Particle Swarm Optimization for Surface Complexation with the PHREEQC geochemical model

MS No: grid-2017-38

Abstract

~~Recently, Particle Swarm Optimization (PSO) techniques have attracted many researchers to optimize model parameters in several fields of research.~~ This article ~~paper~~ explains, for the first time, how to interface the hydroPSO R optimization package and PEST with the PHREEQC geochemical model, version 2.3.1. The main theme involved the sorption of low concentration uranium on quartz at different pH. ~~Sorption of metals on minerals is a key process in treatment water, natural aquatic environments, and other water related technologies. Sorption processes can be simulated by means of surface complexation models. However, determining thermodynamic constants for surface species from batch experiments requires robust parameter estimation tool that does not get stuck in local minima. In this work, uranium at low concentrations was sorbed on quartz at different pH.~~ Results show that hydroPSO delivers more reliable thermodynamic parameter values than PEST when both are coupled to PHREEQC using the same thermodynamic input data ~~(Nair et al., 2014)~~. Besides, ~~Post~~post-processing tools ~~included~~ in hydroPSO are helpful for the interpretation of the results. Thus, hydroPSO is ~~a~~ recommended as an optimization tool for PHREEQC with respect to inverse modeling to determine reliable ~~and meaningful~~ thermodynamic parameter values.

Introduction and Scope

Particle Swarm Optimization technique (PSO) is an evolutionary optimization approach proposed by Eberhart and Kennedy (1995) and was influenced by the activities of flocks of birds in search of corn (Kennedy and Eberhart 1995, and Eberhart and Kennedy 1995). PSO and genetic algorithms (GA) shares a few similarities (Eberhart and Shi 1998). GA has evolutionary operators like crossover or selection, while PSO does not have it (Eberhart and Shi 1998). Recently, PSO 28 has a wide range of applications, including water resources (Zambrano-Bigiarini and Rojas, 2013, Abdelaziz and Zambrano-Bigiarini 2014), geothermal resources (Ma et al., 2013; Beck et al., 2010), structural design (Kaveh and Talatahari, 2009; Schutte and Groenwold, 2003), economics and finance (Huang et

Formatted: Numbering: Continuous
Formatted: Left
Formatted: Font: 14 pt
Formatted: Font: 14 pt
Formatted: Font: 14 pt
Formatted: Justified
Formatted: Font: 14 pt
Formatted: Font: 14 pt
Formatted: Font: 14 pt
Formatted: Font: 14 pt
Formatted: Font: (Default) Times New Roman, 14 pt
Formatted: Font: 14 pt
Formatted: Font: 14 pt
Formatted: Font: 14 pt
Formatted: Font: 14 pt
Formatted: Font: (Default) Times New Roman, 14 pt
Formatted: Font: 14 pt

al., 2006; Das 2012), and applications of video and image analysis (Donelli and Massa, 2005; Huang and Mohan, 2007). For example, the groundwater model MODFLOW 2000/2005 was linked with PSO to estimate permeability coefficients (Sedki and Ouazar, 2010) and a multi-objective PSO code was used to derive a rainfall runoff model parameters (Gill et al., 2006). Notwithstanding PSO recent popularity, the PSO has never been used to calculate the parameters in surface complexation model (SCMs) simulating sorption behavior of metal and metalloids on mineral surfaces. Hence, this article attempts to examine the efficiency and effectiveness of PSO for parameter estimation of a surface complexation model as is PHREEQC (Parkhurst and Appelo, 1999).

Today, 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 the latest Standard Particle Swarm Optimization, SPSO2011 (Clerk, 2012; Zambrano-Bigiarini et al., 2013). This is same as 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 latest 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 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 versatility of *hydroPSO* as parameter estimation tool for geochemical modeling with PHREEQC -3.1.2.

Model description

PHREEQC version 2.3 (Parkhurst and Appelo, 1999) is used to model the sorption and the database of Nuclear Energy Agency thermodynamic NEA_2007 (Grenthe et al., 2007), as well as the LLNL database (Lawrence Livermore National Laboratory) is 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) obtained from the work of Geipel et al. (2008) and Dong and Brooks (2006, 2008). PHREEQC is a geochemical model code, capable of simulating sorption, surface complexation, and other types of reactions. Surface Complexation

Formatted ...
Comment [h1]: What parameters?
Formatted ...
Formatted ...
Formatted ...
Field Code Changed
Formatted: Hyperlink, Font: 14 pt, Font color: Auto
Formatted ...
Comment [h2]: See lines 33-35. Harmonize!
Formatted: Font: 14 pt
Formatted: Font: 14 pt
Formatted: Justified
Formatted ...
Comment [h3]: Check?
Formatted ...

