

Kriging with Nonparametric Variance Function Estimation

J.D. Opsomer, D. Ruppert, M.P. Wand, U. Holst, and O. Hössjer

Working Paper 98-WP 188

February 1998

Kriging with Nonparametric Variance Function Estimation

J.D. Opsomer, D. Ruppert, M.P. Wand,
U. Holst, and O. Hössjer

Working Paper 98-WP 188
February 1998

**Center for Agricultural and Rural Development
Iowa State University
Ames, IA 50011-1070**

The authors are J.D. Opsomer, Department of Statistics, Iowa State University; D. Ruppert, Cornell University; M.P. Wand, Harvard University; U. Holst, University of Lund; and O. Hössjer, University of Lund.

This research was supported by the Swedish National Board for National and Technical Development Grant 91-02637P, National Science Foundation Grant #DMS-9626762, and by the Center for Agricultural and Rural Development, Iowa State University.

The authors thank Wayne Fuller for helpful suggestions in creating the bias-adjusted variogram.

Abstract

A method for fitting regression models to data that exhibit spatial correlation and heteroskedasticity is proposed. A combination of parametric and nonparametric regression techniques is used to iteratively estimate the various components of the model. The approach is demonstrated on a large dataset of predicted nitrogen runoff from agricultural lands in the Midwest and Northern Plains regions of the U.S. For this dataset, the model is comprised of three main components: (1) the mean function which includes farming practice variables, local soil and climate characteristics and the nitrogen application treatment, is assumed linear in the parameters and fitted by generalized least squares, (2) the variance function, which contains a local as well as a spatial component whose shapes are left unspecified, is estimated by local linear regression, and (3) the spatial correlation function is estimated by fitting a parametric variogram model to the standardized residuals and after adjusting the variogram for the presence of heteroskedasticity. The fitting of these three components is iterated until convergence. The model provides an improved fit to the data compared to a previous model that ignored the heteroskedasticity and the spatial correlation.

1 Introduction

For many practical problems, the degree to which components of the statistical model can be specified in a parametric form varies dramatically. When the model is misspecified, the resulting model fit can be biased and the possibility for making wrong inferences exists. On the other hand, when part of the model is amenable to parametric fitting, it is useful to do so in order to have a more analytically tractable model and be able to use traditional inference techniques. Even in the most common form of nonparametric regression where the mean function is left unspecified, it is common to assume that the observations are uncorrelated, which can be viewed as a “parametric” assumption on the distribution of the errors. Violation of that assumption has a serious effect on the optimal bandwidth for estimating that mean function (see Hart (1991) and Opsomer (1996)).

In this article, we consider an application where it appears reasonable to accept a (roughly) linear relationship between dependent and independent variables and the observations clearly display spatial dependence, but where the spatial variance cannot be specified a priori. The proposed approach, which blends elements of parametric and nonparametric fitting, is not only applicable for this specific situation, however. One of the goal of this article is to show that parametric and nonparametric techniques are often complementary and when used appropriately, can lead to better model specification than either set of techniques used in isolation.

We begin by describing the application that motivated this research. Economists at the Center for Agricultural and Rural Development at Iowa State University (CARD) are developing models to evaluate the impact of federal and state agricultural policies on the nitrogen water pollution in the Midwest and Northern Plains of the U.S. (Wu, Lakshminarayan and Babcock (1996)). In a departure from previous economic models, the goal of this project is to predict the environmental impacts at both the regional *and* the local level. Local prediction is achieved by using the 128,591 National Resources Inventory (NRI) points in the region of interest as the basis for the evaluation of pollution impact: in addition to providing detailed information on most of the independent variables (see Table 1), the NRI database provides the sampling weights allowing statistically valid area predictions based

3 Estimation Procedure

3.1 Overview

Let \mathbf{Y} be the $n \times 1$ vector containing the Y_{ij} 's and \mathbf{Z} be the $n \times q$ matrix with (i, j) th row equal to \mathbf{Z}_{ij}^T . Let Σ be the variance-covariance matrix of \mathbf{Y} . Let p be a positive integer-valued tuning parameter. The role of p is to determine the minimum number of replicates at an \mathbf{x}_i in order to use that location for estimating the variance functions. The choice of p is discussed later.

STEP 0: (Initialization step) Set $\hat{\Sigma} = \mathbf{I}$.

STEP 1: Obtain

$$\hat{\beta} = (\mathbf{Z}^T \hat{\Sigma}^{-1} \mathbf{Z})^{-1} \mathbf{Z}^T \hat{\Sigma}^{-1} \mathbf{Y}.$$

STEP 2: Set

$$r_{ij} = Y_{ij} - \mathbf{Z}_{ij}^T \hat{\beta} \quad \text{and} \quad \bar{r}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} r_{ij}.$$

STEP 3: For each \mathbf{x}_i such that $n_i \geq p > 1$ obtain

$$\tilde{v}_u(\mathbf{x}_i) = \frac{1}{n_i - 1} \sum_{j=1}^{n_i} (r_{ij} - \bar{r}_i)^2.$$

STEP 4: Obtain $\hat{v}_u(\mathbf{x}_i)$ for all \mathbf{x}_i by local linear smoothing of the $\tilde{v}_u(\mathbf{x}_i)$.

STEP 5: Obtain $\hat{v}_\varepsilon(\mathbf{x}_i; h)$ by local linear smoothing of $\{\tilde{v}_\varepsilon(\mathbf{x}_i) : n_i > p\}$ where

$$\tilde{v}_\varepsilon(\mathbf{x}_i) = (\bar{r}_i)^2 - \frac{\tilde{v}_u(\mathbf{x}_i)}{n_i}.$$

STEP 6: Define $\hat{v}_r(\mathbf{x}_i) = \hat{v}_\varepsilon(\mathbf{x}_i) + \hat{v}_u(\mathbf{x}_i)/n_i$ and let $\hat{\varepsilon}_i = \bar{r}_i/\hat{v}_r(\mathbf{x}_i)^{1/2}$. Estimate θ in the isotropic correlation model $\rho(\cdot; \theta)$.

STEP 7: Obtain

$$\hat{\Sigma} = \hat{\Sigma}_\varepsilon + \hat{\Sigma}_u$$

where $(\hat{\Sigma}_u)_{ij, i'j'} = \hat{v}_u(\mathbf{x}_i)$ if $i = i', j = j'$ and 0 otherwise, and

$$[\hat{\Sigma}_\varepsilon]_{ij, i'j'} = \hat{v}_\varepsilon(\mathbf{x}_i)^{1/2} \hat{v}_\varepsilon(\mathbf{x}_{i'})^{1/2} \rho(\|\mathbf{x}_i - \mathbf{x}_{i'}\|; \hat{\theta}).$$

on the point predictions (Nusser and Goebel (1997)).

Nitrogen pollution occurs via two primary pathways: by nitrogen runoff into surface waters, and by leaching through the soil into the groundwater. In the current article, we will focus on the prediction of nitrogen runoff. Table 1 shows the variables used in the model. They are further described in Wu *et al.* (1996). A map of the study regions containing the locations of the weather stations is given in Figure 1.

Nitrogen runoff from non-point sources such as agricultural practices is typically unobservable, especially at the scale of interest in this study. The *Water Quality and Erosion Productivity Impact Calculator* (EPIC-WQ, see Sharpley and Williams (1990)), a widely used (deterministic) biogeophysical process model, provides, at least conceptually, a convenient tool for predicting the nitrogen runoff at the NRI points, both for the current situation as well as for scenarios in which agricultural policies change one or several of the variables in Table 1. Unfortunately, running the model for all NRI points would be prohibitively computer-intensive even for establishing a baseline nitrogen runoff level, let alone for evaluating alternative scenarios in which a significant number of points might have changes in their covariate values. It was therefore decided to estimate a statistical “metamodel” on a representative subset of 11,403 data points, and use this metamodel in place of EPIC-WQ to predict nitrogen runoff at the remaining observation points, as well as for scenario evaluation. Another advantage of this approach is the estimation of coefficients and accompanying confidence intervals for the covariate effects, providing additional insights in the nature of the effect of agricultural practices (represented by *NRATE* and the dummy variables in Table 1) on nitrogen pollution.

