BAYESIAN KRIGING AND BAYESIAN NETWORK DESIGN

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

J. Stuart Hunter Lecture
TIES 2004 / ACCURACY 2004 Joint Conference
Portland, Maine
June 30 2004

Acknowledgements to Zhengyuan Zhu,
Elizabeth Shamseldin and Joe Ibrahim

This talk is motivated by spatial interpolation problems that occur very widely in connection with environmental monitoring. There are two broad approaches to these problems:

(a) Geostatistical approach based on covariances or variograms

(b) Lattice approach based on conditional dependence structures

For whatever reasons, (a) is usually approached from a frequentist point of view and (b) from a Bayesian viewpoint.

The main purpose of this talk is to explore Bayesian approaches to (a).

Example 1

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.

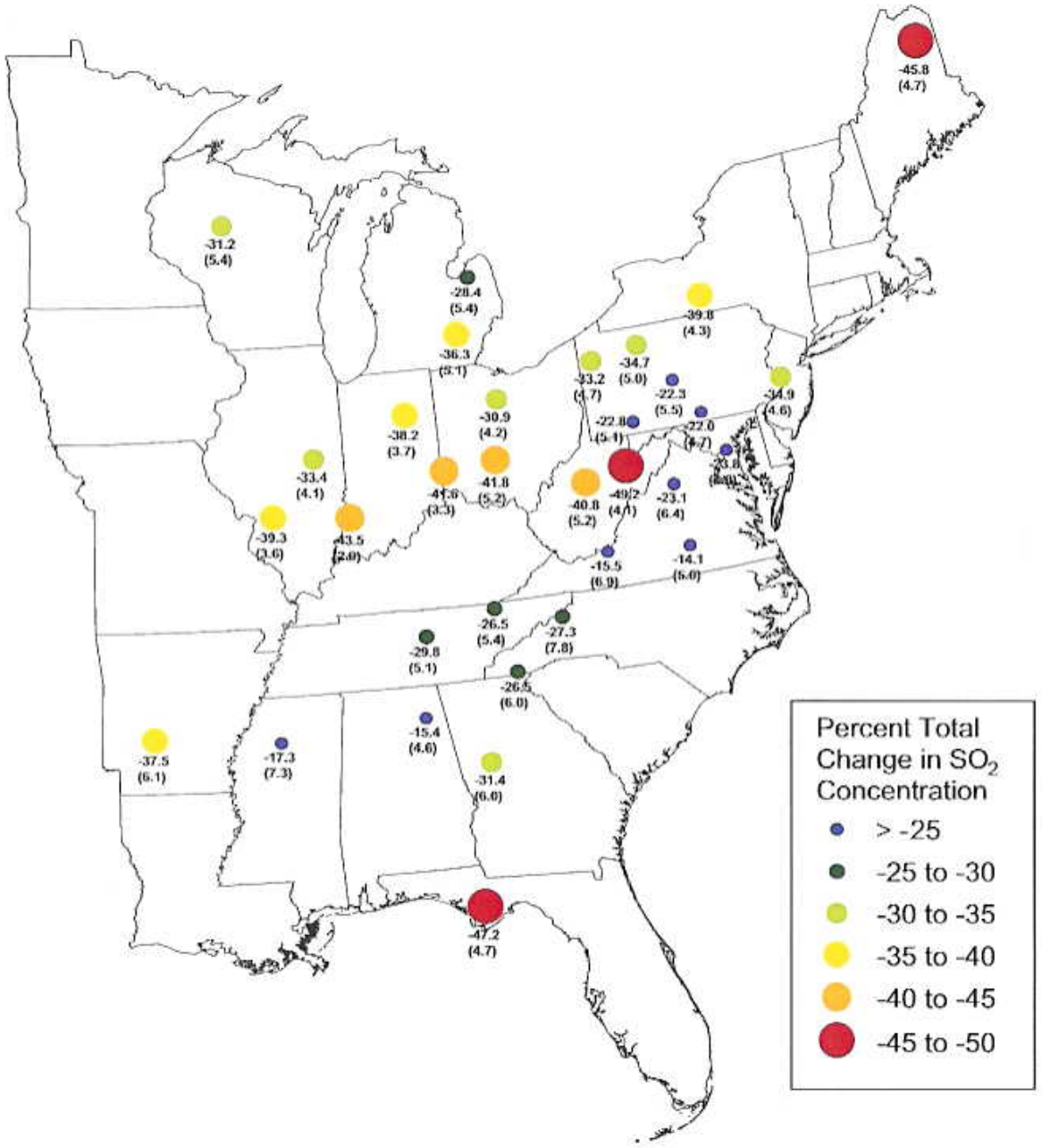


Fig. 1. Trend (percent change) in SO_2 concentrations at CASTNet sites from 1990 to 1999 with standard error (%) in parentheses.

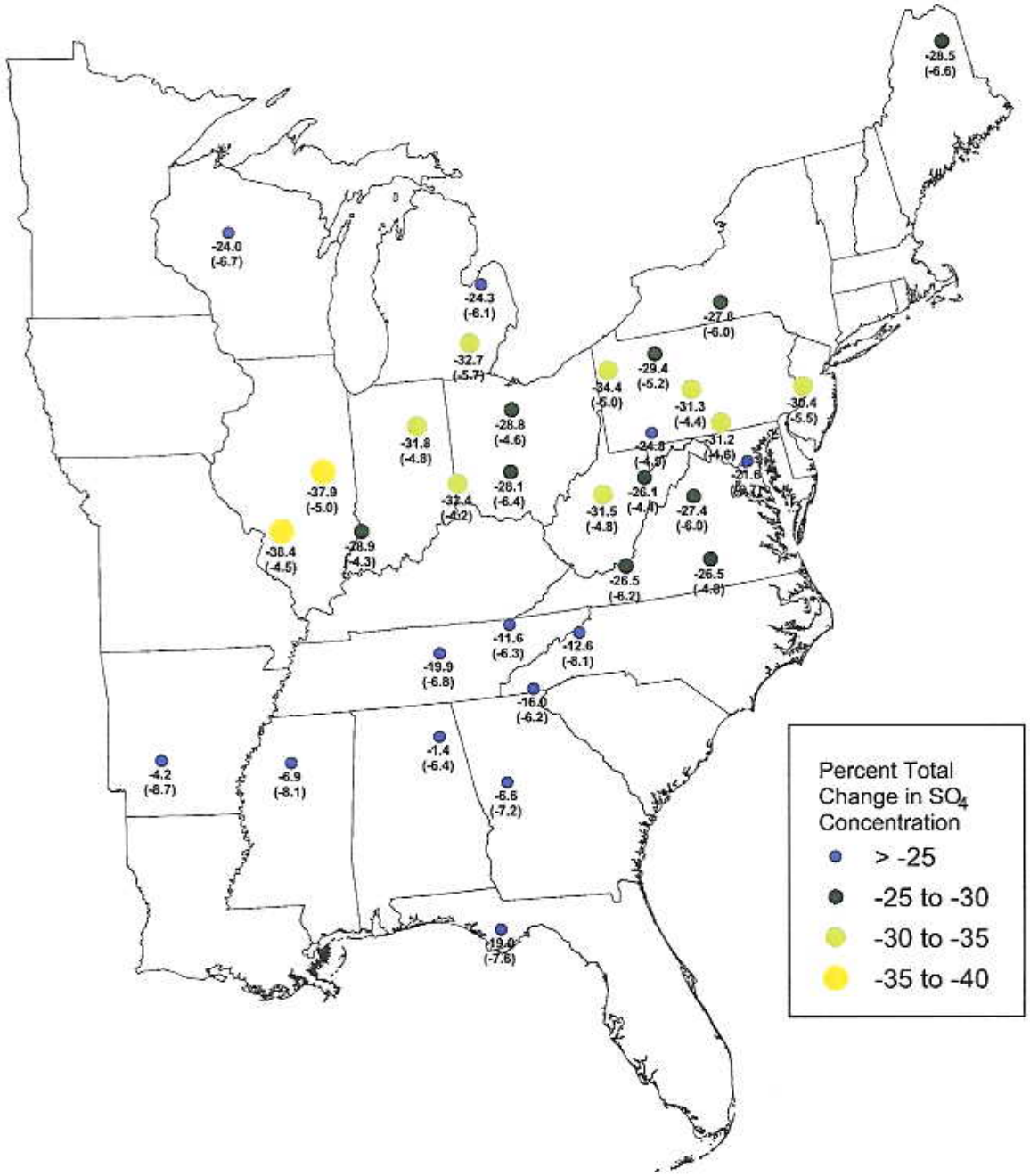


Fig. 2. Trend (percent change) in SO_4^{2-} concentrations at CASTNet sites from 1990 to 1999 with standard error (%) in parentheses.

