Journal of Contaminant Hydrology 164 (2014) 285–298

Contents lists available at ScienceDirect

Journal of Contaminant Hydrology

journal homepage: www.elsevier.com/locate/jconhyd

# Particle Swarm Optimization for inverse modeling of solute transport in fractured gneiss aquifer

Ramadan Abdelaziz [a,*], Mauricio Zambrano-Bigiarini [b,1]

[a] Institute for Geology, Technische Universität Bergakademie Freiberg, Freiberg, Germany
[b] Water Resources Unit, Institute for Environment and Sustainability, Joint Research Centre, European Commission, Via E. Fermi 2749, TP261, 21027 Ispra, VA, Italy

ARTICLE INFO

*Article history:*
Received 28 October 2013
Received in revised form 3 June 2014
Accepted 6 June 2014
Available online 19 June 2014

*Keywords:*
MODFLOW2005
MT3DMS
hydroPSO
Calibration
Inverse modeling
Fractured aquifer
Gneiss aquifer

ABSTRACT

Particle Swarm Optimization (PSO) has received considerable attention as a global optimization technique from scientists of different disciplines around the world. In this article, we illustrate how to use PSO for inverse modeling of a coupled flow and transport groundwater model (MODFLOW2005–MT3DMS) in a fractured gneiss aquifer. In particular, the hydroPSO R package is used as optimization engine, because it has been specifically designed to calibrate environmental, hydrological and hydrogeological models. In addition, hydroPSO implements the latest Standard Particle Swarm Optimization algorithm (SPSO-2011), with an adaptive random topology and rotational invariance constituting the main advancements over previous PSO versions.

A tracer test conducted in the experimental field at TU Bergakademie Freiberg (Germany) is used as case study. A double-porosity approach is used to simulate the solute transport in the fractured Gneiss aquifer. Tracer concentrations obtained with hydroPSO were in good agreement with its corresponding observations, as measured by a high value of the coefficient of determination and a low sum of squared residuals. Several graphical outputs automatically generated by hydroPSO provided useful insights to assess the quality of the calibration results. It was found that hydroPSO required a small number of model runs to reach the region of the global optimum, and it proved to be both an effective and efficient optimization technique to calibrate the movement of solute transport over time in a fractured aquifer. In addition, the parallel feature of hydroPSO allowed to reduce the total computation time used in the inverse modeling process up to an eighth of the total time required without using that feature. This work provides a first attempt to demonstrate the capability and versatility of hydroPSO to work as an optimizer of a coupled flow and transport model for contaminant migration.



## 1. Introduction and scope

Over the last decades, groundwater has got a major role in sustaining the economy of certain regions, by meeting ever-increasing water demands (Shah, 2006). In particular, fractured aquifers are important water resources around the world (Dietrich et al., 2005), playing a significant role in the development of geothermal resources or the safe disposal of high-level radioactive waste (Hurter and Schellschmidt, 2003).

Flow in a fractured granite and gneiss aquifer is highly heterogeneous over a range of scales (Bonnet et al., 2001) and several uncertainties arise from the heterogeneous flow in the aquifer (Neuman, 2005; Zhang, 2001). This may have significant implications for water resources management, from borehole to catchment scale. In addition, a proper understanding of flow heterogeneity in the aquifer is of great

* Corresponding author.
*E-mail addresses:* abdelazi@mailserver.tu-freiberg.de (R. Abdelaziz), mauricio.zambrano@udec.cl (M. Zambrano-Bigiarini).
[1] Present address: Dept. of Environmental Engineering, Faculty of Environmental Sciences and EULA-Chile Centre, University of Concepción, Concepción, Chile.

http://dx.doi.org/10.1016/j.jconhyd.2014.06.003
0169-7722/

importance for groundwater source protection and for our ability to predict the movement and fate of contaminants in the fractured gneiss (Krupka et al., 1999). Consequently, there is a need for better understanding the flow heterogeneity in fractured aquifers and to be able to predict the nature of this heterogeneity (Krásnỳ and Sharp, 2007). Unfortunately, very often the previous task has to be performed on the basis of a relatively sparse data set.

Traditionally, the heterogeneity of the fractured gneiss has been represented as an Equivalent Porous Medium (EPM) (e.g. Pankow et al., 1986; Scanlon et al., 2003). Flow transport in EPM assumes that diffusion rates are fast relative to the fractured groundwater velocity (Pankow et al., 1986). However, for media where the matrix blocks have a high porosity for fluid storage while the fractures have a high permeability for fluid transport, a double-porosity (DP) medium (Barenblatt et al., 1960) is a more suitable approach (Zhang and Sun, 2000). The DP assumes that the medium can be represented as a superposition of these two systems with different hydraulic and solute transport properties, in both of which the fractures are accounted for implicitly (e.g. Gerke and van Genuchten, 1993a,b; Vogel et al., 2000). The two pore systems interact by exchanging water and solute in response to pressure head and concentration gradients. In the dual-porosity model, the groundwater flow, as well as convection and dispersion of solutes, take place only in the fractures, while diffusion of solutes between the fractures and matrix connects the two systems. The dual-porosity approach is selected in this work, instead of EPM, because it is considered to provide a better representation of the transport process in a fractured gneiss rock.

Several aquifer properties are difficult to be directly measured at the correct spatial scale. Therefore they have to be estimated through inverse modeling or a calibration procedure. Inverse modeling techniques are widely used in groundwater, being PEST (Doherty, 2005, 2013) and UCODE (Poeter et al., 2005) two of the most widespread inverse codes. PEST and UCODE perform inverse modeling by estimating parameter values that minimize a weighted least-squares objective function using non-linear regression (Hill and Tiedeman, 2006). However, the main drawback of all gradient-based techniques such as PEST and UCODE2005 are the dependency of the optimization results upon the initial guess about the solution (initial point of the optimization), which may result in finding a local optimum instead of the global one (Hsu et al., 2011).

Particle Swarm Optimization (PSO) is a global optimization technique introduced by Eberhart and Kennedy in 1995, inspired by the social behavior of bird flocks looking for corn (Eberhart and Kennedy, 1995; Kennedy and Eberhart, 1995). PSO shares some similarities with genetic algorithms (Eberhart and Shi, 1998), but it has no presence of evolutionary operators such as crossover or selection. In the last decades, PSO has been applied in several fields (see Poli et al., 2007), e.g. in structural design (Schutte and Groenwold, 2005), geothermal heating planning (Beck et al., 2010; Ma et al., 2013), finance and economics (Huang et al., 2006), and image and video analysis applications (Slade et al., 2004; Wachowiak et al., 2004). In the water resources field, Sedki and Ouazar (2010) coupled MODFLOW2000/2005 with PSO for parameter estimation purposes; Gill et al. (2006) used a multi-objective PSO to calibrate a well-known conceptual rainfall–runoff model; and Zambrano-Bigiarini and Rojas (2013) used a state-of-the art PSO version to calibrate a surface water model (SWAT-2005) and a groundwater model (MODFLOW-2005). However to the best of our knowledge, PSO has never been used to estimate the parameters of a coupled flow and transport model to simulate the transport of pollutants in a fractured gneiss aquifer. Therefore, this work provides a first attempt to demonstrate the effectiveness and efficiency of PSO for calibrating a coupled flow and transport model.

It is worth to note that several implementations of PSO are available in literature, as OSTRICH (Matott, 2005) and squad/MADS (Vesselinov and Harp, 2012), but most of them implement the basic PSO configuration proposed in 1995. In this work we use hydroPSO (Zambrano-Bigiarini and Rojas, 2013), a model-independent R package that implements the latest Standard Particle Swarm Optimization 2011 (SPSO2011). In addition, hydroPSO was selected because it is stand-alone software specifically designed for hydrological and environmental applications, with user-friendly plotting functions which facilitate the interpretation and assessment of the optimization results. hydroPSO is used here, for the first time, to estimate the parameters of a solute transport model in a gneiss fractured aquifer. The main objective of this paper is to test the capability and versatility of hydroPSO to work as an optimizer of a coupled flow and transport model for contaminant migration.

## 2. Methodology

### 2.1. Study area

The experimental area at the TU Bergakademie Freiberg Campus is located in Saxony, Germany, close to a polymetallic sulfide deposit (see Fig. 1). The Inner Grey Gneiss is a metgranite (Tichomirowa, 2003), which is not only the most common rock type within the Freiberg region but also the oldest rock of the Ore Mountains. The density of the Freiberg Grey gneiss is 2.645 kg $m^{-3}$ as determined by the pycnometer method (Kranz and Dillenardt, 2010).

The Freiberg test field contains six wells that were drilled for research and educational purposes. One of the wells was drilled to a depth of 100 m and 5 wells to a depth of 50 m. The upper part of the boreholes (3 to 5 m) is protected by a steel stand pipe; the deeper part was left as open borehole. A major horizontal fracture zone was determined to be 11 m below the ground surface by geophysical and hydraulic tests (Abdelaziz and Merkel, 2012; Abdelaziz et al., 2013), while packer tests, geophysical logs, and tracer tests were conducted in the test field.

