Estimating a non-stationary spatial structure using simulated annealing

Serge IOVLEFF*and Olivier PERRIN†

May 16, 1999

Abstract

During the past decade, a useful model for non-stationary random fields has been developed. This consists of reducing the random field of interest to isotropy *via* a bijective bi-continuous deformation of the index space. Then the problem consists of estimating this space deformation. We propose to estimate this space deformation using a constrained continuous version of the simulated annealing for a Metropolis dynamic. This method provides a non-parametric estimation of the deformation which has the required property to be bijective; so far, the previous non-parametric methods do not guarantee this property. We illustrate our work with two examples, one concerning a precipitation data set. We also give one idea of how spatial prediction should proceed in the new coordinate space.

Key Words: bijective space deformation; constrained minimisation; correlation function; Delaunay triangulation.

1 Introduction

Assumption of isotropy is clearly violated for many, if not most, spatial environmental phenomena. Factors such as topography, local pollutant emissions, and meteorological influences may cause such assumptions to be violated. This has led to research into modelling a spatially non-stationary second order structure, as reviewed in Guttorp and Sampson (1994). To deal with this non-stationarity, Sampson and Guttorp (1992) have developed a model that consists of reducing the correlation function $r(\mathbf{x},\mathbf{x}')$ of the spatial phenomenon of interest, modelled by a random field $Z=\{Z(\mathbf{x}),\mathbf{x}\in G\subseteq\mathbb{R}^2\}$, to an isotropic one as follows:

$$r(\mathbf{x},\mathbf{x}')=\rho_\beta(\|\Phi(\mathbf{x}')-\Phi(\mathbf{x})\|), \tag{1}$$

where $\|.\|$ denotes the Euclidean norm in $\mathbb{R}^2$, Φ represents a bijective deformation of the geographic coordinate system and ρ_β is an isotropic correlation function with parameter $\beta\in\mathbb{R}^q$, $q\geq 1$. In the sequel, we refer to the geographical coordinate system as the G-space $\subseteq\mathbb{R}^2$, where G stands for geographical, and

*Laboratoires SABRES, IUP Vannes-Tohannic, rue Yves Mainguy, 56000 Vannes, France

†Mathematical Statistics, Chalmers University of Technology and Göteborg University, S-412 96 Göteborg, Sweden

to the deformed coordinate representation as the D-space $\subseteq \mathbb{R}^2$, where D stands for deformed. One illustration of model (1) is given in Figure 1 (this example is detailed in Paragraph 3.1): (i) represents the positions $\mathbf{x}_i$, $i = 1, 2, \ldots, n$, of n sites in the G-space; (ii) represents their deformations $\Phi(\mathbf{x}_i)$ in the D-space; (iii) represents the inter-site distances in the G-space *versus* correlations $r(\mathbf{x}_i, \mathbf{x}_j) = \rho_\beta(\|\Phi(\mathbf{x}_j) - \Phi(\mathbf{x}_i)\|)$, $1 \leq i, j \leq n$, where $\rho_\beta(u) = \exp(-\beta\|u\|)$ with $\beta = 1$; (iv) represents the inter-site distances $\|\Phi(\mathbf{x}_j) - \Phi(\mathbf{x}_i)\|$ in the D-space *versus* correlations $r(\mathbf{x}_i, \mathbf{x}_j)$.

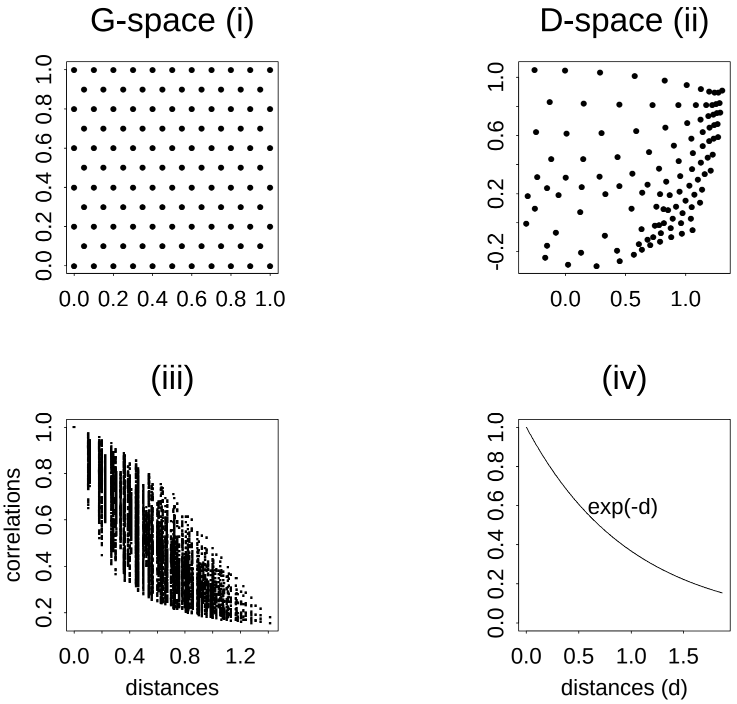

Figure 1: Example of a non-stationary random field using a space deformation.

