

Bayesian spatial prediction

In the conventional geostatistical approaches for interpolation, i.e. kriging, the covariance structure is estimated first, and then the **estimated covariance is used for interpolation**. The properties of the interpolants based on an estimated covariance structure are not well understood, and it is common practice to **ignore the effect of the uncertainty** in the covariance structure on subsequent predictions.

A Bayesian approach to interpolation of spatial processes will provide a general methodology for **taking into account the uncertainty** about parameters on subsequent predictions.

The prediction problem can be stated in the following form: given observations of a vector field,

$$Z = \{Z(\mathbf{x}_0), Z(\mathbf{x}_2), \dots, Z(\mathbf{x}_n)\},$$

predict the value $Z(\mathbf{x}_0) = z_0$, for some $\mathbf{x}_0 \notin \{\mathbf{x}_1, \dots, \mathbf{x}_n\}$. We can take two approaches to this: an approach based on Lagrange multipliers, and a Bayesian approach.

The Lagrange multiplier approach is the most direct derivation of the kriging estimator and is the one most commonly given in textbooks.

The Bayesian approach leads to the **same answers** as the standard kriging predictor when the model parameters are known, but it also extends to the case where these parameters are unknown.

The Bayesian approach generalizes automatically to the case in which the variogram parameters are unknown, whereas the classical approach essentially makes the assumption that these are known and only deals with the question of uncertainty of model parameters in a very peripheral way.

This is one major reason for viewing the problem in Bayesian terms, and the close parallels between this and the more traditional approaches.

When the goal is to predict $Z(\mathbf{x}_0) = z_0$, the Bayesian solution is the posterior predictive distribution of $Z(\mathbf{x}_0)$ given the observations Z ,

$$\pi(Z(\mathbf{x}_0)|Z) \propto \int \pi(Z(\mathbf{x})|Z, \phi) \pi(\phi|Z) d\phi. \quad (12)$$

where $\pi(\phi|Z)$ is the posterior of the model parameters. In the kriging prediction, the model parameters $\phi = (\beta, \alpha, \theta)$ are estimated using a likelihood approach or using empirical methods, and then they are treated as known for the subsequent prediction.

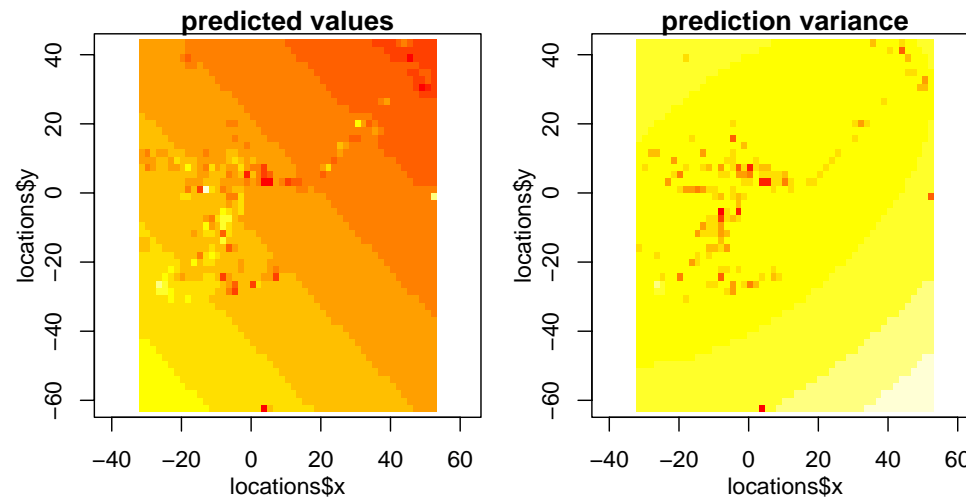


Figure 6: Bayesian kriging for fish data, and estimated variance of the posterior predictive distribution

The model throughout this discussion is the same as in (1). The choice of $\pi(\theta)$ is largely arbitrary, but the equivalence of Bayesian and least-squares approach works only for the “classical” noninformative prior on (β, α) .