### 2.2. Tracer tests

Tracers may be organic or inorganic constituents in water, small particles or isotopes (Käss et al., 1998). They can be introduced into the environment by natural processes, human activities (e.g. nuclear weapon tests, industrial emissions) or by intention during a tracer experiment (Singhal and Gupta, 2010). Tracer tests are used in surface and groundwater to investigate flow and transport processes in heterogeneous porous media, in particular to estimate effective parameters describing the aquifer system (NRC, 1996; Ptak et al., 2004). Tracer tests may be used in a fractured aquifer to provide a

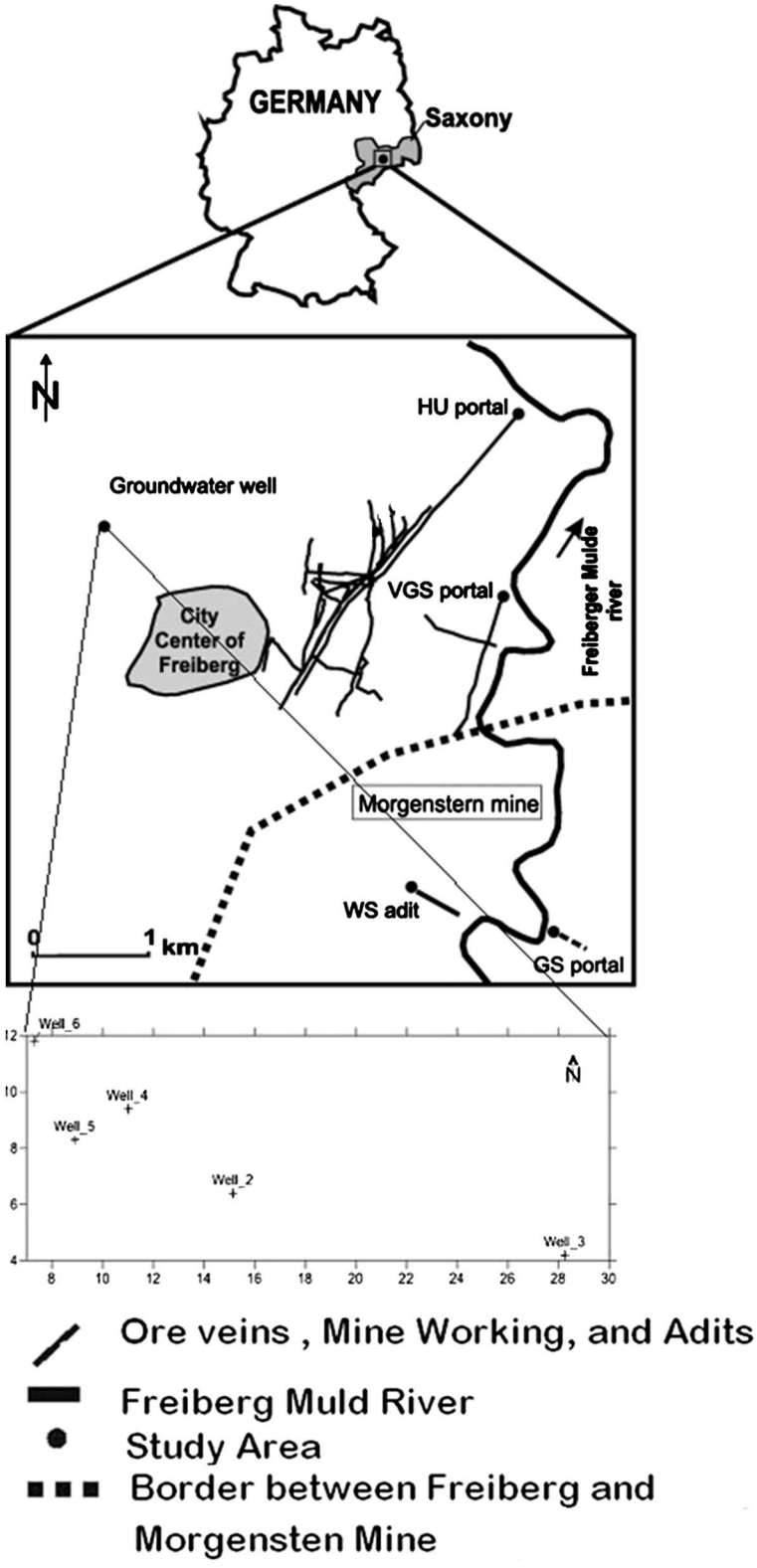


**Fig. 1.** Location of the test site in Freiberg and research wells. Modified after Junghans and Tichomirowa (2009).

better understanding of the connectivity in the fractured aquifer.

Several methods for carrying out tracer tests at field scale are available today, under natural or forced hydraulic gradient conditions (NRC, 1996; Ptak et al., 2004). In the natural gradient method a tracer is injected into the groundwater and moves under natural hydraulic gradient conditions in an undisturbed groundwater flow field (NRC, 1996; Singhal and Gupta, 2010). In the forced gradient method the movement is induced by groundwater pumping or groundwater/tracer solute injection.

Tracer tests can be carried out under steady or transient states, and the interpretation of the test can be done through numerical or analytical analysis. The question that arises is whether there is any ideal tracer for porous and fractured aquifers. Actually, there is no ideal tracer substance for all sites. However, there is always an optimum tracer for a certain site, but it may not necessarily be suitable for another area. The objective of the test plays an important role in the selection of tracer substance, sampling scheme, number of tracers used, tracer properties, and to avoid any complexity in designing the tracer test. A tracer that fits for porous media does not necessarily fits for fractured aquifers, due to the double-porosity behavior of many aquifers (Davis et al., 1980). However, NaCl seems to be an appropriate tracer because of its low cost, conservative character, non-toxicity, and ease to be used and detected (Field, 2002).

The natural gradient technique was used in this research to identify the transport parameters for the fractured zone. The advantage of this technique is its consistency with natural conditions while the disadvantage is the long duration of the experiment. Fig. 2 shows the tracer test with NaCl performed in the fractured Gneiss aquifer, by using a two-packer system to inject the tracer in a single fracture zone. In our experiments, a tracer solution was injected under steady-state conditions, over a period of 6.5 h into the observation well 4, while concentrations were recorded as depth-integrated breakthrough curve at the monitoring well 6 (see Fig. 1). Further information and technical details about the tests and solute analysis are provided in Abdelaziz and Merkel (2012) and Abdelaziz et al. (2013).

### 2.3. Groundwater model

The partial differential equations that govern the solute transport in groundwater are non-linear and have to be solved simultaneously. In this study, the finite-difference model MODFLOW (Harbaugh, 2005) is coupled to the second generation of the transport model MT3D, known as MT3DMS (Zheng, 2006; Zheng and Wang, 1999; Zheng et al., 2010; Abdelaziz and Le, 2014) to implement a three dimensional groundwater flow and solute transport model in a double-porosity fractured aquifer.

MODFLOW simulates both confined and unconfined flow in three-dimensional heterogeneous aquifer systems, while MT3DMS is a three-dimensional model that can simulate multi-species transport of solutes. MT3DS considers advection, dispersion/diffusion, and chemical reactions, by using spatially distributed hydrogeological conditions. MT3DS has been extensively tested, and is known to be robust, accurate, and computationally efficient. MT3DMS was selected to carry out a

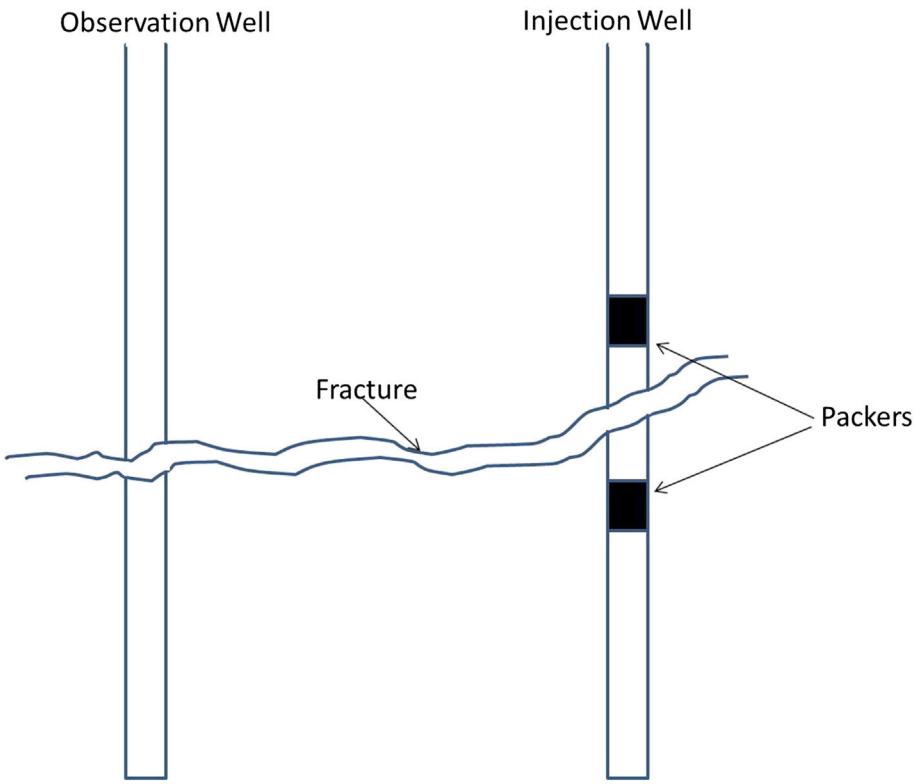


**Fig. 2.** Set up of the tracer test (no scale).

three dimensional simulation of the solute transport model, because it is one of the few three-dimensional and publicly-available model codes implementing the dual-porosity approach directly. Both model codes can use several types of time-dependent boundary conditions and forcing terms, which are implemented with modular structures, allowing users to modify them for any particular application.