Modelling (SCM) is considered a suitable tool to describe the processes at liquid-solid interfaces (Huber and Lützenkirchen, 2009). SCM has been widely employed to simulate the metals sorption from aqueous solution depending on pH and concentration of the solution as well as ionic strength and redox conditions (Davis et al., 2004; Štamberg et al., 2003; Zheng et al., 2003). Thus the result of group reactions within the aqueous species in the surface of the sorbent and the bulk solution, that leads to the surface complexes formation. The constants of surface sorption reaction (log K) values are inevitable for SCM. Such constants are universal constants, not site-specific, and hence transferable. There are different SCMs including two layer model (GTLM), non-electrostatic 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 and 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. Therefore, the surface complexation reactions for quartz were calculated with respect to these species.

The sorption of U (VI) on quartz were investigated and discussed by 80 (Huber and Lützenkirchen 2009). However, formation of Mg-, Ca-, and Sr–Uranyl-Carbonato complexes show a significant impact on sorption of uranium on quartz. This was studied by Nair and Merkel (2011) in batch experiments by adding 10 g of powdered quartz to 0.1 liter of water containing low U concentrations (0.126 9 10-6 M) in the absence and existence of Mg, Sr, and Ca (1 mM) at pH of between 6.5 and 9 in steps of 0.5. $NaHCO_3$ (1 x 10^{-3} M) and NaCl (1.5 x 10^{-3} M) were used as ionic-strength buffers. The low U concentrations were used to avoid precipitation of Ca-U-carbonates. In the non-existence of alkaline earth elements, the percentage of uranium was sorbed on quartz ca. 90% independent from pH. In the presence of Mg, Sr, and Ca, the percentage of sorption of uranium on quartz declined to 50, 30, and 10%, respectively (Nair and Merkel, 2011).

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

Formatted ...

Formatted ...

Formatted ...

Computational implementation

Inverse modeling is a procedure issue for modelers as a result of the numerous uncertainties in model parameters and observations (e.g., Carrera et al., 2005, Beven, 2006). 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 has been successfully applied as calibration tool for both hydrogeological and hydrological models, requiring no instruction or template files as UCODE (Poeter et al., 2005, Abdelaziz and Merkel, 2015) and PEST (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.

In contrast to coupling PEST with PHREEQC, it was required that PEST be run with PHREEQC. PEST needs ASCII output and input files. The four files required were: 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 were built to modify the input files for PHREEQC with other values while an instruction file was employed to extract the simulated values from the output file for PHREEQC. The main control file includes a model application to run, the observations, parameters to be estimated, control data keywords, and etc. Details of PEST are contained in the manual ( ). Figure 1 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.

Formatted: Font: 14 pt
Formatted: Justified
Formatted: Font: 14 pt
Formatted: Highlight
Formatted: Font: 14 pt
Formatted: Font: 14 pt, Highlight
Formatted: Highlight
Formatted: Font: 14 pt, Highlight
Comment [h4]: Harmonize with lines 43-46 of original MS
Formatted: Font: 14 pt
Formatted: Font: 14 pt
Formatted: Font: 14 pt, Highlight
Formatted: Highlight
Formatted: Font: 14 pt, Highlight
Formatted: Highlight
Formatted: Font: 14 pt, Highlight
Formatted: Highlight
Formatted: Font: 14 pt, Highlight
Formatted: Highlight
Formatted: Font: 14 pt
Formatted: Font: 14 pt
Formatted: Font: 14 pt
Formatted: Highlight
Formatted: Font: 14 pt
Formatted: Font: 14 pt, Highlight
Formatted: Font: 14 pt
Formatted: Highlight
Formatted: Font: 14 pt
Formatted: Font: 14 pt

Figure 1: Flow chart with files involved in inverse modeling of surface complexation constants for uranium carbonate (U(VI)) species on quartz with the PHREEQC geochemical model.

