Reviewer, #1: This paper studies the applicability of coupling particle swarm optimization (PSO) to the geochemical PHREEQC code for estimating complexation constants, in this case the reaction of uranium compounds with quartz surfaces. The results are compared to previous estimations where the PEST algorithm was used. The manuscript is within the scope of this journal as this seems to be a new approach to the optimisation of surface complexation constants derived from experimental data. However, as various passages are not clearly written, it is only recommended for publication after major revision.

Response: We appreciate the comment made by the reviewer.

General comments include:

The manuscript would greatly benefit from a short introduction to the principles of PSO in contrast to PEST, as one can assume that the average reader may not be familiar with these codes.

Much effort is put into the technical description of coupling hydroPSO to PHREEQC, e.g. what files are used and how they are formatted (if necessary at all, this could be described in an appendix) , but how the actual coupling (e.g. the fitting process itself) is done remains vague.

It is unclear what the fitting data are and under which assumptions (e.g. ES or NES model of complexation, presence or absence of alkaline earth elements in solution) the fits are done.

Be consistent with citing:

Text: Author (year)

In brackets: (Author 1, year 1; Author 2, year 2;: : :)

Response: The citations were modified. Please check the new version.

Specific comments:

l. 1: The title is misleading. Surface complexation itself cannot be optimised (only the model). Suggest something like “Particle swarm optimization for the estimation of surface complexation constants with the geochemical model PHREEQC<version>”

Response: The title was adapted.

l. 18: Specify what the acronym PEST means.

Response: The acronym for PEST was added.

l. 35: what does “multi-objective” mean in this context?

In line 36, “multi-objective PSO” concept refers to the use of PSO to optimize two objective functions (root-mean-square error and bias) of the Sacramento soil moisture accounting (SACSMA) model, using the “Pareto front” concept (Gill et al., 2006).

l. 53: Need to include that parameter estimations will be compared to PEST. This may also be the right place to give a short introduction to the principles of PEST vs. PSO.

> PEST (Doherty, 2010) is a widely used model-independent local optimisation tool, which requires (input/output) template files to carry out the optimisation. Nair et al. (2014) used PEST in combination with PHREEQC to estimate the surface/sorption reaction constants (log K) for the acid–base reactions of the same surface complexation model (SCM). In this work, surface/sorption reaction constants obtained with hydroPSO are compared to those previously obtained with PEST by Nair et al. (2014), in order to check if the global optimization technique provided by hydroPSO has some advantages for practitioners. Please see the final paragraph of the "Introduction and Scope" section
>
> Doherty, J., 2010. PEST: Model-independent Parameter Estimation. User Manual, fifth ed. Watermark Numerical Computing.

l. 58: This is unclear. Do you mean "databases were modified by adding complexation constants for (: : :) species"? Are these species (and the corresponding alkaline earth ions) constituents of your aqueous model solution?

> Response: yes, PHREEQC version 3.1.2 (Parkhurst and Appelo 1999) and the Nuclear Energy Agency thermodynamic database NEA_2007 (Grenthe et al. 2007) as well as the LLNL database (Lawrence Livermore National Laboratory). Both databases were modified by adding constants for $MUO_2(CO_3)_3^{2-}$ and $M2UO_2(CO_3)_3^{0}$ species (M equal Ca, Mg, Sr) taken from Geipel et al.(2008) and Dong and Brooks (2006, 2008).

l. 60: Define "SCM" here instead of line 62.

> Response: SCM was defined in this line as requested.

l. 65/66: Rephrase sentence with something like "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".

> Response: the sentence modified as requested.

l. 67: Unclear what "site-specific" and "transferable" means.

> Response: Thanks for the comment. That sentence was completely removed, because it does not add any information.

l. 75-77: This passage is not clearly written. Do you mean that the quartz surface is considered to have only one type of binding site? And certainly the database does not determine the reactions that take place (only the reactions and species you may consider in your model).

Response: Quartz is a non-layered and nonporous mineral, and hence the effective surface area is assumed to be equal to the specific surface area. Here, the quartz surface is depicted as single binding site (Q_xOH) and is responsible for all the surface reactions.

l. 78ff.: Please rewrite passage. It is not clearly described what the experimental conditions are (e.g. are alkaline metal ions present or not ?) and which assumptions (e.g. sorption model and PHREEQC parameters used for complexation modelling) are made for parameter estimation. Please state clearly whether the experimental conditions and the PHREEQC modelling assumptions were the same during the PEST optimisation. For instance, in Nair et al. (2014), ES and NES models are used; here you refer to the GTLM model generally, so what is compared to what?

