).$$

The minimization w.r.t. β and then α can be carried out analytically, so the only function that has to be minimized numerically is the *profile neg log likelihood* of θ,

$$\ell^*(\theta) = \text{const } + \frac{n}{2}\log\frac{G^2(\theta)}{n} + \frac{1}{2}\log|V(\theta)|.$$

where $G^2(\theta) = (Z - X\hat{\beta})^T V(\theta)^{-1}(Z - X\hat{\beta})$, $\hat{\beta}$ the GLS estimator of β.

The numerical methods use Cholesky decomposition of V to facilitate computations of $\hat{\beta}$ and also $|V|$, then optimize using a standard nonlinear optimization method, typically quasi-Newton.

Restricted maximum likelihood

(Patterson & Thompson 1971, Harville 1974)

Let $W = A^T Z$ be a vector of $n - q$ linearly independent contrasts, i.e. the $n - q$ columns of A are linearly independent and $A^T X = 0$, then we find that

$$W \sim \mathcal{N}(0, A^T \Sigma A).$$

The density of W is taken to define the neg log likelihood function. After some manipulation, this reduces to

$$\ell_W(\alpha, \theta) = \frac{n-q}{2} \log(2\pi) + \frac{n-q}{2} \log \alpha - \frac{1}{2} \log |X^T X| + \frac{1}{2} \log |X^T V(\theta)^{-1} X| + \frac{1}{2} \log |V(\theta)| + \frac{1}{2\alpha} G^2(\theta).$$

This formula also has a Bayesian interpretation (Harville).

3. Prediction and interpolation

Consider the *universal kriging* model

$$Y = X\beta + \eta,$$

where Y is a vector of observations, β is an unknown vector of regression coefficients and η is a vector of correlated random errors with mean 0 and covariance matrix Σ. Suppose we wish to predict a value y_a, given by

$$y_a = x_a^T \beta + \eta_a,$$

where x_a is known, η_a has mean 0 and variance σ^2, and $E\{\eta_a \eta\} = \tau_a$.

The GLS estimator of β is given by

$$\hat{\beta} = (X^T \Sigma^{-1} X)^{-1} X^T \Sigma^{-1} Y,$$

and the covariance matrix of $\hat{\beta}$ is $(X^T \Sigma^{-1} X)^{-1}$.

From standard theory of the multivariate normal distribution, the conditional expectation of η_a given η is $\hat{\eta}_a = \tau_a^T \Sigma^{-1} \eta$. Therefore, the logical predictor of y_a is

$$\hat{y}_a = x_a^T \hat{\beta} + \tau_a^T \Sigma^{-1} (Y - X\hat{\beta})$$

so that

$$\hat{y}_a - y_a = (x_a^T - \tau_a^T \Sigma^{-1} X)(\hat{\beta} - \beta) + (\tau_a^T \Sigma^{-1} \eta - \eta_a). \quad (1)$$

However, the two terms of (1) are uncorrelated (direct calculation). This facilitates computation of prediction error variances.

If we are considering the covariance of two predictions $y_a = x_a^T \beta + \eta_a$, $y_b = x_b^T \beta + \eta_b$, and denoting the covariance vectors of Y with y_a and y_b respectively by τ_a, τ_b, we calculate

$$\begin{aligned} &Cov\{\hat{y}_a - y_a, \hat{y}_b - y_b\} \\ &= Cov\big\{(x_a^T - \tau_a^T \Sigma^{-1} X)(\hat{\beta} - \beta), \\ &\qquad (x_b^T - \tau_b^T \Sigma^{-1} X)(\hat{\beta} - \beta)\big\} \\ &\quad + Cov\left\{(\tau_a^T \Sigma^{-1} \eta - \eta_a), (\tau_b^T \Sigma^{-1} \eta - \eta_b)\right\} \\ &= (x_a^T - \tau_a^T \Sigma^{-1} X)(X^T \Sigma^{-1} X)^{-1} (x_b - X^T \Sigma^{-1} \tau_b) \\ &\quad - \tau_a^T \Sigma^{-1} \tau_b + Cov\{\eta_a, \eta_b\}. \end{aligned} \quad (2)$$

When $a = b$ we call this the mean squared prediction error, abbreviated MSPE.

