DOWNSCALING OF PRECIPITATION COMBINING KRIGING AND EMPIRICAL ORTHOGONAL FUNCTION ANALYSIS

G. BIAU
Laboratoire de Probabilités et Statistique
Université Montpellier II
Place Eugène Bataillon
34095 Montpellier CEDEX 5

Abstract. The term downscaling denotes a procedure in which local climatic information is derived from large-scale climate parameters. In this paper, we conceive a downscaling procedure based on kriging and 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).

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 so are 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 1991; von Storch 1995). Several general methods, called downscaling methods, have been suggested so far to overcome this scale mismatch (dynamical downscaling, Giorgi and Mearns 1991; statistical downscaling, Karl *et al.* 1990). In this contribution, we explore the possibility to use for downscaling purposes the well-known geostatistical interpolation technique called kriging (for a complete presentation of geostatistics and its methods, see for example Chauvet 1994; Wackernagel 1995). The main idea of our method consists in reducing the spatial dimension of the large-scale Sea Level Pressure (SLP) field by means of Empirical Orthogonal Function Analysis (EOF Analysis, also known as Principal Component Analysis, see *e.g.* Mardia *et al.* 1978; Saporta 1990). Each observed SLP data 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. We present an example of the method by trying to reconstruct the winter precipitation in stations located in the Iberian Peninsula 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 (Zorita *et al.* 1995), which due to its simplicity may be considered in this context as a benchmark.

The paper is organized as follows: after a short presentation of data in section 2, 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 variograms. Section 5 presents then the detailed results of kriging and analog downscaling. Some important aspects of the results obtained are discussed in section 6. The paper is closed by some concluding remarks in section 7. Further, more complete results about the presented material will be published later this year in the *Journal of Climate*.

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.

The SLP data set was obtained from the National Meteorological Centre (Boulder). It consists of monthly mean (DJF) SLP over the same period (1899 - 1989). The original SLP data were interpolated onto a 5° latitude ×

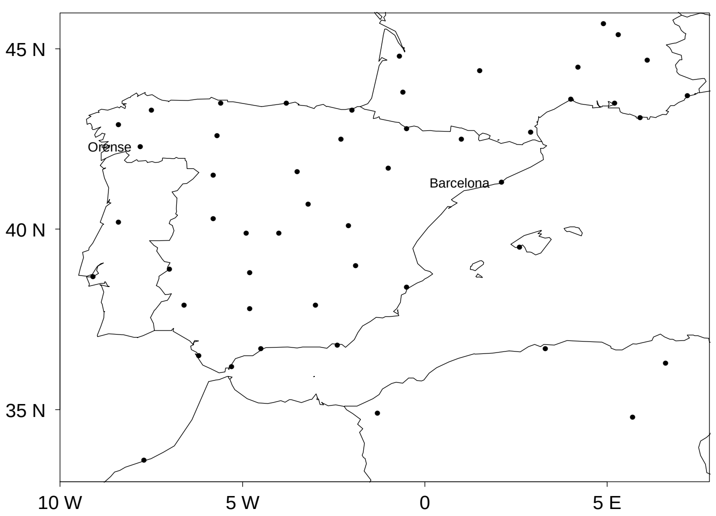

Figure 1. Positions of the 50 stations under study. The names correspond to the stations explicitly cited in the text.

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 (see Figure 2). A wide westward extension was chosen since it is known that the Atlantic weather systems influence Iberian Peninsula rainfall.

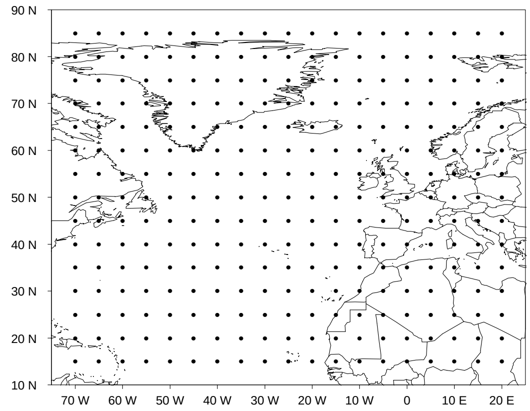

Figure 2. SLP data.

