Variance Modeling for Nonstationary Spatial Processes with Temporal Replications

Doris Damian Paul D. Sampson Peter Guttorp

NRCSE

Technical Report Series

NRCSE-TRS No. 075

January 28, 2003

Variance Modeling for Nonstationary Spatial Processes with Temporal Replications

Doris Damian, Paul D. Sampson and Peter Guttorp

University of Washington

January 28, 2003

Abstract

We have previously formulated a Bayesian approach to the Sampson and Guttorp model for the nonstationary correlation function $r(x, x')$ of a Gaussian spatial process [Damian et al., 2001]. This model assumes that the nonstationarity can be encoded through a bijective space deformation, f, that defines a new coordinate space in which the spatial correlation function can be considered isotropic, namely $r(x, x') = \rho(||f(x) - f(x')||)$, where ρ belongs to a known parametric family. We extend this model to incorporate spatial heterogeneity in site-specific temporal variances. In our Bayesian framework, the variances are considered (hidden) realizations of another spatial process, which we model as log-Gaussian, with correlation structure expressed in terms of the same spatial deformation function underlying that of the observed process. We demonstrate the method in simulations and in an application.

1 Introduction

Nonstationarity (or heterogeneity) of the covariance structure of space-time environmental processes is now widely recognized, although modelling tools remain in their infancy and are not widely available. Among the most recent work are the kernel and/or convolution methods of Higdon et al. [1999] and Fuentes [2002] along with our development of the spatial deformation model introduced by Sampson and Guttorp. Sampson [2002] and Sampson et al. [2001] provide recent reviews.

The fundamental emphasis of the spatial deformation methodology has been the concept of spatially varying, locally anisotropic spatial correlation structure. A second, related feature of the observed covariance structure of many space-time processes is that the variance of the temporal replications (temporal variance) is spatially heterogeneous. We note that this heterogeneity is observed even after response transformations (power or logarithmic) computed to symmetrize the response distribution, often in hope of stabilizing variances, are applied. Nonstationarity in variance appears to be an intrinsic characteristic of environmental processes, not simply related to questions of distributional form.

As an example, figure 1(a) shows a map of variances of log-transformed 10-day aggregate precipitation (mm) at 39 stations in the Languedoc-Roussillon region of France. The data were collected over the months of November and December from 1975 to 1992. The 10-day aggregates over these two month periods at the 39 stations resulted in 6 observations per year, 108 observations in all. (These data have been analyzed previously by Meiring et al. [1997], Iovleff and Perrin [1999] and Sampson et al. [2001], but focusing

only on spatial correlation, disregarding the issue of temporal variance). A log transformation (of the 10-day precipitation aggregates +1) was used to eliminate the mean-variance relationship. Despite this transformation, the map of sample variances shows more variation than would be expected were the true underlying variance process constant. The sample variances range between 2.5 and 6.4, with lower values in the southern part of the region and higher values in the central and northern parts.

When the variance can be assumed constant throughout the region of interest, the variance modeling is straightforward, involving only a single parameter. If the variance is thought to be varying over the region, however, a more sophisticated model is required, especially if one is interested in predictions at locations where observations are not available. We propose in section 3 a preliminary empirical assessment of the magnitude of the heterogeneity in sample temporal variances. A model for non-constant temporal variance is suggested in section 4. We assume the variances at observation locations to be hidden realizations of a nonstationary spatial process. The spatially varying anisotropy of this process is modeled utilizing the same Sampson-Guttorp spatial deformation representing the spatial correlation structure of the observed process, but with different correlation hyperparameters.

We note that another Bayesian setting for the Sampson-Guttorp spatial deformation model was proposed by Schmidt and O'Hagan [2000]. While recognizing the need for modeling the temporal variances and suggesting a log-normal approach, the latter paper does not propose a spatial model like ours. The temporal variances are assumed

exchangeable, with a noninformative prior distribution, and thus remain unstructured. The present paper introduces spatial structure for the temporal variances, taking a step further the correlation modeling of both Damian et al. [2001] and Schmidt and O'Hagan [2000].

2 The Observed Space-Time Process Model

We assume that temporally independent samples $Z_{it} = Z(x_i, t)$ are available at each of N geographic locations and at the same T points in time: $i = 1, 2, \ldots, N$ and $t = 1, 2, \ldots, T$. We consider the following model for the underlying process:

$$Z(x,t) = \mu(x,t) + \nu(x)^{1/2} H_t(x) + \varepsilon(x,t) \tag{1}$$

where x denotes location and t time.

$\mu(x,t)$ represents the spatio-temporal mean field. We focus here on modeling the variance-covariance structure and do not address inference about the spatio-temporal mean; we assume that the mean is constant in time (but not necessarily in space) and use a flat prior to remove it from the likelihood (eq. (5)). This assumption is appropriate for suitably short time periods, or when the temporal trends have been removed at each station and eq. (1) is used to model the residuals. Extensions to incorporate parameterized spatio-temporal trend components are under development.

$H_t(x)$ is a zero mean, variance one, Gaussian spatial process with a correlation function that depends smoothly on the geographic coordinates through the composition of an isotropic correlation function and a spatial deformation of the geographic coordinate

system:

$$\mathrm{Corr}\left(H_t\left(x\right), H_t\left(y\right)\right) = \rho_\theta\left(\left\|f\left(x\right) - f\left(y\right)\right\|\right). \tag{2}$$

This representation of the correlation function was introduced by Sampson and Guttorp [1992]. We assume that $\rho_\theta\left(\cdot\right)$ belongs to a known parametric family with unknown parameter(s) θ and that it is continuous at the origin: $\rho_\theta\left(d\right) \underset{d\to 0}{\longrightarrow} 1$. f represents the spatial deformation from the geographic coordinate system (G-plane) to an hypothesized plane (D-plane) in which the process Z is isotropic. The correlation function of $Z\left(x,t\right)$ does not necessarily converge to 1 at the origin due to the white noise process $\varepsilon\left(x,t\right)$, with variance σ_ε^2, representing small-scale spatial variability and measurement error.

$\nu\left(x\right)$, the variance of the process observed at location x, reflects the temporal fluctuations in the process $Z\left(x,t\right)$. Since we have temporal repetitions, at each geographic location we can compute the sample variance estimate, and thereupon inspect the behavior of these estimates in space. If the variance estimates do not vary too much throughout space, and if there is no apparent spatial pattern in their variability, we model the temporal variance as constant. Otherwise, we represent the variance as a spatial random process. This representation is the subject of our current paper.

Denoting by $\bar{Z}$ the vector of site means, $\bar{Z} = \frac{1}{T}\sum_{t=1}^{T} Z_t$, and by $\mathbf{S}$ the sample covariance matrix, $\mathbf{S} = \frac{1}{T-1}\sum_{t=1}^{T}\left(Z_t - \bar{Z}\right)\left(Z_t - \bar{Z}\right)'$ (which is non-singular if $T > N$), we

can write the likelihood derived from eq. (1) as

$$\begin{aligned}\mathcal{L}\left(\mu,\theta,f,\nu,\sigma_\varepsilon^2;Z_1,\ldots,Z_N\right) &= \left[\bar{Z},\mathbf{S}\mid\mu,\mathbf{\Sigma}\right] \\ &= \left|2\pi\mathbf{\Sigma}\right|^{-T/2}\exp\left\{-\frac{(T-1)}{2}\operatorname{tr}\mathbf{\Sigma}^{-1}\mathbf{S}-\frac{T}{2}\left(\bar{Z}-\mu\right)'\mathbf{\Sigma}^{-1}\left(\bar{Z}-\mu\right)\right\}.\end{aligned} \tag{3}$$

(We use square brackets to denote probability density functions.) Here, μ denotes the N-dimensional vector of means at the geographic locations and Σ represents the covariance matrix with elements

$$\sigma_{ij}=\begin{cases}\sqrt{\nu_i\nu_j}\rho_\theta\left(\left\|f\left(x_i\right)-\left(x_j\right)\right\|\right) & i\neq j\\ \nu_j+\sigma_\varepsilon^2 & i=j\end{cases},\qquad 1\leqslant i,j\leqslant N \tag{4}$$

where ν_i denotes the variance at location i.

The mean μ is viewed in this context as a nuisance parameter. We therefore integrate it out of eq. (3) under a flat (improper) prior distribution, $[\mu]\propto\mathbf{1}$:

$$\left[\mathbf{S}\mid\mathbf{\Sigma}\right]=\int\left[\bar{Z},\mathbf{S}\mid\mu,\mathbf{\Sigma}\right]\left[\mu\right]\ \mathrm{d}\,\mu\propto\left|\mathbf{\Sigma}\right|^{-(T-1)/2}\exp\left\{-\frac{(T-1)}{2}\operatorname{tr}\mathbf{\Sigma}^{-1}\mathbf{S}\right\}. \tag{5}$$

3 Is the Temporal Variance Constant in Space?

Inspection of the sample variances allows us to decide whether it is appropriate to model the temporal variance as constant or as varying in space. In order to explain this, we first review the theoretical properties of the sample variances (e.g., [Anderson, 1984]). Denoted by $S_{ii},\ 1\leq i\leq N$, these are the diagonal elements of the sample covariance matrix $\mathbf{S}$. The diagonal elements of the theoretical covariance matrix $\mathbf{\Sigma}$ are denoted by $\sigma_{ii},\ 1\leq i\leq N$.

Conditional on $\mathbf{\Sigma}$, under the usual normal assumption,

$$\frac{(T-1)\,S_{ii}}{\sigma_{ii}} \sim \chi^2_{(T-1)}. \tag{6}$$

The first two conditional moments of S_{ii} are given by:

$$\mathrm{E}\left(S_{ii} \mid \mathbf{\Sigma}\right) = \sigma_{ii} \tag{7}$$

and

$$\mathrm{Cov}\left(S_{ii}, S_{jj} \mid \mathbf{\Sigma}\right) = \frac{2\,T\sigma_{ij}^2}{(T-1)^2}. \tag{8}$$

In particular,

$$\mathrm{Var}\left(S_{ii} \mid \mathbf{\Sigma}\right) = \frac{2\,T\sigma_{ii}^2}{(T-1)^2}. \tag{9}$$

Therefore, the sample variances are unbiased (eq. (7)), consistent (eq. (9)), with correlations that are functions of the distances of the images of the observation sites in D-plane (eq. (8) and eq. (9)).

