Optimal solute transport in heterogeneous aquifer: coupled inverse modelling

Jie Liang and Guangming Zeng*

College of Environmental Science and Engineering, Hunan University, Changsha, Hunan Province, 410082, China

and

Key Laboratory of Environmental Biology and Pollution Control (Hunan University), Ministry of Education, Changsha 410082, China E-mail: liangjie_ever@yahoo.com.cn E-mail: zgming@hnu.cn *Corresponding author

Shenglian Guo

State Key Laboratory of Water Resources and Hydropower Engineering Science, Luo Jiashan Street, Wuhan, 430072, China

and

College of Environmental Science and Engineering, Hunan University, Changsha, Hunan Province, 410082, China E-mail: Slguo@whu.edu.cn

Anlei Wei, Xiaodong Li, Lin Shi and Chunyan Du

College of Environmental Science and Engineering, Hunan University, Changsha, Hunan Province, 410082, China

and

Key Laboratory of Environmental Biology and Pollution Control (Hunan University), Ministry of Education, Changsha 410082, China E-mail: chinaalwei@gmail.com E-mail: lxdfox@163.com E-mail: shilin307@yahoo.com.cn E-mail: dcyhb220@126.com

Copyright © 2010 Inderscience Enterprises Ltd.

Abstract: Characterisation of transmissivity heterogeneity is critical for groundwater flow and solute transport. The heterogeneity of transmissivity is studied through variogram-based techniques. Conventionally, the parameters in variogram are obtained by fitting measurements to a theoretical variogram. However, conditioning to the sampled geological variables neglects the effects of observed concentration data. This paper presents a coupled inverse modelling system conditioning to both types of measurement. The results of a hypothetical two-dimensional steady flow indicated that the description of transmissivity and solute concentration field was improved when both measured transmissivity data and solute concentration data were combined in the coupled inverse modelling system.

Keywords: groundwater; heterogeneity; inverse modelling; variogram; transmissivity.

Reference to this paper should be made as follows: Liang, J., Zeng, G.M., Guo, S.L., Wei, A.L., Li, X.D., Shi, L. and Du, C.Y. (2010) 'Optimal solute transport in heterogeneous aquifer: coupled inverse modelling', *Int. J. Environment and Pollution*, Vol. 42, Nos. 1/2/3, pp.258–269.

Biographical notes: Jie Liang is a Doctoral candidate at Hunan University. Her main research interest is environmental system analysis.

Guangming Zeng is a Professor and Head of (Chang Jiang Scholar) of the College of Environmental Science and Engineering, Hunan University. His research interests include environmental system analysis, wastewater treatment, waste gas treatment, and municipal solid waste utilisation. His research has been published in journals such as *Environmental Science and Technology, Water Research*.

Shenglian Guo is a Professor of the College of Water Resource and Hydroelectric Engineering in Wuhan University. His main research interests are hydrology and water resources. His research has been published in journals such as *Journal of Hydrology, Hydrological Processes*.

Anlei Wei is a Doctoral candidate at Hunan University. His main research interest is environmental system analysis.

Xiaodong Li is an Assistant Professor of College of Environmental Science and Engineering, Hunan University. His main research interest is environmental system analysis.

Lin Shi is a Doctoral candidate at Hunan University. His main research interest is water resources management.

Chunyan Du is a Doctoral candidate at Hunan University. Her main research interesting is atmospheric systems analysis.

1 Introduction

The dynamic prediction of solute transport in groundwater is one of the most important steps towards effective groundwater management (Li et al., 2007). However, groundwater flow and solute transport predictions are always uncertain due to an

imperfect knowledge of the aquifer properties (Zeng et al., 2004). One of the uncertainties results from aquifer heterogeneity. Heterogeneity includes variations in grain-size, porosity, mineralogy, lithologic texture, rock mechanical properties, structure and digenetic processes. All these factors cause variations in transmissivity, storage, and thus control flow and transport in the aquifer (Timothy, 2006). Among various parameters, the heterogeneity of transmissivity (T) has attracted wide attention for its spatial variability is considerably higher than that of other properties and it can vary by orders of magnitude over a few metres (Feyen et al., 2003).

