Hydrology & Earth
System Sciences

Influence of parameter estimation uncertainty in Kriging: Part 2 – Test and case study applications

Ezio Todini, Federica Pellegrini and Cinzia Mazzetti

Department of Earth and Geo-Environmental Sciences, University of Bologna, Italy
Email for corresponding author: todini@geomin.unibo.it

Abstract

The theoretical approach introduced in Part 1 is applied to a numerical example and to the case of yearly average precipitation estimation over the Veneto Region in Italy. The proposed methodology was used to assess the effects of parameter estimation uncertainty on Kriging estimates and on their estimated error variance. The Maximum Likelihood (ML) estimator proposed in Part 1, was applied to the zero mean deviations from yearly average precipitation over the Veneto Region in Italy, obtained after the elimination of a non-linear drift with elevation. Three different semi-variogram models were used, namely the exponential, the Gaussian and the modified spherical, and the relevant biases as well as the increases in variance have been assessed. A numerical example was also conducted to demonstrate how the procedure leads to unbiased estimates of the random functions. One hundred sets of 82 observations were generated by means of the exponential model on the basis of the parameter values identified for the Veneto Region rainfall problem and taken as characterising the true underlining process. The values of parameter and the consequent cross-validation errors, were estimated from each sample. The cross-validation errors were first computed in the classical way and then corrected with the procedure derived in Part 1. Both sets, original and corrected, were then tested, by means of the Likelihood ratio test, against the null hypothesis of deriving from a zero mean process with unknown covariance. The results of the experiment clearly show the effectiveness of the proposed approach.

Keywords: yearly rainfall, maximum likelihood, Kriging, parameter estimation uncertainty

Introduction

The analysis of the effect of parameter estimation uncertainty on the estimated Kriging variables proposed in Part 1 (Todini, 2001) was applied to the mean annual precipitation of the Veneto region in Northern Italy.

The complexity of the spatial phenomenon, which deals with precipitation influenced by orographic factors, does not warrant the *a priori* assumption of stationarity in the mean. Given the presence of a non-linear gradient of precipitation with elevation, digital elevation model data (with pixels of 1×1 km^2) were therefore used to eliminate this dependence.

The data used in this study consist of mean annual precipitation from 1961 to 1990, collected at 82 recording stations. The area bounded by Latitudes 45°01' and 46°34' N (approximately 172 km) and Longitudes 10°46' and 13°04' E (approximately 183.5 km) contains a suitably dense and well-distributed network of stations, as can be seen from Fig. 1 (which was produced by means of IDRISI, a widely used Geographic Information System (GIS)
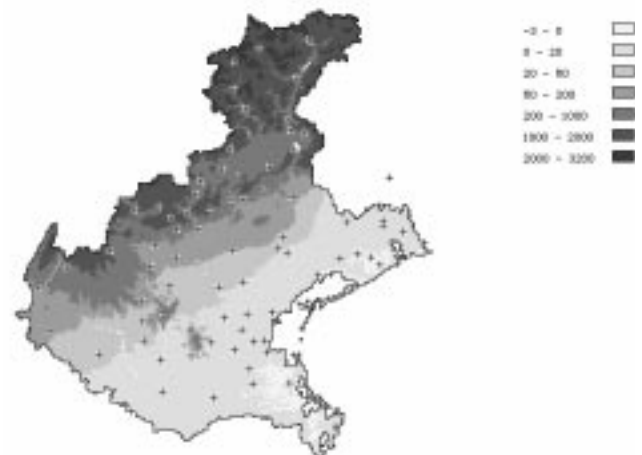

Fig. 1. *Digital Elevation Model of the Veneto Region with rain gauging stations. Elevation in m - Distance ranges: N-S circa 183.5 km; E-W circa 172 km*

software (Eastman, 1992)), where the 1×1 km² digital elevation model of the Veneto Region is shown together with the location of the rain-gauging stations.

As shown in Fig. 2, a non-linear trend with elevation was considered and the following relationship was derived:

$$\bar{r}(h(x,y)) = a_0 + a_1 \{1 - \exp[-h(x,y)/a_2]\} \qquad (1)$$

in which $\bar{r}$ is the deterministic trend in the mean annual precipitation r, $h(x,y)$ is the local terrain elevation, and a_0, a_1, a_2 are three parameters, which were estimated by means of a non-linear regression to give $a_0 = 802.023$, $a_1 = 651.775$, $a_2 = 113.09$.