Extensions

1. Computing an areal average (e.g., estimating the average concentration of ozone over a given region)

Usually compute on a grid, so we predict $\frac{1}{|A|}\sum_{a\in A} y_a$ by $\frac{1}{|A|}\sum_{a\in A}\hat{y}_a$.

The MSPE in this case is

$$\frac{1}{|A|^2}\sum_{a\in A, b\in A}\mathrm{Cov}\left\{\hat{y}_a - y_a, \hat{y}_b - y_b\right\},$$

which is readily calculated from (2).

Computing a temporal average

(This is a less "standard" application, but it's needed for our example in Section 5.)

Suppose we have independent replicates of the random field Y_t at each time $t = 1, ..., T$, and moreover, the regression component $X_t\beta$ has the same coefficients at each t, but it's possible the covariance matrix Σ_t and the vector of cross-covariances $\tau_{a,t}$ may be different for each t (in particular, this will be the case if the network changes during the observation period)

For each time t and location a predict

$$\hat{y}_{at} = x_{at}^T\hat{\beta} + \tau_{at}^T\Sigma_t^{-1}(Y_t - X_t\hat{\beta})$$

where $\hat{\beta}$ is the GLS computed from the whole data set.

We again have (in obvious notation)

$$\begin{aligned}\hat{y}_{at} - y_{at} &= (x_{at}^T - \tau_{at}^T\Sigma_t^{-1}X_t)(\hat{\beta} - \beta)\\ &\quad + (\tau_{at}^T\Sigma_t^{-1}\eta_t - \eta_{at}),\end{aligned}$$

where the two components are independent.

For the temporal average we have

$$\begin{aligned}&\frac{1}{T}\sum_{t=1}^{T}(\hat{y}_{at} - y_{at})\\ &= \frac{1}{T}\sum_{t=1}^{T}(x_a^T - \tau_{at}^T\Sigma_t^{-1}X_t)(\hat{\beta} - \beta)\\ &\quad + \frac{1}{T}\sum_{t=1}^{T}(\tau_{at}^T\Sigma_t^{-1}\eta_t - \eta_{at}).\end{aligned}$$

The variance of the first component is derived directly from that of $\hat{\beta}$, while the variance of the second component is also a sum of independent components and therefore easily calculated.

4. Spatial-temporal models

The direct generalization of spatial statistics to spatial-temporal data is based on finding classes of spatial-temporal covariance functions that obey the positive definiteness property, for which the preceding theories of estimation, interpolation etc., go through directly.

We concentrate here on two specific classes, *separable models* and the *repeated measurements model*. These are the simplest cases and often dismissed as too simple for real applications, but their mathematical properties make them appealing when they are applicable.

The separable model is defined by

$$C(h, u) = C_0(h)\gamma(u)$$

where $C(h, u)$ denotes the covariance between two space-time coordinates with spatial separation h and temporal separation u, $C_0(h)$ is a pure spatial covariance and $\gamma(u)$ is a temporal autocovariance. Since we may always transfer a constant between the functions C_0 and γ, there is no loss of generality in assuming $\gamma(0) = 1$, in other words, that γ is a temporal autocorrelation function.

The special case where $\gamma(u) = 0$ for all $u \neq 0$ was called the *repeated measurements model* by Mardia and Goodall (1993).

5. Application to data on fine particulate matter

Ref.: Smith, Kolenikov and Cox (2003).

In 1997, the U.S. Environmental Protection Agency (EPA) proposed a new air pollution standard for fine particulate matter ($PM_{2.5}$). One of the requirements is that the mean level of $PM_{2.5}$ at any location should be no more than 15 $\mu g/m^3$. A network of several hundred monitors has been set up to assess this.

The present study is based on a small portion of this network, 74 monitors in North Carolina, South Carolina and Georgia. Fig. 4 is a map of the monitor locations. We converted the raw values to weekly averages, but even so more than $\frac{1}{4}$ of the data are missing. The EPA also recorded a "land-use" variable, classified as one of five types of land-use: agricultural (A), commercial (C), forest (F), industrial (I) and residential (R).