Response: please see the answer for comment l. 206.

l. 91: From the given reference it cannot be seen what the parameter ranges mean and how they come about. Do the minima and maxima constitute the parameter space for the particle swarm? Have they been chosen arbitrarily?

Response: Parameter ranges were chosen based on prior expert knowledge from Nair et al. 2014, and they were chosen as wide as possible to allow the optimization engine to choose the final value for each parameter.

l. 92: The caption is a bit confusing. Suggest something like “Complexation reactions with their respective log K range values”. Change the table headings accordingly (the first column is not the name of the parameter but the corresponding reaction). The “calibrated parameters” are results and should be omitted here. They should be presented in the results section alongside the corresponding values from the PEST study. As to the ID, please use capital K throughout the text!

Response: the caption has been rewritten and the table was improved.

l. 98: Do you have references for the successful application of hydroPSO to hydrogeological and hydrological models?

Response: Please check the new version the document, in particular page 5, lines 121 to 123.

l. 100 – 118: For my taste, this is too technical to be included in this place. Suggest to move this and Figure 1 to an appendix, if necessary at all. Instead, a more detailed description of the interaction of hydroPSO and PHREEQC would be appropriate here, especially how the individual “particles” are initially located in the parameter space and what the individual iterations are. A flow chart for this may be helpful. What is the difference in applying PEST, especially the required number of iterations?

The canonical PSO algorithm starts with a random initialization of the particles' positions and velocities within the D-dimensional parameter space. hydroPSO allows to initialize the position of each particle using a random uniform distribution or Latin Hypercube Sampling (LHS), while velocities can be initialized in zero or with two different random distributions or 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). 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 c1 and c2 equal to 2.05, linearly decreasing clamping factor for Vmax in the range [1.0, 0.5], and use of the Clerc's constriction factor instead of the inertia weight. While in Nair et al. 2014 maximum numbers of iterations are set to 30. Only 9 iterations were sufficient to meet stopping criteria (Levenberg-Marquart method as its core).

l. 123: Here and in eq. (1) it is not defined what the model output and the observed values are. Presumably the C_iˆO are the observed U-carbonate concentrations at 6 different pH-values and the C_iˆS are the respective simulated concentrations. This should be stated somewhere.

Response: Please check the new version the line from 132 to line 133.

l. 124: "iteration step", not "time step". If the presumption in comment l. 123 is correct, the index i in eq. (1) cannot be equal to the iteration step index, so please use a different index, e.g. j.

Response: The index is modified.

l. 129 f.: I downloaded the supplementary material and tried to run the batch file but obviously the input file is missing.

Response: The input files were added as requested. Please check the website again.

l. 134 f.: Remove sentence "It is clear: : :". Change next sentence into something like

"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".

Response: The sentence was adapted.

l. 138: It is unclear how the remaining iterations were performed. What does "placed" mean in this context?

Response: A brief description about the PSO iterations was added in the text before the 'Results and Discussion' section. See the new text in lines 168 to 177 in the 'Results and Discussion' section.

l. 146f.: I do not agree that the figure shows the global optimum, because this is the place in parameter space the algorithm wants to approximate. What you probably mean is the RSS.

Response: We appreciate the comment. We meant the global optimum found by all the particles in a given iteration, which does not necessarily coincide with the real and unknown global optimum we are looking for. However, the fact that PSO is a global optimization technique and that all

the particles are “flying” in a small portion of the parameter space with a model performance better than the one obtained with PEST give us confidence in stating that after 100 iterations we are in the neighborhood of the true global optimum. The text was adapted in lines 168 to 182 in the 'Results and Discussion' section.

l. 148f.: Please explain the criteria for convergence. What is the “optimum value” and at which iteration was it reached? Is there any stop criterion for _norm (or rather_norm ?) and/or RSS (RSS)?