Unlike the classical geostatistical models, non-stationarity through second order moments is thus taken into account and model (1) gives the opportunity to enlarge the class of models for studying spatial environmental random fields.

When ρ_β is strictly decreasing, Perrin and Meiring (1998) prove the uniqueness of both the deformation Φ and ρ_β up to a homothetic Euclidean motion for Φ and up to a scaling for ρ_β. Perrin and Senoussi (1998) give the general form of the deformation that reduces a non-stationary random field in the way (1), under smoothness assumptions.

Our concern in this paper is the estimation of both the space deformation Φ and the parameter β. This estimation is based on T repetitions (independent and identically distributed observations) of Z at each of n distinct monitoring sites $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$ in G, which may be irregularly located; G represents the convex hull of these sites. We denote these records using $Z_t(\mathbf{x}_i)$, $t = 1, 2, \ldots, T$, $i = 1, 2, \ldots, n$.

So far, the deformation Φ has mostly been estimated with the help of a non-parametric approach. This non-parametric estimation of Φ has already been extensively developed (Monestiez *et al.* (1993), Meiring (1995), Meiring *et al.* (1997, 1998)), and applied, for instance, in analyses of acid precipitation (Guttorp *et al.* (1992)), solar radiation (Sampson and Guttorp (1992)) and tropospheric ozone (Sampson *et al.* (1994)). The latter development of the non-parametric approach for estimating Φ consists of modelling Φ using a pair of thin-plate splines and minimising a penalised weighted least squares criterion to estimate the D-space coordinates $\Phi(\mathbf{x}_i)$ of the monitoring sites $\mathbf{x}_i$, $i = 1, 2, \ldots, n$. This treatment is carried out with a Marquardt-type algorithm. Several drawbacks are associated with this method: *(i)* bijection condition is not ensured; *(ii)* the fitting of the model becomes a challenging numerical problem with dimensionality roughly proportional to the number of fixed monitoring sites (Meiring *et al.* (1998)); *(iii)* the objective function to be minimised is non-convex and has a lot of local minima; *(iv)* there is a considerable dependence on the starting values in the minimisation procedure.

Perrin and Monestiez (1998) propose a parametric approach. They model Φ using a composition of a “small” number of bijective elementary radial basis deformations. Then they use a least squares criterion to estimate the parameters; this criterion is minimised in a classical way (Marquardt-type algorithm). Let us specify two disadvantages of this method: *(i)* so far a rational choice of the relevant number of elementary deformations has not been implemented yet; *(ii)* there is a considerable dependence on the starting values in the minimisation procedure.

To avoid all these disadvantages, we propose here to estimate, in a first step, the D-space coordinates $\Phi(\mathbf{x}_i)$, $i = 1, 2, \ldots, n$, by minimising an objective function with a stochastic algorithm, the continuous version of the simulated annealing for a Metropolis dynamic. To ensure a bijective correspondence between the n points in the G-space and their estimated deformations in the D-space, we impose some non-folding constraints in the algorithm. In a second step, to estimate the deformation in the whole G-space, we use a piecewise affine interpolation of the points $\mathbf{x}_i$ in the G-space and the estimations of their deformations $\Phi(\mathbf{x}_i)$ in the D-space, $i = 1, 2, \ldots, n$.

The paper is structured as follows. In Section 2, we address the estimation problem of the deformation together with the parameter β. In Section 3, we illustrate our non-parametric approach with two examples: the first one is purely illustrative and is only concerned with the estimation of the deformation when the isotropic correlation is known; the second one shows how our method can apply to precipitation data from 20 sites in the Languedoc-Roussillon region of France. In Section 4 we give one idea of how spatial prediction should proceed in the new coordinate space and propose a cross validation study to demonstrate the possible improvement in predictions due to the deformation. Further, we apply this cross validation study to our precipitation data set through a comparison of four isotropic correlation models. Finally, Section 5 outlines one extension related to this work.

2 Estimation of the model

2.1 Definition of the objective function

The repetitions of Z allow us to define the sample correlation estimates $\hat{r}(\mathbf{x}_i, \mathbf{x}_j)$ for each couple of sites $(\mathbf{x}_i, \mathbf{x}_j)$, $1 \leq i < j \leq n$:

$$\hat{r}(\mathbf{x}_i, \mathbf{x}_j) = \sum_{t=1}^{T} (Z_t(\mathbf{x}_i) - \bar{Z}(\mathbf{x}_i))(Z_t(\mathbf{x}_j) - \bar{Z}(\mathbf{x}_j))/\hat{\sigma}(\mathbf{x}_i)\hat{\sigma}(\mathbf{x}_j), \tag{2}$$

where $\bar{Z}(\mathbf{x}_i)$ are the empirical means and $\hat{\sigma}(\mathbf{x}_i)$ are the empirical standard deviations, $i = 1, \ldots, n$. We denote using $\mathbf{y}_i = \Phi(\mathbf{x}_i)$, $i = 1, 2, \ldots, n$, the deformations of the sites in the D-space and we define $\hat{\mathbf{y}}_1, \hat{\mathbf{y}}_2, \ldots, \hat{\mathbf{y}}_n$ and $\hat{\beta}$ as the estimations of $\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_n$ and β which minimise the following objective function:

$$U(\mathbf{y}^\star, \beta^\star) = \sum_{i<j} [\hat{r}(\mathbf{x}_i, \mathbf{x}_j) - \rho_{\beta^\star}(\|\mathbf{y}_j^\star - \mathbf{y}_i^\star\|)]^2, \tag{3}$$

with respect to the parameters $\mathbf{y}^\star = (\mathbf{y}_1^\star, \mathbf{y}_2^\star, \ldots, \mathbf{y}_n^\star)$ and $\beta^\star$, and subject to some non-folding constraints described in Paragraph 2.2.

