Particle swarm optimization for the estimation of surface complexation constants with the geochemical model PHREEQC-3.1.2

Ramadan Abdelaziz[1,2], Broder J. Merkel[2], Mauricio Zambrano-Bigiarini[3], Sreejesh Nair[4]

[1] *College Of Engineering, A'Sharqiyah University, Oman*

[2] *TU Bergakademie Freiberg, Germany*

[3] *Department of Civil Engineering, Universidad de La Frontera, Chile*

[4] *Institute of Environmental Physics, University of Bremen, Germany*

Email: ramawaad@gmail.com, ramadan.abdelaziz@asu.edu.om

Abstract

Sorption of metals on minerals is a key process in treatment water, natural aquatic environments, and other water related technologies. Sorption processes are usually simulated with surface complexation models; however, identifying numeric values for the thermodynamic constants from batch experiments requires a robust parameter estimation technique that does not get trapped into local minima. Recently, Particle Swarm Optimization (PSO) techniques have attracted many researchers to optimize model parameters in several fields of research. In this work, uranium at low concentrations was sorbed on quartz at different pH, and the hydroPSO R optimization package was used -by the first time- to calibrate the PHREEQC geochemical model, version 3.1.2. . Results show that thermodynamic parameter values identified with hydroPSO are more reliable than those identified with the well-known parameter estimation (PEST) software, when both parameter estimation software are coupled to PHREEQC using the same thermodynamic input data. In addition, post-processing tools included in hydroPSO were helpful for the correct interpretation of uncertainty in the obtained model parameters and simulated values. Thus, hydroPSO proved to be an efficient and versatile optimization tool for identifying reliable thermodynamic parameter values of the PHREEQC geochemical model.

Keywords: particle swam optimization; hydroPSO; PHREEQC; surface complexation; uranium; sorption

Introduction and Scope

Particle Swarm Optimization technique (PSO) is an evolutionary optimization technique proposed by Eberhart and Kennedy (1995) and was influenced by the activities of flocks of birds in search of corn

(Kennedy and Eberhart, 1995; Eberhart and Kennedy, 1995). Both PSO and genetic algorithms (GA) shares a few similarities (Eberhart and Shi, 1998). GA have evolutionary operators like crossover or selection while PSO does not have it (Eberhart and Shi, 1998). Recently, PSO has been implemented in a wide range of applications, e.g. in the water resources (e.g., Bisselink et al. 2016, Zambrano-Bigiarini and Rojas, 2013; Abdelaziz and Zambrano-Bigiarini, 2014), geothermal resources (Ma et al., 2013; Beck et al., 2010), finance and economics (Das, 2012), in structural design (Kaveh and Talatahari, 2009; Schutte and Groenwold, 2003), economics and finance (Huang et al., 2006; Das, 2012), and applications of video and image analysis (Donelli and Massa, 2005; Huang and Mohan, 2007). For example, the groundwater model MODFLOW2000/2005 was linked with PSO to estimate permeability coefficients (Sedki and Ouazar, 2010) and a multi-objective PSO code was used to derive rainfall runoff model parameters by introducing the Pareto rank concept (Gill et al., 2006). Notwithstanding recent popularity, PSO has never been used to calculate the parameters of a surface complexation model (SCMs) simulating sorption behavior of metal and metalloids on mineral surfaces. Hence, this paper attempts to examine the efficiency and effectiveness of PSO for parameter estimation of a surface complexation model as is PHREEQC (Parkhurst and Appelo, 1999).

Nowadays, a number of PSO software codes exist such as MADS (Harp and Vesselinov, 2011; Vesselinov and Harp, 2012) and OSTRICH (Matott, 2005), with most of the codes using the basic PSO formulation developed in 1995. However, in this paper we use the latest Standard Particle Swarm Optimization proposed in literature (Clerk, 2012; Zambrano-Bigiarini et al., 2013), named SPSO2011, as implemented in the hydroPSO R package (R Core team, 2016) version 0.3-3 (Zambrano-Bigiarini and Rojas, 2013; 2014). hydroPSO is an independent R package that includes the newest Standard PSO (SPSO-2011) , which was specifically developed to calibrate a wide range of environmental models. In addition, the plotting functions in hydroPSO are user-friendly and aid the numeric and visual interpretation of the optimization results. The source code, installation files, tutorial (vignette), and manual are available on http://cran.r-project.org/web/packages/hydroPSO.

hydroPSO is used in this paper, for the first time, to estimate the parameters of a surface complexation for U(VI)-Quartz system, to properly capture the non-linear interactions between the model parameters. The aim of this article is to examine the suitability of hydroPSO as a global optimisation tool for parameter estimation of geochemical models, in particular PHREEQC -3.1.2. To this end, surface/sorption reaction constants (log K) of the surface complexation model (SCM) obtained with hydroPSO will be compared to those previously obtained with PEST (Doherty, 2010) by Nair et al. (2014).

PEST and PSO are both model-independent parameter optimizers, i.e., they do not require making any change to the model. PEST is uses the Gauss- Marquardt-Levenberg method to minimize, in the weighted least squares sense, the differences between observations and the corresponding model simulated values(Abdelaziz and Bakr, 2012; Edet et al., 2014). PEST is gradient based algorithm initially calculates the Jacobian Matrix. Then the Jacobian matrix is used to build and upgrade parameter vector to enhance the searching and the ability to acquire a smaller objective function value (Doherty, 2005). The model then iterates towards some parameters adjusting the model parameters on the basis of a new Marquardt lambdas value (Doherty, 2005). Hence, Lambda drives the objective function for faster converging. As a local optimizer, PEST is sensitive to the initial condition (see a complete description in Doherty, 2005, 2010). In contrast, PSO is global optimizers which randomly initialize a population of particles within the D-dimensional parameter space. PSO allows to initializing the position of each particle using a random uniform distribution or Latin Hypercube Sampling (LHS), while velocities can be initialized in zero, with two different random distributions, or with two different LHS strategies (see Zambrano-Bigiarini and Rojas, 2013). Velocity and position of each particle in the parameter space are updated in successive iterations following equations specific to the selected PSO version (see a complete description in Zambrano-Bigiarini and Rojas, 2013 and Abdelaziz and Zambrano-Bigiarini, 2014). As a state-of-the-art global optimizer, PSO is less subject get trapped into local minima compared to PEST.

Model description

PHREEQC version 3.1.2 (Parkhurst and Appelo, 1999), the database of Nuclear Energy Agency thermodynamic NEA_2007 (Grenthe et al., 2007), as well as the LLNL database (Lawrence Livermore National Laboratory) are used to model sorption. Both databases were modified by set constant values for $MUO_2(CO_3)_3^{2-}$ and $M_2UO_2(CO_3)_3^{0}$ species (M equals Ca, Mg, Sr) taken from Geipel et al.(2008) and Dong and Brooks (2006, 2008). PHREEQC is a geochemical code which is capable to simulate sorption, surface complexation, and other types of reactions. SCMs are considered to be suitable tools to describe the processes at liquid-solid interfaces (Huber and Lützenkirchen, 2009). Surface Complexation Modelling (SCM) has been widely employed to simulate the sorption of metal species from aqueous solution depending on solution concentration and pH value as well as ionic strength and redox conditions (Davis et al., 2004; Štamberg et al., 2003; Zheng et al., 2003). A group of reactions of aqueous species from the bulk solution with the surface of the sorbent leads to the formation of surface complexes. The constants for these reactions (surface complexation constants, log K) are indispensable for SCM.

There are different SCMs like generalized two layer model (GTLM), nonelectrostatic model (NEM), constant capacitance model (CCM), diffuse-layer model (DLM), modified triple-layer model (modified TLM). Here, a generalized two layer model (GTLM) (Dzombak and Morel, 1990) was used to simulate the sorption behavior of U(VI) on quartz. The GTLM was used instead of other models because it is relatively simple and can be used in a wide range of chemical conditions. A comprehensive review of GTLM is presented in Dzombak and Morel (1990). Quartz is a nonporous mineral and non-layered, and therefore, the actual area of surface is supposed to be equal to the specific surface area. In this study, the surface of quartz is considered as a single binding site which takes the charge for every surface reaction. The sorption reactions and log K values are related to the aqueous species and thus depend on the thermodynamic database used. Uranyl carbonate complexes—$(UO_2)_2CO_3(OH)_3^-$, $UO_2(CO_3)_2^{2-}$ and $UO_2(CO_3)_3^{4-}$—are the dominant species under our experimental conditions. Therefor, the surface-complexation reactions for quartz were calculated with respect to these species.