The Kriging variables z^* were thus obtained as the deviations from the mean drift:

$$z^*(x,y) = r(x,y) - \bar{r}(h(x,y)) \qquad (2)$$

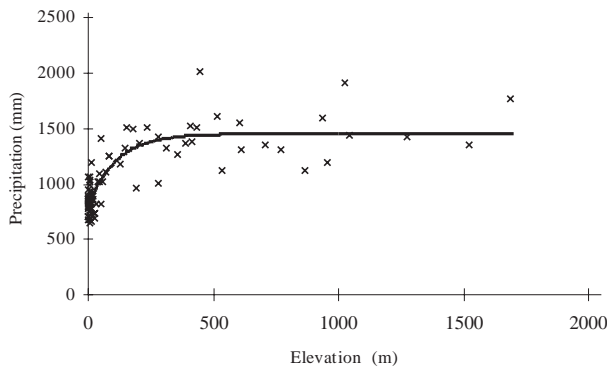and, for the sake of clarity, will be referred in the text as adjusted rainfall.

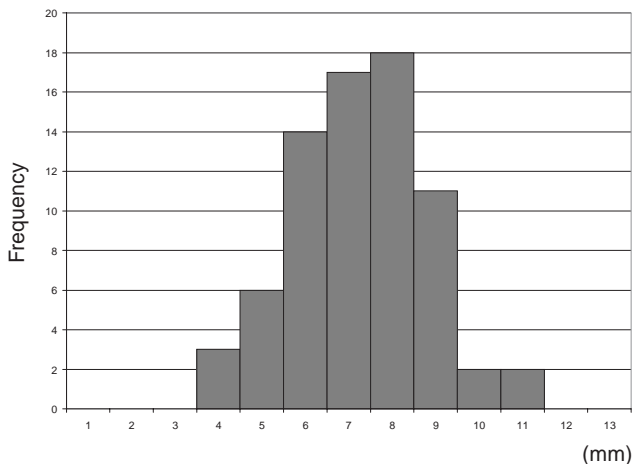Fig. 2. *The non-linear drift of the mean annual precipitation with gauge elevation (a.m.s.l)*

Fig. 3. *The frequency distribution of residuals from the non linear drift (adjusted rainfall) with elevation*

The distribution of the Kriging variables, although this property is not particularly stringent, as already discussed in Part 1, is very close to a Normal distribution, as can be seen from Fig. 3 with classes ranging between –400 to +400 mm.

Parameter estimation

Three theoretical semi-variogram models were tested: exponential, Gaussian and modified spherical, given in Eqns. (3), (4) and (5) respectively, to estimate the parameters of the structural model.

$$\gamma(h) = p + \omega \left[1 - e^{-h/\alpha}\right] \qquad \text{EXPONENTIAL} \qquad (3)$$

$$\gamma(h) = p + \omega \left[1 - e^{-(h/\alpha)^2}\right] \qquad \text{GAUSSIAN} \qquad (4)$$

$$\begin{cases} \gamma(h) = p + \omega \left[1.875\dfrac{h}{\alpha} - 1.25\left(\dfrac{h}{\alpha}\right)^3 + 0.375\left(\dfrac{h}{\alpha}\right)^5\right] & h < \alpha \\ \gamma(h) = p + \omega & h \leq \alpha \end{cases}$$

$$\text{MODIFIED SPHERICAL} \qquad (5)$$

As discussed by Todini (2001), the modified spherical model is used in this application instead of the more classical one of Eqn. (6) because it guarantees the continuity of the function, together with that of its first and second derivatives with respect to the parameters over the entire domain of existence, as it is required by Maximum Likelihood to be minimum variance (Kendall and Stuart, 1973).

$$\begin{cases} \gamma(h) = p + \omega \left[1.5\dfrac{h}{\alpha} - 0.5\left(\dfrac{h}{\alpha}\right)^3\right] & h < \alpha \\ \gamma(h) = p + \omega & h \leq \alpha \end{cases}$$

$$\text{SPHERICAL} \qquad (6)$$

A first attempt at fitting the three models of Eqns. (3), (4) and (5) was made by using the ML estimator under the assumption of spatially independent cross-validation errors, as suggested by Samper and Newman (1989) and by Todini and Ferraresi (1996).

The results of this parameter identification cannot be considered satisfactory. This is evident from Figs. 4a, 4b and 4c, where the estimated semi-variogram is compared to the average sampling values bounded by $\pm 2\sigma$ limits, obtained for the different classes of distance, using the standard Kriging technique. This is obtained by dividing the maximum distance range between measurement points in distance classes (of equal length or of equal frequency of observations); the quantity $g(h_i) = [z(s) - z(s + h_i)]^2$ is

then computed from each couple of points falling in the class and used for estimating the average value and its standard deviation in the class.

