Reliability enhancement of groundwater estimations

Zoltán Zsolt Fehér[1,2], János Rakonczai[1], [1]Institute of Geoscience, University of Szeged, H-6722 Szeged, Hungary, [2]e-mail: zzfeher@geo.u-szeged.hu

Abstract: The Danube Tisza Interflow, located in the middle part of Hungary, was afflicted by a serious groundwater level (abbr.: GWL) reduction. This study aims to analyse the spatiotemporal progress of the change taking care with problems that impairing the precision of spatial extrapolation.

After preliminary analyses of September GWL datasets, 100 realizations of a sequential Gaussian co-simulation were created applying the Markov II type simple cokriging estimator. Since GWL roughly follows the shape of the relief, DEM-1000 is used as auxiliary data.

The expected type grids of e-type estimations can be applied to (1) analyse the trend of absolute groundwater deficit volume, (2) make distinction between areas behaving differently, and (3) continue complex hydrological analyses in the future.

Key words: geostatistics, uncertainty, sequential simulation, auxiliary data, groundwater.

1. INTRODUCTION

Basic thesis in hydrology that multi-year average of the groundwater level (GWL) doesn't show significant changes unless the hydrological elements changes drastically or serious external impacts are taken. The volume, annual distribution and intensity change of precipitation as well as anthropogenic interventions have led to a somewhere excess than 10 meter GWL reduction on the Danube-Tisza interflow, central Hungary.

Several models and studies was dealt with the water level reduction, but without the consideration of spatial relationships. Nevertheless, groundwater flows and transport systems are affected by heterogeneity, namely that the spatial continuity of the property changes with distance and direction.

Geostatistical approaches are designed to handle spatial characteristics of the site using variograms.

2. METHODS

2.1. Multi Gaussian Approach

Kriging as interpolation method is a trivial way to evaluate the GWL with the consideration of spatial heterogeneity. The results depend only on the geometric structures of the property. The kriged results the best local linear estimation of the expected property in the sense of smallest squared differences, however it has several disadvantages. (1) Kriging behaves as low pass filter, thereby can smooth some small details of the GWL. (2) Estimation error depends only on the data configuration and do not on the input values. (3) Only kriging deviation maps can be created, since estimation error is not quantifiable in the absence of exact water level observation at any grid node.

The results of applied sequential Gaussian simulation give more detailed images (realizations) of the GWL. That is, this method can express the uncertainty of estimations at any grid node. Instead of using static data structure, multiple equiprobable realizations are generated on random data point structures, by continuous inclusion of previously simulated grid nodes. As result of the grid node estimations both the mean and standard deviation is known, but rather than choose the mean at each grid node, SGS takes a random value from this distribution along with its probability level (Deutsch and Journel, 1998; Journel, 1993).

Related grid node estimations of generated realizations set up a statistical distribution. These distributions then determine grid by grid the expected value and standard deviation of the GWL (Deutsch and Journel, 1998).

Preliminary studies (Pálfai, 1994) stated a very strong relationship between topography and water reduction therefore an elevation model as secondary information may improve estimation results.

Various methods - kriging with locally varying mean (Goovaerts, 1997), cokriging (Isaaks and Srivastava, 1989), regression kriging (Ahmed and De

Marsily, 1987) were developed for incorporating dense secondary information. Among these methods in this study the simple collocated cokriging approach is compared in interpolation and simulation cases.

Simple cokriging and simple Gaussian co-simulation were performed using dense topography as auxiliary data and GWL presented the primary variable. The *NEWCOKB3D* program (Xianlin and Journel, 1999) is used for cokriging and *SGSIM_FC* (Journel and Shmaryan 1999) to perform co-simulations. Both application is open source and available on the homepage of Stanford Centre for Reservoir Forecasting.

The applied methods in this study is based on simple kriging (Deutsch and Journel, 1998), with the assumption that the trend component is a constant known mean**.** The kriging equation system is expressed as

$$C * w = D, \qquad (1)$$

where w is the kriging weight matrix to be solved, C is the covariance matrix between all measured points and D for the covariance matrix for the estimated grid node versus any measured points. All covariances are calculated based on the variogram model of the property (Isaaks and Srivastava, 1989).

2.2. Determination of cokriging weights

The *classical Simple Cokriging* approach (Isaaks and Srivastava, 1989) uses the Linear Model of Coregionalization (LMC) namely the above mentioned C and D matrices are extended with auxiliary and cross covariates. Thus not only the primary data variogram but all auxiliary data variograms and all cross variograms are necessary to determine the covariances.