The sorption of U(VI) on quartz was investigated and discussed by (Huber and Lützenkirchen, 2009). However, formation of Mg-, Ca-, and Sr–Uranyl-Carbonato complexes shows a significant impact on the

sorption of uranium on quartz. This was studied by Nair and Merkel (2011) in batch experiments adding 10 g of powdered quartz to 0.1 liter of water containing rather low U(VI) concentrations ($0.126X10^{-6}$ M) in the absence and presence of Mg, Sr, and Ca (1 mM) at a pH value between 9 and 6.5 in steps of 0.5. $NaHCO_3$ (1×10^{-3} M) and NaCl (1.5×10^{-3} M) were used as ionic-strength buffers. The low U-concentrations were used to avoid precipitation of Ca-U-carbonates. In the absence of alkaline earth elements, the percentage of uranium was sorbed on quartz ca. 90% independent from pH. In the existence of Mg, Sr, and Ca, the percentage of sorption of uranium on quartz declined to 50, 30, and 10%, correspondingly (Nair and Merkel, 2011).

Table 1 displays the parameter ranges used to optimize the 6 parameters selected to calibrate PHREEQC, based on Nair et al., 2014.

Table 1: Complexation reactions with their respective log K range values.

Corresponding Reaction	ID	Parameter Range values		Calibrated Parameter log K
		Min	Max	
$Q\_xOH + UO_2(CO_3)_3^{4-} + OH^- \rightleftharpoons Q\_xOUO_2(CO_3)_3^{5-} + H_2O$	K1	24	26	25.156
$Q\_xOH + UO_2(CO_3)_2^{2-} + OH^- \rightleftharpoons Q\_xOUO_2(CO_3)_2^{3-} + H_2O$	K2	20	23	21.18
$Q\_xOH + UO_2CO_3 \rightleftharpoons Q\_xOUO_2CO_3^- + H^+$	K3	-8	-5	-5.589
$Q\_xOH + UO_2OH^+ \rightleftharpoons Q\_xOUO_2OH + H^+$	K4	2	4	3.229
$Q\_xOH + (UO_2)_2CO_3(OH)_3^- \rightleftharpoons Q\_xO(UO_2)_2CO_3(OH)_3^{2-} + H^+$	K5	5	8	6.733
$Q\_xOH + Na^+ \rightleftharpoons Q\_xONa + H^+$	K6	-7	-4	-5.842

Q_xOH: Silanol surface site

Computational implementation

Inverse modeling is a complex issue for modelers as a result of the numerous uncertainties in model parameters and observations (e.g., Carrera et al., 2005; Beven, 2006). Particle Swarm Optimisation (PSO) is an evolutionary optimisation algorithm originally developed by Kennedy and Eberhart (1995), which has proven to be highly efficient when solving a large collection of case studies from different disciplines (see, e.g., Poli, 2008). In PSO each individual of the population searches for the global optimum in a multidimensional parameter space, considering the individual and collective past experiences. The canonical PSO algorithm starts with a random initialization of the particles' positions and velocities within the multi-dimensional parameter space. Velocity and position of each particle in the parameter space are updated in successive iterations following equations specific to the selected PSO version, trying to find the minimum (or maximum) of a user-defined objective function (see a complete description in Zambrano-Bigiarini and Rojas, 2013). In the last decades, several improvements have been proposed to the canonical PSO algorithm, and hydroPSO implements several of them in a single piece of software. In particular, hydroPSO implements six PSO variants (equations used to update particles' position and velocities), four different topologies, two different initialization of particles' positions (random uniform distribution or Latin Hypercube Sampling), five different alternatives for initializing particles' velocities, among many other fine-tunning options (see Zambrano-Bigiarini and Rojas, 2013). In the application of hydroPSO to PHREEQC, the following configuration was used: a swarm with 10 particles, 200 iterations, LH initialisation of particle positions and velocities, random topology with 11 informants, acceleration coefficients c_1 and c_2 equal to 2.05, linearly decreasing clamping factor for V_{max} in the range [1.0, 0.5], and use of the Clerc's constriction factor instead of the inertia weight. The hydroPSO R package v0.3-3 (Rojas and Zambrano-Bigiarini, 2012; Zambrano-Bigiarini and Rojas, 2013; 2014) is a model-independent optimization package, which implements a state-of-the-art PSO algorithm to carry out a global parameter optimisation, and it has been successfully applied as calibration tool for both hydrogeological and hydrological models (Zambrano-Bigiarini and Rojas, 2013; Thiemig et al., 2013; Abdelaziz and

Zambrano-Bigiarini, 2014; Bisselink et al., 2016), requiring no instruction or template files as UCODE (Poeter et al., 2005; Abdelaziz and Merkel, 2015) and PEST software (Doherty, 2005; 2013) do. In order to couple hydroPSO with the PHREEQC geochemical model, three text files have to be prepared by the user to handle data transfer between the model code and the optimization engine: (i) '*ParamFiles.txt*', which describes the names of a set of parameters to be estimated and locations in the model input files to be utilized in the inverse procedure, (ii) '*ParamRanges.txt*', which defines the minimum and maximum values that each selected parameter might have during the optimization, and (iii) '*PSO_OBS.txt*', which contains the observations that will be compared against its simulated counterparts. In addition, a user-defined R script file ('*Read_output.R*') have to be prepared, containing the instructions to read model outputs, while an R script template provided by hydroPSO (Rojas and Zambrano-Bigiarini, 2012) has to be slightly modified by the user in order to carry out the optimization. Figure 1a shows a flow chart that depicts how hydroPSO is coupled with PHREEQC to calibrate its parameters. Run-phreeqc.bat is a batch file to run PHREEQC-3.1.2 in the DOS environment, which reads *.phrq files to produce *.prn files as output (simulated data); *.ins files are instructions to read model outputs, by using the Read-output.R script. At each iteration, hydroPSO modifies model parameter values to minimize the value of the user-defined objective function. Finally, the new parameter values are updated following the locations provided in the 'ParamFiles.txt' file. In contrast, to couple PEST with PHREEQC, four files are required: i) template files (*.tpl), ii) instruction files (*.ins), iii) a main control file (*.pst), and iv) a batch file to execute PHREEQC and PEST(*.bat) . Template files are built to modify the input files for PHREEQC with other values while an instruct ion file is employed to extract the simulated values from the output file for PHREEQC. The main control file includes a model application will be run, the observations, parameters to be estimated, control data keywords, and etc. For further information about PEST read the manual is recommended. However, Figure 1a,b shows the key files used to couple PHREEQC with hydroPSO, and explains the flowchart and files involved in the inverse modelling of the surface complexation constants for the U(VI) sorption model.

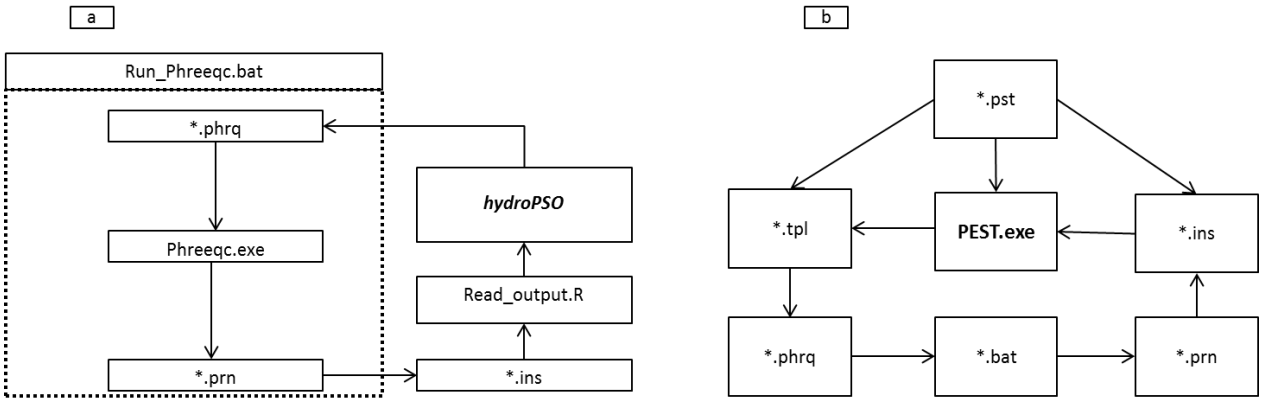


Figure 1: Flow chart used to couple a) PHREEQC with hydroPSO, b) PHREEQC with PEST involved in inverse modeling of surface complexation constants for uranium carbonate (U(VI)) species on quartz with the PHREEQC geochemical model.

