Cahiers de Géostatistique, Fascicule 3
Compte–rendu des Journées de Géostatistique,
25–26 Mai 1993, Fontainebleau, pp. 189–199

DIRECT STRUCTURAL ANALYSIS OF AN IRF-k: CASE STUDIES

Roberto BRUNO, Giuseppe RASPA
Department of I.C.M.M.P.M.
Univ. of Rome "La Sapienza
Via Eudossiana, 18
00184 Rome, Italy

ABSTRACT. After a brief review of the technique for building the authorised measures of order k necessary for studying and directly modelling the Generalised Covariance, the results of applications to some case studies are presented. The classical structural analysis approach (automatic) is compared to the proposed direct modelling approach, basically observing, for several real data sets, the results of test-kriging and the quality of the maps obtained for the two approaches. Finally, the usefulness of getting both "tools" operating on the same package is stressed.

1. INTRODUCTION

Modelling of the Generalised Covariance (GC) is perhaps the trickiest part when one is working in the non-stationary field and chooses to apply the approach based on the Intrinsec Random Functions of order k (IRF-k). The technique developed by the Centre de Geostatique and currently the most used, involves the automatic recognition of the GC essentially by calculating a specific least squares regression starting from a pre-established number of elementary model functions of the GC [Renard, 89]. The selection, carried out on a ranking of the scores, is then verified using other indicators that help to verify the congruences between the rankings and hence the robustness of the selection. Further, a test-kriging can be performed to make a "real" verification of the performances of the different models.

However, this procedure has always been somewhat controversial, mainly with regard to its limitation on the role of the operator who simply selects a model from the set identified by the programme. In reality, even though the number of elementary models (nugget, linear, cubic, spline) is limited *a priori*, modelling a GC

involves identifying: a) the model of the GC (i.e. a combination of elementary models); then b) the coefficients of that model. And these two phases are not always perfectly disjoined and consecutive. With the current procedure, it is not the operator who identifies the type of GC model, nor does he estimate the values of the coefficients of the model. He simply validates this or that model which the computer has fully identified on the basis of the code and of the geometrical parameters given (i.e. shape and dimension of neighbourhood).

The use of the automatic procedure for identifying the covariance model has always been a subject of discussion between the supporters of the automatic and the interactive approaches, respectively. In stationary geostatistics, where the use of both approaches is well consolidated, classical modelling of the mono- and bi-variate cases normally involves the calculation and visualisation on graphics of a set of experimental variograms from which the operator identifies and defines the model variogram. The operator is of course in full control. Automatic modelling seems to be more efficient and is therefore to be preferred for multi-variate cases where, for example, it is necessary (or useful) to adopt a linear model, and a whole series of congruences are needed to be taken into account contemporaneously.

In non-stationary geostatistics, mono-variable applications have been the rule to date. However, automatic modelling has never been seriously queried nor had any alternative general techniques been proposed until the authors proposed direct modelling at IV GEOSTAT [Bruno and Raspa, 1993]. This paper describes the application of direct modelling of the GC to two case studies, illustrating the value of its contribution to the identification of the GC.

2. REVIEW OF THE TECHNIQUE

The main difficulty encountered in direct modelling of a GC has to do with the impossibility of inferring the GC experimentally. In practice, only the variances of the generalised increments are accessible, and these are estimates not of the GC but rather of a function of the GC. With an order $k>0$ and an IRF-k defined in plan, more than two points are needed in order to construct the generalised increment $Z(\lambda) = \Sigma_\alpha \lambda_\alpha Z_\alpha$, where the generic measure (i.e. the set of weights λ_α) must satisfy specific conditions (at least three) for filtering the polynomial terms of a trend of the order $k>0$:

$$\Sigma_\alpha \lambda_\alpha x_\alpha^l = 0 \qquad l=0,k$$

where x_α represents the generic coordinate of the data in the working space. In this case each experimental measure therefore has several, at least three, distances, $h_{\alpha\beta}$, and the bi-univocal link (between the shifting vector and the spatial correlation function) that is exploited, for example, for modelling the stationary variogram, is immediately lost.

