Bayesian Estimation of Semi-Parametric Non-Stationary Spatial Covariance Structures

Doris Damian Paul D. Sampson Peter Guttorp

NRCSE

Technical Report Series

NRCSE-TRS No. 039

The **NRCSE** was established in 1996 through a cooperative agreement with the United States Environmental Protection Agency which provides the Center's primary funding.

EPA

Bayesian Estimation of Semi-Parametric Non-Stationary Spatial Covariance Structures

Doris Damian, Paul D. Sampson, Peter Guttorp

January 25, 2000

ABSTRACT

We use the Sampson and Guttorp approach to model the non-stationary correlation function $r(x, x')$ of a Gaussian spatial process through a bijective space deformation, f, so that in the deformed space 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. Given the locations in the deformed space of a number of geographic sites at which data are available, we smoothly extrapolate the deformation to the whole region of interest. Using a Bayesian framework, we estimate jointly these locations, as well as the parameters of the correlation function and the variance parameters. The advantage of our Bayesian approach is that it allows us to obtain measures of uncertainty of all these parameters. As the parameter space is of a very high dimension, we implement an MCMC method for obtaining samples from the posterior distributions of interest. We demonstrate our method through a simulation study, and show an application to a real data set.

Keywords: Thin-plate splines, Markov Chain Monte Carlo, Gaussian spatial processes.

1 Introduction

The estimation of heterogeneous spatial covariance is an important problem in environmetrics and geostatistics. A general, semi-parametric approach to this, based on repeated observations from a network of monitoring stations, was introduced by Sampson and Guttorp (1992) and developed further in a sequence of papers (Guttorp and Sampson, 1994; Meiring et al., 1996, 1998; Perrin and Meiring, 1999). One of the advantages with the Sampson-Guttorp approach is that it enables estimation of the covariance between spatial locations where no observations have been made, without relying on an assumption of homogeneity or stationarity of the underlying space-time random field. A weakness of the current methodology is its reliance on somewhat ad hoc computationally demanding approaches to estimating the variability of such estimates (Monestiez et al., 1993, 1998).

In the last decade, Bayesian methods have become increasingly popular as an alternative to classical ones in analyses of spatio-temporal data. The traditional method of spatial prediction or interpolation (kriging) assumes knowledge of the parameters of the underlying covariance structure, whereas the Bayesian paradigm incorporates uncertainty in these parameters in predictions. Two main approaches have developed in this context. The first one may be viewed as a direct extension of the traditional kriging. For example, Le and Zidek (1992) propose a fully parametric Gaussian linear model (for a finite number of spatial locations of interest) and Handcock and Stein (1993) use prior distributions for parameters in a wide class of second-order stationary spatial models. The second approach, particularly relevant when temporal correlation as well as spatial correlation must be addressed, is a dynamic one, which makes use of Kalman filter techniques to model the mean of the space-time process recursively. Wikle et al. (1998) develop this method for data on a space-time grid.

In this paper we develop a Bayesian approach to estimating heterogeneous spatial covariance utilizing Markov chain Monte Carlo technology. Among the benefits of this approach are natural estimates of uncertainty for all different aspects of the estimation process, including spatial estimation in the context of either of the two general approaches noted above.

2 General Framework

2.1 The Model

We do not address here issues of temporal correlation and assume that temporally independent samples $Z_{it} = Z(x_i, t)$ from a spatio-temporal process are available at each of N geographic locations and at the same T points in time $i = 1, 2, ..., N; t = 1, 2, ..., T$. We consider the following model for the underlying process:

$$Z(x,t) = \mu(x,t) + \nu(x)^{1/2} E_\tau(x) + E_\epsilon(x,t) \tag{1}$$

where x denotes location and t time. $\mu(x,t)$ represents the spatio-temporal (deterministic) mean field. $E_\tau(x)$ is a zero mean, variance one, Gaussian spatial process with a correlation function that depends smoothly on the geographic coordinates. We assume that $E_\tau(x)$ is second-order continuous, i.e., $Cov(E_\tau(x), E_\tau(y)) \to 1$, as $x \to y$. $\nu(x)$ is a smooth function representing the variance (in time) of the process observed at location x. $E_\epsilon(x,t)$ represents measurement error and small-scale spatial variability. The distribution of E_ϵ is assumed to be independent of location and time, as well as of the process E_τ.

Throughout this paper, we will assume that $\mu(x,t) \equiv \mu(x)$ is constant in time. We focus our attention on modeling the covariance part of (1):

$$Cov(Z(x,t), Z(y,t)) = \begin{cases} (\nu(x)\nu(y))^{1/2} Corr(E_\tau(x), E_\tau(y)) & x \neq y \\ \nu(y) + \sigma_\epsilon^2 & x = y \end{cases} \tag{2}$$

Following Sampson and Guttorp (1992), we model the non-stationary spatial correlation function of E_τ as a function of Euclidean distances between site locations in a bijective transformation of the geographic coordinate system:

$$Corr(E_\tau(x), E_\tau(y)) = \rho_\theta(|f(x) - f(y)|) \tag{3}$$

Here, ρ_θ is a correlation function of a known parametric form (with unknown parameter(s) θ), and f is an unknown function assumed smooth and bijective.

The interpretation of this semi-parametric model is that the correlation structure is isotropic after an appropriate change in the coordinate system. Here (and in all of our applications to date) we assume that it suffices to take f as a mapping from a two-dimensional geographic system (R^2) into another $2D$ coordinate system (R^2). In general, modeling may require mapping R^2 into R^k, $2 \leq k \leq N$.

We note that in Sampson and Guttorp (1992) the model is formulated in terms of the dispersion function, $D(E_\tau(x), E_\tau(y)) = Var(E_\tau(x) - E_\tau(y))$. This function may exist when the spatial correlation function does not. However, this does not happen in our case, where the variance of E_τ is finite (it is equal to 1 by construction). As in Sampson and Guttorp (1992), we will refer to the geographic coordinate system as the $G-$space, and to the deformation of the coordinate system given by f as the $D-$space.

Perrin and Meiring (1999) prove that when ρ is strictly decreasing or when ρ, f and the inverse of f are all differentiable, the model defined in (3) is identifiable, in the sense that ρ is unique up to a scaling factor and f is unique up to homothetic transformations (i.e., translations, rotations, reflections about a line or compositions of those). These assumptions, however, do not ensure identifiability for a finite network of monitoring sites.