It is of crucial importance to improve the characterisation of transmissivity field, for it is served as inputs to the groundwater flow and solute transport simulator. During the last decade, the heterogeneity of transmissivity has been studied based on the statistical analysis because of its random characteristic (Weissmann and Fogg, 1999; Wang et al., 2001; Vrankar et al., 2004; Flipo et al., 2007). The majority of the applications have employed variogram-based techniques (Feyen and Caers, 2006). In the application, the variogram function form of transmissivity should be firstly determined by matching the experimental variogram to a theoretical variogram family either through determinative method or stochastic simulation. Curve-fitting techniques are the most commonly used determinative method, including ordinary or weighted least squares estimation, maximum or restricted maximum likelihood estimation. Recently, Bayesian approach is introduced to estimate the unknown values of parameters in the variogram function, which makes the prediction of groundwater flow and solute transport more conservative (Feyen et al., 2002, 2003; Xu et al., 2005, 2006).

The step described above reflects the degree of experimental variogram matching with a theoretical one. However, it neglects the role of observed concentration (c) data, in spite of the fact that concentration data, for example, from controlled tracer tests, may give important information on spatially variable aquifer prosperities like transmissivity. Rubin (1991) developed an approach that allows the conditioning of concentration ensemble moments on hydraulic head, conductivity, velocity and concentration measurements. But the approach failed to update the transmissity field. Franssen et al. (2003) extended the self-calibrating method to the coupled inverse modelling of flow and transport conditioning to hydraulic conductivity (K) data, hydraulic head data (h) and concentration data (c). Conditional to K data was simulated by geostatistical method and conditional to h and c data was achieved by minimising the objective function consisting of the head and concentration discrepancies between the observed and simulated values. The methodology updated hydraulic conductivity field by perturbation at master blocks. However, the calculation of the optimal perturbation was not straightforward with a complex computation process.

It is believed that in groundwater flow and solute transport simulation at least two types of measurements should be taken into account:

- the sampled geological parameters
- the observed solute concentration data.

In this paper, an optimal groundwater flow and solute transport simulation method in heterogeneous aquifer is developed. The method combines the effects of sampled transmissivity measurements and solute concentration observations and updates the two types of data simultaneously. The method is then applied to a hypothetical two-dimensional steady flow in heterogeneous confined aquifer. The worth of the two types of data has also been illustrated.

2 Methodology

The groundwater flow in a heterogeneous confined aquifer is considered as

$$\nabla \cdot T\nabla h = S\frac{\partial h}{\partial t} + W$$

where T is the transmissivity, h is the potentiometeri head, W represents sources or sinks of water, S is the specific storage of the porous material, and t is the time.

And the solute transport is described as

$$\frac{1}{R}\nabla\cdot(\theta D\cdot\nabla c-\theta vc)=\theta\frac{\partial c}{\partial t}+q_s$$

where *R* is retardation factor, θ is porosity of the subsurface medium, *D* is the hydrodynamic dispersion coefficient tensor, *c* is the solute concentration, *v* is seepage or linear pore water velocity, and *q_s* is volumetric flow rate per unit volume of aquifer representing solute sources and sinks.

To solve the groundwater flow and solute transport model, the transmissivity (*T*) field has to be estimated. The first step for the estimation is to determine the form of variogram by fitting the measured transmissvity data. Generally, transmissivity is found to be log-normally distributed in a heterogeneous aquifer (Freeze, 1975; Sudicky, 1986). Let T(x) denote the stochastic transmissvity field and its log form is denoted as $Y(x) = \log_{10}T(x)$, where x is the unmeasured spatial location. The objective function of the fitting process can be defined as follows:

$$J_{1} = \sqrt{\sum_{i=1}^{N_{iY}} \xi_{Y} (Y_{i}^{\rm sim} - Y_{i}^{\rm obs})^{2}}$$

where N_{iY} is the number of measured Y values, ξ_Y are the weights that chosen inverse-proportional to the average estimated measurements errors, and the superscripts *sim* and *obs* refer to 'simulated' and 'observed' data, respectively.

After the experimental variogram is obtained, the transmissivity field is generated by geostatistical simulation, either determined or stochastic, for instance, Kriging interpolation, Gaussian sequential simulation, sequential indicator simulation and so on. Then the generated T field is used as inputs to groundwater flow and solute transport model. The conditioning to the solute concentration observations is achieved by comparing the measured concentration values with their simulated values at the same locations and times. The objective function defined to measure the mismatch between simulated and measured head and concentration values are as follows:

$$J_{2} = \sqrt{\sum_{t=1}^{N_{tc}} \sum_{i=1}^{N_{ic}} \sum_{j=1}^{N_{jc}} \xi_{c} (c_{ijt}^{sim} - c_{ijt}^{obs})^{2}}$$

where J_2 corresponds to the solute concentration discrepancies at the different time steps and locations; N_{tc} is the simulating time steps of concentration; N_{ic} is the number of observed concentration data; N_{jc} is the number of solutes; *c* represents the solute concentration value, ξ_c are the weights chosen inverse-proportional to the average estimated measurements errors.

With consideration of the two types of data, the updating process is achieved by minimising the mismatch between the theoretical and experimental variogram and the differences between simulated and measured concentration values:

$$J = \varphi_1 J_1 + \varphi_2 J_2$$

where φ_1 and φ_2 are the values of trade-off coefficient that are chosen according to the importance and overall variability within the study area. The objective function *J* is minimised to achieve a satisfactory reproduction of the $\log_{10}T$ and *c* field. When *J* reaches a value below a user-defined one, it is considered that the $\log_{10}T$ and *c* field are reproduced sufficiently close to the reference field. The minimisation process is terminated in case the number of iterations exceeds a user-specified maximum number of iterations or the objective function reduction is very small during the iterations. It is possible that the optimisation finds a local minimum, but in case of a sufficient close reproduction of the measured data, this is not considered to be a problem (Zeng et al., 2007). Nevertheless, the implementation of faster converging optimisation algorithms that are less sensitive to local minimu is subject to future research.

3 Applications

In the paper, a hypothetical two-dimensional steady flow and transport in heterogeneous confined aquifer (Wilson and Miller, 1978) is adopted to illustrate the application of the method. The study area is depicted in Figure 1. The flow model is surrounded by constant-head boundaries on the east and west borders and no-flow boundaries on the north and south borders. The head values at the constant-head boundaries are arbitrarily chosen to establish the required hydraulic gradient and force a flow from west to east. A spill of an inert contaminant occurred in the study area. The study area is discredited into 46 columns along *x*-axis and 31 rows along *y*-axis. Table 1 lists part of the modelling inputs. Groundwater flow is simulated in MODFLOW (Harbaugh and McDonald, 1996) with finite difference method and solute transport is simulated in MT3DMS (Zheng and Wang, 1999) with the method of characteristics solution scheme.

The reference $\log_{10}T$ field is generated using the Sequential Gaussian Simulation (SGSim) algorithm of GSLIB (Deutsch and Journel, 1998) with an average $\log_{10}T$ of 2.0 $\log_{10} (m^2/s)$. The variogram of $\log_{10}T$ is exponential with a range of 50 m, zero nugget and sill of 1.0 $(\log_{10} (m^2/s))^2$. The reference $\log_{10}T$ and the reference concentration field for time steps 100d, 200d and 300d are depicted in Figure 2.



Figure 1 Plan view of study area

 Table 1
 Part of the modelling parameters

Parameter	Value	Unit
Cell width along rows (Δx)	10	m
Cell width along columns (Δy)	10	m
Layer thickness (Δz)	10	m
Groundwater seepage velocity (v_x)	0.33	m/s
Porosity (<i>n</i>)	0.3	_
Dispersivity along x-axis (D_x)	10	m
Dispersivity along y-axis (D_y)	3	m
Volumetric injection rate (q_s)	1	m ³ /d
Concentration of the injected pollutant (c_s)	1000	ppm









Figure 2 Reference $\log_{10}T$ field, solute concentration field for time steps 100d, 200d and 300d (continued)

3.1 Scenarios studied

As the Kriging interpolation is applied to generate Y field, two sets of transimissivity (T) data with different sampling density are used as conditional data for interpolation. The sets are defined so that the smaller set is a subset of the larger one to circumvent the effect of varying T sampling density on the generated Y field and solute concentration (c) field between two sets. With the consideration of the number of Y data used for interpolation and the type of conditional data used for inverse modelling, six different scenarios are calculated. Table 2 illustrates the types and the numbers of data used. In two of the six scenarios (Scenario 1 and 3), 8 Y data are incorporated in the updating process. And in the other two of the six scenarios (Scenario 2, 3, 5 and 6), c data are incorporated in the updating process. The c data are sampled from the reference concentration data for the time step of 100d, 200d and 300d at 11 locations along and perpendicular to the studied field. Among them, 8 Y data are used as conditional data for Kriging interpolation in Scenario 2 and 70 Y data in Scenario 5. Figure 3 shows the location of the conditional data.





 Table 2
 Data sets incorporated in the updating of the different scenarios

	8 Y Data	70 Y Data	11 c Data
Scenario 1	Yes	No	No
Scenario 2	No (used in the Kriging interpolation)	No	Yes

	8 Y Data	70 Y Data	11 c Data	
Scenario 3	Yes	No	Yes	
Scenario 4	No	Yes	No	
Scenario 5	No	No Yes (used in the Kriging interpolation)		
Scenario 6	No	Yes	Yes	

 Table 2
 Data sets incorporated in the updating of the different scenarios (continued)

3.2 Evaluation of results

Each of the realisations is compared with the reference fields and the following performance measures are defined for each of the scenarios:

$$E = \frac{1}{N} \sqrt{\sum_{i=1}^{n} (X_{\text{SIM}} - X_{\text{REF}})^2}$$

where E is the average square root error, N is the number of grid cells, and i is the cell index, X represents either log T or solute concentration in a certain grid cell, the subscripts *SIM* and *REF* refer to the simulated and the reference values, respectively.

3.3 Results

In this paper, Genetic Algorithm (Harrouni et al., 1996; Giacobbo et al., 2002; Zeng et al., 2003) is applied to update the Y data and c data in the inverse modelling process. Table 3 shows the calculated average square root error (E) for the six scenarios. As most previous researches were based on the conditional to Y data method, the calculated E of other scenarios are compared with them Scenario 1 or Scenario 4, respectively.

E(Y)E(c)Scenario 1 0.7371 0.8883 Scenario 2 0.7397 0.9343 Scenario 3 0.7328 0.7767 Scenario 4 0.6088 0.4507 Scenario 5 0.6163 0.4753 0.5923 Scenario 6 0.4043

 Table 3
 Average square root error for different scenarios

3.3.1 Results when a single type of data is used for inverse modelling

In this section, Scenario 1, 2, 4 and 5, where only one type of conditional data is considered in the updating procedure, are analysed. The better results are obtained when Y data are used. When c data are used, both the characterisation of transmissivity field

and solute concentration field become worse. Compared with Scenario 1, the E(Y) and E(c) increases 0.4% and 4.9% for Scenario 2, respectively. And compared with Scenario 4, the E(Y) and E(c) increases 1.2% and 5.5% for Scenario 5, respectively. It indicates the important role of transmissivity data in helping improving the description of transmissivity field and thus solute concentration field. When the *c* data are used as conditional data individually, although the optimisation target could reach the minimum, it takes no effect on the overall solute concentration distribution.

3.3.2 Results when different numbers of transmissivity data are used for interpolation

To further investigate the role of conditional transmissivity data, Scenario 1, 2, 3 are compared with Scenario 4, 5 and 6. In Scenario 1, 2 and 3, 8 Y data are used for Kriging interpolation while in Scenario 4, 5 and 6, 70 Y data are used. In all cases, when conditioning to the same type of data, the characterisation of the transmissivity field and solute concentration field improves for more conditional transmissivity data scenarios. E(Y) is reduced 17.4%, 16.7% and 19.2%, respectively, with respect to 70 Y data scenarios in comparison with 8 Y data scenarios. E(c) is reduced 49.3%, 49.1% and 47.9%, respectively, with respect to 70 Y data scenarios. It shows that conditioning to more transmissivity data helps to yield a better description of the transmissivity field and thus resulting in a noticeable improvement of the characterisation of solute concentration.

3.3.3 Results for the coupled inverse modelling

In case both transmissivity and solute concentration data are available in the updating process (Scenario 3 and 6), better results are obtained. For 8 *Y* data scenario, the *E*(*Y*) and *E*(*c*) reductions are 0.6% and 12.6%. And for 70 *Y* data scenario, the *E*(*Y*) and *E*(*c*) reductions are 2.7% and 10.3%. It should be noticed that the combination use of transmissivity data and solute concentration data yields important *E*(*Y*) reductions and significant *E*(*c*) reductions. The greater reduction of *E*(*Y*) occurs in the more conditional transmissivity scenario. However, the greater reduction of *E*(*c*) occurs in less conditional transmissivity scenario, which is usually the reality in practice. The generated $\log_{10}T$ field and the relative solute concentration field of Scenario 6 can be seen in Figure 4.

Figure 4 $\log_{10}T$ field and the relative solute concentration field for time steps 100d, 200d and 300d of Scenario 6







4 Summary

Instead of traditional conditioning to transmissivity method in solute transport simulation in the groundwater, the coupled inverse modelling method was presented. The developed methodology took the effect of both measured transmissivity data and solute concentration data into account in the procedure of an inverse optimisation. A hypothetical two-dimensional steady flow and transport in a heterogeneous aquifer was used as an example. Six different scenarios were considered in the paper to investigate the role of different types of conditional data and different numbers of conditional data. The results indicated that transmissivity played an important role on charactering transmissivity field and improving the precision for solute concentration simulation when a single type of data was used in the updating process individually. More transmissivity data helped to yield a better description of the transmissivity field and a noticeable improvement of the characterisation of solute concentration. On the contrary, the measured concentration data took no effects on improving the description of either transmissivity field or solute concentration field when it was used as conditional data individually. However, the greater improvement of the description of transmissivity and solute concentration field occurred when both measured transmissivity data and solute concentration data were combined in the coupled inverse modelling system. The coupled inverse modelling provided a trade-off between geological setting and solute transport. It made a more comprehensive use of available data and had potential for a large range of applications.

Acknowledgements

The study was financially supported by National 863 High Technologies Research Foundation of China (No. 2004AA649370), the National Basic Research Program (973 program) (No. 2005CB724203), the National Foundation for Distinguished Young Scholars (Nos. 5042596, 50225926) and the Program for Changjiang Scholars and Innovative Research Team in University (IRT0719).

References

- Deutsch, J.P. and Journel, A.G. (1998) *GSLIB, Geostatistical Software Libarary and User's Guide*, Oxford University Press, New York.
- Feyen, L. and Caers, J. (2006) 'Quantifying geological uncertainty for flow and transport modelling in multi-modal heterogeneous formations', *Advances in Water Resources*, Vol. 29, pp.912–919.
- Feyen, L., Ribeiro Jr., P.J., de Smedt, F. and Diggle, P.J. (2002) 'Bayesian methodology to stochastic capture zone determination: conditional on transmissivity measurements', *Water Resources Researches*, Vol. 38, No. 9, p.1164.
- Feyen, L., Ribeiro Jr., P.J., De Smedt, F. and Diggle, P.J. (2003) 'Stochastic delineation of capture zones: classical versus Bayesian approach', *Journal of Hydrology*, Vol. 281, pp.313–324.
- Flipo, N., Jeannée, N., Poulin, M., Even, S. and Ledoux, E. (2007) 'Assessment of nitrate pollution in the Grand Morin aquifers (France): combined use of geostatistics and physically based modelling', *Environmental Pollution*, Vol. 146, No. 1, pp.241–256.
- Franssen, H.H., Hernández, J.G. and Sahuquillo, A. (2003) 'Coupled inverse modelling of groundwater flow and mass transport and the worth of concentration data', *Journal of Hydrology*, Vol. 281, pp.281–295.
- Freeze, R.A. (1975) 'A stochastic-conceptual analysis of one-dimensional groundwater flow in nonuniform homogeneous media', *Water Resources Research*, Vol. 11, No. 5, pp.725–741.
- Giacobbo, F., Marseguerra, M. and Zio, E. (2002) 'Solving the inverse problem of parameter estimation by genetic algorithms: the case of a groundwater contaminant transport model', *Annals of Nuclear Energy*, Vol. 29, No. 8, pp.967–981.
- Harbaugh, A.W. and McDonald, M.G. (1996) Programmer's Documentation for MODFLOW-96, an Update to the US Geological Survey Modular Finite-Difference Ground-Water Flow Model: US Geological Survey Open-File Report 96-486, p.220.
- Harrouni, K., Ouazar, D., Walters, G.A. and Cheng, A.H.D. (1996) 'Groundwater optimization and parameter estimation by genetic algorithm and dual reciprocity boundary element method', *Engineering Analysis with Boundary Elements*, Vol. 18, No. 4, pp.287–296.
- Li, J.B., Huang, G.H., Zeng, G.M., Maqsood, I. and Huang, Y.F. (2007) 'An integrated fuzzy-stochastic modelling approach for risk assessment of groundwater contamination', *Journal of Environmental Management*, Vol. 82, No. 2, pp.173–188.
- Rubin, Y. (1991) 'The spatial and temporal moments of tracer concentration in disordered porous media', *Water Resources Research*, Vol. 27, No. 11, pp.2845–2854.
- Sudicky, E.A. (1986) 'A natural gradient experiment on solute transport in a sand aquifer: spatial variability of transmissivity and its role in the dispersion process', *Water Resources Research*, Vol. 22, No. 13, pp.2069–2082.
- Timothy, T.E. (2006) 'On the importance of geological heterogeneity for flow simulation', Sedimentary Geology, Vol. 184, pp.187–201.
- Vrankar, L., Turk, G. and Runovc, F. (2004) 'Combining the radial basis function eulerian and Lagrangian schemes with geostatistics for modelling of radionuclide migration through the geosphere', *Computers and Mathematics with Applications*, Vol. 48, Nos. 10–11, pp.1517–1529.
- Wang, Y., Ma. T. and Luo, Z. (2001) 'Geostatistical and geochemical analysis of surface water leakage into groundwater on a regional scale: a case study in the Liulin karst system, northwestern China', *Journal of Hydrology*, Vol. 246, Nos. 1–4, pp.223–234.
- Weissmann, G.S. and Fogg, G.E. (1999) 'Multi-scale alluvial fan heterogeneity modeled with transition probability geostatistics in a sequence stratigraphic framework', *Journal of Hydrology*, Vol. 226, Nos. 1–2, pp.48–65.
- Wilson, J.L. and Miller, P.J. (1978) 'Two-dimensional plume in uniform ground water flow', J. Hyd. Div., ASCE, Vol. 1978, pp.503–514.

- Xu, M., Zeng, G.M., Xu, X.Y., Huang, G.H., Jiang, R. and Sun, W. (2006) 'Application of Bayesian regularized BP neural network model for trend analysis, acidity and chemical composition of precipitation in North Carolina', *Water, Air and Soil Pollution*, Vol. 172, pp.167–184.
- Xu, M., Zeng, G.M., Xu, X.Y., Huang, G.H., Jiang, R., Sun, W. and Jiang, X.Y. (2005) 'Application of Bayesian regularized BP neural network model for analysis of aquatic ecological data – a case study of chlorophyll-a prediction in Nanzui water area of Dongting Lake', *Journal of Environmental Science*, Vol. 17, No. 6, pp.946–952.
- Zeng, G.M., Jiang, R., Huang, G.H., Xu, M. and Li, J.B. (2007) 'Optimization of wastewater treatment alternatives election by hierarchy grey relational analysis', *Journal of Environmental Management*, Vol. 82, pp.250–259.
- Zeng, G.M., Lin, Y.P., Qin, X.S., Huang, G.H., Li, J.B. and Jiang, R. (2004) 'Optimum municipal wastewater treatment plant design with consideration of uncertainty', *Journal of Environmental Science*, Vol. 16, No. 1, pp.126–131.
- Zeng, G.M., Su, X.K., Huang, G.H. and Xie, G.X. (2003) 'A new model for the grid size optimization of the finite element method – based on its application to the water quality modelling of the topographically complicated river', *Process in Natural Science*, Vol. 13, No. 12, pp.920–926.
- Zheng, C. and Wang, P.P. (1999) MT3DMS, A Modular Three-Dimensional Multi-Species Transport Model for Simulation of Advection, Dispersion and Chemical Reactions of Contaminants in Groundwater Systems, Documentation and User's Guide, US Army Engineer Research and Development Center Constract Report SERDP-99-1, Vicksburg, MS.