2.2 Minimising the objective function

The objective function (3) is minimised with respect to $\mathbf{y}^\star$ using a continuous state version of the simulated annealing subject to some non-folding constraints. Simulated annealing is a probabilistic method for finding the global minimum of an objective function that may possess several local minima (Geman and Geman (1984), Kirkpatrick *et al.* (1983), Hajek (1988), Aarts and Korst (1989), Geman (1990) and Azencott (1992) for seminal references). This is motivated by its following advantages:

- it explores the whole objective function's surface and tries to optimise the function while moving both uphill and downhill. Thus, it is largely independent of the starting values, often a critical input in conventional algorithms;
- it can escape from local minima and go on to find the global minimum by the uphill and downhill moves;
- it makes less stringent regularity assumptions regarding the function than do conventional algorithms (it need not even be continuous);
- it is well suited for minimising strongly non-convex functions of several variables ($2n$ variables in our problem) having plenty of local minima;
- it can take intricate constraints into account.

2.2.1 Description of the minimisation algorithm

We describe hereafter the different steps to minimise (3) both with respect to $\mathbf{y}^\star$ and β:

- starting step: set $\mathbf{y}(0) = \{\mathbf{y}_i(0) = \mathbf{x}_i, i = 1, 2, \ldots, n\}$. Then to give an initial value to $\beta^\star$, minimise the objective function (3) with respect to $\beta^\star$ ($\mathbf{y}^\star = \mathbf{y}(0)$ is hold fixed) with a Marquardt-type algorithm. Let $\beta(0)$ be the estimation of β at step 0. Finally, take a sequence of "temperatures" $(c_0, c_1, \ldots, c_k, \ldots)$ decreasing to 0 by step of length n:

$$c_k = \theta^{\lfloor k/n \rfloor} c_0, \quad \theta \in]0, 1[, \ k \in \mathbb{N};$$

- step 0: start from the configuration $\mathbf{y}(0)$ of the sites. The change proposition is: fix $\beta^\star = \beta(0)$ and choose one site (candidate point) j uniformly among the n sites and move it locally and uniformly at a position $\mathbf{y}$ with natural non-folding constraints we precise hereafter. The other sites are hold fixed *i.e.* $\mathbf{y}_i(1) = \mathbf{y}_i(0), \forall i \neq j$. Set $\mathbf{y}(1) = \{\mathbf{y}_i(1), i = 1, 2, \ldots, n\}$ where $\mathbf{y}_j(1)$ is chosen as follows:
  - if $\Delta_0 U = U(\mathbf{y}_1(1), \ldots, \mathbf{y}, \ldots, \mathbf{y}_n(1), \beta^\star) - U(\mathbf{y}(0), \beta^\star) \leq 0$ then take $\mathbf{y}_j(1) = \mathbf{y}$;
  - otherwise sample an uniform law V in $[0, 1]$:
    - if $V \leq \exp(-\Delta_0 U / c_0)$ take $\mathbf{y}_j(1) = \mathbf{y}$;
    - otherwise keep $\mathbf{y}_j(1) = \mathbf{y}_j(0)$.
- step $k > 0$:
  - if $\lfloor k/n \rfloor = \lfloor (k-1)/n \rfloor$ then proceed as in the step 0 by replacing 0 by k and 1 by $k+1$;
  - otherwise minimise (3) with respect to $\beta^\star$ ($\mathbf{y}^\star = \mathbf{y}(k)$ is hold fixed) with a Marquardt-type algorithm. Let $\beta(k)$ be the estimation of β at step k. Then proceed as in the step 0 by replacing 0 by k and 1 by $k+1$.
- stopping criterion: if $U(\mathbf{y}(pn), \beta(pn)) - U(\mathbf{y}((p+1)n), \beta((p+1)n)) < 10^{-8}$ for two consecutive values of the integer p we stop the algorithm.

This algorithm is written in C language.

2.2.2 Description of the non-folding constraints

These constraints account for global and local non-foldings.

First, we embed G in a rectangle $\mathcal{R}$. Second, we construct the Delaunay triangulation for the n geographical sites plus the 4 vertices of $\mathcal{R}$ as well as the 4 mid-points of its 4 edges. To compute the Delaunay triangulation we use a program written by Shewchuk (1996).

At step k, for each site i we identify all the triangles for which this site is a vertex. Then consider the polygon P_i composed with the aggregation of these triangles and denote using $A_{i,1}, A_{i,2}, \ldots, A_{i,q_i}$ its vertices, where q_i is the number of triangles. We assume that these vertices are ordered clockwise and denote using $(A^1_{i,l}, A^2_{i,l})$ their coordinates , $l = 1, 2, \ldots, q_i$.

