SPATIAL COVARIANCE ESTIMATION
FOR MONITORING DATA

CLIVE LOADER
PAUL SWITZER

TECHNICAL REPORT NO. 133
NOVEMBER 1989

SIAM INSTITUTE FOR MATHEMATICS AND SOCIETY

DEPARTMENT OF STATISTICS
Sequoia Hall
Stanford University
Stanford, CA 94305-4065
SPATIAL COVARIANCE ESTIMATION
FOR MONITORING DATA

by

CLIVE LOADER
and
PAUL SWITZER
STANFORD UNIVERSITY

TECHNICAL REPORT NO. 133
NOVEMBER 1989

STUDY ON STATISTICS AND ENVIRONMENTAL
FACTORS IN HEALTH (SIMS)

PREPARED UNDER A GRANT TO SIMS FROM
ENVIRONMENTAL PROTECTION AGENCY (EPA)
SLOAN FOUNDATION
NATIONAL SCIENCE FOUNDATION (NSF)

DEPARTMENT OF STATISTICS
DEPARTMENT OF APPLIED EARTH SCIENCES
STANFORD UNIVERSITY
STANFORD, CALIFORNIA
Spatial Covariance Estimation for Monitoring Data

Clive Loader and Paul Switzer
Stanford University

ABSTRACT

We discuss the role of spatial covariances particularly with regard to estimates of precision for interpolation. Emphasis is placed on the heterogeneous case where precision can vary locally. We describe procedures for modelling heterogeneous covariances and their estimation from repeated observations at fixed sites. Our methods avoid parametric assumptions that are frequently made for this problem. An example shows the implementation of suggested procedures using monthly rainfall acidity data from a monitoring network in the eastern United States obtained over a two year period.

1. Introduction and Summary

Let \( \{Z(x); x \in \Xi\} \) be a real valued spatial field. In our example, \( Z(x) \) represents the concentration of a pollutant in rainfall, and \( \Xi \) will be a two dimensional geographic region. We observe \( Z \) at only at monitored locations \( x_1, \ldots, x_p \), for \( n \) replications of the process. The replications are denoted by \( Z_1(x), \ldots, Z_n(x) \). To motivate the discussion of spatial covariance, suppose we want to estimate or interpolate the values of the random function at unmonitored locations, to estimate the interpolation error or to choose good locations for adding or deleting monitoring locations.

The estimation of interpolation errors, for example, will require some degree of prior specification of properties of the random function. This can be achieved by regarding \( Z(x) \) as a realization of a random field and by changing the problem to one of estimation of expected values of interpolation errors or mean square error. It is convenient to use the notation
\[ \mu(x,j) = E(Z_j(x)), \ x \in \Xi, \ j=1,...,n \]

to denote the possibly replication dependent mean function and
\[ \sigma(x,y) = E(Z_j(x) - \mu(x,j))(Z_j(y) - \mu(y,j)), \ x, y \in \Xi \]
to denote the covariance function of the random field, which we assume to be replication independent. The mean field \( \mu(x,j) \) should capture the smooth and regular part of the spatial field \( Z_j(x) \), while \( \sigma(x,y) \) describes the variability of the residual field. When the interpolation estimates are linear combinations of the monitored values the mean squared errors can be derived explicitly as a function of \( \sigma(x,y) \) and the linear estimates used. To estimate the interpolation precision, we first need to estimate the covariance function.

The usual method in Geostatistics literature is to specify the covariance function up to a small number of unknown parameters, frequently restricting attention to homogeneous models. For example, a stationary variogram \( \gamma(x - y) = E(Z(x) - Z(y))^2 \) which does not depend on \( x \), may be used. These assumptions may conceal important spatially specific information. When we have multiple realizations, the need for strong parametric assumptions seems less compelling.

The multiple realizations of the process may exhibit serial correlation, but we will not address this question in detail. Instead, we remove a smooth time trend, and ignore any serial correlation which remains after removal of the trend.

For our example of rainfall acidity, we have 24 monthly values at each of 19 sites. For any pair of monitoring sites we can calculate the empirical covariance from the corresponding pair of time series. However, site pairs which are separated by the same spatial vector will inevitably have different empirical covariances, reflecting local properties of the field.