Simplest case: β, α, θ all known

The distribution of z_0 given the data and all parameters is:

$$\{z_0|Z, \beta, \alpha, \theta\} \sim \mathcal{N}[(x_0 - X^T \Sigma^{-1} \tau)^T \beta + \tau^T \Sigma^{-1} Z, \sigma_0^2 - \tau^T \Sigma^{-1} \tau]. \quad (13)$$

We shall now improve upon (13) by, successively, removing the conditioning on β , α and θ . We write $\pi(x|y)$ for the generic density of one variable x conditioned on another variable y where the variables x and y will be different from one usage to the next.

To remove the conditioning on β , we write

$$\pi(z_0|Z, \alpha, \theta) = \int \pi(z_0|Z, \beta, \alpha, \theta)\pi(\beta|Z, \alpha, \theta)d\beta \quad (14)$$

Note that the posterior distribution of β , given Z , α and θ , is multivariate normal with mean $\hat{\beta} = \hat{\beta}(\theta)$ (i.e. the GLS estimator of β given the covariance matrix $V(\theta)$), and covariance matrix $\alpha(X^T V(\theta)^{-1} X)^{-1}$ (when the prior for β is uniform).

Then, the conditional distribution of z_0 given α and θ is multivariate normal with mean

$$\begin{aligned} \hat{z}_0(\theta) &= (x_0 - X^T \Sigma^{-1} \tau)^T \hat{\beta} + \tau^T \Sigma^{-1} Z \\ &= (x_0 - X^T V(\theta)^{-1} w(\theta))^T \hat{\beta} + w(\theta)^T V(\theta)^{-1} Z \end{aligned} \quad (15)$$

and covariance matrix

$$\begin{aligned}
& (x_0 - X^T \Sigma^{-1} \tau)^T (X^T \Sigma^{-1} X)^{-1} (x_0 - X^T \Sigma^{-1} \tau) + \sigma_0 - \tau^T \Sigma^{-1} \tau \\
&= \alpha \{ (x_0 - X^T V(\theta)^{-1} w(\theta))^T (X^T V(\theta)^{-1} X)^{-1} (x_0 - X^T V(\theta)^{-1} w(\theta)) \\
&\quad + v_0(\theta) - \tau^T V(\theta)^{-1} \tau \} \\
&= \alpha V_0(\theta) \quad \text{say.}
\end{aligned} \tag{16}$$

The next step is to remove the conditioning on α .

We have,

$$\pi(z_0 | Z, \theta) = \int \pi(z_0 | Z, \alpha, \theta) \pi(\alpha | Z, \theta) d\alpha. \tag{17}$$

Define

$$\hat{\alpha}(\theta) = \frac{G^2(\theta)}{n - q}.$$

Then with slight abuse of notation, we have

$$(\alpha|Z, \theta) \sim \hat{\alpha}(\theta) \frac{n - q}{\chi_{n-q}^2}.$$

Conditionally on Z and θ , we then have

$$\begin{aligned} \frac{z_0 - \hat{z}_0(\theta)}{\sqrt{\hat{\alpha}(\theta)V_0(\theta)}} &= \frac{z_0 - \hat{z}_0(\theta)}{\sqrt{\alpha V_0(\theta)}} \cdot \sqrt{\frac{\alpha}{\hat{\alpha}(\theta)}} \\ &\sim \frac{\mathcal{N}(0, 1)}{\sqrt{\chi_{n-q}^2/(n - q)}} \\ &\sim t_{n-q} \end{aligned} \tag{18}$$

since the numerator and denominator are conditionally independent given Z and θ .

Finally, we integrate over θ to obtain

$$\pi(z_0|Z) = \int \pi(z_0|Z, \theta)\pi(\theta|Z)d\theta \quad (19)$$

This part has to be carried out numerically, but should be straightforward since for most models of interest the dimension of θ is 2 or at most 3.

Approximating the PPD

The predictive distribution can be approximated by the *Rao-Blackwellized estimator*:

$$\pi(z_0|Z) = \frac{1}{m} \sum_{i=1}^m \pi(z_0|Z, \theta^{(i)})$$

where $\theta^{(i)}$ constitute the i -th draw from the posterior distribution.

