

Spatial Stochastic Volatility for Lattice Data

Jun Yan

Department of Statistics and Actuarial Science, University of Iowa,

Iowa City, IA 52242, U.S.A.

March 22, 2006

Abstract

Spatial heteroskedasticity may arise jointly with spatial autocorrelation in lattice data collected from agricultural trials and environmental studies. This leads to spatial clustering not only in the level but also in the variation of the data, the latter of which may be very important, for example, in constructing prediction intervals. This article introduces a spatial stochastic volatility (SSV) component into the widely used conditional autoregressive (CAR) model to capture the spatial clustering in heteroskedasticity. The SSV component is a mean zero, conditionally independent Gaussian process given a latent spatial process of the variances. The logarithm of the latent variance process is specified by an intrinsic Gaussian Markov random field. The SSV model relaxes the traditional homoskedasticity assumption for spatial heterogeneity and brings greater flexibility to the popular spatial statistical models. The Bayesian method is used for inference. The full conditional distribution of the heteroskedasticity components can be shown to be log-concave, which facilitates an adaptive rejection sampling algorithm. Application to the well-known wheat yield data illustrates that incorporating spatial stochastic volatility may reveal the spatial heteroskedasticity hidden from existing analyses.

KEY WORDS: Conditional autoregressive model; Spatial heteroskedasticity; Markov chain Monte Carlo; Stochastic volatility; Wheat yield data

1. INTRODUCTION

Spatial heteroskedasticity may arise jointly with spatial autocorrelation in agricultural trials or environmental studies, where data are typically collected over a regular or irregular lattice. Spatial autocorrelation implies local clustering in the level of the variable under observation and has been studied extensively in the literature. Spatial heteroskedasticity occurs when the variability of the spatial data is higher in one area than another. In contrast to the enormous literature on spatial autocorrelation, the modeling of spatial heteroskedasticity is a much less developed field. For point referenced data, Cowles and Zimmerman (2003) introduced spatial heteroskedasticity by allowing the variance of the error process to differ across subregions. This method falls into the category of group-wise heteroskedasticity. In the econometrics literature, Kelegian and Robinson (1998) suggested tests for spatial heteroskedasticity, assuming that the variance can be modeled by a set of regressors. This method seems unsatisfactory, as it requires a specification of the causes of the changing variance, which may not be possible in general. With the usual assumption of spatial continuity, we expect that the change in variance should occur smoothly over space. A preferable model should, therefore, allow the heteroskedasticity to exhibit a smoothly varying spatial structure, leading to local clustering in the variability. This article, motivated by a further investigation in fitting the well-known wheat yield data of Mercer and Hall (1911), aims to propagate a model which allows joint modeling of spatial autocorrelation and spatial heteroskedasticity within the Bayesian framework.

The wheat yield data of Mercer and Hall (1911) is perhaps the best analyzed lattice dataset in the spatial statistics literature. The data were obtained from a uniformity trial, where all the plots are treated with the same treatment in preparation for future trials. The data consists of wheat yield on a 20×25 lattice of plots, with approximately one acre in total area. The 20 rows run in the east-west direction and the 25 columns run in the north-south direction. The original objective of the experiment was to determine the optimal plot size which would reduce the error within working limits. Such trials today are used to model spatial variation, that is, spatial covariance and possible spatial trend, across the lattice so as to improve the efficiency of the fixed effect estimators. One important aspect of spatial variation is the magnitude of the volatility, which can be very different

from one area to another, making spatial volatility a direct concern. Information about spatial volatility is critical in constructing prediction intervals. For example, on a larger scale, it can be very useful for insurance agencies to assess the variation of yields on insured land.

In the non-Bayesian framework, the wheat yield data has witnessed many of the important contributions to the development of the Gaussian Markov random field (GMRF) models, including the simultaneous autoregression (SAR) (Whittle 1954) and the conditional autoregression (CAR) (Besag 1974, 1977). However, these GMRF models did not fit very well, due at least in part to the nonstationarity in mean and covariance. There has been continuing effort to improve the unsatisfactory fit. To remove the mean nonstationarity before fitting GMRF models, Cressie (1993) detrended the data with median polishing, a robust method that iteratively corrects the data by row and column medians until a stop criterion is met. More recently, Dreesman and Tutz (2001) proposed a spatially varying coefficient model with a locally smoothed pseudolikelihood estimator to handle mean nonstationarity as well as covariance nonstationarity. They considered the spatially varying coefficient model as a useful exploratory tool and found “new strong support for the suspicion that the wheat yield data are not stationary” (Dreesman and Tutz 2001, p.13). Their findings shed lights on desired model properties to improve the fit in a Bayesian framework.

Voluminous literature has been devoted to hierarchical Bayesian modeling and analysis for spatial data. A widely used spatial areal model is given by Besag, York, and Mollié (1991). The model assumes that the detrended small scale variation is composed of two terms: one capturing regional clustering and the other representing unstructured regional heterogeneity. The unstructured regional heterogeneity terms are assumed to be independent and identically distributed normal variables with mean zero. For the wheat yield data, the findings in Dreesman and Tutz (2001) suggest that a good model should address the covariance nonstationarity. When a model of Besag et al. (1991) is fit to the detrended wheat yield data (see more details in Section 5), we find that the unstructured regional heterogeneity terms may not have the same variance across the whole field. In particular, the plot of the square of these terms in Figure 1 shows that variation is much higher in a small region around the center of the field than in other regions, and the variation tends to be locally clustered. The squared unstructured regional heterogeneity terms are found to be

spatially correlated from an exploration analysis using both the Moran and the Geary statistics; see more details in Section 5. This implies that the regional heterogeneity may have local clustering in their variation, and may not be completely “unstructured”. Therefore, one way to extend Besag et al. (1991)’s model is simply to allow the regional heterogeneity terms to have spatially varying variance, as supported by the exploratory analysis.

In this article, we introduce a spatial stochastic volatility (SSV) component into the widely used hierarchical spatial model of Besag et al. (1991) to capture the spatial clustering in heteroskedasticity. The basic idea is adapted from the time series analog. The SSV component is a mean zero, conditionally independent Gaussian process given a latent spatial process of the variances. To ensure positivity of the variance and allow spatial clustering, we use a GMRF to model the logarithm of the latent variance. This GMRF can be intrinsic as for the spatial clustering effect in Besag et al. (1991). With the SSV component, the new spatial model is a more intuitive alternative in the Bayesian framework to the spatially varying coefficient model of Dreesman and Tutz (2001) by providing posterior surfaces of the spatial heteroskedasticity as well as the spatial clustering effect.

The rest of the paper is organized as follows. In Section 2, we briefly review the traditional model of Besag et al. (1991) and then discuss strategies to fix its limitation in handling spatial heteroskedasticity. The proposed SSV model is introduced in Section 3. Implementation details for Bayesian computing are discussed in Section 4. In Section 5, the classic wheat yield data (Mercer and Hall 1911) is reanalyzed using the SSV model and the results are compared to those obtained from the traditional model Besag et al. (1991). The SSV model is found to pick up the spatial clustering of heteroskedasticity hidden from the existing analysis. A summary and discussion of future research directions conclude in Section 6.