For numerical optimization, equation 1 was used to compute the residual sum of the squared (*RSS* or *SSR*):

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

This equation (1) was then utilized to compute the goodness of fit (*GoF*) between the corresponding model outputs (Cs) and observed values (*Co*) for every time step *i*. 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. Detailed and additional information on SPSO 2011 and hydroPSO are contained in Clerc (2012), Zambrano-Bigiarini et. al. (2013), Zambrano-Bigiarini and Rojas (2013). Finally, all the input files required for running PHREEQC and hydroPSO can be found in the supplementary material (https://doi.org/10.5281/zenodo.803874), including all the optimization results.

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 variable observed sorption ratio and the calculated sorption ratio were compared in Figure 2. It is clear that there is a very good fit between the calculated and the experimentally observed values. The coefficient of determination (r^2), in this case almost 0.89 indicates a good match between the observed and calculated values (Figure 2). Only 100 iterations were enough to achieve the region of the global optimum, i.e., 100 x100 =1000 model runs. The rest iterations numbers were placed to refine the search as shown in Figure 3.

Formatted: Font: 14 pt
Formatted ...
Formatted ...
Formatted: Font: 12 pt
Formatted: Font: Not Italic
Formatted ...
Formatted: Font: 14 pt
Formatted: Font: 14 pt
Formatted: Justified
Formatted ...
Formatted: Font: (Default) Times New Roman, 14 pt

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

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

Figure 3 shows the evolution of the global optimum (best model performance for a given iteration, i.e., smallest SSR) and the normalized swarm radius (*δnorm*, a measure of the spread of the population flying over the range of search-space) versus the iterations number. One may observe that both *δnorm* and the global optimum become smaller with an increasing iteration number, which indicates that the main particles are "flying" around a small portion of the solution space. The optimum value was achieved when the SSR was ca. 2.52. Figure 4 presents the values of ……..sampled during optimization.

Figure 4: Boxplots for calibrated parameter. The horizontal red lines indicate the optimum value for each parameter. The bottom and top of the box demonstrate the first and third quartiles, respectively. The horizontal line within the box stands for the median. The points outside the notches are considered to be outliers, where notches are within ±1.58IQR/sqrt(n), while IQR represents the interquartile range and n the number of points. The horizontal red lines represent the optimum value found during optimization for each parameter.

The SSR was chosen as an indicator for goodness of fit (*GoF*). Two dimensional dotty plots in Figure 5 depict the goodness of fit achieved by different parameter sets. They are suitable for identifying parameter ranges, leading to high or roughly the same model performance (Beven and Binley, 1992).

Figure 5 shows the model performance as function of the interaction of different parameter ranges. This figure was used to identify regions of the solution space with good and bad model performances.

Figure 5: Quasi three-172 **dimensional dotty plots.** ***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 (goodness-of-fit value). Parameter values where the model presents high performance are shown in light-blue (high points density), whilst the parameter values where the model shows low performance are shown in dark-red (low points density).***

Formatted ...
Comment [h5]: Include line of fit
Formatted: Font: 14 pt
Formatted ...
Formatted ...
Formatted: Font color: Custom Color(RGB(0,0,10))
Formatted ...
Formatted: Font: 14 pt, Italic
Formatted: Justified
Formatted ...
Formatted ...
Formatted ...

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 shown that the parameter samples are denser around the optimum value (lowest SSR), indicating a low standard deviation around the optimum value. Nevertheless, the optimum value obtained for K3 and K2 indicated the particles were converging into a small region of the solution space.

Figures 6 and 7 give a graphical summary for calibrated parameters.

Empirical Cumulative Density Functions (ECDF) in Figure 6 shows the sampled frequencies for the six parameters.

Figure 6: Empirical cumulative density functions against each parameter of parameter values. ***The horizontal gray dotted lines show a median of the distribution (cumulative probability 180 equal to 0.5), while the vertical gray dotted lines depict a cumulative probability of 0.5, and its value is displayed in the top of every figure.***

Figure 7: Histograms of calibrated 188 parameter values. ***The vertical red line point out the optimum value achieved for each parameter. The histograms and ECDFs show near-normal distributions for K1 and K2, while k4 and k5 follow a skewed distribution with more sampled values near the upper boundary.***

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 show that highest correlation coefficient is among the measure of model performance (SSR) and k4, k6, 193 and k3. In addition, a higher correlation coefficient was observed between k4 and k5, k3 and k4, and k1 and k6.

