(Proceedings of ModelCARE 2007 Conference, held in Denmark, September 2007). IAHS Publ. 320, 2008.

Can conditioning to transmissivity data worsen model predictions?

JAOUHER KERROU[1], HARRIE-JAN HENDRICKS FRANSSEN[2], PHILIPPE RENARD[1] & IVAN LUNATI[3]

1 *Centre for Hydrogeology, University of Neuchâtel, Emile Argand 11 - CP158, CH-2009 Neuchâtel, Switzerland*
jaouher.kerrou@unine.ch

2 *Institute of Environmental Engineering, ETH Zurich, CH-8093 Zurich, Switzerland*

3 *Laboratory of Soil and Environmental Physics, EPF Lausanne, CH-1015 Lausanne, Switzerland*

Abstract It is reasonable to think that spatially variable transmissivity fields often follow non-multi-Gaussian statistics. Nevertheless, in groundwater flow and mass transport studies multi-Gaussian models are very popular. This paper investigates the consequences of adopting a wrong Random Function (RF) model. Previous studies have shown that conditioning to hydraulic head data, adopting a multi-Gaussian approach, only very marginally detects connected structures typical for non-multi-Gaussian fields. In addition, several numerical simulations performed have given us a hint that conditioning on a large number of transmissivity data might prevent head conditioning from being effective. We consider non-multi-Gaussian *T* fields (with braided structures) and compare the results obtained by using the *T* data only for computing the variogram with those obtained by additionally conditioning to *T* data (erroneously, a multi-Gaussian RF model is assumed). The preliminary results presented here do not clearly show an improvement when only part of the data is used for *T* conditioning (this confirms the primary role played by the RF model in deteriorating field characterization). However, evidence is found that conditioning to *T* data yields a systematic loss of connectivity behind a distance of the order of the variogram range. This fact prevents the inverse problem from identifying elongated capture zones. Conditioning to *h* data, instead, generally yields an increase in connectivity, which is more effective at distances larger than the variogram range, and seems to allow a partial recovery of non-Gaussian structures.

Key words aquifer characterisation; inverse modelling; variogram; conditioning data; stochastic simulations; uncertainty; connectivity

INTRODUCTION

Groundwater protection and aquifer remediation are two examples of activities in which modelling the aquifer response to human activity is important. The first step toward accurate modelling is a reliable characterization of the geological medium. The heterogeneous structure of the aquifer has to be inferred on the basis of data that are available at discrete locations and can consist of transmissivity (*T*) and head (*h*) measurements. Groundwater flow and contaminant mass transport predictions are strongly affected by the uncertainty of the transmissivity field. As such, it is important to incorporate transmissivity measurements in flow models, which is achieved by geostatistical simulations. In general, the *T* fields produced by stochastic characterization are conditioned to the available *T* data, which are considered the only reliable values of the unknown field. However, many problems are associated with those measurements (e.g. scale dependence, interpretation or measurement errors, abrupt changes within short distances), which make the use of those values delicate.

