Estimation of precipitation by kriging in the EOF space of the Sea Level Pressure field

Grard BIAU [a,1], Eduardo ZORITA [b], Hans von STORCH [b] and Hans WACKERNAGEL [a]

[a] Centre de Géostatistique, Ecole des Mines de Paris, France;
[b] Institut für Gewässerphysik, GKSS Forschungszentrum, Germany.

Abstract

The term downscaling denotes a procedure in which local climatic information is derived from large-scale climate parameters. In this paper, we explore the possibility to use as downscaling procedure a geostatistical interpolation technique known as kriging. We present an example of the method by trying to reconstruct monthly winter precipitation in the Iberian Peninsula from the North Atlantic Sea Level Pressure (SLP) field in wintertime (December, January and February).

The main idea consists in reducing the spatial dimension of the large-scale SLP field by means of Empirical Orthogonal Function Analysis (EOF Analysis). Each observed SLP field is represented by a point in this low-dimensional space and this point is associated with the simultaneously observed rainfall. New values of the SLP field, for instance simulated by a General Circulation Model (GCM) with modified greenhouse gas concentrations, can be represented by a new point in the EOF space. The rainfall amount to be associated to this point is estimated by kriging interpolation in the EOF space.

The results obtained by this geostatistical approach are compared to the ones obtained by a simpler analog method, by trying to reconstruct the observed rainfall from the SLP field in an independent period. It has been found that generally kriging and the analog method reproduce realistically the long-term mean, that kriging is somewhat better than the analog method in reproducing the rainfall evolution, but that, contrary to the analog method, it underestimates the variance because of the well known smoothing effect. It is argued that there exists an intrinsic incompatibility between the estimation of the mean and replication of the variability.

Finally both methods have been also applied to daily winter rainfall. The methods are also validated by downscaling winter precipitation from SLP. It is concluded that kriging yields a better estimation of daily rainfall than the analog method, but the latter reproduces better the probability distribution of rainfall amounts and of the length of dry periods.

[1] All correspondence to be addressed to this author. Email: biau@math.univ-montp2.fr.

1 Introduction

One major problem concerning climate simulations with General Circulation Models (GCMs) is related to the scale at which GCM simulations can be considered skillful: if it is today accepted that present-day GCMs are able to simulate the large-scale atmospheric states in a generally realistic manner (large-scale refers to processes with a characteristic scale of several grid lengths) and that they are thus a useful tool to predict large-scale climate changes, their implications on regional climate are much more questionable. Thus, for instance, one of the largest uncertainties in climate simulations produced by the present generation of models concerns the small-scale hydrological processes, such as precipitation, cloud formation, infiltration or evaporation. All these processes occur over a much smaller scale than the resolution of today's GCMs, which is limited by computational considerations to typical grid sizes of 200 to 500 km.

The origins of this lack of skill have already been discussed in the literature (Grotch and MacCracken [9], von Storch [25]). Several general methods, called *downscaling methods*, have been suggested so far to overcome this scale mismatch. One category makes use of nested high resolution dynamical models (dynamical downscaling, Giorgi and Mearns [8]) whereas others use empirical-statistical methods (statistical downscaling, Karl *et al* [15]) to link large-scale and small-scale variables. The statistical models are based on observational records. Once the statistical model parameters are estimated from a training set of large-scale and local observations, they may be used to infer changes in the local variables due to changes in the large-scale fields simulated by GCM sensitivity experiments.

Pursuing the latter approach, many recent applications of statistical downscaling techniques can be found in the literature (for a review see *e.g.* Hewitson [10] or Wilby and Wigley [30]). Most of them make use of linear regression methods between the large-scale variables and the local variables. Relatively few studies are based on nonlinear methods, such as neural networks (Hewitson [10]), CART (Classification and Regression Trees, Hughes, Lettenmaier and Guttorp [12]), the analog method (Zorita *et al* [31]), or classification methods (Bardossy and Plate [3], Hulme *et al* [13]). Nonlinear methods are specially necessary when the relationship between the climate large-scale and local parameters follow different probability laws. In this case a linear statistical model cannot describe appropriately the physical link between both sets of variables. This is for instance the case for large-scale atmospheric circulation and daily rainfall. On the other side, the application of nonlinear methods is often more difficult from the theoretical and numerical point of view.

In this paper, we explore the possibility to use for downscaling purposes a well known geostatistical interpolation technique, which behaves non-linearly with respect to the spatial coordinates, known as *kriging* (for a complete presentation of geostatistics and its methods, see for example Chauvet [7] and Wackernagel [28]). Kriging has been originally applied in the mining industry to interpolate spatially data obtained by geological soundings. In climate research it has been used, for instance, to interpolate observed climate fields (*e.g.* Holdaway [11]). Downscaling is a novel application of kriging, but there exists a fairly large amount of theoretical studies and practical applications that may be of considerable support in this respect (Rouhani and Wackernagel [23], Amani and Lebel [1]).

In the normal setting, kriging is used in geostatistics as a two or three-dimensional interpolation technique based on the estimated spatial covariance structure of the observed data. The idea of using kriging to build a statistical model between large-scale climate variables and local variables is based on the fact that in many climate studies the dimensionality of the large-scale climate field is reduced by means of *Empirical Orthogonal Function Analysis* (EOF Analysis, also known as *Principal Component Analysis*, see *e.g.* Mardia, Kent and Bibby [19] or Saporta [24]) to a limited number of characteristic patterns of variability. These EOFs constitute the basis in which predictors or predictands are represented in many linear statistical models. Each configuration of the large-scale field corresponds to a point in the space of reduced dimensionality spanned by its own EOFs. The local variable simultaneously observed with the large-scale field is then associated to this point in the EOF space. Thus, the observations of the local variable form a discrete sampling of a function defined in the EOF space of the large-scale field. New configurations of the large-scale field, for instance observed in another period or simulated by a GCM, may be represented by a point in the EOF space and the associated value of the local variable at this point may be estimated by interpolation or extrapolation in this space using kriging.

It should be underlined that this is a quite unusual application of kriging. This technique is normally used in a two or three-dimensional geographical space, whereas in this case the kriging interpolation is performed in the EOF space of the SLP field, which may in principle have much more dimensions. In the application to monthly rainfall it has been found that the space spanned by the leading two SLP EOFs is enough to obtain reasonable results, but in the case of daily rainfall (see Section 6) kriging is performed in a 15-dimensional space.

We present an example of the method by trying to reconstruct the winter precipitation in stations located in the Western Mediterranean region from the North Atlantic Sea Level Pressure (SLP) field in wintertime (December, January and February). The results obtained by kriging are compared to the ones obtained by a much simpler analog method, which due to its simplicity may be considered in this context as a benchmark. We consider two time scales: the monthly time scale, where the distribution of rainfall amount is more or less normal, and therefore linear methods have been previously applied with success (von Storch, Zorita and Cubasch [27], Busuioc and von Storch [6]), and daily rainfall, which has a much more skewed distribution. Also the requirements for downscaling rainfall at these two time scales are somewhat different. At monthly time scales the most important point is a good estimation of the mean, whereas for daily rainfall other aspects, such as the probability of rainfall amount, extremes, length of dry periods, etc., may be sometimes quite important.

The paper is organized as follows: after a short presentation of data in Section 2, the kriging interpolation technique is briefly described and the kriging setting in the EOF space, a somewhat unusual application, is explained in Section 3. Section 4 focuses on the autocorrelation structure of local precipitation as described in the space of the SLP EOFs by the so-called *variograms*. Section 5 presents then the detailed results of kriging and analog downscaling at monthly time scales. Finally, in Section 6, kriging and analog downscaling are applied to daily data. Some important aspects of the results obtained are discussed in Section 7. The paper is closed by some concluding remarks in Section 8 and by one appendix.

2 Data description

Winter (December, January and February, DJF) precipitation monthly totals records at 50 Spanish, Portuguese, Southern France and North African stations over a period of 91 years (1899-1989) were prepared by the Universidad Complutense (Madrid). Since the relation between SLP and rainfall in Western Europe is known to be stronger in winter, only winter data (DJF) were considered. The positions of stations are shown in Figure 1.

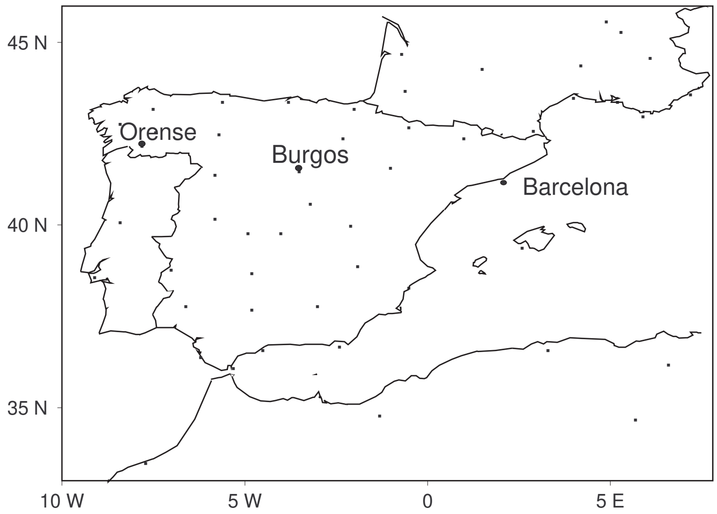

Figure 1: Positions of the 50 stations under study. The names correspond to the stations explicitly cited in the text.