Algorithm to simulate from the posterior distribution

The Hastings-Metropolis algorithm. We start with an arbitrary $\mathbf{x}^{(0)}$ and generate a new “trial value” \mathbf{x}' from some distribution $q(\mathbf{x}'; \mathbf{x}^{(0)})$ which depends on $\mathbf{x}^{(0)}$. Then form the ratio

$$\alpha = \frac{q(\mathbf{x}^{(0)}; \mathbf{x}')l(\mathbf{x}')}{q(\mathbf{x}'; \mathbf{x}^{(0)})l(\mathbf{x}^{(0)})}.$$

If $\alpha \geq 1$ then we accept \mathbf{x}' ; in other words, set $\mathbf{x}^{(1)} = \mathbf{x}'$. If $\alpha < 1$, we perform an independent random drawing: with probability α , accept \mathbf{x}' and set $\mathbf{x}^{(1)} = \mathbf{x}'$; otherwise, reject \mathbf{x}' and set $\mathbf{x}^{(1)} = \mathbf{x}^{(0)}$.

Gibbs Sampling

We describe now another algorithm to efficiently generate the simulations from the posterior of a vector parameter ϕ . The most convenient methods are of Markov chain Monte Carlo (MCMC) type, of which Gibbs sampling is one of the most widely used.

Gibbs sampling. Start with an arbitrary initial value for the vector parameter $\phi^{(0)} = (\phi_1^{(0)}, \dots, \phi_n^{(0)})$. Generate a new value of ϕ_1 , denoted $\phi_1^{(1)}$, from the conditional distribution of Φ_1 given $\Phi_2 = \phi_2^{(0)}, \dots, \Phi_n = \phi_n^{(0)}$. Then generate a new value of ϕ_2 , denoted $\phi_2^{(1)}$, from the conditional distribution of Φ_2 given $\Phi_1 = \phi_1^{(1)}, \Phi_3 = \phi_3^{(0)}, \dots, \Phi_n = \phi_n^{(0)}$. Continue up to the generation of $\phi_n^{(1)}$ from the conditional distribution of Φ_n given $\Phi_1 = \phi_1^{(1)}, \dots, \Phi_{n-1} = \phi_{n-1}^{(1)}$. This completes one iteration of the sampler. Then, starting from the new vector $\phi^{(1)}$, return to ϕ_1 and repeat the whole process to generate $\phi^{(2)}$.

Thus, Gibbs sampling consists purely in sampling from conditional distributions, because instead of updating ϕ *in block*, it is more computationally efficient to divide ϕ into components and then update these components one by one.