The characterization of the uncertain, spatially variable transmissivity field can be further improved by conditioning to hydraulic head data by inverse modelling. An important decision to be taken when modelling groundwater flow concerns the multi-Gaussian assumption of the random field (e.g. Gómez-Hernández & Wen, 1998). Since the connectivity of extreme values of transmissivity has a strong impact on contaminant travel time, neglecting the connectivity (as multi-Gaussian models do) can yield a severe underestimation of travel times (e.g. Zinn & Harvey, 2003). Although inverse modelling methods were developed which can handle non-multi-Gaussian fields (e.g. Capilla *et al.*, 1999; Hendricks Franssen & Gómez-Hernández, 2002), it is cumbersome to infer the Random Function (RF) model from sparse observation data. This is one of the reasons that most spatially variable transmissivity fields are modelled as if they were multi-Gaussian, although this is often not the case. Kerrou *et al.* (2008) investigated how such a wrong decision affects the modelling results. They used a synthetic aquifer with marked non-multi-Gaussianity (Fig. 1), and modelled it with multi-Gaussian methods. They also investigated to what extent inverse modelling (conditioning to hydraulic head data) was able to correct the adoption of a wrong RF model. Note that hydraulic head data could detect channels of extreme hydraulic conductivities, even if such channeling was not detected on the basis of transmissivity data and not imposed by the RF model. Indeed, hydraulic head and transmissivity are correlated through the groundwater flow equation. Kerrou *et al.* (2008) found that hydraulic head data were hardly able to detect such channels, and that connectivity is underestimated even for inverse conditioned realizations. At the same time, they found that transmissivity data might prevent head data from being effective in finding connected channels in inverse conditioning. In case 1000 hydraulic head data were available for inverse conditioning, the results in terms of flow and transport predictions were better (although more uncertain), if 21 transmissivity data were used for conditioning than if 1000 transmissivity data were employed. Figure 2 (after Kerrou *et al.*, 2008) shows how conditioning is able to detect some of the connectivity.

This paper presents some additional simulations aimed at investigating in more detail the role of transmissivity data in inverse conditioning when a non-multi-Gaussian medium is erroneously modelled as being multi-Gaussian. The same synthetic reality is used as by Kerrou *et al.* (2008), in order to directly compare the results obtained by Kerrou *et al.* (2008) and these results.

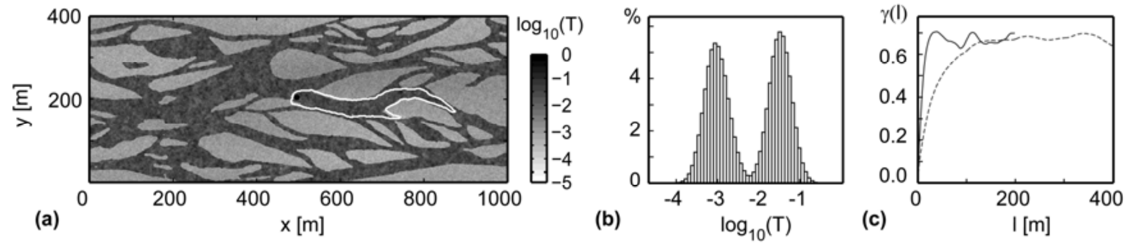

Fig. 1 (a) The synthetic transmissivity field with the 10-days capture zone around the well (small black circle) under steady flow conditions (white contour), (b) histogram of the decimal log of the transmissivities, (c) *x* (dashed line) and *y* (solid line) directional variograms.

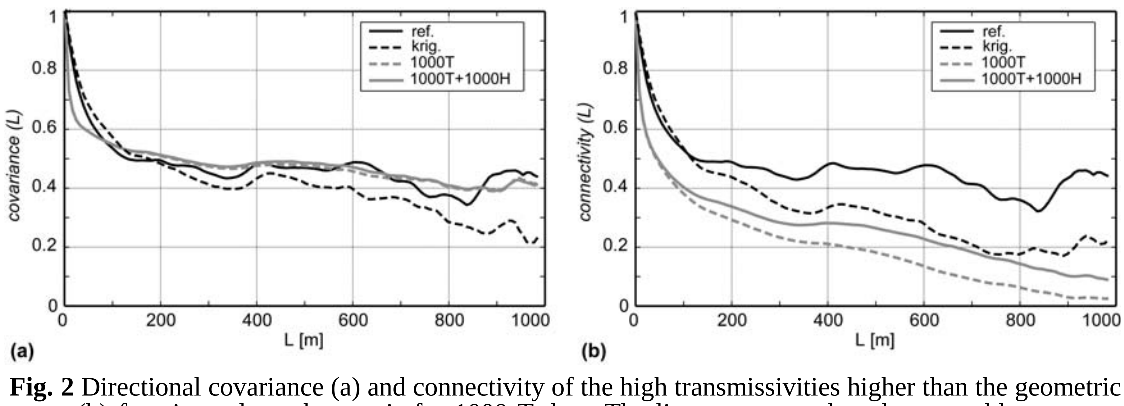

Fig. 2 Directional covariance (a) and connectivity of the high transmissivities higher than the geometric mean (b) functions along the x-axis for 1000 T data. The lines correspond to the ensemble averages over the 100 simulations (1000T, and 1000T+1000h). After Kerrou *et al.* (2008).