Spatial heterogeneity creates special problems for estimation of interpolation precision because mean interpolation errors depend not only on site-pair covariances but also on covariances between sites and the interpolation points, for which empirical estimates are lacking. Furthermore, the empirical site-pair covariances may themselves
be subject to sampling variability when the time series are short. Therefore, some
degree of parametric modelling is required which at the same time respects the
apparent heterogeneity. Proposals for such modelling are described. Basically, a
parametric covariance model is forced on the available empirical covariances and
modified covariance estimates are obtained by shrinking the empirical covariances
towards the parametric covariances.

Our example exhibits different variability characteristics in different parts of the
region, with differences in both the correlation length and the preferred correlation
direction.

2. Covariance Estimation for Monitored Locations

Let $\Sigma$ be the $p \times p$ matrix of covariances between the $p$ observed locations
$x_1, \ldots, x_p$. In this section we assume the process has known mean $\mu(x)$, which
without loss of generality we can take to be $0$. The natural estimate of $\Sigma$ is

$$\hat{\Sigma} = \frac{1}{n} \sum_{j=1}^{n} Z_j Z_j'$$

(2.1)

where $Z_j = (Z_j(x_1), \ldots, Z_j(x_p))'$; $1 \leq j \leq n$. Of course, this tells us nothing about the
covariances involving unobserved locations. Furthermore, although we do not wish to
make parametric assumptions about $\sigma(x, y)$, it is usually reasonable to make some
smoothness assumptions. For example, if the realizations of $Z(x)$ are continuous, the
covariance function must also be continuous. Especially if $n$ is small, the estimate $\hat{\Sigma}$
may not reflect this smoothness.

An appealing alternative is to use a shrinkage model, in which the data is allowed
to select between a parametric model and the empirical estimate $\hat{\Sigma}$. This suggests an
estimate of $\Sigma$ of the form

$$\tilde{\Sigma} = \lambda \hat{\Sigma} + (1-\lambda)C$$

(2.2)

for some $\lambda$, $0 < \lambda < 1$. The parameter $\lambda$ is chosen from the data.
To put this in a more formal setting, we show how the estimate (2.2) may be derived as a Bayes estimate with the additional assumption that \( Z(x) \) is a Gaussian process. More mathematical details may be found in Anderson (1984).

Let \( \Sigma \) have an Inverted Wishart prior distribution, with density

\[
    w^{-1}(\Sigma | \tilde{C}, m) = \frac{1}{\Gamma_p(\frac{1}{2}m) \Gamma_p(\frac{1}{2}m + (m + p + 1))} |\tilde{C}|^{\frac{1}{2}m} |\Sigma|^{-\frac{1}{2}(m + p + 1)} \exp(-\frac{1}{2} \text{tr}(\Sigma^{-1} \tilde{C})) \tag{2.3}
\]

where \( \tilde{C} = (m - p - 1)C \), and \( \Gamma_p(t) = \pi^{p(p-1)/4} \prod_{j=1}^{p} \Gamma(t - \frac{1}{2}(j - 1)) \). The Conditional distribution of \( n \hat{\Sigma} \) given \( \Sigma \) is a Wishart distribution, with density

\[
    w(n \hat{\Sigma} | \Sigma, n) = \frac{1}{\Gamma_p(\frac{1}{2}n) \Gamma_p(\frac{1}{2}n + (n - p - 1))} |\hat{\Sigma}|^{\frac{1}{2}n} |\Sigma|^{\frac{1}{2}n} \exp(-\frac{1}{2} \text{tr}(\Sigma^{-1} \hat{\Sigma})) \tag{2.4}
\]

The posterior mean of \( \Sigma \) is

\[
    E(\Sigma | \hat{\Sigma}) = \frac{1}{n + m - p - 1} (n \hat{\Sigma} + \tilde{C}) = \lambda \hat{\Sigma} + (1 - \lambda)C \tag{2.5}
\]

where \( (n + m - p - 1)\lambda = n \).