The Sea Level Pressure (SLP) data set was obtained from the National Meteorological Center (NMC, Boulder). It consists of monthly mean (DJF) SLP over the same period (1899-1989). The original SLP data were interpolated onto a 5° latitude × 5° longitude grid and we selected a box which covered the North Atlantic sector within 70° W to 20° E and 15° N to 85° N. A wide westward extension was chosen since it is known that the

Atlantic weather systems influence Iberian Peninsula rainfall. The rainfall and SLP data were linearly detrended.

3 Kriging setting

3.1 Short description of kriging

A geostatistical study usually deals with a sample of observations (environmental, chemical, meteorological, etc.) $z(\mathbf{x}_\alpha)$ at different locations $\mathbf{x}_\alpha$. The observation $z(\mathbf{x}_\alpha)$ (in the kriging terminology known as the regionalized variable) is considered as a realization of a random function $Z(\mathbf{x})$.

An important tool in geostatistics is the *experimental variogram*, which is a measure of the variability of the regionalized variable at different spacings. The experimental variogram $\gamma^*(\mathbf{h})$ as a function of the point separation $\mathbf{h}$ is defined as

$$\gamma^*(\mathbf{h}) = \frac{1}{2N(\mathbf{h})} \sum_{\alpha=1}^{N(\mathbf{h})} \Big[z(\mathbf{x}_\alpha + \mathbf{h}) - z(\mathbf{x}_\alpha) \Big]^2,$$

where $\mathbf{h}$ is a vector in geographical space and $N(\mathbf{h})$ is the number of sample points linked by $\mathbf{h}$ - possibly with a certain amount of tolerance on the length and the orientation of the vector.

Normally the experimental variogram is based on too few sample points and it needs to be approximated by a theoretical function $\gamma(\mathbf{h})$, that allows to estimate the variogram analytically for any distance $\mathbf{h}$: this function is called the *variogram model.* In practice, the modeling step is mainly interactive. Using high quality graphical devices here is particularly advantageous for representing and interrogating the data with a set of (geo)statistical tools - histograms, scatter diagrams, variogram clouds... - which help interpret the structure of the stochastic process being studied.

Once the variogram models have been fitted to the experimental variograms, it is possible to perform the estimation by kriging. Kriging is a method to estimate the value of the regionalized variable under study at a location $\mathbf{x}$ where it has not been measured, if its values are known at N locations $x_1, \ldots, x_N$. Normally, a linear estimator is used:

$$Z^*(\mathbf{x}) = \sum_{\alpha=1}^{N} \lambda_\alpha Z(\mathbf{x}_\alpha),$$

where the unknowns of the problem are the weights λ_α. The value of the weights are determined by minimizing the variance of the estimation error $Z^*(\mathbf{x}) - Z(\mathbf{x})$. It can be shown that the minimization of this variance under certain assumptions about the random function Z (see appendix for a slightly more detailed mathematical description) leads to the system of linear equations:

$$\begin{cases} -\sum_{\beta=1}^{N} \lambda_\beta \gamma(\mathbf{x}_\alpha - \mathbf{x}_\beta) + \mu & = & -\gamma(\mathbf{x} - \mathbf{x}_\alpha) \\ \\ \sum_{\beta=1}^{N} \lambda_\beta & = & 1. \end{cases}$$

The observation points that enter these equations may be all the points available or just a neighborhood of the estimation point, depending on the particular application. The unknowns of the system are the weights λ_α and μ is a Lagrangian multiplier. Once the values of the unknown weights have been found it is possible to get the estimated value of the regionalized variable at the point $\mathbf{x}$ by the expression

$$z^*(\mathbf{x}) = \sum_{\alpha=1}^{N} \lambda_\alpha z(\mathbf{x}_\alpha).$$

3.2 Kriging setting in the EOF space of the SLP field

As stated in the introduction the points where kriging is performed are defined in the space of the SLP EOFs. It is possible to obtain a simplified spatial representation of the SLP data set by performing an EOF analysis. The principle of this well known data analysis method (see *e.g.* Mardia, Kent and Bibby [19] or Saporta [24]) is to obtain an optimal approximate representation of the original anomalies of the multivariate field $\mathbf{f}(t)$ by projecting it onto a subspace of lower dimension:

$$\mathbf{f}(t) = \sum_{i=1}^{n} \mathbf{p}_i \alpha_i(t) + \epsilon(t),$$

where n is typically a small number, $\mathbf{p}_i$ are the spatial patterns, $\alpha_i(t)$ are the coordinates of $\mathbf{f}$(t) in the basis spanned by $\mathbf{p}_i$ and $\epsilon(t)$ is the part of $\mathbf{f}$ not described by the patterns $\mathbf{p}_i$.

The EOFs $\mathbf{p}_i$ are defined as the patterns that minimize the variance of ϵ and are also orthogonal to one another. For further information about the use of EOFs in climatology, see von Storch and Navarra [26]. The EOFs $\mathbf{p}_i$ are defined up to a proportionality constant. To define unambiguously the EOFs we additionally require that the coordinates $\alpha_i(t)$ have unit time variance. In this way the EOF patterns contain the information about the typical magnitude of the SLP anomalies.

Figure 2 shows the two leading EOFs of the North Atlantic SLP in winter. The first pattern shows a dipole structure with low pressure centered over Iceland versus high pressure over Azores, and it is associated with the intensity of the geostrophic zonal wind (east-west wind) over the North Atlantic area. This pattern is also closely associated with the North Atlantic Oscillation (Lamb and Peppler [17]). The second SLP pattern shows anomalies of the same sign over the whole North Atlantic, and it has been sometimes denoted as the Western Atlantic pattern.

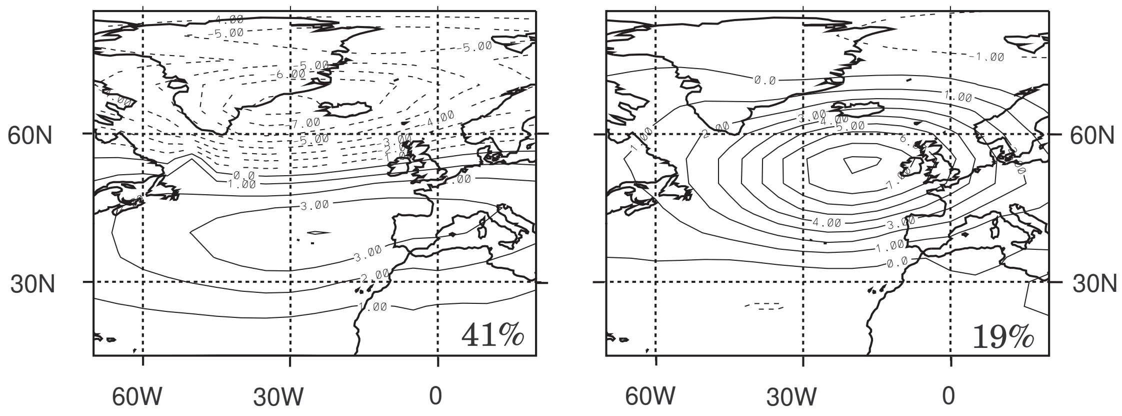

Figure 2: The two leading EOFs of the SLP field in winter, units mb. The percentage of explained variance is given in the lower corner of the diagrams.

The EOF analysis allows the representation of the much of the variability of SLP anomalies by just a few coordinates. In this case just the two leading EOFs, which together explain 60% of the SLP variability, have been retained.

The kriging interpolation is performed separately for each rainfall station in this low-dimensional EOF space as follows: for each rainfall station the value of the precipitation at a month m is attached to a point in a two dimensional space, the coordinates of which are the first two EOF coordinates of the SLP field at the month m. Thus the rainfall observations for each station can be represented by a data cloud in this two-dimensional space. Two examples of this data cloud, for two stations located in the east and in the west of the peninsula, are presented in Figure 3.

The points where the rainfall values have to be estimated may be individual points given by SLP anomalies observed in other period (*e.g.* for validation of this method) or SLP anomalies simulated by a GCM (for downscaling of simulated climate change). Both can be projected onto the leading two SLP EOFs yielding two coordinates in the EOF space. In this way, kriging enables the values of local precipitation associated with the SLP field to be estimated.

To check the validity of this precipitation estimator based on kriging, we tried to reconstruct the winter Iberian rainfall in the period 1969-1989. Thus, the data from 1899 to 1968

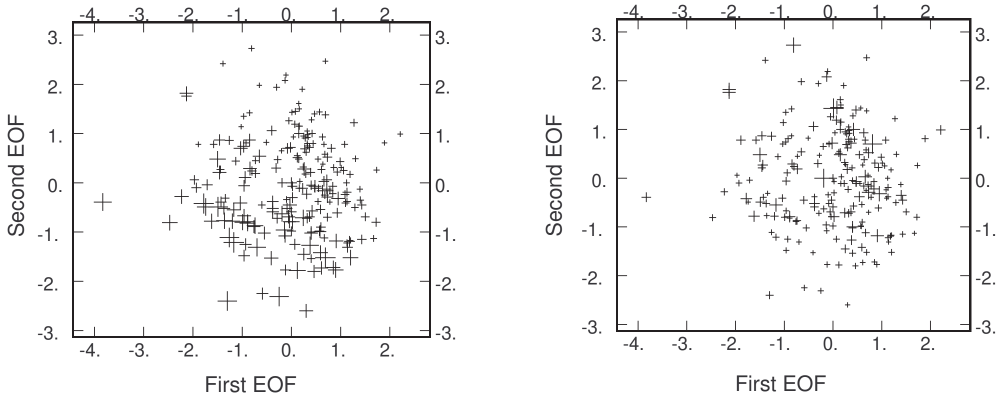

Figure 3: Monthly rainfall in winter represented in the space of the two leading SLP EOFs: Orense (left) and Barcelona (right) data. The size of each cross is proportional to the rainfall amount attached to the point. Note that the locations of the data are the same for the two stations.

were used as input data to the geostatistical analysis, *i.e.* for the determination of the variograms, and the statistical model was validated in the remaining period.