Response: In hydroPSO there are two types of criteria for convergence: i) **absolute**, when the global optimum found in a given iteration is below/above than a user-defined threshold (useful for minimization/maximization problems where the true minimum/maximum is known); ii) **relative**, when the absolute difference between the model performance in the current iteration and the model performance in the previous iteration for the best performing particle is less or equal to a user-defined threshold (useful to prevent too many model runs without any improvement in the optimum found by the algorithm). If none of the two previous criteria are met, then the algorithm stops when the user-defined number of iterations is finally achieved. See lines 168 to 182 in the 'Results and Discussion' section.

l. 153: Quartiles of the distribution densities of log Kj?

Response: Thanks for the comment. In line 153 we meant quartiles of the distribution of each one of the surface/sorption reaction constants (log K) sampled during the optimization. See lines 183 to 185 in the 'Results and Discussion' section.

l. 156: Do you mean “total number of iterations”?

Response: no, it represents the total number of values sampled for each one of the calibrated parameters, i.e, the total number of parameter sets. See lines 187 to 188 in the 'Results and Discussion' section.

l. 158: The vertical axes of the box plots should be denoted by log K1 : : : log K6 and not by RSS.

Response: The vertical axis title was modified.

l. 158: What do these plots show in addition to Figure 7 and how relevant is this?

Response: the box plot and histograms shown in figures provide some insight about which parameters could be better identified (the “peakest” the histogram and the narrowest the boxplot the better identified the parameter.

l. 163: Not sure whether equifinality is the suitable concept here. As far as I know, the term is used if fundamentally different sets of parameters lead to comparable results. In this case, the identification of parameter ranges is a kind of uncertainty estimation. By the way, a correct citation would be Beven (2006), as the term “equifinality” does not appear in the cited paper.

Response: The corrected citation is adapted. See lines 194 to 195 in the 'Results and Discussion' section.

l. 171 (Legend of Figure 5): SSR ranges. Check brackets (round or angular). Suggest using scientific notation and two significant digits throughout .

Response: The legend in figure 5 is modified.

l. 175: Do you mean “showing a small uncertainty range”? The standard deviation can only be defined if the underlying distribution is known (which is obviously not the case).

Response: Thanks for the comment. Actually we meant “showing a small uncertainty range”. See line 206 in the 'Results and Discussion' section.

l. 176 f.: This sentence is unclear. The small uncertainty range has been mentioned before.

Response: Thanks for the comment. The unclear sentence was removed. . See line 206 in the 'Results and Discussion' section.

l. 183: What do you mean by “near-normal”? Did you perform a normality (e.g. KS or Lilliefors) test?

Response: Thanks for the comment. Text was modified. See lines 212 to 213 in the 'Results and Discussion' section.

l. 187: The red line is not discernible in my printout. Please check. Also, the axes should be denoted by log Kj and not Kj .

Response: The red line in figure 1 was made thicker.

l. 188: What are the widths of the bars? It is not quite clear what “frequency” means. Is the integral the total number of iterations? If so, normalizing the frequency to this number may be more appropriate.

Response: The widths of each vertical bar are computed -by default- using the Scott's formula (Scott, 1979), but other alternatives can also be used. Frequency stands for the number of parameter values in each class. See lines 217 to 219 in the 'Results and Discussion' section.

Scott, D. W. (1979) On optimal and data-based histograms. Biometrika 66, 605-610

l. 190 ff.: Not sure whether this kind of representation is necessary. The plots in the lower left half are basically the same as in Figure 5. The diagonal elements do not provide new information, either. What are the red lines? Why do the numbers have different font sizes? - Suggest simply showing the formal correlation (or covariance) matrix (“formal” because the underlying distributions are in most cases non-normal, see e.g. Press et al. “Numerical Recipes in C”, 1992, Sect. 15. 6, p. 695). Also change the caption of Figure 8 as well as the corresponding text accordingly and discuss the correlations between parameters qualitatively.

Response: Thanks for the comment. The reviewer is right in pointing that the plots in the lower left half are basically the same as in Figure 5 (but with less information). However, the red lines and the numbers with different sizes add new information. Red lines represents lowess smoothing, which uses locally-weighted polynomial regression, while the bigger the numbers the bigger the Pearson product-moment correlation coefficient between each pair of parameters.

l. 198: Not sure whether p values are meaningful here, remembering this is only a formal correlation matrix.

Response: Thanks for the comment. The p-values make sense here because they compute a test of the value being zero (i.e., no association). The test statistic is based on Pearson's product moment correlation coefficient and follows a t distribution with 'length(x)-2' degrees of freedom if the samples follow independent normal distributions. If there are at least 4 complete pairs of observation, an asymptotic confidence interval is given based on Fisher's Z transform.

l. 201: (Figure 9) Please give uncertainties for the experimental values. The connecting lines between the simulation results do not add any information and should be omitted.

Response: Thanks for the comment, the lines were replaced by bars. However, uncertainty bands for the simulated outputs were not drawn because they were not available for the PEST results, and therefore it would be impossible to make a comparison.

l. 203: Do you mean the reactions in Tab. 1? There is only one equation in this paper.

Response: lines 203 and 204 of the original manuscript were removed.

l. 204: I am not sure what the meaning of "important in optimizing" is. Do you mean some weighting function should be applied in the optimisation process?

Response: The sentence was removed.

l. 205: I do not quite agree what the most dominant sorbed species are. As this is clear for parameters K1 and K2 as their log K is very high, one can see from Figure 7, that log K5 is substantially higher than log K4 , thus sorption of (UO2)2CO3(OH)32- is preferred to sorption of UO2OH+.

Response: Thanks for the comment the reviewer is right the sentence was modified.

l. 206 ff.: Again this is confusing (see comment l. 78 ff.). Do you refer to Nair et al. (2014)? If so, did you run the PHREEQC model with and without the –noedl option in the SURFACE definition? How was SURFACE defined in the present model (the input file is missing in the supplementary material)? So this section is not informative if it is not specified what is compared.

Response: The reviewer is right. The surface complexation constants are compared to previous values obtained from Nair et al. (2014) which used electrostatic (ES) and nonelectrostatic (NES) model. However, PHREEQC contains three surface-complexation models: (1) the GTLM is used with no explicit calculation of the diffuse-layer composition, (2) the electrostatic double layer model, (3) the NES model (

-no_edl). The electrostatic model is the generalized two-layer model described in Dzombak and Morel (1990). The non-electrostatic model does not consider the effects of the development of surface charge on the formation of surface complexes, with the result that surface complexes are treated mathematically very much like aqueous complexes without activity coefficient terms. In this paper, the SURFACE was generalized two-layer model and was taken from Dzombak and Morel (1990) with no explicit calculation of the diffuse-layer composition. Please see the supplementary material especially PHREEQC file and following keyword data block used in this study!

SURFACE_MASTER_SPECIES

Q_x Q_xOH

Please check the "U.phrq" file in the following website:
https://zenodo.org/record/1044951#.WgVTbVuCzIU

l. 211 f. See last comment.

Response: Response: Thanks for the comment. Text was modified. See lines 243 to 245 in the 'Results and Discussion' section. Also, please see the answer for last comment!

l. 212: change to: ": : :are better estimations than those obtained by PEST, except for pH=7." Can you give a reason why this is different for pH=7 ? Is it possible that PSO did not find the global optimum but the Levenberg-Marquardt algorithm did?

Response: The sentence was adapted. Figure 9 shows that the values found found by PEST provides a better match than hydroPSO to observed values only for PH=7 (i.e., 1 out of 6 cases). The reason is that the two algorithms do their best in providing an optimum fit between all simulations and observations, and probably hydroPSO had to sacrifice one point in order to provide a best fit with the remaining 5 points.

l. 213: change to: "the parameter space as defined by the ranges given in Table 1".

Response: The sentence was adapted.

l. 223: One should be more cautious here. As is stated in l.36, PSO has never been used for estimating surface complexation constants before. So this is a first indication (not a confirmation) that this method could be a "robust tool" for this purpose. The discrepancy at pH=7 adds some doubts concerning robustness.

Response: The sentence was adapted.

l. 227f.: If you compare the coefficients of determination, you also should also show the value for the PEST optimization and include the PEST results in Figure 2. But one should be cautious here, because a higher R2 does not necessarily show the model's superiority because it does not tell you anything about statistical significance! For the comparison of models, tests (e.g. F-Test) are appropriate.

Response: Thanks for the comment. The sentence was removed, because it does not add any information in comparison to the previous sentence.

l. 233: what do you mean by “best fit”?