The results obtained show that, given that the cross-validation errors are actually spatially correlated (de Marsily,

1986), the assumption of independence does indeed affect the parameter and the semi-variogram estimates.

Much better results can be obtained using ML under the assumption of correlated cross-validation errors, as is seen in Figs. 5a, 5b and 5c. These figures show that there is a
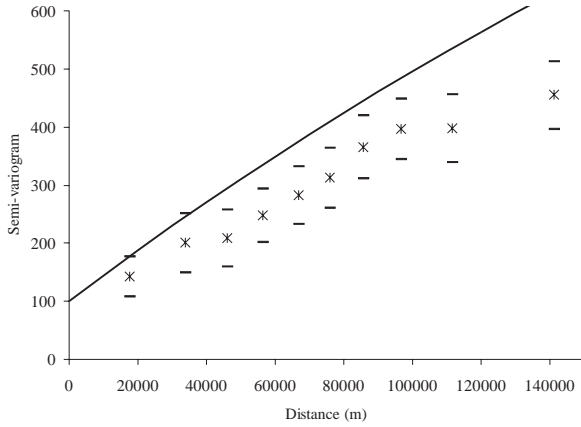

Fig. 4a. *Comparison of the exponential semi-variogram with average classes values assuming the hypothesis of independent cross-validation errors*
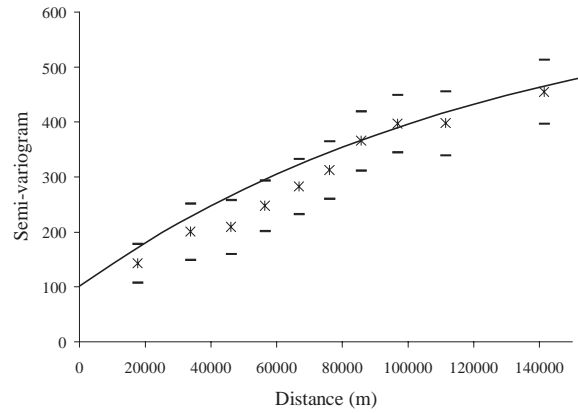

Fig. 5a. *Comparison of the exponential semi-variogram with average classes values assuming the hypothesis of correlated cross-validation errors*
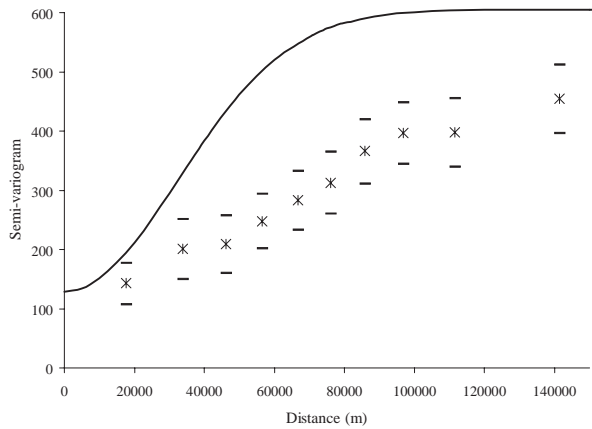

Fig. 4b. *Comparison of the Gaussian semi-variogram with average classes values assuming the hypothesis of independent cross-validation errors*
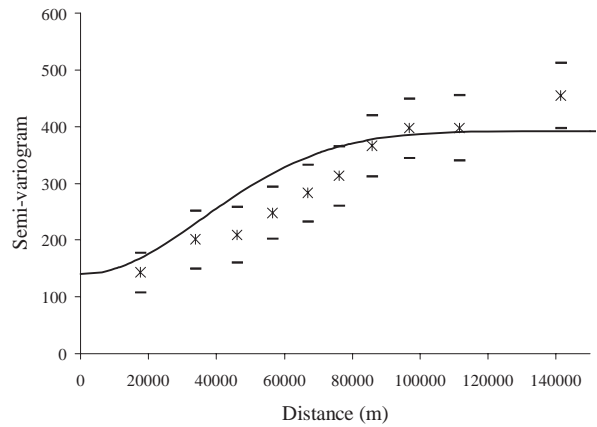