If the temporal variance is constant (ν), then all the S_{ii} have the same chi-squared distribution (conditional on ν and σ_ε^2) with expected value $\nu + \sigma_\varepsilon^2$ and variance $\left(2\,T/(T-1)^2\right)\left(\nu + \sigma_\varepsilon^2\right)^2$. Thus, if the spread of the S_{ii} is "reasonable" (based on the corresponding χ^2 distribution), we will model the variance as constant throughout space; otherwise we will prefer a model that accomodates spatial variability of the variance. By "reasonable" we mean that about $(1-\alpha)100\%$ of the sample variances divided by their average should fall within the interval

$$\left(\chi^2_{(T-1),\alpha/2}/(T-1)\,,\ \chi^2_{(T-1),(1-\alpha)/2}/(T-1)\right). \tag{10}$$

where $\chi^2_{df,p}$ is the p-quantile of the χ^2_{df} distribution.

Figure 2 presents an illustration: 36 sample variances were simulated twice at 36 sites on the unit square ($T = 200$). The first sample, mapped in panel (a), corresponds to a model with constant temporal variance, and the second, (b), to a model with spatially varying temporal variances. The difference between the two graphs is not easily characterized. But if we inspect the boxplots of the sample variances divided by their average, (c) and (d), it is clear that the range and the interquartile range corresponding to (a) are narrower than those corresponding to (b). Indeed, only 2 out of 36 sites (5.6%) in (a) fall outside the interval given by eq. (10) with $\alpha = 0.05\%$, $(0.81, 1.21)$, whereas in (b) 16 out of 36 sites (44.4%) do.

For the French precipitation data presented in the introduction, if the variance were constant in space, we would expect about 95% of the sample variances divided by their average to fall within the interval $(0.75, 1.29)$. However, 7 out of these 39 values (18%) fall outside the interval, thus a model with spatially varying temporal variance appears more appropriate for these data. Figure 1(b) presents the values of the sample variances divided by their average.

In the next section we describe how the variances are modeled when the constant temporal variance assumption does not seem to hold.

4 The Variances as a Random Field

4.1 Motivation

In the general case, we consider the variances at each of the observation sites as unknown parameters and model them as an unobserved vector of random variables, $\boldsymbol{\nu}$. In the same way that we model Z for the whole spatial domain and not only at the N arbitrary observation sites, we model the variance wherever Z is defined. This is necessary in order to predict the Z process (and its variance) at sites where no observations are collected. We model $\boldsymbol{\nu}$ as resulting from a random field, $\nu(x)$, that is defined and varies smoothly over the region of interest. It is reasonable to assume that this process will inherit the nonstationarity of the observed process $Z(x,t)$, in the sense that the spatial deformation f induces a new distance between geographic sites (the distance between their images in the D-plane), that determine the neighborhoods that influence most the process $Z(x,t)$ at any given point. These neighborhoods are in fact our way to express (and simplify) the complex effects of environmental processes and traits, often hidden or unmeasured, that impact $Z(\cdot)$. This temporal variability $\nu(\cdot)$ will be impacted by the same environmental processes and traits assumed to underly the nonstationary spatial correlation structure of $Z(\cdot)$, and therefore we model the nonstationarity in $\nu(\cdot)$ through the same spatial deformation f. Of course, this heuristic argument requires empirical validation.

Because of its interpretation as a variance process, $\nu(\cdot)$ has to be positive everywhere. We chose to use a log-normal model, the properties of which are described

below, through an auxiliary Gaussian spatial process, $\eta(\cdot)$.

4.2 A random field model

Let $\eta(x)$ be a real Gaussian process defined for $x \in \mathbb{R}^2$, with constant spatial mean μ, constant variance $\tilde{\sigma}^2$, and Sampson-Guttorp nonstationary correlation
$\text{Corr}(\eta(x), \eta(y)) = \rho_{\tilde{\theta}}(\|f(x) - f(y)\|)$, where $\rho(\cdot)$ belongs to the same parametric family as the correlation function of the process $Z(\cdot)$, but with different parameter(s) $\tilde{\theta}$. f is the same planar deformation as the one used in modeling the nonstationarity in $Z(\cdot)$. Then, at the observation sites, $\boldsymbol{\eta} = (\eta(x_1), \ldots, \eta(x_N))$ has density of the form

$$[\boldsymbol{\eta}] \propto |\tilde{\boldsymbol{\Sigma}}|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(\boldsymbol{\eta} - \mu \cdot \mathbf{1})' \tilde{\boldsymbol{\Sigma}}^{-1}(\boldsymbol{\eta} - \mu \cdot \mathbf{1})\right\} \tag{11}$$

for $\tilde{\boldsymbol{\Sigma}}$ an $N \times N$ matrix with elements $\tilde{\sigma}_{ij} = \tilde{\sigma}^2 \rho_{\tilde{\theta}}(\|f(x_i) - f(x_j)\|)$.

We define the **variance process**, $\nu(x)$, as $\nu(x) = \exp(\eta(x))$ and we readily obtain from eq. (11) the density of $\nu(\cdot)$ and its moments at the observation sites:

$$[\boldsymbol{\nu}] \propto \frac{1}{\prod_{i=1}^{N} \nu_i} |\tilde{\boldsymbol{\Sigma}}|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(\log(\boldsymbol{\nu}) - \mu \cdot \mathbf{1})' \tilde{\boldsymbol{\Sigma}}^{-1}(\log(\boldsymbol{\nu}) - \mu \cdot \mathbf{1})\right\} \tag{12}$$

where $\log(\boldsymbol{\nu}) = (\log(\nu_1), \ldots, \log(\nu_N))$. The process $\nu(\cdot)$ has constant spatial mean and variance of the form

$$\text{E}(\nu(x)) = \exp\left(\mu + \frac{1}{2}\tilde{\sigma}^2\right) \tag{13}$$

and

$$\text{Var}(\nu(x)) = \exp\left(2\mu + \tilde{\sigma}^2\right)\left(\exp\left(\tilde{\sigma}^2\right) - 1\right). \tag{14}$$

Its nonstationary covariance is given by

$$\mathrm{Cov}\left(\nu\left(x\right),\nu\left(y\right)\right)=\exp\left(2\mu+\tilde{\sigma}^{2}\right)\left\{\exp\left[\tilde{\sigma}^{2}\rho_{\tilde{\theta}}\left(\left\|f\left(x\right)-f\left(y\right)\right\|\right)\right]-1\right\}, \tag{15}$$

and its correlation is

$$\mathrm{Corr}\left(\nu\left(x\right),\nu\left(y\right)\right)=\frac{\exp\left[\tilde{\sigma}^{2}\rho_{\tilde{\theta}}\left(\left\|f\left(x\right)-f\left(y\right)\right\|\right)\right]-1}{\exp\left(\tilde{\sigma}^{2}\right)-1}. \tag{16}$$

The upper row in figure 3 presents three (discretized) realizations of a log-normal nonstationary variance process on the unit square of the type described above. The underlying spatial deformation f is shown in figure 4. Each of the three realizations were sampled and are plotted on a grid of basic unit 0.1×0.1 (so that they are each based on 121 sampled values). The mean of the process is 3, and its variance is 1. The correlation function $\rho_{\tilde{\theta}}\left(d\right)$ is the exponential $\rho_{\tilde{\theta}}\left(d\right)=\exp\left(-\tilde{\theta}d\right)$. The correlation of $\nu\left(x\right)$ at each distance in D-plane is highest for $\tilde{\theta}=0.1$ and lowest for $\tilde{\theta}=10$. Because of this, the realization of the variance process appears smoothest for $\tilde{\theta}=0.1$ and roughest for $\tilde{\theta}=10$. It is also interesting to note that in the middle upper panel (that corresponds to $\tilde{\theta}=1$) it is visible that the spatial correlation is highest in the "northeastern" corner of the square, where the plane is most "shrunk" by the deformation.

The middle row in figure 3 shows histograms of the 121 sampled ν values corresponding to the three values of $\tilde{\theta}$. As expected, the higher the value of $\tilde{\theta}$, the larger the range of the sampled ν values.

We would expect the variability in the values of ν to be reflected in the range of sample variances of realizations of the Z process (since $\nu\left(x\right)$ represents the variance of

$Z(x)$). The panels in the lowest row of figure 3 demonstrate that this indeed is the case: conditional on the respective ν, 36 sample variances were simulated (for time series of length $T = 200$) at the same sites as those shown in figure 2. Boxplots of these sample variances divided by their averages are shown. It is evident that the higher the variability in ν, the higher is the variability in the sample variances. Based on the appropriate confidence interval (eq. (10)) none of the three samples would have been judged a priori to correspond to a model with constant temporal variance: 11.1% (4/36) of the $\tilde{\theta} = 0.1$ sample, 33.3% (12/36) of the $\tilde{\theta} = 1$ sample and 47.2% (17/36) of the $\tilde{\theta} = 10$ sample fall outside the confidence interval.

This illustration emphasizes that besides bringing the model of the observed process Z and the variance model under the same conceptual framework, our approach to modeling the variance is flexible enough to cover a spectrum of temporal variance—from almost uniform to highly irregular.