4 Variographical analyses

4.1 Results

A variographical study should be based on sound physical, chemical or environmental knowledge of the phenomenon: the variography's role in the classical situation consists merely in quantifying the structural information for further estimation procedures. In the present case, the main difficulty of the geostatistical study is precisely the indirect physical meaning of the EOF coordinate space. It is thus quite difficult to refer to external evidence in order to improve - or, at least, to conduct - a variographical analysis. Nevertheless, the experimental variogram certainly remains a convenient tool for exploring the structure of the data described in the previous section.

It was found in practice that the variability of the rainfall data in the EOF space was almost identical in all directions of the EOF space, which means that the phenomenon is isotropic and the variogram depends only on the modulus of $\mathbf{h}$, denoted by $|\mathbf{h}|$. Therefore, in order to simplify the study, we neglect some occasional weak geometric anisotropies, that could be reduced to isotropies by a mere linear transformation of the coordinates (see Journel and Huijbregts [14]).

Inspection of the experimental variogram for each station revealed that they can be divided into three distinct categories (see Figure 4 that presents an example of each category):

- 6 among the 50 variograms appear as *pure nugget effects* (flat variograms). A pure nugget effect corresponds to a total absence of correlation between the $Z(\mathbf{x})$ and $Z(\mathbf{x}+\mathbf{h})$, at least for all available distances. Mathematically, the corresponding variogram model $\gamma_w(\mathbf{h})$ is as follows:

$$\gamma_w(\mathbf{h}) = \begin{cases} 0 & \text{for } |\mathbf{h}| = 0 \\ C_0 & \text{for } |\mathbf{h}| \neq 0, \end{cases}$$

  where C_0 is a constant called the *sill*.

- 9 variograms can be fitted with a *bounded nested structure* (see appendix for an exact definition of bounded variogram models) composed of:

  1. a nugget effect, the amplitude C_0 of the discontinuity at the origin being contained between 40% and 80% of the data total variance (see Figure 4),
  2. and a *spherical model* described by

$$\gamma_s(\mathbf{h}) = \begin{cases} C\left[\frac{3}{2}\frac{|\mathbf{h}|}{a} - \frac{1}{2}\frac{|\mathbf{h}|^3}{a^3}\right] & \text{if } 0 \leq |\mathbf{h}| \leq a \\ C & \text{if } |\mathbf{h}| > a. \end{cases}$$

     Here C is called the *sill* and a is called the *range*. Due to the normalization of the EOF coordinates, $|\mathbf{h}|$ ranges typically between 0 and 9 and in our case $a \simeq 0.5$, making it sometimes difficult to choose between a pure nugget effect and a spherical model.

  The variogram models connected with this second category can then be written as follows:

$$\gamma(\mathbf{h}) = \gamma_w(\mathbf{h}) + \gamma_s(\mathbf{h}) = C_0 + \gamma_s(\mathbf{h}) \text{ for } |\mathbf{h}| \neq 0.$$

- Finally, the 35 remaining observed variograms are similar and can be modeled by an *unbounded model variogram* as follows:

  1. a nugget effect, with the amplitude contained between 25% and 70% of the data variance (see Figure 4),
  2. and an unbounded variogram, which is modeled by an infinite parabolic increase and corresponds to an increasingly smooth regionalized variable. It is then convenient to adopt a so-called *intrinsic-without-drift model* (see appendix for an exact definition of intrinsic-without-drift variogram models). We chose a *power model*:

$$\gamma_p(\mathbf{h}) = \alpha|\mathbf{h}|^\theta \text{ with } 0 < \theta < 2.$$

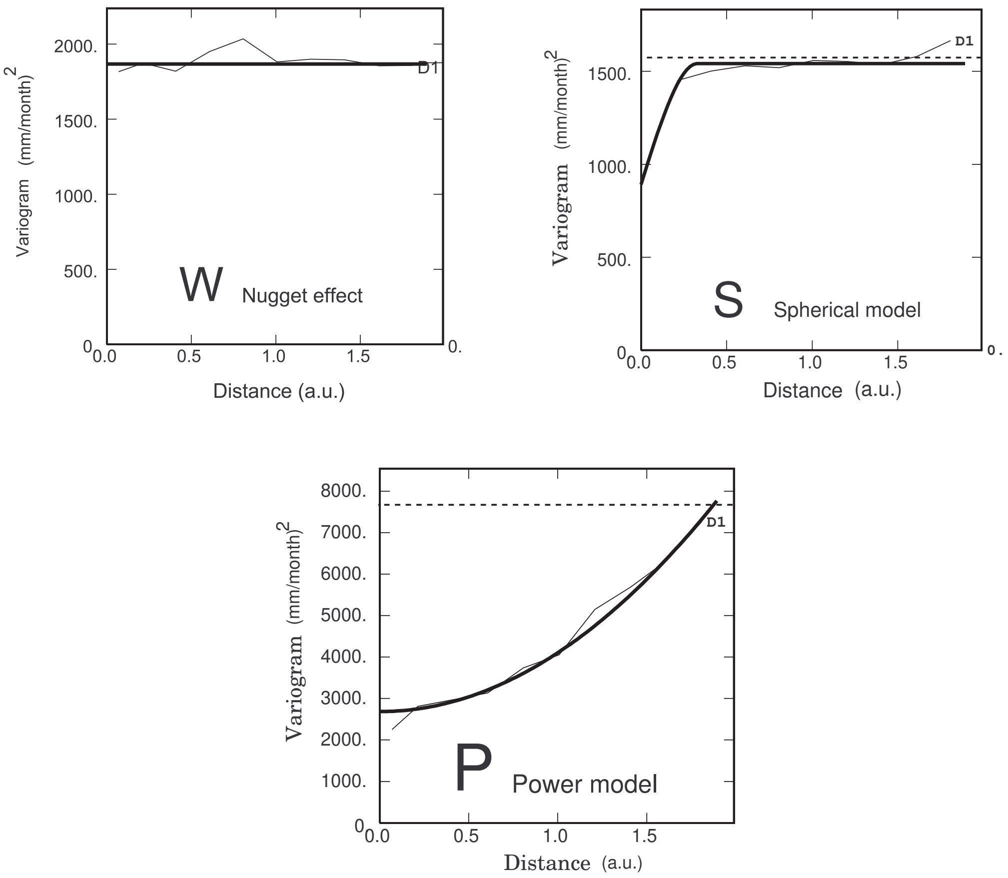

Figure 4: Three typical examples of experimental variograms (continuous lines) with their associated models (thick continuous lines). The notations are as follows: **W** for pure nugget effect ("**W**hite noise"), **S** for spherical model and **P** for power model. We also show the data variance (dotted lines).

Thus, the theoretical model adopted can be represented, for each of the 35 stations, by the equation

$$\gamma(\mathbf{h}) = \gamma_w(\mathbf{h}) + \gamma_p(\mathbf{h}) = C_0 + \gamma_p(\mathbf{h}) \text{ for } |\mathbf{h}| \neq 0.$$

Furthermore, it was observed that the stations with unbounded variogram models are located in the west and center of the Iberian Peninsula (9°W-0°W) as well as in a part of southern France. The non-stationarity increases toward the west (the power of the variogram models approaches the value 2). On the other hand, 90% of the stations modeled with bounded or pure nugget effect variogram models are located around the Mediterranean and in the north of Spain (Gulf of Biscay).

4.2 Interpretation

Through the fitting of several variogram model categories (unbounded models versus bounded models), we observe a discrimination between the stations located in the west and center of the Iberian Peninsula (power models) (and perhaps some stations located in the south of France, too) and those located around the Mediterranean and in the north of Spain (spherical and pure nugget effect models). This phenomenon can be explained in terms of EOF analysis performed on the SLP field. The first and second SLP EOFs discriminate between the winter months of high precipitation and those of weak precipitation for stations located in the west and in the center of the area under study but this effect is much less clear in the stations located in the Mediterranean basin (see Figure 3). It is observed that the negative values of the first and second principal components correspond to high rainfall, whereas the positive values correspond to weaker rainfall. This correlation gets stronger toward the west. The power variograms represent this link between SLP and rainfall. Physically, the rainfall gradient in EOF space can be interpreted with the help of Figure 2, where the physical patterns associated with the first two EOFs are shown. Negative amplitudes of the first of these patterns represent a weakened Azores high that allows the Atlantic weather systems to enter the Iberian Peninsula. Negative amplitudes of the second pattern are linked to geostrophic advection of moist air by low pressure cells located off the British coast. An analogous reasoning explains the link between positive values of the principal components and lower than normal rainfall.

This result is consistent with the conclusions of von Storch, Zorita and Cubasch [27], who noted that a large-scale SLP pattern, essentially the first EOF, is the most important atmospheric phenomenon in the Atlantic area associated with Iberian winter rainfall. The authors also noted that the link between the North Atlantic SLP and the precipitation was higher near the Atlantic coast, with correlations decreasing toward the east.