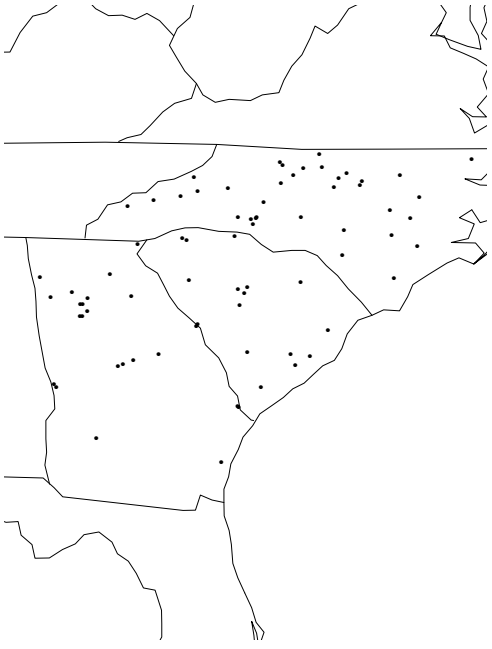

Fig. 4. $PM_{2.5}$ monitors in North Carolina, South Carolina and Georgia: Map of spatial locations

Exploratory data analysis

The first issue considered is whether to make any transformation, such as square roots or logarithms, of the raw $PM_{2.5}$ values. Fig. 5 shows a plot of sample variance against sample mean, across all 74 stations, for each of three transformations, (a) no transformation, (b) square root transformation, (c) logarithmic transformation. On the basis that (b) is the closest fit to a constant-variance model, the rest of the analysis is based on the square root of $PM_{2.5}$ as a variance-stabilizing transformation.

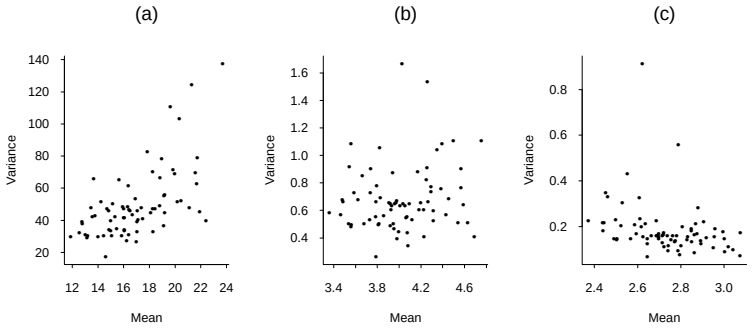

Fig. 5. Variance *vs.* mean plots for each monitor in $PM_{2.5}$ data. (a) Original scale of data. (b) Mean and variance computed after square root transformation. (c) Mean and variance computed after logarithmic transformation.

Time trend:

The time trend was estimated both as a B-spline smooth curve and (more simply) by using a weekly indicator variable to represent the overall mean level for that week.

Fig. 6(a) shows both versions of the fitted time trend, with all data points superimposed. Also shown on Fig. 6 are the same fitted time trend curves, but with different portions of the data superimposed, (b)–(d) corresponding to each of the three states, (e)–(i) corresponding to each of the five land-use variables. The results show a significant discrepancy between states, with Georgia values generally higher than the overall mean, while the land-use variables show significant variations in the directions one would expect.

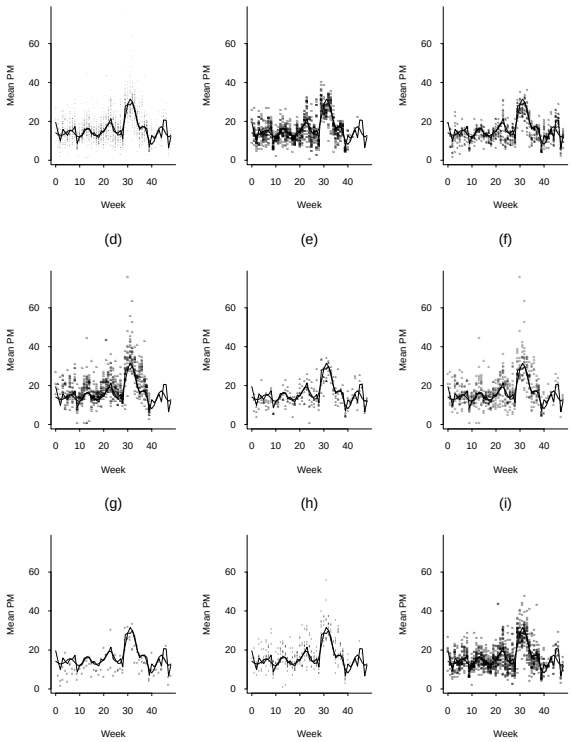

Fig. 6. Fitted common temporal trend with data values from (a) all sites, (b) NC, (c) SC, (d) GA, (e)–(i) five landuse types