By fitting a spatial model (treating the standard error of the trend at each site as an estimate of the measurement error), they constructed a kriged map of trends in SO_2 and SO_4^{2-} across the region, together with estimated prediction standard errors.

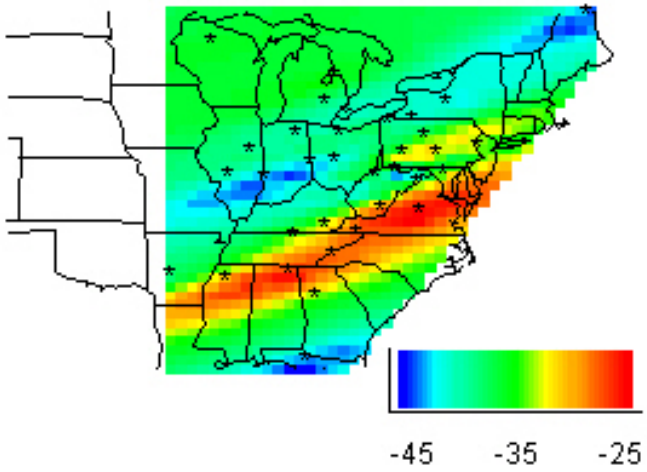
In the next two figures, we show the results constructed by two methods:

- Estimate the model parameters by restricted maximum likelihood (REML), then apply kriging treating these parameters as known.

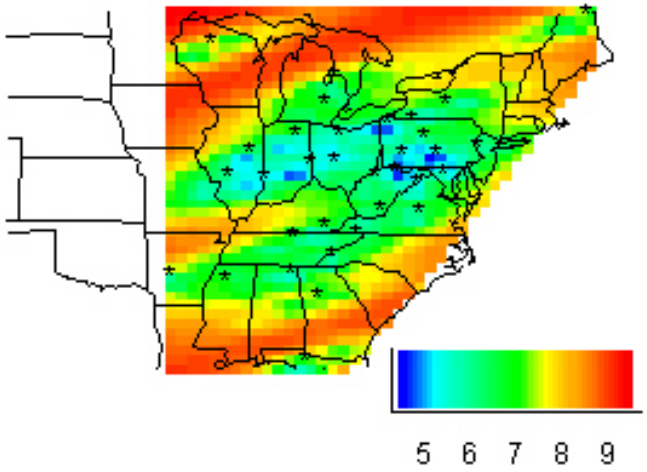
- Fully Bayesian approach, in which the predictive distribution derive from universal is integrated with respect to the posterior distribution