4.3 Choosing the Hyper-Prior Distributions

The variance model (eq. (12)) defines the distribution of $\boldsymbol{\nu}$ conditional on the hyperparameters μ, $\tilde{\sigma}^2$ and $\tilde{\theta}$. In order to render this model complete, we need to specify the prior distributions of the hyperparameters (the hyper-prior distributions). As for the other parameters of our model, we chose diffuse but proper hyper-distributions. In particular, the priors for μ and $\tilde{\sigma}^2$ were chosen as normal and inverse gamma, respectively. This strategy simplifies the corresponding steps in the MCMC algorithm used for obtaining

samples from the posterior distributions since it allows for direct sampling from the full conditionals.

5 Predictive Distributions for the Variance Process

In this section we develop the predictive distribution of the variance field at ungauged locations. We follow the notation of the previous section in which $\log(\nu(x))$ was defined as $\eta(x)$ (the auxiliary Gaussian random field). In addition, the subscripts OL and UL denote observation and ungauged locations respectively. Using this notation, the joint distribution (conditional on the hyperparameters $\tilde{\mu}, \tilde{\theta}, \tilde{\sigma}^2$ and on all the other parameters which we denote by Ω) of $\boldsymbol{\eta}_{OL}$ and $\boldsymbol{\eta}_{UL}$ is multivariate normal of the form

$$\left.\begin{matrix}\boldsymbol{\eta}_{OL} \\ \boldsymbol{\eta}_{UL}\end{matrix}\right| \left(\tilde{\mu}, \tilde{\theta}, \tilde{\sigma}^2, \Omega\right) \sim \text{MVN}\left(\tilde{\mu} \cdot \mathbf{1}, \widetilde{\widetilde{\boldsymbol{\Sigma}}}\right). \tag{17}$$

Similar to the matrix $\widetilde{\boldsymbol{\Sigma}}$ in eq. (11), $\widetilde{\widetilde{\boldsymbol{\Sigma}}}$ has elements $\tilde{\tilde{\sigma}}_{ij} = \tilde{\sigma}^2 \rho_{\tilde{\theta}}(||f(x_i^*) - f(x_j^*)||)$, where x^* denotes now either an observation site or an ungauged site. $\widetilde{\widetilde{\boldsymbol{\Sigma}}}$ can be partitioned in four blocks that represent the covariances between pairs of sites: observation–observation, observation–ungauged, ungauged–observation and ungauged–ungauged. We denote these by $\widetilde{\widetilde{\boldsymbol{\Sigma}}}_{OO}$, $\widetilde{\widetilde{\boldsymbol{\Sigma}}}_{OU}$, $\widetilde{\widetilde{\boldsymbol{\Sigma}}}_{UO}$ and $\widetilde{\widetilde{\boldsymbol{\Sigma}}}_{UU}$, respectively. From eq. (17) we obtain the distribution of $\boldsymbol{\eta}_{UL}$ conditional on $\boldsymbol{\eta}_{OL}$, on the hyperparameters and

on the parameters as

$$\boldsymbol{\eta}_{UL} \,\Big|\, \left(\boldsymbol{\eta}_{OL}, \tilde{\mu}, \tilde{\theta}, \tilde{\sigma}^2, \Omega, \mathbf{Z}_{OL}\right) \sim \boldsymbol{\eta}_{UL} \,\Big|\, \left(\boldsymbol{\eta}_{OL}, \tilde{\mu}, \tilde{\theta}, \tilde{\sigma}^2, \Omega\right) \sim \text{MVN}\left(\boldsymbol{\mu}_{U|O}, \widetilde{\widetilde{\boldsymbol{\Sigma}}}_{U|O}\right), \tag{18}$$

where

$$\boldsymbol{\mu}_{U|O} = \tilde{\mu} \cdot \mathbf{1} + \widetilde{\widetilde{\boldsymbol{\Sigma}}}_{UO} \left(\widetilde{\widetilde{\boldsymbol{\Sigma}}}_{OO}\right)^{-1} \left(\boldsymbol{\eta}_{OL} - \tilde{\mu} \cdot \mathbf{1}\right), \widetilde{\widetilde{\boldsymbol{\Sigma}}}_{U|O} = \widetilde{\widetilde{\boldsymbol{\Sigma}}}_{UU} - \widetilde{\widetilde{\boldsymbol{\Sigma}}}_{UO} \left(\widetilde{\widetilde{\boldsymbol{\Sigma}}}_{OO}\right)^{-1} \widetilde{\widetilde{\boldsymbol{\Sigma}}}_{OU}. \tag{19}$$

The two conditional distributions in eq. (18) are the same because, conditional on $\boldsymbol{\eta}_{OL}$ and on Ω, the distribution of Z_{OL} (the data) is not a function of $\boldsymbol{\eta}_{UL}$.

We may generate K samples from the predictive distribution $\boldsymbol{\nu}_{UL} \,|\, Z_{OL}$ by sampling from the posterior distribution of $\left(\boldsymbol{\eta}_{OL}, \tilde{\mu}, \tilde{\theta}, \tilde{\sigma}^2, \Omega\right) \,|\, Z_{OL}$. For the j^{th} sample, we then draw $\boldsymbol{\eta}_{UL}^{(j)}$ from a multivariate normal distribution with parameters $\mu_{U|O}^{(j)}$ and $\widetilde{\widetilde{\boldsymbol{\Sigma}}}_{U|O}^{(j)}$ and define $\boldsymbol{\nu}_{UL}^{(j)} = \exp(\boldsymbol{\eta}_{UL}^{(j)})$ $(j = 1, \ldots, K)$. Point predictions of $\boldsymbol{\nu}_{UL}$ can then be obtained as the mean (or median) of this sample, and the predictive variance can be estimated as the sample variance.

Alternatively, we can compute point predictions directly from the K samples of $\left(\boldsymbol{\eta}_{OL}, \tilde{\mu}, \tilde{\theta}, \tilde{\sigma}^2, \Omega\right) \,|\, \boldsymbol{Z}_{OL}$ by noting that

$$\text{E}\left[\boldsymbol{\nu}_{UL,i} \,\Big|\, \left(\boldsymbol{\eta}_{OL}, \tilde{\mu}, \tilde{\theta}, \tilde{\sigma}^2, \Omega, \mathbf{Z}_{OL}\right)\right] = \exp\left(\boldsymbol{\mu}_{U|O,i} + \frac{1}{2}\widetilde{\widetilde{\boldsymbol{\Sigma}}}_{U|O,ii}\right), \tag{20}$$

hence,

$$\begin{aligned} \widehat{\boldsymbol{\nu}}_{UL,i} &= \text{E}\left[\boldsymbol{\nu}_{UL,i} \,|\, \boldsymbol{Z}_{OL}\right] \\ &= \widehat{\text{E}}\,\text{E}\left[\boldsymbol{\nu}_{UL,i} \,\Big|\, \left(\boldsymbol{\eta}_{OL}, \tilde{\mu}, \tilde{\theta}, \tilde{\sigma}^2, \Omega, \boldsymbol{Z}_{OL}\right)\right] = \frac{1}{K}\sum_{j=1}^{K} \exp\left(\boldsymbol{\mu}_{U|O,i}^{(j)} + \frac{1}{2}\widetilde{\widetilde{\boldsymbol{\Sigma}}}_{U|O,ii}^{(j)}\right). \end{aligned} \tag{21}$$

Estimated predictive variances can also be obtained directly if we first note that

$$\mathrm{Var}\left[\boldsymbol{\nu}_{UL,i}\,\middle|\left(\boldsymbol{\eta}_{OL},\tilde{\mu},\tilde{\theta},\tilde{\sigma}^2,\Omega,\boldsymbol{Z}_{OL}\right)\right]=\exp\left(2\boldsymbol{\mu}_{U|O,i}+\tilde{\tilde{\boldsymbol{\Sigma}}}_{U|O,ii}\right)\left(\exp\left(\tilde{\tilde{\boldsymbol{\Sigma}}}_{U|O,ii}\right)-1\right), \tag{22}$$

and then make use of the formula

$$\begin{aligned}\widehat{\mathrm{Var}}\left[\boldsymbol{\nu}_{UL,i}|\boldsymbol{Z}_{OL}\right]&=\widehat{\mathrm{E}}\,\mathrm{Var}\left[\boldsymbol{\nu}_{UL,i}\,\middle|\left(\boldsymbol{\eta}_{OL},\tilde{\mu},\tilde{\theta},\tilde{\sigma}^2,\Omega,\boldsymbol{Z}_{OL}\right)\right]\\&\quad+\widehat{\mathrm{Var}}\,\mathrm{E}\left[\boldsymbol{\nu}_{UL,i}\,\middle|\left(\boldsymbol{\eta}_{OL},\tilde{\mu},\tilde{\theta},\tilde{\sigma}^2,\Omega,\boldsymbol{Z}_{OL}\right)\right]\\&=\frac{1}{K}\sum_{j=1}^{K}\exp\left(2\boldsymbol{\mu}_{U|O,i}^{(j)}+\tilde{\tilde{\boldsymbol{\Sigma}}}_{U|O,ii}^{(j)}\right)\left(\exp\left(\tilde{\tilde{\boldsymbol{\Sigma}}}_{U|O,ii}^{(j)}\right)-1\right)\\&\quad+\frac{1}{K}\sum_{j=1}^{K}\left(\exp\left(\boldsymbol{\mu}_{U|O,i}^{(j)}+\frac{1}{2}\tilde{\tilde{\boldsymbol{\Sigma}}}_{U|O,ii}^{(j)}\right)-\widehat{\boldsymbol{\nu}}_{UL,i}\right)^2.\end{aligned} \tag{23}$$

6 A Simulation Study

In this section we assess how well the hyperparameters and the variances can be estimated and use the formulas developed in the previous section to predict variances at ungauged locations. For these purposes, samples from three realizations of the variance process are studied, with exponential correlation functions having parameter $\tilde{\theta}$ set to 0.1, 1 and 10, respectively. The spatial deformation is that of the unit square (figure 4). Models were fitted from data generated at the 16 observation locations ("gauged" sites) marked by diamonds. Twenty locations marked by open circles were considered as "ungauged" sites to assess predictions. The corresponding correlation functions (as functions of distances in the D-plane) are presented in figure 5 along with the correlation function of the Z process.