These comparisons suggest the model

$$y_{xt} = w_t + \psi_x + \theta_x + \eta_{xt} \qquad (3)$$

in which y_{xt} is the square root of $PM_{2.5}$ in location x in week t, w_t is a week effect, ψ_x is the spatial mean at location x (in practice, estimated through a thin-plate spline representation), θ_x is a land-use effect corresponding to the land-use as site x, and η_{xt} is a random error.

So far we have ignored temporal and spatial correlations among the η_{xt}, but we consider these next.

Spatial and temporal dependence

Take residuals from preceding linear regression.

Fig. 7 shows the first five autocorrelations for residuals from each of the 74 stations, superimposed on one graph. Also shown (as horizontal straight lines) are the approximate bounds of the critical region for a hypothesis test of size 0.05 for the null hypothesis that there is no temporal autocorrelation. Very few of the sample autocorrelations are outside the critical region defined by the horizontal straight lines; based on this, we accept the null hypothesis that there is no temporal autocorrelation.

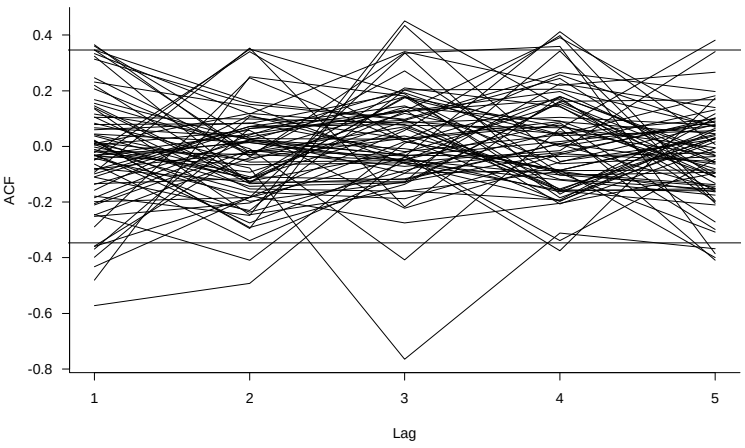

Fig. 7. First 5 autocorrelations for residuals from linear regression in each of the 74 spatial locations, with approximate bounds for the critical region of a hypothesis test of size 0.05 for the null hypothesis that there is no autocorrelation (horizontal lines)

Spatial correlations:

Fig. 8 shows variograms of residuals from simple linear regression, where a number of subsets of the data (classified by state and also by season) have been identified to look for comparability of the estimated variogram among different subsets of data.

Features include

(a) substantial inhomogeneity among subgroups despite initial variance stabilization

(b) does not seem to follow standard nugget-range-sill shape

We fit the power law variogram

$$\gamma(h) = \begin{cases} 0 & \text{if } h = 0, \\ \theta_0 + \theta_1 h^\lambda & \text{if } h > 0, \end{cases} \qquad (4)$$

where $\theta_0 > 0,\ \theta_1 > 0,\ 0 \le \lambda < 2$.

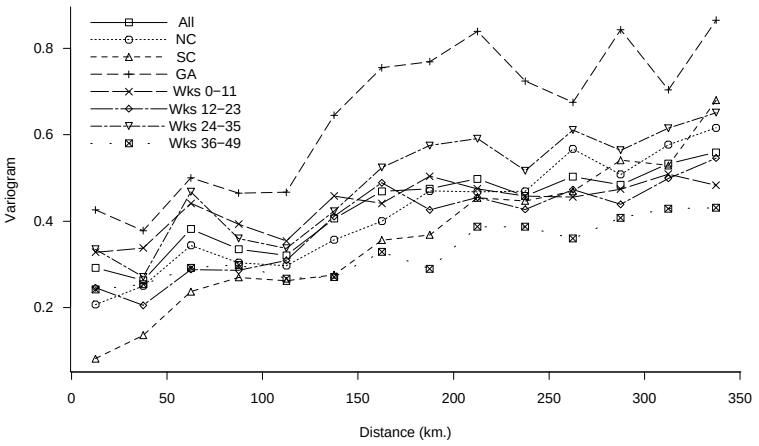

Fig. 8. Spatial variograms computed for various subsets of the data.