For the empirical Bayes approach, we use the data to estimate the parameter \( m \) of the prior distribution of \( \Sigma \). The marginal density of \( n \hat{\Sigma} \) is given by

\[
    M(n \hat{\Sigma} | \Sigma, n, m) = \frac{\Gamma_p(\frac{1}{2}(n + m))}{\Gamma_p(\frac{1}{2}n) \Gamma_p(\frac{1}{2}m)} \frac{|\hat{\Sigma}|^{\frac{1}{2}n} |\Sigma|^{\frac{1}{2}m} |\Sigma|^{\frac{1}{2}(n - p - 1)}}{|\hat{\Sigma} + \tilde{C}|^{\frac{1}{2}(n + m)}} \tag{2.6}
\]

The maximum likelihood estimate of \( m \) is the value of \( m \) that maximizes (2.6), or equivalently maximizes

\[
    \frac{\Gamma_p(\frac{1}{2}(m + n))}{\Gamma_p(\frac{1}{2}m)} (m - p - 1)^{-\frac{1}{2}np} |I + \frac{nC^{-\frac{1}{2}}\hat{\Sigma}C^{-\frac{1}{2}}}{m - p - 1}|^{-\frac{1}{2}(m + n)}, \tag{2.7}
\]

where \( C^{-\frac{1}{2}} \) is any symmetric matrix satisfying \((C^{-\frac{1}{2}})^2 = C^{-1}\). The advantage of (2.7) over (2.6) is computational. The eigenvalues \( \lambda_1, \ldots, \lambda_p \) of \( C^{-\frac{1}{2}}\hat{\Sigma}C^{-\frac{1}{2}} \) are real, and

\[
    |I + \frac{nC^{-\frac{1}{2}}\hat{\Sigma}C^{-\frac{1}{2}}}{m - p - 1}| = \prod_{i=1}^{p} \left( 1 + \frac{n \lambda_i}{m - p - 1} \right),
\]

which simplifies the computation of the determinant when evaluating (2.7) for many
values of \( m \). The maximizing value of \( m \) is used in (2.5) to obtain the estimate \( \tilde{\Sigma} \) of \( \Sigma \).

3. Extension to Unmonitored Locations.

We want to estimate the covariance vectors \( \sigma(x) = (\sigma(x,x_1), \ldots, \sigma(x,x_n))' \), for an unobserved \( x \in X \). Let \( c(x) \) be the estimate from the parametric model. Then Switzer (1989) suggests an estimate of \( \sigma(x) \), which modified for our shrinkage estimate of \( \Sigma \) becomes

\[
\tilde{\sigma}(x) = \tilde{\Sigma} C^{-1} c(x). \tag{3.1}
\]

We note when using (3.1), if \( x = x_i \) for some \( i \), the estimate \( \tilde{\sigma}(x) \) is the \( i \)th column of \( \tilde{\Sigma} \), so the extension is faithful to the covariance estimates for the observed locations.

We want the extended covariance function to be positive definite. In general, we let \( x \) be a vector of \( q \) unobserved locations, and \( \Sigma_{\text{new}}(x) \) be the covariance matrix between these locations. Let \( C(x) \) be the \( p \times q \) matrix of parametric estimate of covariances between locations, using the same parametric model used to estimate \( C \) in section 2. For positive definiteness we require for all \( \lambda_1 \in \mathbb{R}^p \) and \( \lambda_2 \in \mathbb{R}^q \),

\[
0 \leq (\lambda_1', \lambda_2') \begin{bmatrix}
\tilde{\Sigma} & \tilde{\Sigma} C^{-1} c(x) \\
c(x)' C^{-1} \tilde{\Sigma} & \Sigma_{\text{new}}(x)
\end{bmatrix}
\begin{bmatrix}
\lambda_1 \\
\lambda_2
\end{bmatrix}
\]

\[
= (\lambda_1' - \lambda_2' C(x)' C^{-1}) \tilde{\Sigma}(\lambda_1 - C^{-1} c(x) \lambda_2) + \lambda_2' (\Sigma_{\text{new}}(x) - \tilde{\sigma}(x)' \tilde{\Sigma}^{-1} \tilde{\sigma}(x)) \lambda_2. \tag{3.2}
\]