For numerical optimization, the residual sum of squares (*RSS* or *SSR*, see Equation (1) was utilized to compute the goodness of fit (*GoF*) between the corresponding model outputs (C^s_j) and observed U-carbonate concentration values (C^o_j) at different pH values for every iteration step i; n is number of observation points (measured sorption U(VI) onto quartz). Minimizing the residual sum of squares was chosen as the method for estimating the surface/sorption reaction constants in calibrations by Nair et al. (2014) when PEST was combined with PHREEQC. It was decided for consistency to select SSR as the criterion for goodness of fit when applying hydroPSO with PHREEQC. After some initial trials, the number of maximum iterations T was set to 200 and the number of particles used to search for the minimum RSS in the parameter space was fixed at 10 (i.e., 2000 runs of the model). The rest of parameters were set to the default values defined in hydroPSO. More information about SPSO 2011 can be found in Clerc (2012) , Zambrano-Bigiarini et. al. (2013), while detailed information about hydroPSO can be found in Zambrano-Bigiarini and Rojas (2013).All the input files required for running PHREEQC and hydroPSO can be found in the supplementary material (https://zenodo.org/record/1044951#.WgVTbVuCzIU), including all the optimization results.

$$SSR = \sum_{j=1}^{n} \left(C_j^s - C_j^o\right)^2 \qquad (1)$$

Results and Discussion

One of the vital and useful approaches to evaluate the efficacy of model performance is through plotting the simulation against observed values (visualizing outcome of model). The observed and simulated sorption ratio and the calculated sorption ratio are compared in Figure 2. The coefficient of determination (R2) for the relation between calculated and observed values is 0.89, indicating a high linear correlation and thus high model quality (Figure 2).

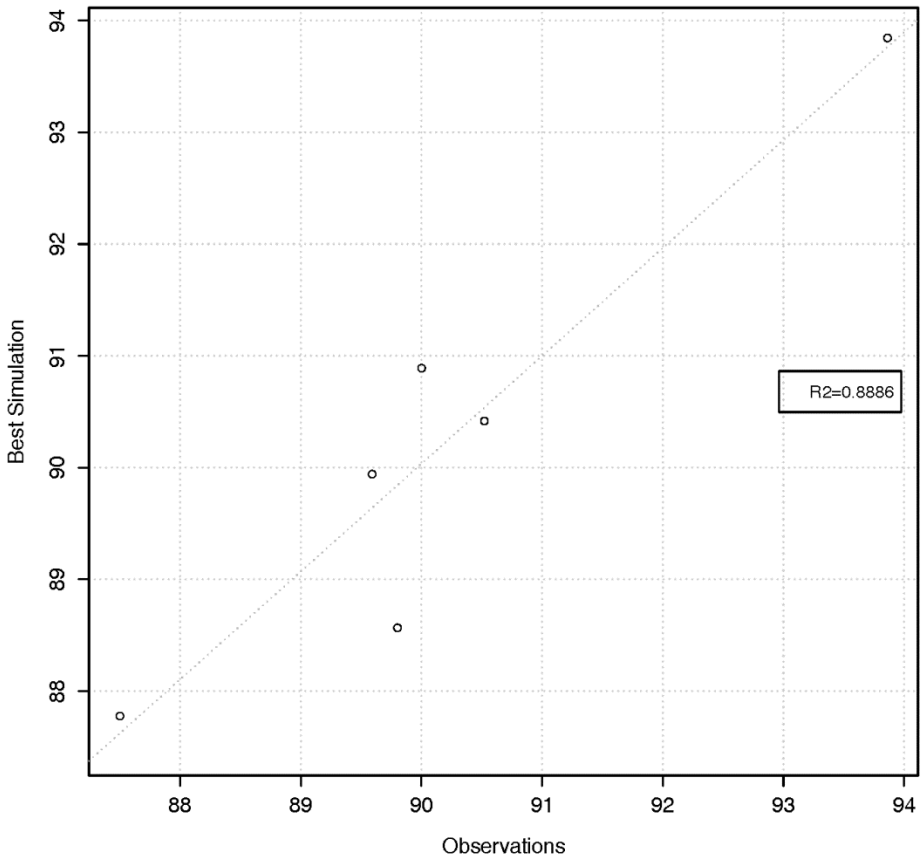
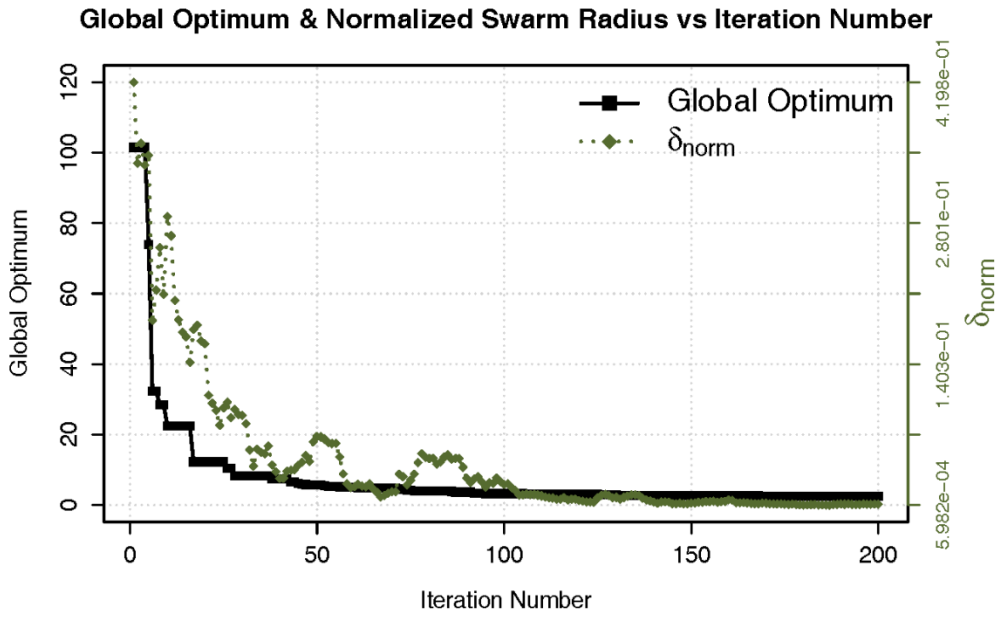


Figure 2: Scatter plot with the experimentally observed and calculated values of uranium carbonate (sorption %).



193

194 **Figure 3: Evolution of the normalized swarm radius (δ norm) and the global optimum (SSR) over**
195 **200 iterations.**

196 In hydroPSO there are two types of criteria for convergence: i) **absolute**, when the global optimum
197 found in a given iteration is below/above than a user-defined threshold (useful for
198 minimization/maximization problems where the true minimum/maximum is known); ii) **relative**, when
199 the absolute difference between the model performance in the current iteration and the model
200 performance in the previous iteration for the best performing particle is less or equal to a user-defined
201 threshold (useful to prevent too many model runs without any improvement in the optimum found by the
202 algorithm). If none of the two previous criteria are met, then the algorithm stops when the user-defined
203 number of iterations is finally achieved. Figure 3 shows the evolution of the best model performance (i.e.,
204 smallest RSS) found by all the particles in a given iteration, and the normalized swarm radius (δ_{norm}, a
205 measure of the spread of the population in the range of search-space) versus the iterations number. One
206 may observe that both δ_{norm} and the best model performance become smaller with an increasing iteration
207 number, which indicates that the main particles are "flying" around a small region in parameter space.
208 Only 100 iterations (i.e., 100x10 =1000 model runs) were enough to reach the region of the global

optimum (i.e., RSS ca. 2.52), and the remaining iterations were just used to refine the search as shown in Figure 3.