2.2 A Bayesian Framework

Our purpose is to estimate, based on the data Z_{it}, the transformation f, the spatial variance ν and the parameters of the correlation function ρ, the form of which must be specified in advance. These underly the matrix of true spatial covariances evaluated at the N sites, which can be expressed as:

$$\Sigma = \Sigma(\theta, \nu_i, \xi_i, i = 1, 2, ..., N) = (\sigma_{ij}),$$

where

$$\sigma_{ij} = (\nu_i \nu_j)^{1/2} \rho_\theta(|\xi_i - \xi_j|)$$

and $\xi_i = f(x_i)$ are the coordinates of site i in the $D-$space representation, θ is the vector of parameters for the class of correlation models under study, ν_i is the variance of the spatial field at site i. In this Bayesian framework, we estimate first the parameters of Σ. After estimating the ξ_i, f can be estimated by smooth interpolation of the finite mapping $x_i \to \xi_i$. In principle, any smooth interpolator may be considered. We use the thin-plate spline (Bookstein, 1991) interpolator because it provides a useful framework for specifying a prior distribution on the coordinate locations ξ_i. The prevailing method (e.g., Meiring et al., 1997) estimates the parameters by a penalized least squares method. In order to assess their variability, a computationally intensive bootstrap approach has been used. In the Bayesian framework, the variability of the

estimates is reflected in their posterior distribution. The dimension of this distribution is proportional to N, the number of geographic locations, and there is no feasible way to evaluate the multi-dimensional integral underlying its density. To avoid this, we follow the common approach of generating samples from the posterior distribution using an MCMC algorithm.

3 Estimation

3.1 The Likelihood

The multivariate normal density of the observations can be expressed in terms of the vector of site means, $\bar{z}$, and the sample covariance matrix S, ($S_{ij} = \sum_{t=1}^{T}(z_{it} - \bar{z}_{i.})(z_{jt} - \bar{z}_{j.})$), (which is nonsingular only if $T > N$):

$$f(\{z_{it}\} \mid \mu, \Sigma) = |2\pi\Sigma|^{-T/2} \exp\left\{-\frac{T}{2} tr\Sigma^{-1}S - \frac{T}{2}(\bar{z}-\mu)^T\Sigma^{-1}(\bar{z}-\mu)\right\}.$$

Setting $\hat{\mu} = \bar{z}$ (the mean estimate under a non-informative prior), the likelihood simplifies to:

$$P(S \mid \Sigma) = |2\pi\Sigma|^{-(T-1)/2} \exp\left\{-\frac{T}{2} tr\Sigma^{-1}S\right\}.$$

We will base inferences on this likelihood.

3.2 Prior Distributions

Our framework requires that we compute a mapping f (with the constraints $f(x_i) = \xi_i$) that is a bijective transformation over the geographic domain of interest. As we compute the mapping using thin-plate splines, there is no guarantee that the result will not "fold" (i.e. that the result will indeed be bijective). Since the folding cannot be assessed analytically, we cannot assign zero prior probability to configurations that result in folding. Instead, we use a prior distribution that is natural for thin-plate spline mappings and which penalizes non-smooth (bending) maps, including those that fold. (This prior was suggested by Mardia, Kent and Walder (1991)).

In the calculation of a thin-plate spline, the bending energy of the spline—reflecting deviation from an affine transformation—can be expressed as a quadratic form in the image coordinates. The bending energy matrix underlying this quadratic form is a function only of the geographic coordinates. Call this matrix $K = K(x_i, i =$

$1, 2, \ldots, N$). K arises naturally in a kriging system with spatial covariance structure of the form $\sigma(d) = |d|^2 \log(|d|^2)$. For further reference see Bookstein (1989), Bookstein (1991), Sampson et al. (1991), or Mardia et al. (1991). Let $\xi_i = (\xi_{i1}, \xi_{i2})$ denote two-dimensional locations, and let $\boldsymbol{\xi_1} = (\xi_{11}, \ldots, \xi_{N1})^T$, $\boldsymbol{\xi_2} = (\xi_{12}, \ldots, \xi_{N2})^T$ and $\Xi = [\boldsymbol{\xi_1}\, \boldsymbol{\xi_2}]$. Then, the bending energy is proportional to

$$\boldsymbol{\xi_1}^T K \boldsymbol{\xi_1} + \boldsymbol{\xi_2}^T K \boldsymbol{\xi_2}$$

We specify a prior on the spatial configuration Ξ as a normal distribution based on the bending energy:

$$\pi(\Xi) = \mathrm{P}(\{\xi_i\}) \propto \exp\{-\frac{1}{2\tau^2}(\boldsymbol{\xi_1}^T K \boldsymbol{\xi_1} + \boldsymbol{\xi_2}^T K \boldsymbol{\xi_2})\}.$$

It should be noted that this density is improper because the matrix K is of rank $N - 3$. This density is flat over the space of $\{\xi_i\}$ that can be obtained from the $\{x_i\}$ via affine transformations. By specifying the scale parameter τ we determine how much the prior (relative to the likelihood) will penalize mappings with high bending energy—the mappings most likely to result in folding. The prior on the other covariance parameters, ν_i and θ, will depend on the covariance structure thought to be feasible for the data under study. We will assign priors for θ and ν_i independently of Ξ_i so that the joint prior, which we will denote by $\pi(\Sigma)$, is:

$$\pi(\Sigma) = \pi(\theta, \nu_i, i = 1, 2, \ldots, N)\pi(\Xi).$$

The posterior to be maximized is:

$$\pi(\Sigma|S) \propto P(S|\Sigma)\, \pi(\Sigma). \tag{4}$$

3.3 The MCMC Algorithm

The algorithm implemented here is a Metropolis-Hastings type one (e.g., Gelman et al. (1996)). It alternates between sampling of new geographic configurations and new values for the other covariance parameters. The geographic configuration is sampled using a multivariate normal kernel (in a Metropolis step); vectors of the x and y coordinates are sampled independently, so that the proposed new configuration is not restricted to any particular direction in the plane. We have found it useful though, in order to avoid proposals of high bending energy and to increase acceptance rates, to restrict points close together in the initial configuration to "move together". This

has been obtained by imposing a spatial structure in the covariance matrix of the multivariate kernel, which decreases with geographic distance. The other covariance parameters are sampled from kernel distributions that are not necessarily symmetric (in Metropolis-Hastings steps) about current values.

Schematically, we let the algorithm start with some initial values of the parameters ν_i and θ, and with an initial configuration Ξ^0 equal to the geographic configuration of the sites. At the m^{th} iteration, the current parameters are denoted $\theta = \theta^{(m-1)}$, $\nu = \nu^{(m-1)}$, $\Xi^{(m-1)} = [\boldsymbol{\xi}_1^{(m-1)}\, \boldsymbol{\xi}_2^{(m-1)}]$, respectively. Then,

- θ^* and ν^* are sampled from a proposal distribution: $Q(\theta^*, \nu^* \mid \theta^{(m-1)}, \nu^{(m-1)})$.
- p_1 is calculated:

$$p_1 = \frac{\pi(\Sigma(\theta^*, \nu^*, \xi^{(m-1)}) \mid S) Q(\theta^{(m-1)}, \nu^{(m-1)} \mid \theta^*, \nu^*)}{\pi(\Sigma(\theta^{(m-1)}, \nu^{(m-1)}, \xi^{(m-1)}) \mid S) Q(\theta^*, \nu^* \mid \theta^{(m-1)}, \nu^{(m-1)})}.$$

