THE DESIGN OF SPATIAL MONITOR NETWORKS

Richard L. Smith and Zhengyuan Zhu
Department of Statistics and Operations Research
University of North Carolina
Chapel Hill, N.C., U.S.A.

Joint Statistical Meetings
Toronto
August 10 2004

This talk is based on the following preprints:

Smith, R.L. (2004), Asymptotic theory for kriging with estimated parameters and its application to network design. Preliminary version, available from
http://www.stat.unc.edu/postscript/rs/supp5.pdf

Zhu, Z. and Stein, M.L. (2004), Two-step spatial sampling design for prediction with estimated parameters. Preprint, University of North Carolina and University of Chicago.

I Background on Spatial Interpolation and Kriging

II "Predictive" and "Estimative" Approaches to Network Design

III A New Viewpoint on "Kriging with Estimated Parameters"

IV Application to Network Design

V Summary and Conclusions

I Background on Spatial Interpolation and Kriging

We assume data follow a *Gaussian random field* with mean and covariance functions represented as functions of finite-dimensional parameters.

Define the prediction problem as

$$\begin{pmatrix} Y \\ Y_0 \end{pmatrix} \sim N\left[\begin{pmatrix} X\beta \\ x_0^T\beta \end{pmatrix}, \begin{pmatrix} V & w^T \\ w & v_0 \end{pmatrix}\right]$$

where Y is an n-dimensional vector of observations, Y_0 is some unobserved quantity we want to predict, X and x_0 are known regressors, and β is a p-dimensional vectors of unknown regression coefficients.

Specifying the Covariances

The most common and widely used spatial models (stationary and isotropic) assume the covariance between components Y_i and Y_j is a function of the (scalar) distance between them, $C_\theta(d_{ij})$. For example,

$$C_\theta(d) \;=\; \sigma \exp\left\{-\left(\frac{d}{\rho}\right)^\kappa\right\},$$

where $\theta = (\kappa, \sigma, \rho)$, or

$$C_\theta(d) \;=\; \frac{\sigma}{2^{\nu-1}\Gamma(\nu)}\left(\frac{2\nu^{1/2}d}{\rho}\right)^\nu \mathcal{K}_\nu\left(\frac{2\nu^{1/2}d}{\rho}\right),$$

where $\mathcal{K}_\nu$ is a modified Bessel function and we have $\theta = (\nu, \sigma, \rho)$ (Matérn).

Estimation

$$Y \sim N[X\beta,\ V(\theta)]$$

1. Curve fitting to the variogram, based on residuals from OLS regression.

2. Maximum likelihood

3. Restricted maximum likelihood (REML): $\widehat{\beta}$ by GLS regression, $\widehat{\theta}$ chosen to maximize

$$\ell_n(\theta) \ = \ -\frac{1}{2}\log|V(\theta)| - \frac{1}{2}\log|X^T V(\theta)^{-1} X| - \frac{G^2(\theta)}{2}$$

where $G^2 = Y^T W Y$, $W = V^{-1} - V^{-1}X(X^T V^{-1} X)^{-1} X^T V^{-1}$, is the generalized residual sum of squares.

Universal Kriging

$$\begin{pmatrix} Y \\ Y_0 \end{pmatrix} \sim N\left[\begin{pmatrix} X\beta \\ x_0^T\beta \end{pmatrix}, \begin{pmatrix} V & w^T \\ w & v_0 \end{pmatrix}\right]$$

Find a predictor $\widehat{Y}_0 = \lambda^T Y$ that minimizes $\sigma_0^2 = E\left\{(Y_0 - \widehat{Y}_0)^2\right\}$ subject to $E\left\{Y_0 - \widehat{Y}_0\right\} = 0$. This leads to

$$\begin{aligned} \lambda &= w^T V^{-1} + (x_0 - X^T V^{-1} w)^T (X^T V^{-1} X)^{-1} X^T V^{-1}, \\ \sigma_0^2 &= v_0 - w^T V^{-1} w + (x_0 - X^T V^{-1} w)^T (X^T V^{-1} X)^{-1} (x_0 - X^T V^{-1} w). \end{aligned}$$

In the traditional formulation, V, w and v_0 are assumed known. When they depend on unknown parameters θ, we first estimate θ and use that to define V, w, v_0. This is called the *plug-in* approach.

Bayesian Reformulation of Universal Kriging

We assume θ has a prior density $\pi(\theta)$, and β has a flat prior independent of θ. The Bayesian predictive density of Y_0 given Y is

$$\begin{aligned} p(Y_0 \mid Y) &= \frac{\int\int f(Y, Y_0 \mid \beta, \theta)\pi(\theta)d\beta d\theta}{\int\int f(Y \mid \beta, \theta)\pi(\theta)d\beta d\theta} \\ &\vdots \\ &= \frac{\int e^{\ell_n(\theta)}\psi(\theta)\pi(\theta)d\theta}{\int e^{\ell_n(\theta)}\pi(\theta)d\theta} \end{aligned} \qquad (1)$$

where $e^{\ell_n(\theta)}$ is the restricted likelihood of θ and
$\psi(\theta) = \frac{1}{\sqrt{2\pi\sigma_0^2}} \exp\left\{-\frac{1}{2}\left(\frac{Y_0 - \lambda^T Y}{\sigma_0}\right)^2\right\}$.