Fig. 5b. *Comparison of the Gaussian semi-variogram with average classes values assuming the hypothesis of correlated cross-validation errors*
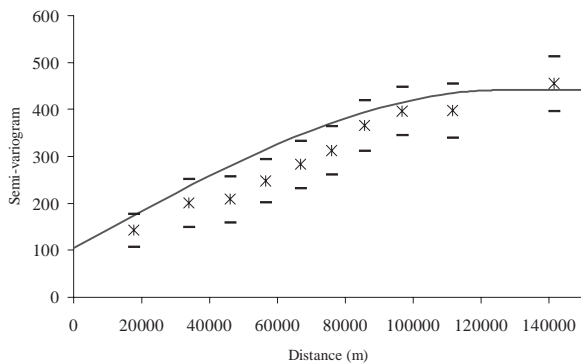

Fig. 4c. *Comparison of the modified spherical semi-variogram with average classes values assuming the hypothesis of independent cross-validation errors*
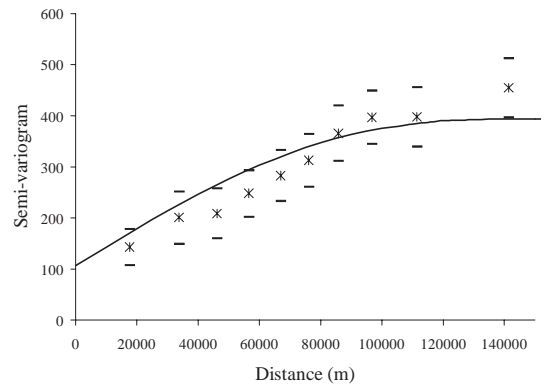

Fig. 5c. *Comparison of the modified spherical semi-variogram with average classes values assuming the hypothesis of correlated cross-validation errors*

Table 1. Parameter values and cross-validation errors statistics (errors in mm) for the three models assuming correlated cross-validation errors

	Parameters			*Statistics*			
	Nugget [cm^{2}]	**Sill** [cm^2]	**Range** [m]	$\frac{1}{n}\sum\limits_{k=1}^{n}\varepsilon_k$	$\sqrt{\frac{1}{n}\sum\limits_{k=1}^{n}\varepsilon_k^2}$	$\frac{1}{n}\sum\limits_{k=1}^{n}\varepsilon_k/\sigma_k$	$\frac{\eta^T \mathbf{V}_\eta^{-1}\eta}{n-1}$
Exponential	101.69	545.55	128768	0.807	13.072	0.00271	1.0000
Gaussian	139.73	252.71	51116	1.077	12.990	0.00377	0.9999
Spherical	107.31	297.84	125045	0.817	13.032	0.00273	1.0000

good agreement between the estimated semi-variograms and their sampling estimates, although no least-squares fit was used on the semi-variogram ordinates.

Table 1 shows the parameter values obtained as well as the classical Kriging statistics computed over the sample of the 82 cross-validation errors, namely the mean, the root mean square error, the re-scaled mean and the re-scaled variance; all the semi-variogram models reconstruct the spatial process with more or less the same statistical properties and variance. Therefore, the selection of the most appropriate model is, or may appear to be, reduced to a subjective choice.

In the approach proposed, the technique of estimating the parameters by fitting the semi-variogram to its sample estimates while using the variance of the cross-validation errors to select the model, has been reversed, leading to a more classical statistical technique. The model parameters are now estimated by applying ML to the distribution of the cross-validation errors, while the adequacy of the model must be checked on the divergence of the theoretical semi-variogram obtained from its sample estimate. In other words, the assumption on a semi-variogram model is followed by the parameter estimation on the basis of the cross-validation errors and by a final acceptance-rejection test on the chosen model with the estimated parameter values retained as the null hypothesis.

This gives rise to the need to define an appropriate statistical acceptance-rejection test for the semi-variogram to account both for the empirical semi-variogram sampling errors and for the uncertainty of the theoretical semi-variogram estimate induced by parameter estimation (Bogaert, 1999). Nevertheless the selection of the most appropriate model falls beyond the scope of the present paper. Once the model has been selected, it can be introduced directly into the Kriging equations to estimate mean annual precipitation of the area under study and to determine the estimation error variance.