The difficulty of LMC is, that all direct $\gamma_{ii}(h)$ and cross $\gamma_{ij}(h)$ variogram structures (1) has to be modelled simultaneously using the same variogram structures and (2) must be adequate to all experimental, direct and cross variograms to guarantee positive definiteness and proper regionalized correlation coefficients (Isaaks and Srivastava, 1989).

Special case of cokriging is a so called *collocated cokriging* while secondary data is known (or previously interpolated) at each node to be estimated. For a situation of strictly collocated cokriging two solutions were developed to ease variogram modelling (Journel, 1999; Xianlin and Journel, 1999; Shmaryan and Journel, 1999).

In Markov Model I. (MM 1) the cross variogram is derived from the primary variogram while in Markov Model II. (MM2) the cross covariance is modelled proportional to the secondary variable variogram (Journel, 1999).

In this study MM2 was preferable, since (1) topography is more densely sampled; therefore its experimental variogram is more reliable than of the GWL data, and (2) Markov Model 1 is based on the Markov filtering hypothesis (Xianlin and Journel, 1999) in which the conditioning of the auxiliary variable by the collocated primary datum filters the influence of farther primary data. This thesis can't be adapted in our case, since topography data is more densely available than groundwater observations.

In case of MM2 the variances and covariances are given by datasets, only the topography variogram and the correlation residual variogram of the GWL is need to be modelled.

The cross variogram can be calculated using the following expression:

$$\gamma_{12}(h) = \sqrt{\frac{C_{11}(0)}{C_{22}(0)}} * \rho_{12}(0) * \gamma_{22}(h) = \alpha * \gamma_{22}(h) \qquad (2)$$

where $\gamma_{12}(h)$ is proportional to the direct variogram $\gamma_{22}(h)$ of the topography data with a proportionality coefficient that includes the correlation coefficient $\rho_{12}(0)$ and the variances of GWL $C_{11}(0)$ and topography $C_{22}(0)$ variable (Shmaryan and Journel, 1999).

The primary variogram covariance function is expressed by the sum of the covariance and covariance residuals (**Eq. 3.**) between the GWL and topography data. Since the secondary variogram is given, only the residual variogram $\hat{\gamma}_R(h)$ has to be calculated. The residual variogram can be modeled; (1) using the experimental variogram of the normalized regression residuals for the GWL; or (2) by analytical approaches. (Xianlin and Journel, 1999)

$$\gamma_{11}(h) = \alpha^2 * \gamma_{22}(h) + \hat{\gamma}_R(h) \qquad (3)$$

where $\gamma_{11}(h)$ is the primary variogram $\hat{\gamma}_R(h)$ is the residual variogram and α is the angular coefficient of the linear regression curve.

2.3. Procedure of the water level estimation

This procedure includes:

1. Transform GWL and topography datasets to normal scores;
2. Determine linear regression residuals and correlation coefficients between GWL and well elevation data;
3. Compute variograms from normal scores for secondary $\gamma_{22}(h)$ and regression residual $\hat{\gamma}_R(h)$;
4. Create 100 sequential Gaussian co-simulation realizations and perform simple collocated cokriging (**Figure 2d.**) of the GWL using the topography and residual variogram model, along with the correlation coefficient;
5. Back transform the results to original dimensions;
6. Produce expected type estimations (**Figure 2c.**)

3. RESULTS

Given September monthly average GWL time series (above sea level) of 213 observations between 1976 and. These wells represent the GWL fluctuation across the study area. In addition, given a 1000-meter resolution DEM. (**Figure 1.**)

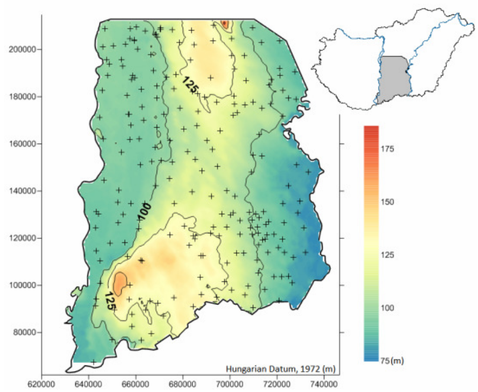

Figure 1: *Well distribution and topography (m.s.l.) of the study area*

The only well that shows outliers systematically was left in the dataset since according to Goovaerts (1997) in environmental applications large values may indicate potentially critical points so they should be removed only if they are clearly proved to be wrong.