The proposed direct modelling technique consists in parametrising the

variances of the increments with a view to identifying those measures that guarantee a pre-established evolution of that variance as a function of a single parameter. In particular, evolutions that substantially coincide with the elementary models were adopted in order to verify whether or not the real GC corresponds to the adopted models (linear, nugget, etc.).

Having identified the order k in the traditional manner, the authorised measures for the order k that ensure a pre-established parametric evolution can be obtained by resolving the following sort of non-linear systems:

$$\begin{aligned} &\Sigma_\alpha \lambda_\alpha x_\alpha^{\,l} = 0 && l=0,k \\ &\Sigma_\alpha \Sigma_\beta \lambda_\alpha \lambda_\beta f_m(h_{\alpha\beta}) = g_m(p) && m=0,1,3,s \end{aligned} \qquad (1)$$

where $f_m()$ represents the elementary model functions of the GC being studied, and $g_m(p)$ represents the corresponding functions of the parameter p. The $g_m()$ do not always coincide with the $f_m()$ for reasons of congruence, but they always respect the latter's behaviour [Bruno and Raspa, 93].

Viewed objectively, therefore, the operator does not identify the GC model on the basis of a straightforward graphical analysis, even with this technique. In reality, he verifies whether or not the pre-selected elementary functions clearly describe the behaviour of the calculated experimental variances. If this behaviour is respected, he goes on to identify the coefficients of the model thus identified; if the behaviour is not respected, he tries again using other elementary functions.

The advantage of the technique at this stage is that it adds depth to a verification of the acceptability of a given model by means of a graphical analysis of the behaviour of the variance as a function of a given parameter. The calculated variances correspond, in practice, to different types of increments; but, more simply, it can be stated that the parameter and hence the evolution of the variances are linked to the average size of the increments themselves.

On the other hand, the operator's contribution is immediate and direct during the identification of the coefficients of the model, when the values of the intercept or the angle coefficient of the model function are selected, one by one, as done in the most classic of stationary structural analyses.

3. DETAILS REGARDING IMPLEMENTATION

For reasons linked to the numerical solution of non-linear systems, it was decided to verify pairs of elementary models of the GC, also because only six pairs are possible. This limitation does not seem to have major practical drawbacks as it is fairly rare to have to use a GC model with more than two elementary models.

To guarantee that a range of values is obtained for the experimental variances and hence avoid the risk of having values all bunched around a narrow interval of values for the parameter, the data needed to construct the increments can be drawn

from two concentric areas, in a manner quite similar to that adopted for the classical procedure for identifying the GC. The parameter p is imposed in the system (1), the solution of which (if it exists) will provide the values for the weights λ_α. Several solutions are generated for a same set of data by adopting different values for p selected within an interval that has been sampled using two steps: a longer one to detect the existance of a solution and a shorter one to generate a range of consistent existing solutions.

Having obtained the measures $\{\lambda_\alpha\}$, the squares of the increments, $Z(\lambda)^2$, are calculated, obtaining what can be considered a first estimate of their experimental variances. Each experimental value corresponds to a parametric value, whose function, applied to the system (1), represents the theoretical variance.

The corresponding graphic is a cloud of points for the variances calculated as a function of the parameter p (Fig. 1a). As it may be difficult to identify a behaviour and model of the GC, or even to estimate the coefficients using this graphic, it is necessary to analyse the mean values of the experimental variances over intervals of value of the parameter p (Fig. 1b).