Figures 6a, 6b and 6c show the interpolated surface of the adjusted rainfall for the three models, namely
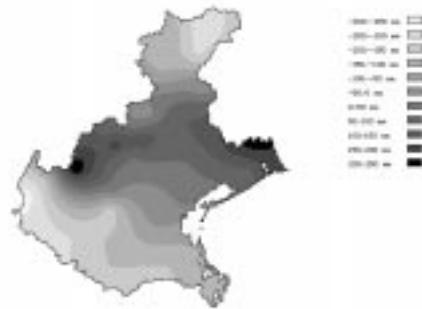

Fig. 6a. *The estimated spatial distribution of the adjusted rainfall in mm, using the exponential model*
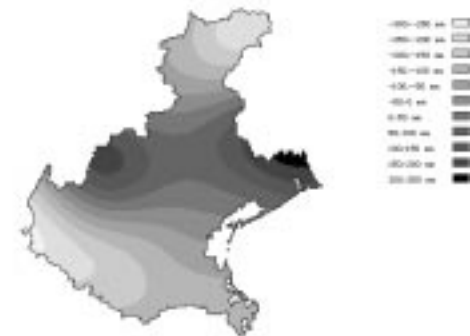

Fig. 6b. *The estimated spatial distribution of the adjusted rainfall in mm, using the Gaussian model*
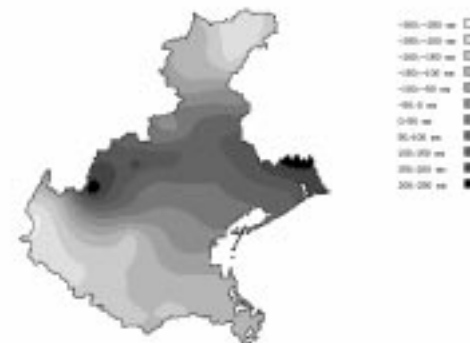

Fig. 6c. *The estimated spatial distribution of the adjusted rainfall in mm, using the spherical model*
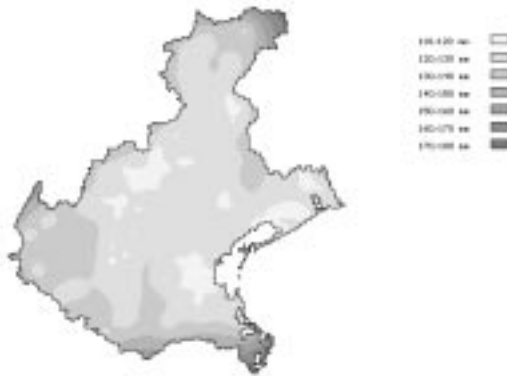
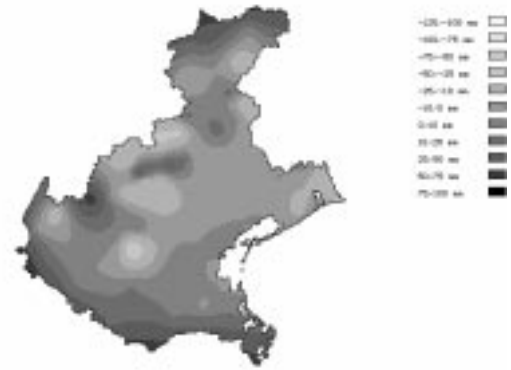Fig. 7a. *The standard deviation of the estimated adjusted rainfall in mm, using the exponential model*

Fig. 8a. *The bias of the estimated adjusted rainfall in mm, using the exponential model*

Fig. 7b. *The standard deviation of the estimated adjusted rainfall in mm, using the Gaussian model*
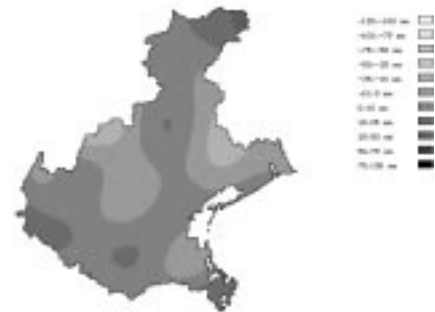

Fig. 8b. *The bias of the estimated adjusted rainfall in mm, using the Gaussian model*

Fig. 7c. *The standard deviation of the estimated adjusted rainfall in mm, using the spherical model*
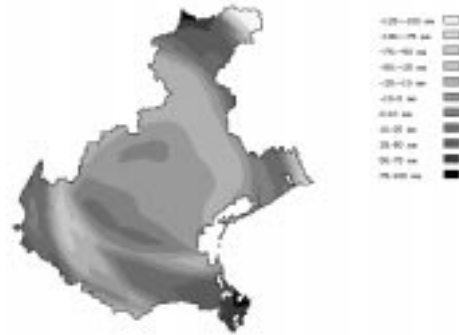

Fig. 8c. *The bias of the estimated adjusted rainfall in mm, using the spherical model*

exponential, Gaussian and modified spherical, while Figs. 7a, 7b and 7c show the relevant standard deviations.