To ensure non-folding constraints in our algorithm when the candidate point j is moved at a position $\mathbf{y}$, in other words to avoid the case where triangles overlap, we impose that this new position is chosen uniformly in a convex set $\mathcal{C}^k_j$ which is the set of points $M_j = (M^1_j, M^2_j)$ such that, for $l = 1, 2, \ldots, q_j$:

$$M^1_j(A^2_{j,l} - A^2_{j,l+1}) + M^2_j(A^1_{j,l+1} - A^1_{j,l}) + A^1_{j,l}A^2_{j,l+1} - A^1_{j,l+1}A^2_{j,l} < 0, \qquad (4)$$

with the convention $q_j + 1 = 1$. In other words, this set is the kernel of P_j that is the set of all points $M_j \in P_j$ such that the line segment $[M_j, \mathbf{x}] \subset P_j$ for each $\mathbf{x} \in P_j$. The set $\mathcal{C}^k_j$ corresponds to the marked area of Figure 2. Note that the computation of this region is fairly easy since it consists of a set of linear inequalities like (4).

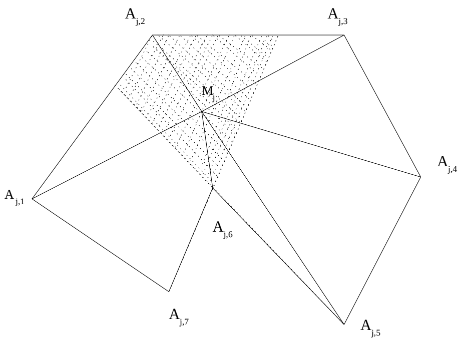

Figure 2: The marked area corresponds to the acceptable move for M_j.

These constraints mean that we impose moves that preserve the topological structure of the Delaunay triangulation the same.

Note that the rectangle $\mathcal{R}$ and its 8 points are hold fixed in the minimisation procedure. As illustrated by the left-hand plot of Figure 4, the embedding of G in $\mathcal{R}$ makes it possible to write the previous constraints (4) in a similar way for all the n sites: indeed, any one of the n site is included in a polygon for which the vertices are also vertices of the Delaunay triangulation for the n sites plus the 8 points of $\mathcal{R}$ represented by $\blacksquare$. Without this embedding we should distinguish between the boundary points of G and its interior points in the determination of the acceptable move in the algorithm. Therefore, with this embedding it is no longer necessary to distinguish between the boundary points of G and its interior points. Moreover, the virtual links between the sites in G and the 8 points of $\mathcal{R}$ prevent our algorithm from global folding of the triangulation in itself.

2.2.3 Estimating the deformation on the whole G-space

Any point $\mathbf{x}$ in G belongs to a unique triangle for which the vertices are three monitoring sites, say $\mathbf{x}_i$, $\mathbf{x}_j$ and $\mathbf{x}_k$. Thus any point $\mathbf{x}$ is uniquely defined by its barycentric coordinates in the triangle $\{\mathbf{x}_i, \mathbf{x}_j, \mathbf{x}_k\}$ as follows:

$$\mathbf{x} = a_{\mathbf{x},i}\mathbf{x}_i + a_{\mathbf{x},j}\mathbf{x}_j + a_{\mathbf{x},k}\mathbf{x}_k,$$

where $(a_{\mathbf{x},i}, a_{\mathbf{x},j}, a_{\mathbf{x},k}) \in [0,1]^3$ such that $a_{\mathbf{x},i} + a_{\mathbf{x},j} + a_{\mathbf{x},k} = 1$. We define the estimation $\hat{\Phi}(\mathbf{x})$ at point $\mathbf{x}$ of the deformation Φ as follows:

$$\hat{\Phi}(\mathbf{x}) = a_{\mathbf{x},i}\hat{\mathbf{y}}_i + a_{\mathbf{x},j}\hat{\mathbf{y}}_j + a_{\mathbf{x},k}\hat{\mathbf{y}}_k,$$

where $\hat{\mathbf{y}}_i$, $\hat{\mathbf{y}}_j$ and $\hat{\mathbf{y}}_k$ are estimated with the simulated annealing algorithm. So defined, the estimation $\hat{\Phi}$ holds the following features: continuous; bijective; it is an interpolant, *i.e.* $\hat{\mathbf{y}}_i = \hat{\Phi}(\mathbf{x}_i)$, $i = 1, 2, \ldots, n$.

3 Applications

3.1 Illustrative example

In this example we suppose that the correlation is of the form:

$$r(\mathbf{x}, \mathbf{x}') = \exp(-\|\Phi(\mathbf{x}') - \Phi(\mathbf{x})\|),$$

where Φ is a composition of three bijective functions of the type:

$$\begin{array}{rcl} \mathbf{f}_\alpha : \mathbb{R}^2 & \longrightarrow & \mathbb{R}^2 \\ \mathbf{x} & \longmapsto & \mathbf{b} + (\mathbf{x} - \mathbf{b})(1 + a_1 \exp(-a_2\|\mathbf{x} - \mathbf{b}\|^2)), \end{array}$$

with $\alpha = (a_1, a_2, \mathbf{b})$ a four-dimensional parameter, where $\mathbf{b} \in \mathbb{R}^2$ is the centre of the deformation, $a_1 > 0$ is a range parameter and $a_2 \in]-1, \frac{1}{2}\exp(\frac{3}{2})[$ denotes the intensity of stretching ($a_2 > 0$) or shrinking ($a_2 < 0$) ($a_2 = 0$ corresponds to the identity function). The form of the three functions $\mathbf{f}_\alpha$ we consider as well as the values of the parameters for these functions are given in Figure 3.