The original approach of Wu *et al.* (1996) was to fit the metamodel by OLS after transforming the dependent variable and adding a limited number of interaction terms. The model was:

$$(Y_{N03})^{1/3.5} = \alpha + \mathbf{Z}_1\boldsymbol{\beta}_{z_1} + NRATE * \mathbf{Z}_2\boldsymbol{\beta}_{z_2} + \mathbf{X}\boldsymbol{\beta}_x + \text{iid errors}, \quad (1)$$

where \mathbf{Z}_1 contains the values for the covariates from Table 1 except the weather station location, \mathbf{Z}_2 the same as \mathbf{Z}_1 except for the removal of the covariate $NRATE$, and $\mathbf{X} = (LAT, LONG)$ the location of the nearest weather station. We will let $\mathbf{Z} = [\mathbf{Z}_1 \ NRATE * \mathbf{Z}_2 \ \mathbf{X}]$ and for simplicity refer to \mathbf{Z} as the covariates for this model, and let $\boldsymbol{\beta} = [\boldsymbol{\beta}_{z_1}^T \ \boldsymbol{\beta}_{z_2}^T \ \boldsymbol{\beta}_x^T]^T$. The location and interaction terms were included to improve the fit of the model, and the transformation was selected to remove some of the observed departures from the usual assumptions that the errors are homoscedastic and normally distributed. Nevertheless, the residuals still exhibited both severe heteroskedasticity, as well as spatial correlation. As noted in Carroll and Ruppert (1988), transformations of the dependent variable only remove heteroskedasticity when it depends on the mean. They are therefore not appropriate in cases where spatial location appears to cause most of the variance effects.

In the current paper, we demonstrate how a combination of universal kriging and nonparametric variance function estimation can be used to develop an improved regression model for this problem, while maintaining the interpretability of the mean function model (1). The choice of kriging is motivated by the fact that one of primary uses of this model is the prediction of $YN03$ at the large number of points not included in the regression observations, a situation for which kriging has well-known optimality properties (Cressie (1993)). Since the residuals of the OLS fit of model (1) exhibited significant heteroskedasticity as well, the explicit inclusion of a spatial variance function is expected to further improve the fit of both the mean and the correlation function. Because the specific shape of this function was not of particular interest to the CARD researchers, a generalization of the nonparametric variance estimation approach of Ruppert, Wand, Holst and Hössjer (1997) is used. This has the advantage of avoiding to introduce bias in the estimation due to inappropriate choice of a parametric form for the variance function.

Section 2 proposes a model that explicitly accounts for the heteroskedasticity and spatial correlation in the data, and Section 3 provides a description of the approach used in estimating its various components. In Section 4, the model estimates are discussed. Section 5 briefly discusses the use of universal kriging for predicting the nitrogen runoff values at the remaining NRI points

not included in the metamodel.

2 The Model

The data consist of n_i scalar response measurements Y_{ij} (the $YN03$ measurements from Section 1) and covariates \mathbf{Z}_{ij} recorded at N distinct geographic sites \mathbf{x}_i (the weather stations from Section 1), and the total number of observations is denoted by $n = \sum_{i=1}^N n_i$.

The model is

$$Y_{ij} = \mathbf{Z}_{ij}^T \boldsymbol{\beta} + v_\varepsilon(\mathbf{x}_i)^{1/2} \varepsilon_i + v_u(\mathbf{x}_i)^{1/2} u_{ij} \quad (2)$$

for $j = 1, \dots, n_i$, $i = 1, \dots, N$. Here $\boldsymbol{\beta}$ is a $q \times 1$ vector of parameters, v_ε and v_u are bivariate variance functions, the errors u_{ij} are independent and identically distributed with $E(u_{ij}) = 0$, $\text{var}(u_{ij}) = 1$ and the ε_i are such that $E(\varepsilon_i) = 0$, $\text{var}(\varepsilon_i) = 1$ and $\text{cov}(\varepsilon_i, \varepsilon_{i'}) = \rho(\|\mathbf{x}_i - \mathbf{x}_{i'}\|; \boldsymbol{\theta})$, where $\rho(\cdot; \boldsymbol{\theta})$ represents a parametric family of stationary, isotropic correlation functions indexed by the (possibly multi-dimensional) parameter $\boldsymbol{\theta}$. The $\{u_{ij}\}$ are independent of the $\{\varepsilon_i\}$.

This model is easily adapted to apply to other situations. The mean function $\mathbf{Z}_{ij}^T \boldsymbol{\beta}$ can be replaced by any other parametric model, including $\mathbf{Z}_{ij}^T \boldsymbol{\beta} \equiv \mu$ if ordinary kriging is used. Similarly, if there are no replicates at the geographic sites \mathbf{x}_i (i.e. $n_i = 1$ for all i), the term $v_u(\mathbf{x}_i)^{1/2} u_{ij}$ can be subsumed into $v_\varepsilon(\mathbf{x}_i)^{1/2} \varepsilon_i$.

As mentioned above, many points share the same “location” \mathbf{x}_i , with n_i ranging from 1 to 221 for the $N = 329$ weather stations in our dataset. There is also an important computational reason for working with these approximate locations instead of the actual point locations: only this reduction in the true dimension of the spatial variance-covariance matrix allows us to use “off-the-shelf” statistical packages to perform the estimation and prediction computations for this problem. The remaining errors u_{ij} at a given weather station location \mathbf{x}_i were assumed to be independent, since the correlation is taken to be spatial. In the kriging context, the variance function associated with the u_{ij} is referred to as the *nugget effect*. If no replicates are available, the nugget effect would have to be estimated directly from the spatial error process $v_\varepsilon(\mathbf{x}_i)^{1/2} \varepsilon_i$.

STEP 8: Repeat Steps 1–7 R_{iter} times.

Of course, the local linear smoothing in Steps 4 and 5 could be replaced by more higher degree local polynomial regression. After the estimation steps have been completed, predictions can be made as discussed in Section 5.

3.2 Details on the Implementation

3.2.1 Generalized Least Squares

In step 1, computations involving the inverse of the $11,403 \times 11,403$ matrix $\hat{\Sigma} = \text{cov}(\mathbf{Y})$ are avoided by noting that, because of the assumed model (2),

$$\Sigma = \Sigma_u + \mathbf{K}^T \mathbf{V}_\varepsilon \mathbf{K},$$

where Σ_u is a diagonal matrix with repeating “blocks” of length n_i :

$$\Sigma_u = \text{diag}\{v_u(\mathbf{x}_i), j = 1, \dots, n_i, i = 1, \dots, N\},$$

\mathbf{V}_ε is the $N \times N$ covariance matrix of the ε_i and \mathbf{K} is an $N \times n$ matrix with (i, i') entry equal to 1 for $i' = 1 + \sum_{k=1}^{i-1} n_k, \dots, \sum_{k=1}^i n_k$ and zero otherwise. The inverse of Σ is therefore equal to

$$\Sigma^{-1} = \Sigma_u^{-1} - \Sigma_u^{-1} \mathbf{K}^T (\mathbf{V}_\varepsilon^{-1} + \mathbf{K} \Sigma_u^{-1} \mathbf{K}^T)^{-1} \mathbf{K} \Sigma_u^{-1}.$$

(Horn and Johnson (1985)), which can be rapidly computed since the largest non-diagonal matrix to invert is only $N \times N$.