Figure 8: Correlation matrix among model performance (SSR) and 196 calculated parameters. Vertical and horizontal axes illustrate the physical range utilized for parameter identification. * stands for a p < 0.001 ; ** stands for p < 0.01, according to level of statistical significance**

Formatted: Font: 14 pt
Formatted: Font: 14 pt
Formatted: Highlight
Formatted: Font: 14 pt
Formatted: Font: 14 pt
Formatted: Highlight
Formatted: Font: 14 pt
Formatted: Font: 14 pt
Formatted: Font: 14 pt
Formatted: Font: 14 pt, Highlight
Formatted: Highlight
Formatted: Font: 14 pt, Highlight
Formatted: Font: 14 pt
Formatted: Font: 14 pt
Formatted: Font: 14 pt
Comment [h6]: Which six
Formatted: Highlight
Formatted: Highlight
Formatted: Font: 14 pt
Formatted: Font: Bold, Italic
Formatted: Font: 14 pt
Formatted: Font: 14 pt
Formatted
Formatted: Font: 14 pt
Formatted: Font: Bold, Italic
Formatted: Justified
Formatted: Font: 14 pt
Formatted: Font: 14 pt
Formatted: Font: 14 pt
Formatted: Font: 14 pt
Formatted
Formatted: Font: 14 pt
Formatted: Font: 14 pt
Formatted: Font: 14 pt
Formatted
Formatted: Font: 14 pt
Formatted: Font: 14 pt

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 4 are more important and the equations that are less important are 3, 5, 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_2OH^+ are the dorminant species sorbed 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, 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.

Comparing the results of optimized log-K values with hydroPSO with previous work of Nair et al. (2014). PEST was applied for the similar case and from the same data, it can be shown that the log k values obtained with hydroPSO are better than those obtained with PEST. The main reason is that PSO is a global optimization technique, which searches for optimum values in the whole parameters space, while PEST searches on a neighborhood of the initial solution. In particular, PEST carries out inverse modelling by computing value of parameter that minimizes a weighted least-squares objective function via 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 done to estimate surface complexation constants for uranium (VI) species on quartz. The open-source hydroPSO R package was confirmed to be a robust tool for inverse modeling of surface complexation models with PHREEQC and allowed a prompt

Formatted: Font: 14 pt
Formatted ...
Formatted ...
Comment [h7]: Where are equations 2, 3, 4, 5 and 6?
Formatted: Font: 14 pt
Formatted ...
Formatted: Font: 14 pt
Formatted: Justified
Formatted ...
Formatted: Font: 14 pt
Formatted: Justified
Formatted ...

evaluation of the calibration results. Furthermore, thermodynamic values obtained with *hydroPSO* provided a better match to observed sorption rate in comparison to those obtained with PEST, using the same input data. This is documented by the higher coefficient of determination for the results based on *hydro*PSO.

Finally, the paper basically treats the coupling of a parameter estimation code with PHREEQC. A limited data set was used from the work of Nair and Merkel (2011) and Nair et al. (2014) to demonstrate the ability of PSO as an optimizer for a geochemical model as PHREEQC. The examples were used only for testing the coupled-codes to show the link between PHREEQC and hydroPSO. Indeed, it is obvious that more comprehensive data sets in the future are needed to get a best-fit and smaller degree of uncertainty.

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/803874#.WTigbY26zIV).

Formatted: Font: 14 pt
Formatted: Font: 14 pt
Formatted: Highlight
Formatted: Font: 14 pt
Formatted: Font: 14 pt
Formatted: Font: 14 pt
Formatted: Font: 14 pt
Formatted: Font: 14 pt
Formatted: Font: 14 pt, Highlight
Formatted: Highlight
Formatted: Highlight
Formatted: Font: 14 pt
Formatted: Font: 14 pt
Formatted: Font: 14 pt
Formatted: Font: 14 pt
Formatted: Font: 14 pt
Formatted: Font: 14 pt
Formatted: Justified
Formatted: Font: 14 pt, Font color: Auto
Formatted: Font color: Auto
Formatted: Font: 14 pt, Font color: Auto
Formatted: Font color: Auto
Formatted: Font: 14 pt, Font color: Auto
Formatted: Font color: Auto
Formatted: Font color: Auto
Formatted: Font: 14 pt, Font color: Auto
Formatted: Font color: Auto
Formatted: Font: 14 pt, Font color: Auto
Formatted: Font color: Auto
Formatted: Font: 14 pt, Font color: Auto
Formatted ...
Formatted ...
Formatted ...