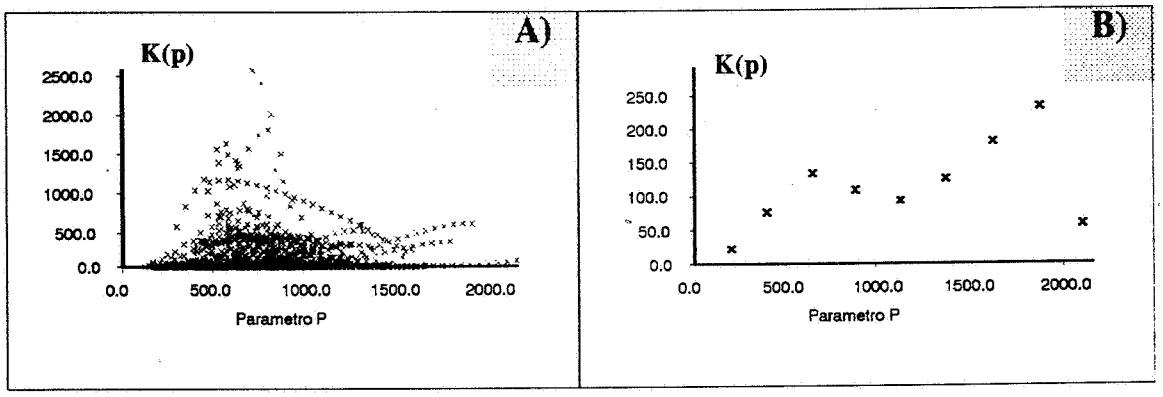

Fig.1a - Cloud of the squares of the increments as a function of the parameter p.

Fig.1b - Diagram of the estimated variance of the increments *versus* the parameter p.

This operation is quite similar to the one applied, for example, to the variograms. In that case, however, the interval is on the step (h) and the measure (λ_α) remains constant (+1,-1), whereas in the present case, it is the value of the parameter that remains constant because one is taking the average of increments that have several values of distances ($h_{\alpha\beta}$) and different measures (λ_α). The procedure of estimating the theoretical value of the variance by this average value seems to be practically correct.

4. THE TOP OF AN OIL RESERVOIR

This case study focuses on the characterisation of the top of an oil reservoir where the stratigraphic level has been explored by means of over 100 drillholes (Fig. 2).

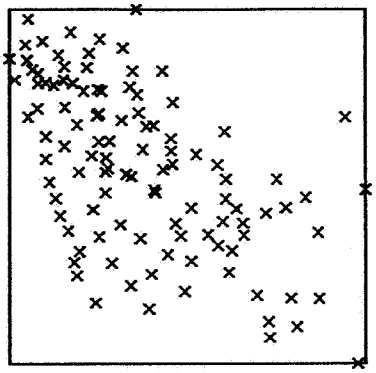

Fig.2 - Locations of the drillholes that intercept the top of the oil reservoir.

Classical structural analysis is strongly influenced by the working scale: e.g. a reconnaissance area of 6000x6000 m is a dimension that allows both the contemporaneous processing of a consistent number of data items (50) using the FAIPACK package [Bruno and Raspa, 91] on the one hand and the local investigation of the whole surface using an adequate number of elementary areas on the other hand. In this case, automatic recognition identified an order k = 1 for all the calculation options, namely:

- with or without 50% overlapping of the elementary recognition areas;
- with or without selection by octants of the information.

The use of a smaller scale, 4000x4000 m, increases the level of local detail while still guaranteeing a satisfactory density of information with regard to the whole surface. In this case, an order k = 0 is just as valid if not more valid than an order k = 1.

For neighbourhood sizes greater than 6000x6000 m, k = 2 would probably be more suitable, vice-versa, neighbourohod sizes smaller than 4000x4000 m could justify the use of an order k = 0, but the amount of information per neighbourhood, in both cases, would prevent consistent quality calculations. Anyhow choice of working scale is closely linked to the purpose of the analysis. Considering that the objective in our case was simply to generate a spatial reconstruction of the reservoir top by kriging, the choice was conditioned by the size of the estimating neighbourhood: order k and GC had to be identified on elementary areas that were congruent with the downstream working scale.

In reality, both the sizes proposed above would have been valid since a large neighbourhood makes it possible to consider a higher number of data items and hence, in theory, provides greater precision, whereas the need for a globally coherent model, calling for the use of a higher order k, results in less precision of the different inferences, from the model of the GC down to the subsequent estimates.

In this case, direct recognition of the IRF-k was found to be helpful in choosing between the alternative models of the GC. In fact, with neighbourhood of 6x6 km^2 and an order k = 1, the best model of the GC generated by the automatic structural recognition procedure was a pure spline model, with coefficients varying

between 7 and $10x10^{-5}$ as a function of the activated options. The jacknife presented consistent values, although they were quite similar to those of the models of the GC ranked next to it. The calculations were made with and without selection by octants. A cubic and a linear model, respectively, were ranked in second position, whereas a nugget + spline and a pure cubic model were ranked in third position (Fig. 3).