3.2.2 Variance Function Estimation by Local Polynomial Regression

If we assume normality of the errors, the $\hat{v}_u(\mathbf{x}_i)$ in Step 4 are independently distributed, heteroskedastic random variables, with variance equal to $2v_u(\mathbf{x}_i)^2/(n_i - 1)$, so that the theory developed by Ruppert *et al.* (1997) is directly applicable here. While $p = 2$ observations are sufficient for computing $\hat{v}_u(\mathbf{x}_i)$ at a location, there is clearly much more information about v_u at the locations with more observations. Since $n = 11,403$ and $N = 329$, we have $\bar{n} = 35$ and it might make sense to use only locations where n_i is “not too small.” We experimented with $p = 2, 3, 4$ and found that $p = 3$ gave the best estimates, in terms of speed of convergence of the algorithm.

lack of boundary problems as well as avoiding negative variance estimates (see below). The number of locations where with $n_i \geq 3$ is 290. The special structure between the estimator and its variance is used in the bandwidth selection of the EBBS algorithm (Ruppert (1997)). More specifically, let $\hat{v}_u(\mathbf{x}_i; h)$ be the local linear estimator of $v_u(\mathbf{x}_i)$. EBBS separately estimates the squared bias and variance of $\hat{v}_u(\mathbf{x}_i)$. These quantities are added together and their sum is minimized over a grid of h -values to produce the EBBS bandwidth at \mathbf{x}_i . The bias estimate is exactly as in Ruppert (1997). The estimate of $\text{var}(\hat{v}_u(\mathbf{x}_i; h))$ uses the relation

$$\text{var}(\hat{v}_u(\mathbf{x}_i; h)) = \mathbf{s}(\mathbf{x}_i; h)^T \text{diag}\{\text{var}(\hat{v}_u(\mathbf{x}_i))\} \mathbf{s}(\mathbf{x}_i; h),$$

where $\mathbf{s}(\mathbf{x}_i; h)$ is the N by 1 local polynomial ‘‘smoother vector’’ such that $\hat{v}_u(\mathbf{x}_i; h) = (\hat{v}_u(\mathbf{x}_1), \dots, \hat{v}_u(\mathbf{x}_N)) \mathbf{s}(\mathbf{x}_i; h)$. EBBS estimates $\text{var}(\hat{v}_u(\mathbf{x}_i; h))$ by

$$\widehat{\text{var}}(\hat{v}_u(\mathbf{x}_i; h)) = \mathbf{s}(\mathbf{x}_i; h)^T \text{diag}\{2(\hat{v}_u(\mathbf{x}_i; h))/(n_i - 1)\} \mathbf{s}(\mathbf{x}_i; h).$$

In Step 5, we obtain \hat{v}_ε by smoothing $\{\hat{v}_\varepsilon(\mathbf{x}_i) : n_i > 1\}$ again using EBBS to select the bandwidth. We will ignore the error in $\hat{\beta}$ so that

$$r_{ij} = v_\varepsilon(\mathbf{x}_i)^{1/2} \epsilon_i + v_u(\mathbf{x}_i)^{1/2} u_{ij},$$

and therefore

$$\bar{r}_i = v_\varepsilon(\mathbf{x}_i)^{1/2} \bar{\epsilon}_i + v_u(\mathbf{x}_i)^{1/2} \bar{u}_i.$$

Then $\hat{v}_u(\mathbf{x}_i)$ is unbiased for $v_u(\mathbf{x}_i)$, and therefore

$$E(\hat{v}_\varepsilon(\mathbf{x}_i)) = E(\bar{r}_i)^2 - \frac{v_u(\mathbf{x}_i)}{n_i} = v_\varepsilon(\mathbf{x}_i).$$

Therefore, when we smooth the $\{\hat{v}_\varepsilon(\mathbf{x}_i)\}$ there is no bias term involving v_u and EBBS will properly estimate the bias of our final estimate of v_ε . One might consider estimating v_ε by smoothing the $\{\bar{r}_i\}$ and then subtracting off an estimate of $v_u(\mathbf{x}_i)/n_i$; however, in this case, the bandwidth optimal for smoothing the $\{\bar{r}_i\}$ will not be optimal for the final estimate of v_ε . The EBBS bandwidth for smoothing $\hat{v}_\varepsilon(\mathbf{x}_i)$ requires an estimate of $\text{var}(\hat{v}_\varepsilon(\mathbf{x}_i; h))$. Estimation of this variance is based upon the following results.

Let $H = \mathbf{Z}(\mathbf{Z}^T \boldsymbol{\Sigma}^{-1} \mathbf{Z})^{-1} \mathbf{Z}^T \boldsymbol{\Sigma}^{-1}$ represent the ‘‘hat’’ matrix from the estimation of the mean. Let $\boldsymbol{\kappa}$ be the $N \times n$ matrix with (i, j) entry equal

$$\mathbf{s}^T(\mathbf{x}; h) \Sigma_{t^\varepsilon}(\mathbf{x}; h) \mathbf{s}$$

where, as with $\hat{r}_n(\mathbf{x}; h)$ is a smoother vector. Then, the estimate of $\text{var}(\hat{r}_n(\mathbf{x}; h))$ used by EBBS is

$$\hat{r}_\varepsilon = (r^\varepsilon(\mathbf{x}_1; h), \dots, r^\varepsilon(\mathbf{x}_n; h))$$

by $\hat{\mathbf{V}}^\varepsilon$ and $\hat{\mathbf{V}}^n$. Suppose that $r^\varepsilon(\mathbf{x}_t)$ by \hat{r}_t and r_t . Then let Σ_{t^ε} be given by (5) with \mathbf{V}^ε and \mathbf{V}^n replaced by $\hat{\mathbf{V}}^\varepsilon$ and $\hat{\mathbf{V}}^n$ be obtained from \mathbf{V}^ε and \mathbf{V}^n by replacing $r_n(\mathbf{x}_t)$ and where $\mathbf{E}_2 = \text{diag}(m_t - 1)^{-1}$.

$$(5) \quad \Sigma_{t^\varepsilon} = 2 \left\{ (\mathbf{V}^\varepsilon + \mathbf{V}^n \mathbf{E}_1)_{[2]} + \mathbf{V}_{[2]}^n \mathbf{E}_{[2]} \mathbf{E}_2 \right\}$$

is given by

$$r^\varepsilon(\mathbf{x}_t) = (r_t)_{[2]} - \frac{r_n(\mathbf{x}_t)}{m_t} \quad t = 1, \dots, N$$

containing

Y_{tj}^ε 's are normally distributed, the covariance matrix of the random vector $\mathbf{Z}\boldsymbol{\beta}$ is ignored due to estimation of the mean $\mathbf{Z}\boldsymbol{\beta}$ is ignored and the

Result 2: If the error due to estimation of the mean $\mathbf{Z}\boldsymbol{\beta}$ is ignored and the

and $\mathbf{A}_{[2]} = \mathbf{A} \cdot \mathbf{A}$ where $\mathbf{V}^n = \text{diag}\{r_n(\mathbf{x}_t), t = 1, \dots, N\}$ and $\mathbf{E}_1 = \text{diag}\{1/m_t, t = 1, \dots, N\}$

$$(4) \quad 2(\mathbf{V}^\varepsilon + \mathbf{V}^n \mathbf{E}_1)_{[2]}$$

$\mathbf{Z}\boldsymbol{\beta}$ is ignored, then the above covariance matrix simplifies to

where $\mathbf{m} = \mathbf{Z}\boldsymbol{\beta}$, $\mathbf{A} = \boldsymbol{\kappa}(\mathbf{I} - \mathbf{H})$. If the error due to estimation of the mean

$$(3) \quad 2\mathbf{A}\Sigma\mathbf{A}^T \cdot (\mathbf{A}\Sigma\mathbf{A}^T + 2\mathbf{A}\mathbf{m}\mathbf{m}^T\mathbf{A}^T)$$

is given by

$$r_t^2, \quad t = 1, \dots, N$$

random vector containing

Result 1: Assuming normality of the Y_{tj}^ε 's, the covariance matrix of the

and \mathbf{B} .