- $\theta^{(m)}$ and $\nu^{(m)}$ are set:

$$(\theta^{(m)}, \nu^{(m)}) = \begin{cases} (\theta^*, \nu^*) & \text{with probability } min(p_1, 1) \\ (\theta^{(m-1)}, \nu^{(m-1)}) & \text{otherwise} \end{cases}$$

- A new configuration, $\Xi^* = [\boldsymbol{\xi_1}^*\, \boldsymbol{\xi_2}^*]$, is sampled by combining $\boldsymbol{\xi_1}^* \sim MVN(\boldsymbol{\xi_1}^{(m-1)}, B)$ with $\boldsymbol{\xi_2}^* \sim MVN(\boldsymbol{\xi_2}^{(m-1)}, B)$. As explained above, the elements of the matrix B are decreasing functions of the geographic distances.
- p_2 is calculated:

$$p_2 = \frac{\pi(\Sigma(\theta^{(m)}, \nu^{(m)}, \Xi^*) \mid S)}{\pi(\Sigma(\theta^{(m)}, \nu^{(m)}, \Xi^{(m-1)}) \mid S)}.$$

- $\Xi^{(m)}$ is set:

$$\Xi^{(m)} = \begin{cases} \Xi^* & \text{with probability } min(p_2, 1) \\ \Xi^{(m-1)} & \text{otherwise} \end{cases}$$

These steps are repeated until the chain is judged to have converged. Thereafter, samples from the posterior distributions can be obtained. We assess the convergence of the chain both by examining the behavior of the log(posterior), and by monitoring changes in the ratio of between to within chains variability (as expressed by the Gelman and Rubin $\hat{R}$ statistic, which should be close to one at convergence (Gelman et al., 1996)), when running the chain with several different starting configurations.

4 Simulation

Model (1) accommodates heterogeneity in the variance of the spatial process, $\nu(x)$, as well as the spatial correlation structure. In this first simulation we have tested the performance of the MCMC algorithm in a simpler case, in which the spatial variance is constant ($\nu(x) \equiv \nu$), and the covariance structure is anisotropic but stationary. We also assume that there is no small-scale spatial variability ($E_\epsilon(x,t) \equiv 0$).

At each of $N = 10$ monitoring sites, $x_1, ..., x_N$, $T = 400$ independent observations $(z^*_{i1}, ...z^*_{iT})$ were sampled, $z^*_{ij} \sim N(0,1)$. The bijective transformation f for a stationary, anisotropic correlation structure is affine and specified by a 2×2 matrix, A. The form of the correlation model is exponential:

$$\sigma_{ij} = \nu \exp(-\theta\,|A(x_i) - A(x_j)|). \tag{5}$$

We set the parameters $\nu = 1$ and $\theta = 0.003$. Figure 1 shows the true and the deformed configurations (the G and the D planes), and a biorthogonal grid for the affine transformation. The idea behind biorthogonal grids (see Sampson et al. (1991)) is that a smooth transformation has at any point a locally linear approximation, given by the matrix of its partial first derivatives at the point, Δ. The eigenvectors of $\Delta^T\Delta$ are called principal axes. The direction of the axis corresponding to its largest eigenvalue is the direction in which the plane is (locally) most stretched — this is the direction of the weakest spatial correlation. The strength of the (local) stretching is given by the square roots of the eigenvalues of $\Delta^T\Delta$ (referred to as gradients,"grad", in our figures). In the case of an affine transformation, the linear approximation is of course exact. The square roots of the eigenvalues of A^TA are 2.3 and 0.9.

For both ν and θ we use exponential prior distributions with parameter 1, mutually independent and independent of the prior distribution for the $\boldsymbol{\xi_i}$. These have been chosen purely for convenience and, as shown in figure 2, have little influence on the posterior distributions. The proposal distribution for ν^* is $\Gamma(40, \nu/40)$ and for θ^* —

$\Gamma(30, \theta/30)$. (The parameters of the two Gamma distributions were chosen in order to provide MCMC acceptance rates of about 20%).

The form of the prior distribution of the $\boldsymbol{\xi_i}$ has been discussed in section 3.2. In this example, we set the scale parameter $\tau = 1$. We intend to investigate the influence of this parameter on the prior distribution in future work. As explained in section 3.3, $\boldsymbol{\xi_1}^*$ and $\boldsymbol{\xi_2}^*$ are sampled independently from multivariate normal distributions with parameters $(\boldsymbol{\xi_1}, B)$ and $\boldsymbol{\xi_2}, B)$ respectively. The components of the covariance matrix, B, decrease exponentially with the (geographic) distance between sites ($b_{ij} = v \exp(-t|x_i - x_j|)$. We have used here $v = 4$ and $t = 0.02$ in order to ensure that about 30% of the proposed configurations will be accepted. The first two rows of B have been set to 0, so that two points are held fixed.

The distribution of the log(posterior) seems to stabilize after the first 5,000 iterations. The maximum square-root of $\widehat{R}$ when running the chain with five different starting configurations decreases from 3.0 in 5,000 to 1.17 in 25,000 iterations. We have therefore decided to discard the first 50,000 iterations. Thereafter, samples of length 500 from the posterior distributions of ν, θ and ξ were obtained, by sampling every 100^{th} iteration. We thus estimate the spatial configuration, ξ, by the mean of 500 sampled configurations. ν and θ are estimated similarly by the means of the 500 sampled parameters . The estimated ν is 0.996 ($95\% C.I. = (0.926, 1.069)$), and the estimated θ is 0.0029 ($95\% C.I. = (0.0025, 0.0033)$). (C.I. = Credibility Interval). Figure 2 shows histograms of the empirical posterior distributions of ν and θ. The plots in figure 3 show the estimated transformation and its biorthogonal grid; these clearly resemble very much those in figure 1. The sample covariances as functions of the distance in the G, the D and the estimated D planes are shown in figure 4 – their variability clearly decreases in the D plane (the true as well as the estimated one). We have superimposed on these plots the true and the estimated covariance functions.

One way of depicting the uncertainty in the estimated transformation is by plotting the principal axes of the 500 sampled transformations at given points. In figure 5 we have superimposed the 500 pairs of principal axes at monitoring site 3. From this figure, it can be seen that the transformation in the vicinity of the point is satisfactorily stable.

We can also assess the variance and the bias of the estimated covariances, both among monitoring sites and between locations at which there are no observations. The box-plots in figure 6 summarize the estimated 500 covariances between site 3 and other locations and between location (0,0) (at which no observations are available) and other locations. The inter-quartile ranges are small (about 0.1), and the true

covariances are quite close to the estimated ones, as expected.