When $\tilde{\theta}$ is the smallest, the correlation in the variance process is high even at large distances. When $\tilde{\theta}$ is equal to 10, the correlation decays rapidly. The intermediate value of 1 approximately corresponds to the correlation in the Z process. The correlation parameter $\tilde{\theta}$ affects the smoothness of the realizations of the variance surface: these are smoothest for small values of $\tilde{\theta}$ and roughest when $\tilde{\theta}$ has a large value. For $\tilde{\theta}$ we have used a diffuse prior (uniform on the interval $(0, 50)$) that is flat at all three values of $\tilde{\theta}$.

We note, that even though the emphasis of the present paper is variance estimation, all the many parameters of the model, including the site variance, are estimated jointly. The estimates are obtained through a Metropolis-Hastings algorithm, the basic steps of which are described by Damian et al. [2001]. The extension of this algorithm to include the additional variance parameters is straightforward, so we do not discuss details here here. As explained in Damian et al. [2001], in the early stages of the development of the algorithm, we monitor its convergence both by studying the behavior (until stabilization) of the log-posterior distribution and by means of the Gelman-Rubin $\widehat{R}$ statistic that compares between- to within-chain variability for sufficiently distant initial configurations [Gelman and Rubin, 1992].

Smoothed histograms of the posterior distributions of the hyperparameters are shown in figure 6. Maximum aposteriori (MAP) estimates and 95% credible intervals are marked on the histograms. μ and $\tilde{\sigma}^2$ are well estimated in all three cases. When $\tilde{\theta}$ is either 0.1 or 1, the MAP estimate is close to the true value. When $\tilde{\theta}$ is 10 the posterior is almost flat over the interval (10,50) resulting in a poor estimate. The reason for this behavior is

that for large values of $\tilde{\theta}$ the sample variances are practically uncorrelated (figure 5) and thus could have resulted from a range of values of $\tilde{\theta}$.

The variances at the observation locations are estimated quite well for all three values of $\tilde{\theta}$, as can be seen from figure 7 in which MAP estimates of the temporal variances are plotted against the true sampled values – in all cases the true values fall within the 95% C.I.s.

The plots in figure 8 show the simulated variance values at 20 "ungauged" sites vs. the predicted values of these variances. The predicted values obtained by sampling from the predictive distribution are quite similar to the ones obtained directly through eq. (21), and all the predicted variances are within one estimated standard deviation of the simulated values.

In figure 9 we present maps of variance contours representing the different levels of variability in the spatial process Z. The left panel of the figure is based on the simulated variance values and the right panel is based on the 16 estimated and 20 predicted values. Both panels exhibit the same behavior: in an east—west direction the variance levels first decrease then increase again (by the same order of magnitude).

We conclude that our method performs satisfactorily in estimating the variance field at observation locations and in predicting it at ungauged locations.

7 The French Precipitation Data

We have divided the 39 observation locations of the Languedoc-Roussillon database into two groups: the first group, consisting of 31 locations, was used for model fitting and the the remaining 8 locations were set aside for the evaluation of spatial predictions.

We fitted two models to the data from the 31 locations, one assuming a nonstationary variance process and one assuming constant temporal variance. Samples of 2,000 observations from the posterior distributions of the various parameters of each of these models were used to estimate the Bayes factor comparing them [Kass and Raftery, 1995]. Its estimated value is greater than 150, the threshold value for decisive evidence in favor of the model with nonstationary variance, so we conclude that the model with constant temporal variance is not suitable for the precipitation process. A comparison of the plots of estimated covariances versus the sample ones for these two models (figure 10) supports the choice of a nonstationary temporal variance in this application, since the variability of the covariance cloud is notably lower. Our results are thus based on a model similar to the one presented in section 4. Figure 11 depicts the estimated spatial deformation as given by the mean of the posterior distribution. Dots mark the 31 locations used for model fitting while numbered open circles identify the sites set aside for the evaluation of predictions.

The estimated variances at the 31 locations are shown in figure 12. The pattern in their spatial distribution is similar to the one observed in the sample variances (figure 1): the lowest values are in the southwest, and the highest in the central and northern regions.

The estimated mean of the variance field is 4.31 with interquartile range $(3.94, 4.66)$. The estimated correlation in the variance field decreases from 0.85 between the two closest locations in D-plane to 0.0002 between the most distant pair (figure 12).

Figure 13 is a plot of the observed sample variances (including nugget effects) vs. the corresponding estimated (or fitted) variances for the 31 sites used in model fitting and the variances predicted for the remaining 8 sites. Most of the predictions fall as close to the sample values as do the estimated variances, but there are two predicted variances, at sites 19 and 41, that are clearly further away from the reference line. The sample variance at site 19 is surprising large compared that of a nearby observation site while the sample variance at site 41 is over-estimated. A plot of variance contours based on the estimated and the predicted values is shown in figure 14.

8 Discussion

In this paper we have proposed a model for spatially heterogeneous temporal variance, integrated within the conceptual framework of the spatial deformation approach of Sampson and Guttorp for structuring spatial correlations. Software for the MCMC estimation of this model is available from the authors. These variances, combined with the spatial correlation structure to define spatial covariance, are fundamental in the prediction of the observable process, $Z(x,t)$, at ungauged sites. To fully enable this spatial prediction we have under development the incorporation of an additional (multivariate) random field to represent $\mu(x,t)$ through temporal trend models at the monitoring sites with spatially

varying parameters. This random field model may incorporate physical attributes of the domain, such as slope and elevation, for example, as spatially varying covariates. The accommodation of simple temporal autocorrelation structure will follow. The model is being applied now to daily observations of tropospheric ozone across the U.S.

Acknowledgement

The research described in this article has been funded in part by the United States Environmental Protection Agency through agreement CR825173-01-0 to the University of Washington. It has not been subjected to the Agency's peer and policy review and therefore does not necessarily reflect the views of the Agency and no official endorsement should be inferred.

References

Anderson, T. (1984). *An Introduction to Multivariate Statistical Analysis*. Wiley, New York, second edition.

Damian, D., Sampson, P., and Guttorp, P. (2001). Bayesian estimation of semi-parametric non-stationary spatial covariance structures. *Environmetrics*, 12:161–178.

Fuentes, M. (2002). Spectral methods for nonstationary spatial processes. *Biometrika*, 89(1):197–210.

Gelman, A. and Rubin, D. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*, 7(4):457–473.

Higdon, D., Swall, J., and Kern, J. (1999). Non-stationary spatial modeling. In Bernardo, J., JO., B., David, A., and Smith, A., editors, *Bayesian Statistics 6*, pages 761–768.

Iovleff, S. and Perrin, O. (1999). Estimating a non-stationary spatial structure using simulated annealing. http://www.math.chalmers.se/ olivier/paper/recuit.ps.

Kass, R. and Raftery, A. (1995). Bayes factors. *Journal of the American Statistical Association*, (90):773–795.

Meiring, W., Monestiez, P., Sampson, P., and Guttorp, P. (1997). Developments in the modelling of nonstationary spatial covariance structure from space-time monitoring data. In Baafi, E. and Schofield, N., editors, *Geostatistics Wallongong '96*, volume 1, pages 162–173. Kluwer Dordrecht.

Sampson, P. (2002). *Encyclopedia of Environmetrics*, chapter Spatial Covariance. John Wiley & Sons.

Sampson, P., Damian, D., and Guttorp, P. (2001). Advances in modelling and inference for environmental processes with nonstationary spatial covariance. In Allard, D., Monestiez, P., and Froidevaux, R., editors, *GeoENV2000: Third European Conference on Geostatistics for Environmental Applications*.

Sampson, P. and Guttorp, P. (1992). Nonparametric estimation of nonstationary spatial covariance structure. *Journal of the American Statistical Association*, 87(417):108–119.

Schmidt, A. and O'Hagan, A. (2000). Bayesian inference for nonstationary spatial covariance structures via spatial deformations. Technical report, University of Sheffield.

Figure Captions

Figure 1 Sample variances of log-transformed French precipitation data: (a) original values, (b) values divided by their average (all values multiplied by 10 to facilitate visualization). The axes define a rectangular coordinate system measured in 100's of meters.

Figure 2 Maps of simulated sample variances, (a) and (b), and corresponding boxplots of sample variances divided by their average, (c) and (d). (Values in the upper panels are multiplied by 10 to facilitate visualization.) The contour lines in the upper panels correspond to values from 20 to 44 at intervals of 2. The data in (a) and (c) were generated from a constant temporal variance model (equal to 3) and the data in (b) and (d) were generated from a log-normal variance model with mean equal to 3 and exponential hyper-parameter $\tilde{\theta} = 2.5$.

Figure 3 Variance processes for different values of the correlation parameter $\tilde{\theta}$ (exponential correlation). The first row illustrates three realizations of the process. The second row shows histograms of the resulting 121 variances, and the third row presents boxplots of sample variances divided by their average.

Figure 4 A non-linear deformation of the unit square: (a) G-plane, (b) D-plane. Solid diamonds mark the 16 observation locations used in model fitting. Open circles mark 20 points used to assess predictions.

Figure 5 Correlation functions of the variance process for $\tilde{\theta} = 0.1, 1, 10$. The solid line is

the correlation function of the Z process.

Figure 6 Posterior distributions of the hyperparameters (the parameters of the variance process). The first column corresponds to the model with $\tilde{\theta} = 0.1$, the second to $\tilde{\theta} = 1$ and the third to $\tilde{\theta} = 10$. The true values are marked by solid lines. MAP estimates are marked by dotted lines. The dashed lines depict 95% credible intervals.

Figure 7 Estimated variances at 16 sites used for model fitting vs. corresponding true variances underlying the simulation. The vertical bars represent 95% credible intervals. The $y = x$ line is superimposed.