Based on the characterization of the study site (Abdelaziz and Merkel, 2012; Abdelaziz et al., 2013) and using the double-porosity approach, the fractured aquifer is represented as three main layers, where fractures have different hydraulic and solute transport properties than the two matrix blocks (Fig. 3). The implementation of the groundwater flow and solute transport model is based upon a three dimensional finite difference grid, considering the geological setting described in Abdelaziz et al. (2013) and Abdelaziz and Merkel (2012). Since the inverse problem is ill-posed and affected by model uncertainties, the parameters used to characterize the spatial distribution of hydrogeological parameters are estimated through calibration. Physical ranges used in the calibration process are based on prior expert knowledge, which is subject of ongoing research. Among the calibration parameters, equivalent hydraulic conductivity (*HK*), porosity (*p*), immobile porosity (*imp*), horizontal transverse dispersivity ratio (*TRPT*), and vertical transverse dispersivity ratio (*TRPV*) are assumed to be unknown for both the matrix blocks and the fractured zone. Longitudinal dispersivity was assumed to a known parameter, with a value of 0.08 [m] (Abdelaziz and Merkel, 2012). Table 1 shows the set of 8 parameters selected to be estimated through inverse modeling and the physical ranges used for calibration, based on prior expert knowledge (Abdelaziz and Le, 2014).

Notwithstanding usually the Darcy's law is not appropriate for modeling flow into fractures, the Reynolds number for both the matrix block and the fractured layer was within the range of validity of the Darcy's law.

### 2.4. Inverse modeling

Parameter estimation is a challenging task for groundwater modelers due to the many uncertainties associated with the conceptual model, observations and model parameters (Carrera et al., 2005), which usually leads to an ill-posed problem (Yeh, 1986). In contrast to gradient-based techniques,

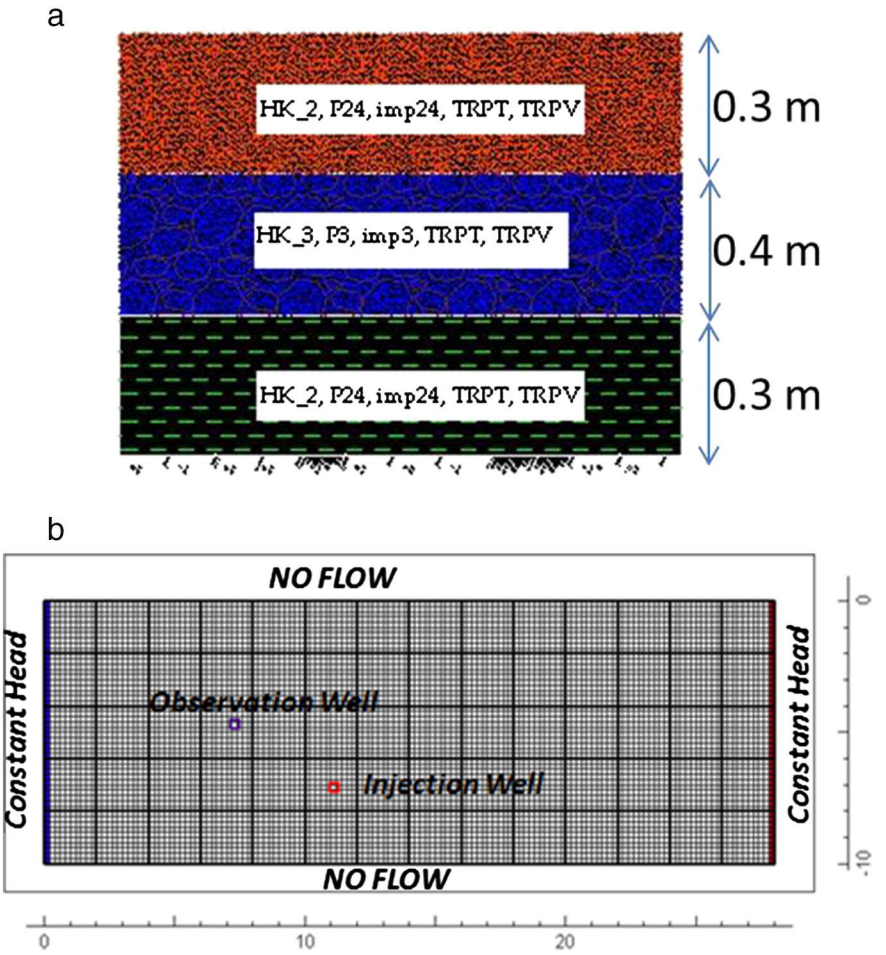


**Fig. 3.** Conceptual model used to represent the groundwater flow and solute transport model in the fractured gneiss aquifer system. (a) Conceptual model used and parameters being calibrated. (b) Boundary conditions selected for the conceptual model.

**Table 1**
Parameters of the flow and transport model to be estimated through inverse modeling. Meaningful physical ranges used for calibration are based on prior expert knowledge.

| Parameter name | Units | ID | Calibration range | | Calibrated |
|---|---|---|---|---|---|
| | | | Min | Max | Value |
| Hydraulic conductivity of the matrix blocks | [m day$^{-1}$] | *HK_2* | 9.00E−01 | 1.55E+00 | 1.02E+00 |
| Hydraulic conductivity of the fractured layer | [m day$^{-1}$] | *HK_3* | 3.70E + 01 | 5.00E+01 | 3.98E+01 |
| Porosity of the matrix blocks | [–] | *P24* | 5.50E−02 | 6.70E−02 | 5.63E−02 |
| Porosity of the fractured layer | [–] | *P3* | 8.70E−02 | 1.20E−01 | 1.06E−01 |
| Horizontal transverse dispersivity ratio | [–] | *TRPT* | 1.80E−01 | 2.10E−01 | 2.03E−01 |
| Vertical transverse dispersivity ratio | [–] | *TRPV* | 2.70E−01 | 3.10E−01 | 2.79E−01 |
| Immobile porosity of the matrix blocks | [–] | *imp24* | 1.80E−02 | 2.30E−02 | 2.04E−02 |
| Immobile porosity of the fractured layer | [–] | *imp3* | 5.00E−02 | 6.00E−02 | 5.14E−02 |

such as the widespread PEST (Doherty, 2005, 2013) and UCODE (Poeter et al., 2005), global-search methods are able to identify parameter values that produce the best fit to observations and prior information, regardless of the degree of model nonlinearity and the presence of local minima (Hill and Tiedeman, 2006).

Within the recently developed global optimization techniques, Particle Swarm Optimization (PSO, (Eberhart and Kennedy, 1995; Kennedy and Eberhart, 1995)) has received a surge of attention given its flexibility, ease of implementation, and efficiency (Poli et al., 2007), and it was selected to calibrate the coupled MODFLOW–MT3DMS solute transport model in the gneiss fractured aquifer. PSO is a population-based technique which shares some similarities with genetic algorithms (GA), but without crossover or selection (Eberhart and Shi, 1998). In PSO, a "swarm" of candidate solutions (called "particles" in PSO terminology) flies around the user-defined parameter space looking for the best model performance. The position of each particle is updated taking into account its own best "position" and the best position in its neighborhood.

Since 1995, a plethora of variants of the canonical PSO algorithm have appeared in the literature (see e.g. Poli et al., 2007), aimed at improving its performance. In particular, the Standard Particle Swarm Optimization 2011 (SPSO-2011) is a major improvement over previous PSO versions (Clerc, 2012), with an adaptive random topology and rotational invariance constituting the main advancements (Clerc, 2012; Zambrano-Bigiarini et al., 2013). For each particle, a random point is defined in the hypersphere $\mathcal{H}\left(\vec{G}_i^t, ||\vec{G}_i^t - \vec{X}_i^t||\right)$,

$$\vec{V}_i^{t+1} = \omega\vec{V}_i^t + \mathcal{H}\left(\vec{G}_i^t, ||\vec{G}_i^t - \vec{X}_i^t||\right) - \vec{X}_i^t \quad (1)$$

while the particle's position is updated following Eq. (2).

$$\vec{X}_i^{t+1} = \vec{X}_i^t + \vec{V}_i^{t+1} \quad (2)$$

where $\vec{V}_i^t$ and $\vec{V}_i^{t+1}$ are the previous and new particle velocities, respectively; $\vec{X}_i^t$ and $\vec{X}_i^{t+1}$ are the previous and new position of each particle, respectively; $\vec{G}_i^t$ is the previous center of gravity of each particle; $i = 1, 2, \ldots, N$, with $N$ equal to the swarm size, and $t = 1, 2, \ldots, T$, with $T$ equal to the maximum number of iterations. $\omega$ is the *inertia weight* that controls the impact of the previous velocity of the particle on its current one.

At each iteration the *swarm radius* ($\delta^t$) is computed as the median Euclidean distance between the position of the swarm optimum ($\vec{G}$) and any given particle ($\vec{x}_i$), as follows (modified after Evers and Ghalia, 2009):

$$\delta^t = median\|\vec{X}_i^t - \vec{G}^t\| \quad (3)$$

where $\|\cdot\|$ denotes the Euclidean norm. Let $diam(\Psi) = ||range_d(\Psi)||$ be the diameter of the search space, then the *normalized swarm radius* ($\delta_{norm}^t$) is used as a measure of convergence, indicating how close particles are to each other:

$$\delta_{norm}^t = \frac{\delta^t}{diam(\Psi)}. \quad (4)$$

Further details about SPSO-2011 can be found on Clerc (2012) and Zambrano-Bigiarini et al. (2013).