The REML estimator $\widehat{\theta}$ is the value of θ that maximizes $\ell_n(\theta)$. We also write (1) as $\tilde{\psi}$, to distinguish it from the plug-in rule $\widehat{\psi} = \psi(\widehat{\theta})$.

An Example

Holland, Caragea and Smith (*Atmospheric Environment*, 2004), interested in long-term time trends in atmospheric SO_2 and particulate SO_4^{2-}, at 30 long-term monitor stations (CASTNet). Estimated trend (percent change from 1990 to 1999), with standard errors, are shown on the next two figures.

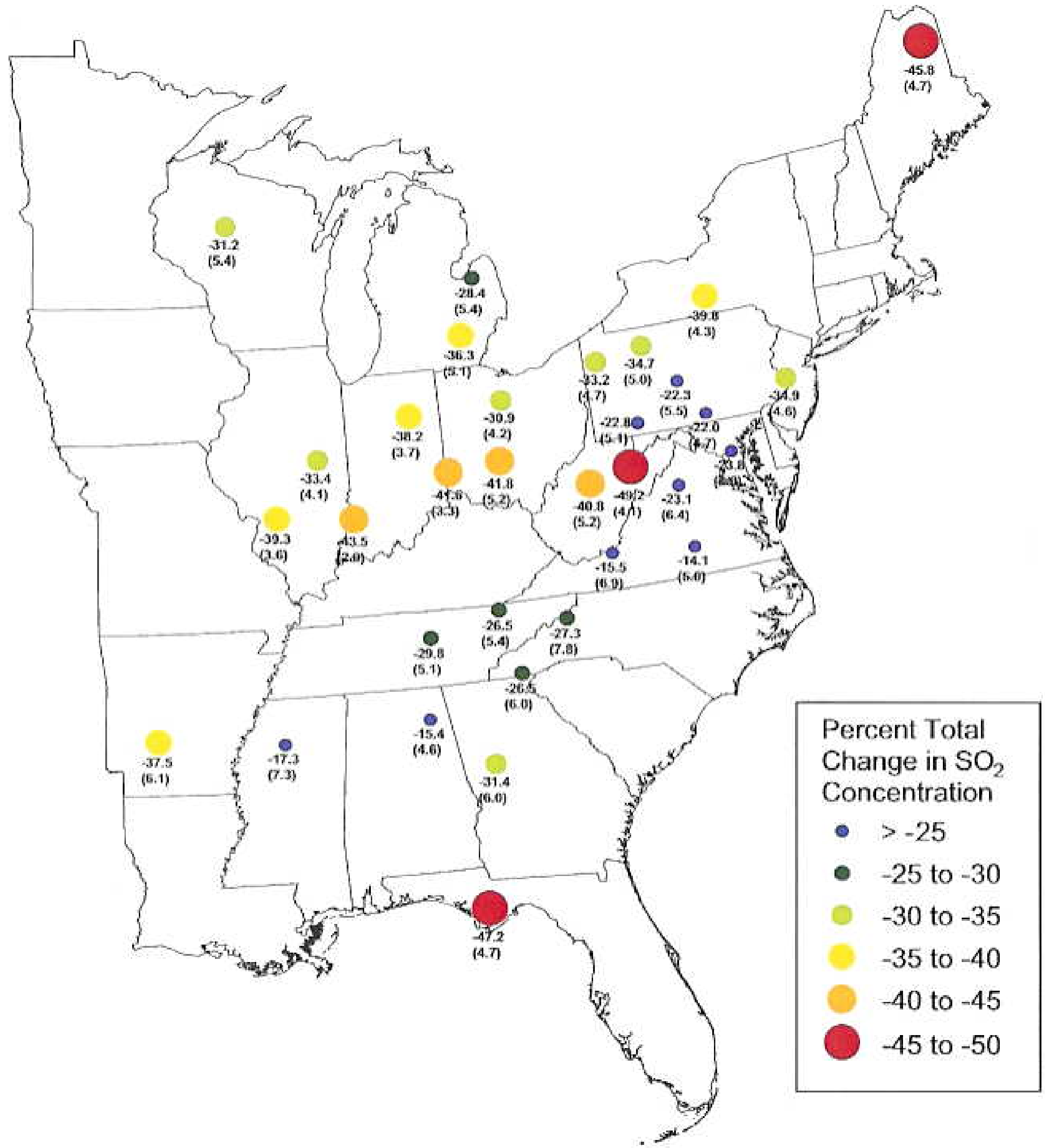


Fig. 1. Trend (percent change) in SO_2 concentrations at CASTNet sites from 1990 to 1999 with standard error (%) in parentheses.