A necessary and sufficient condition for (3.2) to be satisfied for all \( \lambda_1 \) and \( \lambda_2 \) is that

\[
\Sigma_{\text{new}}(x) - \tilde{\sigma}(x)' \tilde{\Sigma}^{-1} \tilde{\sigma}(x) \tag{3.3}
\]

be non-negative definite. Estimating \( \tilde{\Sigma}_{\text{new}}(x) \) for multiple sites simultaneously appears to be difficult; we suggest a sequential approach where one site is added at a time. Suppose we have two additional sites \( x \) and \( y \). We want to compare the estimates obtained by adding \( x \) first with the estimates obtained by adding \( y \) first. We have
\[
\begin{bmatrix}
I & C^{-1}c(x)
\end{bmatrix}
= 
\begin{bmatrix}
C^{-1} & 0
\end{bmatrix}
\begin{bmatrix}
C & c(x)
\end{bmatrix}
\begin{bmatrix}
c(x)' & c(x,x)
\end{bmatrix}
\]

so
\[
\begin{bmatrix}
\tilde{\Sigma} & \sigma(x)
\end{bmatrix}
\begin{bmatrix}
C & c(x)
\end{bmatrix}
\begin{bmatrix}
c(x)' & c(x,x)
\end{bmatrix}^{-1}
= 
\tilde{\Sigma}
\begin{bmatrix}
I & C^{-1}c(x)
\end{bmatrix}
\begin{bmatrix}
C & c(x)
\end{bmatrix}
\begin{bmatrix}
c(x)' & c(x,x)
\end{bmatrix}^{-1}
\]
\[
= \tilde{\Sigma}
\begin{bmatrix}
C^{-1} & 0
\end{bmatrix}.
\]

From (3.4), we see the estimate \( \tilde{\sigma}(y) \) will be the same whether or not we have added \( x \) to the model first. This shows if we let the permutation group \( S_q \) operate on the new locations \( x_{p+1}, \ldots, x_{p+q} \), then the estimates of covariance between observed and unobserved locations generated sequentially are invariant with respect to elements of \( S_q \) of the form \( (j, j+1), j = 1, \ldots, q-1 \). Since these elements generate \( S_q \) (Lang (1984) p57), it follows that we get the same estimate of covariance between observed and unobserved locations regardless of the order in which new locations are added.

By adding new locations to the model sequentially we only need to consider additional methods to estimate the diagonal elements of \( \Sigma_{\text{new}}(x) \); estimates of the off diagonal elements are obtained automatically. However, the order invariance will not extend to the estimates of covariance between unobserved locations, which will depend on the estimates of diagonal elements of \( \Sigma_{\text{new}}(x) \). There are several possibilities for estimating the diagonal elements of \( \Sigma_{\text{new}}(x) \). The simplest is to just choose a constant, for example, the mean of the diagonal elements of \( \tilde{\Sigma} \). When there is some spatial structure to the variance, it may be preferable to smooth the diagonal elements of \( \tilde{\Sigma} \). In either case, we may get problems with positive definiteness if there is a strong correlation between locations.

The problem of estimating spatially heterogeneous covariances may also be approached by rescaling procedures. The basic idea is to map locations \( x \in \mathbb{R}^2 \) to new locations \( f(x) \in \mathbb{R}^k \) such that spatial covariances in the image space are stationary and isotropic. Such rescaling procedures may be described as follows:
1) Pick a strictly monotone isotropic covariance function $\tilde{C}(h)$ where $h$ is distance in the image space. e.g. $\tilde{C}(h) = e^{-h}$. This choice is arbitrary and can affect the subsequent covariance estimate.

2) If $c_{ij}$ are the empirical covariances (or correlations) between monitoring sites $x_i$ and $x_j$, then

$$\tilde{C}^{-1}(c_{ij}) = ||f(x_i) - f(x_j)||, \quad i, j = 1, \ldots, p \tag{3.5}$$

where $||.||$ denotes distance in $R^k$, $f$ is the desired mapping and $\tilde{C}^{-1}$ is the inverse function of $\tilde{C}$. If we choose the image space dimension, $k$, to be less than $p-1$, then we will not in general be able to choose the transformation $f$ such that (3.5) holds exactly for all $i$ and $j$. Some approximation will be needed. See, for example, Sampson (1987), who considers mappings from $R^2$ to $R^2$.

3) For an unmonitored location $x_0$, we interpolate to obtain $f(x_0)$. If the interpolation is linear, we will obtain

$$f(x_0) = \sum_{i=1}^{p} w_i(x_0)f(x_i)$$

where $w_i(x_0)$ is a $k \times k$ matrix.

4) The estimated covariance is

$$\hat{\sigma}_{0j} = \tilde{C}(||f(x_0) - f(x_j)||).$$

Using such procedures readily gives covariance estimates between unmonitored locations. However, the choice of $\tilde{C}$ and $f$ is somewhat arbitrary.