We assume that the correlations between the $n = 116$ sites located in the upper-left plot of Figure 1 are known. Our goal is the estimation of the deformation Φ. First we embed these sites in a rectangle and we build the Delaunay triangulation which is represented by the left-hand plot of Figure 4. Then we estimate the positions of the n sites in the D-space with the method described in the previous section (but without the estimation of $\beta = 1$ we suppose known in this example). In the cooling schedule we take $c_0 = 100$ and $\theta = 0.999$, and it takes 730 seconds CPU (on a Sun Ultra 10 model Creator, with a 300 MHz UltraSPARC-II i processor) to get the estimation of the 2×116 coordinates. The right-hand plot of Figure 4 represents both the true deformation and its estimation. The quality of our estimation method can be observed by comparing these two plots. Moreover, in the estimated D-space the isotropic structure is fairly well achieved, as it is shown by Figure 5.

3.2 Application to a precipitation data set

The data consist of monitored precipitation data from $n = 20$ sites in Languedoc-Roussillon in the south of France, for which the geographical configuration is shown in the left-hand plot of Figure 6.

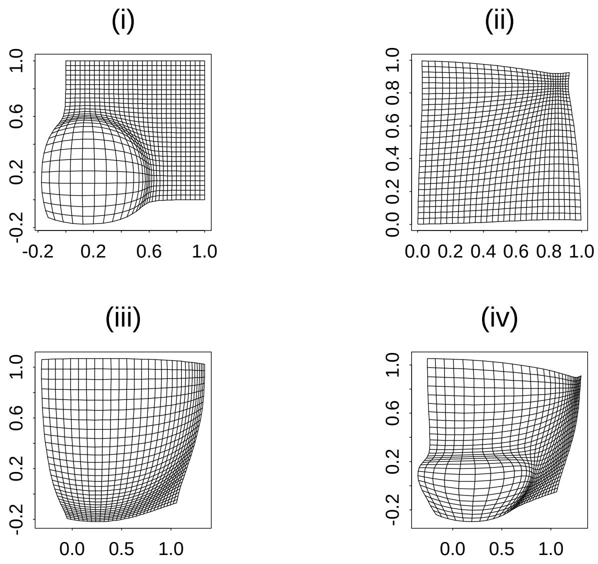

Figure 3: (i) $\mathbf{f}_{\alpha_1}$ with $\alpha_1 = (14, 1.6, 0.15, 0.15)$; (ii) $\mathbf{f}_{\alpha_2}$ with $\alpha_2 = (4, -0.6, 0.85, 0.85)$; (iii) $\mathbf{f}_{\alpha_3}$ with $\alpha_3 = (2, 1.4, 0.25, 0.95)$; (iv) their composition $\Phi = \mathbf{f}_{\alpha_3} \circ \mathbf{f}_{\alpha_2} \circ \mathbf{f}_{\alpha_1}$.

These data are in the form of 10-day aggregates, giving 6 records during November and December each year from 1975 through 1992. These data hold the following features: *(i)* a low frequency of missing values; *(ii)* a similar altitude (between 0 and 200 meters) for all the sites so that an altitude correction is pointless; *(iii)* a homogeneous period in the year so that a seasonal adjustment is not necessary. As pointed out by Meiring (1995), means and variances are positively related, and therefore, sample correlations between observations are calculated on the log scale, after adding 1 to all observations. The sample correlation estimates for a pair of sites is based on all the time points for which both sites have observations. The sample correlations for different pairs of sites are sometimes based on different numbers of observations, so that the sample correlations matrix is not positive definite. However, the model (1) fitted to the sample correlation remains positive definite.

The right-hand plot in Figure 7 represents the geographical inter-site distances *versus* the corresponding correlations $\hat{r}(\mathbf{x}_i, \mathbf{x}_j)$, $1 \leq i < j \leq n$, defined by (2), to which one would fit an isotropic correlation model if we assume isotropy for Z (note this would mean that Φ is the identity function in (1)). According to this plot, we decide to use the exponential model:

$$\rho_\beta(u) = \exp\left(-\beta u\right), \quad \beta > 0,$$

so that the objective function (3) is re-written as follows:

$$U(\mathbf{y}_1^\star, \mathbf{y}_2^\star, \dots, \mathbf{y}_n^\star, \beta^\star) = \sum_{i<j} [\hat{r}(\mathbf{x}_i, \mathbf{x}_j) - \exp\left(-\beta^\star \|\mathbf{y}_j^\star - \mathbf{y}_i^\star\|\right)]^2. \tag{5}$$