Finally, let $\mathbf{A} \cdot \mathbf{B}$ denote the elementwise product of equi-sized matrices \mathbf{A}

$$\boldsymbol{\kappa} = \begin{bmatrix} \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & 0 & 0 & 0 & 0 \\ \frac{3}{4} & \frac{3}{4} & \frac{1}{4} & \frac{1}{4} & 0 & 0 & 0 & 0 \\ \frac{3}{4} & \frac{1}{4} & \frac{3}{4} & \frac{1}{4} & 0 & 0 & 0 & 0 \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & 0 & 0 & 0 & 0 \end{bmatrix}$$

in the case $N = 3$, $m_1 = 3$, $m_2 = 1$, $m_3 = 4$ we have:

to $1/m_t$ for $j = 1 + \sum_{k=1}^{t-1} n_k, \dots, \sum_{k=1}^t n_k$ and zero otherwise. For example,

Our estimate of v_ε does not use locations where $n_i < p$, but these are locations where there is relatively little information about v_ε . Since the $\hat{v}_\varepsilon(\mathbf{x}_i)$ smoothed in Step 5 are possibly negative, there is a positive probability that $\hat{v}_\varepsilon(\mathbf{x}_i)$ is negative. As n_i increases, the probability that $\hat{v}_\varepsilon(\mathbf{x}_i)$ is negative decreases. While negative values for $\hat{v}_\varepsilon(\mathbf{x}_i)$ are in principle not a problem, it is highly undesirable to have negative variance estimates \hat{v}_ε , since they would result in a negative definite covariance matrix $\hat{\mathbf{V}}_\varepsilon$. For $p = 3$, only 9 locations had negative variance estimates, and all were located at the boundary of the estimation region (see Figure 2), making it likely that they are the result of “boundary effects,” a common nuisance in nonparametric regression similar to extrapolation problems in parametric regression. We therefore decided to add a local averaging step at each iteration of the algorithm to “correct” any negative estimates. Note that this step only changes the negative estimates and leaves all the other ones unchanged.

3.2.3 Estimation of the Correlation Function by Variogram Fitting

In Step 6, the correlation function is estimated parametrically by variogram fitting. Because heteroskedasticity is known to cause spurious patterns in variograms, it is important to remove that effect before estimating the correlation function. Hence, the spatial residuals \bar{r}_i have to be normalized. It is not a priori clear what the normalizing constants should be, however. If we ignore the errors caused by the estimation of the mean and variance functions, and use $\hat{\varepsilon}_i^* = \bar{r}_i / v_\varepsilon(\mathbf{x})^{1/2}$, the variogram will estimate

$$\begin{aligned} 2\hat{\gamma}(\mathbf{x}_i - \mathbf{x}_{i'}) &= \text{E}(\hat{\varepsilon}_i^* - \hat{\varepsilon}_{i'}^*)^2 = \text{var}(\hat{\varepsilon}_i^*) + \text{var}(\hat{\varepsilon}_{i'}^*) - 2\text{cov}(\hat{\varepsilon}_i^*, \hat{\varepsilon}_{i'}^*) \\ &= \frac{v_r(\mathbf{x}_i)}{v_\varepsilon(\mathbf{x}_i)} + \frac{v_r(\mathbf{x}_{i'})}{v_\varepsilon(\mathbf{x}_{i'})} - 2\rho(\mathbf{x}_i - \mathbf{x}_{i'}) \end{aligned}$$

while if we use $\hat{\varepsilon}_i^* = \bar{r}_i / v_r(\mathbf{x}_i)^{1/2}$, then

$$\begin{aligned} 2\hat{\gamma}'(\mathbf{x}_i - \mathbf{x}_{i'}) &= \text{E}(\hat{\varepsilon}_i^* - \hat{\varepsilon}_{i'}^*)^2 \\ &= 2 - 2\rho(\mathbf{x}_i - \mathbf{x}_{i'}) \sqrt{\frac{v_r(\mathbf{x}_i) v_r(\mathbf{x}_{i'})}{v_\varepsilon(\mathbf{x}_i) v_\varepsilon(\mathbf{x}_{i'})}}. \end{aligned}$$

Neither $\tilde{\gamma}(\cdot)$ nor $\tilde{\gamma}'(\cdot)$ are in general equal to $\gamma(\cdot) = 1 - \rho(\cdot)$, so they cannot be directly used to fit the correlation function. However, it is easy to see that

$$\gamma(\mathbf{x}_i - \mathbf{x}_{i'}) = 1 - \frac{1 - \tilde{\gamma}'(\mathbf{x}_i - \mathbf{x}_{i'})}{\sqrt{\frac{v_r(\mathbf{x}_i) v_r(\mathbf{x}_{i'})}{v_\varepsilon(\mathbf{x}_i) v_\varepsilon(\mathbf{x}_{i'})}}}. \quad (6)$$

We can therefore construct a “bias-corrected” variogram based on (6). Let $\hat{\varepsilon}_i = \bar{v}_i/v_r(\mathbf{x}_i)^{1/2}$. For a given distance t , let $S(t) = \{(i, i') : \|\mathbf{x}_i - \mathbf{x}_{i'}\| \in (t \pm \delta)\}$ with δ a given bin size and $n(t) = |S(t)|$. Then,

$$\hat{\gamma}(t) = 1 - \frac{1 - \frac{1}{2n(t)} \sum_{S(t)} (\hat{\varepsilon}_i - \hat{\varepsilon}_{i'})^2}{\frac{1}{n(t)} \sum_{S(t)} \sqrt{\frac{\hat{v}_r(\mathbf{x}_i) \hat{v}_r(\mathbf{x}_{i'})}{\hat{v}_\varepsilon(\mathbf{x}_i) \hat{v}_\varepsilon(\mathbf{x}_{i'})}}}.$$

The following parametric model is used for $\rho(\cdot)$:

$$\rho(t; \boldsymbol{\theta}) = 1 - \theta_3 e^{-\theta_1 t} - (1 - \theta_3) e^{-\theta_2 t}$$

with $\theta_1, \theta_2 > 0$ and $0 \leq \theta_3 \leq 1$. This is a mixture of two exponential functions, which was chosen to guarantee the positive definiteness of the variance/covariance matrix estimate. Clearly, other parametric models, including mixtures of larger numbers of exponentials, could be selected as correlation functions for other datasets. The parameters $\theta_1, \theta_2, \theta_3$ are estimated by weighted least squares minimization following Cressie (1993, p.96).

The estimate of the spatial variance covariance matrix \mathbf{V}_ε is computed by setting

$$[\hat{\mathbf{V}}_\varepsilon]_{ii'} = \rho(\mathbf{x}_i - \mathbf{x}_{i'}; \hat{\boldsymbol{\theta}}) \sqrt{\hat{v}_\varepsilon(\mathbf{x}_i) \hat{v}_\varepsilon(\mathbf{x}_{i'})}.$$

4 Results

The model was run on the CARD dataset, using both the transformed and untransformed EPIC-WQ predicted nitrogen runoff values as dependent variables. The transformation is no longer necessary to reduce the heteroskedasticity. This is not too surprising, since the heteroskedasticity was now explicitly accounted for in the model itself. A methodological advantage of the untransformed model is that the predictions computed as in Section 5 are unbiased, while the ones found by using transformed runoff observations are biased after inverting the transformation. We will therefore only report the results for the untransformed model.

The model converges in 2-10 iterations, depending on the strictness of the convergence criterion and the choice of some of the tuning parameters. For $p = 3$ the model converges after 4 iterations, which takes approximately 10 minutes to run on a DEC 3000 Model 900 AXP workstation, with the bulk of the computing time taken by the GLS fitting (Step 1 in Subsection 3.1).

Figures 3 and 4 show the nonparametric estimates of the variance functions $v_u(\cdot)$ and $v_\varepsilon(\cdot)$ at the weather station locations. Both estimates show a pattern of low values in the center. The patterns also show some interesting differences: while the Great Lakes region exhibits high local and spatial variance, the spatial variance is also high in the South-most part of the study region, while the local variance is high at the Western edge. Most of the variability in the data is explained by the local variance v_u , with the mean value of $\hat{v}_u(\mathbf{x}_i)$ equal to 4.789, while that for $\hat{v}_\varepsilon(\mathbf{x}_i)$ is 0.404.

In Figure 5, the bias-adjusted variogram of the standardized residuals $\hat{\varepsilon}_i$ is displayed as well as the weighted least squares fitted variogram function. The spatial correlation decreases rapidly as distance increases, and is only important for points at short distances of each other.