3. 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 indeed 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 *et al.* 1978; Saporta 1990) is to obtain an optimal approximate representation of the original data set by projecting it onto a subspace of lower dimension (here, the winter months from 1899 to 1989 are considered as the observations and the SLP spatial measurements as the variables).

The EOF Analysis allows the representation of much of the variability of the SLP data by only a few coordinates. In this case only the two leading EOFs have been retained, that together explain 60% of the SLP variability.

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 west and in the east of the peninsula, are presented in Figure 3.

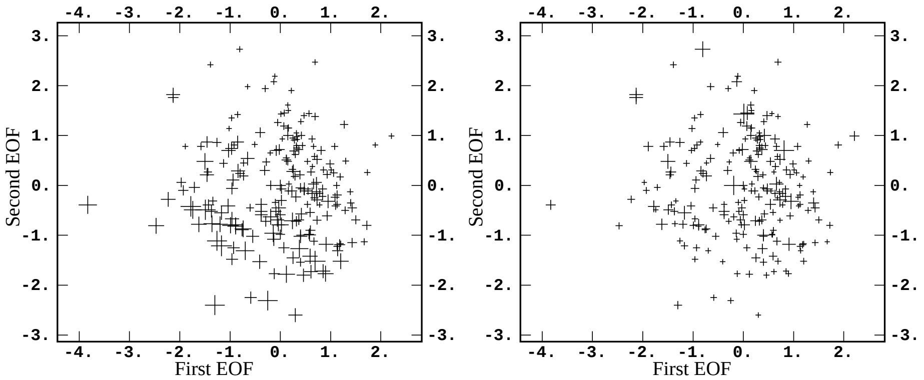

Figure 3. Monthly rainfall in winter represented in the space of the two leading SLP EOFs: Orense (left) and Barcelona (right) data. The size of each cross is proportional to the rainfall amount attached to the point. Note that the locations of the data are the same for the two stations.

The points where the rainfall values may be estimated are points given either by SLP extra individuals (*i.e.* individuals that did not enter the initial EOF Analysis) observed in other time period (*e.g.* for validation of this method) or by SLP extra individuals simulated by a GCM (for downscaling of simulated climate change). In both cases, we project each SLP extra individual onto the leading two SLP EOFs yielding two coordinates in the EOF space. Then kriging enables the values of local precipitation associated with the new SLP situations 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 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. Structural analyses and kriging

Sometimes a variographical study is 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 h of $\mathbf{h}$. 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 1978). Inspection of the experimental variogram for each station revealed that the 50 variograms 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 effect (flat variograms).
- 9 variograms can be fitted with a bounded nested structure composed of 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), and a spherical model. Due to the normalization of the EOF coordinates, h ranges typically between 0 and 9 and in our case the range $a \simeq 0.5$, making it sometimes difficult to choose between a pure nugget effect and a spherical model.
- Finally, the 35 remaining observed variograms are similar and can be modeled by an unbounded model variogram as follows: a nugget effect, with the amplitude contained between 25% and 70% of the data vari-