Figure 8 Predicted variances at 20 ungauged sites vs. corresponding true (simulated) variances ($\tilde{\theta} = 1$). (a) The predictions are based on means of a sample of 2,000 values from the predictive distribution. (b) The predictions are calculated using eq. (21). The bars represent $\pm$ one standard deviation.

Figure 9 Variance contour maps: (a) based on true (simulated) values at 36 sites and (b) based on estimated values at 16 gauged sites and predicted values at 20 ungauged sites. The contour lines correspond to values from 2.5 to 4.0 at intervals of 0.3.

Figure 10 Estimated vs. sample covariances: (a) in a constant variance model and (b) in a nonstationary variance model. The solid line is the graph of $y = x$.

Figure 11 (a) G-plane representation of 31 Languedoc-Roussillon precipitation monitoring sites. (b) D-plane representation according to the mean of the posterior distribution.

Figure 12 (a) Estimated French precipitation variances at 31 stations (values multiplied by 10 to facilitate visualization). (b) The correlation function of the variance field (as a function of distance in D-plane).

Figure 13 Estimated ($\cdot$) and predicted ($\circ$) variances (including nugget) vs. the sample variances. The solid line is the graph of $y = x$ and the dashed line represents the mean value of the variances. Vertical lines are $\pm$ one predictive standard deviation.

Figure 14 Contour plot of the estimated variance field for the French precipitation data.

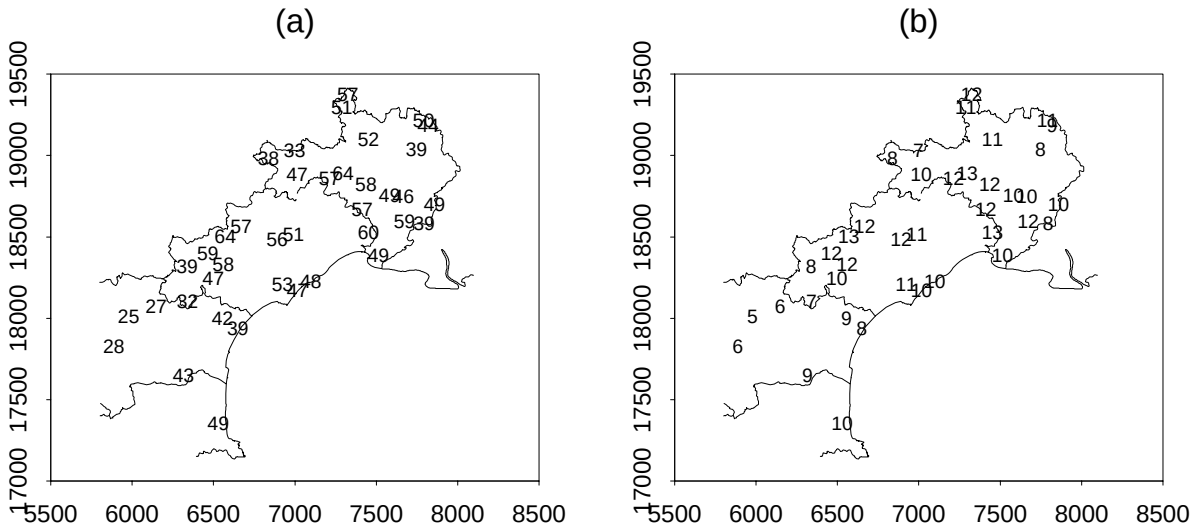
(a)
(b)
17000
17500
18000
18500
19000
19500
5500
6000
6500
7000
7500
8000
8500

Figure 1:

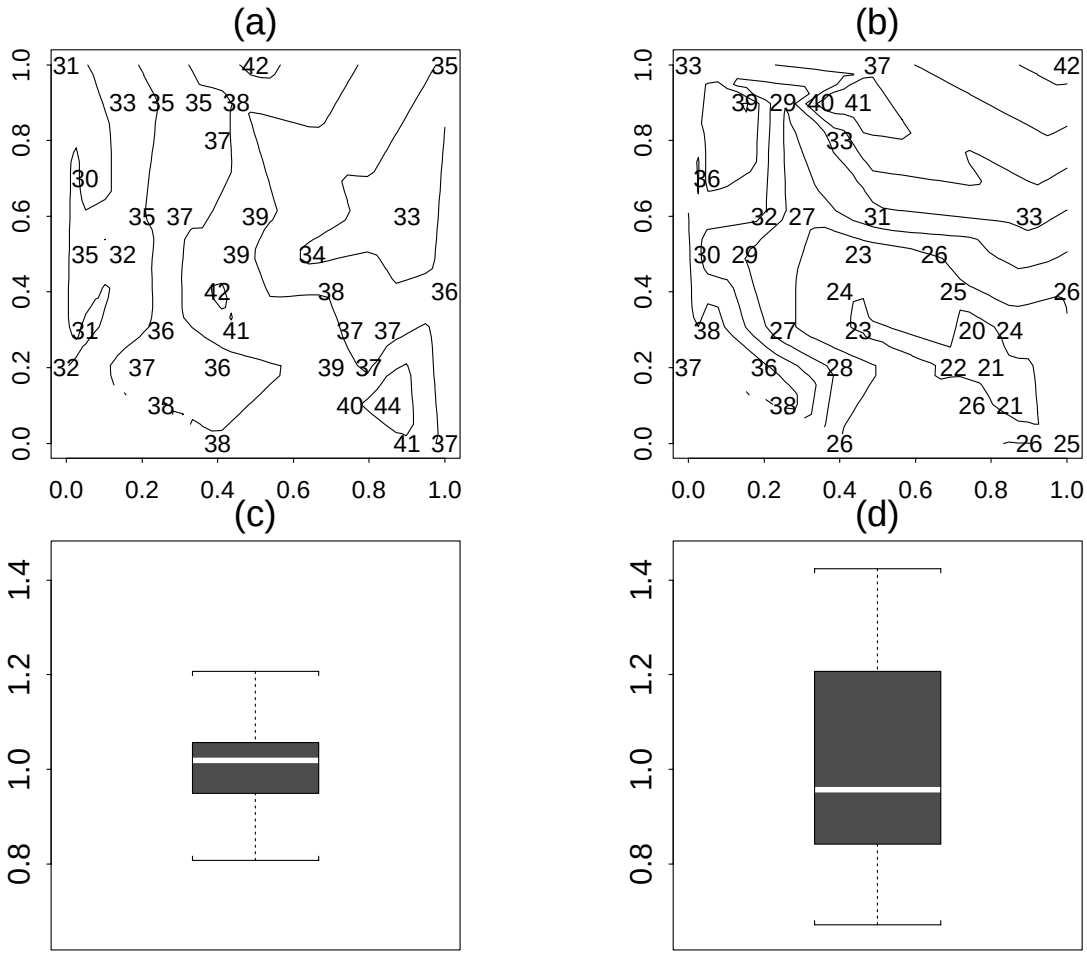

Figure 2:

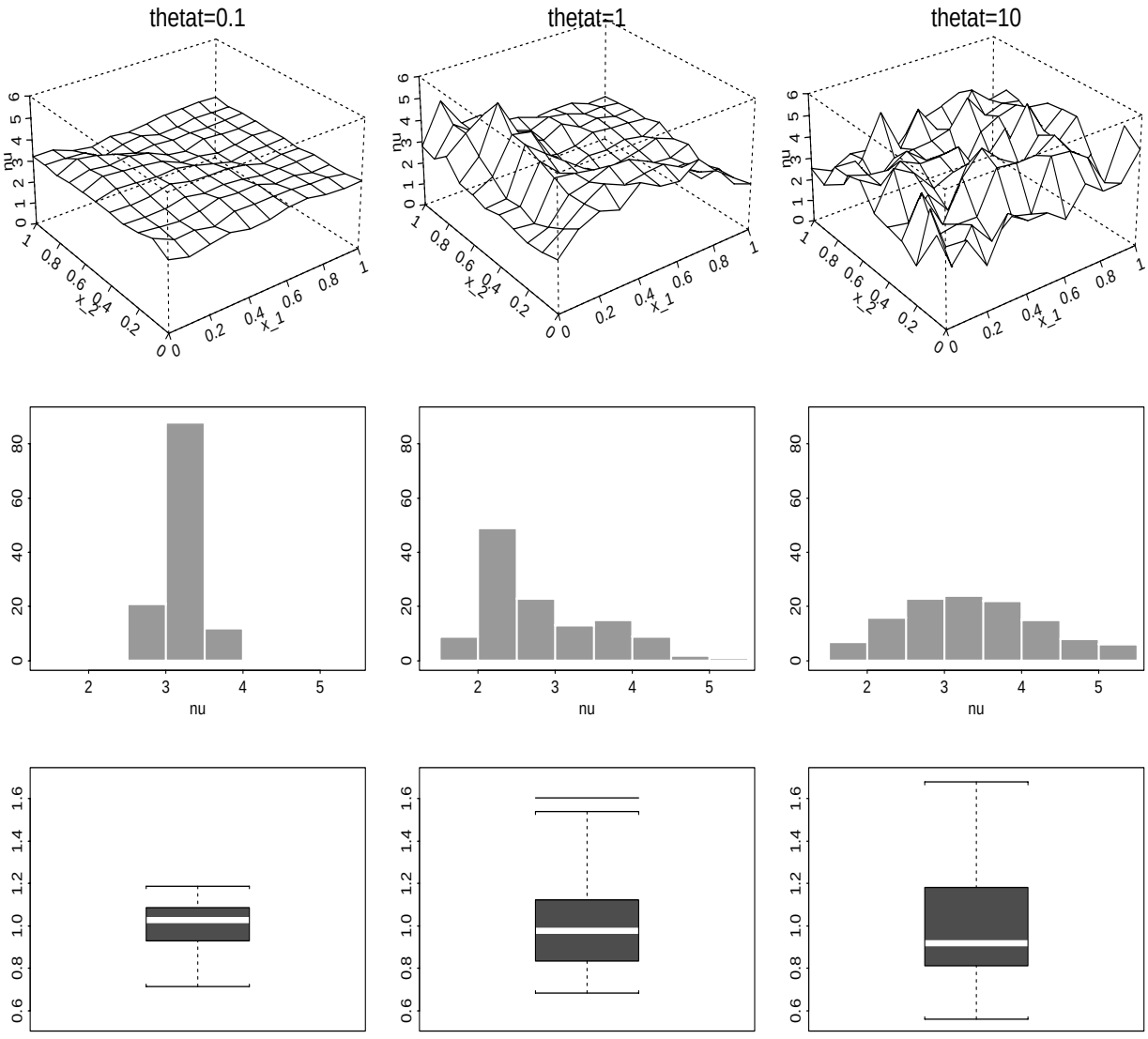

Figure 3:

