Available online at www.sciencedirect.com

ScienceDirect

Energy Procedia 76 (2015) 616 – 622

Energy Procedia

European Geosciences Union General Assembly 2015, EGU

Division Energy, Resources & the Environment, ERE

GCex: a Visual Analytics approach for interactive exploration of geochemical models

Marco De Lucia[a,*], Janis Jatnieks[b], Mike Sips[b]

[a] *GFZ German Research Centre for Geosciences - Section 5.3 Hydrogeology, Telegrafenberg, 14473 Potsdam*
[b] *GFZ German Research Centre for Geosciences - Section 1.5 Geoinformatics, Telegrafenberg, 14473 Potsdam*

Abstract

GCex is a graphical tool implementing a Visual Analytics approach to support interactive exploration of geochemical models. It encodes many-to-many input/output relationships by the simple yet effective approach called Stacked Parameter Relation (SPR). GCex assists in the setup of simulations, model runs, data collection and results exploration, greatly enhancing the user experience in tasks such uncertainty and sensitivity analysis, inverse modeling and risk assessment. While in principle model-agnostic, the toolchain currently supports the geochemical code `PHREEQC`. For further information and download refer to the project site: http://www.gfz-potsdam.de/en/section/geoinformatics/projects/gcex/.

Peer-review under responsibility of the GFZ German Research Centre for Geosciences

Keywords: Geochemical modeling, `PHREEQC`, Uncertainty analysis, Visual Analytics

1. Introduction

Geochemical models are vastly employed for evaluation and predictions of fluid-fluid and fluid-rock interactions, finding wide application in subsurface settings: groundwater modeling and remediation, evaluation of geothermal resource extraction, long term CO_2 storage, hydrothermal systems among others.

Given the large uncertainties which affect the thermodynamic and chemical databases and the typical scarcity of data regarding the investigated formations, sensitivity and uncertainty analysis are routinely required in order to assess the validity of the simulated outcomes [1,2]. Moreover, it is often useful to investigate the thermodynamics of chemical reactions in complex systems across a broad range of conditions, such as temperature, redox potential or background salinity: pe-pH and activity diagrams are typical bivariate representations of such calculations.

From a practical point of view, all these tasks share a single common root: the variation of some model's input parameters followed by the analysis and representation of the corresponding changes in the model's outputs. Through

[*] Corresponding author. Tel.: +49-331-288-2829 ; fax: +49-331-288-1529.
E-mail address: delucia@gfz-potsdam.de

Peer-review under responsibility of the GFZ German Research Centre for Geosciences
doi:10.1016/j.egypro.2015.07.881

the explicitation of the relationships between the outputs and the varying inputs, often in form of diagram, the modelers gain insights about the investigated systems.

Geochemical models, however, are highly non-linear and *massively multivariate*, as sketched in figure 1. Total element concentrations in aqueous solution, the mineral phases in the rock, redox potential, pH and possibly kinetic parameters as well as state variables such as pressure and temperature constitute the inputs. Even larger is the number of distinct values calculated by one model run: the speciation of the solution (the concentration of the compounds actually present in the solution), the mass changes in the minerals, their saturation indices and the overall element concentrations after reactions are the outputs; they can easily be few dozens overall.

In such a multivariate context, a bivariate representation such as the activity diagrams is in general not powerful enough to explore the many responses across ranges of many varying inputs. In presence of so many output values, it may be challenging to screen the ensemble of calculated simulations for finding out which ones are significantly sensitive to the variation of particular inputs. A synthetic representation of such complex, non linear many-to-many relationships - between many inputs and many outputs - is therefore extremely beneficial for deep understanding of model's behavior and for a great number of practical tasks related to geochemical modeling.

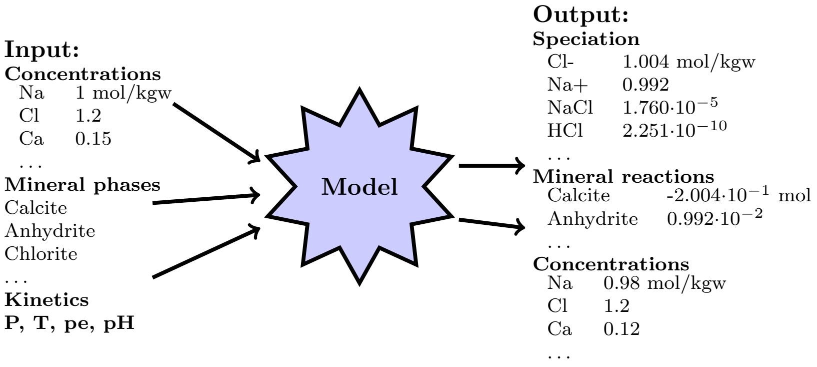


Fig. 1: Illustration of the multivariate nature of input/output data for geochemical models. Total aqueous concentrations, phases in mineral assemblages, pressure, temperature, redox state counts amoung the inputs; speciation, mass changes in mineral phases and total element concentrations constitute the outputs. A dozen of inputs and few dozens of outputs are typical for geochemical models.