5 Model Predictions

As mentioned in Section 1, the purpose for developing this metamodel is to be able to predict the potential nitrogen runoff at the set of 128,591 NRI points. If we let \mathbf{Z}^* and \mathbf{x}^* represent the matrices of covariates and locations at any n^* prediction points, the universal kriging prediction equation for the points $\hat{\mathbf{Y}}^* = (\hat{Y}_1^*, \dots, \hat{Y}_{n^*}^*)^T$ is given by

$$\hat{\mathbf{Y}}^* = \mathbf{Z}^* \hat{\beta} + \mathbf{C}_\varepsilon \Sigma^{-1} (\mathbf{Y} - \mathbf{Z} \hat{\beta}), \quad (7)$$

where \mathbf{C}_ε is a $n^* \times n$ matrix with elements $[\mathbf{C}_\varepsilon]_{i'i'} = \rho(x_i^* - x_{i'}, \boldsymbol{\theta}) \sqrt{\hat{v}_\varepsilon(x_i^*) \hat{v}_\varepsilon(x_{i'})}$ (Cressie (1993 p.173)). Since the true values for the variance and covariance

functions are unknown, we replace them by their estimators obtained by the procedure described in Section 3.

An alternative approach for prediction uses the fact that the prediction and estimation data are at the same set of weather station locations, so that the spatial residuals ε_i can be considered a *lattice* process (Cressie (1993)). The vector of spatial errors $\boldsymbol{\varepsilon} = (\varepsilon_1, \dots, \varepsilon_N)^T$ can therefore be predicted by a “shrunk” version of the spatial residuals r_i :

$$\hat{\boldsymbol{\varepsilon}} = \hat{\mathbf{V}}_{\varepsilon} \left(\hat{\mathbf{V}}_{\varepsilon} + \hat{\mathbf{V}}_u \mathbf{E}_1 \right)^{-1} \mathbf{r}, \quad (8)$$

with $\mathbf{r} = (r_1, \dots, r_N)^T$, by a straightforward application of conditional expectations (Fuller (1987)). Hence the spatial “correction” $[\hat{\mathbf{C}} \hat{\boldsymbol{\Sigma}}^{-1} (\mathbf{Y} - \mathbf{Z} \hat{\boldsymbol{\beta}})]_i$ at a weather station location \mathbf{x}_i^* can be predicted directly by the corresponding element of the vector $\hat{\boldsymbol{\varepsilon}}$. This approach is computationally much more efficient than the “full” universal kriging approach described above, because the size of the matrices in (7) increases with n^* , while those in (8) only depend on N . Figure 6 shows a plot of the values of the spatial corrections $\hat{\varepsilon}_i$.

A Proofs

Proof of Result 1: The vector of residuals, r_{ij} , can be written as $\mathbf{r} = (\mathbf{I} - \mathbf{H})\mathbf{Y}$. From the definition of $\boldsymbol{\kappa}$ we have the vector of \bar{r}_i values equalling $\boldsymbol{\kappa}(\mathbf{I} - \mathbf{H})\mathbf{Y}$. The stated result in (3) then follows directly from Lemma 1 of Ruppert *et al.* (1997) for the special case of normal Y_i .

If the bias due to estimation of $\mathbf{Z}\boldsymbol{\beta}$ is ignored, then this expression simplifies to

$$2(\boldsymbol{\kappa}\boldsymbol{\Sigma}\boldsymbol{\kappa}^T)^{[2]}.$$

Expression (4) follows directly after noting that we are indexing the matrix $\boldsymbol{\Sigma}$ as follows:

$$\boldsymbol{\Sigma}_{ij,i'j'} = \text{cov}(Y_{ij}, Y_{i'j'})$$

for $j = 1, \dots, n_i, i = 1, \dots, N$. Hence,

$$(\boldsymbol{\kappa}\boldsymbol{\Sigma}\boldsymbol{\kappa}^T)_{i'i'} = (n_i n_{i'})^{-1} \sum_{j=1}^{n_i} \sum_{j'=1}^{n_{i'}} \boldsymbol{\Sigma}_{ij,i'j'}$$

so that

$$(\boldsymbol{\kappa}\boldsymbol{\Sigma}\boldsymbol{\kappa}^T)_{ii'} = \begin{cases} \frac{v_u(\mathbf{x}_i)}{n_i} + v_\varepsilon(\mathbf{x}_i) & i = i' \\ v_\varepsilon(\mathbf{x}_i)^{1/2}v_\varepsilon(\mathbf{x}_{i'})^{1/2}\rho(\|\mathbf{x}_i - \mathbf{x}_{i'}\|; \theta) & i \neq i' \end{cases}$$

■

Proof of Result 2: Recall that

$$\tilde{v}_u(\mathbf{x}_i) = \frac{1}{n_i - 1} \sum_{j=1}^{n_i} (r_{ij} - \bar{r}_i)^2 = \sum_{j=1}^{n_i} (u_{ij} - \bar{u}_i)^2.$$

Because the $\{u_{ij}\}$ are iid normals, \bar{u}_i is independent of $\tilde{v}_u(\mathbf{x}_i)$. Therefore, \bar{r}_i is independent of $\tilde{v}_u(\mathbf{x}_i)$. Let $\boldsymbol{\Sigma}_r$ be the covariance matrix of the vector $((\bar{r}_1)^2, \dots, (\bar{r}_N)^2)^T$, $\boldsymbol{\Sigma}_{\tilde{v}_\varepsilon}$ the covariance matrix of $(\tilde{v}_\varepsilon(\mathbf{x}_1), \dots, \tilde{v}_\varepsilon(\mathbf{x}_N))^T$, and $\boldsymbol{\Sigma}_{\tilde{v}_u}$ the covariance matrix of $(\tilde{v}_u(\mathbf{x}_1), \dots, \tilde{v}_u(\mathbf{x}_N))^T$. Since $\tilde{v}_u(\mathbf{x}_i)$ is $v_u(\mathbf{x}_i)/(n_i - 1)$ times at $\chi^2(n_i - 1)$ random variable, we have

$$\boldsymbol{\Sigma}_{\tilde{v}_u} = \text{diag} \left(\frac{2v_u^2(\mathbf{x}_i)}{n_i - 1} \right) = 2\mathbf{V}_u^{[2]}\mathbf{E}_2.$$

By (4), and ignoring the error caused by using $\hat{\boldsymbol{\beta}}$ in place of $\boldsymbol{\beta}$, we have

$$\boldsymbol{\Sigma}_{\tilde{v}_\varepsilon} = 2(\mathbf{V}_r + \mathbf{V}_u\mathbf{E}_1)^{[2]} + \boldsymbol{\Sigma}_{\tilde{v}_u}\mathbf{E}_1^{[2]} = 2\left\{(\mathbf{V}_r + \mathbf{V}_u\mathbf{E}_1)^{[2]} + \mathbf{V}_u^{[2]}\mathbf{E}_1^{[2]}\mathbf{E}_2\right\}. \quad (9)$$

■

References

- [1] R.J. Carroll and D. Ruppert. *Transformation and Weighting in Regression*. Chapman and Hall, New York, 1988.
- [2] N. A. C. Cressie. *Statistics for Spatial Data*. John Wiley & Sons, New York, 2 edition, 1993.
- [3] W. A. Fuller. *Measurement Error Models*. John Wiley & Sons, New York, NY, 1987.
- [4] J. D. Hart. Kernel regression estimation with time series errors. *Journal of the Royal Statistical Association, Series B*, 53:173–187, 1991.