Analysis of parameter estimation uncertainty

After determining the surface of the adjusted rainfall z^* and the standard deviation of their estimation error using the three different semi-variogram models, the analysis of the effect of parameter estimation uncertainty was carried out along the lines described in Part 1.

The bias corrections computed using the proposed methodology (Eqn. (13) in Todini, 2001) for the three models mentioned above are given in Figs. 8a, 8b and 8c, while the increase in the standard deviation of the estimation error (obtained by means of Eqn. (14) in Todini, 2001) is plotted in Figs. 9a, 9b and 9c. Finally the bias corrected adjusted rainfall values are given in Figs. 10a, 10b and 10c. To complement the areal spread information, Table 2 gives the minimum and maximum values obtained for the different quantities.

Prior to bias correction all the models give similar results

Fig. 9a. *The increase in standard deviation of the estimated adjusted rainfall in mm, using the exponential model*
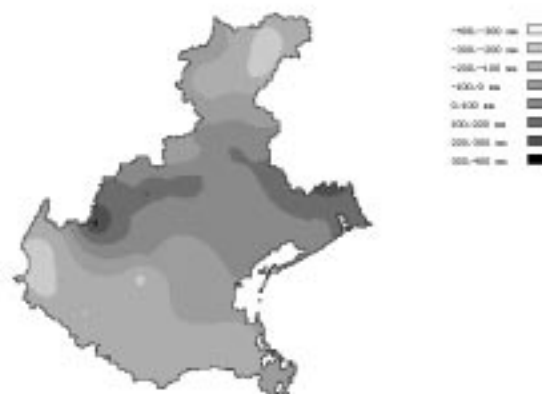

Fig. 10a. *The estimated spatial distribution of the bias corrected v alue of the adjusted rainfall in mm, using the exponential model*

Fig. 9b. *The increase in standard deviation of the estimated adjusted rainfall in mm, using the Gaussian model*
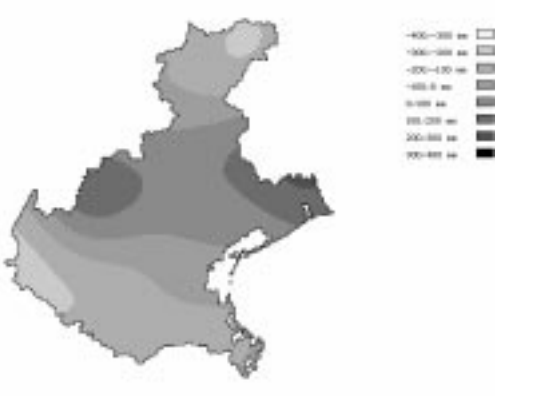

Fig. 10b. *The estimated spatial distribution of the bias corrected value of the adjusted rainfall in mm, using the Gaussian model*

Fig. 9c. *The increase in standard deviation of the estimated adjusted rainfall in mm, using the spherical model*
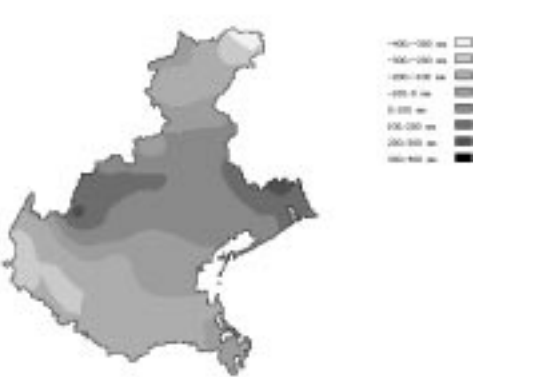
Fig. 10c. *The estimated spatial distribution of the bias corrected value of the adjusted rainfall in mm, using the modified spherical model*

Table 2. Minimum and maximum values of several statistics for the different models

Model	Exponential Min/Max [mm]	Gaussian Min/Max [mm]	Modified spherical Min/Max [mm]
Adjusted rainfall	-268/246	-267/230	-250/181
Std. Dev. of adjusted rainfall	115/181	124/178	116/176
Bias	-45/79	-19/29	-121/100
Increment of Std. Dev.	0/37	0/8.5	0/28
Bias corrected adjusted rainfall	-264/313	-247/220	-335/265

in terms of adjusted rainfall. Figures 6a, 6b and 6c show similar patterns, more smoothed in the case of the Gaussian model, with ranges from –268 to 246 mm for the exponential, from –267 to 230 mm for the Gaussian and from –250 to 245 mm for the spherical.