Exploratory analyses of annual cross-selections established that each annual dataset (1) differs significantly from any theoretical distribution; (2) correlates strongly with the others; (3) the corresponding experimental variograms are quasy similar; (4) correlates significantly with well elevations.

A significant difference was revealed between the experimental variogram surface of well elevations and a more densely sampled topography data. This statement along with the strong correlation between GWL data and elevation reveals the unreliability of the existing observation network.

As presented before the direct variogram model for DEM (**Eq. 4.**), the correlation coefficient and the variogram model of regression residuals (**Eq. 5.**) are required for Markov Model 2 to determine the primary and cross variograms.

$$\gamma_{22}(\boldsymbol{h}) = 0.1 + 0.53 * Sph\left(\frac{177°}{32000m}\frac{87°}{9697m}\right) + 0.46 * Sph\left(\frac{29°}{31360m}\frac{119°}{9224m}\right) \quad (4)$$

$$\hat{\gamma}_R(\boldsymbol{h}) = 0.38 + 0.46 * Sph(9350m) + 0.16 * Sph(32300m) \quad (5)$$

Contour maps of different approaches show that sequential Gaussian simulation (**Figure 2a.**) whereas highlights local details, it is still significantly determined by the groundwater variogram model applied. Examining on large scale, even application of Ordinary Kriging (**Figure 2b.**) may give better results on the water level but without the consideration of elevation information. Simple Cokriging overestimated and still smoothed the water level but now rough contours of the elevation can be observed (**Figure 2d.**). Only sequential co-simulation seems to honour the reality of the inputs (**Figure 2c.**) while results better local estimates than Simple Cokriging (**Figure 3.**).

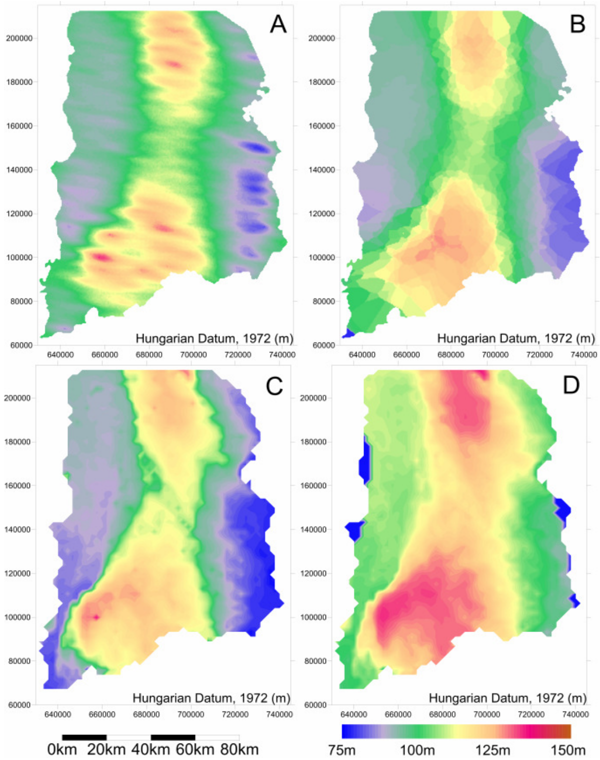

Figure 2: *Results for sequential Gaussian simulation (A), Ordinary Kriging (B), co-simulation with MM2 (C) and Simple Cokriging with MM2 (D)*