Interpolated Surfaces, MLE Method

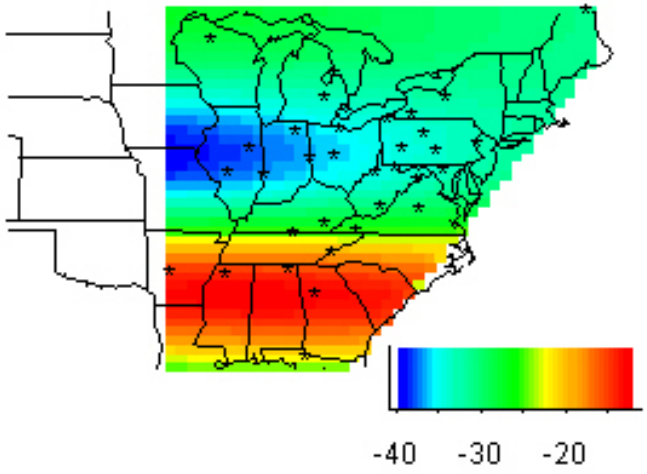
(a) SO2 Trend



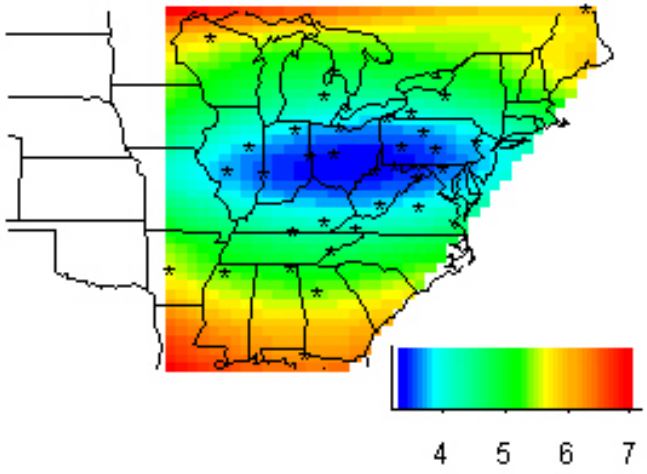
(b) SO2 SD



(c) SO4 Trend

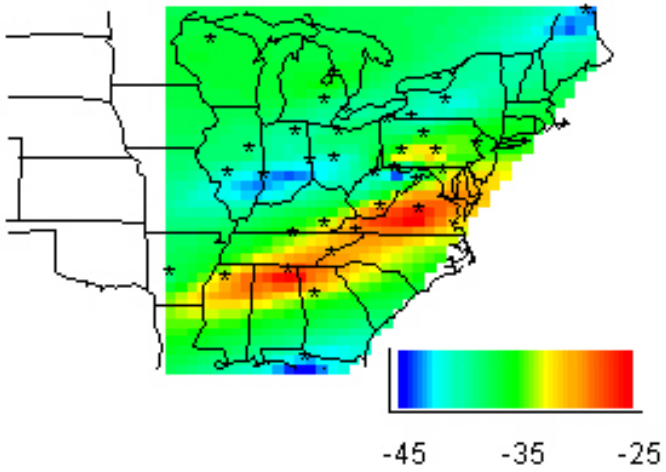


(d) SO4 SD

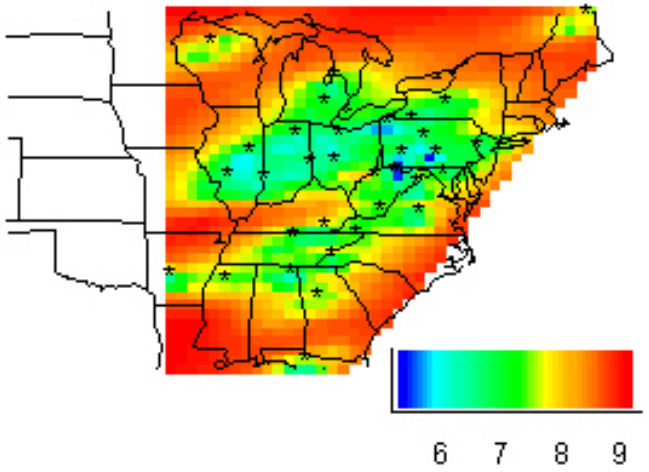


Interpolated Surfaces, Bayesian Method

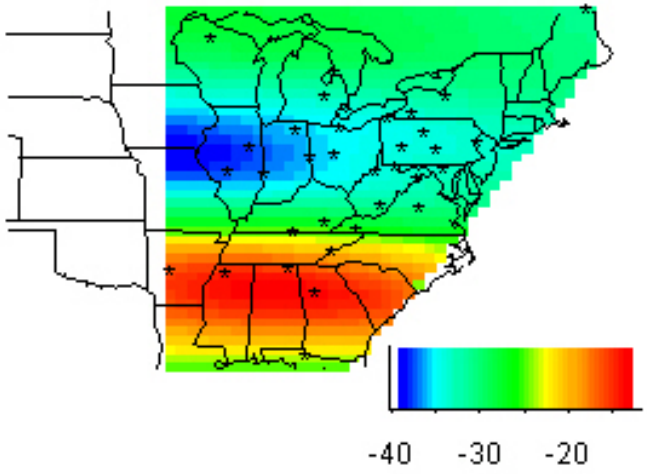
(a) SO2 Trend



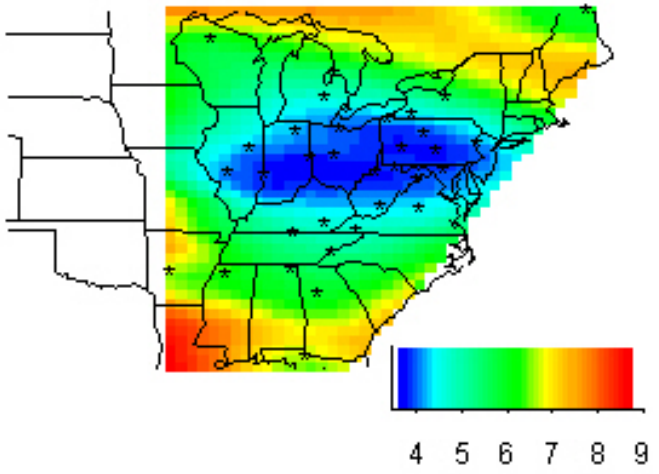
(b) SO2 SD



(c) SO4 Trend



(d) SO4 SD



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

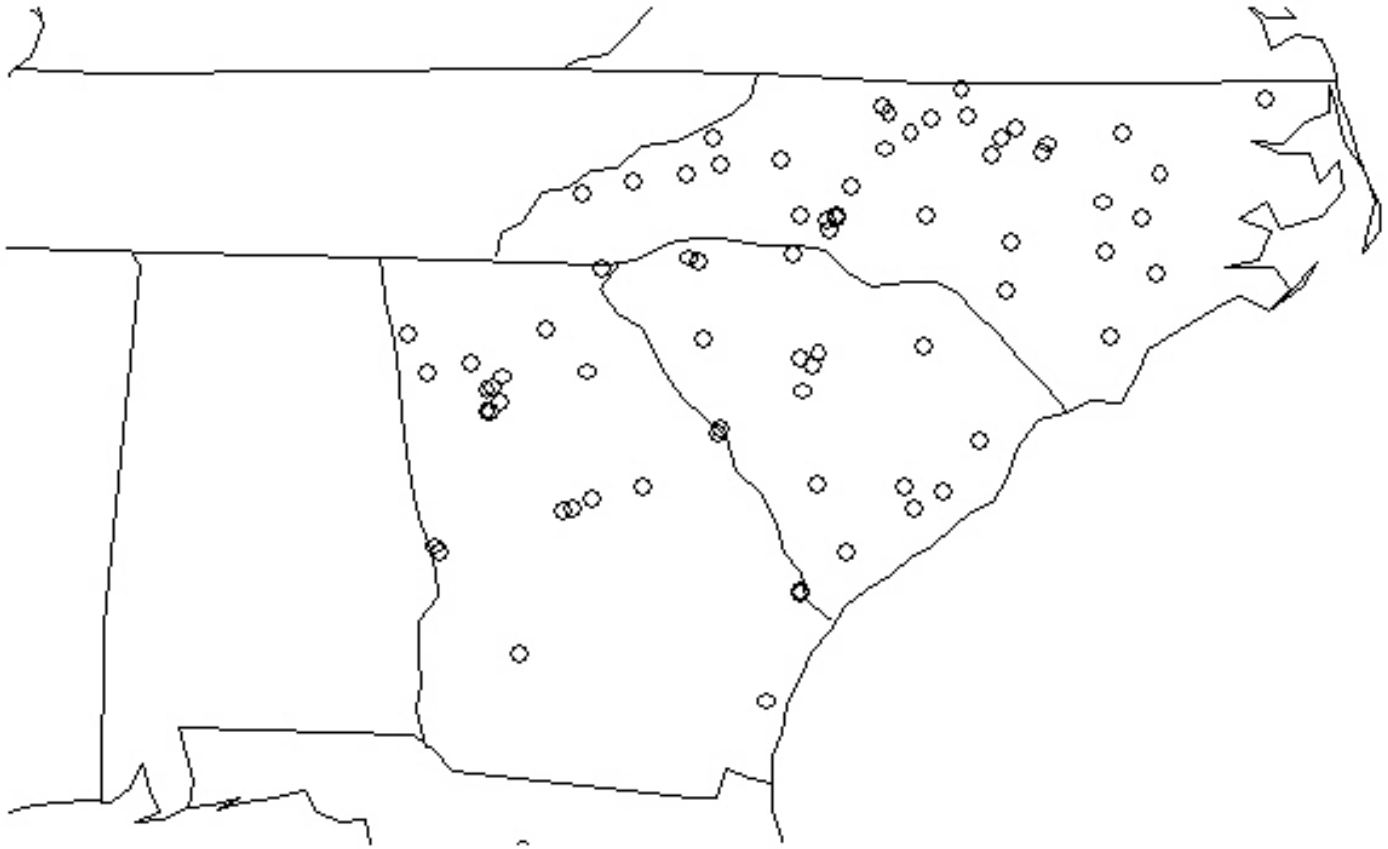
Example 2

Smith, Kolenikov and Cox (2003) developed a spatial-temporal model for $PM_{2.5}$ ("fine" particulate matter of diameter $\leq 2.5\mu$m), restricted to the states of North Carolina, South Carolina and Georgia, and the year 1999. The model used linear regression to represent fixed temporal and spatial means and a landuse effect, and spatially correlated random errors estimated with a power-law variogram.

They used weekly aggregated data from 74 stations, and constructed interpolated maps of $PM_{2.5}$ for week 33 (the week with the highest levels) and for the average over the whole year.

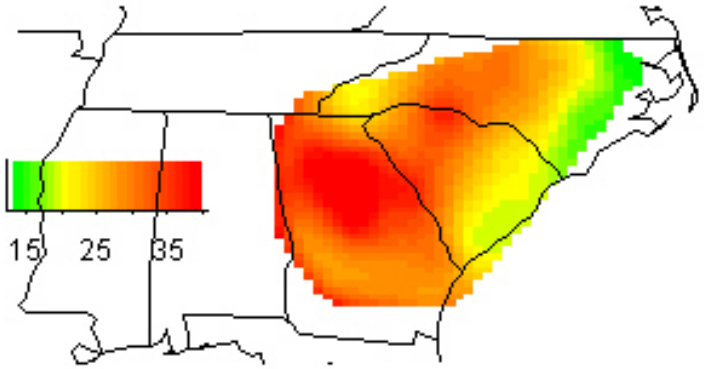
Although the published paper didn't include any Bayesian methodology, the present also includes a parallel Bayesian analysis, to illustrate the differences between the two. The prediction variances are typically about one-third higher using the Bayesian approach compared with the MLE.