SO₂ Concentrations

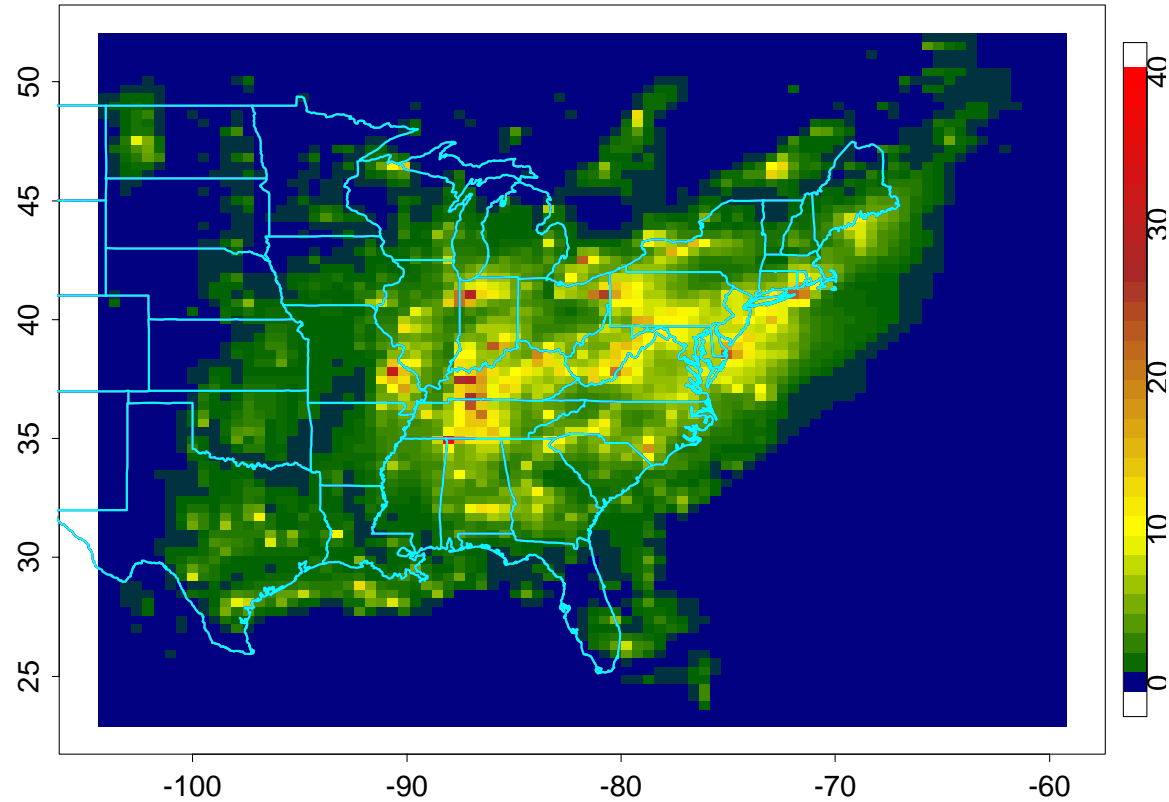


Figure 7: Output of Models-3, weekly average of SO_2 concentrations (ppb), for the week of July 11, 1995. The resolution is 36 km^2 .

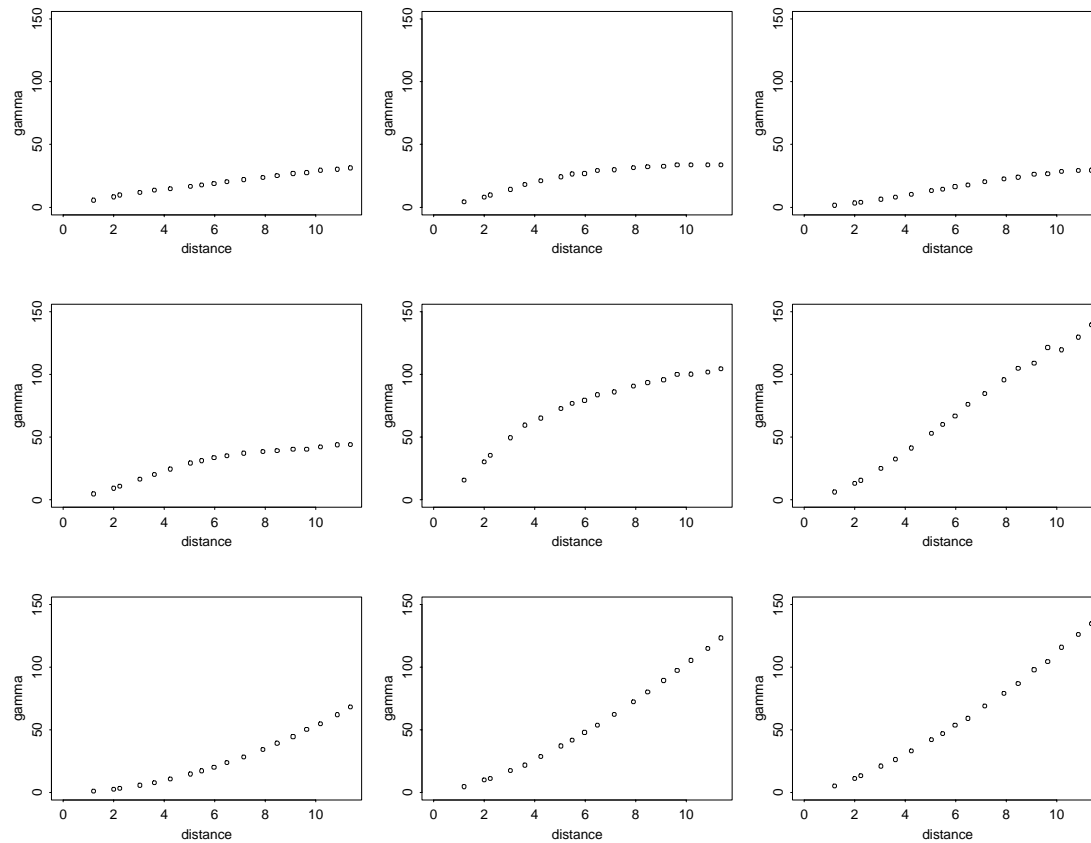


Figure 8: We divided the domain shown in the previous figure in 9 subregions and we calculated the empirical semivariograms. There is clear evidence of lack of stationarity.