METHODOLOGY

This study is conducted on a synthetic transmissivity field (*T*) built from an aerial photograph displaying braided channels and lenses in the Ohau River, New Zealand (Mosley, 1982). Two unconditional multi-Gaussian simulations were generated to separately populate with *T* values channels and lenses. This yielded an aquifer whose structure consists of channels and lenses displaying internal heterogeneities (Fig. 1). By imposing constant head boundary conditions on the east (4.24 m) and west (0 m) boundaries and no flow boundary conditions in the north (positive y-axis direction) and south of the reference *T* field (assumed as a confined aquifer), a uniform steady-state flow was solved.

The synthetic *T* field was sampled to provide data for both direct and inverse characterization methods; also the reference hydraulic head field was sampled for inverse conditioning. The transmissivity data sets consist of 21, 250 and 1000 *T* measurements; the hydraulic head data set consists of 1000 *h* data. These data were used in direct conditioning with the turning bands method and in inverse conditioning with *INVERTO* (Hendricks Franssen, 2001). Some of the simulations were performed without conditioning to transmissivity data, aiming at investigating whether conditioning to transmissivity data prevents head data from revealing non-multi-Gaussian structures. Note that in those cases a variogram model estimated with the help of 250 or 1000 *T* data was adopted. In all cases 100 equally likely realisations were generated. These realisations were used as input for a forecasting problem. Figure 3 shows one of the generated realisations for two of the studied scenarios.

For the forecasting problem a well was added in the middle of the domain (the boundary conditions being the same as for the reference head field). The 10-days capture zone of the well was calculated by solving the Kolmogorov backward equation with the *GroundWater* finite element code (Cornaton, 2007).

RESULTS

A large number of comparison measures were calculated, but here we compare the results only on the basis of the well capture zone maps and the measures of connectivity. The other measures, especially the average ensemble absolute error and the average ensemble standard deviation (comparing each set of realizations to the *T* and *h* reference fields), showed in most cases better results when more transmissivity data were used for conditioning. These measures do not indicate that the effect of conditioning to transmissivity data limits the impact of head data.

The 10-days capture zone probability maps calculated for different datasets are shown in Fig. 4. Visual inspection of the capture zone maps indicates that conditioning to *T* data tends to reduce uncertainty. However, the identification of the capture zone is not clearly improved. In particular, it is surprising that probability spreads in the direction transversal to the flow. This is evident when the variogram estimated on the basis of 1000 transmissivity data was used (Fig. 4(d)–(h)).