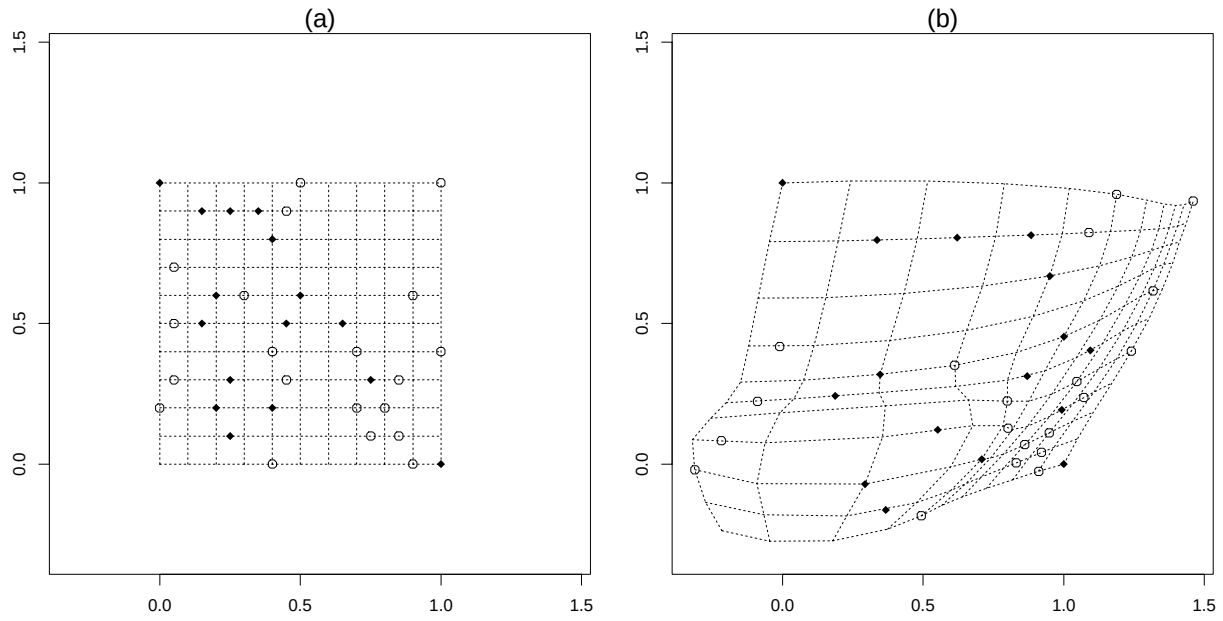

Figure 4:

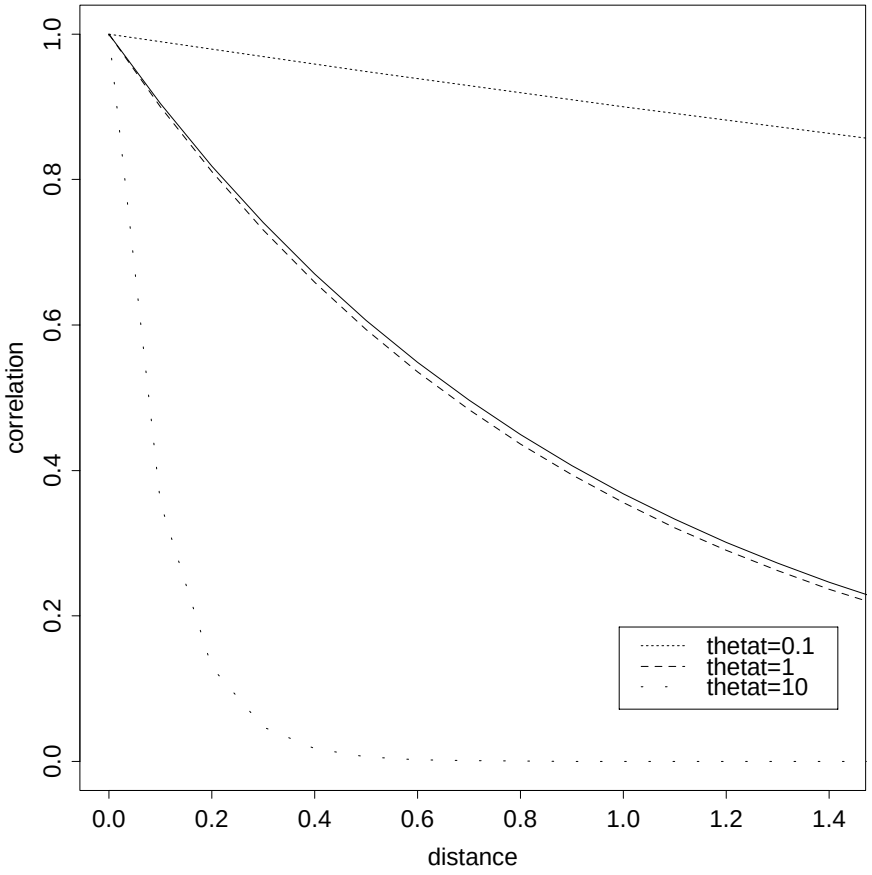
correlation
distance
0.0
0.2
0.4
0.6
0.8
1.0
1.2
1.4
thetat=0.1
thetat=1
thetat=10

Figure 5:

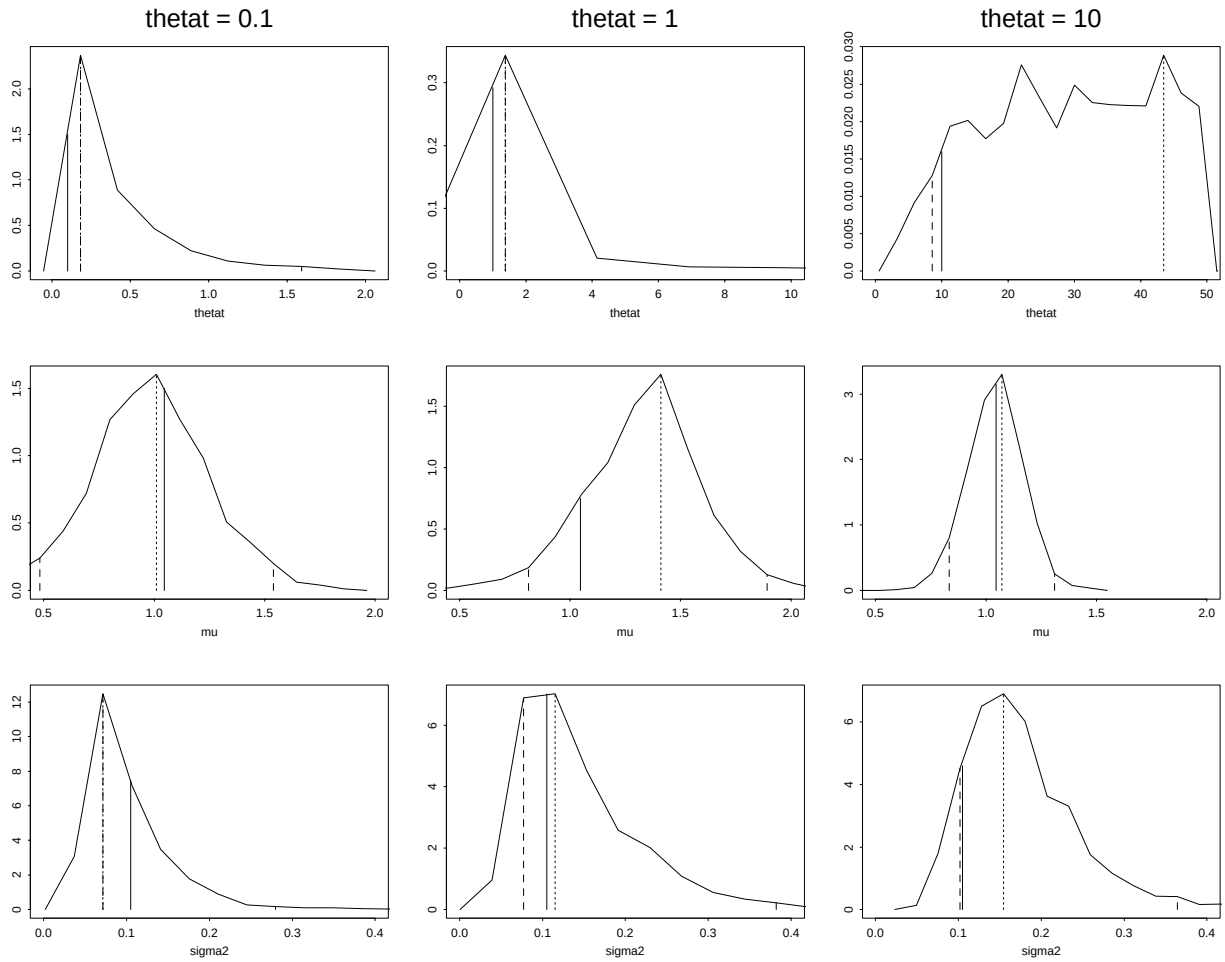
thetat = 0.1
thetat = 1
thetat = 10
thetat
thetat
thetat
mu
mu
mu
sigma2
sigma2
sigma2