4. Interpolation and Estimation of the mean

In the preceding section we assume $E(Z(x)) = 0$. This will usually mean an estimate of $E(Z(x))$ has been subtracted before estimating the covariance, and $n$ is reduced to compensate for this estimation. In this section, we consider estimation of $\mu(x, j) = E(Z_j(x))$ and $\hat{Z}(x) = E(Z(x)|Z(x_1), \ldots, Z(x_p))$. We also discuss the precision of these estimates.
We assume a generalized additive model (Hastie and Tibshirani, 1987) for $\mu(x,j)$,

$$\mu(x,j) = \mu + X(x) + T(j) \quad (4.1)$$

where $X$ and $T$ are suitably smooth functions, subject to the constraints

$$\sum_{i=1}^{p} X(x_i) = 0 \text{ and } \sum_{j=1}^{n} T(j) = 0.$$

Assuming an orthogonal design (i.e. observations are made at the same sites at each time) we can estimate $X$ and $T$ by applying a suitable smoother to the marginal effects of the array $Z_j(x_i); i=1, \ldots, p; j=1, \ldots, n$. For the purposes of this paper, we use the loess smoother (Cleveland and Devlin, 1988), but many other smoothers or parametric regression models could be used.

The MLE of $Z_j(x)$ for unobserved $x$ if $\mu(x,t)$ and $\sigma(x,y)$ were known is

$$\hat{Z}_j(x) = \mu(x,j) + \sigma(x)\Sigma^{-1}(Z_j - \mu(.,j)) \quad (4.2)$$

where $\mu(.,j) = (\mu(x_1,j), \ldots, \mu(x_p,j))'$. Using the estimators proposed for $\mu$ and $\sigma$, (4.2) becomes

$$\hat{Z}_j(x) = \hat{\mu}(x,j) + c(x)'C^{-1}(Z_j - \hat{\mu}(.,j)). \quad (4.3)$$

We note if $x = x_1$, then $c(x)'C^{-1} = (1,0, \ldots, 0)$, so (4.3) reproduces the observed data. Also, (4.3) depends only on the parametric model used to derive $C$, and not on $\Sigma$.

The prediction error is defined by

$$\eta(x,t) = E((Z_t(x) - \hat{Z}_t(x))^2) \quad (4.4)$$

From standard probability theory and the assumption of no serial correlation, we have

$$\eta(x,t) \geq \text{var}(Z_t(x)|Z_t)$$

$$= \sigma(x,x) - \sigma(x)\Sigma^{-1}\sigma(x). \quad (4.5)$$

Note that (4.4) is the prediction error of $\hat{Z}_j(x)$. We might hope the prediction error could be estimated from (4.5). However, this makes no allowance for variability in $\mu(x,t)$, and our example in section 5 will show that this may increase prediction error substantially.
We derive the covariance function of $\bar{\mu}(x,t)$ for fixed $t$. The marginal effects of the array $Z_j(x_i)$ are given by

$$Z_i(x_i) = \frac{1}{n} \sum_{j=1}^{n} Z_j(x_i) - \mu = X(x_i) + \frac{1}{n} \sum_{j=1}^{n} \epsilon_j(x_i) - \bar{\mu} \quad (4.6)$$

and

$$Z_j(.) = \frac{1}{p} \sum_{i=1}^{p} Z_j(x_i) - \bar{\mu} = T(j) + \frac{1}{p} \sum_{i=1}^{p} \epsilon_j(x_i) - \bar{\mu}, \quad (4.7)$$

where

$$\bar{\mu} = \frac{1}{np} \sum_{i=1}^{p} \sum_{j=1}^{n} Z_j(x_i) = \mu + \frac{1}{np} \sum_{i=1}^{p} \sum_{j=1}^{n} \epsilon_j(x_i). \quad (4.8)$$