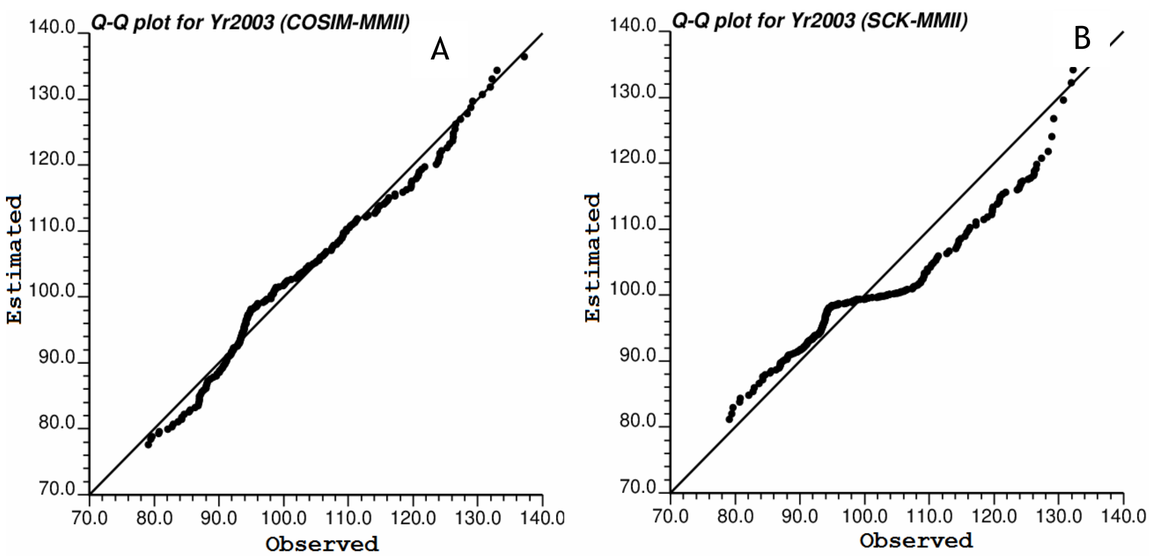

Figure 3: *Comparison of expected vs. observed grid node values for Gaussian Co-simulation with Markov Model 2 (A) and Cokriging with Markov Model 2 (B)*

Analysis of trends revealed, that neglecting secondary information could lead to overestimation of groundwater change. Furthermore results of simple cokriging drew the attention to the risk of acceptance any singular estimation of a property (**Figure 4.**).

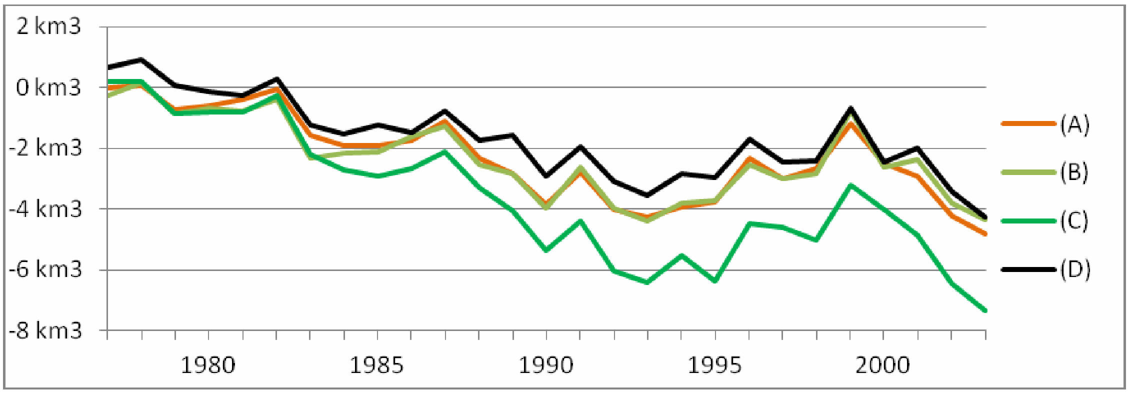

Figure 4: *Calculation of the groundwater volume reduction (km3) using Sequential Gaussian simulation (A), Ordinary Kriging (B), Simple Cokriging (C) and sequential Gaussian co-simulation with Markov Model 2 (D) on the study area*

4. REFERENCES

AHMED, S. & DE MARSILY G. (1987): Comparison of geostatistical methods for estimating transmissivity using data on transmissivity and specific capacity. Water Resources Research 23, 9, 1717-1737.

DEUTSCH, C. V. & JOURNEL, A. (1998): GSLIB, Geostatistical Software Library and User's Guide, Oxford University Press, New York, 340p.

GOOVAERTS, P. (1997): Geostatistics for Natural Resources Evaluation. Oxford University Press, New York, 483 p.

ISAAKS, E. H. & SRIVASTAVA M. R. (1989): An Introduction to Applied Geostatistics, Oxford University Press, New York, 561p.

JOURNEL, A. (1999): Markov Models for Cross-Covariances. Mathematical Geology, 31, 8, 955-964.

PÁLFAI I. (1994): Water Management Problems of the Danube-Tisza Interflow (in Hungarian), Nagyalföld Alapítvány kötetei, Békéscsaba, 126p.

SHMARYAN, L. E. & JOURNEL, A. (1999): Two Markov Models and their Application. Mathematical Geology, 31, 8, 965-988.

XIANLIN, M. & JOURNEL, A. (1999): An expanded GSLIB Cokriging Program Allowing for two Markov Models, Computers and Geosciences, 25, 6, 627-639.