The boxplots in Figure 4 are graphical representations of the values sampled during optimization. The bottom and top of the box show the first and third quartiles of the distribution of each one of the surface/sorption reaction constants (log K) sampled during the opimization, respectively. The horizontal line within the box denotes the median of the distribution. Points outside the whiskers are considered to be outliers, where notches are within $\pm 1.58 IQR/\sqrt{n}$, IQR represents the interquartile range and n the total number of parameter sets used in the optimization. The horizontal red lines in Figure 4 point out the optimum value found during optimization for each parameter.

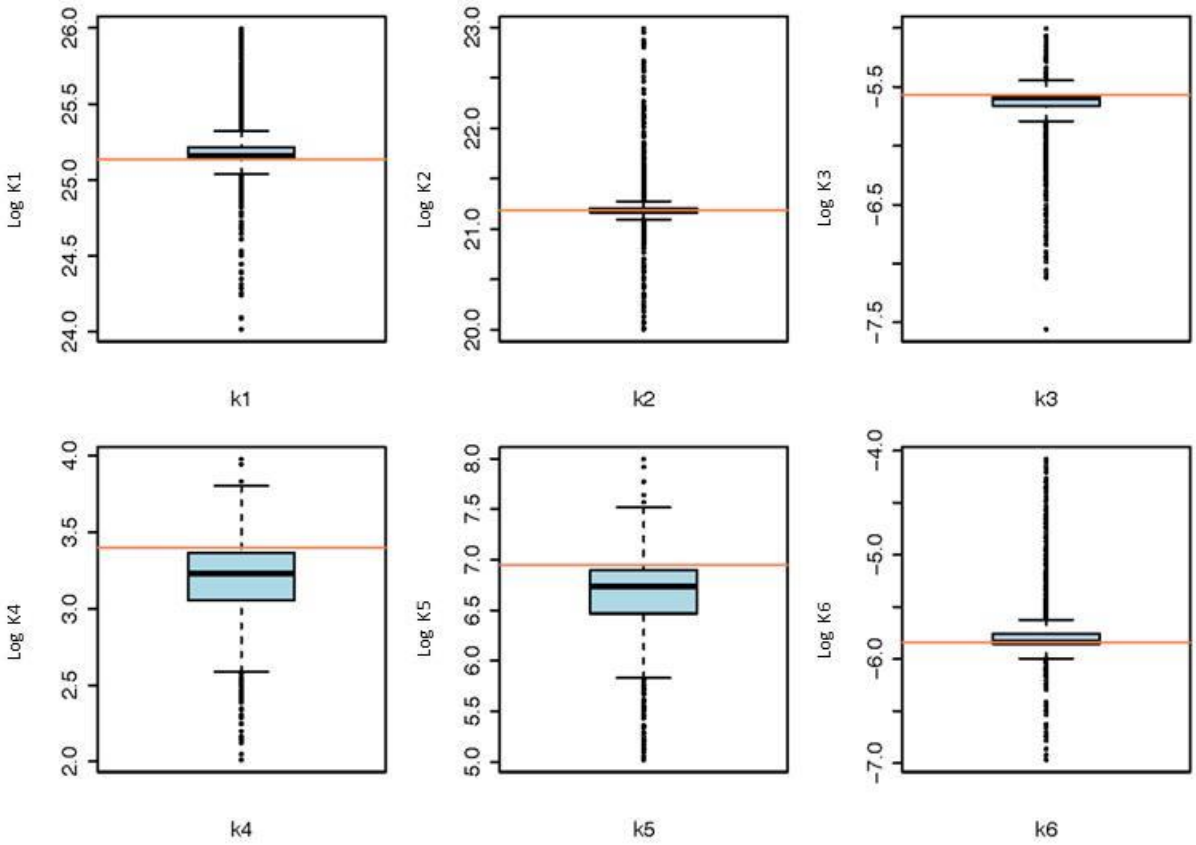


Figure 4: Boxplots for the optimised parameters. The horizontal red lines indicate the optimum value for each parameter. Parameter names are defined in Table 1.

Two dimensional dotty plots in Figure 5 depict the goodness-of-fit values achieved by different parameter sets. They are suitable for identifying ranges where different sets of parameters lead to the same model performance (equifinallity, Beven , 2006).

Figure 5 shows the model performance as function of the interaction of different parameter ranges. The (quasi) three-dimensional dotty plot shown in Figure 5 is a projection of the values of pairs of parameters onto the model response surface "RSS". Parameter values where the model presents high performance are shown in light-blue (points density), whilst the parameter values where the model shows low performance are shown in dark-red (points density). This figure was used to identify regions of the solution space with good and bad model performances (Figure 5).

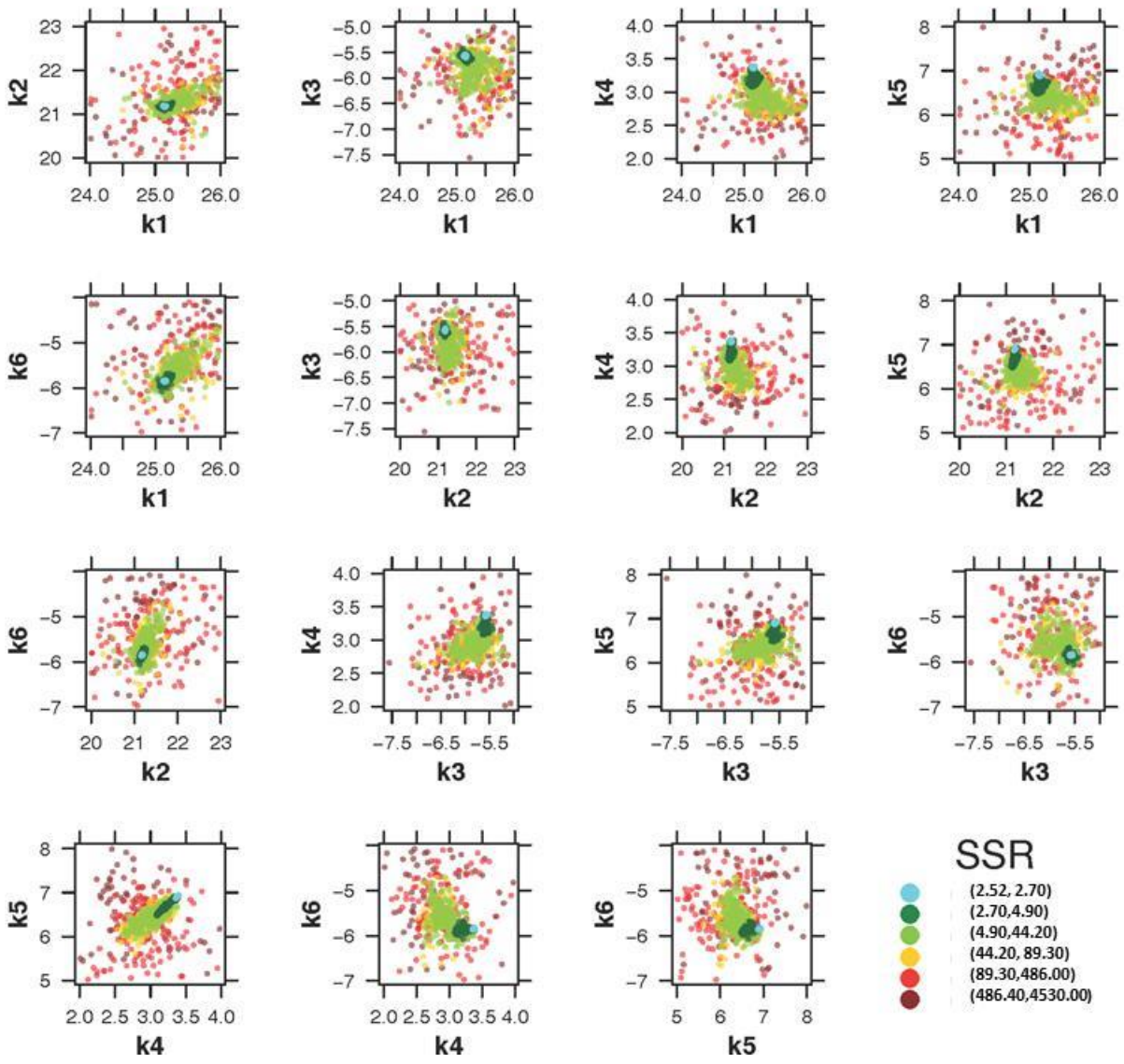


Figure 5: Quasi three-dimensional dotty plots.