Map of 74 Stations

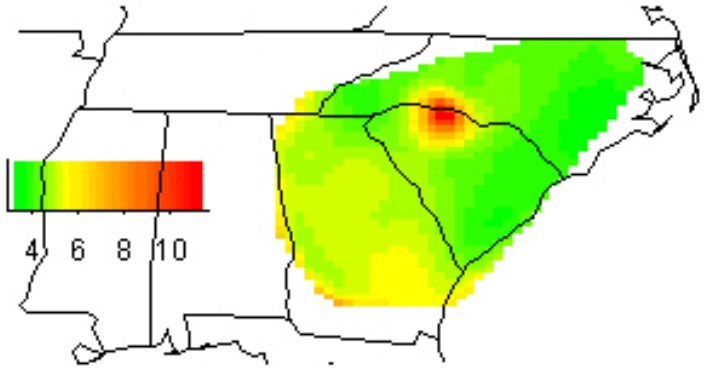


MLE Analysis

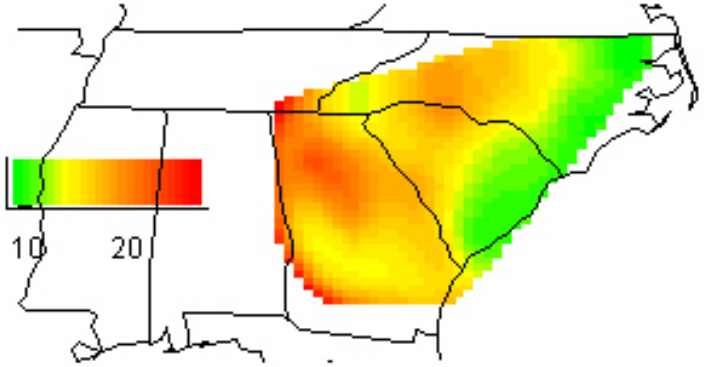
Week 33



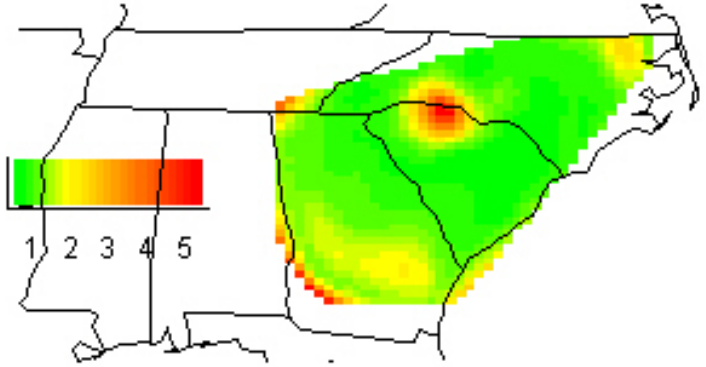
Week 33 SE



Annual Mean

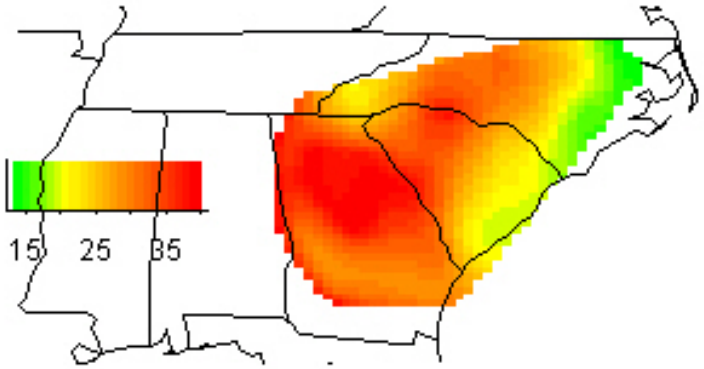


Annual Mean SE

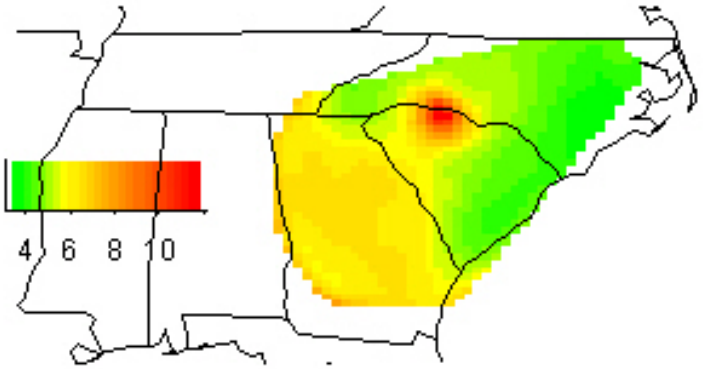


Bayes Analysis

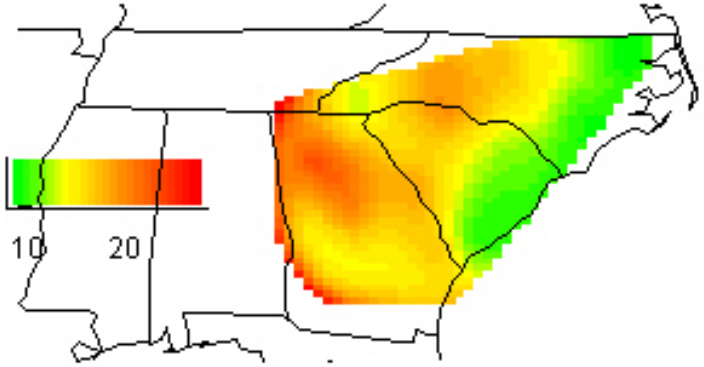
Week 33



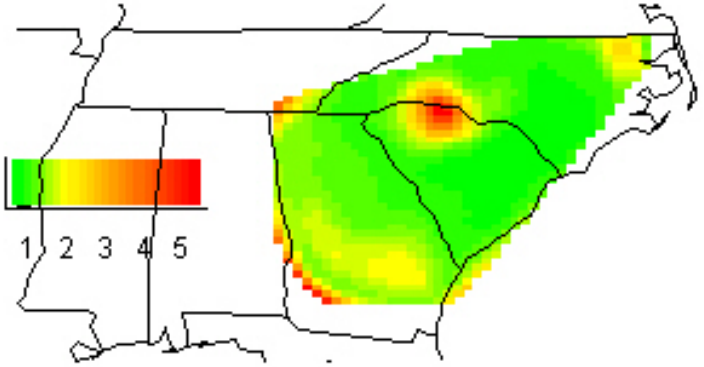
Week 33 SE



Annual Mean



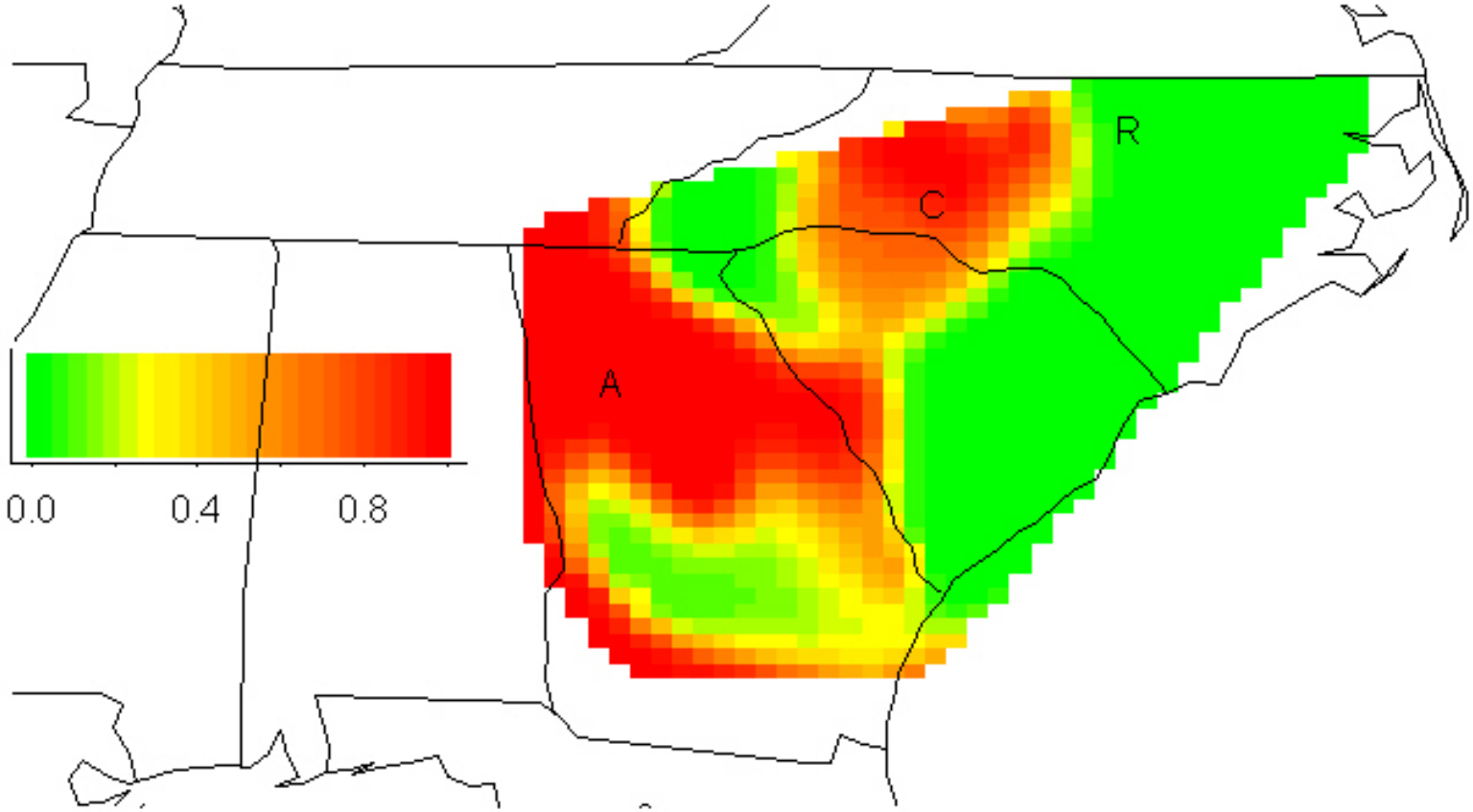
Annual Mean SE



It's also possible to interpret these results as the probability that the annual mean $PM_{2.5}$ at any location exceeds a threshold, which we have here taken to be 16.5 μm. Once again we have constructed spatial maps of this quantity using both the MLE and the Bayesian approach. The estimated probability is typically higher using the Bayesian method than the MLE.