SPSO-2011 was run using the open source implementation provided by the hydroPSO R package v0.3-3 (Rojas and Zambrano-Bigiarini, 2012; Zambrano-Bigiarini, 2014; Zambrano-Bigiarini and Rojas, 2013), which is a model- and platform-independent optimization software specifically designed for environmental applications. To run hydroPSO with MODFLOW–MT3DMS only a few basic text files have to be prepared by the user, with no need of template or instruction files as in PEST or UCODE. Two (problem-specific) ASCII files are: (i) ParamFiles.txt, which defines the names and location of the parameters to be calibrated and, (ii) ParamRanges.txt, which provides meaningful parameter ranges for the calibration. In addition, the name of the executable model code and instructions for reading model outputs have to be provided in a separated R script (R. Core Team, 2013), which is easily adapted from an existing template provided with hydroPSO (Rojas and Zambrano-Bigiarini, 2012). For this work, a batch file was used for running in sequence MODFLOW and MT3DMS, while a short FORTRAN code was written to speed up the reading of parameters being optimized (preproc.exe). Fig. 4 describes the main files used for coupling the MODFLOW and MT3DMS model with hydroPSO.

For optimization, the sum of the squared residuals (SSR, Eq. (5)) was selected to measure the goodness-of-fit (GoF) value between observed concentrations ($C_i^o$) and its corresponding model outputs ($C_i^s$) at each time step. After some preliminary trials, the maximum number of iterations $T$ was

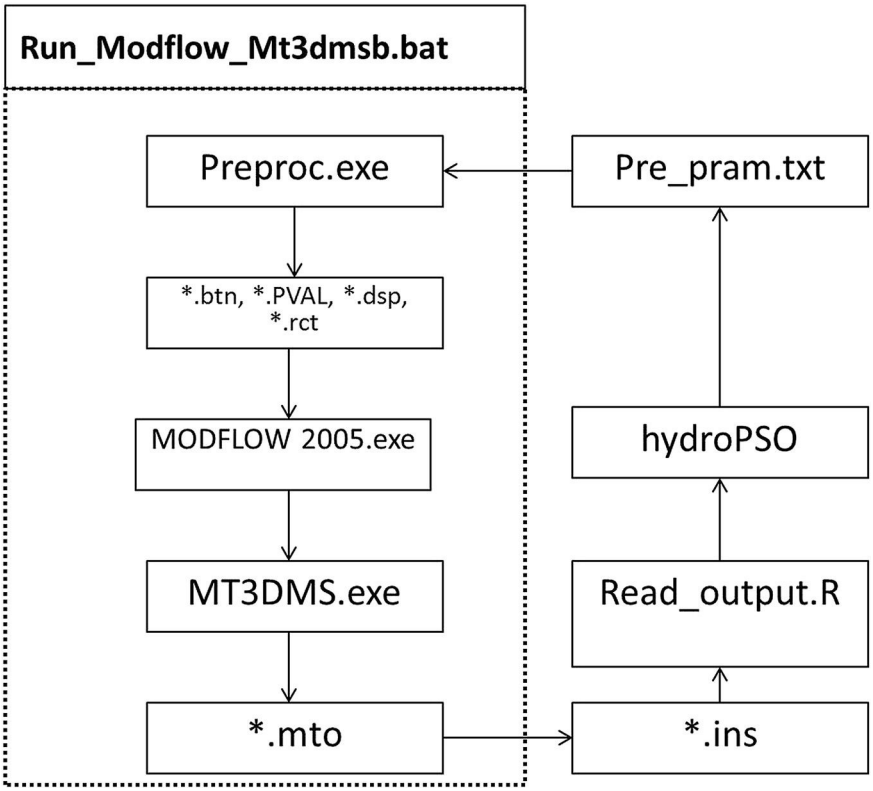


**Fig. 4.** Flow chart showing the functions and files involved in the optimization of the groundwater flow and solute transport model. *.xxx stands by all the input files with extension xxx. In particular, *.ins are all the instruction files, *.mto are all the files with simulated concentrations, while *.dsp, *.btn, and *.rct represent the setting up the transport model, and *. PVAL is the input file with flow parameters.

set to 35 and the number of particles was set to 8 (i.e., 280 model runs), while other PSO arguments were kept in the default values provided by hydroPSO. Parallel computations were activated in order to accelerate the optimization process in an Intel® Core™ i7-875K machine (four cores and 8 threads), 2.93 GHz.

$$SSR = \sum_{i=1}^{n} \left(C_i^s - C_i^o\right)^2 \quad (5)$$

## 3. Results and discussion

One of the most common ways to measure the performance of any model is by plotting the simulated values against its corresponding observations. Fig. 5a and b show breakthrough curves and a scatter plot, respectively, for the best model output obtained with hydroPSO and the observed concentrations at observation well 6. Fig. 5b depicts a good correspondence between the monitored solute concentration and their corresponding best simulated values, with a small over-estimation of low and high concentration values, and a slight underestimation of medium concentrations. In addition, the coefficient of determination ($R^2$) is equal to 0.933, indicating that a high proportion of the variance of the response variable is explained by the model.

Fig. 6 shows the evolution of the *global optimum* (best model performance for a given iteration, i.e., smallest SSR) and the *normalized swarm radius* ($\delta_{norm}$, a measure of swarm spread on the search space) against the number of iterations. It is clear that both the global optimum and the normalized swarm radius decrease with increasing number of iterations, indicating that most of the particles converged into a small region of the solution space. Moreover, Fig. 6 shows that only 5 iterations (i.e., $5 \times 40 = 200$ model runs) were required for finding the region of the global optimum, and the remaining iterations were just used for refining the search. The optimum value found for the sum of the squared residuals was 2.97E+03 [$mg^2\ l^{-2}$].

Fig. 5a shows that the breakthrough curve is positively skewed with a rapid rise to the peak, which lies after the mean tracer transit time. The long tail on the right is assumed to be due to dead end pores and retention of the tracer in the fractured rock (Abdelaziz et al., 2013). The long tail may also be linked to low hydraulic conductivity in some parts of the fractured rock. In addition, solute transport in the matrix block plays a major role on the diffusive transport of the tracer. Immobile porosity has a major impact on the late arrival of the tracer, but only a minor effect on the peak of the breakthrough curve. Further information on the analysis of the breakthrough curve can be found in Abdelaziz et al. (2013).

Boxplots in Fig. 7a are useful non-parametric diagrams for summarizing the statistical distribution of the sampled values. The top and bottom of the box show the first and third quartiles, respectively, while the horizontal line inside the box represents the second quartile (the median). The notches extend to $+/-1.58 \cdot IQR/sqrt(n)$, where $IQR$ is the interquartile range and $n$ is the number of points. Finally,