Visual inspection of Figure 5 shows a good exploratory capability of PSO because the particles are well spread over the entire range space. It is clearly visible that the parameter samples are denser around the optimum value (lowest SSR), showing a small uncertainty range around the optimum value.

Figure 6 and Figure 7 give a graphical summary for optimised parameters. Empirical Cumulative Density Functions (ECDF) in Figure 6 shows the sampled frequencies for the six calibrated parameters. The horizontal gray dotted lines show the median of the distribution (cumulative probability equal to 0.5) whiles the vertical gray dotted lines depict the corresponding parameter value, display at the top of every

figure (Figure 6). The thin vertical red line in Figure 7 points out the optimum value achieved for each parameter. Histograms in figure 7 show near-normal distributions for K1 and K2, while k4 and k5 follow a skewed distribution with sampled values concentrated near the upper boundary of each parameter.

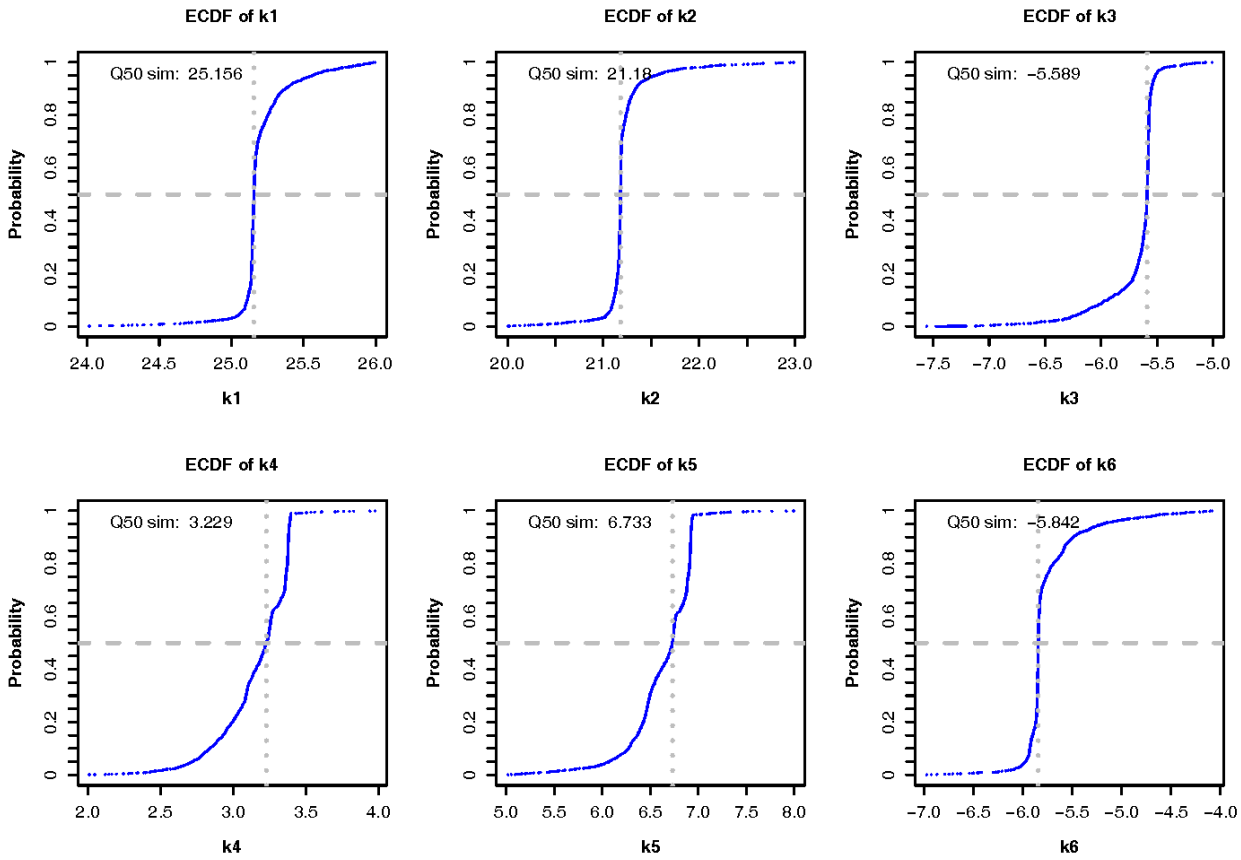
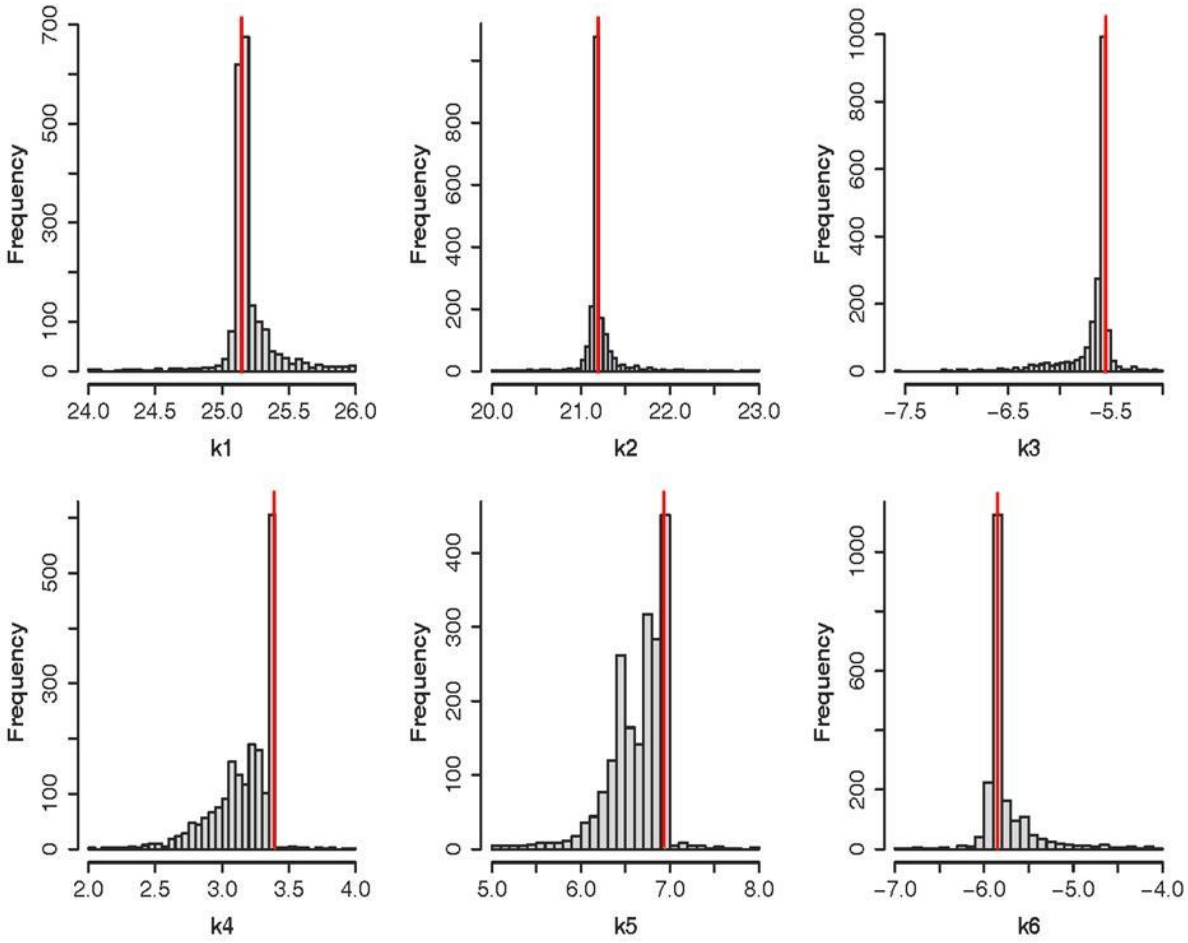