MODEL	B0	B1	B3	Bs	Jacknife
1	0.000000	0.000000	0.000000	8.4221e-05	1.259467
2	0.000000	0.000000	3.11016e-08	0.000000	1.313214
7	44.771564	0.000000	0.000000	5.5797e-05	1.393335

A) Processing **without** selection by octant

MODEL	B0	B1	B3	Bs	Jacknife
1	0.000000	0.000000	0.000000	1.0541e-04	0.997354
3	0.000000	-0.230219	0.000000	0.000000	1.039165
2	0.000000	0.000000	3.6634e-08	0.000000	1.123639

B) Processing **with** selection by octant

Fig.3 - Models of the Generalised Covariance identified automatically.

Summary analysis shows that all the solutions are acceptable in terms of absolute values of scores, and that the spline is the most frequently present model, followed by the cubic then by the linear and nugget effect ones. Test-kriging reverses these evaluations, showing the linear model to have the best performances, with the lowest variance of estimating error, the most congruent mean square standardised error (MSSE).

With an elementary area of 4000m x 4000m and keeping the order $k=1$, the ranking of the identified models of the GC was as follows: cubic, cubic + spline, linear + cubic + spline, confirming the trends identified by structural analysis at the larger scale. But in this case, test-kriging provided results that were globally less satisfactory. With an order $k=0$, the classical stationary analysis reveals: a) a clear trend in the direction of 90 degrees, although its numerical influence on the experimental variograms did not seem excessively strong; b) a variogram that tended to be linear although it may have anisotropies, probably linked to the trend and with evidence of sills in some directions without trends. In this case the results of test-kriging were comparable to those obtained with the order $k=1$ and the linear model with neighbourhoods of 6000m x 6000m.

A solution to the problem of how to identify the GC is, therefore, neither immediate nor univocal. First of all, although it is clear that the regionalised variable is studied at a transition scale, the selection of a smaller order k (as advisable where there is uncertainty) would be pragmatically acceptable but

difficult to justify both geologically and geostatistically. Further, it is necessary to choose the selection criterion separately with respect to the working scale (and to the order k considered). If preference is given to the ranking criterion, there is no doubt but that the spline model is the one that provides the best performances, followed by the cubic model. If preference is given to the results of test-kriging, the linear model seems to provide the best assurances of efficiency. And the above is true regardless the analysis of the maps of the top generated by applying the different models.

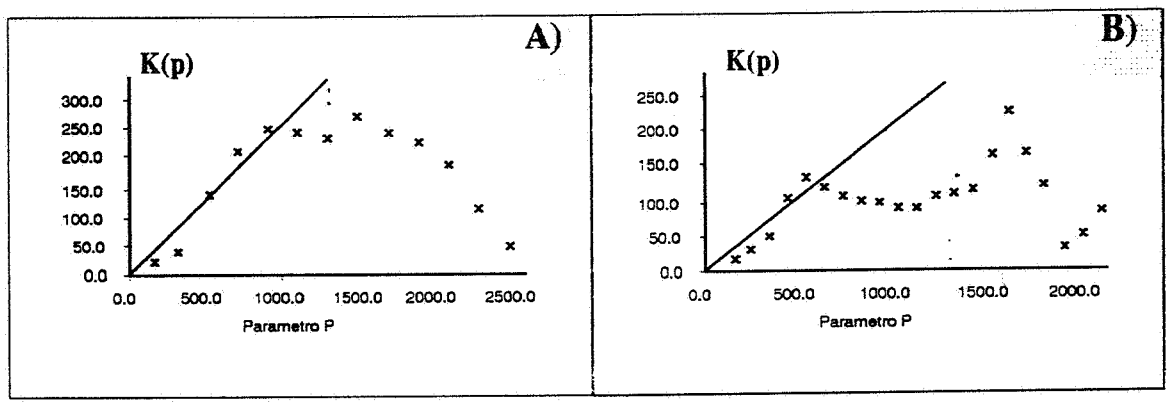

Fig.4a - Direct modelling of the Generalised Covariance: order k = 0, B1 = -0.25.

Fig.4b - Direct modelling of the Generalised Covariance: order k = 1, B1 = -0.20.