Let $l_i(x), i=1, \ldots, p$ be the smoothing weights for estimating $X(x)$ and $m_j(t), j=1, \ldots, n$ be the smoothing weights for estimating $T(t)$. Then $\hat{X}(x) = \sum_{i=1}^{p} l_i(x) Z_i(x)$ and $\hat{T}(t) = \sum_{j=1}^{n} m_j(t) Z_j(.)$. We assume the smoothing procedure used to estimate $X(x)$ and $T(t)$ will reproduce constants. This implies $\sum_{i=1}^{p} l_i(x) = \sum_{j=1}^{n} m_j(t) = 1$ for all $x$ and $t$. Then from (4.6), (4.7) and (4.8) we get

$$\bar{\mu}(x,t) = \mu + \sum_{i=1}^{p} l_i(x) X(x_i) + \sum_{j=1}^{n} m_j(t) T(j) + \sum_{i=1}^{p} \sum_{j=1}^{n} \left[ \frac{l_i(x)}{n} + \frac{m_j(t)}{p} - \frac{1}{np} \right] \epsilon_j(x_i), \quad (4.9)$$

giving (with our assumption $\text{cov}(\epsilon_j(x_i), \epsilon_{j'}(x_{i'})) = 0$ if $j \neq j'$),

$$\text{cov}(\bar{\mu}(x,t), \bar{\mu}(x',t)) = \sum_{j=1}^{n} \sum_{i=1}^{p} \sum_{i'=1}^{p} \left[ \frac{l_i(x)}{n} + \frac{m_j(t)}{p} - \frac{1}{np} \right] \left[ \frac{l_{i'}(x')}{n} + \frac{m_j(t)}{p} - \frac{1}{np} \right] \sigma(x_i, x_{i'}) \quad (4.10)$$

where $l(x) = (l_1(x), \ldots, l_p(x))$ and $\mathbf{1}_p$ represents a vector of 1's of length $p$.

We can now derive a more accurate formula for the prediction error (4.4). Assuming there is no bias, (4.4) is the variance of $Z_i(x) - \bar{Z}_i(x)$. Then

$$\eta(x,t) = \sigma(x,x) - 2 \text{cov}(Z_i(x), \bar{Z}_i(x)) + \text{var}(\bar{Z}_i(x)) \quad (4.11)$$

Neglecting variability in fitting the parametric model, we use (4.7) to get
\[ \text{cov}(Z_t(x),\tilde{Z}_t(x)) = \text{cov}(Z_t(x),\bar{\mu}(x,t)) + c(x)YC^{-1}\text{cov}(Z_t(x),Z_t-\bar{\mu}(.,t)) \]

\[ = \left[ l(x)' + (m_t(t) - \frac{1}{n}) \frac{1}{p} + c(x)YC^{-1}(I - A_t) \right] \sigma(x) \quad (4.12) \]

where

\[ A_t = \frac{1}{n}L + \left( \frac{m_t(t)}{p} - \frac{1}{np} \right)J. \]

Here, \( L \) is the \( p \times p \) matrix with \( L_{i,j} = l_j(x_i) \) and \( J \) is the \( p \times p \) matrix of 1's. Similarly,

\[ \text{var}(\tilde{Z}_t(x)) = \text{var}(\bar{\mu}(x,t)) + c(x)YC^{-1}\text{var}(Z_t-\bar{\mu}(.,t))C^{-1}c(x) \]

\[ + 2c(x)YC^{-1}\text{cov}(\bar{\mu}(x,t),Z_t-\bar{\mu}(.,t)) \]

\[ = \frac{1}{n}l(x)'\Sigma l(x) + \frac{1}{p^2} \sum_{j=1}^{n} (m_j(t) - \frac{1}{n})^2 \]

\[ + c(x)C^{-1} \left[ (I - A_t)\Sigma(I - A_t)' - A_t\Sigma A_t' + \frac{L\Sigma L'}{n} + \frac{J\Sigma J}{p^2} \sum_{j=1}^{n} (m_j(t) - \frac{1}{n})^2 \right] C^{-1}c(x) \quad (4.13) \]

\[ + 2c(x)YC^{-1} \left[ \Sigma(\frac{l(x)}{n} + (m_t(t) - \frac{1}{n}) \frac{1}{p}) - \frac{L\Sigma l(x)}{n} - \frac{J\Sigma J}{p^2} \sum_{j=1}^{n} (m_j(t) - \frac{1}{n})^2 \right] \]

It is easily verified from (4.11), (4.12) and (4.13) that \( \eta(x_i,t) = 0; \ i=1, \ldots, p \).