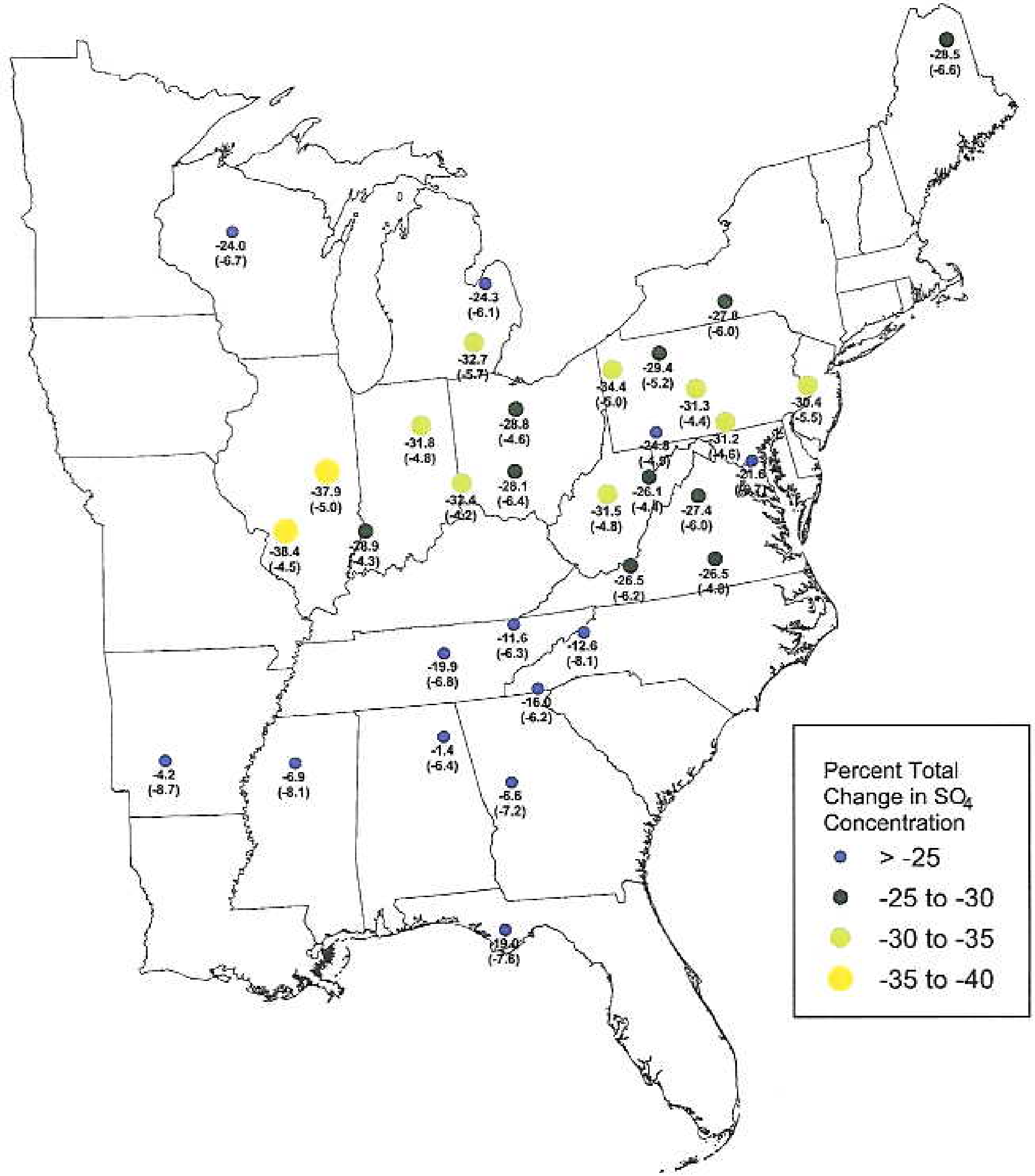
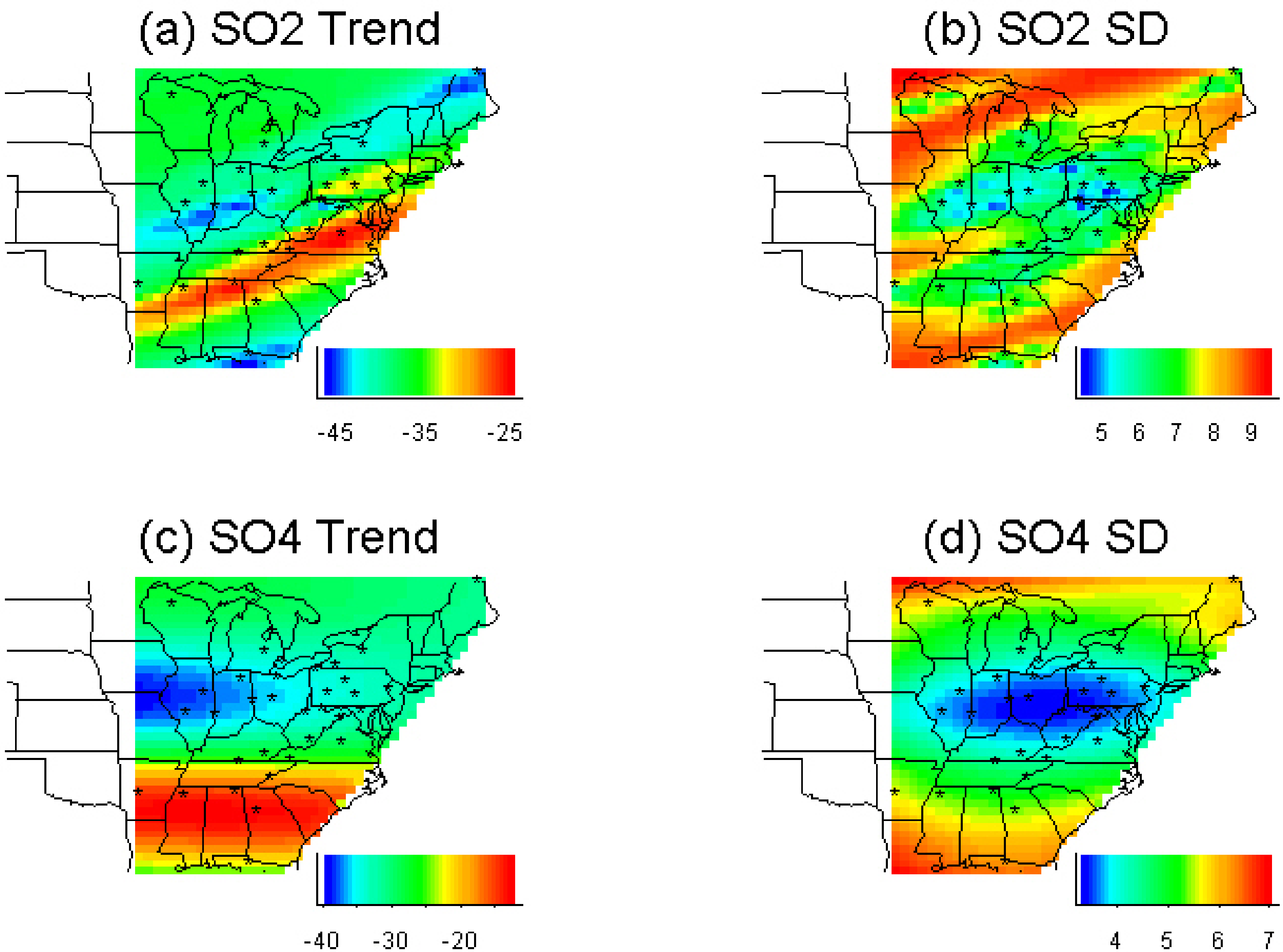