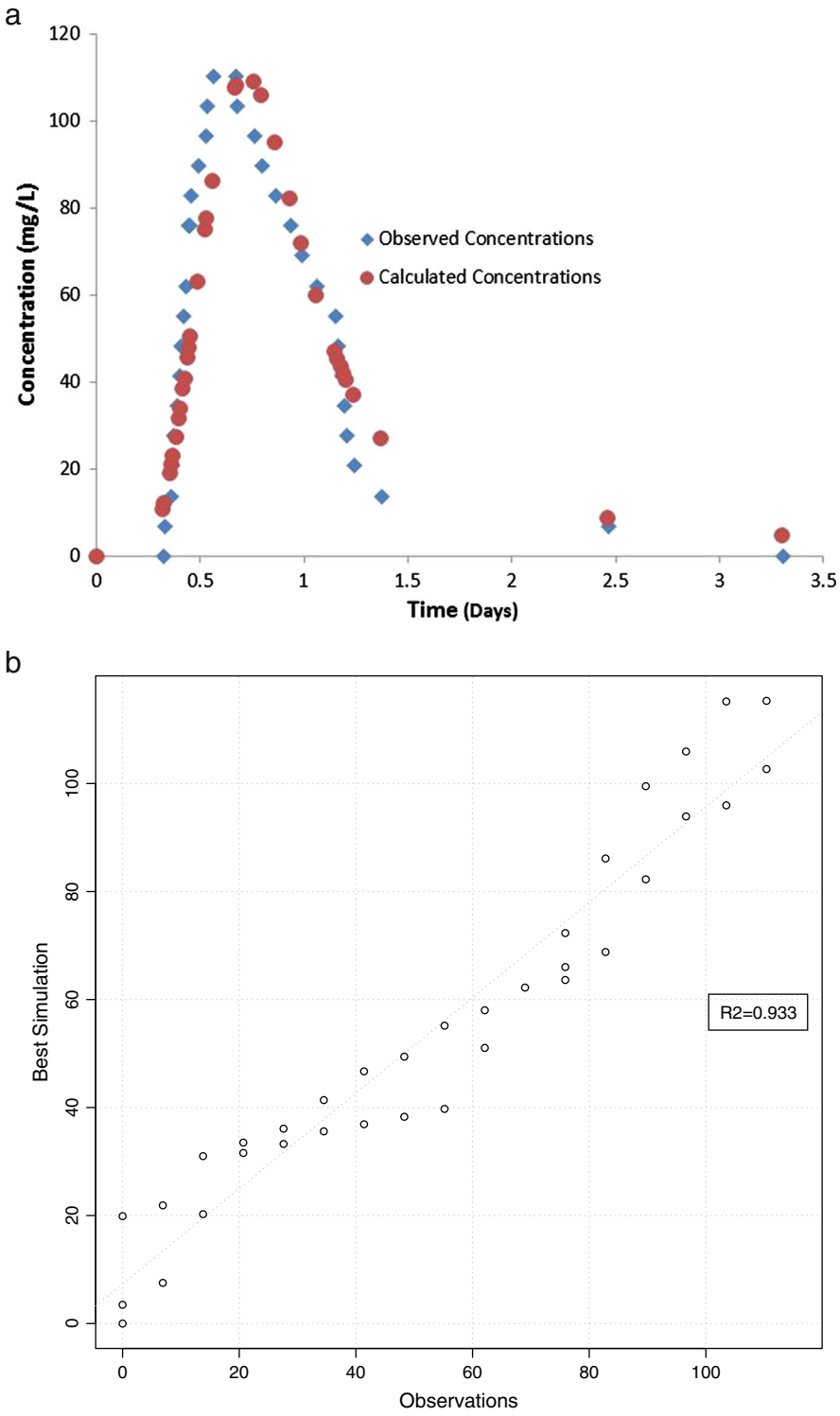


**Fig. 5.** Comparison of observed tracer concentrations at observation well 6 with the corresponding best simulated values obtained with hydroPSO. (a) Observed and best simulated breakthrough curves. (b) Scatter plot with the best model output and the observed concentrations.

points outside the notches are considered to be *outliers*. Dotty plots in Fig. 7b show the parameter values against its corresponding goodness-of-fit value (SSR) obtained during the optimization. They are useful for identifying parameter ranges that produce the best model performance or with roughly the same model performance (*equifinality*, Beven and

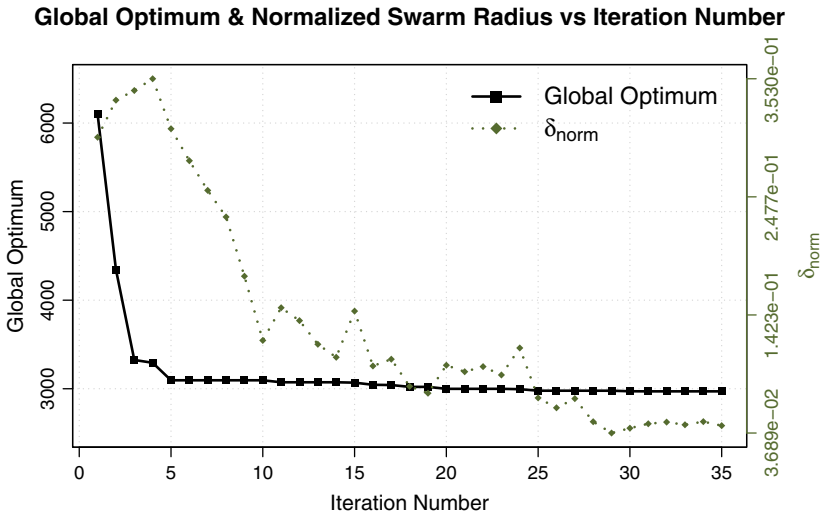


**Fig. 6.** Evolution of the global optimum (best sum of squared residuals) and the normalized swarm radius ($\delta_{norm}$, a measure of swarm spread on the search space) against the number of iterations.

Binley, 1992). Visual inspection of Fig. 7a shows that for 5 out of 8 parameters (*HK_2*, *HK3*, *P3*, *imp24* and *imp3*) the optimum value found during the optimization coincides with the median of all the sampled values, while for *TRPT* and *TRPV* those values were very close, confirming that most of the particles converged into a small region of the solution space. On the other hand, Fig. 7b shows that the optimum values found for parameters *HK_3* and *P3* are really well defined, whereas *HK_2*, *imp24* and *imp3* show a wider region around their optimum.

Histograms in Fig. 8a are a graphical representation of sampled frequencies for each parameter, and are used to provide an estimate of the probability density function thereof. In addition, empirical cumulative distribution functions (ECDFs) in Fig. 8b are used to estimate the true underlying cumulative distribution function (CDF) of the sampled points. Both figures confirm that parameters *P3* and *imp24* follow a near-normal distribution, while only *imp3* shows a sampled distribution highly skewed towards the lower boundary used for calibration.

Fig. 9 shows (quasi)-three dimensional dotty plots, which summarize interactions among parameters by projecting the SSR response surface onto the parameter space (for different pairs of parameters). In this figure low model performance is represented with dark-red color while high model performance is shown with light-blue color. In general, it can be seen that particles are spread all over the parameter space, indicating a good exploratory capability of PSO. However, regions with bad model performance have a low density of points while regions with better model performance are more densely sampled, confirming the good ability to exploitation of PSO. This figure shows that the optimum values found for *HK_3* and *P3* define a narrow range of the parameter space with high model performance. On the other hand, the coupled flow and transport model can achieve a good model performance for a wide range of values of *imp24*, confirming that this parameter is not well identified.

Finally, a correlation matrix among parameters values and model performance (SSR) is shown in Fig. 10. The panel above the diagonal shows the numeric (absolute) value of the Pearson's product moment correlation coefficient between paired samples plus its statistical significance as stars. The bottom panel shows bivariate scatterplots between each column and row of the matrix, with a fitted line obtained using locally-weighted polynomial regression (*loess*, Cleveland, 1979). On the diagonal, a histogram of each variable sampled during the optimization. This figure shows that the highest (linear) correlation between the measure of model performance (SSR) is obtained for the hydraulic conductivity (xtitHK_3) and immobile porosity (*imp3*) of the fractured layer, followed by the hydraulic conductivity of the matrix blocks *HK_2*. In addition, a high (linear) correlation is observed between the hydraulic conductivity of the matrix blocks *HK_2* and *imp3*; and the porosity of the matrix blocks *P24* and the vertical transverse dispersivity ratio *TRPV*. The same figure shows that nonlinear relations are the dominating behavior between most of the calibrated parameters, e.g. *HK_2* vs *P24*, *HK_2* vs *imp3*, and *HK_3* vs *imP3*.

The transport of contaminants in the fractured aquifer is represented in MT3DMS by a convection–dispersion reaction (CDR) equation, whose solution is affected by several well-known problems of oscillation and artificial dispersion. In particular, even small negative concentration values may lead to nonlinear amplification of the reaction source/sink term of the coupled reactive-transport model, resulting in physically erroneous and unstable results (Herrera and Valocchi, 2006). In this work, the advective term of the CDR equation was solved by using the Generalized Conjugate-Gradient (GCG) method, with upstream third-order total-variation-diminishing (TVD) scheme, because it helps to diminish the artificial oscillations. The initial transport step size (*DT0*) was specified as zero in the Basic Transport Package (BTN). Therefore, for each model run *DT0* was automatically computed by MT3DMS to meet various stability constraints