2. BESAG, YORK, AND MOLLIE’S (1991) MODEL

Consider a spatial process $\{Y_i : i = 1, \dots, n\}$ on a lattice D . Without loss of generality, assume that there is no fixed effect in the sequel except an overall level μ . Fixed effects capturing the large scale variation can be added to the model by replacing μ with $X_i^\top \beta$ for covariate vector X_i and

coefficient vector β . A popular formulation given by Besag et al. (1991) is:

$$\begin{aligned} Y_i &= \mu + \phi_i + \varepsilon_i, \\ \phi_i | \phi_{j \neq i} &\sim N\left(\frac{\sum_{j \neq i} b_{ij} \phi_j}{\sum_{j \neq i} b_{ij}}, \frac{\sigma_\phi^2}{\sum_{j \neq i} b_{ij}}\right), \\ \varepsilon_i &\sim N(0, \sigma_\varepsilon^2), \end{aligned} \tag{1}$$

where ϕ_i is a random effect to capture the small-scale variation in region i attributable to regional clustering, ε_i represents unstructured noise, b_{ij} 's are known weights with $b_{ij} = b_{ji}$, and σ_ϕ^2 and σ_ε^2 are variance parameters. It is the conditionally specified $\{\phi_i\}$, which follows a Gaussian Markov random field, that make model (1) a spatial model. A common choice for weight b_{ij} is 1 if site i and site j are immediate neighbors and 0 otherwise. This choice will be used in the sequel. Note that the specification of $\{\phi_i\}$ leaves the overall mean level of the GMRF unspecified. It can be shown that only the pairwise differences between neighboring sites are specified (Besag and Kooperberg 1995). The GMRF in model (1) therefore corresponds to an improper prior joint distribution of $\{\phi_i\}$; see more discussion in Section 3. The heterogeneity effects $\{\varepsilon_i\}$ are independent and identically distributed normal variables. This model separates the variability of the observations into local clustering and global heterogeneity, which can both be of scientific interest. The two-component error decomposition can be used in more general settings, such as generalized linear models and survival models, finding applications in a variety of fields including economics, epidemiology, and social science, among others.

Successful as model (1) has been, it provides only limited accommodation to spatial heteroskedasticity, which may arise jointly with spatial autocorrelation in many spatial settings. Similar to a spatial process, spatial heteroskedasticity itself may exhibit local clustering. That is, sites with high volatility tend to have neighbors with high volatilities and those with low volatility tend to have neighbors with low volatilities. Heteroskedasticity clustering occurs naturally when there are measurement errors. If a variable is hard to measure precisely in a region, then it may be more likely to be hard to measure in the surrounding regions. For example, the difficulty level of measuring snow water equivalent (Cowles et al. 2002) may cluster according to the geographical contiguity. Model (1) implicitly specifies conditional heteroskedasticity through the spatial clustering effects $\{\phi_i\}$. It postulates that the conditional variance of ϕ_i given $\phi_{j \neq i}$ depends inversely on

the number of its neighbors. However, there is no spatially varying structure explicitly imposed on the marginal variance of $\{\phi_i\}$. The spatial heterogeneity effects $\{\varepsilon_i\}$ are assumed to have constant volatility (CV), which cannot reflect the local clustering effect of heteroskedasticity for many environmental datasets with clustered volatility of measurement errors.

Ideas of heteroskedasticity modeling in time series are to be adapted to the spatial analog. There are two classes of heteroskedasticity models in time series: autoregressive conditional heteroskedasticity (ARCH) and stochastic volatility (SV); see survey articles, for example, by Bollerslev, Engle, and Nelson (1994) for ARCH and Ghysels, Harvey, and Renault (1996) for SV. In an ARCH model, the conditional variance is modeled as a deterministic function of the available information; while in an SV model, the conditional variance is modeled as a latent stochastic process. These modeling strategies may be applied to $\{\varepsilon_i\}$ in model (1). In a spatial ARCH (SARCH) model, one would like to assume the conditional distribution of ε_i given $\varepsilon_{j \neq i}$ to be normal with mean zero and variance a function of $\varepsilon_{j \neq i}$. However, the existence of the joint distribution of $\{\varepsilon_i\}$ determined by these assumed full conditional distributions is unknown. On the other hand, in a spatial stochastic volatility model, one ends up with a hierarchical model which has been extensively studied in the Bayesian paradigm. Assume there is a latent volatility process $\{\sigma_i^2\}$ which has local clustering. Given this volatility process, the conditional distribution of ε_i is independent $N(0, \sigma_i^2)$, $i = 1, \dots, n$. When the SSV model for $\{\varepsilon_i\}$ is introduced into model (1), one more layer of the hierarchical structure is formed from the $\{\sigma_i^2\}$; see details of the revised model in Section 3.

3. SPATIAL STOCHASTIC VOLATILITY

We start by modifying the traditional model (1). Let $\mathbf{Y} = \{Y_i\}$ denote the vector of observed responses, $\boldsymbol{\phi} = \{\phi_i\}$ the vector of spatial clustering random effects, and $\boldsymbol{\varepsilon} = \{\varepsilon_i\}$ the vector of the unstructured heterogeneity (often measurement errors). Conditioning on $\boldsymbol{\phi}$, \mathbf{Y} in model (1) follows a multivariate normal distribution $N(\boldsymbol{\mu} + \boldsymbol{\phi}, \sigma_\varepsilon^2 I)$. To allow spatially varying heteroskedasticity, we introduce a latent variance process $\boldsymbol{\sigma}^2 = \{\sigma_i^2\}$, and use it to replace the constant conditional variance $\sigma_\varepsilon^2 I$ of \mathbf{Y} . That is,

$$\mathbf{Y} | \boldsymbol{\mu}, \boldsymbol{\phi}, \boldsymbol{\sigma}^2 \sim N(\boldsymbol{\mu} + \boldsymbol{\phi}, \text{diag } \boldsymbol{\sigma}^2).$$

To impose a spatially smooth structure on σ^2 , we let $\sigma_i^2 = \exp(\mu_h + h_i)$, where μ_h is an overall level of the log volatility, and $\mathbf{h} = \{h_i\}$ follows another CAR model in the same way as ϕ .