In this case, direct structural analysis of the IRF-k would seem to be a useful tool for decision taking. In fact, analysis of the behaviour of the experimental variances as a function of the parameter p helps to understand the spatial variability: it always shows a linear growth at small scales, then stopping and keeping constant for larger values of the parameter. This regardless of which order k is adopted and which calculating modality is pre-selected. Estimating the coefficient of the linear component of the GC is as easy as for a stationary variogram: it is close to -0.25 when k = 0 and -0.20 when k = 1 (Figures 4a and 4b). Verification by test-kriging confirmed the adequacy of the model in both cases and the results compare well with the best models identified by means of the automatic procedure.

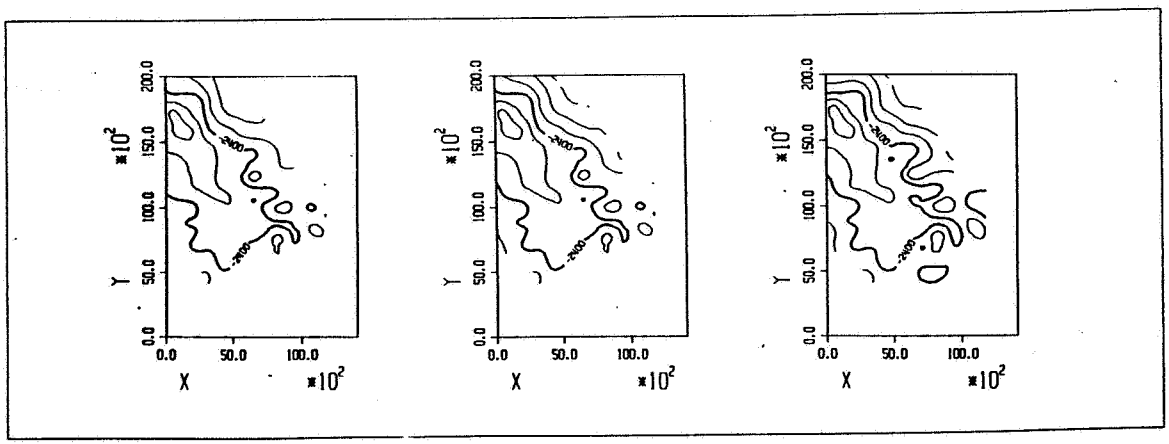

Fig.5 - Contour line of the top of the reservoir: a) order k = 0, linear GC; b) order k = 1, linear GC; c) order k = 1, spline GC.

Figures 5 a-c present the maps generated on the basis of the orders k = 0 and k = 1 with neghbourhood of 6x6km^2 and, respectively, by means of the linear models confirmed by direct structural analysis and by means of the spline model identified by the classical structural analysis.

5. A REGIONAL WATER TABLE

The object of this case study focused on the average piezometrics of a water table in the provinces of Parma and Piacenza (Italy), that has been investigated by means of over 80 wells distributed over an area of about 4,500 km^2 (Fig. 6).

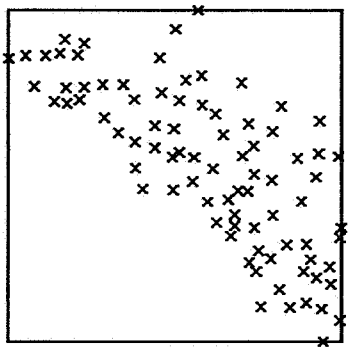

Fig.6 - Locations of the wells from which piezometric data on the water table were taken.

Analysis of the order k, regardless of which neighbourhood size was adopted, clearly identified an order k = 2. Recognition of the GC for a neighbourhood size of 30x50 km^2 generated the following ranking of models: linear, linear + spline, nugget + spline. But test-kriging reversed the scale of values by attributing the best performances to the nugget + spline model. However, whereas the jacknife of the automatic recognition degenerates rapidly between the best model and the next best one, the experimental variances of the estimating error generated by applying the one or the other model did not vary consistently (Fig. 7).

MODEL	B0	B1	B3	Bs	Jacknife
3	0.000000	-7.189630	0.000000	0.000000	0.871668
4	0.000000	-2.735236	0.000000	0.736495	0.633367
8	3.357861	0.000000	0.000000	0.975136	0.616054

A) Automatic identification and jacknife ranking of first three G.C. models.