To fit this model by maximum likelihood, we need to concept of *generalized covariances*, introduced by Matheron (1973). For modern references see Cressie (1993), Chilès and Delfiner (1999) or Stein (1999). In the present context the key formula is the following: for an intrinsically stationary process defined by a semivariogram γ,

$$Cov\left\{\sum_x \nu_x \eta_{x,t}, \sum_{x'} \kappa_{x'} \eta_{x',t}\right\} = \sum_x \sum_{x'} \nu_x \kappa_{x'} G(||x - x'||),$$

provided $\sum_x \nu_x = \sum_{x'} \kappa_{x'} = 0$. Here G is known as the generalized covariance function: however for an intrinsically stationary process, it suffices to take $G = -\gamma$.

Practical implementation:

In (3), replace each y_{xt} by $y^*_{xt} = y_{xt} - \frac{1}{n_t}\sum_{x'} y_{x't}$ where the second sum is over all x' values available in week t; n_t is the number of such x' values in a given week. With some further simplifications we replace (3) by

$$y^*_{xt} = \psi^*_x + \theta^*_x + \eta^*_{xt} \qquad (5)$$

where

$$Cov\{\eta^*_{x,t}, \eta^*_{x',t}\} = \frac{1}{n_t}\sum_{x_1} \gamma(||x - x_1||) + \frac{1}{n_t}\sum_{x_1} \gamma(||x' - x_1||) - \gamma(||x - x'||) - \frac{1}{n_t^2}\sum_{x_1}\sum_{x_2} \gamma(||x_1 - x_2||). \qquad (6)$$

The model defined by (4)—(6) may now be fitted by maximum likelihood.

There are additional complications because of the missing values, which mean that n_t and the fitted covariance matrix are different from week to week. The present data set is relatively small and we were still able to compute exact maximum likelihood, but some variants of the EM algorithm (Little and Rubin 1987, McLachlan and Krishnan 1997) were also used, and remain the focus of further research.

Results

The model (5) was fitted to the data values from which each weekly mean had been subtracted. The residuals η^*_{xt} were assumed independent at different time points but with spatial covariances given by (6) with (4). As an example of the results, the maximum likelihood of the parameter θ_2 was 0.92 with standard error 0.097. Since a linear variogram corresponds to $\theta_2 = 1$, this shows that the spatial dependence is not significantly different from a linear variogram.

The fitted model was then used to construct a predicted surface, with estimated root mean squared prediction error (RMSPE), for each week of the year and also for the average over all weeks. The latter is of greatest interest in the context of EPA standards setting.

Fig. 9 shows the predicted surface and RMSPE for week 33 (the week with highest average $PM_{2.5}$) and overall for the annual mean. Fig. 10 shows the estimated probability that any particular location exceeds the 15 $\mu g/m^3$ annual mean standard. These maps are based on kriging the residuals η^*_{xt} in (4) and then combining them with the estimated fixed effects for ψ^*_x and θ^*_x, transforming back to the original scale of the data for the actual plots. The RMSPE values used here take into account the averaging of kriged values, but do not take account of the additional uncertainty in estimating the parameters θ_1 and θ_2. Fig. 10 is based on the assumption that (on a square root scale) the difference between the predicted and true values, scaled by the RMSPE, has a standard normal distribution.