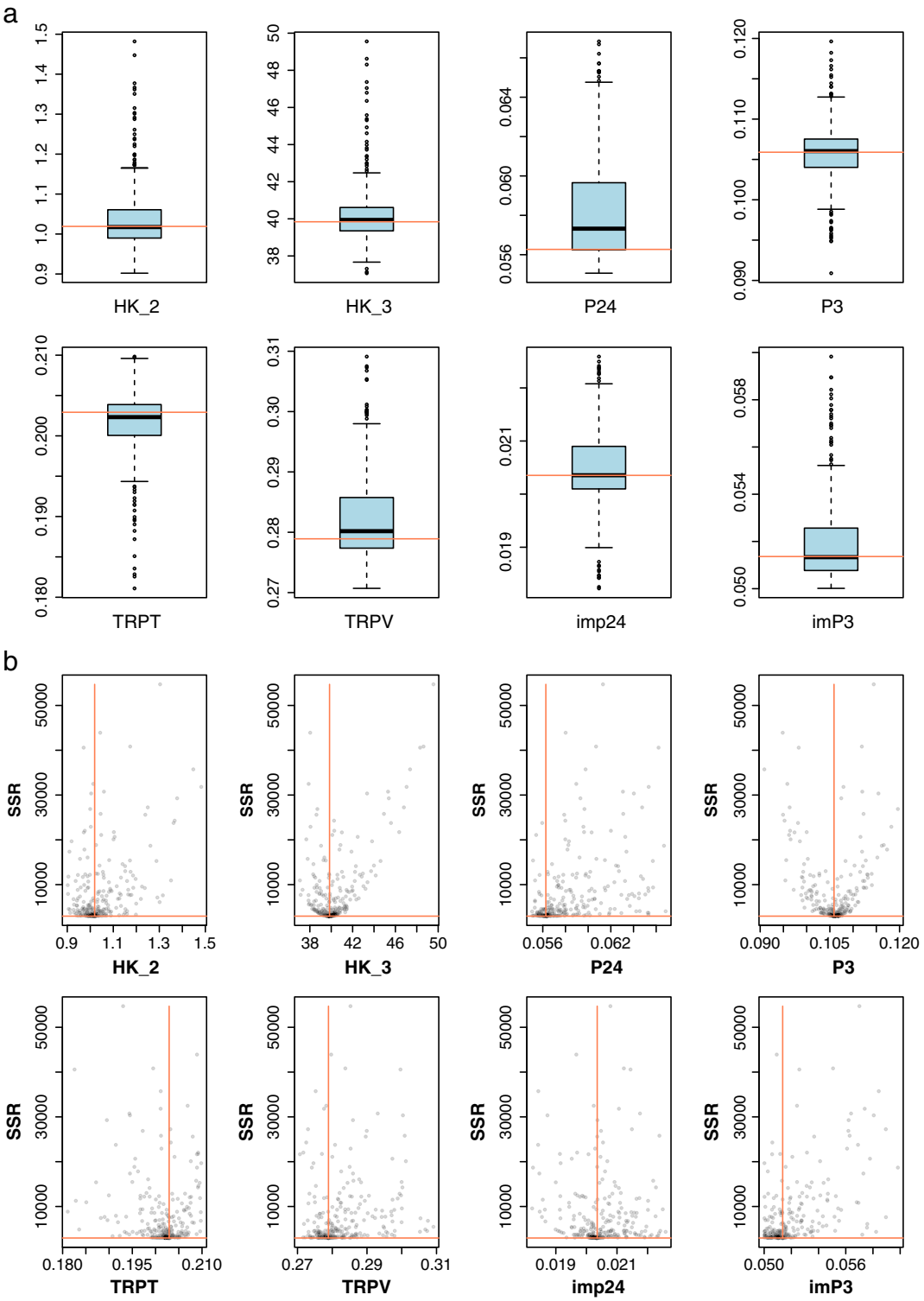


**Fig. 7.** Graphical summary of parameter values sampled during the optimization. (a) Box-and-whisker plots (or boxplots). Horizontal red line indicates the *optimum* value found for each parameter. (b) Parameter values against its corresponding goodness of fit (SSR). Vertical red line indicates the *optimum* value found for each parameter.

that are solution dependent, using a maximum Courant number equal to 1.0. In spite of using the GCG–TVD solver, oscillation and artificial dispersion problems were detected in our simulations, posing an additional difficulty to allow hydroPSO to find a (near-)optimum parameter set. For that reason, we suggest to other global optimization users to define –

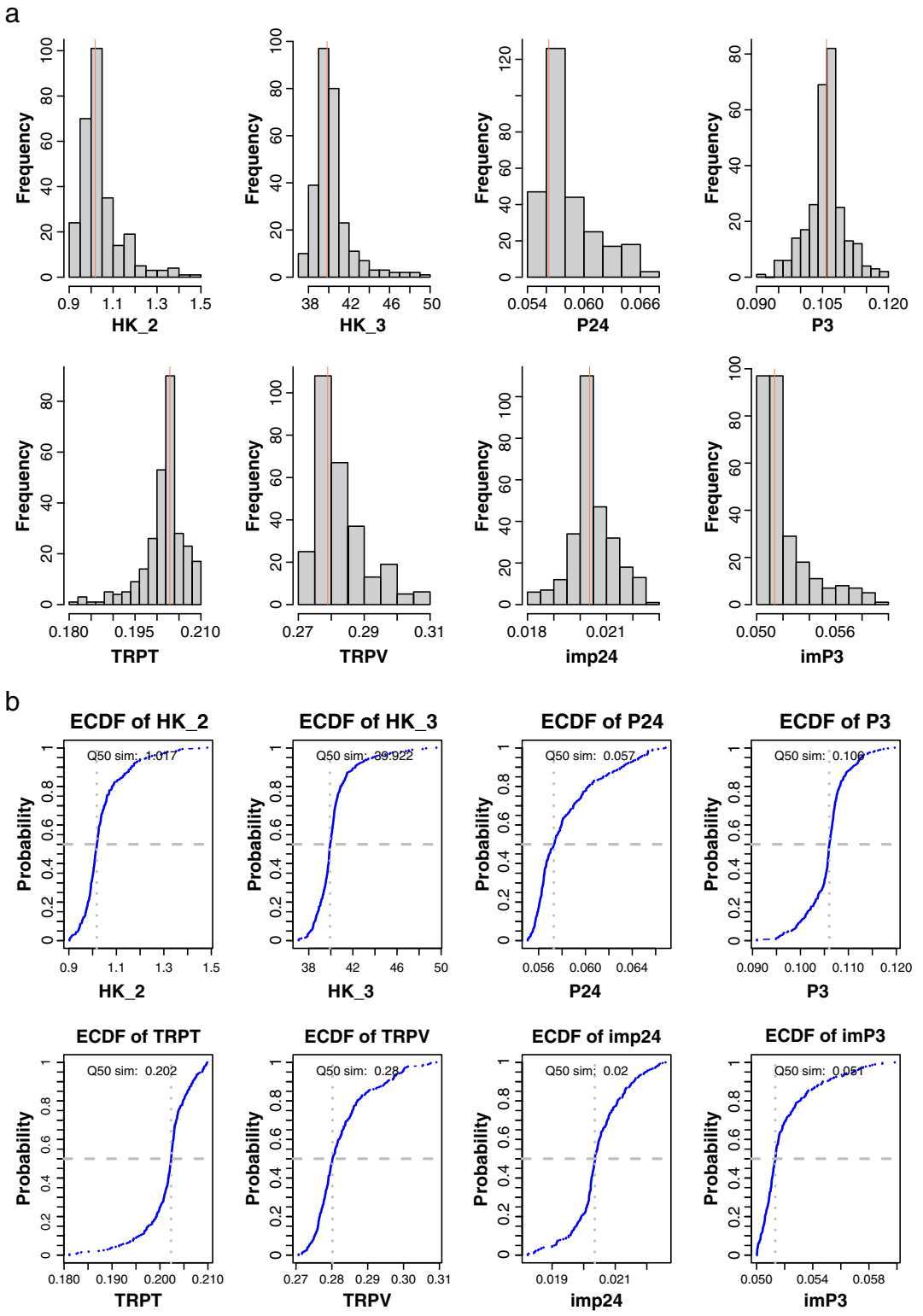


**Fig. 8.** Graphical summary of parameter values sampled during the optimization. (a) Histograms showing the frequencies of the parameter values. Vertical red line indicates the *optimum* value found for each parameter. (b) ECDFs of parameter values. Horizontal gray dotted lines represent a cumulative probability equal to 0.5 (median of the distribution). Vertical gray dotted lines represent a cumulative probability of 0.5 (its value is shown in the upper part of each figure).

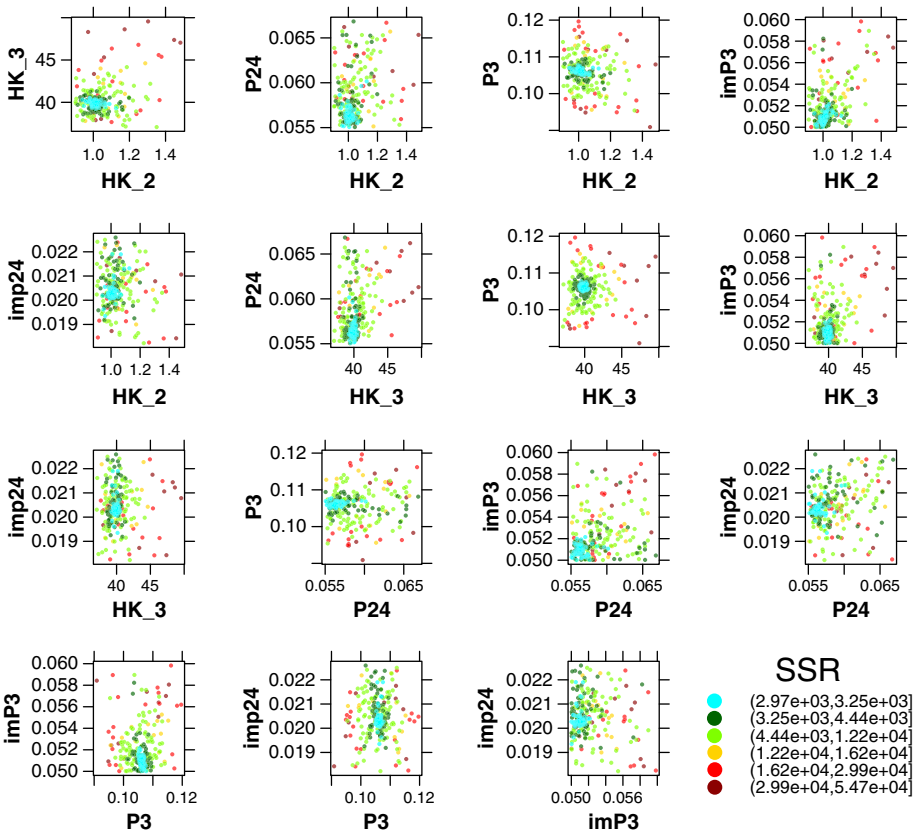


**Fig. 9.** Model performance (SSR) projected onto the parameter space (for different pairs of parameters). Low model performance is represented with dark-red color while high model performance is shown with light-blue color. Parameter names are described in Table 1. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

a priori – a small value for *DT0*, in order to avoid instability during the optimization.

Comparing the results shown in this section with previous work done by Abdelaziz et al. (2013), where Lattice Boltzmann was implemented for the same case, we found that the use of hydroPSO provided a great improvement in the simulated concentration values. Therefore, using a global optimizer as hydroPSO proved to be important for simulating the coupled flow and solute transport mechanism. This approach opens promising possibilities for similar groundwater applications.