Figure 6:

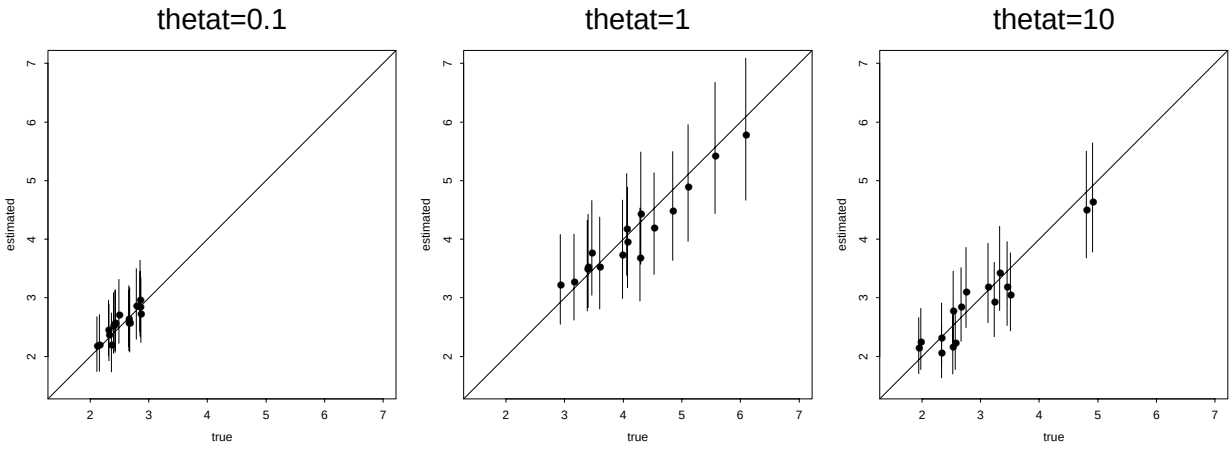

Figure 7:

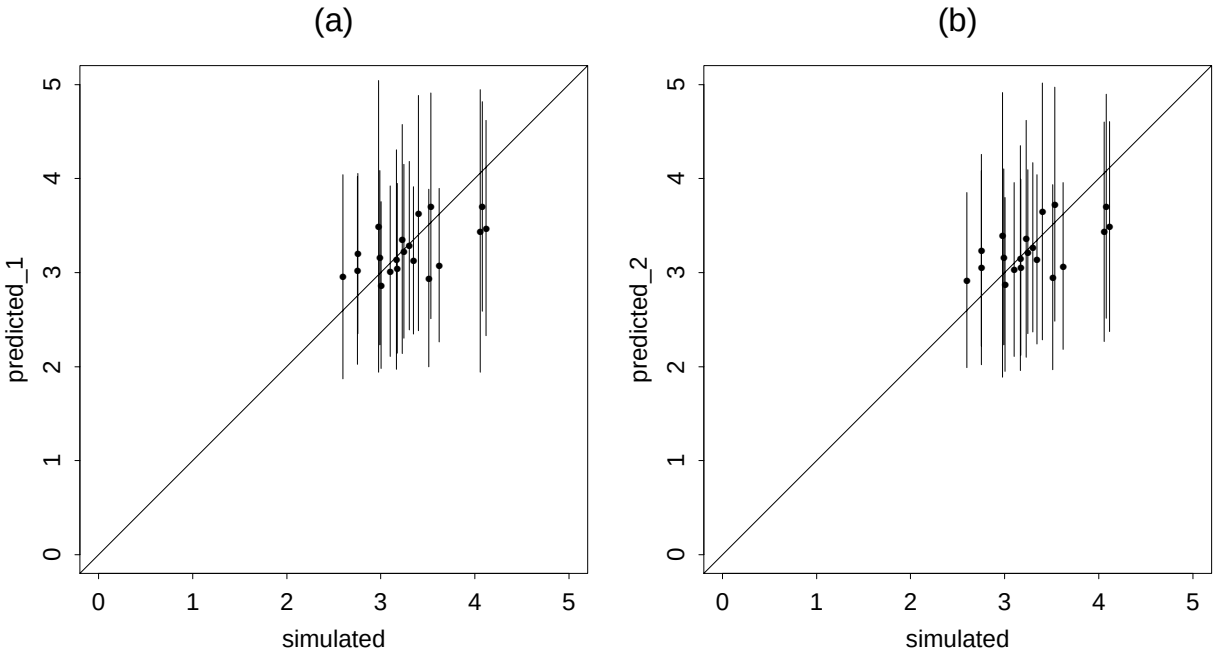

Figure 8:

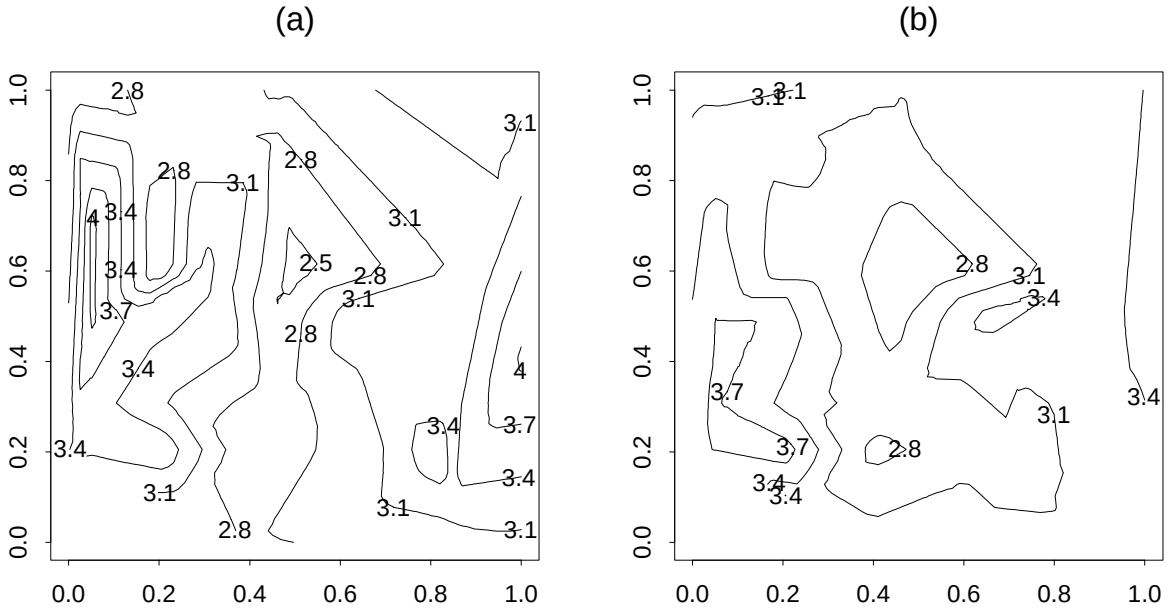

Figure 9:

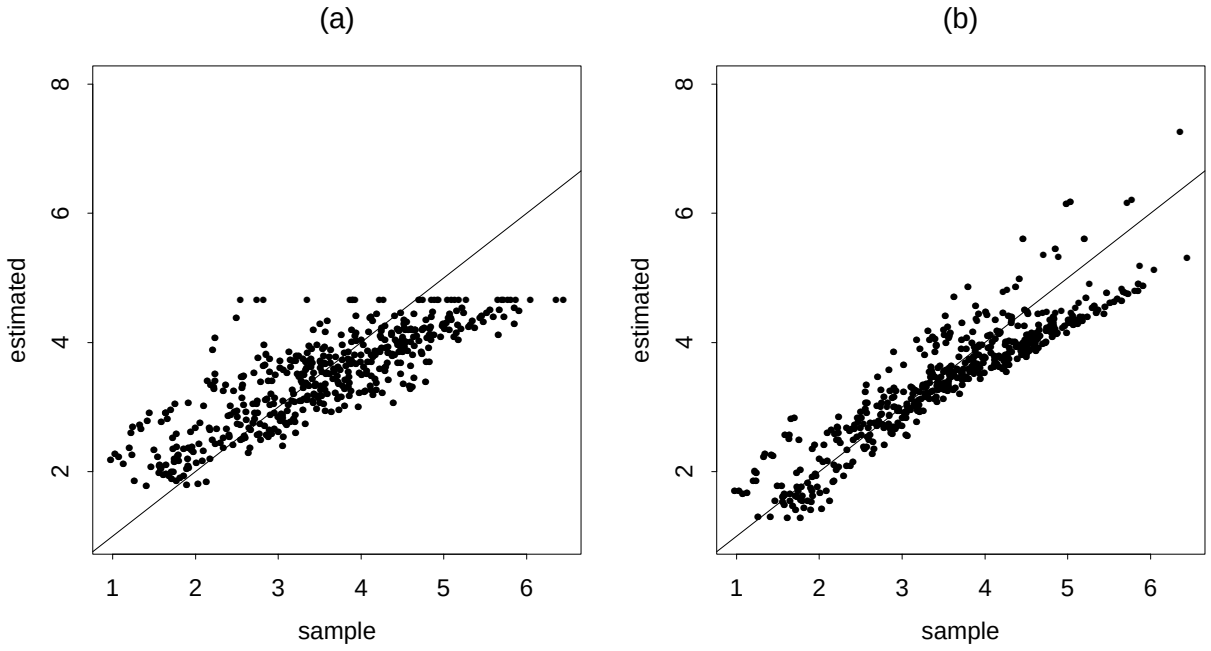

Figure 10:

(a) (b)