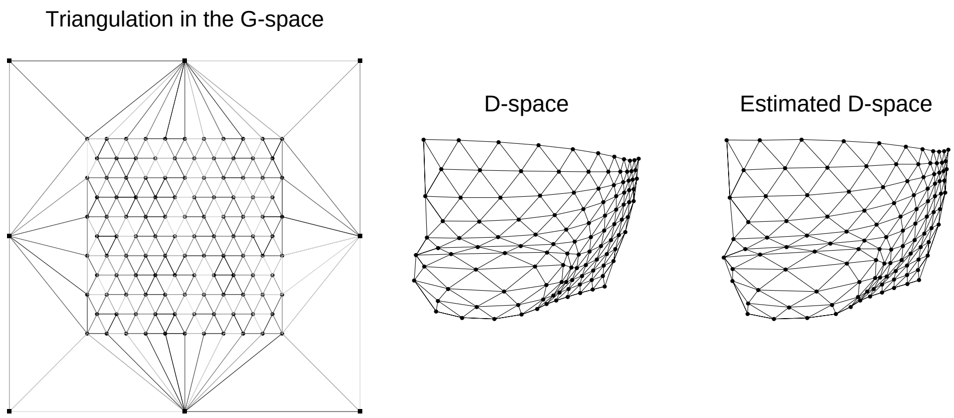

Figure 4: Delaunay triangulation for the $n = 116$ sites plus the 4 vertices of the embedding rectangle as well as the 4 mid-points of its 4 edges. True deformation (D-space) of the Delaunay triangulation restricted to the $n = 116$ sites and its estimation.

To minimise (5) we apply our constrained procedure described in Paragraph 2.2. The Delaunay triangulation for the 20 sites is shown in Figure 6. In the cooling schedule we take $c_0 = 1000$ and $\theta = 0.99$. The right-hand plot of Figure 6 represents the estimated deformation of the Delaunay triangulation. The upper right-hand plot of Figure 7 shows the fitted exponential correlation as a function of the D-space coordinates. The G-space has been stretched in regions of relatively lower correlations, and shrunk in regions of relatively higher correlations, so that the exponential correlation better models the correlations in the D-space representation (minimum of the objective function is equal to 0.023) than in the G-space system (minimum of the objective function is equal to 0.155). Improvement of this fitting is illustrated by the interquartile intervals: the empirical correlations are less scattered in the D-space than in the G-space.

4 Application to the prediction

We point out one possible application where the space deformation model (1) may be useful. This is the prediction using Kriging. We refer to Cressie (1993) for a presentation of the Kriging method.

In practice, the mean and the variance fields are very seldom known, and must be predicted. As Høst *et al.* (1995) illustrate, separate modelling of the mean, variance and residual fields from monitoring data collected in space and time, may give very valuable information about the standard errors in spatial interpolation. Prediction, of each of the mean, variance and residual fields contributes to the overall spatial interpolation errors. However, estimation of the mean and of the variance is not the topic of this work and we concentrate only on the prediction of the centred and standardised random field Z.

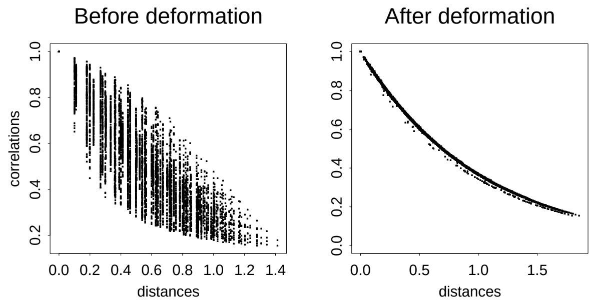

Figure 5: On the left: inter-site distances $\|\mathbf{x}_j - \mathbf{x}_i\|$ in the G-space *versus* correlations $r(\mathbf{x}_i, \mathbf{x}_j)$. On the right: inter-site distances $\|\Phi(\mathbf{x}_j) - \Phi(\mathbf{x}_i)\|$ in the estimated D-space *versus* correlations $r(\mathbf{x}_i, \mathbf{x}_j)$.

To predict the centred and standardised random process Z at any location $\mathbf{x} \in G$ we use the simple Kriging predictor:

$$\hat{Z}(\mathbf{x}) = \sum_{l=1}^{n} \lambda_l Z(\mathbf{x}_l),$$

with $(\lambda_l) = (\hat{r}_{i,j})^{-1} \left(\rho_{\hat{\beta}}(\|\hat{\Phi}(\mathbf{x}_l) - \hat{\Phi}(\mathbf{x})\|)\right)$ where (λ_l) denote the vector of the Kriging coefficients, $(\hat{r}_{i,j})$ the empirical correlation matrix of the vector $(Z(\mathbf{x}_1), Z(\mathbf{x}_2), \ldots, Z(\mathbf{x}_n))$ and $\left(\rho_{\hat{\beta}}(\|\hat{\Phi}(\mathbf{x}_l) - \hat{\Phi}(\mathbf{x})\|)\right)$ the correlation vector between the "known sites" $(Z(\mathbf{x}_1), Z(\mathbf{x}_2), \ldots, Z(\mathbf{x}_n))$ and the "unknown site" $Z(\mathbf{x})$.