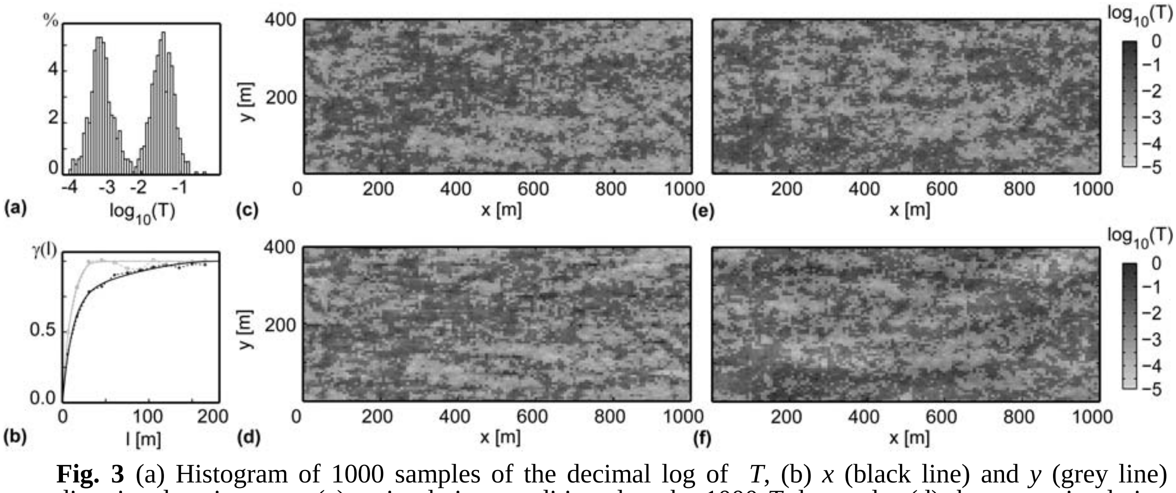

Fig. 3 (a) Histogram of 1000 samples of the decimal log of *T*, (b) *x* (black line) and *y* (grey line) directional variograms; (c) a simulation conditioned to the 1000 *T* data only; (d) the same simulation after conditioning to 1000 *T* and *h* measurements; (e) an unconditional simulation generated using only the variogram of 1000 *T* data; (f) the same unconditional simulation generated using the variogram of 1000 *T* data and conditioning to 1000 *h* measurements.

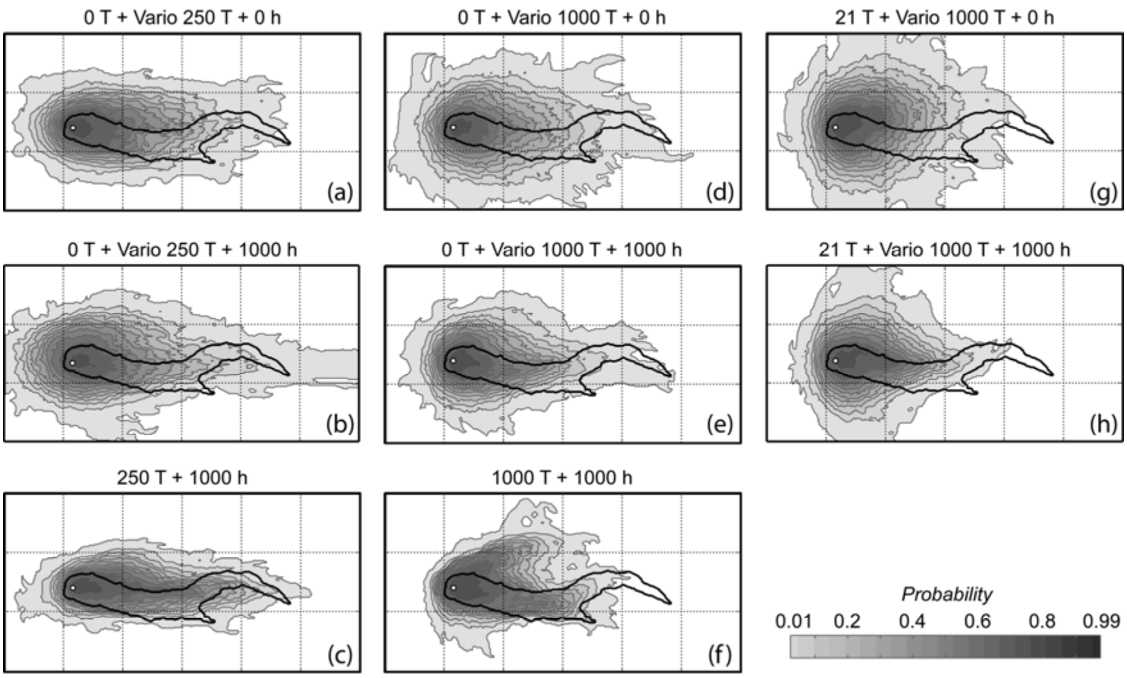

Fig. 4 Probability maps of the 10-days capture zone. The black line represents the reference 10-days capture zone. The grey levels represent the isoprobability contours for all the combinations of *T* and *h* datasets, and variogram models (i.e., estimated on the basis of 250 or 1000 transmissivity data), according to the map title.

Conditioning to 21 *T* data yields a probability area that is larger in the *y* direction (compare Fig. 4(d), resp. Fig. 4(e), with Fig. 4(g), resp. Fig. 4(h)), whereas it becomes shorter in the *x* direction and fails to identify the most upstream part of the capture zone. We attribute this to a lack of connectivity typical of multi-Gaussian models: if the connectivity in the flow direction is underestimated, the forecasted capture zone becomes less elongated and spreads in the transversal direction to satisfy mass balance (the area is solely determined by the well rate and must be the same regardless of the specific *T* field). Also the results obtained conditioning to 1000 *T* data can be expressed by this argument. The lack of connectivity makes it more convenient for the well to extract water from the northern channel than from far upstream in the southern channel (compare Fig. 1 and Fig. 4(f)). Therefore, the upstream part of the channel is not captured anymore in any of the realisations, whereas a high probability of flow from the northern channel is erroneously predicted (Fig. 4(f)). Conditioning to *h* data, instead, always helps in finding the elongated channel, as it can be observed by comparing Fig. 4(d), resp. Fig 4(g), with Fig. 4(e), resp. Fig. 4(h). So, a visual inspection of the well capture zones gives hints that transmissivity data make it more difficult to find the parts of the well capture zone which are related to the non-multi-Gaussian properties of the *T* field.

To further investigate these aspects, we have calculated the connectivity of the *T* fields (Fig. 5). The multi-Gaussian simulations systematically underestimate the connectivity of the non-multi-Gaussian reference *T* field. When more conditioning data are used, the characterisation of connectivity improves. Although the best results are obtained if both transmissivity and head data are used for conditioning, it is evident that conditioning to head data increases the long-range connectivity. Note that this effect can only be appreciated at distance larger than the variogram range (about 100 m, see Fig. 3).

DISCUSSION AND CONCLUSIONS

This paper analysed the role of transmissivity data in inverse conditioning, if a multi-Gaussian model was erroneously adopted. Kerrou *et al*. (2008) found evidence that head data were less able

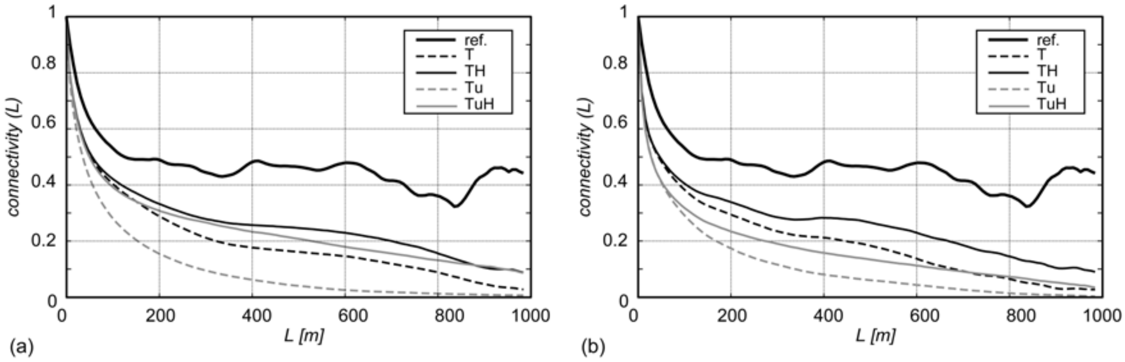

Fig. 5 Connectivity functions (of transmissivities higher than the mean) along the *x*-axis for (a) 250 *T* data and (b) 1000 *T* data. The lines corresponding to the simulations are ensemble averages over the 100 simulations. Index corresponds to: *T* transmissivity data plus their variogram, *Tu* only variogram, *TH* transmissivity and heads, *TuH* only variogram plus heads and *ref* reference field.

to "correct" the wrong Random Function model in case many transmissivity data were used for conditioning. This paper points again in that direction, and addresses the question whether using only the head data for conditioning and a small number of transmissivity data (whereas all transmissivity data are used to compute the variogram) can improve the prediction.

The results presented here do not indicate such an improvement, and suggest that it is rather the RF model based on the computed variogram and on the multi-Gaussian assumption which is responsible for unsatisfactory predictions. This is confirmed by the fact that, in general, conditioning to *T* data improves the identification of the capture zone in the vicinity of the well, i.e. within a distance comparable with the variogram range, but yields a systematic loss of connectivity behind this distance. Conditioning to *h* data, instead, yields an increase in connectivity that is more effective at distances larger than the variogram range. We can suggest that, in general, head data tend to help to detect the elongated channels of the well capture zone, whereas transmissivity data have an opposite impact and yield a less elongated and larger probability map.

In order to assess impact of conditioning to transmissivity data when a multi-Gaussian model is erroneously applied to a non-multi-Gaussian reality, further investigation is needed. The present work indicated that connectivity plays a primary role and therefore suggested that the wrong RF model is the first source of deterioration for the capture-zone prediction. To circumvent these difficulties, we argue that two solutions may be proposed, one is to use alternative RF models that can handle different connectivity patterns such as the multiple point statistics or to use a multi-Gaussian model but with variograms that would have smaller ranges than the experimental variograms. In that way, the inverse method may have more degrees of freedom and may be able to compensate the inconsistencies between the head data and the stochastic model for the transmissivity field.

Acknowledgments J. Kerrou and P. Renard would like to acknowledge the Swiss National Science Foundation for funding part of this research under Grant nos. 207020-110017/1 and PP002-106557. Furthermore, we thank Fabien Cornaton for providing the GroundWater code and supporting us while preparing the simulations.

REFERENCES

Capilla, J., Rodrigo, J. & Gómez-Hernández, J.J. (1999) Simulation of non-Gaussian transmissivity fields honouring piezometric data and integrating soft and secondary information. *Math. Geology* **31**(7), 907–927.

Cornaton, F. (2007) GroundWater: A 3-D ground water flow and transport finite element simulator. Reference manual. http://www1.unine.ch/chyn/php/softwares.php.

Gómez-Hernández, J. J. & Wen, X. -H. (1998) To be or not to be multi-Gaussian? A reflection on stochastic hydrogeology. *Adv. Water Resour.* **21**(1), 47–61.

Hendricks Franssen, H. J. (2001). Inverse stochastic modeling of groundwater flow and mass transport. PhD dissertation Technical University of Valencia, Spain.

Hendricks Franssen, H. J. & Gómez-Hernández, J. J. (2002). 3D Inverse modelling of groundwater flow at a fractured site using a stochastic continuum model with multiple statistical populations. *Stoch. Env. Res. & Risk A.* **16**, 155–174.

Kerrou, J., Renard, P., Hendricks Franssen, H. J. & Lunati, I. (2008). Issues in characterizing heterogeneity and connectivity in non-multi-Gaussian media. *Adv. Water Resour.* **31**(1), 147-159 doi:10.1016/j.advwatres.2007.07.002.

Mosley, M. P. (1982). Analysis of the effect of changing discharge on channel morphology and instream uses in a braided river, Ohau River, New Zealand. *Water Resour. Res.* **18**(4), 800–812.

Zinn, B. & Harvey, C. F. (2003) When good statistical models of aquifer heterogeneity go bad: A comparison of flow, dispersion, and mass transfer in connected and multivariate Gaussian hydraulic conductivity fields *Water Resour. Res.* **39**(3), 1051. doi:10.1029/2001WR001146.