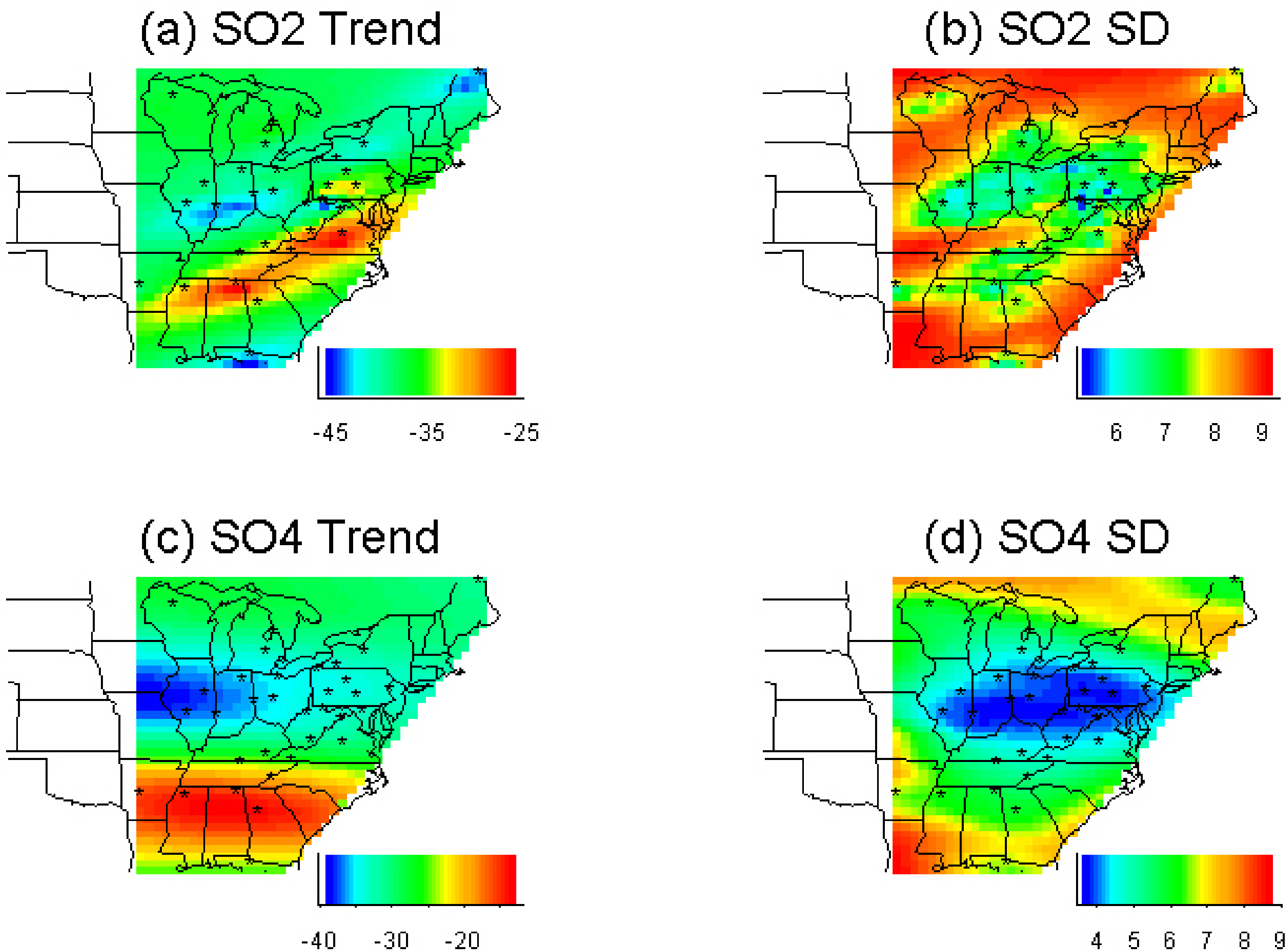
Fig. 2. Trend (percent change) in SO_4^{2-} concentrations at CASTNet sites from 1990 to 1999 with standard error (%) in parentheses.