One way to check whether the deformation may improve the prediction is a cross validation study. More precisely, for each of the T repetitions, we set aside a site and we predict the value of Z at this site with the $n-1$ remaining sites using the simple Kriging predictor described above. We repeat this operation for each of the n sites T times. Then we calculate the mean square error prediction ($MSEP$) that has to be compared with the one we would obtain by using the fitted isotropic correlation function if we assume isotropy for Z. We apply this cross validation study to the precipitation data set for each of the four following isotropic correlation models:

$$\begin{aligned}
\rho_E(u) &= \exp(-\beta_1 u), \text{ exponential model } (E);\\
\rho_{EN}(u) &= \beta_2 \exp(-\beta_1 u), \text{ exponential model with nugget } (EN);\\
\rho_G(u) &= \exp(-\beta_1 u^2), \text{ Gaussian model } (G);\\
\rho_{GN}(u) &= \beta_2 \exp(-\beta_1 u^2), \text{ Gaussian model with nugget } (GN);
\end{aligned}$$

where $\beta_1 > 0$ and $\beta_2 \in]0, 1[$.

The results of the cross validation study are given in the following Table which gives the $MSEP$ before and after the deformation for each of the four previous models:

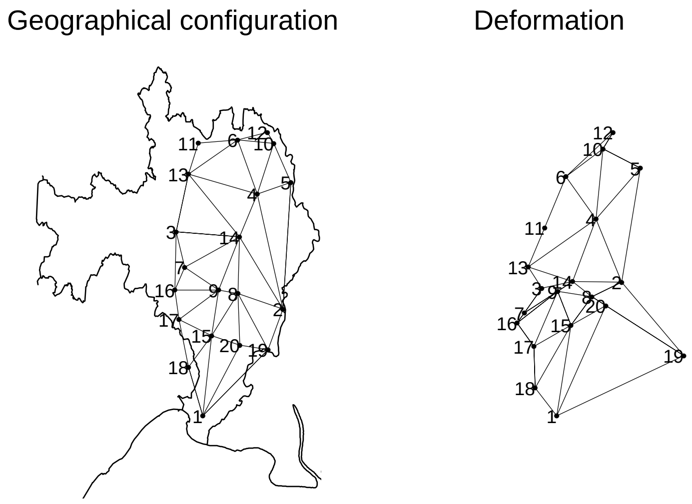

Figure 6: On the left: site locations and the corresponding Delaunay triangulation without the rectangle (the outlines indicate the French department of Gard and the coast). On the right: deformation of the triangulation.

	before deformation	after deformation	% of improvement
E	0.224	0.129	42.4
EN	0.203	0.136	33.1
G	0.271	0.132	51.3
GN	0.180	0.115	36.1

We deduce that the best improvement in terms of prediction is obtained for the Gaussian model: on average the prediction at the monitoring sites is 51.3 % better in the D-space than in the G-space. We also deduce that the best model is the Gaussian model both with and without deformation: nevertheless, with this model the prediction at the monitoring sites is 36.1 % better in the D-space than in the G-space. In conclusion, we can claim that the model with deformation for the correlation of the random field is better than the model without deformation.

5 Discussion

Simulated annealing appears to be a suitable tool for estimating a non-stationary structure. Combined with non-folding constraints, it gives the opportunity to estimate non-parametrically the bijective spatial deformation Φ, in model (1), in such a way that the estimation is bijective; so far the previous method using a pair of thin-plate splines does not guarantee this feature. However we wish to note one research question: so far, the choice of the isotropic correlation model

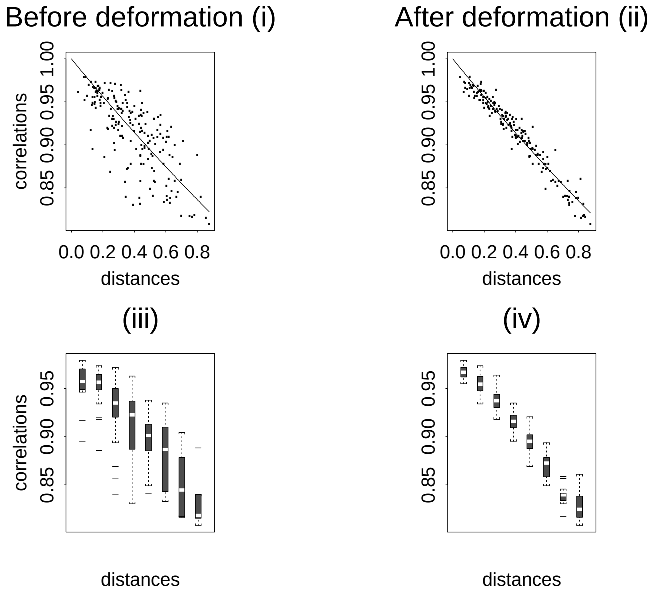

Figure 7: (i) Inter-site distances $\|\mathbf{x}_j - \mathbf{x}_i\|$ in the G-space *versus* empirical correlations $\hat{r}(\mathbf{x}_i, \mathbf{x}_j)$ and the fitted exponential correlation model (solid line) (ii) Inter-site distances $\|\hat{\Phi}(\mathbf{x}_j) - \hat{\Phi}(\mathbf{x}_i)\|$ in the estimated D-space *versus* empirical correlations $\hat{r}(\mathbf{x}_i, \mathbf{x}_j)$ and the fitted exponential correlation model (solid line) (iii) Box-plots of the empirical correlations as a function of the distances in the G-space (iv) Box-plots of the empirical correlations as a function of the distances in the D-space.

has been made by visual inspection. One future direction would be to establish a rational choice, by using for instance an Akaïke-type criterion. Furthermore, an attempt to use a non-parametric family of correlation models has yet to be developed.