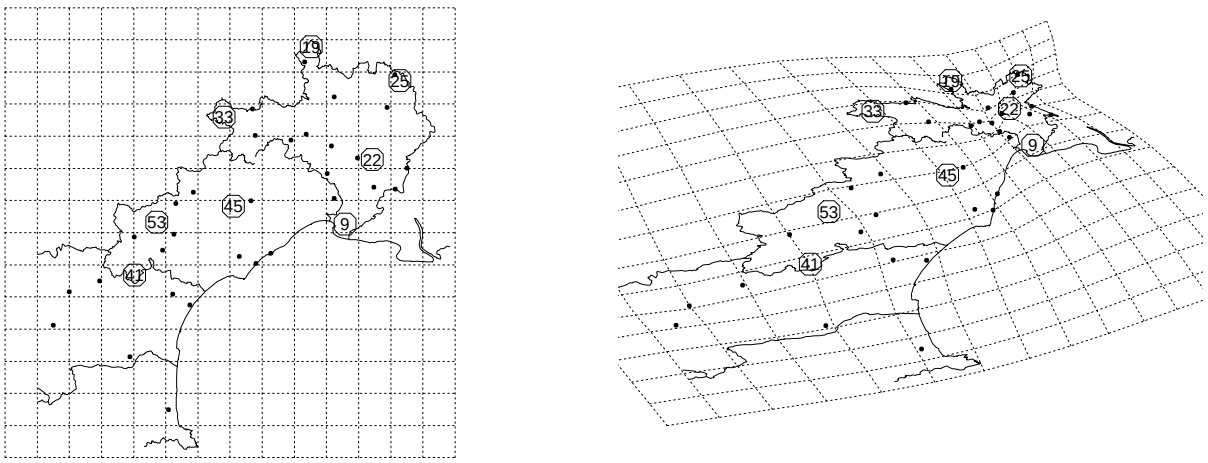

Figure 11:

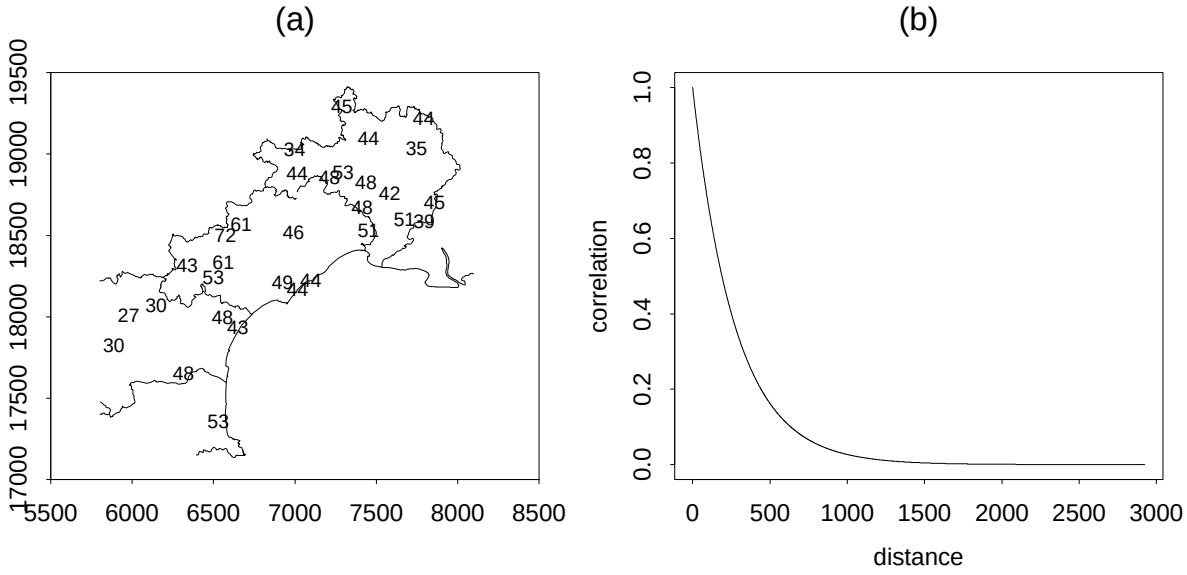
(a)
(b)
19500
19000
18500
18000
17500
17000
5500
6000
6500
7000
7500
8000
8500
correlation
distance
0.0
0.2
0.4
0.6
0.8
1.0
0
500
1000
1500
2000
2500
3000

Figure 12:

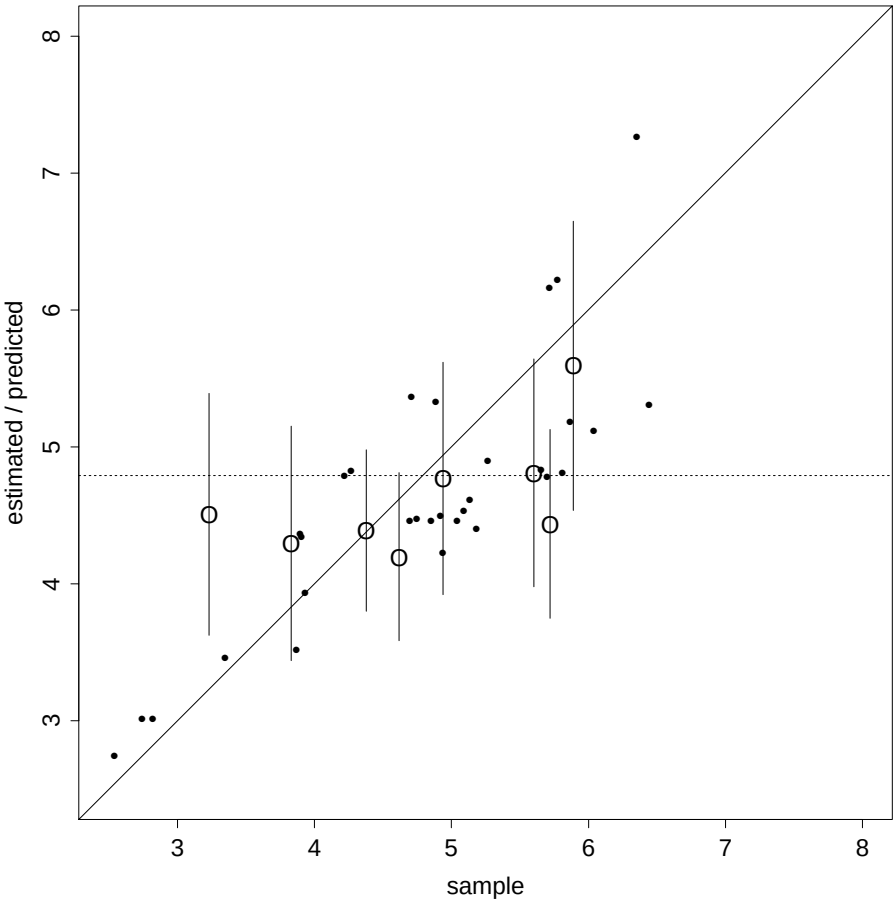

Figure 13:

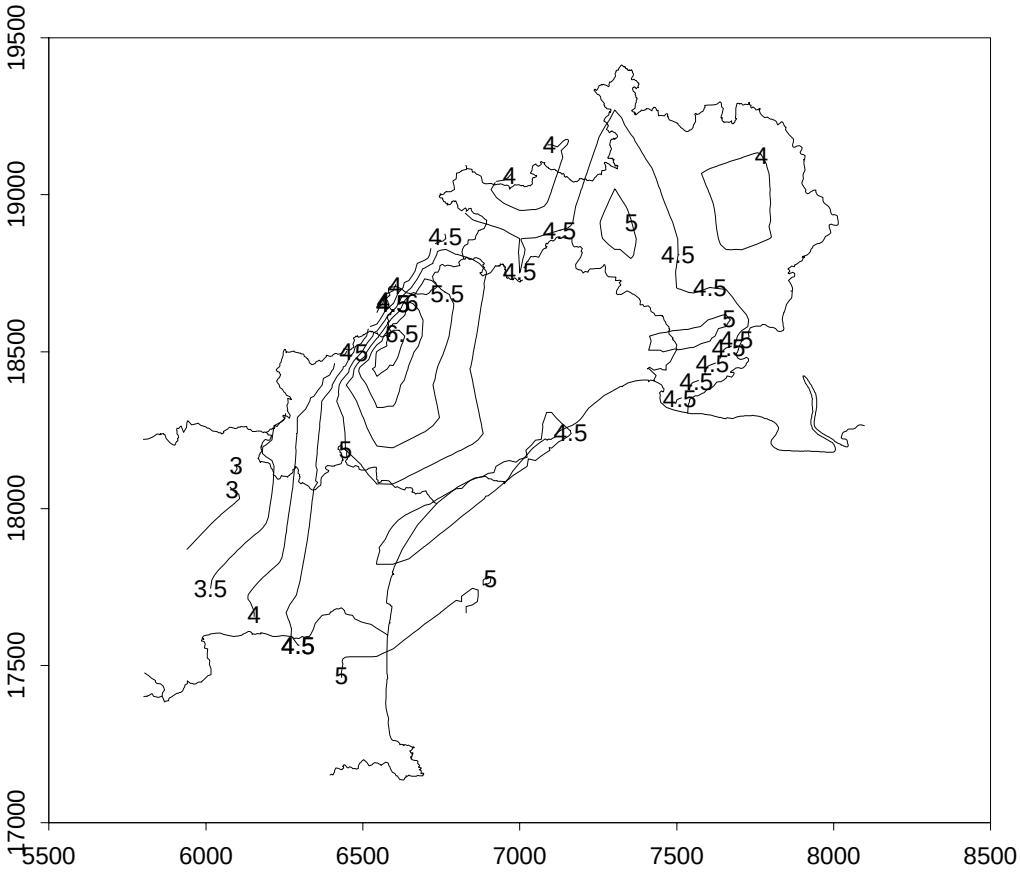
19500
19000
18500
18000
17500
17000
5500
6000
6500
7000
7500
8000
8500

Figure 14: