Estimation of hydrogeological parameters for groundwater modelling with fuzzy geostatistics: closer to nature?

JAN A. PIOTROWSKI
Institute of Geology and Palaeontology, University of Kiel, Olshausenstrasse 40-60, D-24118 Kiel, Germany

FRANK BARTELS
Institute of Informatics, University of Kiel, Olshausenstrasse 40-60, D-24118 Kiel, Germany

ARKADIUSZ SALSKI
Ecosystem Research Centre, University of Kiel, Schauenburgerstrasse 112, D-24118 Kiel, Germany

GUNTHER SCHMIDT
Institute of Geology and Palaeontology, University of Kiel, Olshausenstrasse 40-60, D-24118 Kiel, Germany

Abstract One uncertainty in groundwater flow modelling is regionalization of hydrogeological parameters which are needed to define boundary conditions for the model area. In many instances, due to scarce and unevenly distributed boreholes, coupled with high heterogeneity of the geological composition, no standard regionalization techniques such as kriging can be reliably implemented. Here, we demonstrate an application of the fuzzy kriging method in regionalization of hydraulic conductivities ($K$) in a 1700 km$^2$ large area in northwestern Germany, in which the set of conventional, crisp parameters obtained from boreholes is supplemented by imprecise (fuzzy) information subjectively estimated by experts. As a result, a set of maps is obtained which represents the entire range of possible $K$-values in the study area. This enables groundwater modelling with a range of realistically estimated conductivity conditions, which is relevant for e.g. "worst and best case scenarios" of contaminant transport. It is believed that such an approach may eventually reflect the real-world conditions more closely than a traditional crisp-value approach, because the former does not impose exactness artificially on phenomena which are diffuse by their nature.

INTRODUCTION

Kriging is one of the most common methods of data regionalization, i.e. of extrapolating data from randomly spaced points onto a regular grid. Kriging applications, however, are sometimes restricted owing to insufficient amount of input data. The number of available measurements is often too low and data scatter too high to obtain reliable variogram analysis, from which basic parameters for kriging are derived. In such a case the data set can be supplemented by additional, imprecise data subjectively estimated by an expert.
Fuzzy kriging is a modification of the conventional kriging procedure which utilizes exact (crisp) measurement data as well as imprecise estimates obtained from an expert (Bárdossy et al., 1989). For handling imprecision and uncertainty of estimates, a fuzzy set approach (Zadeh, 1965) is used here. Bárdossy et al. (1989) distinguish several types of fuzzy kriging, according to different types of input data and types of variograms. In this study, fuzzy kriging with both crisp and fuzzy data, and a crisp variogram was used. This regionalization procedure is used to calculate interpolated values as fuzzy numbers, reflecting the imprecision of the input data set and providing means of quantifying estimate uncertainty.

**GEOLOGICAL SETTING**

The study area of about 1700 km$^2$ is located in northwestern Germany in the vicinity of Kiel, south of the Baltic Sea basin (Fig. 1). Surficial sediments are of Quaternary age. Their sequence comprises primarily glacial sediments deposited during at least three major cold stages, that is the Elsterian, the Saalian and the Weichselian Glaciations (Grube et al., 1986; Stephan et al., 1983). Each of those glaciations consisted of multiple ice advances and retreats which left behind a full spectrum of sediments with tills, coarse-grained outwash material (sand and gravel) and glaciolacustrine silt and clay. In the interglacial periods, primarily lacustrine and marine clays accumulated in deep, elongated incisions of glacial origin (Ehlers & Linke, 1989; Piotrowski, 1994) or in shallow basins of large lateral extent. In numerous places the sediment sequence is glaciotectonically disturbed (e.g. Piotrowski, 1993). All this amounts to a rather complicated and heterogenous picture of the geology. The thickness of the Quaternary deposits varies between several hundreds of metres to just a few metres, and is commonly in the range of 100-150 m.
FUZZY KRIGING REGIONALIZATION

The above mentioned heterogeneity causes difficulties in spatial interpolation of geological data acquired from boreholes. One way to improve the interpolation quality is to involve expert knowledge in the form of subjective estimates of a given parameter where the boreholes are sparse. These estimates are expressed as fuzzy numbers.

Fuzzy number is a fuzzy set with convex membership function, which assumes a maximum value of 1 for one parameter value only (Zadeh, 1965). In this study, a linear (triangular) form of the membership function was applied, whereby a fuzzy number is defined by three values of the estimated parameter (Fig. 2):

- the most likely value of the estimated parameter ("a" in Fig. 2); the membership function is 1 for this value,
- the highest possible value of the estimated parameter ("b" in Fig. 2), and
- the lowest possible value of the estimated parameter ("c" in Fig. 2).

All three values are subjectively defined by an expert based on his experience. The highest and lowest possible values define the support (the interval of the possible values) of the fuzzy number. The form of the membership function does not have to be symmetrical. The membership function equals 0 outside the support.

The logical structure of kriging with fuzzy numbers applied here is shown in Fig. 3. Kriging is based on a theoretical variogram, which describes the spatial structure of the parameter considered. An experimental variogram with the input data is used as a statistical tool to find the theoretical variogram. Based on the theoretical variogram, interpolation of values onto a uniform grid is performed by minimizing the kriging variance.

Fuzzy data, marked in Fig. 3 with zigzag lines, enter the calculation procedure in two stages. In the first stage, a fuzzy experimental variogram is created from the input data, some of which are imprecise. The expert takes the fuzziness of the experimental variogram into account when determining the theoretical variogram. In the second stage at the final step of kriging calculation, the fuzzy input values are used again. This leads to kriging results which are fuzzy, too. It should be noted that kriging variance does not depend directly on the input values. It depends only on the spatial arrangement of the input data and on the theoretical variogram. Since in the kriging procedure used, the theoretical variogram is crisp and the input coordinates are crisp, the kriging variance is crisp, too.
For regionalization of hydraulic conductivities with fuzzy kriging as outlined above, a software package FUZZEKS (Fuzzy Evaluation and Kriging System; Bartels, 1996) was used, in which both exact (crisp) and fuzzy data are entered in one unified format. FUZZEKS is written in C++ and it operates in Windows 3.1 environment. Kriging results have the form of regularly rastered fuzzy numbers in a two-dimensional space. Some output options include viewing the results at a particular point of the investigated field, as a cross section through the field, and as contour lines, colours or greyscales of the regionalized values, their fuzziness and variance.

REGIONALIZATION OF HYDRAULIC CONDUCTIVITIES

In this study, regionalization of hydraulic conductivity values ($K$) for the major aquifer in the vicinity of Kiel was performed in order to obtain a regular finite-difference mesh of conductivities as an input to a numerical model of groundwater flow in this area. The aquifer consists of Elsterian and Saalian outwash sediments sandwiched between till layers, normally some 65-85 m below the ground surface. The aquifer thickness varies between 0 and 96 m and is in most cases some 15 to 25 m. Where no direct data on hydraulic conductivities were available at boreholes, weighted $K$-values were estimated for the borehole points from the grain-size distribution along the borehole profile. $K$ ranges in the model area between $5 \times 10^{-5}$ m s$^{-1}$ for fine-grained silty sand and $1 \times 10^{-2}$ m s$^{-1}$ for gravel.
Altogether 557 boreholes with reliable data exist in the study area. The boreholes are randomly distributed, with the greatest density at the city of Kiel and the lowest density in rural areas to the south and to the west of Kiel (Fig. 4). The experimental variogram parameters calculated for this crisp data set after a logarithmical transformation of $K$-values are: range 9296 m, sill 0.825 and nugget effect 0.716. The experimental variogram consisted of 20 classes.

Kriging was performed for a mesh of $64 \times 54 = 3456$ nodes which builds a block network composed of $1 \times 1$ km large blocks. The interpolated $K$-values are shown as contour lines in Fig. 4. Kriging variance ranging from 0.03 to 0.09 is relatively high and it reaches highest values in areas of scarce borehole density.

Subsequently, 30 additional data points with hydraulic conductivities were added in areas of greatest kriging variance. Each additional data point is a fuzzy number whose most likely value, the highest possible value and the lowest possible value (cf. Fig. 2) are estimated based on our subjective knowledge of the regional geology and depositional processes in glaciofluvial environments. Kriging results with fuzzy points are shown in Fig. 5, which exhibits a different arrangement of contour lines especially in the vicinity of new points. Kriging variance (max. about 0.06) is significantly lower than in the calculation without fuzzy data.

In order to visualize the possible divergence of hydraulic conductivities from the most likely values, contour lines of the interpolated highest and lowest values are given in Figs 6 and 7. This is a very useful presentation which enables a rapid spatial assessment of the expert-estimated, possible range of hydraulic conductivities in the

Fig. 4 Contour lines of the interpolated hydraulic conductivity ($K$) values, log-transformed; kriging of crisp values from boreholes (squares). Contour lines are 1: $K = 4.0 \times 10^3$ m s$^{-1}$, 2: $K = 1.6 \times 10^2$ m s$^{-1}$, 3: $K = 6.3 \times 10^4$ m s$^{-1}$, 4: $K = 2.5 \times 10^4$ m s$^{-1}$, 5: $K = 1.0 \times 10^5$ m s$^{-1}$, 6: $K = 4.0 \times 10^5$ m s$^{-1}$.
Fig. 5 Contour lines of the interpolated hydraulic conductivity (K) values, log-transformed; kriging of crisp values from boreholes (squares) and fuzzy values from data points estimated with expert knowledge (triangles). Contour lines as in Fig. 4. A-A' to C-C' and D: location of sections on the single point from Fig. 8.