We fitted a spatial model to the estimated trend at each location and constructed an interpolated surface by universal kriging, using both plug-in and Bayesian approaches. We also estimated the prediction standard errors by both methods.

Interpolated Surfaces, MLE Method



Interpolated Surfaces, Bayesian Method



Although the two maps are not very different, there are perceptible differences, with the prediction standard errors being larger under the Bayesian approach.

Next, we looked at *regional average* trend by averaging over a lattice.

Fig. 3. Lattice for the Midwest region.

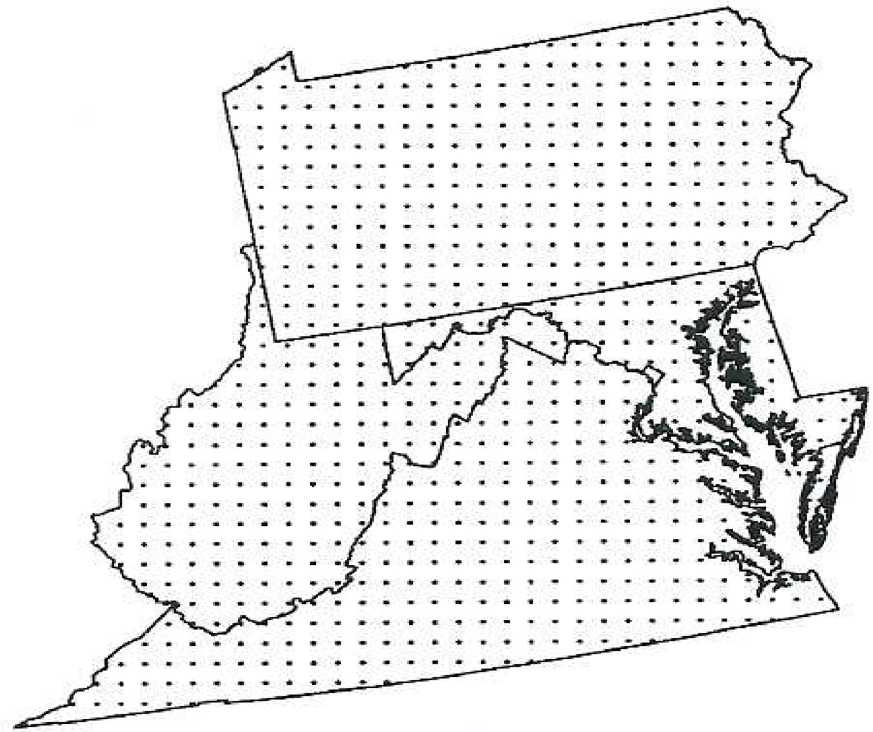


Fig. 4. Lattice for the Mid-Atlantic region.