SO₂ concentrations (CASTNet)

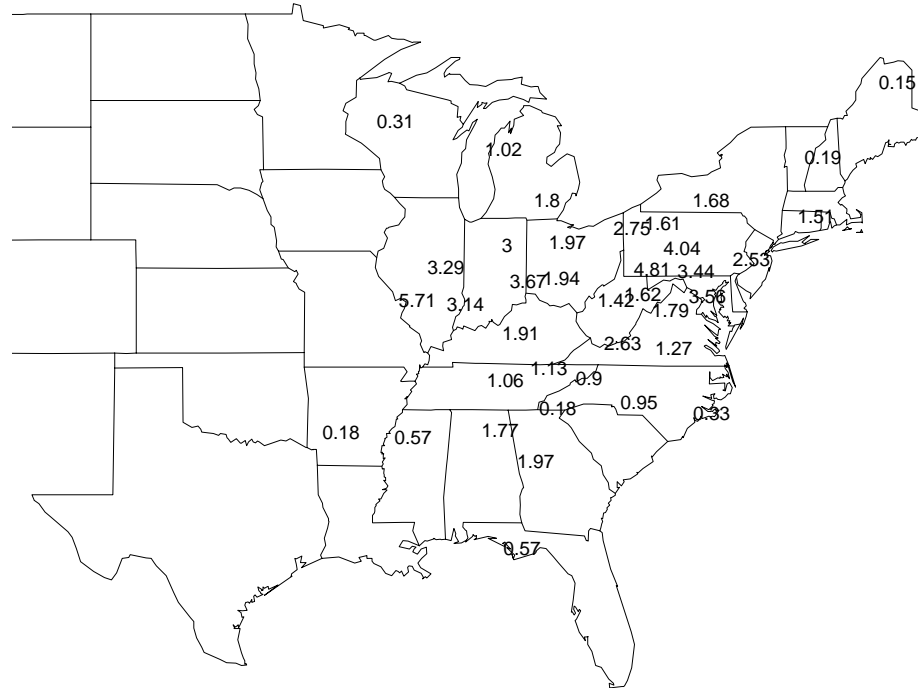


Figure 9: Weekly average of SO_2 concentrations (ppb) at the Clean Air Status and Trends Network (CASTNet) sites, for the week of July 11, 1995.