5 Validation of the method and comparison with the analog method

5.1 Results

As a logical continuation to the previous sections, the amount of rainfall at each of the 50 Iberian Peninsula stations for the winter months (DJF) of the years 1969 to 1989 was estimated by kriging. For the previous variographical analysis and for the calculation of the SLP EOFs only the years 1899 to 1968 were used. It is thus possible to compare the precipitation estimates obtained by kriging with the *in situ* observations. The kriging operation will be repeated at each of the 63 time steps (once again, winter months from 1969 to 1989: see Figure 5 for the coordinates of these estimation points in EOF space). A simpler method has been proposed that uses the precipitation information carried by only one point chosen in the EOF space. This alternative downscaling scheme is called the analog method (Zorita *et al* [31]). As applied by these authors the analog method operates

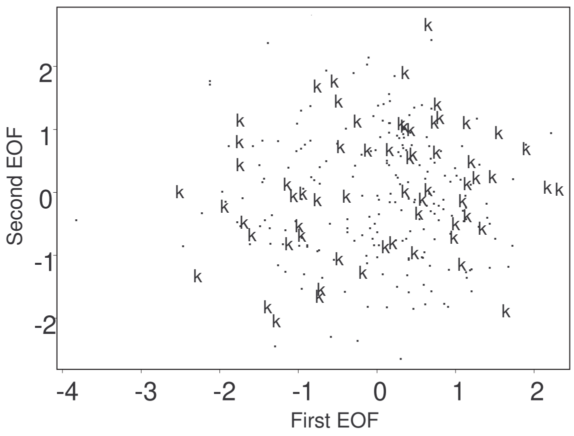

Figure 5: Position of the coordinates in the EOF space. Points indicate the data used for the variographical analysis (**1899-1969**), "k" denotes the points where the estimation was performed (**1969-1989**).

in the same EOF space as in the present paper, but simply chooses as an estimate at a given point the closest observation. If several estimation points have the same observation point as closest, they will of course get the same value[1].

The main difference between kriging and analog estimation is that an analog estimation makes use of information carried by only one point (the closest) whereas kriging uses a linear combination of the information at all data points. For these reasons, and in order to check to what extent the additional technical complications of kriging in an unusual application really bring an improvement over a simpler method, it was decided to compare the results obtained by kriging with the results obtained by the analog method.

For each station three measures have been used to compare the results of both methods:

- The long-term observational means of the reconstructed rainfall and of the *in situ* observations.
- The observed and reproduced standard deviations.
- The linear correlation coefficients between the observed and simulated rainfall time series.

It has to be noted that when the variogram model adopted for a station is a pure nugget effect, kriging yields the mean of the observation points as output, independently of the location of the estimation point. It follows, for such stations, that the reconstructed time series are constant, making meaningless the calculation of linear correlation coefficient. In

[1]The analog method is called "Polygon method" in mining (see *e.g.* Wackernagel [28]).

such cases, the presence of a pure nugget effect was depicted with "W" (white noise) on the figures showing the correlation coefficients.

Figures 6 and 7 depict respectively the linear correlation coefficients between reconstructions and observations, and the long-term means and the standard deviations estimated by kriging and analog method. We note that:

- The kriging linear correlation coefficients are good in the center of the Iberian Peninsula (5°W-0°), with a typical value of 0.7; they get better toward the west and reach values higher than 0.85 near the western coast (Portugal and Spain Atlantic coasts).
- All the stations located around the Mediterranean coast (North African, Spanish and Southern France coasts) exhibit low correlation between kriging estimation and observations (with sometimes negative values). The same is observed for the stations located in the north of Spain (Gulf of Biscay).
- The analog method linear correlation coefficients follow the same spatial distribution as those obtained with kriging but are nearly everywhere significantly lower (with typically a difference of 0.2).

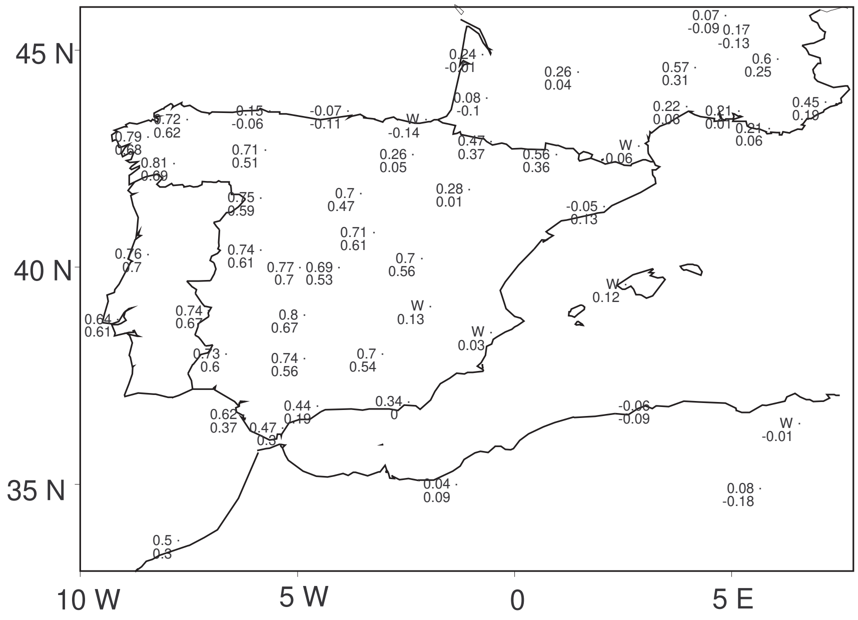

Figure 6: Linear correlation coefficients between kriging reconstructions and observations (top) and between analog method reconstructions and observations (bottom). Stations with a "W" have a white noise variogram model.

- The long-term means are very well reproduced by both downscaling methods.

- The standard deviations are acceptably reconstructed by the analog method but underestimated by kriging.

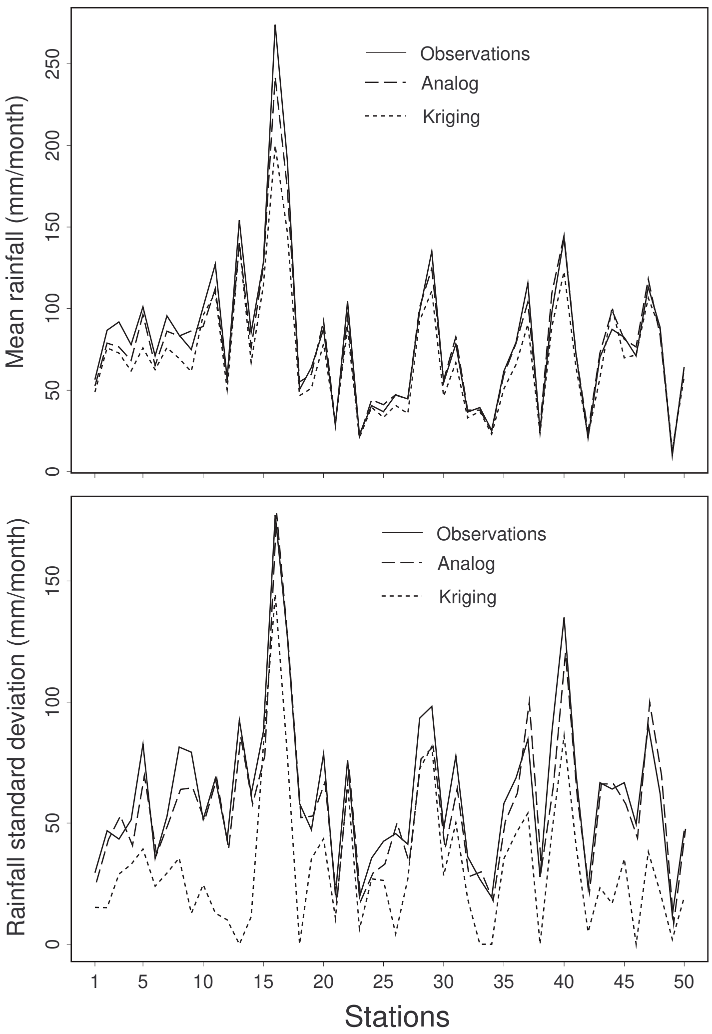

Figure 7: **(Top)** Means of observations (continuous line), kriging reconstructions (dotted line) and analog method reconstructions (dashed line). **(Bottom)** Standard deviations of observations (continuous line), kriging reconstructions (dotted line) and analog method reconstructions (dashed line).

As a complementary remark it should be mentioned that we conducted the same geostatistical study using, instead the pair (EOF1,EOF2) as basis vectors, the basis (EOF1,EOF3), (EOF2,EOF3) and (EOF1,EOF2,EOF3), respectively. The main conclusion is that the most important structural information is carried by the first two EOFs and that the spatial information carried by the third and following EOFs decreases with the rank of the

EOF.

Another approach of the problem by means of universal kriging (Matheron [20]) was also attempted. Concerning the linear correlation coefficients and the reproduction of the mean, the results we got were relatively close to those described in this section. On the other hand, we observed that the variance of the estimated values was of the same order as the variance of the observations. Unfortunately, the reproduction of the variance is definitely not an intrinsic property of universal kriging and is probably due to extrapolation effects at the border of the data region.

5.2 Illustrations

In this paragraph the functional form of the estimated rainfall as a function of the amplitudes of the two leading SLP EOFs is illustrated. In the previous paragraph the rainfall associated with isolated points in the EOF space was estimated by kriging and analog methods, but it is also possible to perform this estimation at each node of a regular grid in the whole low dimensional EOF space, thus yielding numerical representation of the estimated rainfall surface as a function of the two SLP EOF coordinates.