The distribution of the standard deviations is also very similar, with again more smoothness shown by the Gaussian model. The standard deviations vary from 115 to 181 mm for the exponential, from 120 to 178 for the Gaussian and from 115 to 170 mm for the spherical.

Figures 8a, 8b and 8c show that all the models are biased, as was anticipated in Part 1, and the bias tends to be larger on the edges because the variance is larger. The range of the bias of the Gaussian model is quite small (from –19 to 29 mm) when compared to that of the exponential (from –45 to 79 mm), and of the spherical (from –121 to 100 mm), (see Table 2), particularly if the scale of the bias is compared to the estimated deviations: the bias for the spherical model reaches almost half of the signal's standard error. An increase in standard deviation as calculated from Eqn. (14) of Todini (2001) is observed, but is relatively modest, particularly for the Gaussian model. The maximum increase is 37 mm for the exponential model, but most of this increase is smaller than 5 mm (Fig. 9a). The maximum is of 8.5 mm for the Gaussian (Fig. 9b) and reaches 28 mm for the spherical (Fig. 9c), but again in this last case, most of the increase is smaller than 2 mm.

Finally, because of the largely different spatial pattern of the bias (shown in Figs. 8a, b and c) the corrected adjusted rainfall values (shown in Figs. 10a, b and c) now differ from each other, with ranges of –264 to 313 mm for the exponential, of –247 to 220 mm for the Gaussian and of –335 to 265 mm for the spherical.

This paper does not deal with the model selection problem but, in the light of the results obtained , model selection is extremely important because of its strong conditioning of
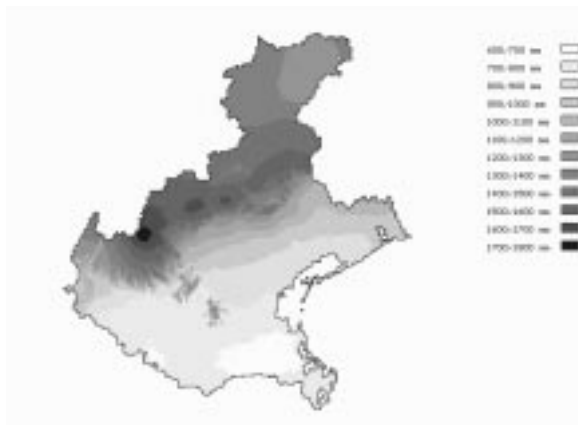

Fig. 11. *The estimated spatial distribution of the average yearly totals of precipitation over the Veneto region in Italy in mm, using the exponential model*
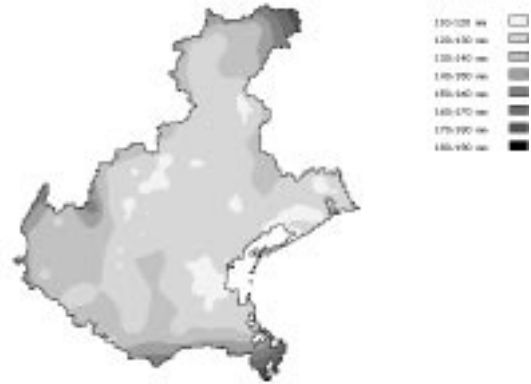

Fig. 12. *The estimated standard deviation of the average yearly totals of precipitation over the Veneto region in Italy in mm, using the exponential model*

the bias corrected results. To illustrate the final results and provide a basis for the numerical experiment, the exponential model was selected subjectively because it provided a larger bias than the Gaussian and the largest increase in standard error. Consequently, Fig. 11 shows the reconstructed field of yearly average rainfall over the Veneto region in Italy, with a standard deviation field given in Fig. 12.

The numerical experiment

As pointed out in Part 1, the Kriging estimates at a point can be demonstrated to be unbiased if the true model and its parameter values are known; if one has to choose the model and/or has to estimate the model parameters, this property does not hold as shown in the previous section, with a resulting small increase in the variance and mainly the introduction of a relatively large bias. While the effect of model selection, shown to be important in the Veneto Region example, will be dealt in a successive work, this paper focuses on the effect induced by parameter estimation.