Two lattices for “regional average” calculations

Variable	Region	Model	MLE	Bayes 1	Bayes 2
SO_2	Midwest	(b)	−40.03 (3.37)	−38.41 (3.64)	−39.80 (3.55)
SO_2	Midwest	(c)	−40.27 (3.20)	−37.87 (3.60)	−40.18 (3.44)
SO_2	Mid-Atlantic	(b)	−33.94 (3.56)	−36.56 (3.85)	−34.78 (3.84)
SO_2	Mid-Atlantic	(c)	−32.84 (3.48)	−36.79 (3.67)	−34.11 (3.70)
SO_4^{2-}	Midwest	(b)	−33.91 (3.95)	−34.59 (4.12)	−34.50 (4.02)
SO_4^{2-}	Midwest	(c)	−36.16 (3.53)	−34.88 (3.94)	−35.72 (3.72)
SO_4^{2-}	Mid-Atlantic	(b)	−31.19 (3.56)	−32.00 (3.74)	−31.64 (3.66)
SO_4^{2-}	Mid-Atlantic	(c)	−31.76 (3.54)	−33.19 (3.83)	−32.14 (3.66)

Regional estimates using two spatial models, by maximum likelihood and two Bayesian analyses. Predictive standard deviations in parentheses.

This example illustrates the application of spatial interpolation using both the maximum likelihood plug-in approach, and a Bayesian approach.

It also illustrates the point that there may be one or a small number of specific quantities of interest, such as the regional average for either the Midwest or Mid-Atlantic grid.

This raises the question of how we might *design* a network to optimize the interpolation or prediction of specific quantities of interest.

II "Predictive" and "Estimative" Approaches to Network Design (Zhu and Stein 2004)

Suppose we are interested in a specific Y_0, e.g. the average trend over the midwest region. How could we design a network specifically to estimate that variable as accurately as possible?

The most obvious design criterion is σ_0^2, the prediction error variance. However, this ignores the effect of having to estimate θ.

Harville and Jeske (1992) and Zimmerman and Cressie (1992) proposed the following correction to the mean squared prediction error:

$$V_1 = \sigma_0^2 + \mathrm{tr}\left\{\mathcal{I}^{-1}\left(\frac{\partial\lambda}{\partial\theta}\right)^T V \left(\frac{\partial\lambda}{\partial\theta}\right)\right\}$$

where $\mathcal{I}$ is the observed information matrix for θ. This formula corrects for the error in specifying the kriging weights λ.

However, in calculating a prediction interval for Y_0, it is also necessary to consider the effect of σ_0^2 being unknown. Zhu and Stein defined

$$V_2 \; = \; \left(\frac{\partial \sigma_0^2}{\partial \theta}\right)^T \mathcal{I}^{-1} \left(\frac{\partial \sigma_0^2}{\partial \theta}\right).$$

They suggested that some linear combination of V_1 and $\frac{V_2}{\sigma_0^2}$ would be a suitable design criterion taking account of both predictive and estimative considerations. Following arguments of Stein (1999), they suggested

$$V_3 \; = \; V_1 + \frac{V_2}{2\sigma_0^2}$$

as a suitable combined criterion. However, it's not clear exactly why this particular linear combination is appropriate.

Alternative Bayesian Approach

- For any data set, use MCMC to construct the Bayesian predictive distribution
- For any given design, run the Bayesian analysis on simulated data sets to determine the expected length of Bayesian prediction intervals
- Use an optimization algorithm (e.g. simulated annealing) to find the optimal design

It is attractive to use Bayesian prediction intervals because they automatically take account of parameter estimation. However, implementation of this approach seems too computationally intensive, and Zhu and Stein dismiss it in favor of their approximate approach. A further technical point is that there is no proof that Bayesian prediction intervals actually outperform the plug-in approach, when assessed for example by how close the actual coverage probability is to the nominal coverage probability.

III Kriging with Estimated Parameters

Redefine

$$\tilde{\psi}(z, Y) \;=\; \frac{\int e^{\ell_n(\theta)+Q(\theta)}\psi(z, Y; \theta)d\theta}{\int e^{\ell_n(\theta)+Q(\theta)}d\theta} \tag{2}$$

where $e^{\ell_n(\theta)}$ is the restricted likelihood of θ, $Q(\theta) = \log \pi(\theta)$ and $\psi(z, Y; \theta) = \Phi\left(\frac{Y_0 - \lambda^T Y}{\sigma_0}\right)$. Also let $\tilde{\psi}^{-1}$ be inverse function, i.e. $\tilde{\psi}^{-1}(P, Y)$ is the value of z for which $\tilde{\psi}(z, Y) = P$.

For $P \in (0, 1)$ define

$$\begin{aligned} z_P(Y; \theta) &= \lambda^T Y + \sigma_0 \Phi^{-1}(P), \\ \hat{z}_P(Y) &= \hat{\lambda}^T Y + \hat{\sigma}_0 \Phi^{-1}(P), \\ \tilde{z}_P(Y) &= \tilde{\psi}^{-1}(P, Y). \end{aligned}$$