We now present the model with spatial stochastic volatility:

$$\begin{aligned}
Y_i &= \mu + \phi_i + \varepsilon_i, \\
\phi_i | \phi_{j \neq i} &\sim N\left(\frac{\sum_{j \neq i} b_{ij} \phi_j}{\sum_{j \neq i} b_{ij}}, \frac{\sigma_\phi^2}{\sum_{j \neq i} b_{ij}}\right), \\
\varepsilon_i &\sim N(0, \exp(\mu_h + h_i)) \\
h_i | h_{j \neq i} &\sim N\left(\frac{\sum_{j \neq i} c_{ij} h_j}{\sum_{j \neq i} c_{ij}}, \frac{\sigma_h^2}{\sum_{j \neq i} c_{ij}}\right),
\end{aligned} \tag{2}$$

where the c_{ij} 's are pre-specified weights playing the same role for \mathbf{h} as the b_{ij} 's do for ϕ in model (1), σ_h^2 is the variance parameter for \mathbf{h} , and all other parameters have the same meaning as in model (1). To make the parameters μ and μ_h identifiable, we add the constraints $\sum_{i=1}^n \phi_i = 0$ and $\sum_{i=1}^n h_i = 0$, respectively. The weights $\{c_{ij}\}$ can be different from $\{b_{ij}\}$ in general, but in practice, we can choose them to be the same. Model (2) distinguishes itself from model (1) by putting a smoothly-changing spatial structure on the variance of the ε_i 's. When σ_h^2 approaches zero, the variation in \mathbf{h} vanishes, and model (2) reduces to the constant volatility model (1). The process $\{\varepsilon_i\}$ in (2) can be used as an error process alone even without the presence of spatial clustering effects ϕ ; see Table 2 in Section 5. In a generalized linear models setup, we can replace the first equation in (2) with

$$\eta_i = \mu + \phi_i + \varepsilon_i, \tag{3}$$

where η_i is the linear predictor with $\eta_i = g\{E(Y_i)\}$ for link function g . However, we limit the scope in this paper to the case of normal responses.

The incorporation of a latent spatial process \mathbf{h} in the log volatility has the same flavor as the spatially varying coefficient processes in the literature. For point referenced data, Gelfand et al. (2003) viewed the regression coefficients as realizations from latent spatial processes, and applied their method to explain housing prices. Assunção (2003) dealt with the counterpart for areal data. These methods offer great flexibility in regression for spatial data, providing maps of spatially varying regression coefficients that researchers like to see. However, they do not address the clustering of volatilities, which can be important for many applications.

The SSV process has a variety of characteristics that make it attractive for spatial applications. In the case of volatility clustering, researchers may want to detect “hot spots” of volatilities, and monitor these spots more closely in the future. This can be done naturally by the SSV model. When prediction is of interest, the SSV model may be preferred to CV models by allowing the variance to vary spatially. SSV can also be an approximation of a more complex model and can pick up the effects of omitted variables. The existence of SSV would be interpreted as evidence of misspecification, either by omitted variables or through structural change.

4. BAYESIAN IMPLEMENTATION

Model (2) can be straightforwardly implemented in a Bayesian framework using MCMC methods. Let $\theta = (\mu, \sigma_\phi^2, \mu_h, \sigma_h^2)^\top$. Let $p(\cdot)$ in general represent the density of its arguments. The joint posterior distribution of interest is given by

$$p(\theta, \boldsymbol{\phi}, \mathbf{h} | \mathbf{y}) \propto L(\mu, \boldsymbol{\phi}, \mu_h, \mathbf{h}; \mathbf{y}) p(\boldsymbol{\phi} | \sigma_\phi^2) p(\mathbf{h} | \sigma_h^2) p(\mu) p(\sigma_\phi^2) p(\mu_h) p(\sigma_h^2), \quad (4)$$

where the first term L on the right-hand side is the likelihood, and the remaining terms are prior densities. Marginal posteriors of interest can be obtained from (4) by integrating out any unwanted parameters. For the normal case,

$$L(\mu, \boldsymbol{\phi}, \mu_h, \mathbf{h}; \mathbf{y}) \propto \sum_{i=1}^n \exp \left\{ -\frac{1}{2} \sum_{i=1}^n (\mu_h + h_i) \right\} \exp \left\{ -\frac{1}{2} \sum_{i=1}^n \frac{(y_i - \mu - \phi_i)^2}{\exp(\mu_h + h_i)} \right\}. \quad (5)$$

It can be shown (e.g. Hodges, Carlin, and Fan 2003) that the improper joint distributions $p(\boldsymbol{\phi} | \sigma_\phi^2)$ and $p(\mathbf{h} | \sigma_h^2)$ are

$$p(\boldsymbol{\phi} | \sigma_\phi^2) \propto \sigma_\phi^{-(n-1)} \exp \left\{ -\frac{\sum_{i \sim j} (\phi_i - \phi_j)^2}{2\sigma_\phi^2} \right\}, \quad (6)$$

and

$$p(\mathbf{h} | \sigma_h^2) \propto \sigma_h^{-(n-1)} \exp \left\{ -\frac{\sum_{i \sim j} (h_i - h_j)^2}{2\sigma_h^2} \right\}, \quad (7)$$

respectively, where $i \sim j$ if i and j are neighboring sites. Note that the exponents are $n - 1$ instead of n , as rigorously shown by Hodges et al. (2003). Prior specification of θ completes the Bayesian setup. For μ and μ_h , a flat (improper uniform) prior can be used. For variance parameters σ_ϕ^2 and σ_h^2 , a vague but proper prior is chosen. Let $IG(a, b)$ denote an inverse gamma distribution with

mean $b/(a-1)$. Choosing $IG(a_\phi, b_\phi)$ and $IG(a_h, b_h)$ as priors for σ_ϕ^2 and σ_h^2 has the advantage of being semi-conjugate in terms of their full conditionals.

MCMC algorithms for drawing samples from posterior distribution of the parameters in the CV model (1) have become standard and are available in the widely used software BUGS (Spiegelhalter et al. 1995). Therefore, here we only examine whether there is any extra difficulty brought by the SSV in model (2). Reparameterize μ_h and \mathbf{h} as $\boldsymbol{\lambda} = \{\mu_h + h_i\}$. Once we have $\boldsymbol{\lambda}$, μ_h and \mathbf{h} can be obtained by applying the constraint $\sum_{i=1}^n h_i = 0$. The full conditional of λ_i is

$$p(\lambda_i | \lambda_{-i}, \mu, \boldsymbol{\phi}, \sigma_h^2, \mathbf{y}) \propto p(\lambda_i | \lambda_{-i}, \sigma_h^2) p(y_i | \mu, \phi_i, \lambda_i), \quad (8)$$

where λ_{-i} denotes the vector $\boldsymbol{\lambda}$ with λ_i excluded. The first term $p(\lambda_i | \lambda_{-i}, \sigma_h^2)$ is simply the density of $N(\lambda_i^*, \nu_i^2)$, where $\lambda_i^* = \sum_{j \sim i} \lambda_j / m_i$ and $\nu_i^2 = \sigma_h^2 / m_i$ with m_i being the number of neighbors of site i . The second term can be bounded up to a scaling constant (Kim, Shephard, and Chib 1998, p.365),

$$\begin{aligned} p(y_i | \mu, \phi_i, \lambda_i) &\propto \exp \left[-\frac{1}{2} \lambda_i - \frac{(y_i - \mu - \phi_i)^2}{2} \exp(-\lambda_i) \right] \\ &\leq \exp \left[-\frac{1}{2} \lambda_i - \frac{(y_i - \mu - \phi_i)^2}{2} \{ \exp(-\lambda_i^*) (1 + \lambda_i^*) - \lambda_i \exp(\lambda_i^*) \} \right]. \end{aligned}$$