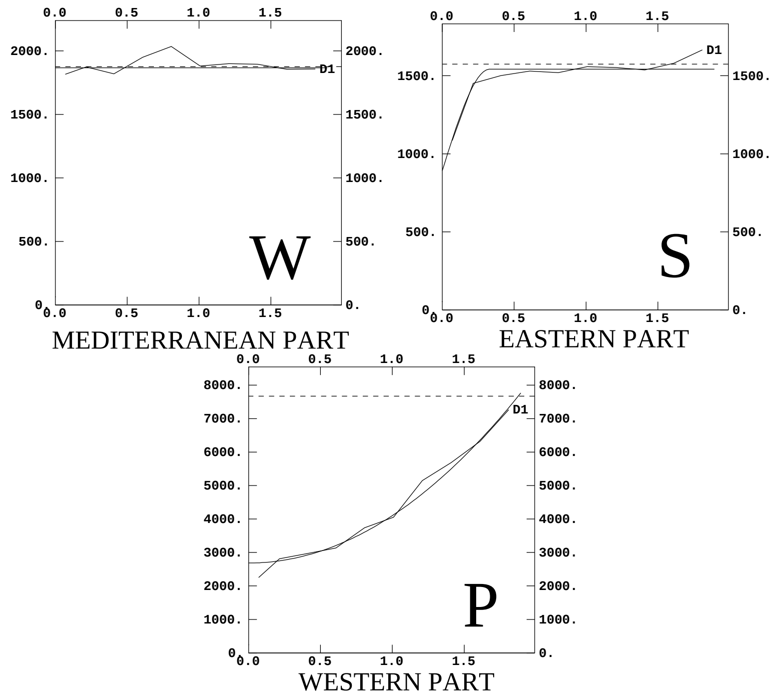

Figure 4. Three typical examples of experimental variograms (solid lines) with their associated models (thick solid lines). The notations are as follows: **W** for pure nugget effect ("**W**hite noise"), **S** for spherical model and **P** for power model. We also show the data variance (dotted lines).

ance (see Figure 4), and an unbounded variogram, which is modeled by an intrinsic-without-drift model (we chose a power model).

Furthermore, it was observed that the stations with unbounded variogram models are located in the west and centre 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). 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 centre 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

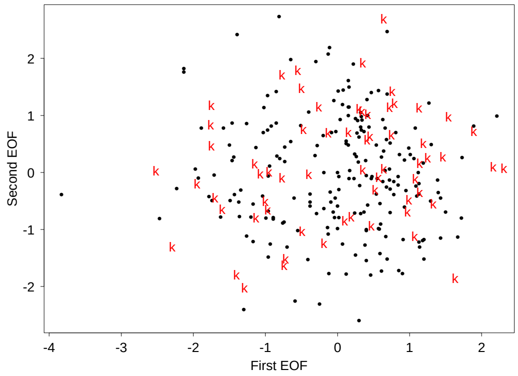

Figure 5. Position of the coordinates in the EOF space. Points indicate the data used for the variographical analysis (1899-1969), "k" denotes the points where the estimation was performed (1969-1989).

the winter months of high precipitation and those of weak precipitation for stations located in the west and in the centre 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.

5. Validation of the method and kriging

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 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). In this study, as we have less than 100 points, we have used unique neighbourhood throughout. With more data locations, a sliding neighbourhood would have been appropriate.

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.* 1995). As applied by these authors the analog method operates 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:

1. The linear correlation coefficients between the observed and simulated rainfall time series.
2. The observational means of the reconstructed rainfall and of the *in situ* observations.
3. The observed and reproduced precipitation standard deviations.

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

- The kriging linear correlation coefficients are good in the centre 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).
- The means are very well reproduced by both downscaling methods.
- The standard deviations are acceptably reconstructed by the analog method but underestimated by kriging.

As a complementary remark it should be mentioned that we conducted the same geostatistical study using, instead the pair (EOF1,EOF2) as basis

[1]The analog method is called Polygon method in mining (see *e.g.* Wackernagel 1995) where it was abandoned because it neglects the support effect.

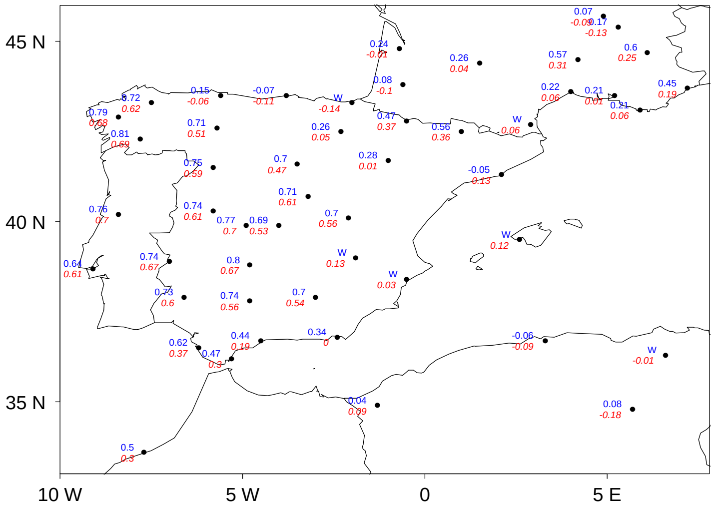

Figure 6. Linear correlation coefficients between kriging reconstructions and observations (top) and between analog method reconstructions and observations (bottom). Stations with a “W” have a white noise variogram model.