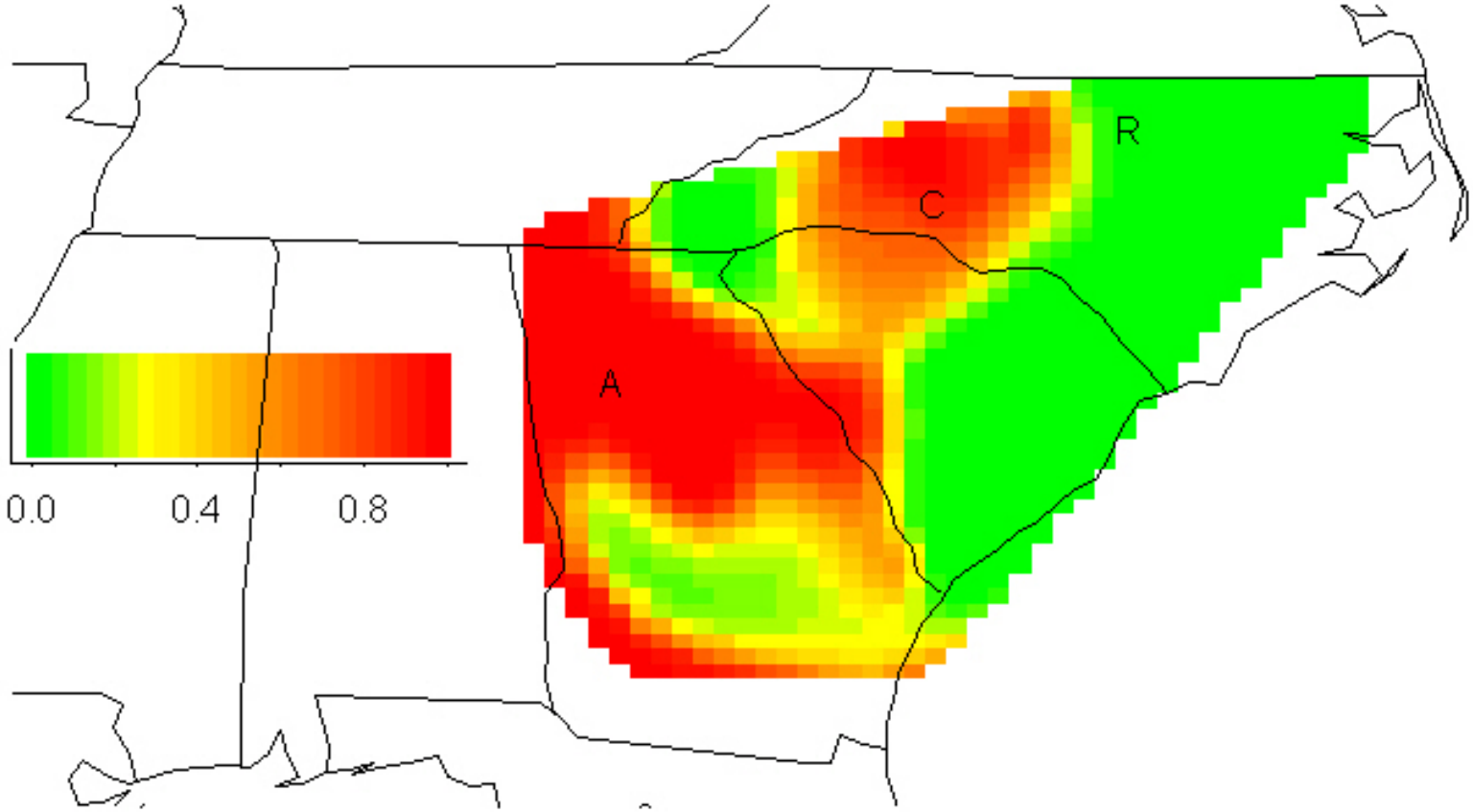
MLE Analysis

Probability of Exceeding Threshold 16.5



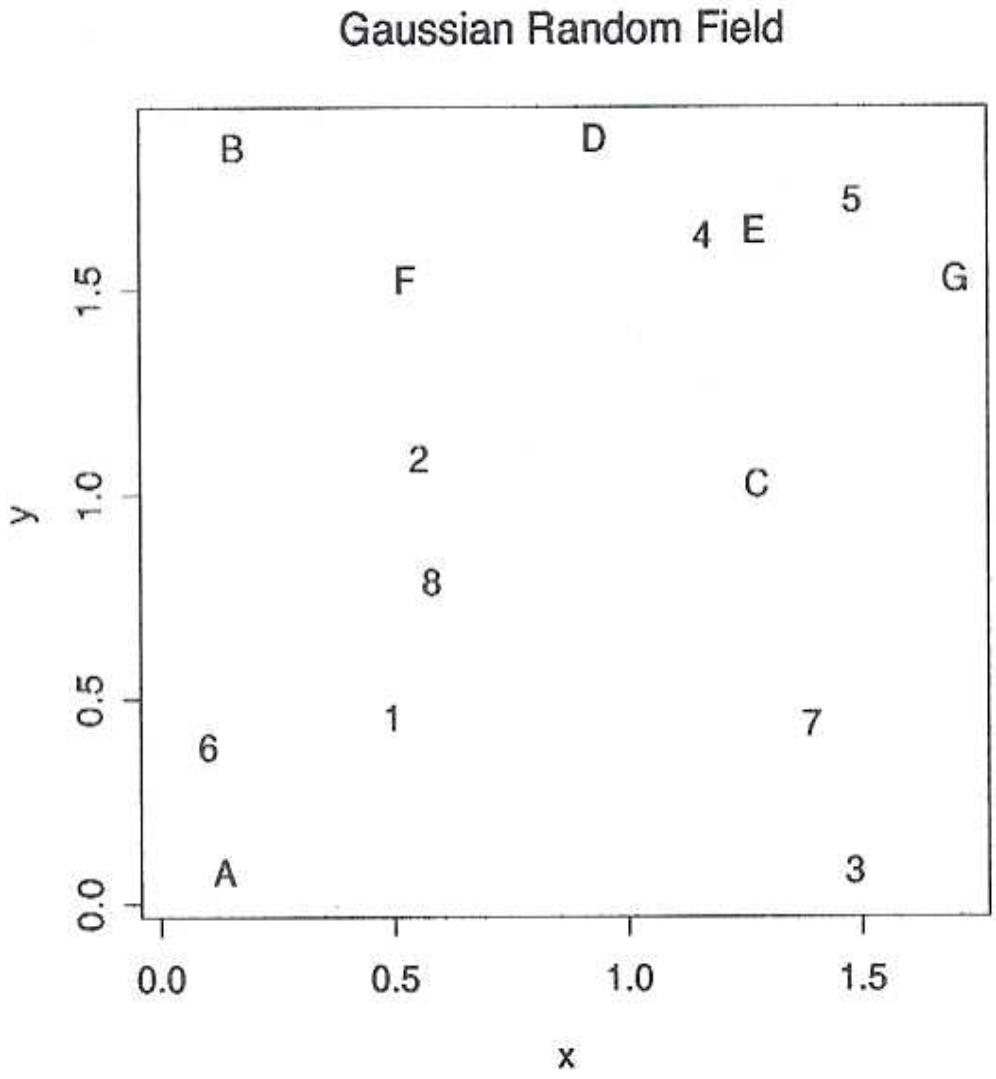
Bayes Analysis

Probability of Exceeding Threshold 16.5

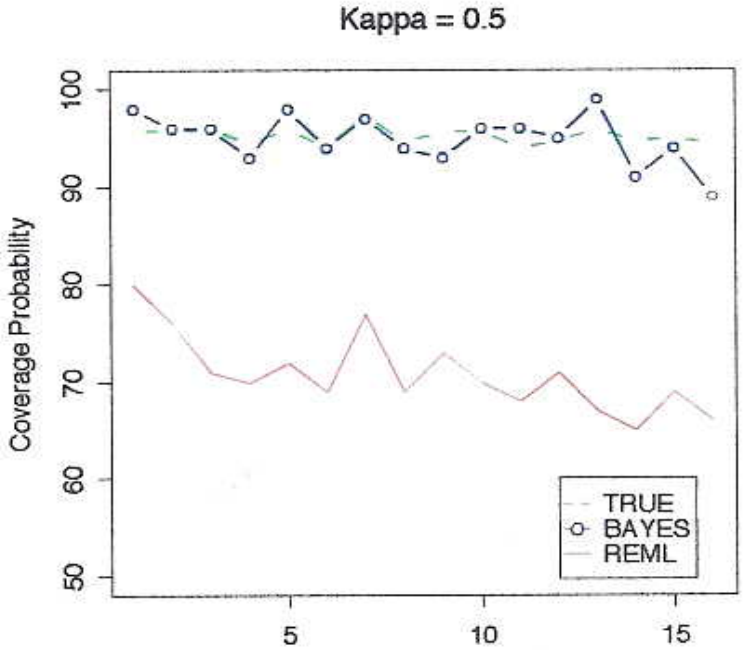


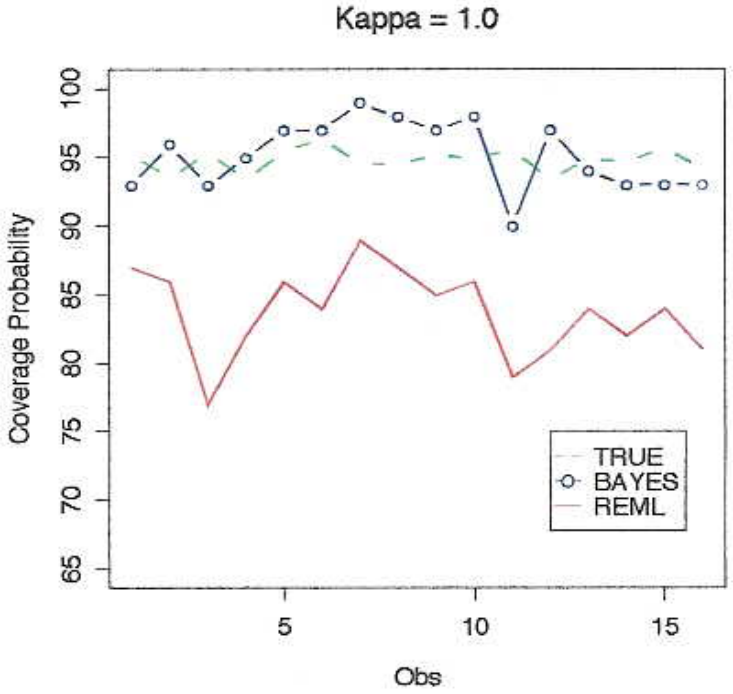
Next, I show some simulations produced by Elizabeth Shamseldin. She constructed a simulated random field on 16 locations, using an exponential-power correlation proportional to $\exp\left\{-\left(\frac{d}{\rho}\right)^{\kappa}\right\}$ where d is the distance between two locations, ρ is the