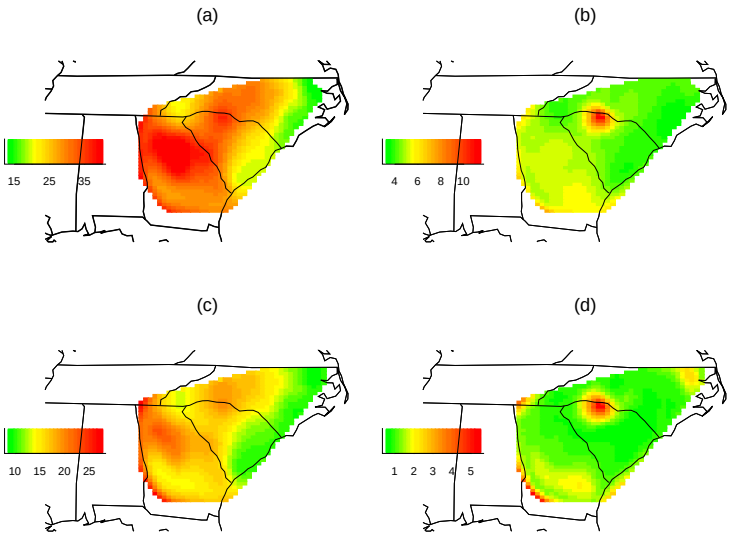

Fig. 9. (a) Map of estimated $PM_{2.5}$ surface in residential locations for week 33 of data. (b) RMS prediction errors for map in (a). (c) Map of estimated $PM_{2.5}$ surface in residential locations averaged over all weeks of data. (d) RMS prediction errors for map in (c).

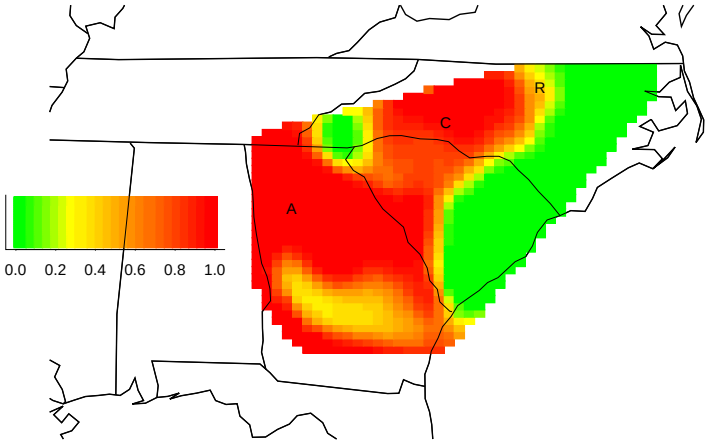

Fig. 10. Estimated probability that the annual average $PM_{2.5}$ exceeds the EPA standard of 15 $\mu g/m^3$ at each location; the letters A, C and R indicate the positions of Atlanta, Charlotte, and Raleigh.

It can be seen that substantial parts of the region, including the western portions of North and South Carolina and virtually the whole of the state of Georgia, appear to be in violation of the standard. Of the three major cities marked on Fig. 10, Atlanta and Charlotte are clearly in the "violation" zone; Raleigh is on the boundary of it.

In future work, we hope to extend this analysis to other parts of the country (this will certainly involve consideration of nonstationary spatial models), to analyze more recent data, and to consider the associated "network design" questions.

REFERENCES

Chilès, J.-P. and Delfiner, P. (1999). *Geostatistics: Modeling Spatial Uncertainty.* John Wiley, New York.

Cressie, N. (1993), *Statistics for Spatial Data.* 2nd edition, Wiley, New York.

Harville, D.A. (1974), Bayesian inference for variance components using only error contrasts. *Biometrika* **61**, 383-385.

Little, R.J.A., and Rubin, D.B. (1987). *Statistical Analysis with Missing Data.* Wiley Series in Probability and Statistics, John Wiley, New York.

Mardia, K.V. and Goodall, C.R. (1993), Spatial-temporal analysis of multivariate environmental monitoring data. In *Multivariate Environmental Statistics*, eds. G.P. Patil and C.R. Rao, Elsevier Science Publishers, pp. 347–386.

Mardia, K.V. and Marshall, R.J. (1984), Maximum likelihood estimation of models for residual covariance in spatial regression. *Biometrika* **71**, 135-146.

Matérn, B. (1986), *Spatial Variation.* Lecture Notes in Statistics, Number 36, Springer Verlag, New York. (Second edition: originally published in 1960).

Matheron, G. (1973), The intrinsic random functions and their applications. *Adv. Appl. Prob.* **5**, 439–468.

McLachlan, G.J., and Krishnan, T. (1997) *The EM Algorithm and Extensions.* Wiley Series in Probability and Statistics, John Wiley, New York.

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

Smith, R.L., Kolenikov, S. and Cox, L.H. (2003), Spatio-temporal modeling of $PM_{2.5}$ data with missing values. *Journal of Geophysical Research–Atmospheres*, to appear. http://www.stat.unc.edu/postscript/rs/pm25.pdf

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