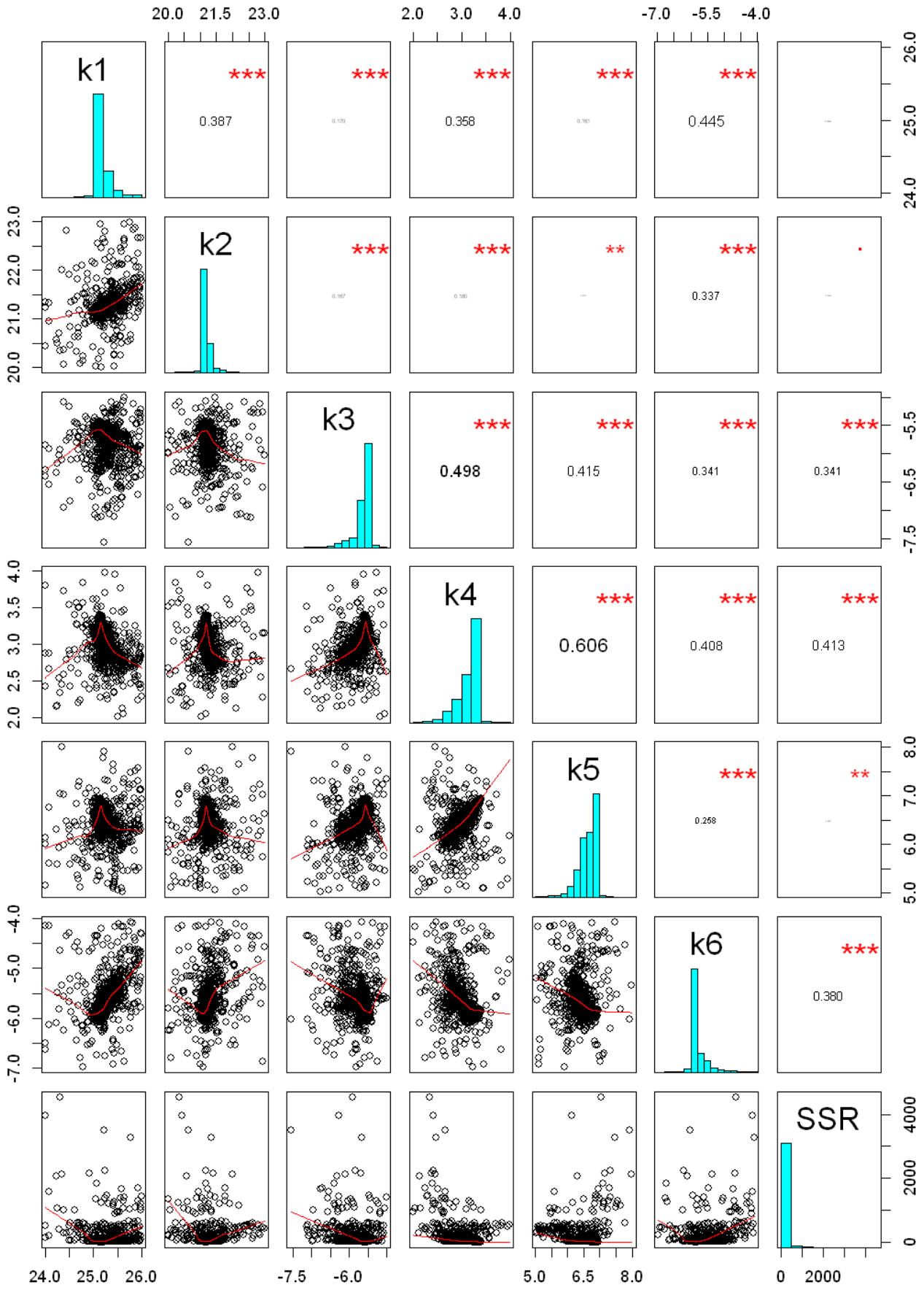
Figure 6: Empirical cumulative density functions for each calibrated parameter.



244

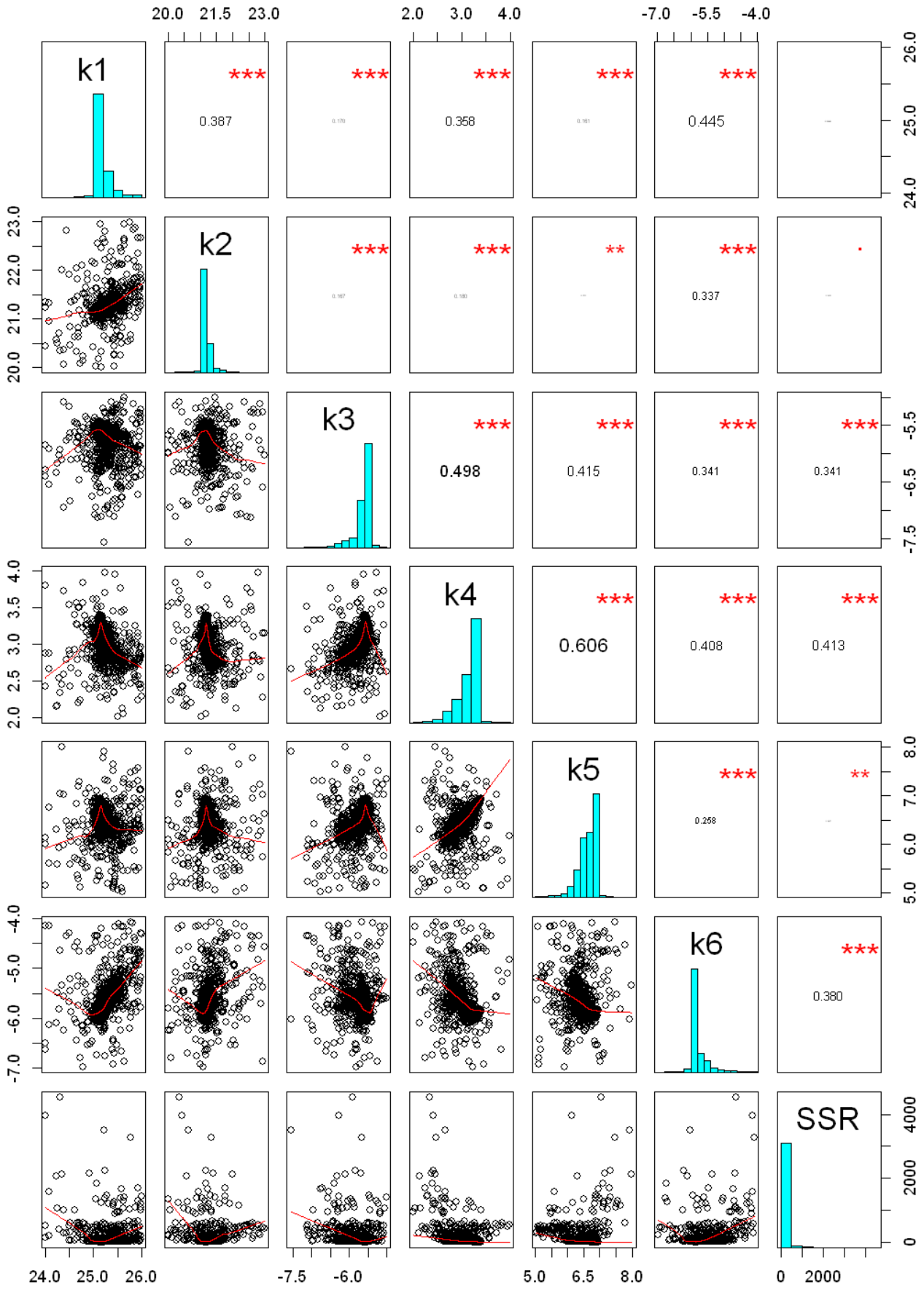
245 **Figure 7: Histograms of calibrated parameter values. Horizontal axis shows the sampled range for**
246 **each parameter and vertical axis represents the amount of parameter sets in each of the classes used**

to divide the horizontal axis.


k1
k2
k3
k4
k5
k6
SSR
0.387
0.358
0.445
0.337
0.498
0.415
0.341
0.341
0.606
0.408
0.413
0.258
0.380

248

Figure 8 illustrates the correlation matrix among K values and model performance (SSR), with horizontal and vertical axes displaying the ranges used for the calibration of each parameter. The figure represents that highest correlation coefficient occurred among the measure of model performance (SSR) and K4, K6, and K3. In addition, a higher correlation coefficient was observed between K4 and K5, K3 and K4, and K1 and K6.


k1
k2
k3
k4
k5
k6
SSR
0.387
0.358
0.445
0.337
0.498
0.415
0.341
0.341
0.606
0.408
0.413
0.258
0.380
20.0 21.5 23.0
2.0 3.0 4.0
-7.0 -5.5 -4.0
24.0 25.0 26.0
20.0 21.0 22.0 23.0
-5.5 -6.5 -7.5
2.0 2.5 3.0 3.5 4.0
5.0 6.0 7.0 8.0
-4.0 -5.0 -6.0 -7.0
0 2000 4000
24.0 25.0 26.0
-7.5 -6.0
5.0 6.5 8.0
0 2000

254

Figure 8: Correlation matrix between model performance (SSR) and calibrated parameters. Red lines represents lowess smoothing, using locally-weighted polynomial regression, and numbers in the upper panel represents the Pearson-moment correlation coefficient between each pair of parameters. Vertical and horizontal axes illustrate the physical range utilized for parameter optimization. * stands for a p < 0.001 ; ** stands for p < 0.01, according to level of statistical significance**

Figure 9 shows the model output using hydroPSO fitted log-K values and the monitored sorption ratio.