5 Application

We have compared our Bayesian estimation approach to the penalized least squares estimation, by applying it to the French precipitation data described in Meiring et al. (1997). The 108 altitude-adjusted 10-day aggregated precipitation records for 39 sites in the Languedoc-Roussillon region have been first log-transformed and the site-specific means have been removed. Unlike Meiring et al. (1997), we have not standardized the time series by dividing by site-specific standard deviations. We believe that the assumption of constant spatial variance may not be adequate in this example — as the spread of the sample variances (covariances at distance zero) indicates in figure 8; a more thorough analysis of these data will have to account for this aspect. The deformed map representing the posterior mean configuration (as shown in figure 7) has the same features as the one in Meiring et al. (1997), figure 2: the central region has been "stretched", indicating a relatively low spatial covariance, whereas the north-western region has been "shrunk". The estimated exponential covariance function superimposed on the plot of the sample covariances (in figure 8) lies below the main cloud of points — we have checked through other simulations (the results of which are not presented here) that this feature is not an indicator of poor estimation, but rather a side-effect of the dependences among the sample covariances.

Posterior credibility intervals can be used to assess variability in the estimated covariances, or any function of them. Figure 9 is an example: following figure 4 in Meiring et al. (1997), it depicts boxplots of the dispersions between site 45 and other sites. The bootstrap method implemented in Meiring et al. (1997) for this purpose is extremely computationally expensive, and requires additional, post-estimation computations. Our boxplots are based on the same samples from the posterior distribution that have been stored for obtaining the estimates. We note that the relative positions of these boxplots are similar to the ones shown in Meiring et al. (1997), figure 4. (There are scale differences due to the fact that in Meiring et al. (1997) the series have been previously standardized). The number of empirical dispersions that fall outside the interquartile ranges is comparable as well (6 in the present paper, 5 in Meiring et al. (1997)).

6 Discussion

We have introduced a method of estimating non-stationary semi-parametric covariance structures within a Bayesian framework. The main advantage of this method is the ease with which measures of variability of any quantity of interest can be obtained through sampling from their posterior distributions. The isotropic correlation function applied to the D-plane, here assumed exponential, can be readily replaced with other models (as described, for example in Matérn, (1986), chapter 2). Fully general (nonparametric) isotropic correlation functions might also be considered, in principle, although their computation and interpretation poses a number of challenges.

Due to the high dimensionality of the posterior distribution of the parameters involved, we have obtained samples from this distribution through a Markov Chain Monte Carlo algorithm. We intend to investigate further the specification of the prior distribution on the D-plane — or, equivalently, the smoothness of the underlying deformation representing nonstationarity — with a view to implementing it as a truly informative and interpretable prior.

The statistical model used in this preliminary work incorporated two simplyfying assumptions: constant spatial variance ($\nu(x) \equiv \nu$) and no temporal trend ($\mu(x,t) \equiv \mu(x)$). Both of these will be relaxed in future work. In addition, we will address the interpolation of the spatially varying mean structure in order to provide a complete Bayesian modeling and analysis framework for non-stationary spatio-temporal data.

Acknowledgments

The authors wish to thank Ruth Grossmann and Noam Shoresh for valuable help in developing the code for the simulations, and Wendy Meiring and Pascal Monestiez for providing the French precipitation data used in the Application section.

Although 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 required peer and policy review and therefore does not necessarily reflect the views of the Agency and no official endorsement should be inferred.

References

[1] Bookstein, F.L. (1989): Principal warps: Thin-plate splines and the decomposition of deformations. *I.E.E.E. Trans. Patt. Anal. Mach. Intell.* 11, pp. 567-585.

[2] Bookstein, F.L. (1991): *Morphometric Tools for Landmark Data*, Cambridge University Press.

[3] Cressie, N.A.C. (1991): *Statistics for Spatial Data*, Wiley Series in Probability and Mathematical Statistics.

[4] Handcock, M.S. and Stein, M.L. (1993): A Bayesian Analysis of Kriging. *Technometrics* 35, No. 4, pp. 403-410.

[5] Gelman, A., Carlin, J.B., Stern, H.S., and Rubin, D.B. (1996): *Bayesian Data Analysis*, Chapman and Hall.

[6] Le, N.D. and Zidek, J.V. (1992): Interpolation with Uncertain Spatial Covariances: A Bayesian Alternative to Kriging. *Journal of Multivariate Analysis* 43, No. 2, pp. 351-374.

[7] Mardia, K.V., Kent, J.T., and Walder, A.N. (1991): Statistical shape models in image analysis. *Proceedings of the 23rd Symposium on the Interface between Computing Science and Statistics*, pp. 550-557. Interface Foundation, Fairfax Station.

[8] Matérn, B. (1986): *Spatial Variation*, Springer-Verlag.

[9] Meiring, W., Monestiez, P., Sampson, P.D. and Guttorp, P. (1997): Developments in the Modeling of Nonstationary Spatial Covariance Structure from Space-Time Monitoring Data. *Geostatistics Wollongong '96*, vol.1, pp. 162-173. E.Y. Baafi and N. Schofield (eds.), Kluwer Academic Publishers.

[10] Perrin, O. and Meiring, W. (1999): Identifiability for Non-Stationary Spatial Structure. To appear in *Applied Probability*.

[11] Sampson, P.D., Lewis, P., Guttorp, P., Bookstein, F.L., and Hurley, C. (1991): Computation and interpretation of deformations for landmark data in morphometrics and environmetrics. In *Proceedings of the 23rd Symposium on the Interface between Computing Science and Statistics*, pp. 534-541. Interface Foundation, Fairfax Station.

[12] Sampson, P.D., and Guttorp, P. (1992): Nonparametric estimation of nonstationary spatial covariance structure. *J. Amer. Statist. Assoc.* 87, pp 108-119.

[13] Wikle, C.K., Berliner, L.M. and Cressie, N. (1998): Hierarchical Bayesian Space-Time Models. *Environmental and Ecological Statistics* 5, pp. 117-154.