For an estimator z_P^* (could be $\widehat{z}_P$ or $\widetilde{z}_P$) we would like to calculate

$$E\left\{\psi(z_P^*(Y)\ ;\ Y,\theta) - \psi(z_P(Y;\theta)\ ;\ Y,\theta)\right\} \tag{3}$$

and

$$E\left\{z_P^*(Y) - z_P(Y;\theta)\right\} \tag{4}$$

(3) is called the coverage probability bias (CPB). (4) leads to the expected length of a prediction interval (our proposed design criterion) because for a $100(P_2 - P_1)\%$ interval,

$$\begin{aligned} & E\left\{z_{P_2}^*(Y) - z_{P_1}^*(Y)\right\} \\ = \ & E\left\{z_{P_2} - z_{P_1}\right\} + E\left\{z_{P_2}^* - z_{P_2}\right\} - E\left\{z_{P_1}^* - z_{P_1}\right\} \\ = \ & \sigma_0\{\Phi^{-1}(P_2) - \Phi^{-1}(P_1)\} + E\left\{z_{P_2}^* - z_{P_2}\right\} - E\left\{z_{P_1}^* - z_{P_1}\right\} \end{aligned}$$

Define $U_i = \frac{\partial \ell_n(\theta)}{\partial \theta^i}$, $U_{ij} = \frac{\partial^2 \ell_n(\theta)}{\partial \theta^i \partial \theta^j}$, $U_{ijk} = \frac{\partial^3 \ell_n(\theta)}{\partial \theta^i \partial \theta^j \partial \theta^k}$.

Other quantities $Q(\theta) = \log \pi(\theta)$, $\lambda(\theta)$, $\sigma_0(\theta)$. Suffixes denote partial differentiation, e.g. $Q_i = \frac{\partial Q}{\partial \theta^i}$, $\sigma_{0ij} = \frac{\partial^2 \sigma_0}{\partial \theta^i \partial \theta^j}$.

Let

$$\begin{aligned}
\kappa_{i,j} &= n^{-1} E\left\{U_i U_j\right\}, \\
\kappa_{ijk} &= n^{-1} E\left\{U_{ijk}\right\}, \\
\kappa_{i,jk} &= n^{-1} E\left\{U_i U_{jk}\right\}.
\end{aligned}$$

Suppose inverse of $\{\kappa_{i,j}\}$ matrix has entries $\{\kappa^{i,j}\}$. We assume all these quantities are of $O(1)$ as $n \to \infty$ and we employ the summation convention.

Results

$$
\begin{aligned}
&nE\left\{\psi(\widehat{z}_P(Y)\ ;\ Y,\theta) - \psi(z_P(Y)\ ;\ Y,\theta)\right\} \\
&\sim\ \phi(\Phi^{-1}(P))\Phi^{-1}(P)\left[-\frac{1}{2}\Phi^{-1}(P)^2\kappa^{i,j}\frac{\sigma_{0i}\sigma_{0j}}{\sigma_0^2}\right. \\
&\qquad +\kappa^{i,j}\kappa^{k,\ell}\left(\kappa_{jk,\ell}+\frac{1}{2}\kappa_{jk\ell}\right)\frac{\sigma_{0i}}{\sigma_0}+\frac{1}{2}\kappa^{i,j}\left\{\frac{\sigma_{0ij}}{\sigma_0}-\frac{\lambda_i^T V\lambda_j}{\sigma_0^2}\right\} \\
&\qquad \left.-\frac{1}{2}\kappa^{i,k}\kappa^{j,\ell}\cdot\frac{1}{n\sigma_0^2}\left(\lambda_i^T V\frac{\partial W}{\partial\theta^k}V\frac{\partial W}{\partial\theta^\ell}V\lambda_j+\lambda_i^T V\frac{\partial W}{\partial\theta^\ell}V\frac{\partial W}{\partial\theta^k}V\lambda_j\right)\right],
\end{aligned}
$$

$$
\begin{aligned}
&nE\left\{\psi(\tilde{z}_P(Y)\ ;\ Y,\theta) - \psi(z_P(Y)\ ;\ Y,\theta)\right\} \\
&\sim\ \phi(\Phi^{-1}(P))\Phi^{-1}(P)\left[\kappa^{i,j}\kappa^{k,\ell}\left(\kappa_{jk,\ell}+\kappa_{jk\ell}\right)\frac{\sigma_{0i}}{\sigma_0}\right. \\
&\qquad -\kappa^{i,j}\left(\frac{\sigma_{0i}\sigma_{0j}}{\sigma_0^2}-\frac{\sigma_{0ij}}{\sigma_0}\right)+\kappa^{i,j}\frac{\sigma_{0i}}{\sigma_0}Q_j \\
&\qquad \left.-\frac{1}{2}\kappa^{i,k}\kappa^{j,\ell}\cdot\frac{1}{n\sigma_0^2}\left(\lambda_i^T V\frac{\partial W}{\partial\theta^k}V\frac{\partial W}{\partial\theta^\ell}V\lambda_j+\lambda_i^T V\frac{\partial W}{\partial\theta^\ell}V\frac{\partial W}{\partial\theta^k}V\lambda_j\right)\right].
\end{aligned}
$$