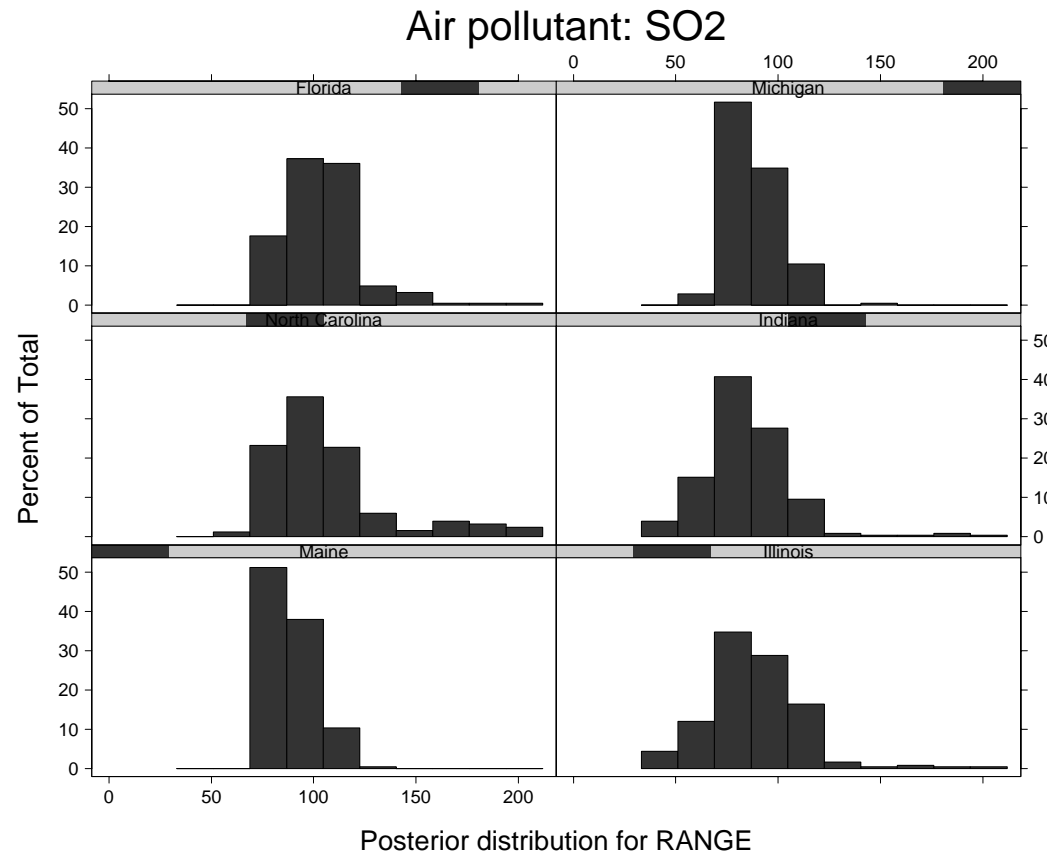


Figure 10: Posterior distributions for the range parameter (km) of the Matérn covariance for Models-3 SO_2 concentrations, for the week starting July 11, 1995. At 6 selected locations.

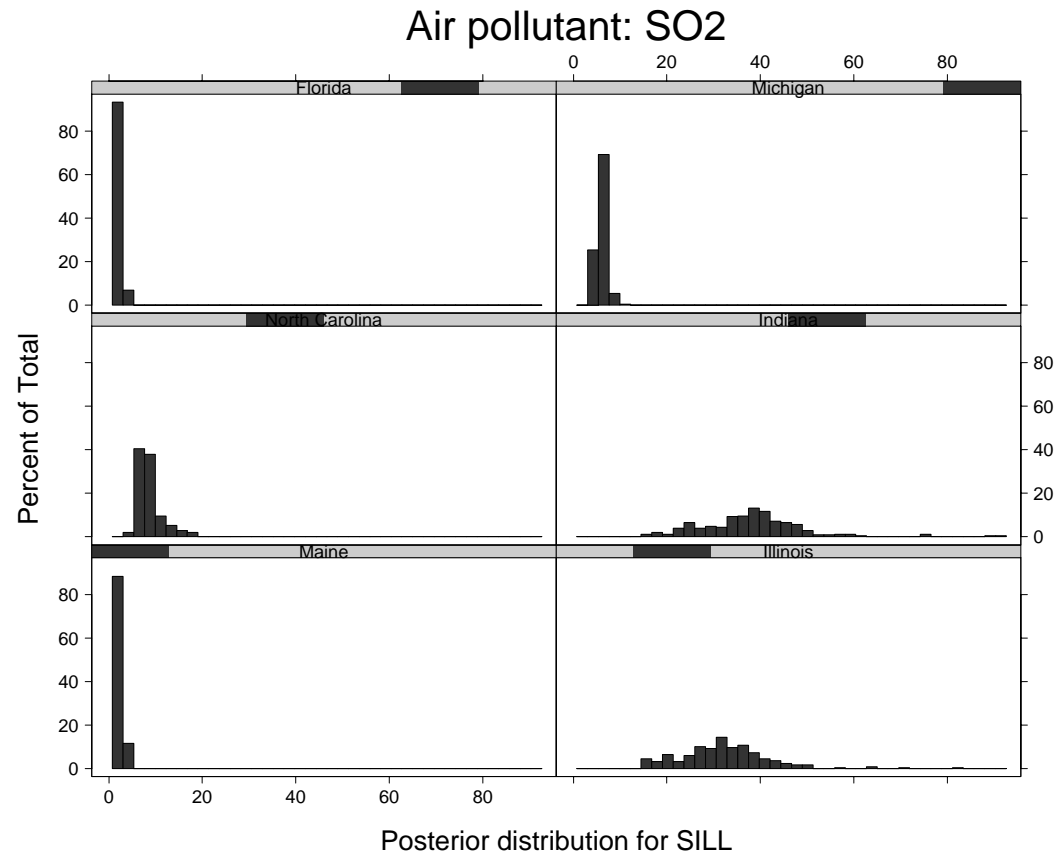


Figure 11: Posterior distributions for the sill parameter of the Matérn covariance for Models-3 SO_2 concentrations, for the week starting July 11, 1995. At the 6 selected locations.