In this paragraph the results of the rainfall estimation in the EOF space obtained by kriging, analog method and, for illustration purposes, a linear regression method are presented. For these three methods the rainfall estimation was performed on a 0.1×0.1 regular grid (see Figure 8) in the two-dimensional EOF space. We selected the station Orense, which is located in the west of the Iberian Peninsula (Figure 1). As representative of the linear regression models we chose the Canonical Correlation Analysis (CCA), that is quite often applied in the context of statistical downscaling (von Storch, Zorita and Cubasch [27], Kushnir *et al* [16]). Given two multivariate random variables, CCA finds the linear combination of the components of each of the variables that maximizes the linear correlation (Bretherton, Smith and Wallace [4]). This property has been used to construct an optimized linear downscaling model between the North Atlantic SLP field and rainfall in the Iberian Peninsula (von Storch, Zorita and Cubasch [27]). For comparison with the kriging and analog methods the SLP field has been projected only onto the two leading EOFs and the CCA was performed in this two-dimensional space in the same fitting period as for the other two methods.

The results are shown in Figure 8. The kriging surface is more regular than the analog one. This is a good illustration of the underestimation of the variance by kriging (see discussion in Section 7). The analog surface presents a variation in steps. This is not surprising according to the definition of the analog method. This method chooses as estimation of a point the closest one, thus making it likely that several grid points are connected with the same estimator. The representation of precipitation estimated by the CCA method (this being linear) is obviously a plane in EOF space. The CCA finds the plane that would optimally fit the rainfall in all stations as a function of the two leading SLP EOFs, if the rainfall time series were standardized to unit variance.

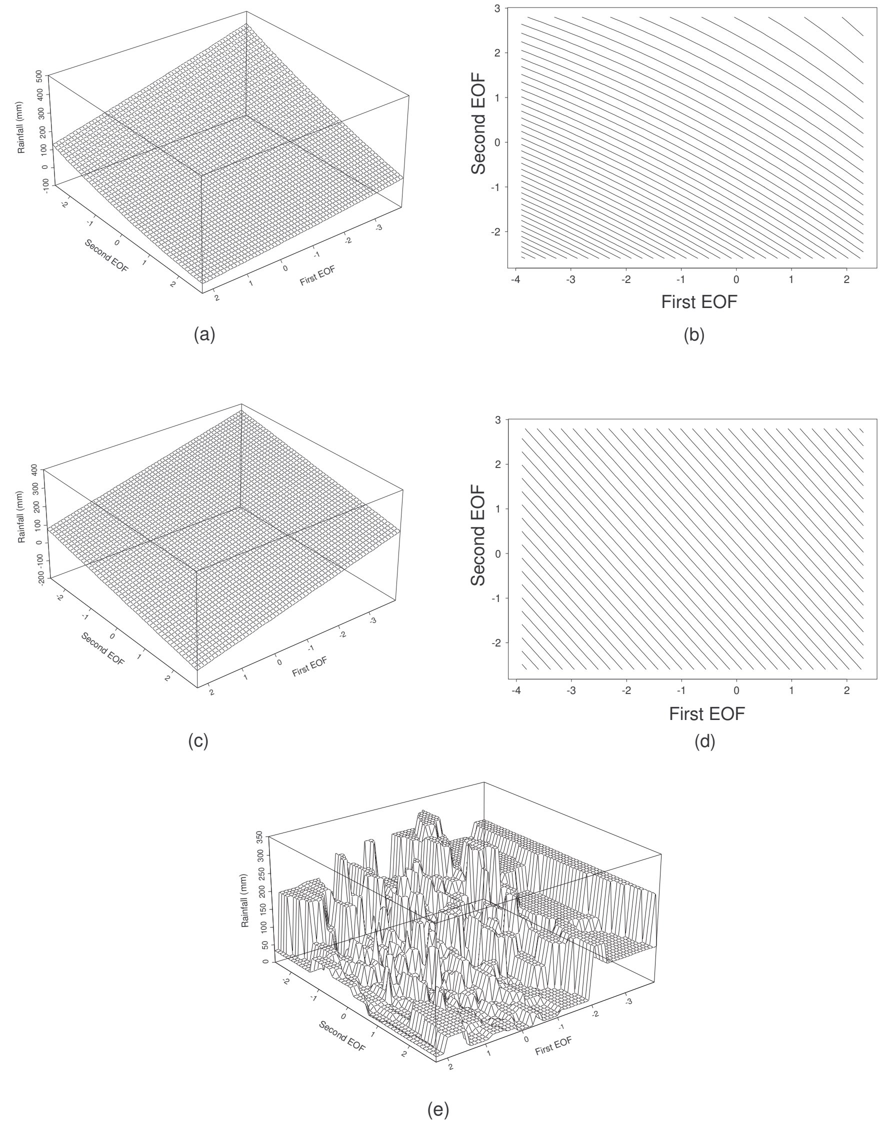

Figure 8: Three-dimensional and isolines plots of kriging (a) and (b), Canonical Correlation Analysis (c) and (d) and analog method (e) results for the station Orense. All interpolations were performed on a $\mathbf{0.1} \times \mathbf{0.1}$ regular grid. The variogram model is a power model (with nugget effect). Note that the axes of the three-dimensional views are inverted.

Figure 8 illustrates clearly the different behavior of the three estimation methods. The analog method is strongly nonlinear in the EOF space and its stepwise variations are somewhat unrealistic. This could be avoided by averaging over a neighborhood closest to the estimation point, and not taking just the closest point. But then the replication of the variance would have been impaired. CCA tries to find just the best regression plane, and the weakness of the method should be more evident at the edges of the estimation region if the relationship between rainfall and SLP is not linear. Kriging produces a nice smoothed approximation of the analog surface and the nonlinear dependency of rainfall on the SLP for large negative values of the EOF coordinates is quite clear. An interesting point would be to try to explain in physical terms this nonlinear dependency.

6 Daily data

6.1 Data description

In this section both the kriging and analog downscaling procedures are tested with winter (DJF) *daily* data. In the previous sections, we dealt with winter (DJF) *monthly* data, whose variability is much weaker than the variability of winter daily data. Also the probability distributions of monthly and daily rainfall are quite different.

The winter (DJF) daily precipitation data in the Iberian Peninsula were kindly supplied by the Instituto Nacional de Meteorología (Madrid) but were unfortunately available only for 10 stations. They covered the period from 1$^{\text{st}}$ of January 1969 to 31$^{\text{st}}$ of December 1989. The winter (DJF) SLP data are the same as those described in Section 2, except that we now consider the daily means. We selected the same time period as for the precipitation (winter days from 1$^{\text{st}}$ of January 1969 to 31$^{\text{st}}$ of December 1989).

For the specification of daily rainfall, it will be assumed that the occurrence of precipitation at a day d depends not only on the state of the atmosphere at the same day d but also on the previous states of the atmosphere, as for instance in Zorita *et al* [31]. These authors concluded that the persistence of the precipitation process may be better captured by weather generators that take into account the evolution of the daily SLP field and that using information from the SLP in the roughly five previous days can improve the results.

Therefore we tried to apply the kriging and analog downscaling as described in the previous sections, but with some modifications. Firstly, the five leading SLP EOFs were considered, since at daily time scales the variability of the SLP field is spread over a wider range of patterns. Secondly, the three-day evolutions were also considered, so that instead of connecting the precipitation observed at a day d with a point embedded in a two-dimensional EOF space, the daily rainfall amount was attached to a point in a 15-dimensional space whose coordinates are defined as follows: the first five coordinates are the first five EOF coordinates of the SLP anomalies at day d; the following five coordinates are the first five EOF coordinates of the SLP situation at day $d-1$ and the last five coordinates are the first five EOF coordinates of the SLP situation at day $d-2$.

In this way it was hoped that the relationship between daily SLP and precipitation could be better captured. But the price to be paid is double. Firstly, the spatial distribution of the data in the EOF space is much more difficult to visualize because of the very high dimension of the space: in other words, the variographical analyses are going to be totally blind. And secondly, this will require - at least for kriging - longer computation time.

6.2 Variographical analyses for daily rainfall

As we did for monthly data, we split the data sets in two parts, in order to perform a validation. The data from 1969 to 1979 were reserved for the variographical analyses, and we tried to reconstruct the rainfall time series from 1980 to 1989. The variographical analyses were performed in the 15-dimensional EOF-time space. We found that the 10 experimental variograms could be divided into two groups (Figure 9 presents an example of each category):

- 4 variograms are fitted with a power model and an additional nugget effect (see Section 4 for the corresponding definitions). The powers are close to the value 2, what probably indicates the presence of a very strong gradient in the (time-augmented) EOF space.

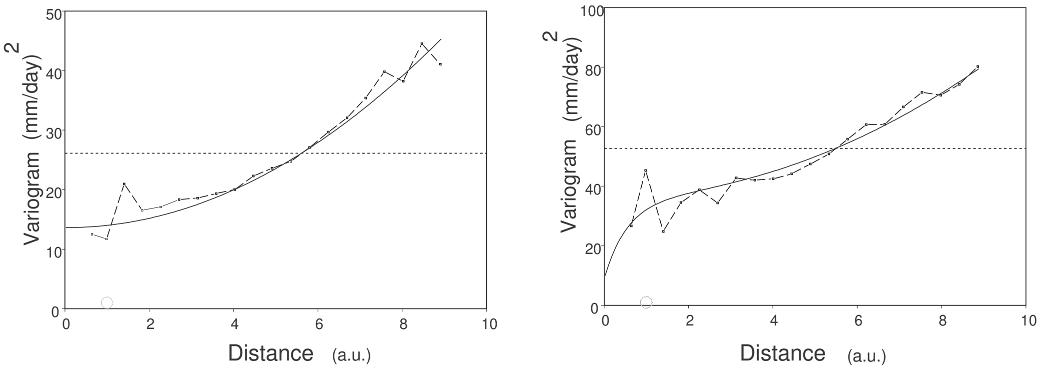

Figure 9: Two examples of experimental variograms (dashed lines) with their associated variogram models (continuous lines) [power (left) and exponential + power (right) model]. We also show the data variance (dotted lines).

