

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, Postpost-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: (Default) Times New Roman, 14 pt

Formatted: Font: 14 pt

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

51
52 Today, a number of PSO software codes exist such as MADS (Harp and
53 Vesselinov, 2011; Vesselinov and Harp, 2012) and OSTRICH (Matott, 2005),
54 with most of the codes using the basic PSO formulation
55 developed in 1995. However, in this paper the latest Standard Particle Swarm
56 Optimization, SPSO2011 (Clerk, 2012; Zambrano-Bigiarini et al., 2013). This is
57 same as in the *hydroPSO* R package (R Core team, 2016) version 0.3-3
58 (Zambrano-Bigiarini and Rojas, 2013; 2014). *hydroPSO* is an independent R
59 package that includes the latest Standard PSO (SPSO-2011), which was
60 specifically developed to calibrate a wide range of environmental models. In
61 addition, the plotting functions in *hydroPSO* are user-friendly and aid the numeric
62 and visual interpretation of the optimization results. The source code, installation
63 files, tutorial (vignette), and manual available on [http://cran.r-](http://cran.r-project.org/web/packages/hydroPSO)
64 [project.org/web/packages/hydroPSO](http://cran.r-project.org/web/packages/hydroPSO). *hydroPSO* is used in this paper, for the first
65 time, to estimate the parameters of a surface complexation for U(VI)-Quartz
66 system, to properly capture the non-linear interactions between the model
67 parameters. The aim of this article is to examine the versatility of *hydroPSO* as
68 parameter estimation tool for geochemical modeling with PHREEQC -3.1.2.

71 **Model description**

72
73 PHREEQC version 2.3 (Parkhurst and Appelo, 1999) is used to model the sorption
74 and the database of Nuclear Energy Agency thermodynamic NEA 2007 (Grenthe
75 et al., 2007), as well as the LLNL database (Lawrence Livermore National
76 Laboratory) is used to model sorption. Both databases were modified by set
77 constant values for $\text{MUO}_2(\text{CO}_3)_3^{2-}$ and $\text{M}_2\text{UO}_2(\text{CO}_3)_3^0$ species (M equals Ca, Mg,
78 Sr) obtained from the work of Geipel et al. (2008) and Dong and Brooks (2006,
79 2008). PHREEQC is a geochemical model code, capable of simulating sorption,
80 surface complexation, and other types of reactions. Surface Complexation

Formatted

Commented [h1]: What parameters?

Formatted

Formatted

Formatted

Field Code Changed

Formatted: Hyperlink, Font: 14 pt, Font color: Auto

Commented [h2]: See lines 33-35. Harmonize!

Formatted: Font: 14 pt

Formatted: Font: 14 pt

Formatted: Justified

Formatted

Commented [h3]: Check?

Formatted

81 Modelling (SCM) is considered a suitable tool to describe the processes at liquid-
82 solid interfaces (Huber and Lützenkirchen, 2009). SCM has been widely employed
83 to simulate the metals sorption from aqueous solution depending on pH and
84 concentration of the solution as well as ionic strength and redox conditions (Davis
85 et al., 2004; Stamberg et al., 2003; Zheng et al., 2003). Thus the result of group
86 reactions within the aqueous species in the surface of the sorbent and the bulk
87 solution, that leads to the surface complexes formation. The constants of surface
88 sorption reaction (log K) values are inevitable for SCM. Such constants are
89 universal constants, not site-specific, and hence transferable. There are different
90 SCMs including two layer model (GTLM), non-electrostatic model (NEM),
91 constant capacitance model (CCM), diffuse-layer model (DLM), modified triple-
92 layer model (modified TLM). Here, a generalized two layer model (GTLM)
93 (Dzombak and Morel, 1990) was used to simulate the sorption behavior of U(VI)
94 on quartz. The GTLM was used instead of other models because it is relatively
95 simple and can be used in a wide range of chemical conditions. A comprehensive
96 review of GTLM is presented in Dzombak and Morel (1990). Quartz is a
97 nonporous mineral and non-layered, and therefore, the actual area of surface
98 is supposed to be equal to the specific surface area. In this study, the surface of quartz
99 is considered as a single binding site and takes the charge for every surface
100 reaction. The sorption reactions and log K values are related to the aqueous species
101 and thus depend on the thermodynamic database used. Uranyl carbonate
102 complexes— $(\text{UO}_2)_2\text{CO}_3$, $(\text{OH})_3^-$, $\text{UO}_2(\text{CO}_3)_2^{2-}$ and $\text{UO}_2(\text{CO}_3)_3^{4-}$ are the dominant
103 species under our experimental conditions. Therefore, the surface complexation
104 reactions for quartz were calculated with respect to these species.

106 The sorption of U (VI) on quartz were investigated and discussed by 80 (Huber
107 and Lützenkirchen 2009). However, formation of Mg-, Ca-, and Sr-Uranyl-
108 Carbonato complexes show a significant impact on sorption of uranium on quartz.
109 This was studied by Nair and Merkel (2011) in batch experiments by adding 10 g
110 of powdered quartz to 0.1 liter of water containing low U concentrations (0.126 9
111 10^{-6} M) in the absence and existence of Mg, Sr, and Ca (1 mM) at pH of between
112 6.5 and 9 in steps of 0.5. NaHCO_3 (1×10^{-3} M) and NaCl (1.5×10^{-3} M) were used
113 as ionic-strength buffers. The low U concentrations were used to avoid
114 precipitation of Ca-U-carbonates. In the non-existence of alkaline earth elements,
115 the percentage of uranium was sorbed on quartz ca. 90% independent from pH. In
116 the presence of Mg, Sr, and Ca, the percentage of sorption of uranium on quartz
117 declined to 50, 30, and 10%, respectively (Nair and Merkel, 2011).

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

Formatted

Formatted

Formatted

121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160

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
- Commented [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
- 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

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

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

$$169 \text{SSR} = \sum_{i=1}^n (C_i^s - C_i^o)^2 \quad (1)$$

173 This equation (1) was then utilized to compute the goodness of fit (GoF) between
174 the corresponding model outputs (Cs) and observed values (Co) for every time
175 step i. After some initial trials, the number of maximum iterations T was set to 200
176 and the number of particles used to search for the minimum RSS in the parameter
177 space was fixed at 10 (i.e., 2000 runs of the model). The rest of parameters were
178 set to the default values defined in hydroPSO. Detailed and additional information
179 on SPSO 2011 and hydroPSO are contained in Clerc (2012), Zambrano-Bigiarini
180 et. al. (2013), Zambrano-Bigiarini and Rojas (2013). Finally, all the input files
181 required for running PHREEQC and hydroPSO can be found in the supplementary
182 material (<https://doi.org/10.5281/zenodo.803874>), including all the optimization
183 results.

186 Results and Discussion

188 One of the vital and useful approaches to evaluate the efficacy of model
189 performance is through plotting the simulation against observed values (visualizing
190 outcome of model). The variable observed sorption ratio and the calculated
191 sorption ratio were compared in Figure 2. It is clear that there is a very good fit
192 between the calculated and the experimentally observed values. The coefficient of
193 determination (r^2), in this case almost 0.89 indicates a good match between the
194 observed and calculated values (Figure 2). Only 100 iterations were enough to
195 achieve the region of the global optimum, i.e., $100 \times 100 = 10000$ model runs. The
196 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

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

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

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

215 **Figure 4: Boxplots for calibrated parameter. The horizontal red lines indicate**
216 **the optimum value for each parameter. The bottom and top of the box**
217 **demonstrate the first and third quartiles, respectively. The horizontal line within**
218 **the box stands for the median. The points outside the notches are considered to**
219 **be outliers, where notches are within $\pm 1.58 IQR / \sqrt{n}$, while IQR represents the**
220 **interquartile range and n the number of points. The horizontal red lines**
221 **represent the optimum value found during optimization for each parameter.**

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

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

233 **Figure 5: Quasi three-172 dimensional dotty plots. The (quasi) three-**
234 **dimensional dotty plot shown in Figure 5 is a projection of the values of pairs of**
235 **parameters onto the model response surface (goodness-of-fit value). Parameter**
236 **values where the model presents high performance are shown in light-blue (high**
237 **points density), whilst the parameter values where the model shows low**
238 **performance are shown in dark-red (low points density).**

Formatted

Commented [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

239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277

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
- Commented [h6]: Which six parameters?
- 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: (Default) Times New Roman, 14 pt
- Formatted: Font: (Default) Times New Roman, 14 pt
- Formatted: Font: 14 pt
- Formatted: Font: 14 pt

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

Formatted: Font: 14 pt

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

Formatted

283 It is worthwhile to mention that the surface complexation constants for the
284 equations 1, 2, and 4 are more important and the equations that are less important
285 are 3, 5, and 6 in optimizing the “log K” values. It proves that $UO_2(CO_3)_3^{4-}$, UO_2
286 $(CO_3)_2^{2-}$, and UO_2OH^+ are the dominant species sorbed on quartz.

Formatted

Commented [h7]: Where are equations 2, 3, 4, 5 and 6?

289 From the optimized model, the surface complexation constants for the equations 2
290 and 4 was optimized to be 21.18 and 3.229 respectively, which is higher than the
291 electrostatic (ES) and nonelectrostatic (NES) models, while the optimized value for
292 equation 1 is 25.156, which is higher than the NES model and almost the same as
293 the ES model.

Formatted: Font: 14 pt

296 Comparing the results of optimized log-K values with hydroPSO with previous
297 work of Nair et al. (2014). PEST was applied for the similar case and from the
298 same data, it can be shown that the log k values obtained with hydroPSO are better
299 than those obtained with PEST. The main reason is that PSO is a global
300 optimization technique, which searches for optimum values in the whole
301 parameters space, while PEST searches on a neighborhood of the initial solution.
302 In particular, PEST carries out inverse modelling by computing value of parameter
303 that minimizes a weighted least-squares objective function via Gauss-Marquardt-
304 Levenberg non-linear regression method (Marquardt, 1963). Actually, a major
305 drawback of PEST, as of all gradient-based techniques, is the dependency of the
306 quality of the optimization results upon the initial point used for the optimization,
307 which might lead to a local optimum rather than the global one. Thus, PSO
308 techniques offer promising possibilities for similar surface complexation and
309 reactive transport applications in hydrogeology and hydrochemistry.

Formatted: Font: 14 pt

Formatted: Justified

Formatted

312 **Conclusions**

Formatted: Font: 14 pt

313 The coupling of hydroPSO and PHREEQC was successfully done to estimate
314 surface complexation constants for uranium (VI) species on quartz. The open-
315 source hydroPSO R package was confirmed to be a robust tool for inverse
316 modeling of surface complexation models with PHREEQC and allowed a prompt
317

Formatted: Justified

Formatted

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

322
323 Finally, the paper basically treats the coupling of a parameter estimation code with
324 PHREEQC. A limited data set was used from the work of Nair and Merkel (2011)
325 and Nair et al. (2014) to demonstrate the ability of PSO as an optimizer for a
326 geochemical model as PHREEQC. The examples were used only for testing the
327 coupled-codes to show the link between PHREEQC and *hydroPSO*. Indeed, it is
328 obvious that more comprehensive data sets in the future are needed to get a best-fit
329 and smaller degree of uncertainty.

330 Data availability

331
332
333 PHREEQC is available in the following
334 <http://www.hydrochemistry.eu/ph3/index.html>. Source code, tutorials, and
335 reference manual of *hydroPSO* can be obtained from [https://CRAN.R-](https://CRAN.R-project.org/package=hydroPSO)
336 [project.org/package=hydroPSO](https://CRAN.R-project.org/package=hydroPSO). The PHREEQC model input files along with the R
337 scripts used for coupling it with *hydroPSO* and the model outputs can be obtained
338 from the Zenodo repository (<https://zenodo.org/record/803874#.WTigbY26zIV>).
339

- Formatted: Font: 14 pt
- Formatted: Font: 14 pt
- Formatted: Highlight
- Formatted: Font: 14 pt
- Formatted: Font: 14 pt, Highlight
- Formatted: Highlight
- Formatted: Highlight
- 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: (Default) Times New Roman, 14 pt, Font color: Auto
- Formatted: Font: 14 pt, Font color: Auto
- Formatted: Font: Times New Roman