All the optimization was run taking advantage of 8 threads available in the Intel® Core™ i7-875K machine used for the experiments. A single model run of the coupled flow and transport model takes 14 min, and therefore the 280 model runs involved in the optimization (8 particles × 35 iterations) would have taken 2.72 days (65.33 h or 3920 min). However, setting parallel = "parallelWin" and nnodes=8 in hydroPSO reduced the total computation time up to 8.19 h, what represents (roughly) an eighth of the total time required without using the parallel argument. Therefore, hydroPSO allowed a significant decrease of the total computation time required for the inverse modeling process.

## 4. Conclusions

Despite the challenges involved in the representation of fractured rocks due to the heterogeneity and uncertainty in the fracture properties, the double-domain approach is still capable of making a reasonable and accurate prediction of the plume distribution.

The hydroPSO R package was successfully applied to calibrate a coupled groundwater flow and solute transport model for conservative tracer tests. hydroPSO proved to be both an efficient and effective tool for inverse modeling of the coupled MODFLOW–MT3DMS model, allowing a quick assessment of the calibration results.

The main conclusions and recommendations are as follows:

- A high value of the coefficient of determination ($R^2 = 0.93$) and relatively low sum of squared residuals (2.97E+03 [$mg^2\ l^{-2}$]) indicate that simulated values obtained after calibration matched the corresponding observed concentrations quite well.
- The parallel feature of hydroPSO allowed to reduce the total computation time used in the inverse modeling process up to an eighth of the total time required without using the parallel argument.
- Solver packages for flow (MODFLOW) and solute transport (MT3DMS) should be choosing carefully because their set up plays an important role in the stability thereof.
- The initial transport step size (*DT0*) of MT3DMS is also an important (indirect) parameter to stabilize hydroPSO, and it should be chosen as small as possible.

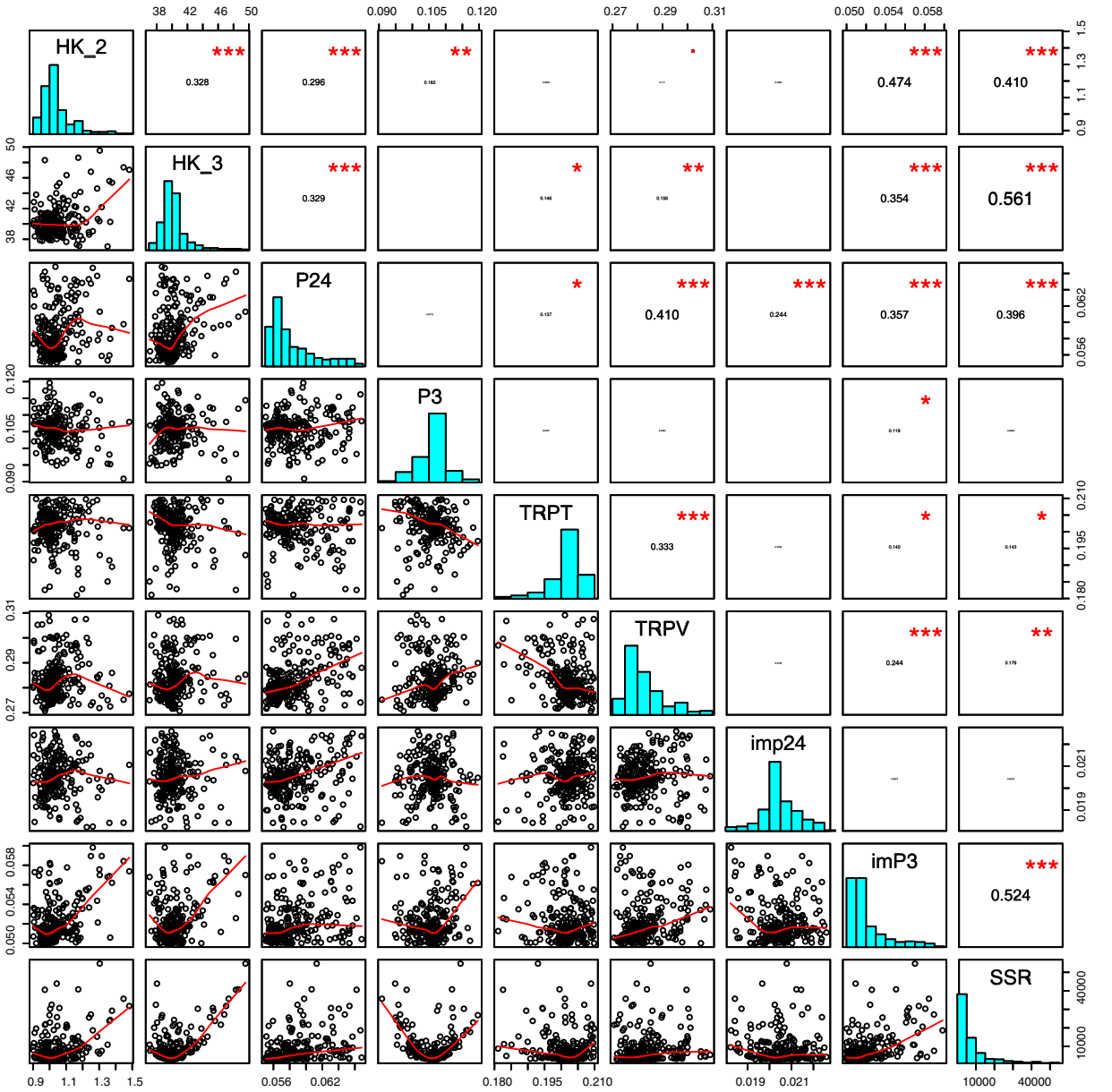


**Fig. 10.** Correlation matrix between parameters and model performance (SSR). Horizontal and vertical axes show the physical range used for the calibration of each parameter.

## Acknowledgments

The authors thank the associate editor and two anonymous reviewers whose constructive comments significantly improved the original manuscript. The authors wish to thank Rodrigo Rojas for technical support and Prof. Broder J. Merkel for his help and advice.

## References

Abdelaziz, R., Merkel, B.J., 2012. Analytical and numerical modeling of flow in a fractured gneiss aquifer. J. Water Res. Prot. 4 (8). http://dx.doi.org/10.4236/jwarp.2012.48076.

Abdelaziz, R., Le, H.H., 2014. Sensitivity analysis of transport modeling in a fractured gneiss aquifer. J. Afr. Earth Sci. http://dx.doi.org/10.1016/j.jafrearsci.2014.06.006.

Abdelaziz, R., Pearson, A.J., Merkel, B.J., 2013. Lattice boltzmann modeling for tracer test analysis in a fractured gneiss aquifer. Nat. Sci. 5, 368–374. http://dx.doi.org/10.4236/ns.2013.53050.

Beck, M., Hecht-Méndez, J., de Paly, M., Bayer, P., Blum, P., Zell, A., 2010. Optimization of the energy extraction of a shallow geothermal system. 2010 IEEE Congress on Evolutionary Computation (CEC), pp. 1–7.

Barenblatt, G.I., Zheltov, Iu P., Kochina, I.N., 1960. Basic concepts in the theory of seepage of homogeneous liquids in fissured rocks. J. Appl. Math. Mech 24 (5), 1286–1303.

Beven, K.J., Binley, A., 1992. The future of distributed models — model calibration and uncertainty prediction. Hydrol. Process. 6, 279–298.

Bonnet, E., Bour, O., Odling, N., Davy, P., Main, I., Cowie, P., Berkowitz, B., 2001. Scaling of fracture systems in geological media. Rev. Geophys. 39, 347–383.

Carrera, J., Alcolea, A., Medina, A., Hidalgo, J., Slooten, L.J., 2005. Inverse problem in hydrogeology. Hydrogeol. J. 13, 206–222.

Clerc, M., 2012. Standard Particle Swarm Optimisation. Technical Report. Particle Swarm Central. http://clerc.maurice.free.fr/pso/SPSO_descriptions.pdf (Online. Last accessed 24-Sep-2012).

Cleveland, W.S., 1979. Robust locally weighted regression and smoothing scatterplots. J. Am. Stat. Assoc. 74, 829–836.