Acknowledgements

Olivier Perrin has been supported by INRA, France; and by the European Union's research network "Statistical and Computational Methods for the Analysis of Spatial Data", Grant #ERB-FMRX-CT960095, while he was visiting the department of Mathematical Statistics at Chalmers University of Technology and Göteborg University in Sweden. The authors are grateful to Dominique Courault from Unité de Bioclimatologie, INRA-Avignon, France, for providing the precipitation data set, and to Jonathan Richard Shewchuk (1996) for providing the program that performs the Delaunay triangulation. We are also grateful to Xavier Guyon for giving us the idea of this work.

References

Aarts, E. and Korst, T.J. (1989), *Simulated annealing and Boltzman machines: stochastic approaches to combinatorial optimization and neural computing*, Wiley.

Azencott, R. (ed) (1992), *Simulated annealing: parallelization techniques*, Wiley.

Cressie, N.A.C. (1993), *Statistics for spatial data*, Revised Edition, Wiley-Interscience.

Geman, D. (1990), *Random fields and inverse problem in imaging*, L.N.M. no. 1427, Springer.

Geman, S. and Geman, D. (1984), "Stochastic relaxation, Gibbs distribution, and the Bayesian restoration of images," *IEEE Trans. Pattern Anal. Machine Intell.*, 6, 721-741.

Guttorp, P. and Sampson, P.D. (1994), "Methods for estimating heterogeneous spatial covariance functions with environmental applications," in: Patil, G.P. and Rao, C.R. (eds), *Handbook of Statistics XII: Environmental Statistics*, Elsevier/North Holland, New York, 663-690.

Guttorp, P., Sampson, P.D. and Newman, K. (1992), "Nonparametric estimation of non-stationary spatial covariance structure with application to monitoring network design," in: Walden, A. and Guttorp, P. (eds), *Statistics in Environmental and Earth Sciences*, Edward Arnold, London, 39-51.

Hajek, B. (1988), "Cooling schedules for optimal annealing," *Mathematics of operations research*, 13 (2), 311-329.

Høst, G., Omre, H. and Switzer, P. (1995), "Spatial interpolation errors for monitoring data," *Journal of the American Statistical Association*, 90, 853-861.

Kirkpatrick, S., Gelatt, Jr., C.D., and Vecchi, M.P. (1983), "Optimization by simulated annealing," *Science*, 220, 671-679.

Meiring, W. (1995), "Estimation of heterogeneous space-time covariance," *PhD thesis*, University of Washington, Seattle.

Meiring, W., Guttorp, P. and Sampson, P.D. (1998), "Computational issues in fitting spatial deformation models for heterogeneous spatial correlation," *Computing Science and Statistics*, 29 (1), (Scott, D.W. (ed)), 409-417.

Meiring, W., Monestiez, P., Sampson, P.D. and Guttorp, P. (1997), "Developments in the modelling of nonstationary spatial covariance structure from space-time monitoring data," in: Baafi, E.Y. and Schofield, N. (eds), *Geostatistics Wollongong '96*, 1, Kluwer Academic Publishers, 162-173.

Monestiez, P., Sampson, P.D. and Guttorp, P. (1993), "Modelling of heterogeneous spatial correlation structure by spatial deformation," *Cahiers de Geostatistique*, Fascicule 3, Compte Rendu des Journées de Géostatistique, 25-26 May 1993, Fontainebleau. Published by the École Nationale Supérieure des Mines de Paris.

Perrin, O. and Meiring, W. (1998), "Identifiability for non-stationary spatial structure," to appear in *Journal of Applied Probability*, 36 (4).

Perrin, O. and Monestiez, P. (1998), "Modelling of non-stationary spatial structure using parametric radial basis deformations," to appear in: Soares, A., Gómez-Hernandez, J. and Froidevaux, R. (eds), *geoENV98*, Kluwer Academic Publishers.

Perrin, O. and Senoussi, R. (1998), "Reducing non-stationary random fields to stationarity or isotropy using a space deformation," submitted for publication.
Sampson, P. and Guttorp, P.D. (1992), "Nonparametric estimation of nonstationary spatial covariance structure," *Journal of the Americal Statistical Association*, 87, 108-119.
Sampson, P.D., Guttorp, P. and Meiring, W. (1994), "Spatio-temporal analysis of regional ozone data for operational evaluation of an air quality model," *Proceedings of the Section on Statistics and the Environment*, American Statistical Association.
Shewchuk, J. R. (1996), Triangle: engineering a 2*D* quality mesh generator and Delaunay triangulator, *First Workshop on Applied Computational Geometry* (Philadelphia, Pennsylvania), 124-133, ACM. Available online from http://www.cs.cmu.edu/ quake/triangle.research.html.