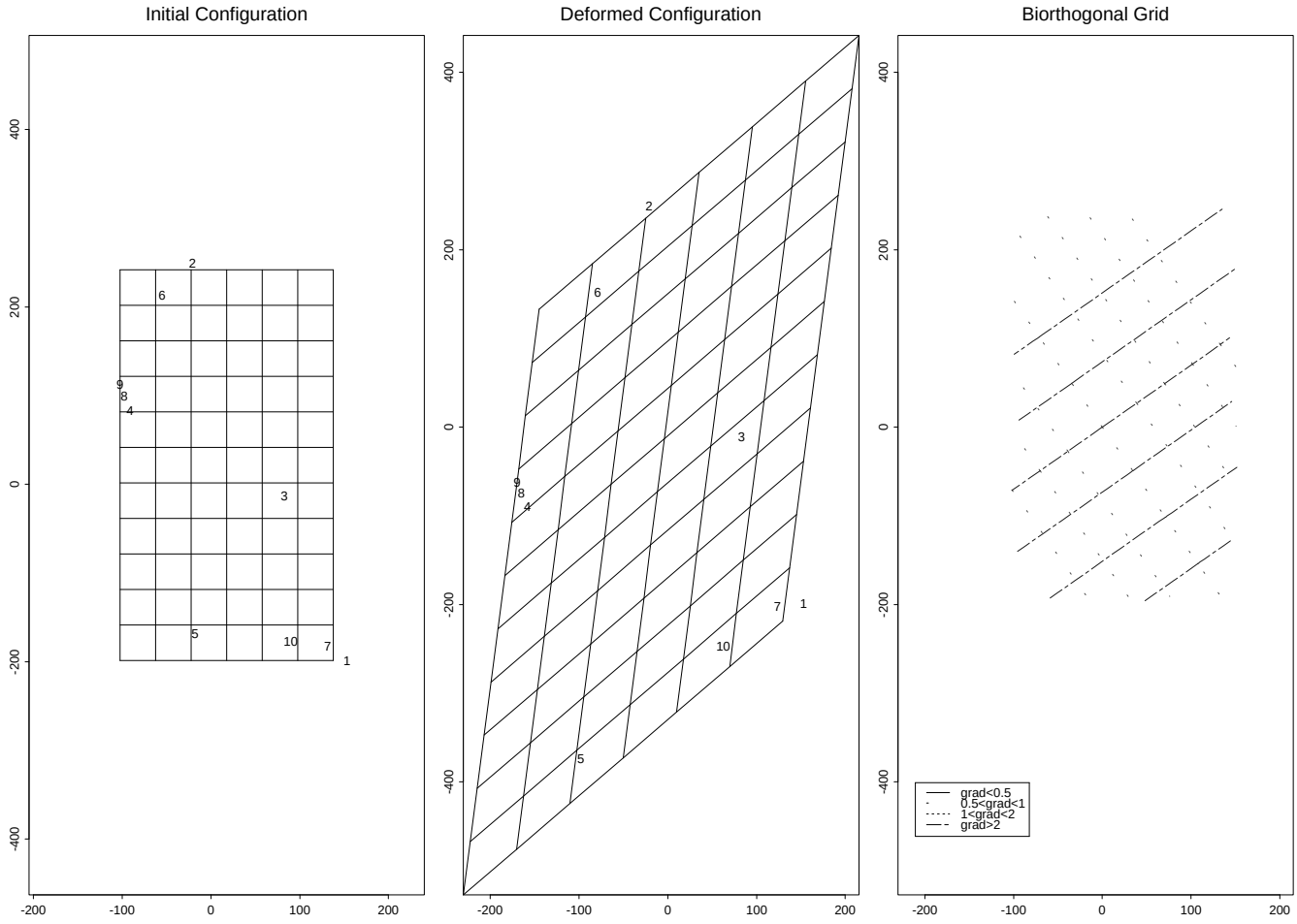

Figure 1: The Affine Transformation of the 10 Sites and Its Biorthogonal Grid

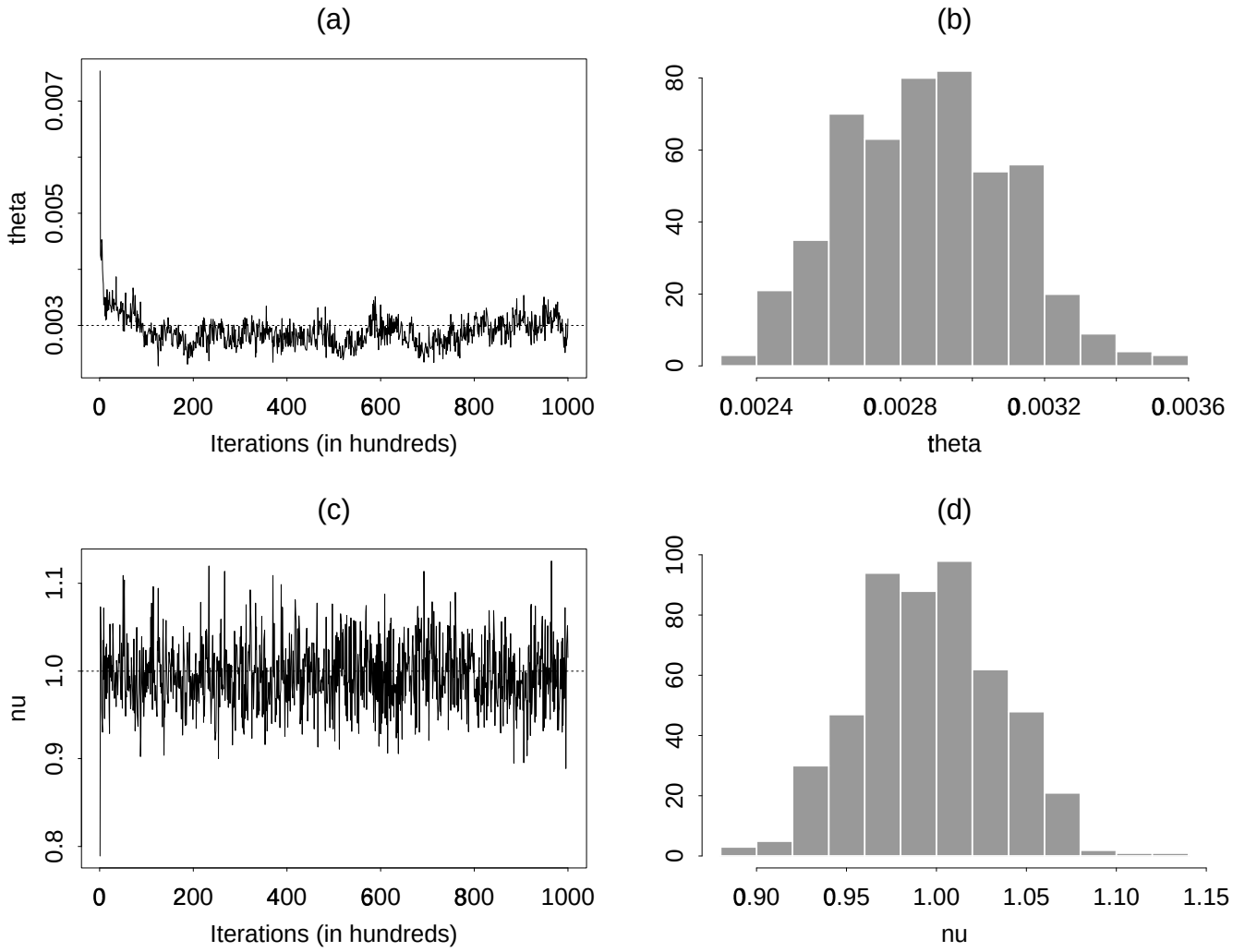

Figure 2: Profile plots of theta and nu show a quick stabilization around the true values ((a) and (c)) and histograms of the 500 samples from the respective posterior distributions ((b) and (d)) are bell-shaped, centered close to the true values