To assess statistically the validity of the proposed bias correction, the following numerical experiment was set up. One hundred sets of 82 "adjusted rainfall observations" were extracted, at the same locations of the Veneto case study, from the following Gaussian process:

$$z = B\delta \qquad (7)$$

where δ is an [82] dimension NIP(0,1) and B the square root of the [82, 82] covariance matrix, defined as a function of h, the mutual distances among the 82 raingauges, reflecting the assumed exponential type semi-variogram, with the same set of parameters of Table 1 found for the Veneto case study:

Table 3. Likelihood Ratio Test Results for original and corrected cross-validation errors

H_0 : Mean = 0; Unknown Variance	$\frac{(100-82)}{82} m^T S^{-1} m$	$F_{82,18}(90\%) = 1.706162$	$F_{82,18}(95\%) = 1.990905$
Original Ordinary Kriging cross-val. errors	21.84675	REJECTED	REJECTED
Corrected Ordinary Kriging cross-val. errors	1.623827	ACCEPTED	ACCEPTED

$$Cov\{z(s), z(s+h)\} = \sigma^2 - \gamma(h) = \sigma^2 - \left[p + \omega \left(1 - e^{-h/a} \right) \right] \quad (8)$$

Matrix B is the symmetrical matrix obtained by the eigenvalue-eigenvector decomposition of the estimated $Cov\{z(s), z(s+h)\}$ assumed to be equal to B^2. Each set was then used for estimating the parameter values as well as the cross-validation errors. The latter were computed in two ways namely without and with the proposed correction given in Eqn. (13) of Part 1 (Todini, 2001), thus producing two samples of 82x100 cross-validation errors. Each of the two samples was then tested against the null hypothesis of deriving from a zero mean multivariate process with unknown covariance. Given that the cross-validation errors are not independent, as previously discussed, the Likelihood Ratio Test (LRT) was considered the most appropriate (Mardia *et al.*, 1979). Under the null hypothesis of zero mean and unknown variance (Hotelling one-sample T^2 test), the quantity $(n-1) m^T S^{-1} m$, with m the estimated sample mean, S the estimated sample covariance, n the sample size ($n = 100$ in the proposed experiment) is a $T^2(p,n-1)$ with $p = 82$ the multivariate space dimension and the quantity $\frac{(n-p)}{p} m^T S^{-1} m$ has a Fisher $F_{p,n-p}$ distribution. Table 3 shows the results of the test at the 90 and 95% levels based upon an $F_{82,18}$. As one can clearly observe, while the test on the "conventional" cross-validation sample is highly unlikely, there is no evidence to reject the null hypothesis either at the 10 or 5% levels, thus validating once more the proposed approach.

Conclusions

The theoretical approach to assess the effect of parameter estimation uncertainty in ordinary Kriging presented in Part 1, which was shown to produce a bias and an increase in estimation variance, was used to analyse the behaviour of three semi-variogram models by means of yearly precipitation data over the Veneto region in Italy.

The numerical results are coherent with what were anticipated in the theoretical development; namely that the parameter estimation uncertainty introduces a bias, which can be relatively large depending upon the non-linearity of the model used, and an increase in variance of the errors of estimate.

Finally, given the strong dependence of the bias and the increase in variance on the model used, it is felt that there is an urgent need for the development of an objective acceptance or rejection test on the basis of the ML estimated semi-variogram taking into account on the one hand the uncertainty induced by the parameter estimates and, on the other hand, the sampling distribution of the empirical semi-variogram estimate.

Acknowledgement

References

Bogaert, P., 1999. Assessing the variability of the variogram estimator. In: *geoENV II – Geostatistics for Environmental Applications*, J.Gòmez-Hernàndez, A. Soares and R. Froidevaux (Eds.), Kluwer Academic Publishers, Dordrecht, The Netherlands. 479–490.

Eastman, J.R., 1992. IDRISI, Clark University, Graduate School of Geography, Worchester, Massachussetts, 01610, USA.

de Marsily, G., 1986. *Quantitative hydrogeology*, Academic Press, London.

Kendall, M.G. and Stuart, A., 1973. *The advanced theory of statistics*, Vol. 2, Griffin.

Mardia, K.V., Kent, J.T. and Bibby, J.M., 1979. *Multivariate Analysis*, Academic Press, London.

Samper, F.J. and Neuman, S.P., 1989. Estimation of spatial covariance structures by adjoint state Maximum Likelihood cross validation. 1: Theory. *Water Resour. Res.*, **25**, 351–362; 2: Synthetic experiments. *Water Resour. Res.*, **25**, 363–371; 3: Application to hydrochemical and isotopic data. *Water Resour. Res.*, **25**, 373–384.

Todini, E. and Ferraresi, M., 1996. Influence of parameter estimation uncertainty in Kriging, *J. Hydrol.*, **175**, 555–566.

Todini, E., 2001. Influence of parameter estimation uncertainty in Kriging. Part 1 – Theoretical Development. Hydrol. Earth System Sci., **5**, 215–232.