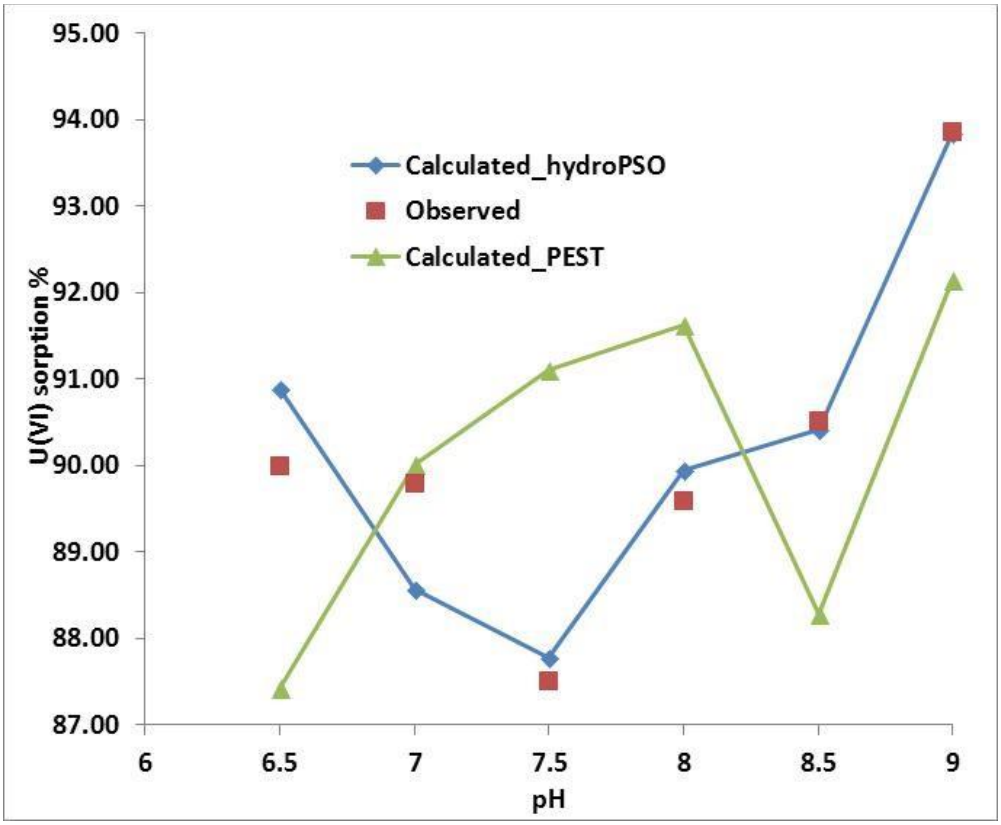


Figure 9: Observed and simulated sorption of uranium in quartz vs pH with both PEST and hydroPSO calibrated log-k values.

It is worthwhile to mention that the surface complexation constants for the equations 1, 2, and 5 are more important and the equations that are less important are 3, 4, and 6 in optimizing the "log K" values. It proves that $UO_2(CO_3)_3^{4-}$, $UO_2(CO_3)_2^{2-}$, and $(UO_2)CO_3(OH)_3^-$ are the most dominate species sorption on quartz. From the optimized model, the surface complexation constants for the equations 2 and 4 was optimized to be 21.18 and 3.229 respectively (Nair et al., 2014), which is higher than the electrostatic (ES)

and nonelectrostatic (NES) models, while the optimized value for equation 1 is 25.156, which is higher than the NES model and almost the same as the ES model (Nair et al., 2014).

Comparing the results of optimized log-K values for GTLM as sorption model which obtained by hydroPSO versus previous work done by Nair et al. (2014). The experimental conditions and the PHREEQC modelling assumptions were the same during the PEST optimisation. In other words, PEST was applied for the similar case and the same data, we can show that the log k values obtained with hydroPSO are better estimations than those obtained by PEST, except for pH=7. The main reason is that PSO is a global optimization technique, which searches for optimum values in the parameter space as defined by the ranges given in Table 1, while PEST searches on a neighborhood of the initial solution. In particular, PEST carries out inverse modelling by computing value of parameters that minimize a weighted least-squares objective function via the Gauss-Marquardt-Levenberg non-linear regression method (Marquardt, 1963). Actually, a major drawback of PEST, as of all gradient-based techniques, is the dependency of the quality of the optimization results upon the initial point used for the optimization, which might lead to a local optimum rather than the global one. Thus, PSO techniques offer promising possibilities for similar surface complexation and reactive transport applications in hydrogeology and hydrochemistry.

Conclusions

The coupling of hydroPSO and PHREEQC was successfully carried to estimate surface complexation constants for uranium (VI) species on quartz, based on a data set published by Nair and Merkel(2011), and Nair et al.(2014). The open-source hydroPSO R package proved to be a useful tool for inverse modeling of surface complexation models with PHREEQC and allowed a prompt evaluation of the calibration results. Furthermore, thermodynamic values obtained with *hydroPSO* provided a better match to observation sorption rates in comparison to those obtained with PEST, using the same input data.

Data availability

PHREEQC is available in the following http://www.hydrochemistry.eu/ph3/index.html. Source code, tutorials, and reference manual of hydroPSO can be obtained from https://CRAN.R-project.org/package=hydroPSO. The PHREEQC model input files along with the R scripts used for coupling it with hydroPSO and the model outputs can be obtained from the Zenodo repository (https://zenodo.org/record/1044951#.WgVTbVuCzIU).

References

Abdelaziz, R.; ZAMBRANO-BIGIARINI, M. Particle Swarm Optimization for inverse modeling of solute transport in fractured gneiss aquifer. Journal of contaminant hydrology, **2014**, 164. Jg., S. 285-298.

Abdelaziz, R.; Bakr. M.I. Inverse modeling of groundwater flow of Delta Wadi El-Arish. J. Water Resour. Protect., **2012**, 4 (07), p. 432.

Abdelaziz, R.; Merkel, B.J., 2015. Sensitivity analysis of transport modeling in a fractured gneiss aquifer. Journal of African Earth Sciences. **2015**, 103, pp.121-127.

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

Beven, K. A manifesto for the equifinality thesis. Journal of hydrology, **2006**, 320. Jg., Nr. 1, S. 18-36.

Bisselink, B., Zambrano-Bigiarini, M., Burek, P., & de Roo, A. Assessing the role of uncertain precipitation estimates on the robustness of hydrological model parameters under highly variable climate conditions. Journal of Hydrology: Regional Studies. **2016**, 8, 112-129.

Clerc, M. Standard Particle Swarm Optimisation. Technical Re-port. Particle Swarm Central. http://clerc.maurice.free.fr/pso/SPSO_descriptions.pdf. [Online. Last accessed 24-Sep-2012], **2012**.

Carrera, J; Alcolea, A.; Medina, A.; Hidalgo, J.; Slooten, L. J.. Inverse problem in hydrogeology. Hydrogeology journal, **2005**, 13. Jg., Nr. 1, S. 206-222.

Das, Parichay. Economics of Distributed Generation Using Particle Swarm Optimization: A Case Study. Economics, **2012**, 1. Jg., Nr. 5.

Davis, J.A.; Meece, DE, Kohler M, Curtis GP. Approaches to surface complexation modeling of uranium(VI) adsorption on aquifer sediments. Geochimica Et Cosmochimica Acta, **2004**,68 (18):3621-364.1.