An estimate of \( \eta(x,t) \) is found by replacing \( \sigma(x) \) by \( \hat{\sigma}(x) \) in (4.12), \( \Sigma \) by \( \hat{\Sigma} \) in (4.13) and estimating \( \sigma(x,x) \) as suggested in section 3. After some algebra, we get

\[ \hat{\eta}(x,t) = \hat{\sigma}(x,x) - c(x)YC^{-1}\hat{\Sigma}C^{-1}c(x) + \frac{1}{p^2} \sum_{j=1}^{n} (m_j(t) - \frac{1}{n})^2 \]

\[ + \frac{1}{n}(l(x) - c(x)YC^{-1}L)\hat{\Sigma}(l(x)' - L'C^{-1}c(x)). \]

5. Example: Acid Rain Data

We have monthly average sulphate concentrations at 19 sites in the Eastern United States, measured over a two year period. A log transformation has been applied to the data. The sites are shown in Figure 1.
Both components of (4.1) are estimated using the loess smoother. Local quadratic fitting is used to estimate \( T(t) \) and bivariate local linear fitting is used to fit \( X(x) \). In Figure 2, we show \( M_f \) plots (Cleveland and Devlin, 1988) for both cases. We select the smoothing parameters \( f = 18/24 \) for estimating \( T(t) \) and \( f = 10/19 \) for estimating \( X(x) \). Figure 3 plots the sulphate concentrations for each site, together with the smoothed time trend. Figure 4 is a plot of estimated contours for \( X(x) \). The highest sulphate concentrations occur in the Midwest and Mid Atlantic areas, while the lowest sulphate concentrations are in the Great Plains States (The high concentrations in the Atlantic Ocean are presumably spurious extrapolations).

We use the parametric model \( C(x,y) = ae^{-\lambda h} \) where \( h \) is the grand circle distance between \( x \) and \( y \). Estimates of \( a \) and \( \lambda \) are found by least squares applied to the \( p(p+1)/2 \) variances and covariances between sites. Figure 5 shows the column of \( \hat{\Sigma} \) for each site, and the lines represent the fitted model.

The shrinkage parameter found using (2.5) and (2.6) was \( \lambda = 0.6 \). In Figure 6 we plot estimated contours of \( \sigma(x,y) \) over the eastern half of the United States, using (2.8).

We use cross validation to assess how well the estimates of accuracy are performing. We delete each of the sites in turn, and use the remaining 18 sites to estimate the sulphate concentrations at the deleted sites. We compute the site bias,

\[
\beta(x) = \frac{1}{n} \sum_{j=1}^{n} (Z_j(x) - \bar{Z}_j(x)) \quad (5.1)
\]

and variance,

\[
\nu(x) = \frac{1}{n} \sum_{j=1}^{n} \left[ Z_j(x) - \bar{Z}_j(x) - \beta(x) \right]^2. \quad (5.2)
\]

The expected values are estimated by splitting (4.14) and summing over \( t \). This gives

\[
E(\beta(x)^2) = n \sigma(x,x) - c(x)'C^{-1} \bar{\Sigma} C^{-1} c(x) + \frac{1}{p^2} \sum_{t=1}^{n} \sum_{j=1}^{n} (m_j(t) - \frac{1}{n})^2
\]

and
E(v(x)) = (l(x) - c(x)'C^{-1}L)\tilde{\Sigma}(l(x)' - LC^{-1}c(x)).

The results are given in table 1. We include estimates of (4.5) for comparison.