TESTK - Prder83.dat

MODEL	ND	M.S.E.	M.S.S.E.	M.E.
3	84	34.153957	0.916279	0.409774
4	84	31.698935	0.8966430	0.316807
8	84	31.550737	0.902285	0.312345

B) Test-kriging characterization of the best three G.C. models.

Fig.7 - Characterisation of the models of the Generalised Covariance identified automatically.

Then, the contribution of information given by direct recognition was verified.

First, it is noteworthy that no linear model was identified and that it is only with some difficulty that a linear + cubic model could be modelled. A better fitting was obtained using a nugget component, as for the nugget + cubic and nugget + spline pairs. The latter seems to be more coherent, at least at the scales investigated by means of direct analysis (Fig. 8 a,b,c), but this nugget + spline model does not have the same coefficients as the model that had been automatically defined.

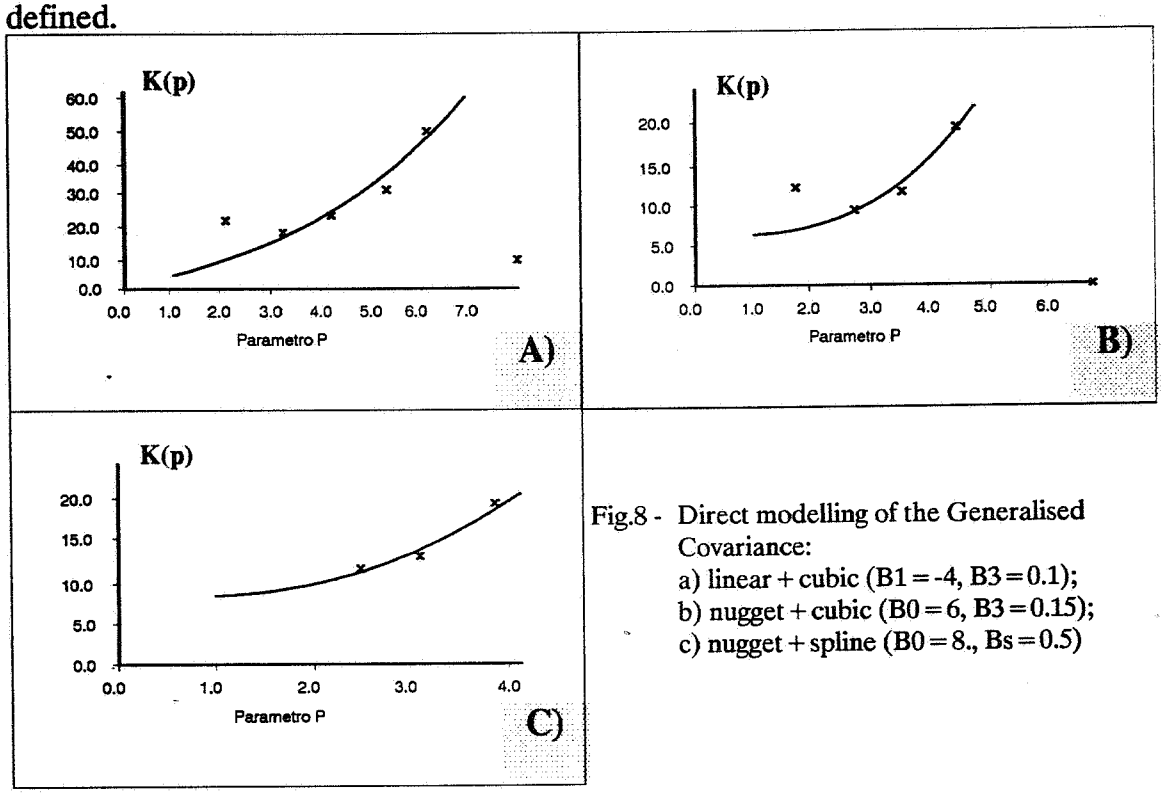

Fig.8 - Direct modelling of the Generalised Covariance:
a) linear + cubic (B1 = -4, B3 = 0.1);
b) nugget + cubic (B0 = 6, B3 = 0.15);
c) nugget + spline (B0 = 8., Bs = 0.5)

Confirmation by test-kriging was again sought but the results obtained for the three interactively identified cases did not significantly improve the quality of the estimates although the results were comparable to the best results obtained by automatic recognition. The three models identified by direct analysis seem to behave in a manner that is practically equivalent, with only very small variations in the control parameters between one model and another (Fig. 9).