By completing the square of λ_i , the term on the right-hand side of (8) can then be shown to be bounded, up to a scale, by the density of

$$N \left(\lambda_i^* + \frac{\nu_i}{2} [(y_i - \mu - \phi_i)^2 \exp(-\lambda_i^*) - 1], \quad \nu_i^2 \right). \quad (9)$$

An accept-rejection algorithm can then be used to sample λ_i from its full conditional (e.g. Robert and Casella 1999, p.49). Furthermore, the density in (8) is a log-concave density, as

$$\frac{\partial^2}{\partial \lambda_i^2} \log p(\lambda_i | \lambda_{-i}, \mu, \boldsymbol{\phi}, \sigma_h^2, \mathbf{y}) = -\frac{1}{\nu_i^2} - \frac{(y_i - \mu - \phi_i)^2}{2} \exp(-\lambda_i) < 0. \quad (10)$$

Therefore, a more efficient adaptive rejection sampling (ARS) method can be used (Robert and Casella 1999, section 2.3.3). This is within the capability of BUGS, which means that the SSV model is easily accessible by scientists with little programming efforts.

Appropriate choice of prior distribution for the hyperparameters σ_ϕ^2 and σ_ε^2 in the CV model and σ_ϕ^2 and σ_h^2 in the SSV model. For the CV model, Best et al. (1999) find that $IG(0.5, 0.0005)$,

suggested by Kelsall and Wakefield (1999), may be a more reasonable prior for the variance parameters σ_ϕ^2 and σ_ε^2 . On the scale of standard deviation, which is the easiest to interpret, the resulted 0.01, 0.50, and 0.99 quantiles are 0.012, 0.047, and 2.52. To get an idea of how this prior compares to the priors originally used by Best et al. (1999), one can look at the 0.01 and 0.99 quantiles on the scale of the standard deviation. For $IG(0.1, 0.1)$, these quantiles are 0.251 and 4.057×10^9 . For $IG(0.001, 0.001)$, the 0.01 quantile is 6.240 while the 0.99 quantile is beyond the numerical precision of most softwares — for example, it is computed as infinity in R (R Development Core Team 2005). Therefore, the $IG(0.5, 0.0005)$ prior puts considerable greater prior mass near zero, which is reasonable for many applications. This prior is used in the analysis of the wheat yield data in Section 5, given that the data has sample standard deviation 0.371.

For the SSV model, the prior for σ_ϕ^2 can be specified similarly as for the CV model, but care is needed to in specifying the prior of σ_h^2 since $\{h_i\}$ is on the log scale of the variance $\{\sigma_i\}$. Because of amplification effect of log transformation around the neighborhood of zero, the log scale is very sensitive to numbers close to zero. Clearly, the overall level μ_h affects the magnitude of the variation level of $\{h_i\}$. In order to make a fair prior between σ_ϕ and $\{\sigma_i\}$, one million realizations of σ_ϕ were generated from the prior $IG(0.5, 0.0005)$, log transformed, and then centered on the log scale. The observed 0.01 and 0.99 quantiles of the resulted quantity is 0.028 and 7.536 in absolute value. A convenient choice of prior for σ_h^2 is $IG(0.5, 0.0005)$, which yields 0.01 and 0.99 quantiles on the scale of standard deviation 0.039 and 7.979, respectively.

It would be interesting to compare the fit of the constant volatility model (1) and the SSV model (2). A convenient model choice criterion is the Deviance Information Criterion (DIC) (Spiegelhalter et al. 2002) defined as

$$DIC = \bar{D} + p_D,$$

where \bar{D} is the posterior expectation of the deviance, and p_D is the effective number of parameters. The first term represents the fit of the model, and the second term measures the complexity of the model. The effective number of parameters p_D is obtained as the difference between the posterior expected deviance and the deviance evaluated at the posterior expectations. This number is typically smaller than the actual total number of parameters with random effects included, since the

random effects have dependence structure. The DIC is particularly useful in comparing models with a large number of random effects. Lower DIC implies preferred model. It has been applied in both spatial frailty models (Banerjee, Wall, and Carlin 2003) and stochastic volatility models in time series (Berg, Meyer, and Yu 2004), although the use of DIC is controversial for its lack of rigorous theoretical foundation in general settings; see the discussion of Spiegelhalter et al. (2002), for example, Smith (2002). We adopt the DIC in this article for what it is worth as an informal model comparison tool in the data analysis in Section 5. Further investigation with other model comparison methods would certainly be helpful.

5. WHEAT YIELD DATA REVISITED

Peculiar trends in this data have been documented in the literature (Cressie 1993, p.250), probably due to an earlier ridge and furrow pattern of plowing on the field. Neither a linear trend nor a periodic component was found flexible enough to capture the large-scale variation in the data. The median-polish algorithm is a robust way to decompose the trend into an overall effect, row effects, and column effects. It has attractive outlier-resistance properties and may yield less-biased residuals than the mean based method. Cressie (1993) remarked that the spatial trend, or large scale variation, must be taken into account before parameters of the small scale variation can be interpreted. To focus on the modeling of spatial heteroskedasticity, the data used in this paper is the median-polish residuals (Cressie 1993, Section 3.5) with spatial trend removed. The median-polish surface of the data can be found on page 252 of Cressie (1993). The residuals $\{y_i\}$ from median-polishing, shown in Figure 1, are the starting point of our analysis. They have median 0, mean -0.012 , and standard deviation 0.371.

[Figure 1 about here.]

5.1 Constant Volatility Model

The first model we fit is the widely used CV model (1) of Besag et al. (1991). The parameters of interest are $\{\mu, \phi, \sigma_\phi^2, \sigma_\varepsilon^2\}$. A non-informative flat prior is chosen for μ . The prior distribution of σ_ϕ^2 and σ_ε^2 are both chosen to be $IG(0.5, 0.0005)$, following the suggestion of Kelsall and Wakefield

(1999). Given the magnitude 0.371 of the standard deviation of $\{y_i\}$ this is a reasonable choice.

We ran 500,000 MCMC samples, discarded the first 100,000, and recorded every 100th of the remaining samples, resulting in 4,000 samples for summarizing. Table 1 gives the 2.5, 50, and 97.5 posterior percentiles for the parameters μ , σ_ϕ , and σ_ε . Since we started with the detrended residuals, our focus here is σ_ϕ , and σ_ε . Although these quantities give a rough idea about the decomposition of the variation in y_i 's, they are not directly comparable, since one is for the conditional specification and the other is for the usual marginal specification (Banerjee, Carlin, and Gelfand 2004, p.164).