To tackle this challenging issue, we introduce a newly developed software tool, GCex, which implements a conceptually simple yet powerful Visual Analytics approach for the interactive exploration of large ensembles of geochemical simulations: the Stacked Parameter Relation (SPR). Furthermore, the software package offers the complete toolchain required for the whole analysis: simulations set up, running the calculations and collecting the results is done through convenient graphical user interfaces, thus greatily easing the workflow of the modelers. GCex is based on open source third party components combining python and R languages, and, while in principle completely model-agnostic, it currently implements `PHREEQC` [3] as chemical speciation software through the `Rphree` interface [1].

In the following section, the core Visual Analytics method SPR (section 2.1) is introduced, followed by details about GCex design principles (section 2.2) and some technical implementation details (section 2.3). The tool itself is finally showcased on a real-life application in section 3.

2. GCex: a Visual Analytics approach

2.1. The Stacked Parameter Relation encoding

The main component of GCex is the hyerarchical, nested axes visualization called Stacked Parameter Relation (SPR) [4], of which an illustrative example with fictitious values is given in figure (2). Its function is to visually relate multiple input parameters to multiple output variables at once, thus offering an overview of model responses over whole ranges of the selected inputs. It consists of three essential parts:

- visualization of input parameter combinations along a single, hierarchical horizontal axis;

- visualization of multiple output variables on a line graph, showing their response to the input combinations;
- interactivity: the user selects the ordering of parameter stacking on the hierarchical x axis and can filter input and output values arbitrarily.

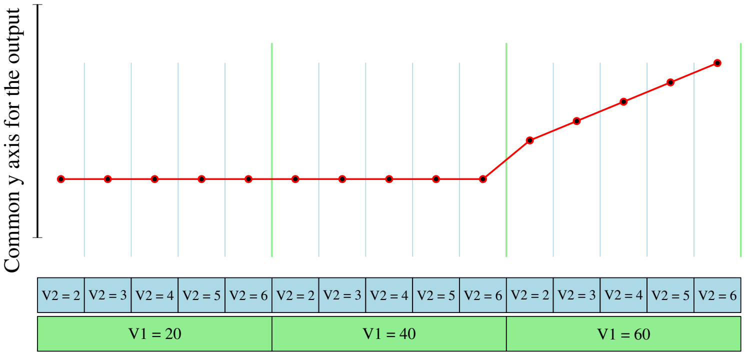


Fig. 2: Sketch of SPR encoding with fictitious parameters. The diagram highlights the repetition of the second input range V2 for as many values the first (V1) takes in the analysis. A common ordinate axis allows the user to read from the SPR the exact value taken by the output at each parameters combination. In this very simple example, it is apparent how the output variable is changing only for large values of V1, and when it happens it is in linear relationship with the parameter V2.