- The 6 remaining experimental variograms present a sill at small scale followed by a parabolic increase. They were consequently fitted with a nested structure composed of a nugget effect, an exponential and a power model:

$$\gamma(\mathbf{h}) = C_0 + C\left[1 - \exp\left(-\frac{|\mathbf{h}|}{\beta}\right)\right] + \alpha|\mathbf{h}|^\theta \text{ with } 0 < \theta < 2 \text{ and } |\mathbf{h}| \neq 0.$$

6.3 Results and comparison with the analog method

We analyzed the results of kriging and analog method according to the three criteria (linear correlation coefficients, means and variances) previously used for the monthly data, and additionally the replication of the daily rainfall histograms and of the probability of the storm interarrival times was also investigated. The latter probability, p(τ), is defined as the probability that a certain period with no rain lasts at least τ days.

It was found that the long-term means for all 10 stations are well reproduced by both the kriging and analog methods (not shown). Furthermore, most of the linear correlation coefficients obtained by kriging downscaling are acceptable considering the noisy character of daily rainfall, especially in the western part of the Iberian Peninsula. The maximum value is reached is the center-north of Spain with a value equal to 0.34. The analog method correlation coefficients are poor everywhere and never are higher than 0.15. If the time series are smoothed with a three-day-running-mean filter the correlations obtained by kriging are improved, but those obtained by the analog method remain low (Figure 10).

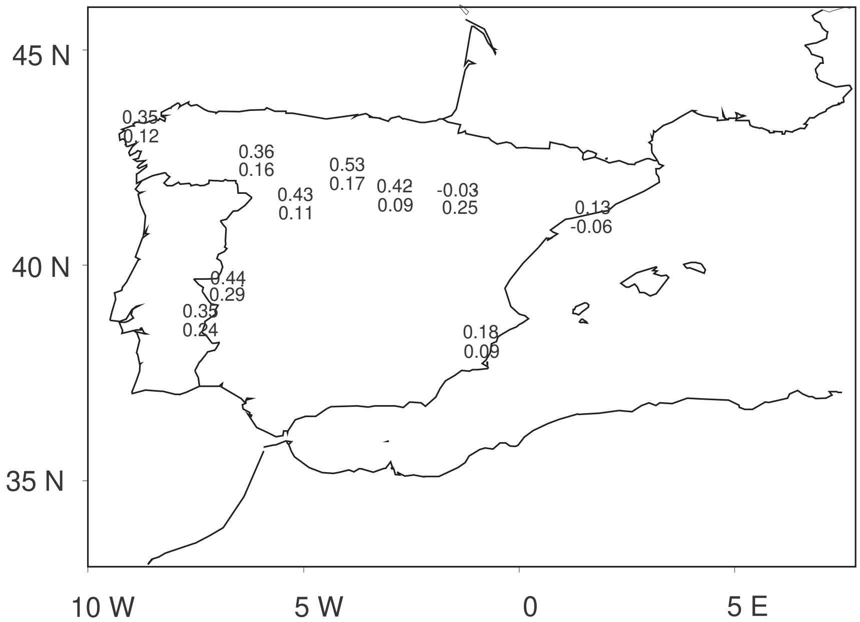

Figure 10: Linear correlation coefficients after the time series have been smoothed by a three-day-running mean between kriging reconstructions and observations (top) and between analog method reconstructions and observations (bottom).

However, the reconstructed variances are acceptable for the analog method but again kriging underestimates the rainfall variance (not shown). Also the probability distribution of daily rainfall is not satisfactorily replicated. Figure 11 shows the histogram of daily rainfall in Burgos, a station located in the center of the peninsula (see Figure 1). Kriging is not able to reproduce the probability distribution of daily rainfall, specially in the lower range (0-3 mm/day). This is due to the kriging smoothing effect (see Section 7) and is specially clear in a severe underestimation of the number of dry days. The analog method is able to replicate quite nicely the probability distribution in the independent period, suggesting that 10 years of winter daily values are able to sample sufficiently this high-dimensional EOF-time space (see Section 7). Kriging also clearly underestimates the frequency of dry periods for all time lengths, whereas the analog method is able to produce time series with the right distribution of the length of dry periods (Figure 12).

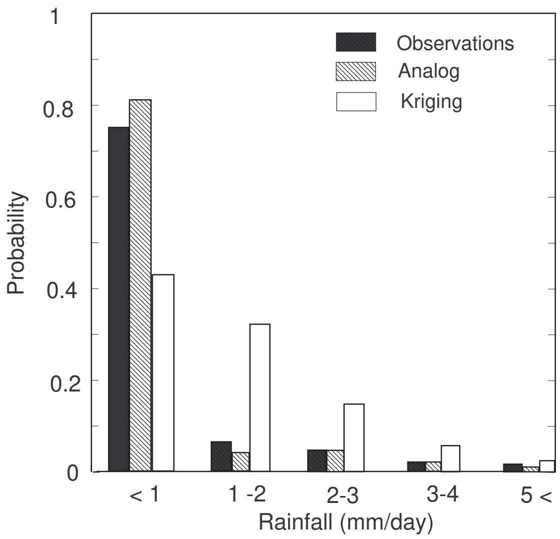

Figure 11: Histogram of daily rainfall at Burgos (center of the peninsula), as observed and simulated by the kriging and analog methods.

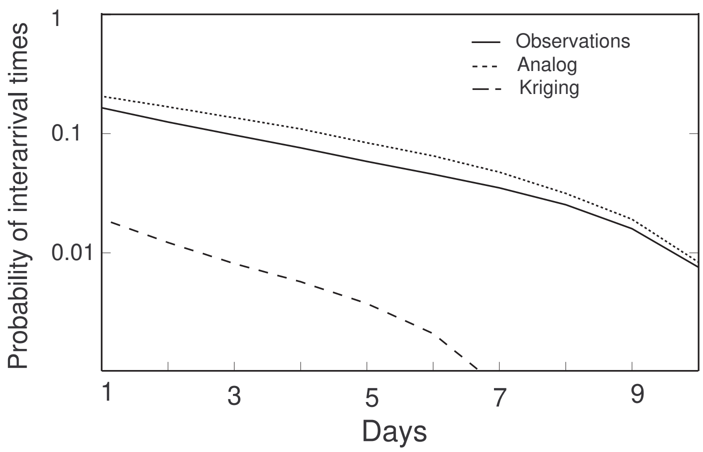

Figure 12: Probability of length of dry periods at Burgos (center of the peninsula), as observed and simulated by the kriging and analog methods.

It can be mentioned here that the histograms of the daily rainfall data are right-skew so that a nonlinear kriging method (Rivoirard [21] and [22]) could provide some improvement.

7 Discussion

7.1 Variographical analyses and kriging results

The linear correlation coefficients between observed and reconstructed rainfall obtained by both kriging and the analog method enable us to distinguish between three geographical areas:

- The west and center of the Iberian Peninsula and some part of the south of France, where the correlations between estimations and observations are good and decrease from west to east. It seems that specification of precipitation from SLP seems to be here most efficient.
- The Mediterranean coast: both methods are poor. This can be explained by the weakening influence of North Atlantic sea level pressure on precipitation toward the east.
- The north of Spain (Gulf of Biscay), which exhibits low linear correlation coefficients, probably due orographic effects ($\sim$ 1500 m) along the coast.

The results are consistent with the comments to the variographical analysis in Section 4. There seems to be a direct relation between the types of variogram models fitted in Section 4 and the values of the linear correlation coefficients. In summary, we have roughly the following scheme:

$$\begin{array}{ccc} \text{power model} & \Leftrightarrow & \text{good or very good estimation} \\ \text{stationary model} & \Leftrightarrow & \text{less good estimation} \\ \text{pure nugget effect} & \Leftrightarrow & \text{poor estimation.} \end{array}$$

Hence, the kriging results are consistent with the variographical analyses. We thus note that the efficiency of kriging downscaling is closely bound to the structure of the rainfall data in the EOF space: for stations with a strong rainfall gradient (Figure 3) kriging is much more successful than for stations with a flatter rainfall distribution. Furthermore the three geographical areas have already been put forward during the variographical analyses. It was found in Section 4 that most western stations and those at the center of the Iberian Peninsula (as well as some stations in the south of France) were fitted with power models, whereas the stations located around the Mediterranean and in the north of Spain (Gulf of Biscay) were fitted with bounded or pure nugget effect variograms.

7.2 Replication of the mean and variance

We noticed in Section 5 that the analog method is able to reconstruct in a very realistic manner both *in situ* observed mean and variance. Actually, this phenomenon is much more due to a favorable spatial distribution of the data under study than to an intrinsic property of the method. If the space is sufficiently covered by the observed data points, a random attribution of a rainfall amount to the estimation points will automatically reproduce the right mean and variance. However, if the estimation points lie outside the cloud of observed data points, the right estimation of the mean and the variance is not obvious any more.

The performance of the analog method is thus closely bound to the distribution in EOF space and the number of estimation and observation points. In our case, for example, there are less estimation points than observations, and all the data are relatively well spread over the region. This explains the success of the method. But we could however imagine some much more unfavorable situations could occur: either because the estimation points are clustered in a single region of space or because there are simply less observations than estimation points.