[Table 1 about here.]

The surface of the posterior medians of $\{\phi_i\}$, $\{\varepsilon_i\}$, and $\{\varepsilon_i^2\}$, denoted by $\{\hat{\phi}_i\}$, $\{\hat{\varepsilon}_i\}$, and $\{\hat{\varepsilon}_i^2\}$, respectively, are plotted in Figure 1. The surface of $\{\hat{\phi}_i\}$ is smooth. Though it is hard to observe any pattern in the plot of $\{\hat{\varepsilon}_i\}$, the plot of $\{\hat{\varepsilon}_i^2\}$ does show some pattern of clustering. That is, high (low) values of $\hat{\varepsilon}_i^2$ tend to have high (low) values in the surrounding vicinity, potentially contradicting the constant volatility assumption on $\{\varepsilon_i\}$.

Based on 10,000 random permutations, the observed Moran and the Geary statistics have P-values 0.9983 and 0.9974 for the $\hat{\varepsilon}_i$ process, and 0.0142 and 0.0134 for the $\hat{\varepsilon}_i^2$ process, respectively. These tests suggest no evidence of clustering in $\hat{\varepsilon}_i$'s, but strong evidence of clustering in $\hat{\varepsilon}_i^2$'s. This motivates us to fit a stochastic volatility model (2).

5.2 Spatial Stochastic Volatility

Model (2) is now fitted, with the hope to better describe the data. Non-informative flat priors are chosen for μ and μ_h . The prior distributions of the variance parameters σ_ϕ^2 and σ_h^2 are $IG(0.5, 0.0005)$ and $IG(0.5, 0.005)$, respectively.

The same number of MCMC samples are collected for parameters in the SSV model. The 2.5, 50, and 97.5 posterior percentiles for the parameters are also presented in Table 1. The posterior median of spatial clustering effects $\{\hat{\phi}_i\}$ and the posterior median of latent standard deviation $\{\hat{\sigma}_i\}$ are plotted in Figure 2.

[Figure 2 about here.]

Compared to the spatial clustering effect in the CV model, the surface of $\{\phi_i\}$ in the SSV model is similar. except that it is maybe slightly smoother. This is consistent with the similar posterior median of σ_ϕ in both models, 0.394 in the CV model and 0.386 in the SSV model. A wide range of priors for σ_h^2 has been tried, such as $IG(0.5, 0.05)$, $IG(0.1, 0.1)$, and $IG(1, 1)$, and the results in Table 1 does not change much.

The essential utility of the SSV model is seen from the posterior median surface of $\{\sigma_i\}$, which closely resembles the surface shape of the $\{\varepsilon_i^2\}$ in Figure 1, picking up the spatial clustering in volatility. Clearly from Figure 2, the volatility in the north-east corner is much lower than that in the center and the north-west corner of the field. We can summarize the posterior median of the spatially varying standard deviation $\{\sigma_i\}$ from the MCMC samples. The median of these medians is 0.209 in the SSV model, compared to median 0.242 in the CV model. Their range is from 0.115 to 0.486, thus differing by a factor of 4.2.

It is also worth noting from Figure 1 and Figure 2 that the clustering patterns of the spatial effects $\{\phi_i\}$ and the stochastic volatility $\{\sigma_i^2\}$ are generally different. A high value in ϕ_i does not necessarily imply high volatility, since its neighbor may also have high values of ϕ_i . It is possible that high volatility appears at positions with ϕ_i 's low in magnitude. Actually, a closer investigation of the center of the field, where the volatility is the highest, reveals that wheat yield residuals in this area are not high in magnitude but are very different from one plot to another, implying high volatility.

To compare the error decomposition of the two models, Figure 3 plots the posterior median of $\{\mu + \phi_i\}$ from the CV model against that from SSV model, together with the 45 degree line. Most of the points lie about the 45 degree line except for some points at the tails. The range is $(-0.648, 0.914)$ from the CV model and $(-0.721, 0.729)$ from the SSV model, suggesting that the SSV model is pushing variations of a large residual from the spatial clustering effect $\{\phi_i\}$ to the spatial heteroskedastic effect $\{\varepsilon_i\}$. Figure 3 also plots the posterior median spatial heteroskedastic effect $\{\varepsilon_i\}$ and the posterior median latent variance $\{\sigma_i^2\}$ in the SSV model. It is clear that the SSV model provides larger variance for larger errors.

[Figure 3 about here.]

Table 2 compares the CV and SSV models in terms of the DIC and effective number of parameters p_D . It is worth noting that the deviances are computed only up to an additive constant. Therefore the DIC can be negative and can only be compared across models under the same distributional assumption. As seen in Figure 3 (b), most sites have relatively low volatilities except for those clustered “hot spots”. Therefore, putting a structure on the volatility via SSV improves the fit significantly. The CV model is soundly beaten by the SSV model in terms of their DIC values. Both models have effective number of parameters p_D much smaller than their nominal parameter counts, which would include 500 and 1000 random effects in the CV and SSV models, respectively. The more complicated SSV model has even smaller p_D than the simpler CV model. This is, however, not unusual; see for example the analysis of Scotland lip cancer data in Spiegelhalter et al. (2002).

[Table 2 about here.]

To illustrate that the spatial volatility process ε_i can be used alone in a model without the spatial clustering effects ϕ_i , we also fit a model of the form

$$Y_i = \mu + \varepsilon_i, \tag{11}$$

denoted as SSV0 in Table 2, where $\{\varepsilon_i\}$ is the same as those in the SSV model (2). This model ignores the spatial clustering in the mean, and does not fit the data as well as the CV model (1) from the DIC comparison.

Table 2 also presents DIC results for two other competitive models suggested by a referee. The CV-t model is the same as the CV model except that the conditional distribution of Y_i given ϕ_i is t with degrees of freedom df , scaled by σ_ε . The SV-iid model is the same as the SSV model except that the prior distribution of $\{h_i\}$ is independent $N(0, \sigma_h^2)$, without any spatial varying structure. These two models both allow large residuals. In terms of DIC, the CV-t model fits better than the CV model, and the SV-iid model fits better than the SSV model. During the experimental fitting process, the DIC of the SSV model and SV-iid model are not robust to prior changes, although the value of p_D are. Even though the SV-iid model always has smaller DIC than the SSV model under

all the experimented priors in this example, the SSV model may still be preferred, as illustrated below, in situations such as prediction for its spatially varying nature in the latent variance $\{\sigma_i^2\}$.