Davis, S.N., Thompson, G.M., Bentley, H.W., Stiles, G., 1980. Ground-water tracers — a short review. Ground Water 18, 14–23.
Dietrich, P., Helmig, R., Sauter, M., Hötzl, H., Köngeter, J., Teutsch, G., 2005. Flow and Transport in Fractured Porous Media. Springer.
Doherty, J., 2005. PEST: model-independent parameter estimation, user manual, Technical Report5th ed. Watermark Numerical Computing, Brisbane, Queensl., Australia.
Doherty, J., 2013. Addendum to the PEST manual. Technical ReportWatermark Numerical Computing, Brisbane, Queensl., Australia.
Eberhart, R., Kennedy, J., 1995. A new optimizer using particle swarm theory. Proceedings of the Sixth International Symposium on Micro Machine and Human Science, 1995. MHS'95, pp. 39–43.
Eberhart, R.C., Shi, Y., 1998. Comparison between genetic algorithms and particle swarm optimization. In: Porto, V., Saravanan, N., Waagen, D., Eiben, A. (Eds.), Evolutionary Programming VII, vol. 1447. Springer Berlin, Heidelberg, pp. 611–616.
Evers, G.I., Ghalia, M.B., 2009. Regrouping particle swarm optimization: a new global optimization algorithm with improved performance consistency across benchmarks. IEEE International Conference on Systems, Man and Cybernetics, 2009. SMC 2009, pp. 3901–3908.
Field, M.S., 2002. The QTRACER2 Program for Tracer–Breakthrough Curve Analysis for Tracer Tests in Karstic Aquifers and Other Hydrologic Systems. National Center for Environmental Assessment — Washington Office, Office of Research and Development, US Environmental Protection Agency.
Gerke, H., van Genuchten, M., 1993a. A dual-porosity model for simulating the preferential movement of water and solutes in structured porous media. Water Resour. Res. 29, 305–319.
Gerke, H., van Genuchten, M., 1993b. Evaluation of a first-order water transfer term for variably saturated dual-porosity flow models. Water Resour. Res. 29, 1225–1238.
Gill, M.K., Kaheil, Y.H., Khalil, A., McKee, M., Bastidas, L., 2006. Multiobjective particle swarm optimization for parameter estimation in hydrology. Water Resour. Res. 42, W07417.
Harbaugh, A.W., 2005. MODFLOW-2005, the US Geological Survey Modular Ground-water Model: The Ground-water Flow Process. US Department of the Interior, US Geological Survey.
Herrera, P., Valocchi, A., 2006. Positive solution of two-dimensional solute transport in heterogeneous aquifers. Ground Water 44, 803–813.
Hill, M.C., Tiedeman, C.R., 2006. Effective Groundwater Model Calibration: With Analysis of Data, Sensitivities, Predictions, and Uncertainty. John Wiley & Sons.
Hsu, K., Chang, L., Jung, C., Huang, C., Chen, J., Tsai, P., Chen, Y., Chang, P., 2011. Applying expert system on the development of a robust model for groundwater parameter identification. AGU Fall Meeting Abstracts, p. 1386.
Huang, F.Y., Li, R.J., Liu, H.X., Li, R., 2006. A modified particle swarm algorithm combined with fuzzy neural network with application to financial risk early warning. IEEE Asia-Pacific Conference on IEEE Services Computing, 2006. APSCC'06, pp. 168–173.
Hurter, S., Schellschmidt, R., 2003. Atlas of geothermal resources in Europe. Geothermics 32, 779–787.
Junghans, M., Tichomirowa, M., 2009. Using sulfur and oxygen isotope data for sulfide oxidation assessment in the Freiberg polymetallic sulfide mine. Appl. Geochem. 24, 2034–2050.
Käss, W., Behrens, H., Himmelsbach, T., Hötzl, H., 1998. Tracing Technique in Geohydrology. Balkema Rotterdam.
Kennedy, J., Eberhart, R., 1995. Particle swarm optimization, in: neural networks, 1995. Proceedings. IEEE International Conference on Neural Networks, pp. 1942–1948.
Kranz, K., Dillenardt, J., 2010. Mine water utilization for geothermal purposes in Freiberg, Germany: determination of hydrogeological and thermophysical rock parameters. Mine Water Environ. 29, 68–76.
Krásnỳ, J., Sharp, J.M., 2007. Groundwater in Fractured Rocks: IAH Selected Paper Series, vol. 9. CRC Press.
Krupka, K., Kaplan, D., Whelan, G., Serne, R., Mattigod, S., 1999. Understanding variation in partition coefficient, Kd, values. Review of Geochemistry and Available Kd Values, for Cadmium, Cesium, Chromium, Lead, Plutonium, Radon, Strontium, Thorium, Tritium (3H), and Uranium, vol. II. EPA.
Ma, R.J., Yu, N.Y., Hu, J.Y., 2013. Application of particle swarm optimization algorithm in the heating system planning problem. Sci. World J. 11. http://dx.doi.org/10.1155/2013/718345.
Matott, L., 2005. Ostrich: An Optimization Software Tool, Documentation and User's Guide, Version 1.6. Department of Civil, Structural and Environmental Engineering, University at Buffalo, Buffalo, NY.
Neuman, S.P., 2005. Trends, prospects and challenges in quantifying flow and transport through fractured rocks. Hydrogeol. J. 13, 124–147.
NRC, 1996. Rock Fractures and Fluid Flow: Contemporary Understanding and Applications. National Academies Press.
Pankow, J.F., Johnson, R.L., Hewetson, J.P., Cherry, J.A., 1986. An evaluation of contaminant migration patterns at two waste disposal sites on fractured porous media in terms of the equivalent porous medium (EPM) model. J. Contam. Hydrol. 1, 65–76.
Poeter, E., Hill, M., Banta, E., Mehl, S., Christensen, S., 2005. UCODE 2005 and six other computer codes for universal sensitivity analysis, calibration, and uncertainty evaluation. US Geological Survey Techniques and Methods, vol. 6-A11.
Poli, R., Kennedy, J., Blackwell, T., 2007. Particle swarm optimization: an overview. Swarm Intell. 1, 33–57.
Ptak, T., Piepenbrink, M., Martac, E., 2004. Tracer tests for the investigation of heterogeneous porous media and stochastic modelling of flow and transport — a review of some recent developments. J. Hydrol. 294, 122–163.
R. Core Team, 2013. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.
Rojas, R., Zambrano-Bigiarini, M., 2012. Tutorial for interfacing hydroPSO with SWAT-2005 and MODFLOW-2005. Technical Report (Online. Last accessed 03-Feb-2014).
Scanlon, B.R., Mace, R.E., Barrett, M.E., Smith, B., 2003. Can we simulate regional groundwater flow in a karst system using equivalent porous media models? Case study, Barton Springs Edwards aquifer, USA. J. Hydrol. 276, 137–158.
Schutte, J.F., Groenwold, A.A., 2005. A study of global optimization using particle swarms. J. Glob. Optim. 31, 93–108.
Sedki, A., Ouazar, D., 2010. Swarm Intelligence for Groundwater Management Optimization.
Shah, T., 2006. Groundwater and human development: challenges and opportunities in livelihoods and environment. Groundwater Research and Management: Integrating Science Into Management Decisions.
Singhal, B., Gupta, R.P., 2010. Applied Hydrogeology of Fractured Rocks, second edition. Springer.
Slade, W.H., Ressom, H.W., Musavi, M.T., Miller, R.L., 2004. Inversion of ocean color observations using particle swarm optimization. IEEE Trans. Geosci. Remote Sens. 42, 1915–1923.
Tichomirowa, M., 2003. Die Gneise des Erzgebirges-hochmetamorphe Äquivalente von neoproterozoisch-frühpaläozoischen Grauwacken und Granitoiden der Cadomiden. Freib. Forsch. C 495.
Vesselinov, V.V., Harp, D.R., 2012. Adaptive hybrid optimization strategy for calibration and parameter estimation of physical process models. Comput. Geosci. 49, 10–20.
Vogel, T., Gerke, H., Zhang, R., Van Genuchten, M.T., 2000. Modeling flow and transport in a two-dimensional dual-permeability system with spatially variable hydraulic properties. J. Hydrol. 238, 78–89.
Wachowiak, M.P., Smolíková, R., Zheng, Y., Zurada, J.M., Elmaghraby, A.S., 2004. An approach to multimodal biomedical image registration utilizing particle swarm optimization. IEEE Trans. Evol. Comput. 8, 289–301.
Yeh, W.W., 1986. Review of parameter identification procedures in groundwater hydrology: the inverse problem. Water Resour. Res. 22, 95–108.
Zambrano-Bigiarini, M., 2014. hydroPSO: Particle Swarm Optimisation, With Focus on Environmental Models (R package version 0.3-3).
Zambrano-Bigiarini, M., Rojas, R., 2013. A model-independent Particle Swarm Optimisation software for model calibration. Environ. Model Softw. 43, 5–25.
Zambrano-Bigiarini, M., Clerc, M., Rojas, R., 2013. Standard Particle Swarm Optimisation 2011 at CEC-2013: a baseline for future PSO improvements. 2013 IEEE Congress on Evolutionary Computation. IEEE-INST Electrical Electronics Engineers Inc., Piscataway, USA, pp. 2337–2344.
Zhang, D., 2001. Stochastic Methods for Flow in Porous Media: Coping With Uncertainties. Academic press.
Zhang, D., Sun, A.Y., 2000. Stochastic analysis of transient saturated flow through heterogeneous fractured porous media: a double-permeability approach. Water Resour. Res. 36, 865–874.
Zheng, C., 2006. MT3DMS v5.3: A Modular Three-dimensional Multispecies Transport Model for Simulation of Advection, Dispersion and Chemical Reactions of Contaminants in Groundwater Systems. Supplemental User's Guide Software Management Plan for MODIFLOW and Related Codes CHPRC-00258 Rev 0.
Zheng, C., Wang, P.P., 1999. MT3DMS: a modular three-dimensional multispecies transport model for simulation of advection, dispersion, and chemical reactions of contaminants in groundwater systems; documentation and user's guide. Technical Report. DTIC Document.
Zheng, C., Weaver, J., Tonkin, M., 2010. MT3DMS, A Modular Three-dimensional Multispecies Transport Model — User Guide to the Hydrocarbon Spill Source (hss) Package. Prepared under contract to the US Environmental Protection Agency, Athens, Georgia.