With respect to kriging, the situation is somewhat different, since kriging does not provide any partition of the space, but uses the information carried by all the observation points. Therefore, it can be expected *a priori* that kriging might be more robust with respect to the sampling density in EOF space, specially when the data show a strong gradient in EOF space. However, if the variogram can be approximated by a nugget effect, or close to it, and the EOF space is scarcely sampled the estimation will tend to become worse.

By construction, kriging is an unbiased estimator (see appendix), which means that the average error on point estimations is zero over a large area. This is coherent with our results, since we found that the kriging reproduced well the rainfall means. On the other hand, the reproduction of the variance is not part of the criterion searched for in kriging. It is indeed well known that kriging "smoothes" and that the histogram of the values estimated by kriging will display more values around the mean but less extreme values than the histogram of the observations, as observed in the histograms of the daily rainfall amounts (Figure 11). In general it is possible to show (Chauvet [7]) that in the random function model:

$$\mathrm{Var}\big[Z^*(\mathbf{x})\big] \leq \mathrm{Var}\big[Z(\mathbf{x})\big],$$

if the expectation of $Z(\mathbf{x})$ is supposed being equal to zero (simple kriging). This last relation is called the *smoothing relation* of kriging.

If the aim of the model is to preserve the variance of the phenomenon, the smoothing effect can become undesirable (this is for instance the case if the output of the statistical model is used to drive an ecosystem or sector model). To deal with this objective, attention is drawn to a different set of geostatistical techniques called *conditional simulations* (Journel and Huijbregts [14], Armstrong and Dowd [2]). The idea of conditional simulations consists in drawing another realization $\{z_s(\mathbf{x})\}$ from the random function model Z and to impose that this realization meets the experimental values at the data locations $\mathbf{x}_\alpha$, *i.e.* we look for a conditionally simulated $z_{sc}(\mathbf{x})$ satisfying

$$z_{sc}(\mathbf{x}_\alpha) = z(\mathbf{x}_\alpha).$$

Considering the "true" regionalized variable z at each point of a data set, the geostatistical approach consists in interpreting z as a particular realization of a random function Z. This random function is characterized by its first two moments and its distribution function, which are estimated, in a first step, from the experimental data. After an anamorphosis into a Gaussian random function Y, in a second step, we draw a realization $\{y_s(\mathbf{x})\}$. This realization has the advantage of being known at all points $\mathbf{x}$ and not only at the experimental data points $\mathbf{x}_\alpha$ and it is also called a *numerical model* of the real variable. We get in a third step the required conditional simulation by writing

$$y_{sc}(\mathbf{x}) = y^*(\mathbf{x}) + \big[y_s(\mathbf{x}) - y_s^*(\mathbf{x})\big],$$

where $y_s(\mathbf{x})$ represents any simulation (*i.e.* non conditional) of the random function Y under study and $y_s^*(\mathbf{x})$ represents the result of kriging procedure applied to $y_s(\mathbf{x})$ with the initial data configuration $\mathbf{x}_\alpha$ of the observations. In the fourth step, the conditional simulation is back-transformed into $z_{sc}(\mathbf{x})$.

The requirements of a conditional simulation are satisfied, and in particular, this entails that for all $\mathbf{x}_\alpha$

$$z_{sc}(\mathbf{x}_\alpha) = z(\mathbf{x}_\alpha).$$

By construction, the variance of the simulated points should be identical to the variance of observations. On the other hand, the price to be paid is that at each point $\mathbf{x}$, the simulated value $z_{sc}(\mathbf{x})$ is not the best possible estimator of $z(\mathbf{x})$. Hence, though the simulation values $z_{sc}(\mathbf{x})$ reproduce better the variance of the real data, the estimations $z^*(\mathbf{x})$ are "closer" (in the least squares sense) to the real values $z(\mathbf{x})$. A similar situation has been found in the context of estimation by linear regression models $\big($Bürger [5]$\big)$.

Thus, in general, an optimal estimation and the replication of the variance seem not to be compatible. This can be illustrated as follows: a statistical model for one predictor y and a multivariate predictand $\mathbf{x}$ can be written in general as

$$y(t) = F\big(\mathbf{x}(t)\big) + \epsilon(t),$$

where ϵ describes the part of y that cannot be taken into account by $\mathbf{x}$. In our case y represents local rainfall and $\mathbf{x}$ represents the SLP field. But the SLP field cannot explain all of the rainfall variability; there will always be a part of the rainfall variability due to local noise, such as measurement errors, convection, localized winds, that are totally or to a great extent independent of the large-scale SLP. Therefore any statistical model in similar situation as this (*i.e.* linking two physically distinct variables) will tend to replicate a smaller variance than observed.

This comment does not apply to statistical models that incorporate stochastically the observed local noise ϵ, such as the analog method. By sampling from a pool of past observations, the analog method introduces this noise implicitly. This makes the analog method worse than other "deterministic" methods in specifying the rainfall at time t, $y(t)$, since the stochastically introduced noise for one estimation at time t, $\epsilon(t_{ana})$, will not be the same as the *observed* noise at time t, $\epsilon(t)$; but on the other hand the analog method will reproduce the right variance.

8 Conclusion

The main goal of this paper was to conceive and to test a downscaling procedure based on the geostatistical technique kriging. Noteworthy is the application of the kriging technique in a space without geographical meaning. Over the last half of this century, geostatistics has found many new domains of application, the latest to date being perhaps in the field of electromagnetism (Lefèbvre *et al* [18], Walter and Pronzato [29]), where kriging is also applied in a non-geographical space.

From a technical point of view, it was found that both kriging and the simpler analog method were able to reproduce in a very realistic manner both rainfall evolution and observational means at most of the stations under study. On the other hand, kriging underestimates the observation variance because of a well known "smoothing effect". Furthermore, it was pointed out that the results of kriging may be more robust than those of the analog method to the distribution of the observations in EOF space and the number of the observational data points.

We checked the validity of both kriging and analog downscaling approaches by trying to downscale winter daily precipitation from daily SLP. It was concluded that kriging was able to replicate better the time evolution of rainfall, whereas the analog method yielded a more realistic variance and probability histograms. In the corresponding section, we performed variographical analyses and kriging in a 15-dimensional space, which is a very unusual situation for geostatistics.

We also found that the performance of the downscaling based on kriging seem to depend strongly on the distribution of the local variable, for example monthly versus daily rainfall, at least in the normal kriging setting used in this paper. This property may hinder the application of the method to other environmental variables that are usually non-normally distributed. In those cases other kriging techniques that do not rely heavily on the probability distribution could be worth to be applied.

Thus, the geostatistical downscaling process designed in this study, using North Atlantic SLP data as input and yielding Iberian precipitation estimations as output, was successfully validated by reconstructing precipitation observations. It could be hence possible, as the last step of downscaling, to apply this new downscaling scheme to simulated SLP data derived from a GCM climate change experiment.

Acknowledgments. Many results presented in this paper were achieved using the geostatistical software Isatis and the statistical macrolanguage S-Plus. We would like to thank Pierre Chauvet and Reiner Schnur for their helpful comments. We are also grateful to Fidel González from the Universidad Complutense (Madrid) for providing part of the Iberian precipitation data. Several research visits of one of the authors (Gérard Biau) were financed by the GKSS-Research Centre. The cooperation between the Ecole des Mines and the GKSS-Research Centre was initiated at the 6^{th} International Meeting on Statistical Climatology in Galway (Ireland).

9 Appendix

Definition 9.1 *The random function* Z *is said to be* second-order stationary *if for every point* $\mathbf{x}$:

1. *the mean* $\mathbf{E}\big[Z(\mathbf{x})\big]$ *exists and does not depend on* $\mathbf{x}$:
$$\mathbf{E}\big[Z(\mathbf{x})\big] = m,$$
2. *and the covariance between* $Z(\mathbf{x})$ *and* $Z(\mathbf{x}+\mathbf{h})$ *exists and depends only on the separation vector* $\mathbf{h}$:
$$\mathrm{Cov}\big[Z(\mathbf{x}), Z(\mathbf{x}+\mathbf{h})\big] = C(\mathbf{h}).$$

In this situation the variogram is bounded and, conversely, only when the variogram is bounded does a corresponding covariance function exists. With second-order stationarity, it follows that:

- $\mathrm{Var}\big[Z(\mathbf{x})\big] = \mathrm{Cov}\big[Z(\mathbf{x}), Z(\mathbf{x})\big] = C(\mathbf{0})$
- $\gamma(\mathbf{h}) = \frac{1}{2}\mathrm{Var}\big[Z(\mathbf{x}+\mathbf{h}) - Z(\mathbf{x})\big] = C(\mathbf{0}) - C(\mathbf{h}).$

In the general case the variogram is unbounded and an *intrinsically stationary* random function model is adopted, where the differences $Z(\mathbf{x}+\mathbf{h}) - Z(\mathbf{x})$ are assumed second-order stationary, while $Z(\mathbf{x})$ itself is not necessarily stationary.