Fig. 6 Contour lines of the highest possible K-values interpolated with crisp and fuzzy data, log-transformed. Contour lines as in Fig. 4.
Fig. 7 Contour lines of the lowest possible $K$-values interpolated with crisp and fuzzy data, log-transformed. Contour lines as in Fig. 4.

area. Furthermore, these two sets of regionalized $K$-values can be used as end members of boundary conditions for groundwater modelling. It should be emphasized that these two data sets differ from the results that would be obtained by conventional kriging carried out separately for (a) highest possible and (b) lowest possible $K$-values. This is because in fuzzy kriging, calculation of any raster point value always takes into account the entire membership function there.

Similar procedure can be applied to cross sections through the study area. Figure 8 A-C depicts three sections through the contour surface of $K$-values from Fig. 5. In each section the central thick line represents the most likely value, the line above is the highest possible and the line below the lowest possible value. Section in Figure 8A runs through numerous fuzzy points, which results in a broad spread between the upper and the lower lines (high fuzziness), whereas section in Fig. 8C runs primarily through crisp data points (low fuzziness). Finally, a similar visualization can be also applied to any single points within the area. Figure 8D depicts a randomly chosen point whose most likely hydraulic conductivity is 0.0005 m s$^{-1}$, the highest possible conductivity is 0.0017 m s$^{-1}$ and the lowest is 0.0002 m s$^{-1}$. Note that these values correspond to the notation a, b and c in Fig. 2.

In order to illustrate the spatial distribution of conductivities above and below any given boundary value taking into account the subjectivity of our estimate, the areas with $K$-values above (dark grey) and below (light grey) a randomly chosen value of $5 \times 10^{-4}$ m s$^{-1}$ are shown in Fig. 9. The two areas are separated with a contour line representing the value of $5 \times 10^{-4}$ m s$^{-1}$ with most likelihood. In places where the original data set was significantly modified by adding imprecise data, there is a white stripe on one or
Fig. 8 A–C: Sections through the contour surface of $K$-values from Fig. 5. Thick central line: most likely $K$-value; lower thin line: lowest possible $K$-value; upper thin line: highest possible $K$-value. The degree of fuzziness is greatest along the A–A’ section (lines lie far apart) and smallest along the C–C’ section (lines are closely spaced). $K$-values are on the vertical axis in m s$^{-1}$. C: Membership function of a single data point D. The most likely $K$-value is at this place 0.0005 m s$^{-1}$, the lowest possible is 0.0002 m s$^{-1}$ and the highest possible is 0.0017 m s$^{-1}$. Location of sections and point D is given in Fig. 5.

both sides of the contour line. This stripe represents a zone where there is an uncertainty regarding the affiliation of hydraulic conductivities to any of the two fields.

**CONCLUSIONS**

We consider fuzzy kriging for interpolation of sparse and unevenly distributed data such as, for instance, hydrogeological parameters of aquifers in northwestern Germany a useful tool with a potential of utilizing vast pieces of information available as inexact expert knowledge. In this approach inexactness accounts for the inherent vagueness of nature (Burrough, 1989) and does not force one to describe objects with largely artificial precision typical for conventional methods.

This method has several advantages when applied to creating input parameter matrix for groundwater flow modelling. By taking into account the estimate fuzziness, it enables one to run models with extreme — but yet still conceivable — boundary conditions, useful for determining e.g. worst case scenarios of contaminant migration. By displaying the spatial distribution of data fuzziness, it secures a semi-quantitative,
Estimating hydrogeological parameters for groundwater modelling with fuzzy geostatistics

Fig. 9 Areas where $K$-values are higher than $5 \times 10^{-4}$ m s$^{-1}$ (light grey), and lower than $5 \times 10^{-4}$ m s$^{-1}$ (dark grey). The two areas are separated with a contour line of $K = 5 \times 10^{-4}$ m s$^{-1}$. In places of high fuzziness there is a white stripe on one or on both sides of the boundary contour line. The stripe is the zone where there is an uncertainty in assigning the $K$-values to any of the two areas.

largely unbiased evaluation of groundwater simulation reliability in different regions of the model area. Furthermore, it provides means for "intelligent" sensitivity analysis of simulation results, where the possible range of the input parameter is determined by expert knowledge of a particular area.

Although in the example presented here, application of fuzzy kriging was a necessity caused by sediment heterogeneity and very unevenly distributed data points, determination of parameters for groundwater modelling which accounts for usual uncertainties caused by e.g. inaccurate methods or imprecise measuring devices, can be universally advocated.

Acknowledgement The authors thank Dr S. Christensen, Director of the Geological Survey of Schleswig-Holstein for providing them with borehole data.

REFERENCES