<table>
<thead>
<tr>
<th>Site</th>
<th>(\beta(x)^2)</th>
<th>(\nu(x))</th>
<th>(E\beta(x)^2)</th>
<th>(Ev(x))</th>
<th>(4.5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Turners Falls, MA</td>
<td>0.1445</td>
<td>0.2057</td>
<td>0.0156</td>
<td>0.1993</td>
<td>0.2074</td>
</tr>
<tr>
<td>Tunkhannock, PA</td>
<td>0.0873</td>
<td>0.2470</td>
<td>0.0164</td>
<td>0.2924</td>
<td>0.3034</td>
</tr>
<tr>
<td>Zanesville, OH</td>
<td>0.0019</td>
<td>0.0958</td>
<td>0.0025</td>
<td>0.0481</td>
<td>0.0499</td>
</tr>
<tr>
<td>Rockport, IN</td>
<td>0.0031</td>
<td>0.2061</td>
<td>0.0141</td>
<td>0.2233</td>
<td>0.2317</td>
</tr>
<tr>
<td>Fort Wayne, IN</td>
<td>0.1498</td>
<td>0.2290</td>
<td>0.0120</td>
<td>0.2187</td>
<td>0.2269</td>
</tr>
<tr>
<td>Raleigh, NC</td>
<td>0.3545</td>
<td>0.2080</td>
<td>0.0236</td>
<td>0.2457</td>
<td>0.2532</td>
</tr>
<tr>
<td>Gaylord, MI</td>
<td>1.0999</td>
<td>0.4184</td>
<td>0.0492</td>
<td>0.3646</td>
<td>0.3767</td>
</tr>
<tr>
<td>Clearfield, KY</td>
<td>0.1225</td>
<td>0.1823</td>
<td>0.0101</td>
<td>0.2062</td>
<td>0.2146</td>
</tr>
<tr>
<td>Alamo, TN</td>
<td>0.0009</td>
<td>0.4928</td>
<td>0.0205</td>
<td>0.4126</td>
<td>0.4288</td>
</tr>
<tr>
<td>Winterport, ME</td>
<td>0.3621</td>
<td>0.2744</td>
<td>0.0281</td>
<td>0.2479</td>
<td>0.2561</td>
</tr>
<tr>
<td>Uvalda, GA</td>
<td>0.2542</td>
<td>0.4279</td>
<td>0.0238</td>
<td>0.3658</td>
<td>0.3789</td>
</tr>
<tr>
<td>Selma, AL</td>
<td>0.0584</td>
<td>0.1885</td>
<td>0.0130</td>
<td>0.2197</td>
<td>0.2277</td>
</tr>
<tr>
<td>Clinton, MS</td>
<td>0.0156</td>
<td>0.3156</td>
<td>0.0144</td>
<td>0.2701</td>
<td>0.2804</td>
</tr>
<tr>
<td>Marshall, TX</td>
<td>0.1404</td>
<td>0.2462</td>
<td>0.0268</td>
<td>0.2668</td>
<td>0.2747</td>
</tr>
<tr>
<td>Lancaster, KS</td>
<td>0.0573</td>
<td>0.2407</td>
<td>0.0178</td>
<td>0.2543</td>
<td>0.2617</td>
</tr>
<tr>
<td>Brookings, SD</td>
<td>0.0588</td>
<td>0.5707</td>
<td>0.0446</td>
<td>0.4374</td>
<td>0.4519</td>
</tr>
<tr>
<td>Underhill, VT</td>
<td>0.0008</td>
<td>0.2509</td>
<td>0.0112</td>
<td>0.2083</td>
<td>0.2170</td>
</tr>
<tr>
<td>Big Moose, NY</td>
<td>0.0019</td>
<td>0.2423</td>
<td>0.0130</td>
<td>0.2669</td>
<td>0.2780</td>
</tr>
<tr>
<td>McArthur, OH</td>
<td>0.0227</td>
<td>0.1531</td>
<td>0.0076</td>
<td>0.1726</td>
<td>0.1800</td>
</tr>
</tbody>
</table>

Table 1. Cross Validated prediction errors and estimates.

From table 1, we see the site biases are in many cases substantially larger than their estimated values. This means there is bias in the estimation of \(X(x)\), so the neighbourhoods used for the loess procedure are too large. Due to the small number of sites involved, it is not feasible to reduce the neighbourhood size further; to get better estimates, we would require data from additional sites.

The estimates derived using (4.5) are adjusted by a factor of 4, to allow for the model fitting with 24 times and 18 sites. For most sites, estimates of variability have performed well. In Figure 7, we plot the interpolated values and actual data for each of the 19 sites. In most cases we see a strong correlation between the two, indicating the fitting is performing well.
6. References


Figure 1. Acid Rain Sulfate Concentrations: 19 Recording Locations
Figure 2. Mf Plots to select the smoothing parameters to estimate $T(t)$ and $X(x)$
Figure 3. Sulphate Concentrations and Smoothed trends
Figure 5. Raw Covariance Estimates

x axis is distance (k.m.)
Figure 6. Interpolated Covariances
Figure 7. Cross Validation

x-axis = Interpolated Data
y-axis = Actual Data
with line y=x.