- [5] R. A. Horn and C. A. Johnson. *Matrix Analysis*. Cambridge University Press, Cambridge, U.K., 1985.
- [6] S.M. Nusser and J.J. Goebel. The national resources inventory: a long-term multi-resource monitoring programme. *Environmental and Ecological Statistics*, 4:181-204, 1997.
- [7] J.-D. Opsomer. Nonparametric regression in the presence of correlated errors. In *Proceedings of the Conference on Modelling Longitudinal and Spatially Correlated Data, Nantucket, October 15-18, 1996*. Springer-Verlag, 1996. Submitted.
- [8] D. Ruppert. Empirical-bias bandwidths for local polynomial nonparametric regression and density estimation. *Journal of the American Statistical Association*, 92:1049-1062, 1997.
- [9] D. Ruppert, M.P. Wand, U. Holst, and Ola Hössjer. Local polynomial variance function estimation. *Technometrics*, 39:262-273, 1997.
- [10] A.N. Sharpley and eds. J.R. Williams. Epic-erosion/productivity impact calculator: 1. model documentation. Tech. Bull. 1768. USDA, Washington, DC, 1990.
- [11] J. Wu, P.G. Lakshminarayan, and B.A. Babcock. Impacts of agricultural practices and policies on potential nitrate water pollution in the Midwest and the Northern Plains of the United States. Working Paper 96 WP 148, Center for Agricultural and Rural Development, Iowa State University, February 1996.

Table 1: Model variables.

Tillage, conservation and irrigation practice dummies (reference: conventional tillage):		Hydrology dummies (reference: DHYGA):		Location of closest weather station:	
YNO3	nitrogen runoff (predicted by EPIC-WQ)	NRATE	nitrogen application rate	LAT	latitude
				LONG	longitude
DRT		reduced tillage		DHYCB	hydrologic group B
DNT		no till		DHYCC	hydrologic group C
DCONTR		contouring			
DSTRIP		strip-cropping			
DTERRA		terracing			
DIRTYP		irrigation			
DRO11		continuous corn			
DRO12		continuous soybeans			
DRO13		continuous wheat			
DRO14		continuous sorghum			
DRO15		corn-soybeans			
DRO16		corn-corn-soybeans			
DRO17		corn-soybeans-wheat			
DRO18		continuous corn			
DRO19		continuous soybeans			
DRO20		continuous wheat			
DRO21		continuous sorghum			
DRO22		corn-soybeans			
DRO23		corn-corn-soybeans			
DRO24		corn-soybeans-wheat			
DRO25		corn-corn-3 years alfalfa			
DRO26		wheat-sorghum			
DRO27		wheat-soybeans			
DRO28		wheat-sorghum-fallow			
DRO29		wheat-fallow			
DRO30		soybeans-soybeans-corn			
BD		bulk density			
PH		soil pH			
PERM		soil permeability			
AWC		available water capacity			
DHYGD		hydrologic group D			

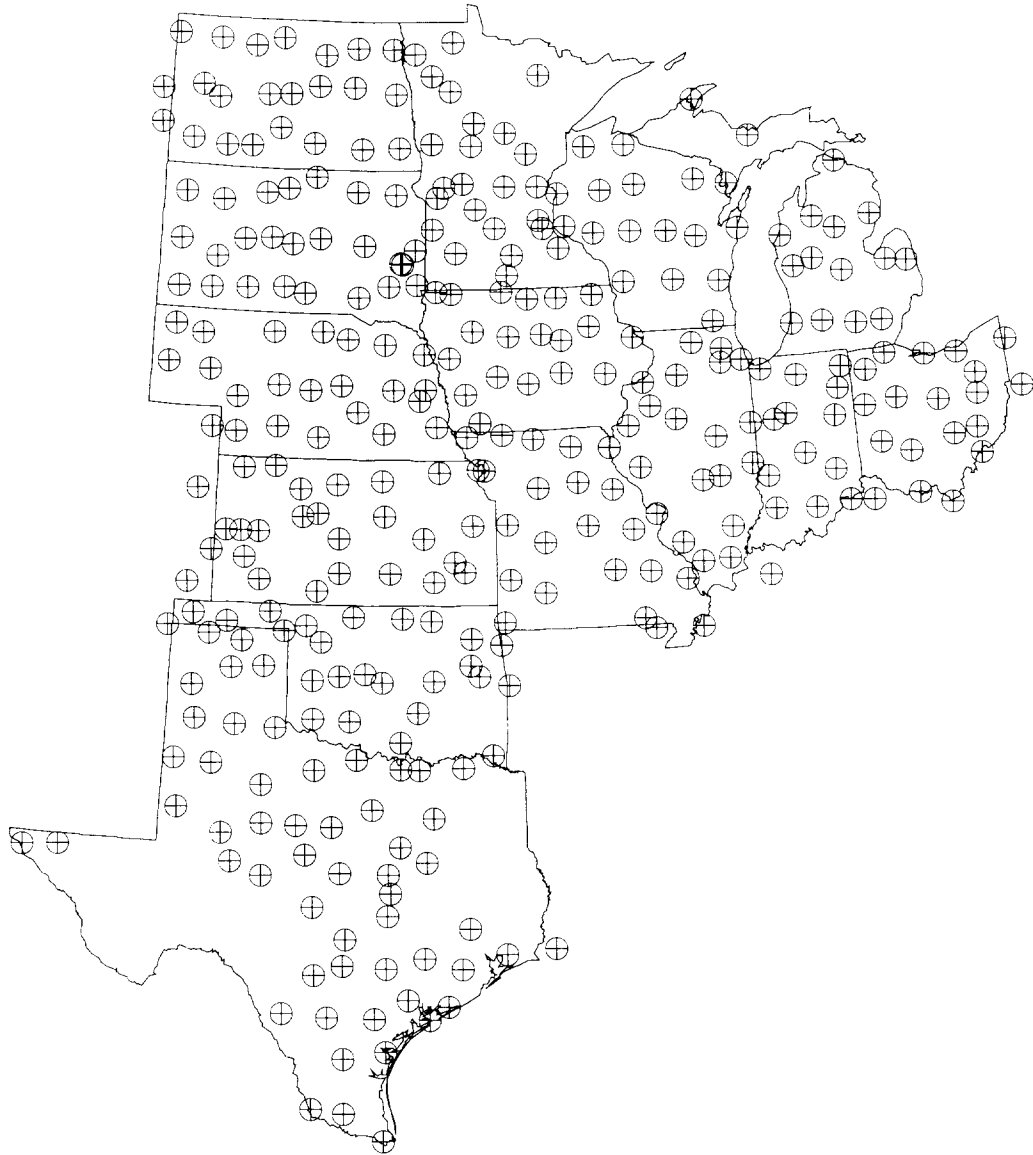


Figure 1: Map of the study region (+ denotes weatherstation location).