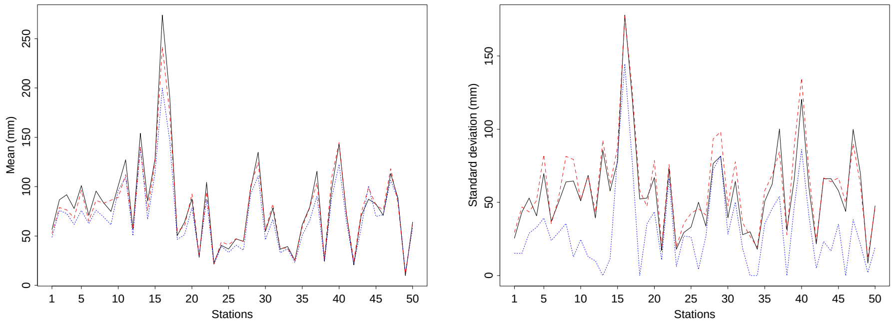

Figure 7. **(Left)** Means of observations (solid lines), kriging reconstructions (dotted lines) and analog method reconstructions (dashed lines). **(Right)** Standard deviations of observations (solid lines), kriging reconstructions (dotted lines) and analog method reconstructions (dashed lines).

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.

6. Discussion

6.1. VARIOGRAPHICAL ANALYSES AND KRIGING RESULTS

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

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

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.

6.2. REPLICATION OF THE MEAN AND VARIANCE

We noticed in 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.

More precisely, 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. By construction, kriging is an unbiased estimator, 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. In general it is possible to show (Chauvet 1994) that in the random function model:

$$Var[Z^*(\mathbf{x})] \leq Var[Z(\mathbf{x})], \tag{1}$$

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 correct for the smoothing effect of kriging, attention is drawn to a different set of geostatistical techniques called conditional simulations (Journel and Huijbregts 1978; Armstrong and Dowd 1994). 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 $\mathbf{x}_\alpha$.

7. 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.* 1996; Walter and Pronzato 1997), 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 as well the validity of both the kriging and analog downscaling approaches by trying to downscale winter daily precipitation from daily SLP (not shown). 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 this approach, we performed variographical analyses and kriging in a 15-dimensional space, which is a very unusual situation for geostatistics. Thus, the geostatistical downscaling process designed in the present paper, 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

This work has been achieved at GKSS Research Centre in close collaboration with Hans von Storch (GKSS Research Centre, Germany), Hans Wackernagel (Centre de Géostatistique, École des Mines de Paris, France) and Eduardo Zorita (GKSS Research Centre, Germany). I thank all of them.

References

Armstrong, M. and Dowd, P.A. (1994) *Geostatistical simulation*, Kluwer Academic Publisher, Dordrecht.
Chauvet, P. (1994) Aide-mémoire de géostatistique linéaire, *École des Mines de Paris, Cahiers de Géostatistique*, **Fasc. 2**.
Giorgi, F. and Mearns, L.O. (1991) Approaches to the simulations of regional climate change: a review, *Rev. Geophys.*, **Vol. 29**, pp. 191–216.
Grotch, S. and MacCracken, M. (1991) The use of General Circulation Models to predict regional climate change, *J. Climate*, **Vol. 4**, pp. 286–303.
Journel, A.G. and Huijbregts, C.J. (1978) *Mining geostatistics*, Academic Press, London.
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, *J. Climate*, **Vol. 3**, pp. 1053–1079.
Lefèbvre, J., Roussel, H., Walter, E., Lecointe, D. and Tabbara, W. (1996) Prediction from wrong models, the kriging approach, *Ant. and Prop. Mag., IEEE*, **Vol. 38**, pp. 33–45.
Mardia, K.V., Kent, J.T. and Bibby, J.M. (1978) *Multivariate analysis*, Academic Press, London.
Saporta, G. (1990) *Probabilités, analyse des données et statistique*, Technip, Paris.
von Storch, H. (1995) Inconsistencies at the interface between climate research and climate impact studies, *Metereologische Zeitschrift*, **Vol. 4**, pp. 72–80.
Wackernagel, H. (1995) *Multivariate Geostatistics*, Springer-Verlag, Berlin.
Walter, E. and Pronzato, L. (1997) *Identification of parametric models from experimental data*, Springer-Verlag, Berlin.
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, *J. Climate*, **Vol. 8**, pp. 1023–1042.