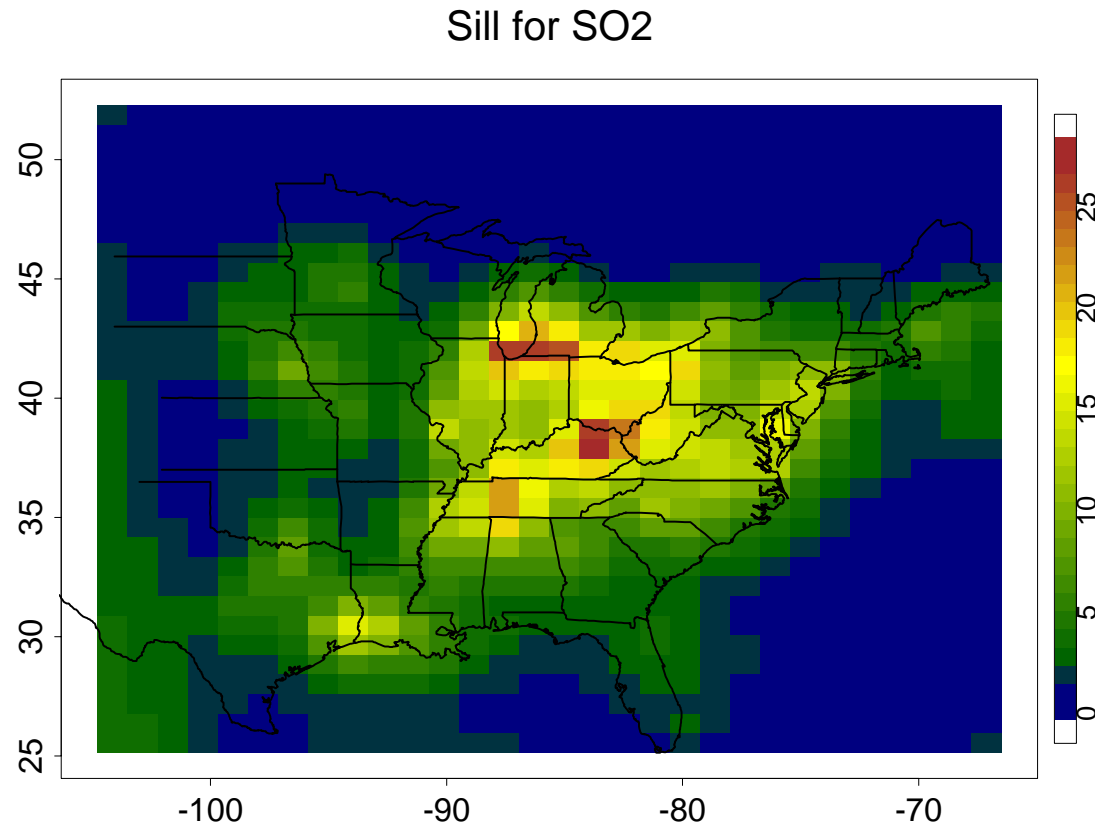
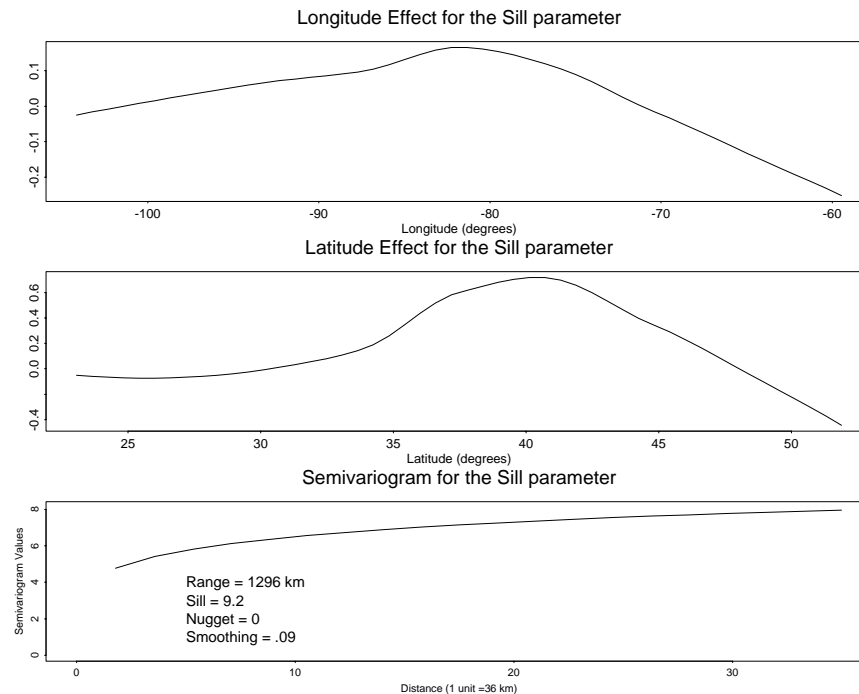


Figure 12: Map of the modes of the posterior distributions for the sill parameter of the Matérn covariance for Models-3 SO_2 concentrations, for the week starting July 11, 1995.