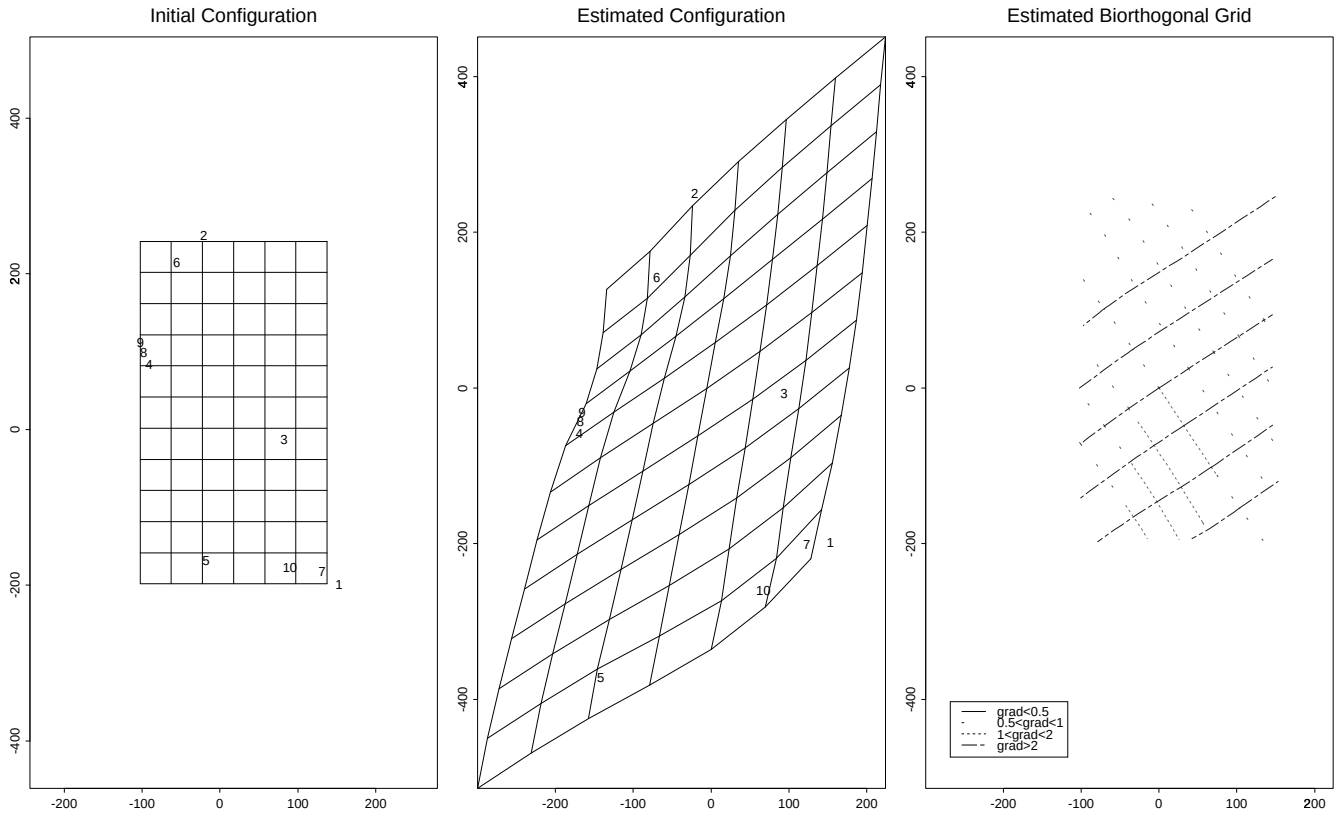

Figure 3: Estimated Transformation and Its Biorthogonal Grid

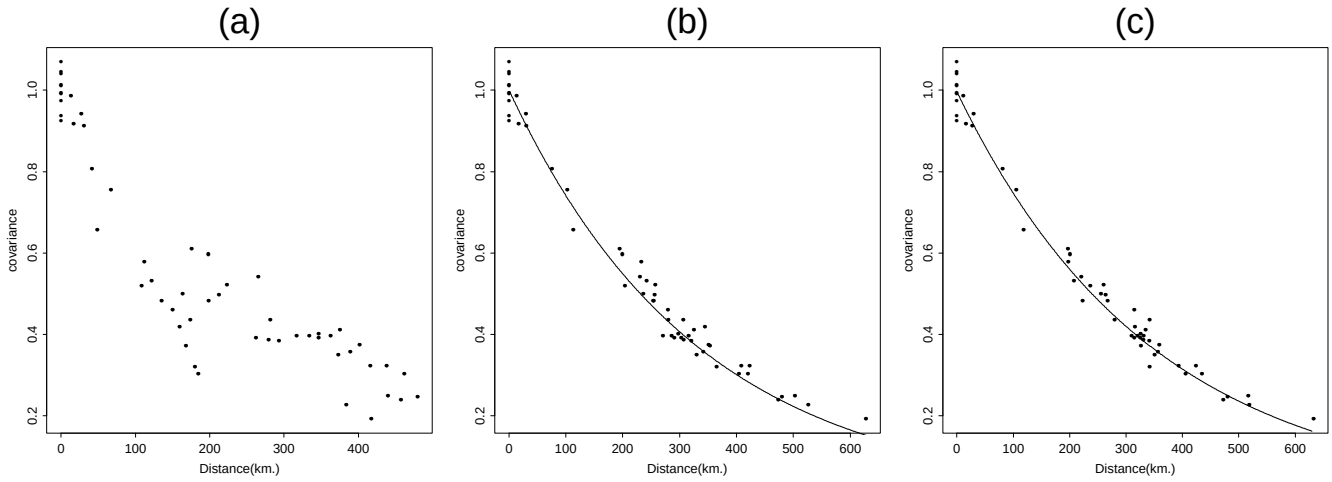

Figure 4: Sample Covariances as a Function of Distances - (a) in the "G" plane, (b) in the "D" plane, and (c) in the estimated "D" plane