Response: Thanks for the comment. The last paragraph was completely re-written, and the previous paragraph was slightly modified.

Technical comments:

l. 4: Country is missing.

Response: The country is added.

l.19: Remove citation.

Response: The citation is removed.

l. 24: The particle swarm: : :.

Response: The sentence was adapted.

l. 27: : : :shares a few: : :. GA have: : :

Response: The sentence was adapted.

l. 35: : : :derive rainfall runoff: : :

Response: The sentence was adapted.

l. 36: Notwithstanding recent popularity, PSO has never been: : :

Response: The sentence was adapted.

l. 48: : : :are available: : :

Response: The sentence was adapted.

l. 52: Suggest “suitability” instead of “versatility”.

Response: The sentence was adapted.

l. 55: In l. 53, the version number is 3.1.2 !

Response: The corrected vesion was adapted.

l. 55: Delete “is used the sorption by and”. Insert comma.

Response: The sentence was adapted.

l. 57: : : :are used for sorption modelling.

Response: The sentence was adapted.

l. 59: Delete “model”.

Response: The word “model” was removed.

l. 62: : : :simulate the sorption of metal species: : :.

Response: The sentence was adapted.

l. 75: : : :which take the charge: : :

Response: The sentence was adapted.

l. 79: : : :with respect to these species.

Response: The sentence was adapted.

l. 80: The sorption (: : :) was investigated: : :. See also general comment 4.

Response: The sentence was adapted.

l. 81: : : :complexes shows a significant impact on the sorption: : :.

Response: The sentence was adapted.

l. 82: Use “absence” and “presence” instead of “existence” and “non-existence”, respectively.

Response: The sentence was adapted.

l. 83: Check number of U concentration.

Response: The number of U was added.

l. 90: Delete “however”.

Response: The word “however” is removed.

l. 109-114: Be consistent in using past or present tense.

Response: The tense of sentences was adapted.

l. 122: : : :residual sum of squares (RSS): : : Please be consistent using RSS throughout the text

Response: The sentence was adapted.

l. 128: Remove “finally”.

Response: The word “finally” removed.

l. 134: : : :are compared in Figure 2.

Response: The sentence was adapted.

l. 137: : : :reach the region of the global optimum.

Response: The sentence was adapted.

l. 148: : : :population in the range: : :

Response: The sentence was adapted.

l. 150: : : :a small region in parameter space.

Response: The sentence was adapted.

l. 152: The boxplots in Figure 4 are graphical representations: : :

Response: The sentence was adapted.

l. 154: Change the end of the sentence to ": : :within the box denotes the median of the distribution."

Response: The sentence was adapted.

l. 155: Remove "finally"; replace "notches" by "whiskers".

Response: The sentence was adapted.

l. 159: Caption: "Boxplots for the optimised parameters. The: : :"

Response: The caption was adapted.

l. 161: Suggest removing the first sentence as this has already been explained in l. 122f.

Response: The first sentence was moved.

l. 162: Suggest something like: "They are suitable for identifying ranges where different sets of parameters lead to the same goodness of fit."

Response: The sentence was adapted.

l. 167: Replace the contents of the brackets by "RSS".

Response: The content of the brackets was modified.

l. 168/169: point density

Response: The sentence was adapted.

l. 178: Suggest replacing "calibrated" by "optimised" throughout text.

Response: The word was replaced.

l. 180: : : :the median: : :

Response: The sentence was adapted.

l. 181: : : :depict the corresponding parameter value, displayed at the top: : :

Response: The sentence was adapted.

l. 182: The vertical red line in Figure 7 points out: : : (delete “Figure 7” at end)

Response: The sentence was adapted.

l. 215: : : :parameters that minimize : : : and : : :via the Gauss-: : :.

Response: The sentence was adapted.

l. 226: sorption rates

Response: The sentence was adapted.

l. 232 f.: remove hyphens

Response: The hyphens were removed.

l. 249: Reference does not appear in text.

Response: The reference was removed.

l. 251: Check title of article.

Response: The article title was adapted.

l.298: Should be listed before Huber et al. (2009).

Response: The reference for Huang and Mohan (2007) was moved up.

l. 323: Reference does not appear in text.

Response: The reference was removed.

l. 330: Reference does not appear in text.

Response: The reference was removed.