Longitude effect (r_i 's), latitude effect (c_j 's) and semivariogram of the error term (ϵ_σ) for the sill parameter (σ), where $\sigma(s_i, s_j) = a + r_i + c_j + \epsilon_\sigma(s_i, s_j)$ and the process ϵ_σ has a Matérn stationary covariance with parameters $\boldsymbol{\tau}_0$. In the bottom graph we plot the **Matérn covariance model** of the the error term (ϵ_σ), the covariance parameters for this Matérn model are the modes of the posterior for the hyperparameters $\boldsymbol{\tau}_0$.

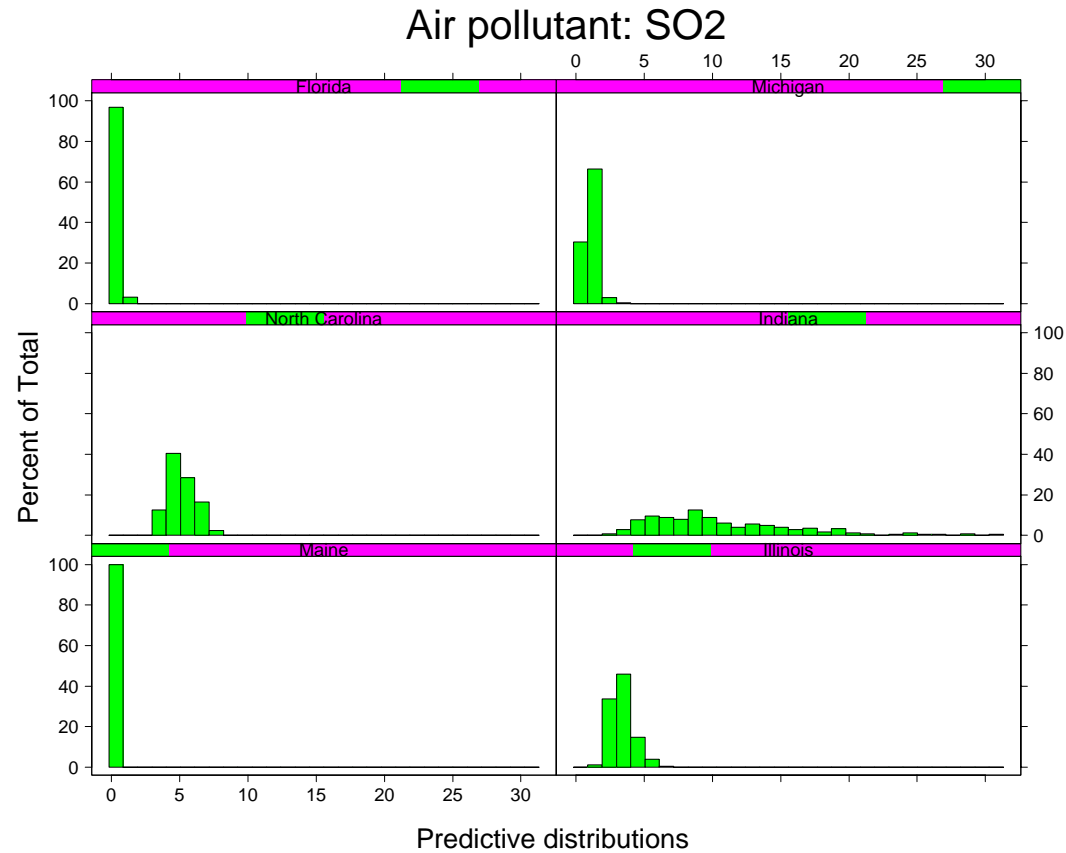


Figure 13: Predictive distributions for the Models-3 SO_2 concentrations, at the 6 selected locations, for the week starting July 11, 1995.