range parameter, and $0 < \kappa \leq 2$. The model also assumed a spatial trend linear in the latitude and longitude coordinates.

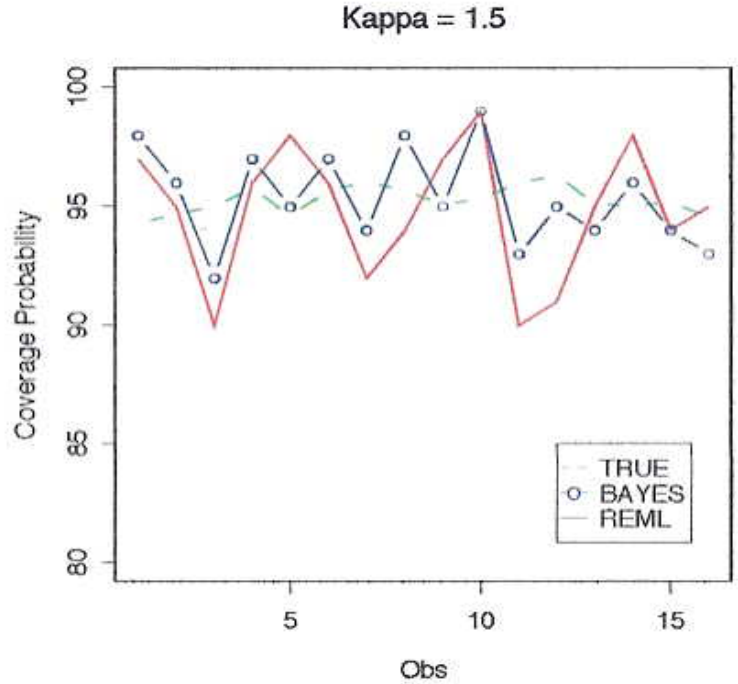
A cross-validation exercise was performed in which each location was omitted an predicted from the other 15 data points. This was repeated 100 times to calculate an empirical coverage probability for the nominal 95% prediction intervals. Three methods were used: universal kriging based on the known model parameters, a "plug-in" approach using the REML estimators, and a fully Bayesian analysis.