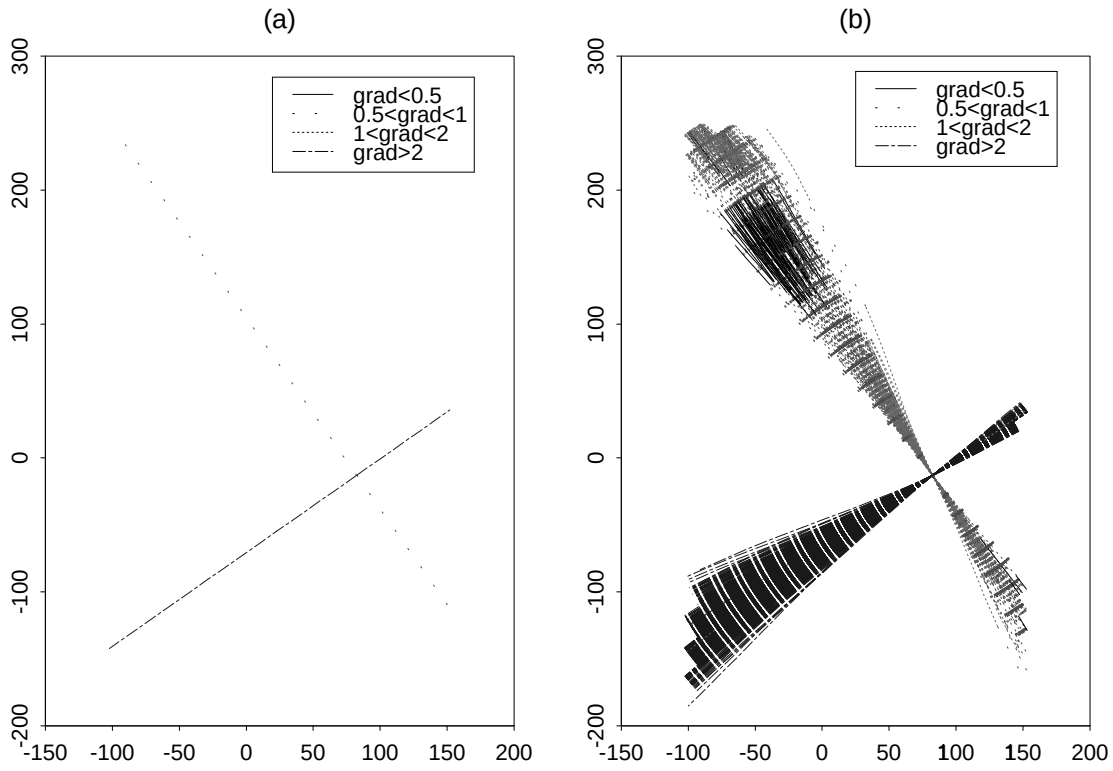

Figure 5: Assessing Uncertainty in the Transformation at Site 3: (a) true principal axes; (b) estimated principal axes.

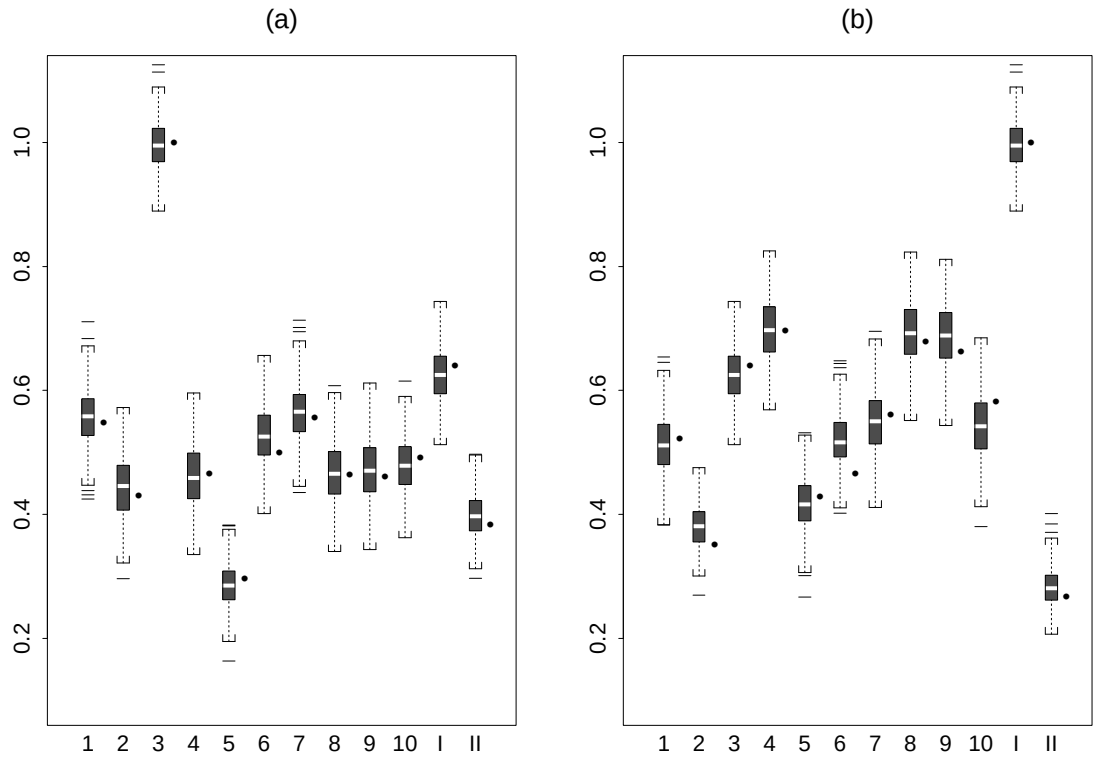

Figure 6: Estimated Covariances: (a) between site 3 and other locations; (b) between location (0,0) and other locations. The numbers denote monitoring sites. I and II denote locations (0,0) and (100,150), respectively. The dots represent the true values.

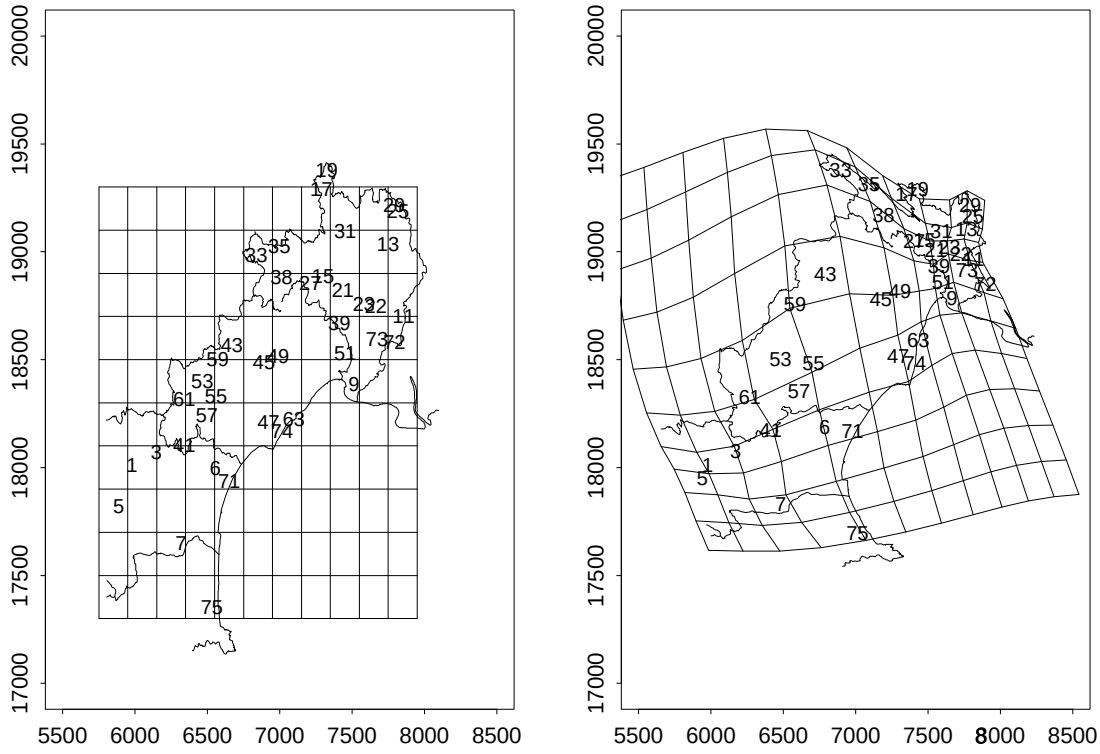

Figure 7: Geographic and Deformed Maps of Languedoc-Roussillon.