As a pilot study to compare the predicting capability of the CV model and the SSV model, Table 3 presents the leaving-one-out 95% prediction intervals of some selected points in the plot. Three sites in row 12 and three sites in column 18 are chosen, representing high volatility and low volatility regions, respectively. For each site, the observed value is replaced with missing value and then the two models are fitted. The reported prediction intervals are obtained from the 4,000 posterior samples. One observe that center of the prediction interval from the two models are very close. However, the standard deviation of the prediction at all sites are about the same from the CV model, but vary noticeably from the SSV model. Consequently, compared to the CV model, the prediction intervals from the SSV model is wider for larger residuals and narrower for smaller residuals. A full comparison of the predicting capability at all 500 sites with this leaving-one-out procedure is beyond the scope of this paper but is of great interest in future studies.

[Table 3 about here.]

6. DISCUSSION

In this paper we have introduced a SSV component into the widely used model of Besag et al. (1991). The SSV component allows the regional heterogeneity effects to have variance clustering spatially. The model is applied to the well-known wheat yield data and compared with the traditional constant volatility model. Volatility clustering is found to be present in the data. The SSV model is easy to implement and can be useful in a wide range of spatial applications, particularly when the volatility of measurement errors clusters.

As pointed by a referee, the identification issue of the SSV model needs to addressed. That is, given a large residual $e_i = \phi_i + \varepsilon_i$, how do we know if it is due to a large spatial clustering effect ϕ_i or spatial heteroskedastic effect ε_i ? Actually, this question applies not only to the SSV model but also to the CV model. The data Y_i cannot inform about ϕ_i or ε_i , but only about their sum e_i . Therefore, ϕ_i and ε_i are Bayesian unidentifiable. However, Bayesian learning, that is, prior to posterior movement, is still feasible (Banerjee et al. 2004, p.165). When the true model has a SSV

component, a large residual e_i is likely to be surrounded by residuals with large magnitude, but not necessarily of the same sign. If a large residual is surrounded by large residuals of the same sign, then it may be more likely to be picked up by the spatial clustering effect ϕ_i . It is the spatial structure imposed by the priors of $\{\phi_i\}$ and $\{\varepsilon_i\}$ that makes “borrowing strength from neighbors” possible. Vague, fair and proper priors for the variance parameters σ_ϕ^2 and σ_h^2 are important. Further investigation along this line may be worthwhile.

Several directions of extending the SSV model are immediate. We demonstrated the SSV model assuming that the large scale variation has been removed. With the overall level μ in model (2) replaced by a linear predictor $X_i^\top \beta$, the algorithm in Section 4 can be easily adapted to include the unknown parameter vector β in the constructed linear structure. Although we focused on a lattice setup, the essentials of the SSV can be carried over to irregularly spaced data, in which case, weight matrices must be appropriately chosen, similar to what Pettitt, Weir, and Hart (2002) did for a CAR model. The SSV can also be used as a random effect which replaces the commonly used homoskedastic term in a model for non-normal data, for example, count data for disease mapping as Haran, Hodges, and Carlin (2003) did in a generalized linear model setting.

The other class of models for volatility in time series, the ARCH model, is promising to provide an alternative way for volatility clustering. Bera and Simlai (2003) got the specification of a spatial ARCH model as a byproduct of an information matrix (IM) test. The error term is a replacement for the i.i.d. error in a Simultaneous Autoregressive (SAR) model, which does not have much overlap with the general hierarchical model (1).

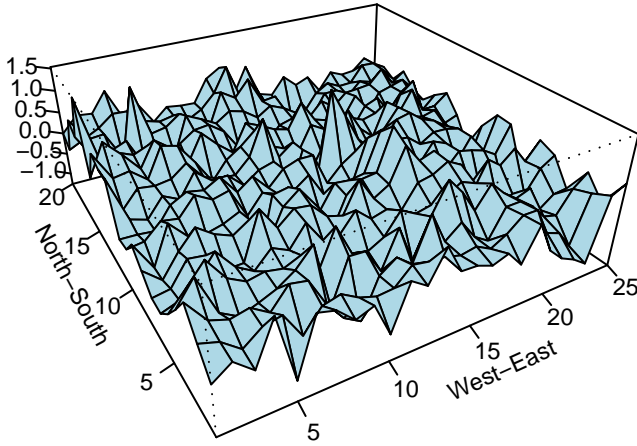
REFERENCES

- Assunção, R. (2003), “Space varying coefficient models for small area data,” *Environmetrics*, 14, 453–473.
- Banerjee, S., Carlin, B. P., and Gelfand, A. E. (2004), *Hierarchical modeling and analysis for spatial data*, Chapman & Hall / CRC.
- Banerjee, S., Wall, M. M., and Carlin, B. P. (2003), “Frailty Modeling for Spatially Correlated

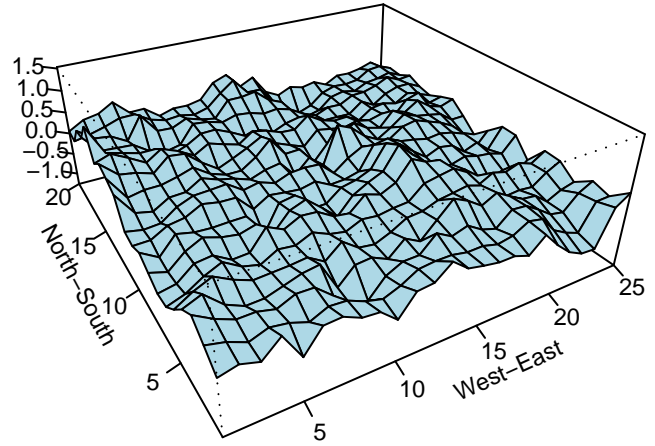
- Survival Data, with Application to Infant Mortality in Minnesota,” *Biostatistics (Oxford)*, 4, 123–142.
- Bera, A. K. and Simlai, P. (2003), “Testing for spatial dependence and a formulation of spatial ARCH (SARCH) model with applications,” *Thirteenth Annual Meeting of the Midwest Econometrics Group*.
- Berg, A., Meyer, R., and Yu, J. (2004), “Deviance Information Criterion for Comparing Stochastic Volatility Models,” *Journal of Business & Economic Statistics*, 22, 107–120.
- Besag, J. (1974), “Spatial interaction and the statistical analysis of lattice systems (with discussion),” *Journal of the Royal Statistical Society, Series B, Methodological*, 36, 192–236.
- (1977), “Errors-in-variables Estimation for Gaussian Lattice Schemes,” *Journal of the Royal Statistical Society, Series B: Methodological*, 39, 73–78.
- Besag, J. and Kooperberg, C. (1995), “On conditional and intrinsic autoregressions,” *Biometrika*, 82, 733–746.
- Besag, J., York, J., and Mollié, A. (1991), “Bayesian image restoration, with two applications in spatial statistics (Disc: p21-59),” *Annals of the Institute of Statistical Mathematics*, 43, 1–20.
- Best, N. G., Arnold, R. A., Thomas, A., Waller, L. A., and Conlon, E. M. (1999), “Bayesian Models for Spatially Correlated Disease and Exposure Data,” in *Bayesian Statistics 6 – Proceedings of the Sixth Valencia International Meeting*, eds. Bernardo, J. M., Berger, J. O., Dawid, A. P., and Smith, A., Clarendon Press [Oxford University Press], pp. 131–156.
- Bollerslev, T., Engle, R., and Nelson, D. (1994), “ARCH models,” in *The Handbook of Econometrics*, eds. Engle, R. and McFadden, D., Amsterdam: North-Holland, pp. 2959–3038.
- Cowles, M. K. and Zimmerman, D. L. (2003), “A Bayesian space-time analysis of acid deposition data combined from two monitoring networks,” *Journal of Geophysical Research*, 108, doi:10.1029/2003JD004001.

- Cowles, M. K., Zimmerman, D. L., Christ, A., and McGinnis, D. L. (2002), “Combining snow water equivalent data from multiple sources to estimate spatio-temporal trends and compare measurement systems,” *Journal of Agricultural, Biological, and Environmental Statistics*, 7, 536–557.
- Cressie, N. A. C. (1993), *Statistics for spatial data*, John Wiley & Sons.
- Dreesman, J. M. and Tutz, G. (2001), “Non-stationary Conditional Models for Spatial Data Based on Varying Coefficients,” *Journal of the Royal Statistical Society, Series D: The Statistician*, 50, 1–15.
- Gelfand, A. E., Kim, H.-J., Sirmans, C. F., and Banerjee, S. (2003), “Spatial Modeling with Spatially Varying Coefficient Processes,” *Journal of the American Statistical Association*, 98, 387–396.
- Ghysels, E., Harvey, A., and Renault, E. (1996), “Stochastic volatility,” in *Statistical Methods in Finance*, eds. C.R., R. and G.S., M., Amsterdam: North-Holland, pp. 119–191.
- Haran, M., Hodges, J. S., and Carlin, B. P. (2003), “Accelerating computation in Markov Random Fields Models for Spatial Data via Structured MCMC,” *Journal of Computational and Graphical Statistics*, 12, 249–264.
- Hodges, J. S., Carlin, B. P., and Fan, Q. (2003), “On the precision of the conditional autoregressive prior in spatial models,” *Biometrics*, 59, 317–322.
- Kelegian, H. H. and Robinson, D. P. (1998), “A suggested test for spatial autocorrelation and/or heteroskedasticity and corresponding Monte Carlo results,” *Regional Science and Urban Economics*, 28, 389–417.
- Kelsall, J. E. and Wakefield, J. C. (1999), “Comment on “Bayesian Models for Spatially Correlated Disease and Exposure Data”,,” in *Bayesian Statistics 6 – Proceedings of the Sixth Valencia International Meeting*, eds. Bernardo, J. M., Berger, J. O., Dawid, A. P., and Smith, A., Clarendon Press [Oxford University Press], p. 151.

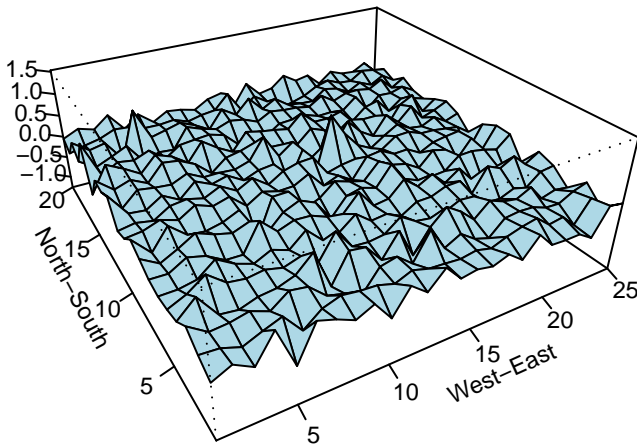
- Kim, S., Shephard, N., and Chib, S. (1998), “Stochastic Volatility: Likelihood Inference and Comparison with ARCH Models,” *Review of Economic Studies*, 65, 361–393.
- Mercer, W. B. and Hall, A. D. (1911), “The experimental error of field trials,” *Journal of Agricultural Science (Cambridge)*, 4, 107–132.
- Pettitt, A. N., Weir, I. S., and Hart, A. G. (2002), “A conditional autoregressive Gaussian process for irregularly spaced multivariate data with application to modelling large sets of binary data,” *Statistics and Computing*, 12, 353–367.
- R Development Core Team (2005), *R: A language and environment for statistical computing*, R Foundation for Statistical Computing, Vienna, Austria, ISBN 3-900051-07-0.
- Robert, C. P. and Casella, G. (1999), *Monte Carlo Statistical Methods*, Springer-Verlag Inc.
- Smith, J. (2002), “Comment on “Bayesian Measures of Model Complexity and Fit” (Pkg: P583-639),” *Journal of the Royal Statistical Society, Series B: Statistical Methodology*, 64, 619–620.
- Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and van der Linde, A. (2002), “Bayesian Measures of Model Complexity and Fit (Pkg: P583-639),” *Journal of the Royal Statistical Society, Series B: Statistical Methodology*, 64, 583–616.
- Spiegelhalter, D. J., Thomas, A., Best, N. G., and Gilks, W. R. (1995), “BUGS: Bayesian inference using Gibbs sampling, Version 0.50,” Technical report, Medical Research Council Biostatistics Unit, Institute of Public Health, Cambridge University.
- Whittle, P. (1954), “On stationary processes in the plane,” *Biometrika*, 41, 434–449.



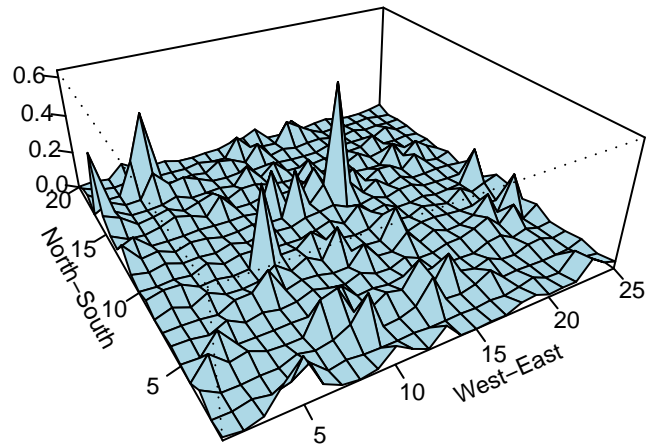
(a) wheat yield residual $\{y_i\}$ from median-polish