Sampling configuration for simulation experiment (E. Shamseldin)







Simulated coverage probabilities for nominal 95% PIs

In these simulations, we can see that the Bayesian prediction intervals have coverage probabilities close to their nominal 95% values, but the prediction intervals based on plugging the REML estimates into the universal kriging formula have too small a coverage probability, especially for small κ.

The undercoverage of plug-in prediction intervals is well known, and several proposals have been made over the years. Harville and Jeske (1992), Zimmerman and Cressie (1992) proposed approximations to the prediction variance based on the delta method, and Stein (1999) developed a number of alternatives.

From various points of view, it is attractive to use Bayesian methods in this context, but there is no proof in general that they actually solve the problem of undercoverage of prediction intervals.

Bayesian network design

(Joint work with Zhengyuan Zhu)

There are many criteria for design of a monitoring network, including

- Entropy approaches (J. Zidek and co-authors)
- Extensions of classical design criteria such as D-optimality and A-optimality (Federov and W. Müller (1989); Müller (2000); Berliner, Lu and Snyder (1999))
- "Pure Bayesian" approaches (Sansó and P. Müller)
- Designs that attempt to trade off between "optimality for prediction" and "optimality for estimation" (Zimmerman; Zhu)

An alternative suggestion (containing elements of all of the above):

Since Bayesian prediction intervals supposedly "correct" for the model parameters being unknown, we might use the expected length of a Bayesian prediction interval, directly, as a criterion for experimental design.

Practical point: It is extremely unlikely, in practice, that anyone would agree on a single "variable of interest" whose prediction forms the sole purpose of a network design. However, the same method could be applied to a suite of possible prediction variables, using either the average or the maximum length of a Bayesian prediction interval as the criterion for the overall design. This is quite close to what is already done in several network design criteria.

What would be needed for a brute force implementation of this 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

Zhu (2002, PhD thesis, University of Chicago) dismissed this possibility as too computationally intensive, and instead proposed various approximate schemes

Basics of Spatial Prediction

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

$$\begin{pmatrix} Y \\ Y_0 \end{pmatrix} \sim N\left[\begin{pmatrix} X\beta \\ x_0^T\beta \end{pmatrix}, \begin{pmatrix} V & w^T \\ w & v_0 \end{pmatrix}\right] \tag{1}$$

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. For the moment, we assume V, w and v_0 are known.

Where notationally convenient, we also define $Y^* = \begin{pmatrix} Y \\ Y_0 \end{pmatrix}$ and write (1) as

$$Y^* \sim N[X^*\beta, V^*]. \tag{2}$$

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,

$$C_\theta(d) = \sigma \exp\left\{-\left(\frac{d}{\rho}\right)^\kappa\right\}, \tag{3}$$

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

$$C_\theta(d) = \frac{\sigma}{2^{\nu-1}\Gamma(\nu)}\left(\frac{2\nu^{1/2}d}{\rho}\right)^\nu \mathcal{K}_\nu\left(\frac{2\nu^{1/2}d}{\rho}\right), \tag{4}$$

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

Estimation

Model of form

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

where the unknown parameters are (β, θ) and $V(\theta)$ is a known function of finite-dimensional parameters θ.

Methods of estimation:

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

2. MLE: choose (β, θ) to maximize log likelihood function

$$-\frac{1}{2}\log|V| - \frac{1}{2}(Y - X\beta)V^{-1}(Y - X\beta).$$

Maximize w.r.t. β first, set $\widehat{\beta} = (X^T V^{-1} X)^{-1} X^T V^{-1} Y$, then choose θ to maximize

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

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

3. Restricted maximum likelihood (REML estimation). Theoretically, this is defined as maximum likelihood estimation in the subset of data that lies orthogonal to the X-space. In practice, $\widehat{\theta}$ is chosen to maximize

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

The Main Prediction Problem

Assume model (1) where the covariances $V,\ w,\ v_0$ are known but β is unknown. The classical formulation of *universal kriging* asks for a predictor $\hat{Y}_0 = \lambda^T Y$ that minimizes $\sigma_0^2 = E\left\{(Y_0 - \hat{Y}_0)^2\right\}$ subject to the unbiasedness condition $E\left\{Y_0 - \hat{Y}_0\right\} = 0$.

The classical solution:

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

Bayesian Reformulation of Universal Kriging

Assume the model (1) or equivalently (2). Suppose β (the only unknown parameter, for the moment) has a prior density which is assumed uniform across $\mathcal{R}^p$. The Bayesian predictive density of Y_0 given Y is then

$$p(Y_0 \mid Y) = \frac{\int f(Y^* \mid \beta) d\beta}{\int f(Y \mid \beta) d\beta}. \tag{5}$$

After some manipulation, this may be rewritten in the form

$$p(Y_0 \mid Y) = \frac{1}{\sqrt{2\pi\sigma_0^2}} \exp\left\{ -\frac{1}{2} \left(\frac{Y_0 - \lambda^T Y}{\sigma_0} \right)^2 \right\} \tag{6}$$

Thus, in the case where β is the only unknown, we have rederived universal kriging as a Bayesian predictor. Moreover, because of the usual (frequentist) derivation of universal kriging, it follows that in this case, Bayesian procedures have exact frequentist properties.

Now consider the case where θ is also unknown. We assume θ has a prior density $\pi(\theta)$, independent of β.

The Bayesian predictive density of Y_0 given Y is now

$$\begin{aligned} p(Y_0 \mid Y) &= \frac{\int\int f(Y^* \mid \beta, \theta)\pi(\theta)d\beta d\theta}{\int\int f(Y \mid \beta, \theta)\pi(\theta)d\beta d\theta} \\ &\vdots \\ &= \frac{\int e^{\ell_n(\theta)}\psi(\theta)\pi(\theta)d\theta}{\int e^{\ell_n(\theta)}\pi(\theta)d\theta} \end{aligned} \tag{7}$$

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 (7) as $\tilde{\psi}$, to distinguish it from the plug-in rule $\widehat{\psi} = \psi(\widehat{\theta})$.

Bullet Point Number 1

Universal kriging has a nice Bayesian interpretation, as the exact solution of a Bayesian prediction problem for the linear model parameters, but only under the assumption that the covariance parameters are *known* — an unrealistic assumption in practice.

When the covariance parameters are unknown, the REML estimator has the interpretation of a *maximum a posteriori* estimator, after integrating out the linear model parameters. However, the traditional plug-in approach to prediction does not have a rigorous Bayesian interpretation.

Solution of (7): Use *Laplace approximation*.