Now let us discuss the construction of an estimator Z^*. We are dealing with a minimization without constraint, when the random function Z is assumed to have a zero mathematical expectation. However, this is an exceptional case and in most cases the expectation - assumed constant in the second-order stationary model - is unknown in the intrinsically stationary model. Consequently, in order to ensure the estimation error $Z^*(\mathbf{x}) - Z(\mathbf{x})$ will satisfy
$$\mathbf{E}\big[Z^*(\mathbf{x}) - Z(\mathbf{x})\big] = 0$$
we have to impose the constraint
$$\sum_{\alpha=1}^{N} \lambda_\alpha = 1.$$

This constraint ensures the unbiased behavior of the kriging estimator and is called the *universality condition*. Thus, the variance of the estimation error $Z^*(\mathbf{x}) - Z(\mathbf{x})$ can then be written as:

$$\begin{aligned}\mathrm{Var}\big[Z^*(\mathbf{x}) - Z(\mathbf{x})\big] &= \mathbf{E}\Big\{\big[Z^*(\mathbf{x}) - Z(\mathbf{x})\big]^2\Big\} \\ &= 2\sum_{\alpha=1}^{N}\lambda_\alpha\gamma(\mathbf{x}-\mathbf{x}_\alpha) - \sum_{\alpha=1}^{N}\sum_{\beta=1}^{N}\lambda_\alpha\lambda_\beta\gamma(\mathbf{x}_\alpha-\mathbf{x}_\beta)\end{aligned}$$

by simple application of the general formula. It is a quadratic equation, according to the unknown coefficients λ_α which have to be minimized. The kriging problem amounts then to minimize a quadratic form under a linear constraint. We obtain a system of linear equations:

$$\begin{cases} -\sum\limits_{\beta=1}^{N}\lambda_\beta\gamma(\mathbf{x}_\alpha-\mathbf{x}_\beta)+\mu &= -\gamma(\mathbf{x}-\mathbf{x}_\alpha) \\ \\ \sum\limits_{\beta=1}^{N}\lambda_\beta &= 1. \end{cases}$$

The unknowns of the system are the weights λ_α and the Lagrangian multiplier μ. The kriging variance σ_k^2 is the minimum of $\mathrm{Var}\big[Z^*(\mathbf{x}) - Z(\mathbf{x})\big]$, and is given by

$$\sigma_k^2 = \sum_{\alpha=1}^{N}\lambda_\alpha\gamma(\mathbf{x}-\mathbf{x}_\alpha) - \mu,$$

where the λ^α and μ are the solutions to the kriging system.

A weaker condition than second-order stationarity is given by the *intrinsic-without-drift* condition: the function Z is said to be *intrinsic* if for any vector $\mathbf{h}$ the increment $Z(\mathbf{x}+\mathbf{h}) - Z(\mathbf{x})$ is second-order stationary. The function $m(\mathbf{h}) = \mathbf{E}\big[Z(\mathbf{x}+\mathbf{h}) - Z(\mathbf{x})\big]$ is then called the *drift* and verifies the linear relation $m(\mathbf{h}+\mathbf{h}') = m(\mathbf{h}) + m(\mathbf{h}')$.

Furthermore, when the drift function is supposed equal to zero, the random function is said to be intrinsic-without-drift. This important assumption is made here. Therefore, for any vector $\mathbf{h}$, the increment $Z(\mathbf{x}+\mathbf{h}) - Z(\mathbf{x})$ has a finite variance which does not depend on $\mathbf{x}$, and we get:

$$\gamma(\mathbf{h}) = \frac{1}{2}\mathrm{Var}\big[Z(\mathbf{x}+\mathbf{h}) - Z(\mathbf{x})\big] = \frac{1}{2}\mathbf{E}\Big\{\big[Z(\mathbf{x}+\mathbf{h}) - Z(\mathbf{x})\big]^2\Big\}.$$

Note that the second-order stationarity implies the intrinsic condition but that the converse is not true. However, by a mathematical transformation it is possible to show that the resulting kriging equations and the kriging variance are given by the same equations as in the second-order stationary case.

Under quite reasonable conditions kriging estimation has interesting properties:

- By construction, kriging is an unbiased estimator, *i.e.*, the estimation error has zero expectation in the probabilistic model. When considering the regionalized variable, this means that the average error on point estimations is zero over a large area.
- Kriging is an exact interpolator, which means that when we perform kriging at a sampling point, the kriging system will return the sample value as the estimator (and a kriging variance of zero).

References

[1] Amani, A. and Lebel, T. (1997) Lagrangian kriging for the estimation of Sahelian rainfall at small time steps, *Journal of Hydrology*, **Vol. 192**, pp. 125–157.

[2] Armstrong, M. and Dowd, P.A. (1994) *Geostatistical Simulation*, Kluwer Academic Publisher, Dordrecht.

[3] Bardossy, A. and Plate, J. (1992) Space-time model for daily rainfall using atmospheric circulation patterns, *Water Resource Research*, **Vol. 28**, pp. 1247–1259.

[4] Bretherton, C., Smith, C. and Wallace, J. (1992) An intercomparison of methods to find coupled patterns in climate data, *Journal of Climate*, **Vol. 5**, pp. 541–560.

[5] Bürger, G. (1996) Expanded downscaling for generating local weather scenarios, *Climate Research*, **Vol. 7**, pp. 111–128.

[6] Busuioc, A. and von Storch, H. (1996) Changes in winter precipitation in Romania and its relation to the large-scale circulation, *Tellus*, **Vol. 48A**, pp. 538–552.

[7] Chauvet, P. (1994) *Aide-Mémoire de Géostatistique Linéaire*, Cahiers de Géostatistique, Ecole des Mines de Paris, **fasc. 2**.

[8] Giorgi, F. and Mearns, L.O. (1991) Approaches to the simulations of regional climate change: a review, *Revue of Geophysics*, **Vol. 29**, pp. 191–216.

[9] Grotch, S. and MacCracken, M. (1991) The use of General Circulation Models to predict regional climate change, *Journal of Climate*, **Vol. 4**, pp. 286–303.

[10] Hewitson, B. (1996) Climate downscaling: techniques and application, *Climate Research*, **Vol. 7**, pp. 85–95.

[11] Holdaway, M.R. (1996) Spatial modelling and interpolation of monthly temperature using kriging, *Climate Research*, **Vol. 6**, pp. 215–225.

[12] Hughes, J., Lettenmaier, D. and Guttorp, P. (1993) A stochastic approach for assessing the effect of changes in regional circulation patterns on local precipitation, *Water Resource Research*, **Vol. 29**, pp. 3303–3315.

[13] Hulme, M., Briffa, K., Jones, P. and Senior, C. (1993) Validation of GCM control simulations using indices of daily airflow types over the British Isles *Climate Dynamics*, **Vol. 9**, pp. 95–105.

[14] Journel, A.G. and Huijbregts, C.J. (1978) *Mining Geostatistics*, Academic Press, London.

[15] Karl, T.M., Wang, W.C., Schlesinger, M.E., Knight, D.E. and Portman, D. (1990) A method of relating General Circulation Model simulated climate to the observed local climate. Part I: seasonal statistics, *Journal of Climate*, **Vol. 3**, pp. 1053–1079.

[16] Kushnir, Y., Cardonne, V.J., Greenwood, J.G. and Cane, M.A. (1997) The recent increase in North Atlantic wave heights, *Journal of Climate*, **Vol. 10**, pp. 2107–2113.

[17] Lamb, P. and Peppler, R. (1987) The North Atlantic Oscillation: concept and an application, *Bulletin of the American Meteorological Society*, **Vol. 68**, pp. 1218–1225.

[18] Lefèbvre, J., Roussel, H., Walter, E., Lecointe, D. and Tabbara, W. (1996) Prediction from wrong models, the kriging approach, *IEEE Antenna and Propagation Magazine*, **Vol. 38**, pp. 33–45.

[19] Mardia, K.V., Kent, J.T. and Bibby, J.M. (1978) *Multivariate Analysis*, Academic Press, London.

[20] Matheron, G. (1969) *Le Krigeage Universel*, Cahiers du Centre de Morphologie Mathématique de Fontainebleau, Ecole des Mines de Paris, **fasc. 1**.

[21] Rivoirard, J. (1990) A review of log-normal estimators for in situ reserves, *Mathematical Geology*, **Vol. 22**, pp. 213–221.

[22] Rivoirard, J. (1994) *Introduction to Disjunctive Kriging and Nonlinear Geostatistics*, Oxford University Press, Oxford.

[23] Rouhani, S. and Wackernagel, H. (1990) Multivariate geostatistical approach to space-time data analyses, *Water Resource Research*, **Vol. 26**, pp. 585–591.

[24] Saporta, G. (1990) *Probabilités, Analyse des Données et Statistique*, Technip, Paris.

[25] von Storch, H. (1995) Inconsistencies at the interface between climate research and climate impact studies, *Metereologische Zeitschrift*, **Vol. 4**, pp. 72–80.

[26] von Storch, H. and Navarra, A. (1995) *Analysis of Climate Variability. Applications of Statistical Techniques*, Springer–Verlag, Berlin.

[27] von Storch, H., Zorita, E. and Cubasch, U. (1993) Downscaling of global climate change estimates to regional scales: an application to Iberian rainfall in wintertime, *Journal of Climate*, **Vol. 6**, pp. 1161–1171.

[28] Wackernagel, H. (1995) *Multivariate Geostatistics*, Springer–Verlag, Berlin.

[29] Walter, E. and Pronzato, L. (1997) *Identification of Parametric Models from Experimental Data*, Springer–Verlag, Berlin.

[30] Wilby, R.L. and Wigley, T.M.L. (1997) Downscaling General Circulation Model output: a review of methods and limitations, *Progress in Physical Geography*, **Vol. 21**, pp. 530–548.

[31] Zorita, E., Hugues, P.J., Lettenmaier, D.P. and von Storch, H. (1995) Stochastic characterization of regional patterns for climate model diagnosis and estimation of local precipitation, *Journal of Climate*, **Vol. 8**, pp. 1023–1042.