Doherty, J. PEST: model-independent parameter estimation, user manual. Technical Report (5th ed.)Watermark Numerical Computing, Brisbane, Queensl., Australia, **2005**.

Doherty, J. PEST: Model-independent Parameter Estimation. User Manual, fifth ed. Watermark Numerical Computing, **2010**.

Doherty, J. Addendum to the PEST manual.Technical ReportWatermark Numerical Computing, Brisbane, Queensl., Australia, **2013**.

Dong, W.M.; Brooks, S.C. Determination of the formation constants of ternary complexes of uranyl and carbonate with alkaline earth metals ($Mg^{2+,}$ Ca^{2+}, Sr^{2+}, and Ba^{2+}) using anion exchange method. Environ Sci Technol, **2006**, 40:4689–4695.

Dong, W.M.; Brooks, S.C. Formation of aqueous $MgUO2(CO_3)_3^{2-}$ complex and uranium anion exchange mechanism onto an exchange resin. Environ Sci Technol, **2008**, 42:1979–1983.

Donelli, M., Massa, A. Computational approach based on a particle swarm optimizer for microwave imaging of two-dimensional dielectric scatterers. IEEE Transactions on Microwave Theory and Techniques, **2005**,53(5), 1761-1776.

Dzombak, D.A.; Morel, F.M. Surface complexation modeling: Hydrous ferric oxide. John Wiley & Sons, New York, **1990**.

Eberhart, R.; Kennedy, J. A new optimizer using particle swarm theory. Proceedings of the Sixth International Symposium on Micro Machine and Human Science, 1995. MHS'95, 1995,pp. 39–43.

Eberhart, R.C.; Shi, Y.Comparison between genetic algorithms and particle swarm optimization, in Evolutionary Programming VII, V. Porto, N. Saravanan, D. Waagen, and A. Eiben, Eds. Springer Berlin / Heidelberg, **1998**,vol. 1447, pp. 611–616. doi: 10.1007/BFb0040812.

Edet, A. ; Abdelaziz, R.; Merkel, B.; Okereke, C.; Nganje, T. Numerical groundwater flow modeling of the coastal plain sand aquifer, Akwa Ibom State, SE Nigeria. J. Water Resour. Protect. **2014**, 6 (04)), p. 193

Geipel, G.; Amayri, S.; Bernhard, G. Mixed complexes of alkaline earth uranyl carbonates: a laser-induced time-resolved fluorescence spectroscopic study. Spectrochimica Acta Part A-Mol Biomol Spectrosc, **2008**, 71:53–58.

Gill, M. K.; Kaheil, Y. H.; Khalil, A.; McKee, M.; Bastidas, L. Multiobjective particle swarm optimization for parameter estimation in hydrology. Water Resources Research, **2006**,42(7). doi:10.1029/2005WR004528.

Grenthe, I.; Fuger, J.; Konings R.; Lemire, R.J.; Muller, A.B.; Wanner, J. The Chemical Thermodynamics of Uranium. Elsevier: New York, **2007**.

Harp, D., Vesselinov, V.V., Recent developments in MADS algorithms: ABAGUS and Squads, EES-16 Seminar Series, LA-UR-11-11957, **2011**.

Huang, F. Y.; Li, R. J.; Liu, H. X.; Li, R. A modified particle swarm algorithm combined with fuzzy neural network with application to financial risk early warning. In Services Computing, 2006. APSCC'06. IEEE Asia-Pacific Conference, IEEE, **2006**, 168-173.

Huang, T., & Mohan, A. S. A microparticle swarm optimizer for the reconstruction of microwave images. IEEE Transactions on Antennas and Propagation, **2007**,55(3), 568-576.

Huber, F.; Lutzenkirchen, J., Uranyl Retention on Quartz-New Experimental Data and Blind Prediction Using an Existing Surface Complexation Model. Aquatic Geochemistry, **2009**, 15, (3), 443-456.

Kaveh, A., Talatahari, S. A particle swarm ant colony optimization for truss structures with discrete variables. Journal of Constructional Steel Research, **2009**,65(8), 1558-1568.

Kennedy, J.; Eberhart, R. Particle swarm optimization, in: neural networks, 1995. Proceedings. IEEE International Conference on Neural Networks, **1995**, pp. 1942–1948.

Ma, R.J.; Yu, N.Y.; Hu, J.Y. Application of particle swarm optimization algorithm in the heating system planning problem. Sci. World J. 11., **2013**, http://dx.doi.org/10.1155/2013/718345.

Marquardt, D. An algorithm for least-squares estimation of nonlinear parameters. Journal of the Society for Industrial and Applied Mathematics, **1963**,11. pp. 431–441.

Matott, L. 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, **2005**.

Nair, S.; Merkel B. J. Impact of Alkaline Earth Metals on Aqueous Speciation of Uranium(VI) and Sorption on Quartz. Aquatic Geochemistry, **2011**,1 17 (3):209-219

Nair, Sreejesh; Karimzadeh, Lotfollah; Merkel, Broder J. Surface complexation modeling of Uranium (VI) sorption on quartz in the presence and absence of alkaline earth metals. Environmental Earth Sciences, **2014**, 71. Jg., Nr. 4, S. 1737-1745.

Parkhurst, D.L.; Appelo, C.A. User's Guide to PHREEQC (version 2). A Computer Program for Speciation, Batch-Reaction, One-Dimensional Transport, and Inverse Geochemical Calculation. USGS, Water Resources Investigation Report, **1999**, 99 – 4259.

Poli, R. Analysis of the publications on the applications of particle swarm optimisation. Journal of Artificial Evolution and Applications, **2008**.

Poeter, E.; Hill, M.; Banta, E.; Mehl, S.; Christensen, S. UCODE 2005 and six other computer codes for universal sensitivity analysis, calibration, and uncertainty evaluation. US Geological Survey Techniques and Methods, **2005**, vol. 6-A11.

R Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. Vienna, Austria. URL: http://www.R-project.org/, **2016**.

Rojas, R.; Zambrano-Bigiarini, M. Tutorial for interfacing hydroPSO with SWAT-2005 and MODFLOW-2005. Technical Report. URL: http://www.rforge.net/hydroPSO/files/hydroPSO_vignette.pdf. [Online. Last accessed 03-Feb-2014], **2012**.

Schutte, J. F., Groenwold, A. A. Sizing design of truss structures using particle swarms. Structural and Multidisciplinary Optimization, **2003**, 25(4), 261-269.

Sedki, A.; Ouazar, D. Swarm intelligence for groundwater management optimization. Journal of Hydroinformatics, **2011**,13(3), 520-532.

Štamberg, K.; Venkatesan, K. A.; Vasudeva Rao, P. R. Surface complexation modeling of uranyl ion sorption on mesoporous silica. Colloids and Surfaces A: Physicochemical and Engineering Aspects, **2003**, 221. Jg., Nr. 1, S. 149-162.

Thiemig, V., Rojas, R., Zambrano-Bigiarini, M., De Roo, A. Hydrological evaluation of satellite-based rainfall estimates over the Volta and Baro-Akobo Basin. Journal of Hydrology, **2013**,499, 324-338.

Vesselinov, V.V.; Harp, D.R. Adaptive hybrid optimization strategy for calibration and parameter estimation of physical process models. Computers & Geosciences, **2012**, 49, 10–20. doi:10.1016/j.cageo.2012.05.027.

Zambrano-Bigiarini, M.; Rojas, R. A model-independent Particle Swarm Optimisation software for model calibration. Environmental Modelling & Software, **2013**, 43, 5-25. doi:10.1016/j.envsoft.2013.01.004.

Zambrano-Bigiarini, M.; Rojas, R. hydroPSO: Particle Swarm Optimisation, with focus on Environmental Models. URL: http://www.rforge.net/hydroTSM/, http://cran.r-project.org/web/packages/hydroTSM/ .R package version 0.3-3, **2014**.

Zambrano-Bigiarini, M., Clerc, M., & Rojas, R. (2013, June). Standard particle swarm optimisation 2011 at cec-2013: A baseline for future pso improvements. In *Evolutionary Computation (CEC), 2013 IEEE Congress on* (pp. 2337-2344). IEEE.

Zheng, Z.; Tokunaga, T. K.; Wan, J. Influence of calcium carbonate on U (VI) sorption to soils. Environmental science & technology, **2003**, 37. Jg., Nr. 24, S. 5603-5608.