First, some notation. Let

$$\begin{aligned} U_i &= \frac{\partial \ell_n(\theta)}{\partial \theta^i}, \\ U_{ij} &= \frac{\partial^2 \ell_n(\theta)}{\partial \theta^i \partial \theta^j}, \\ U_{ijk} &= \frac{\partial^3 \ell_n(\theta)}{\partial \theta^i \partial \theta^j \partial \theta^k}, \end{aligned}$$

where θ^i, θ^j... denote components of the vector θ.

Suppose inverse of $\{U_{ij}\}$ matrix has entries $\{U^{ij}\}$.

We shall introduce other quantities such as $Q(\theta) = \log \pi(\theta)$ and $\psi(\theta) = \frac{1}{\sqrt{2\pi\sigma_0^2}} \exp\left\{-\frac{1}{2}\left(\frac{Y_0 - \lambda^T Y}{\sigma_0}\right)\right\}$ that are functions of θ, and where needed, we use suffixes to denote partial differentiation, for example $Q_i = \partial Q/\partial\theta^i$, $\psi_{ij} = \partial^2\psi/\partial\theta^i\partial\theta^j$. All these quantities are evaluated at the true θ unless denoted otherwise. The REML estimator is denoted $\widehat{\theta}$ with components $\widehat{\theta}^i$. The REMLE of ψ is $\widehat{\psi} = \psi(\widehat{\theta})$. Any expression with a hat on it, such as $\widehat{U}_{ijk}$, means that it is to be evaluated at the REMLE $\widehat{\theta}$ rather than the true value θ.

Using *summation convention*, define

$$\mathcal{D} = \frac{1}{2} U_{ijk} U^{ik} U^{j\ell} \psi_\ell - \frac{1}{2} (\psi_{ij} + 2\psi_i Q_j) U^{ij} \tag{8}$$

and let $\widehat{\mathcal{D}}$ denote the same expression where all terms have hats. With these conventions, Smith (1997, 1999) used Laplace's intergal approximation to derive the formula

$$\tilde{\psi} = \hat{\psi} + \widehat{\mathcal{D}}, \tag{9}$$

accurate to $O_p(n^{-1})$.

We can also apply (9) to the *predictive distribution function*

$$\psi(z; Y, \theta) = \Phi\left(\frac{z - \lambda^T Y}{\sigma_0}\right)$$

and invert the resulting $\hat{\psi}$ or $\tilde{\psi}$ to compute prediction intervals.

Bullet Point Number 2

Laplace integration is a plausible alternative to MCMC. It is not clear whether it is a better approximation to the true posterior distribution than MCMC, but it could be much more convenient to apply.

Parenthetical remark: A different form of Laplace approximation was proposed by Tierney and Kadane (1988). The present version is more convenient if the objective is to compute the predictive distribution function at many values of z.

Mathematical properties of Bayesian predictors (Cox 1975; Smith 1997, 1999)

Suppose $\psi^*(z)$ (could be $\widehat{\psi}$ or $\tilde{\psi}$) is an estimator of the conditional prediction distribution function $\psi(z) = \psi(z \ ; \ Y, \theta)$ that has an expansion

$$\psi^*(z) \ = \ \psi(z) + n^{-1/2}R + n^{-1}S + o(n^{-1}). \qquad (10)$$

Define predictive quantile z_P^* by $\psi^*(z_P^*) = P$. Then

$$z_P^* - z_P \ = \ -n^{-1/2}\frac{R}{\psi'} - n^{-1}\left(\frac{RR'}{\psi'^2} - \frac{R^2}{\psi'^3} - \frac{S}{\psi'}\right) + o_p(n^{-1}), \qquad (11)$$

$$\psi(z_P^*) - \psi(z_P) \ = \ -n^{-1/2}R - n^{-1}\left(\frac{RR'}{\psi'} - S\right) + o_p(n^{-1}). \qquad (12)$$

By taking expectations in (11) and (12) respectively, we derive expressions for the expected length of a prediction interval and the coverage probability bias (CPB).

More notation:

$$\begin{aligned}
\kappa_{i,j} &= n^{-1}E\left\{U_iU_j\right\}, \\
\kappa_{ij} &= n^{-1}E\left\{U_{ij}\right\} = -\kappa_{i,j}, \\
\kappa_{ijk} &= n^{-1}E\left\{U_{ijk}\right\}, \\
\kappa_{i,jk} &= n^{-1}E\left\{U_iU_{jk}\right\}, \\
W &= V^{-1} - V^{-1}X(X^TV^{-1}X)^{-1}X^TV^{-1}, \\
\lambda_i &= \frac{\partial\lambda}{\partial\theta^i}, \quad \lambda_{ij} = \frac{\partial^2\lambda}{\partial\theta^i\partial\theta^j}, \\
\sigma_{0i} &= \frac{\partial\sigma_0}{\partial\theta^i}, \quad \sigma_{0ij} = \frac{\partial^2\sigma_0}{\partial\theta^i\partial\theta^j}.
\end{aligned}$$

Suppose inverse of $\{\kappa_{i,j}\}$ matrix has entries $\{\kappa^{i,j}\}$. We assume all these quantities are of $O(1)$ (or $O_p(1)$) as $n \to \infty$.

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

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

Bullet Point Number 3

Even if...

- you subscribe to the belief that "subjective probability does not exist";
- you haven't a clue how to choose a prior distribution;
- you can't or won't use MCMC (because Laplace approximation is an alternative);

you should still use Bayesian methods to solve spatial interpolation problems with estimated parameters.

Expected length of a prediction interval

(This page has been corrected since the original version of the talk)

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

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

For a $100(P_2 - P_1)\%$ Bayesian prediction interval,

$$\begin{aligned} E\left\{\tilde{z}_{P_2}(Y) - \tilde{z}_{P_1}(Y)\right\} &= E\left\{z_{P_2}(Y\ ;\ \theta) - z_{P_1}(Y\ ;\ \theta)\right\} \\ &\quad + E\left\{\tilde{z}_{P_2}(Y) - z_{P_2}(Y\ ;\ \theta)\right\} \\ &\quad - E\left\{\tilde{z}_{P_1}(Y) - z_{P_1}(Y\ ;\ \theta)\right\} \\ &= \sigma_0\{\Phi^{-1}(P_2) - \Phi^{-1}(P_1)\} \\ &\quad + E\left\{\tilde{z}_{P_2}(Y) - z_{P_2}(Y\ ;\ \theta)\right\} \\ &\quad - E\left\{\tilde{z}_{P_1}(Y) - z_{P_1}(Y\ ;\ \theta)\right\} \end{aligned}$$

Application to network design

Recall from earlier discussion, a "brute force" approach to Bayesian network design would require three nested computations, (a) MCMC to calculate a predictive distribution, (b) simulation to determine the expected length of a Bayesian prediction interval, (c) simulated annealing (or some equivalent method) to determine the optimal design.

The methods described here create the possibility of replacing (a) and (b) by a single analytic approximation, so that only (c) needs iterative numerical methods.

Bullet Point Number 4

The expected length of a Bayesian prediction interval may be used as the basis of a network design criterion, combining "predictive" and "estimative" approaches to design. The new approximation takes into account the prior density, as well as the desired coverage level of the prediction interval. Because it is formally justified as the $O\left(\frac{1}{n}\right)$ approximation to the exact expected length, it may also have a more rigorous mathematical justification than the earlier approaches.

Conclusions

1. Universal kriging is a very elegant technique for spatial prediction in the present of unknown linear regression parameters, but it fails to allow correctly for the spatial covariance parameters also being unknown.

2. Bayesian prediction intervals provide an alternative to this, with numerical solution by MCMC or Laplace methods.

3. Asymptotic results suggest that Bayesian prediction intervals have better sampling properties than conventional approaches, regardless of the prior.

4. The same ideas suggest a new design criterion for monitoring networks.

Thank you to TIES for the opportunity to give this talk!