(b) posterior median of spatial effect $\{\phi_i\}$

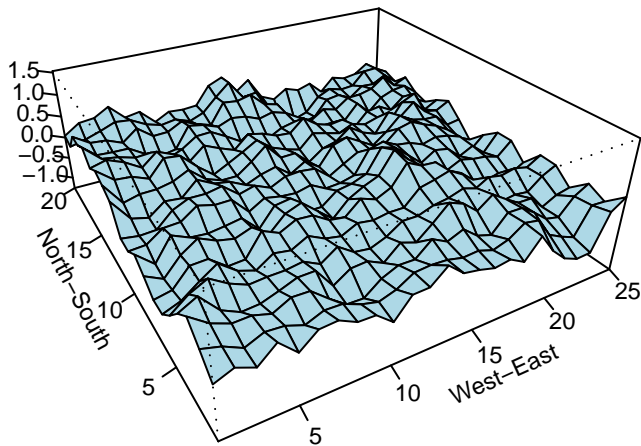


(c) posterior median of residual $\{\epsilon_i\}$

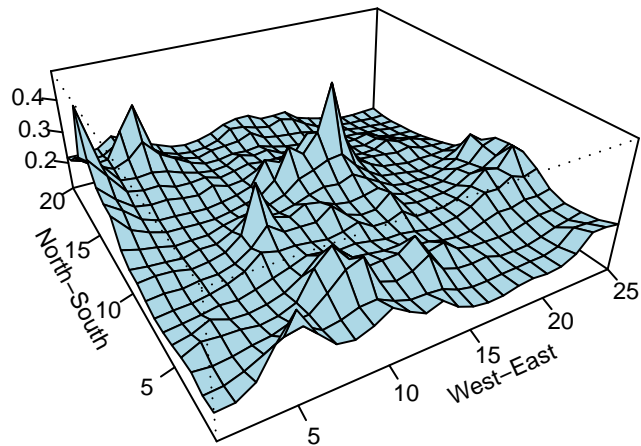


(d) squared residual $\{\epsilon_i^2\}$

Figure 1: Three-dimensional perspective of the wheat yield data and results from the constant volatility model of Besag, York, and Mollié (1991), $y_i = \mu + \phi_i + \epsilon_i$: (a) Wheat yield residual $\{y_i\}$ from Median-Polish; (b) Posterior median of the spatial clustering effect $\{\phi_i\}$; (c) Posterior median of the spatial heterogeneity effect $\{\epsilon_i\}$; (d) Squared posterior median of the spatial heterogeneity effect $\{\epsilon_i^2\}$.

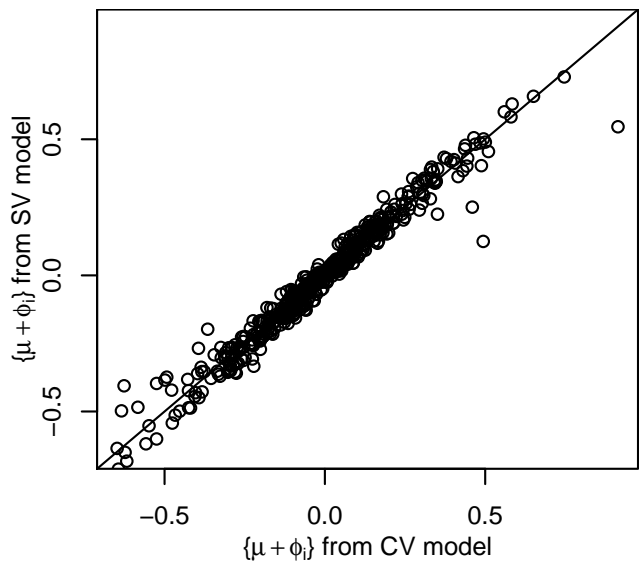


(a) posterior median of spatial effect $\{\phi_i\}$

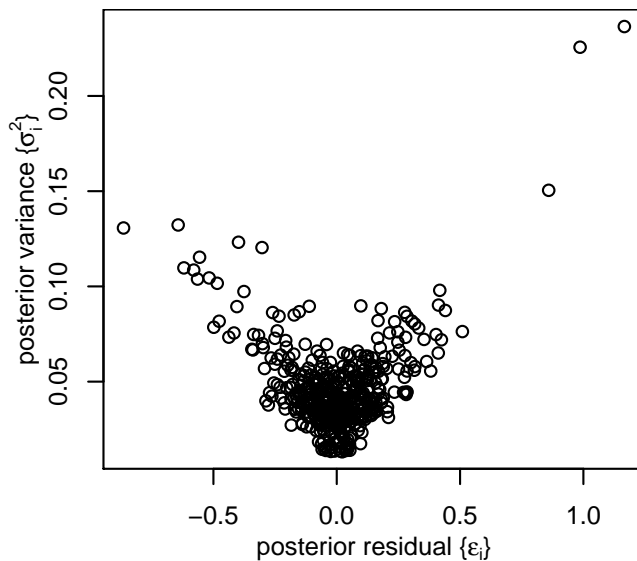


(b) posterior median of standard deviation $\{\sigma_i\}$

Figure 2: Three-dimensional perspective of results from the stochastic volatility model: (a) Posterior median of the spatial clustering effects $\{\phi_i\}$; (b) Posterior median of stochastic volatility $\{\sigma_i\}$.



(a) spatial clustering effect from two models



(b) spatially varying variance in error

Figure 3: Diagnostic plot of the SSV model: (a) Comparison of the posterior median of spatial clustering effects, $\{\mu + \phi_i\}$, from CV and SSV, the solid line being the 45 degree line; (b) Posterior median of spatial heterogeneity effect $\{\varepsilon_i\}$ and their variance $\{\sigma_i^2\}$ in the SSV model.

Table 1: Posterior percentiles for parameters in CV and SSV Models.

Parameter	Constant Volatility Model			Stochastic Volatility Model		
	2.5%	50%	97.5%	2.5%	50%	97.5%
μ	-0.034	-0.012	0.010	-0.033	-0.012	0.009
σ_ϕ	0.293	0.394	0.508	0.291	0.386	0.480
σ_ε	0.170	0.242	0.292			
μ_h				-4.398	-3.192	-2.630
σ_h				0.262	1.141	2.182

Table 2: Model comparison via DIC and effective number of parameters p_D .

	CV	CV-t	SSV0	SSV	SV-iid
p_D	237.9	292.5	54.4	227.8	231.9
DIC	216.9	197.2	383.7	7.7	-26.3

Table 3: Leaving-one-out prediction at selected points from CV and SSV models.

Observation			Constant Volatility Model				Stochastic Volatility Model			
col	row	value	sd	2.5%	50%	97.5%	sd	2.5%	50%	97.5%
14	12	-0.690	0.324	-0.695	-0.049	0.584	0.473	-1.028	-0.128	0.804
15	12	1.291	0.314	-0.784	-0.151	0.450	0.344	-0.836	-0.174	0.482
16	12	0.643	0.325	-0.547	0.111	0.733	0.419	-0.795	-0.031	0.760
18	14	0.055	0.325	-0.704	-0.089	0.546	0.289	-0.665	-0.099	0.482
18	15	-0.005	0.323	-0.835	-0.196	0.436	0.282	-0.791	-0.210	0.354
18	16	-0.383	0.326	-0.845	-0.224	0.416	0.272	-0.761	-0.222	0.295