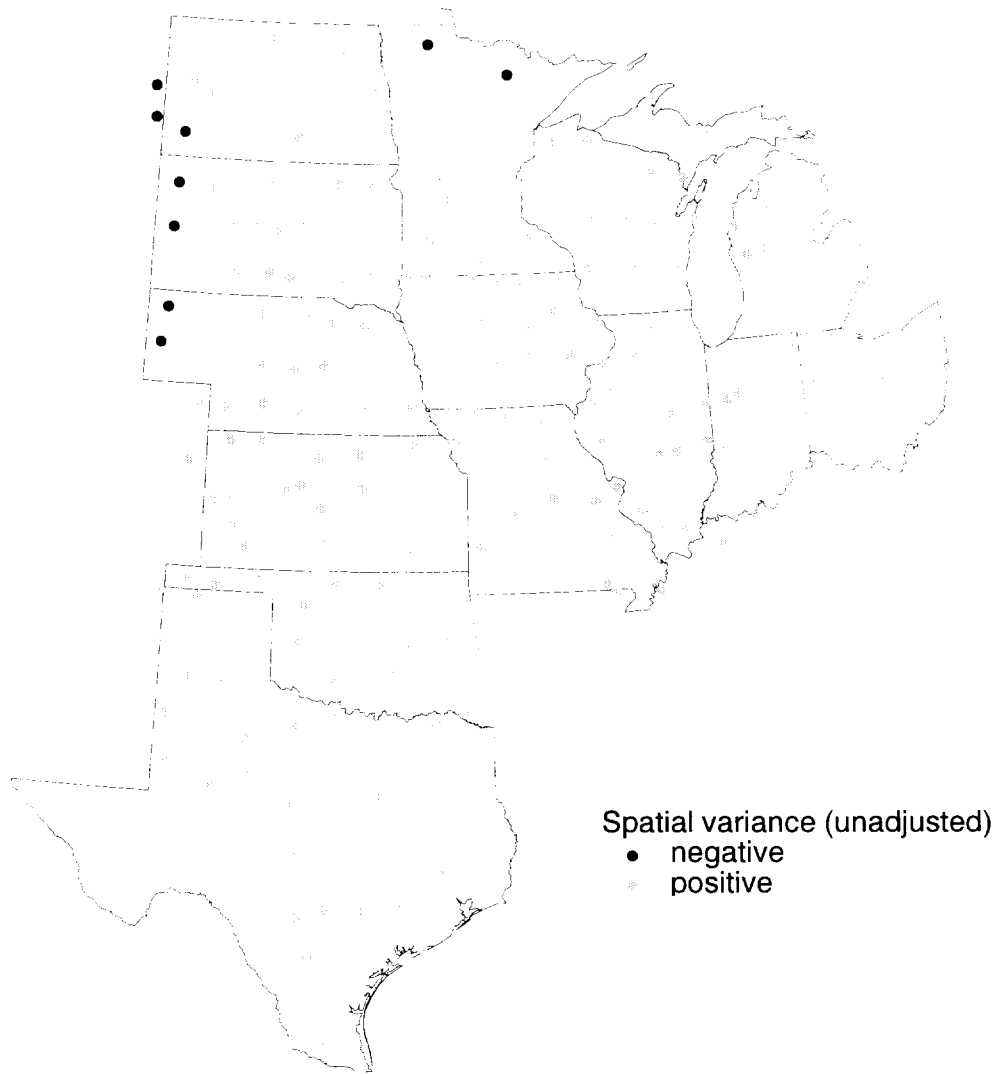


Figure 2: Locations of negative and positive estimates of $v_\varepsilon(\cdot)$ before local averaging adjustment.

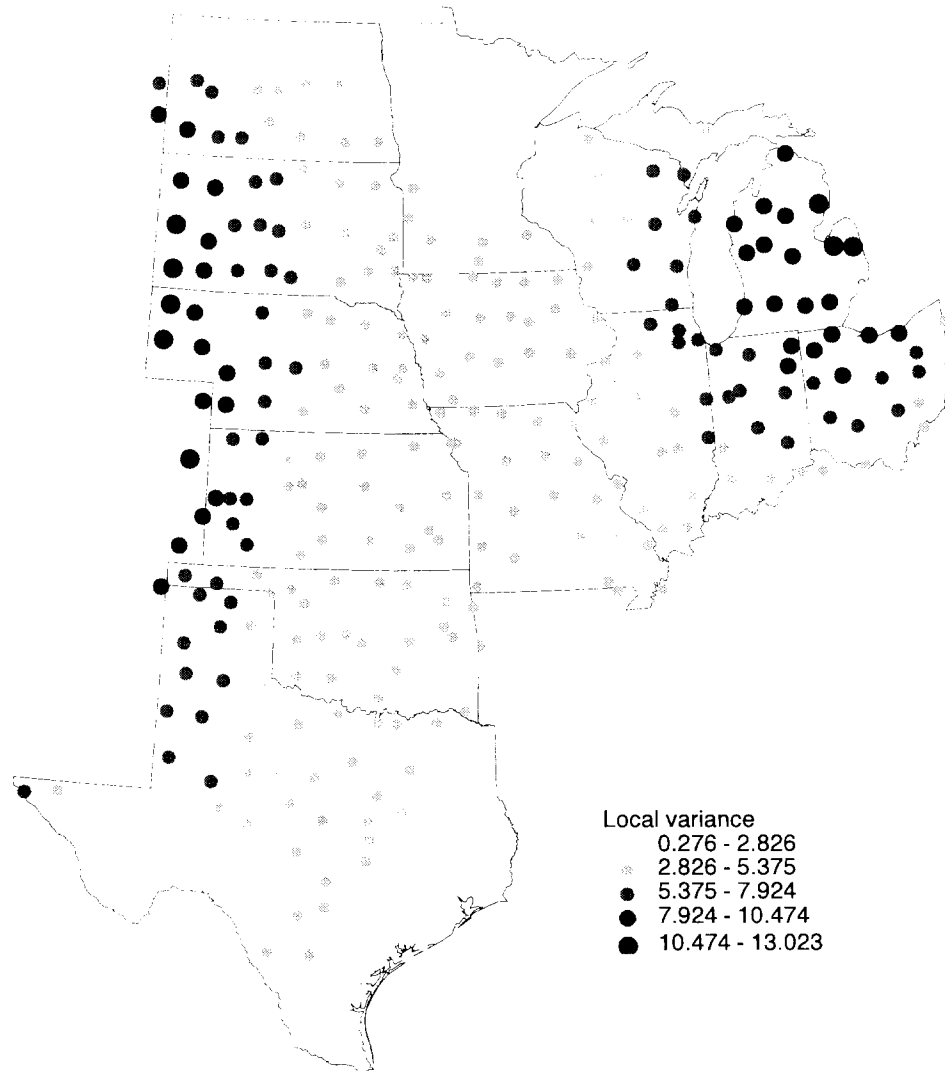


Figure 3: Estimate of the variance function $v_u(\cdot)$ at the weatherstation locations.

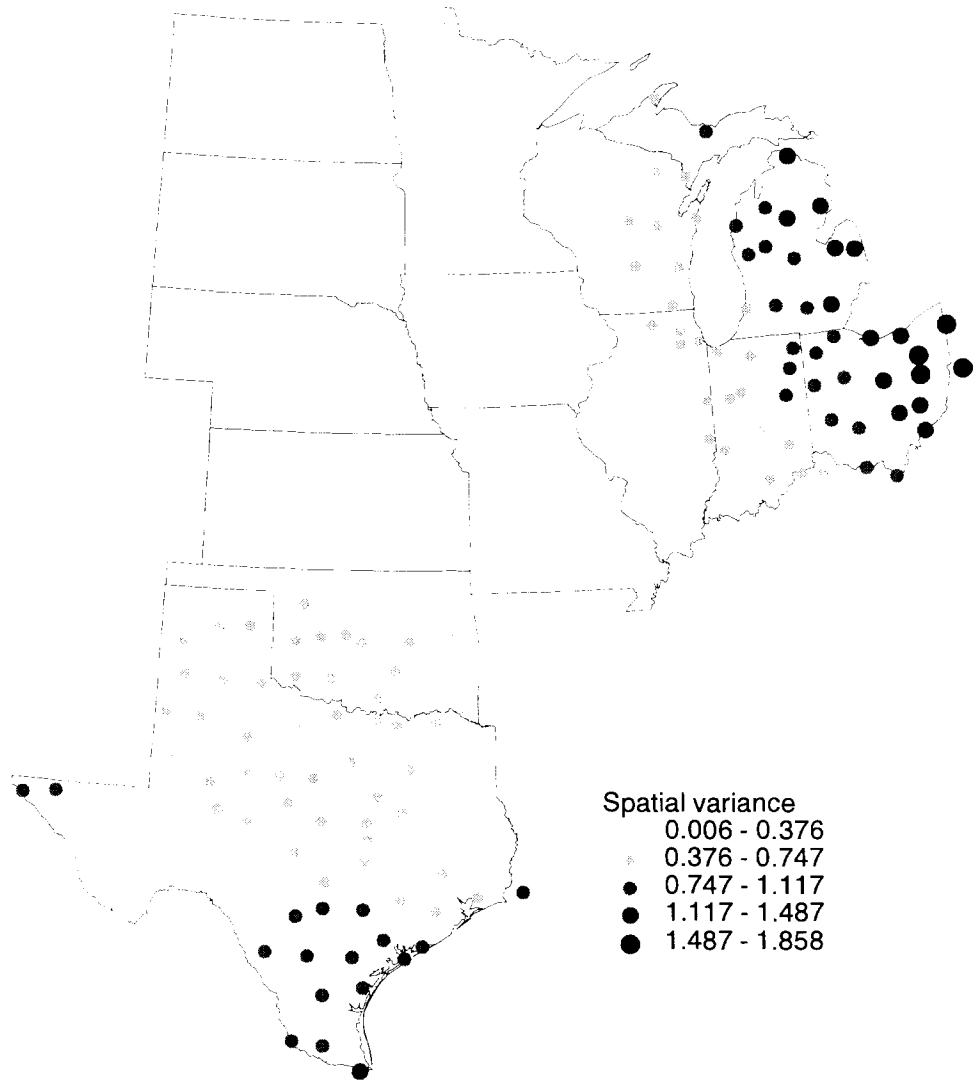


Figure 4: Estimate of the variance function $v_{\varepsilon}(\cdot)$ at the weatherstation locations.