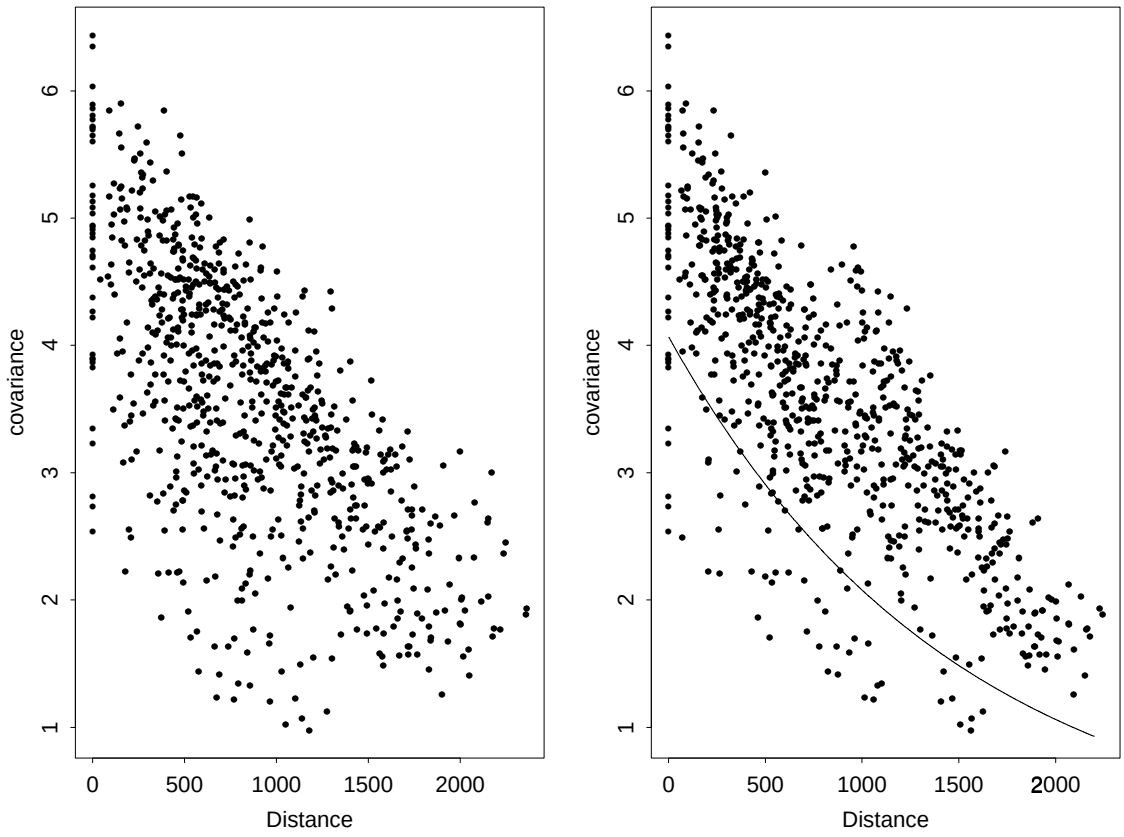

Figure 8: Sample Covariances as a Function of Distances in the "G" and estimated "D" planes

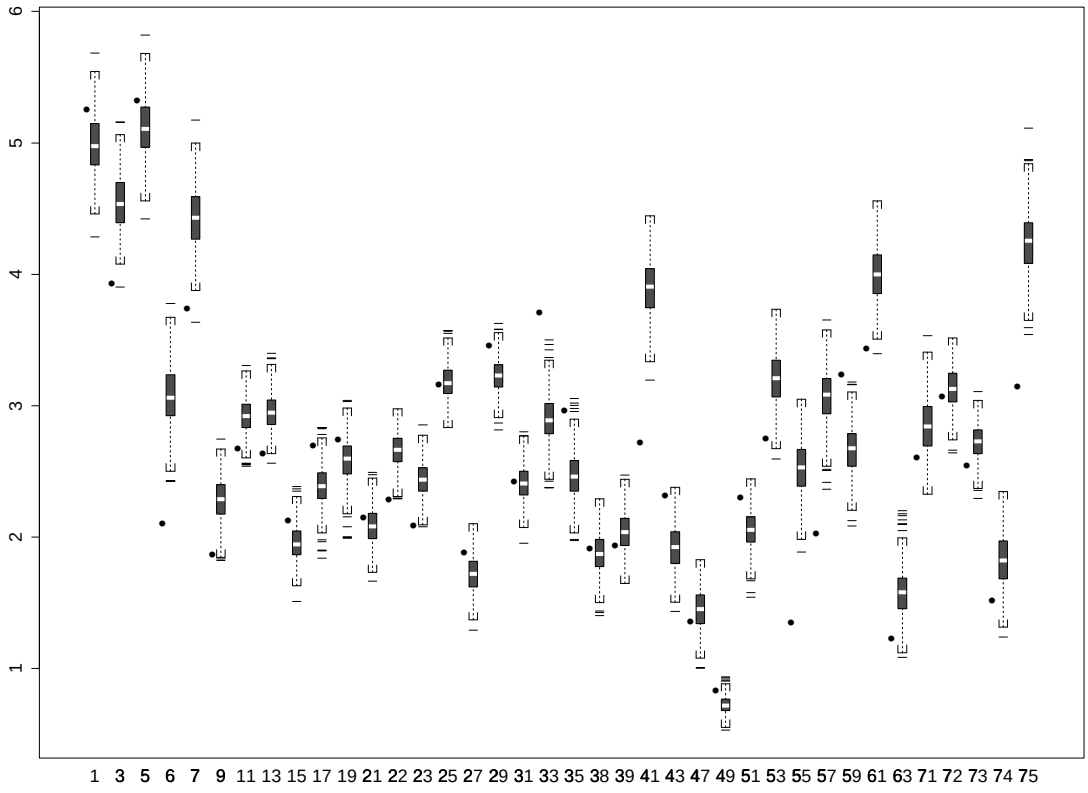

Figure 9: Posterior Dispersions between site 45 and other sites. The dots represent the empirical values.