TESTK - Prder83.dat

MODEL	ND	M.S.E.	M.S.S.E.	M.E.
$B_1 + B_3$	84	31.597124	0.895402	0.269147
$B_0 + B_3$	84	32.277241	0.960806	0.245460
$B_0 + B_s$	84	32.375854	1.048095	0.358999

Fig.9 - Results of test-kriging on the three GC models identified by direct analysis.

This substantial equivalence between the models is coherent with the findings of an analysis of the results of automatic recognition carried out at a larger scale, with neighbourhoods of 45x50 km^2 with 50% overlapping. Nine models were identified within a very small field of values of the jacknife, all of them with good levels (0.89-0.92). It is seen that the first two positions on the ranking are occupied by two elementary models, namely the spline and the linear. These models had not been identified by direct analysis and they had generated the "worst" results during test-kriging. It is also seen that the results obtained by test-kriging using the models identified by direct inference are in line with the best obtained by means of automatic recognition for these large neighbourhoods.

In conclusion, the large body of information in hand for this case study coupled to the specific spatial variability of the water table assure a good quality of the results of the recognition of the GC, regardless of which technique is adopted. However, the results of direct analysis seem to be more stable and of a higher average quality. Further, direct analysis seems capable of rejecting immediately those models which may perhaps be easy to infer automatically but which are less coherent and efficient in application.

Figure 10 presents the maps of the results obtained using the different models, confirming the substantial equivalence of the possible choices.

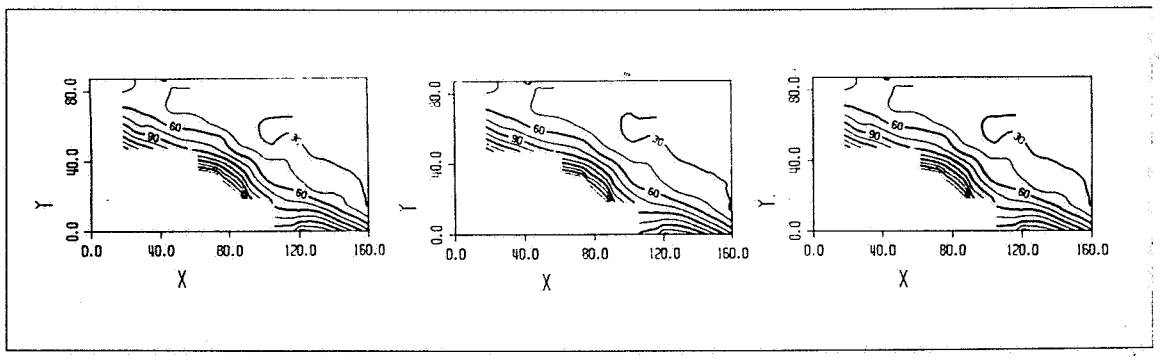

Fig.10 - Contour line of the piezometric of the water table for different models of the Generalised Covariance: a) direct analysis, linear + cubic; b) direct analysis, nugget + spline; c) automatic analysis, nugget + spline.

CONCLUSION

The direct recognition of Generalised Covariance has been tested on two case studies with a relatively large set of data, but still with some problems of modelling the spatial variability.

The proposed technique for direct recognition of the GC is an interesting analytical tool that is undoubtedly useful as a complement to the classical automatic technique. In the future, it may become a valid alternative for the identification of the GC, but further research is required to fully verify the conditions of applicability and reliability, especially under extreme conditions.

LIST OF REFERENCES

1) Renard, D. (1989), "Automatic structure recognition", in Geostatistics, Vol. II, Kluwer Academic Publishers, Dordrecht, pp. 579-590.

2) Bruno, R. and Raspa, G. (1993), "Towards a Direct Structural Analysis of an IRF-k", in Geostatistics Troia '92, Vol. I, Kluwer Academic Publishers, Dordrecht, pp. 49-60.

3) Bruno, R. and Raspa, G. (1991), Manual FAIPACK, Vers. 1.2, Vol. I and II, Dip. ICMMPM, Univ. of Rome "La Sapienza", Rome.