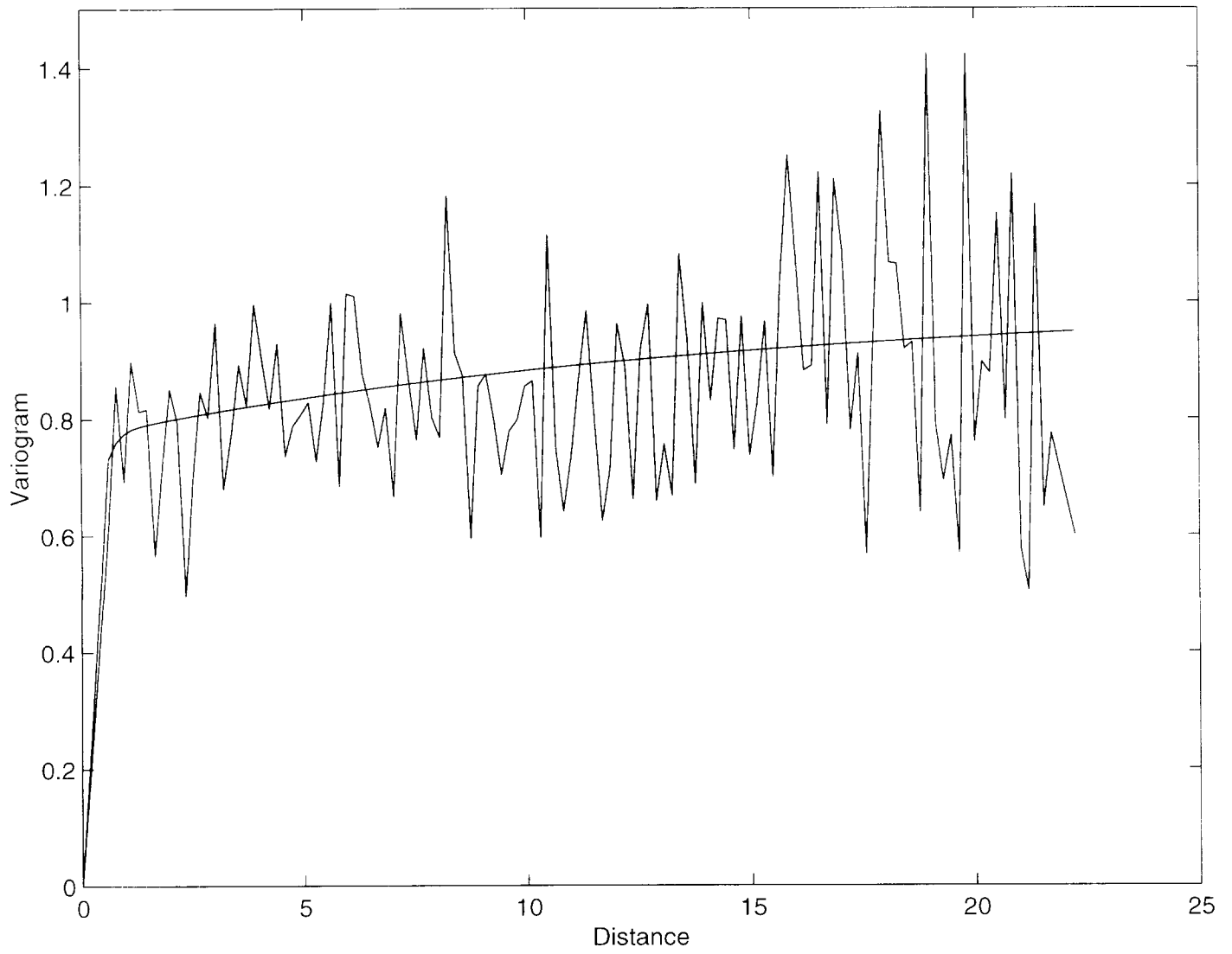


Figure 5: Variogram and estimated variogram function $\hat{\rho}(\cdot; \hat{\theta})$.

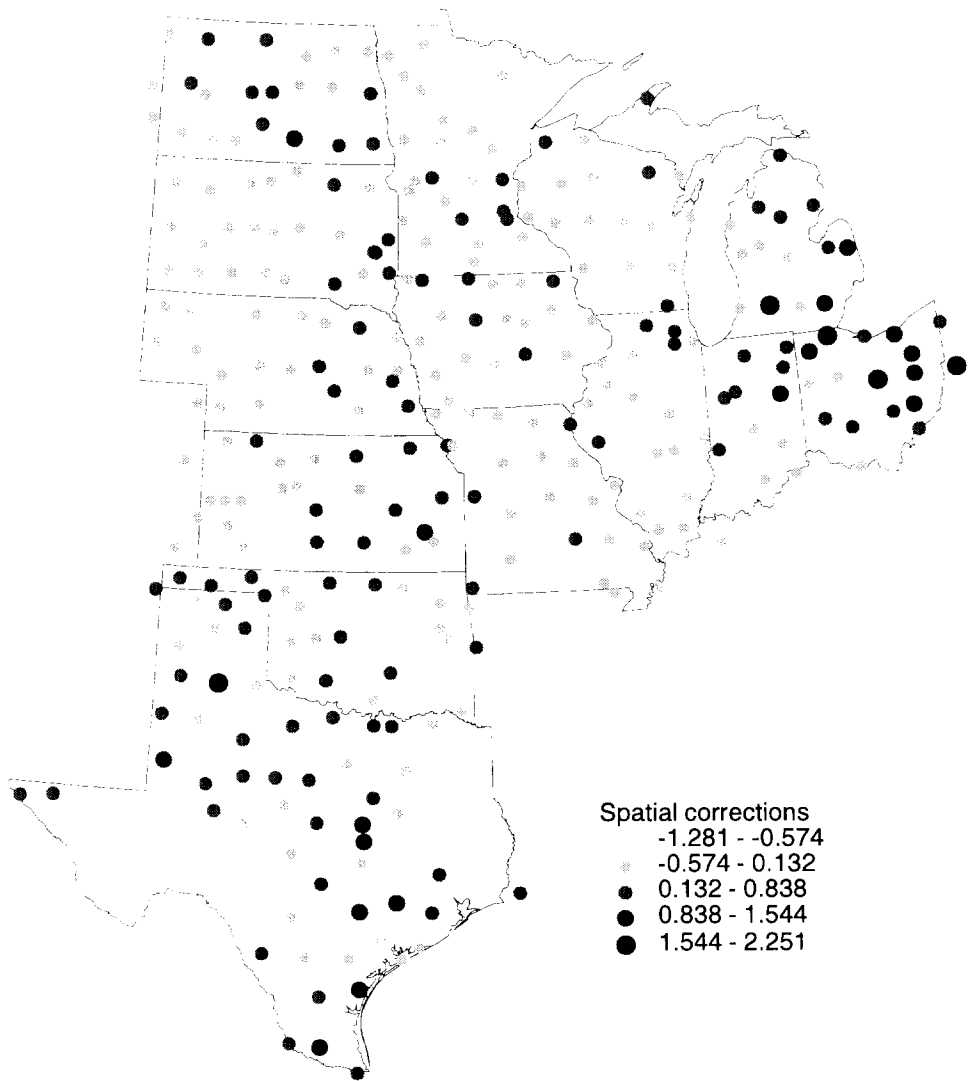


Figure 6: Spatial corrections $\hat{\varepsilon}_i$ at the weatherstation locations.