$$nE\{\widehat{z}_P - z_P\} \;\approx\; \Phi^{-1}(P)\left\{\kappa^{i,j}\kappa^{k,\ell}\sigma_{0\ell}\left(\kappa_{ik,j} + \frac{1}{2}\kappa_{ijk}\right) + \frac{1}{2}\kappa^{i,j}\sigma_{0ij}\right\}$$

$$\begin{aligned} nE\{\tilde{z}_P - z_P\} \;\approx\; & \Phi^{-1}(P)\Big\{\kappa^{i,j}\kappa^{k,\ell}\sigma_{0\ell}(\kappa_{ik,j} + \kappa_{ijk}) \\ & + \kappa^{i,j}\left(\sigma_{0ij} - \frac{\sigma_{0i}\sigma_{0j}}{\sigma_0}\right) + \kappa^{i,j}Q_j\sigma_{0i} \\ & + \frac{1}{2}\Phi^{-1}(P)^2\kappa^{i,j}\frac{\sigma_{0i}\sigma_{0j}}{\sigma_0} + \frac{1}{2}\kappa^{i,j}\frac{\lambda_i^T V \lambda_j}{\sigma_0}\Big\}. \end{aligned}$$

We can find an estimator of z_P whose second-order CPB is 0 in two ways: either a Bayesian estimator with the matching prior, or directly, by

$$\begin{aligned} z_P^\dagger &= \widehat{z}_P - n^{-1}\Phi^{-1}(P)\left\{\widehat{\kappa}^{i,j}\widehat{\kappa}^{k,\ell}\widehat{\sigma}_{0\ell}\left(\widehat{\kappa}_{ik,j} + \frac{1}{2}\widehat{\kappa}_{ijk}\right)\right. \\ &\quad + \frac{1}{2}\widehat{\kappa}^{i,j}\left(\widehat{\sigma}_{0ij} - \frac{\widehat{\sigma}_{0i}\widehat{\sigma}_{0j}}{\widehat{\sigma}_0}\Phi^{-1}(P)^2\right) - \frac{1}{2\widehat{\sigma}_0}\widehat{\kappa}^{i,j}\widehat{\lambda}_i^T\widehat{V}\widehat{\lambda}_j \\ &\quad \left. - \frac{1}{2n\widehat{\sigma}_0}\widehat{\kappa}^{i,j}\widehat{\kappa}^{k,\ell}\left(\widehat{\lambda}_j^T\widehat{V}\frac{\partial\widehat{W}}{\partial\theta^i}\widehat{V}\frac{\partial\widehat{W}}{\partial\theta^k}\widehat{V}\widehat{\lambda}_\ell + \widehat{\lambda}_j^T\widehat{V}\frac{\partial\widehat{W}}{\partial\theta^k}\widehat{V}\frac{\partial\widehat{W}}{\partial\theta^i}\widehat{V}\widehat{\lambda}_\ell\right)\right\}. \end{aligned}$$

IV Application to Network Design

Suppose we use an estimator of z_P whose second-order CPB is 0 (e.g. either the Bayes estimator with matching prior, or $z_P^\dagger$). Use this to construct a two-sided prediction interval, with tail probability $1-P$ in each tail. The approximate expected length of this prediction interval is

$$2\Phi^{-1}(P)\sqrt{\sigma_0^2 + n^{-1}\kappa^{i,j}\lambda_i^T V \lambda_j + n^{-1}\Phi^{-1}(P)^2\kappa^{i,j}\sigma_{0i}\sigma_{0j}}.$$

In the notation of Zhu and Stein (2004), the quantity under the square root sign is

$$V_1 + \frac{\Phi^{-1}(P)^2}{4} \cdot \frac{V_2}{\sigma_0^2}.$$

Recall their own criterion was $V_3 = V_1 + \frac{1}{2} \cdot \frac{V_2}{\sigma_0^2}$.

Two formulae for V_3

$$V_3 = V_1 + \frac{1}{2} \cdot \frac{V_2}{\sigma_0^2} \quad \text{(Zhu and Stein)}$$

$$V_3 = V_1 + \frac{\Phi^{-1}(P)^2}{4} \cdot \frac{V_2}{\sigma_0^2} \quad \text{(this talk)}$$

The present formula has the unusual feature that the design might depend on the desired coverage probability of a prediction interval.

It is also tied directly to two specific methods of constructing a prediction interval whose second-order coverage probability bias is 0, whereas previous approaches have not shown how to construct such an interval.

V Summary and Conclusions

1. The second-order coverage probability bias of the Bayes estimator of z_P is smaller than that of the plug-in estimator in the limit as $P \to 0$ or 1, regardless of the prior.

2. For the Bayesian predictive distribution there is a matching prior, i.e. one for which the second-order CPB of $\tilde{z}_P$ is 0.

3. However we can also achieve the same second-order properties directly, using the estimator $z_P^\dagger$.

4. For any of these estimators of predictive quantiles, we have an approximation for the expected length of a prediction interval, and this can be used as a design criterion.

5. In the case of an estimate whose second-order CPB is 0, we obtain a design criterion very similar to that of Zhu and Stein, but adapted to a specific construction of a prediction interval.