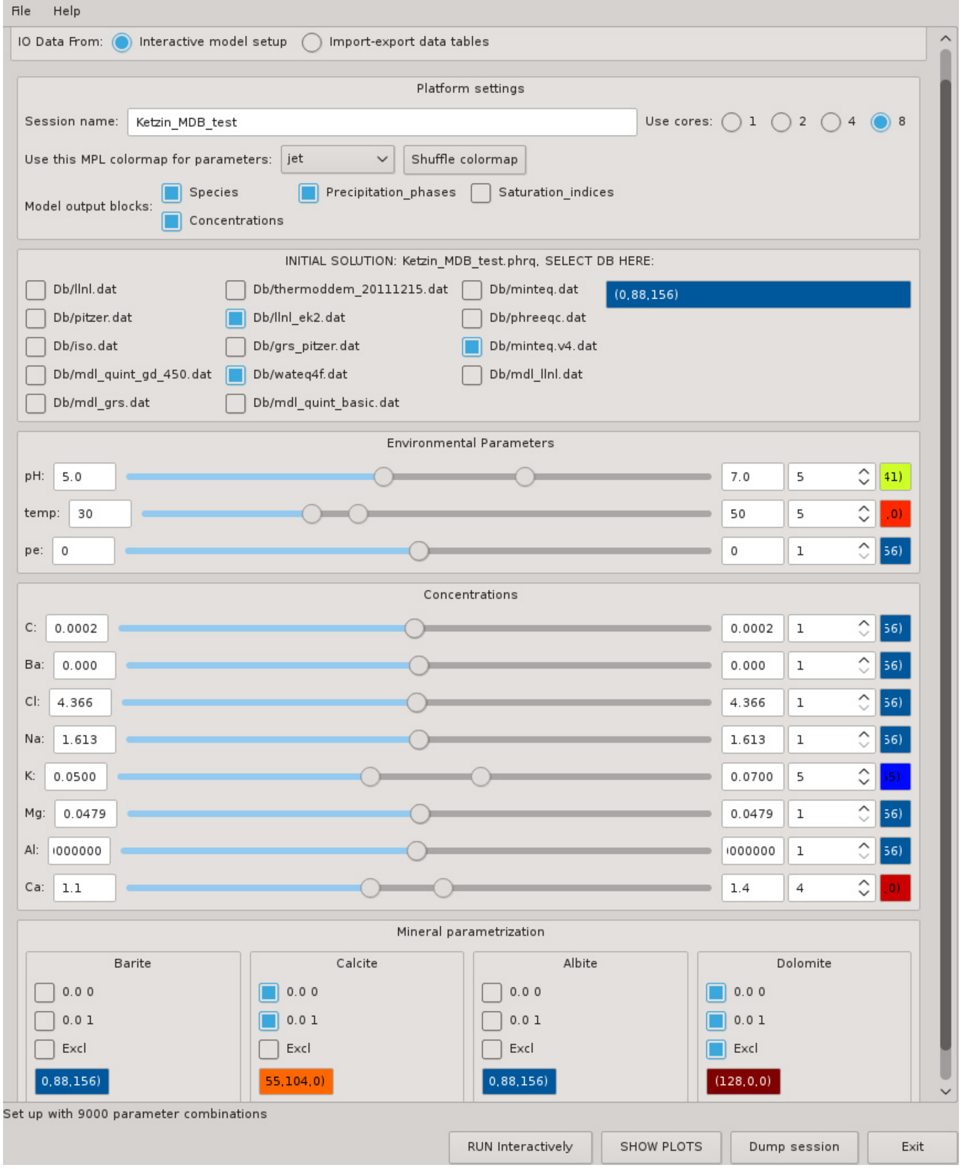
SPR can be thought of as a multivariate generalization of scatter plots, with inputs on the abscissa and responses (outputs) on the ordinate axis. In SPR, the visualization of a further input parameter after the first (V2 in figure 2) is obtained “embedding” a corresponding abscissa at each value point of the hierarchically preceding input (V1). The second input range is therefore repeated along the first, bottom axis. It is at this point clear that adding a further input only requires nesting another hierarchical level at each value of the previous one, recursively. The axes hierarchy is displayed at the bottom of the SPR visualization component. Finally, multiple output variables corresponding to each input combination are plotted using lines of different colors. The design of the SPR visual encoding enables the user to read out the exact input and output values from the diagram itself. This approach enables users to immediately relate changes in output values to combinations of input parameters, effectively highlighting trends, non-linearities and quantitative differences in the outcome of the model. In our experience, up to 5 input parameters can be successfully visualized in SPR at once, depending on the number of value points for each parameter.

The ordering of the axes hierarchy is decisive for the effective interpretation and understanding of the results using the SPR encoding. While it is possible to offer some automatic guidance for the exploration, as will be showed in the next section, the user must be able to adjust the axis hierarchy order to achieve the most meaningful overview of the model. This interactivity is crucial to bring the user’s own expert knowledge into the exploration process and is a fundamental inspiring principle for the implementation of GCex.

2.2. *Design principles*

GCex has been developed to support the analysis of large ensembles of massively multivariate geochemical models resulting from large numbers of input parameters combinations. It has been designed with interactivity, modularity and flexibility in mind, providing both an efficient, powerful and extensible backend for calculations and a graphical user interface to assist users in otherwise tedious and repetitive tasks.

GCex offers a graphical tool for interactive model set-up, allowing the user to interactively specify the parametrization of the models (parameters ranges) and executing the simulation runs (figure 3) starting from a user-provided `PHREEQC` input script. However, the user is free to generate the parameter combinations and the resulting simulations ensembles externally and import it into GCex afterwards for visualization; this possibility makes GCex truly model-agnostic. Some detail about the internal conventions and data formats of the input/output tables expected by GCex are given in subsection (2.3).



IO Data From: ● Interactive model setup ○ Import-export data tables

Platform settings

Session name: Ketzin_MDB_test Use cores: ○ 1 ○ 2 ○ 4 ● 8

Use this MPL colormap for parameters: jet

Model output blocks: ☑ Species ☑ Precipitation_phases □ Saturation_indices ☑ Concentrations

INITIAL SOLUTION: Ketzin_MDB_test.phrq, SELECT DB HERE:

□ Db/llnl.dat □ Db/thermoddem_20111215.dat □ Db/minteq.dat (0,88,156)
□ Db/pitzer.dat ☑ Db/llnl_ek2.dat □ Db/phreeqc.dat
□ Db/iso.dat □ Db/grs_pitzer.dat ☑ Db/minteq.v4.dat
□ Db/mdl_quint_gd_450.dat ☑ Db/wateq4f.dat □ Db/mdl_llnl.dat
□ Db/mdl_grs.dat □ Db/mdl_quint_basic.dat

Environmental Parameters

pH: 5.0 7.0 5 41)
temp: 30 50 5 ,0)
pe: 0 0 1 56)

Concentrations

C: 0.0002 0.0002 1 56)
Ba: 0.000 0.000 1 56)
Cl: 4.366 4.366 1 56)
Na: 1.613 1.613 1 56)
K: 0.0500 0.0700 5 55)
Mg: 0.0479 0.0479 1 56)
Al: ı000000 ı000000 1 56)
Ca: 1.1 1.4 4 ,0)

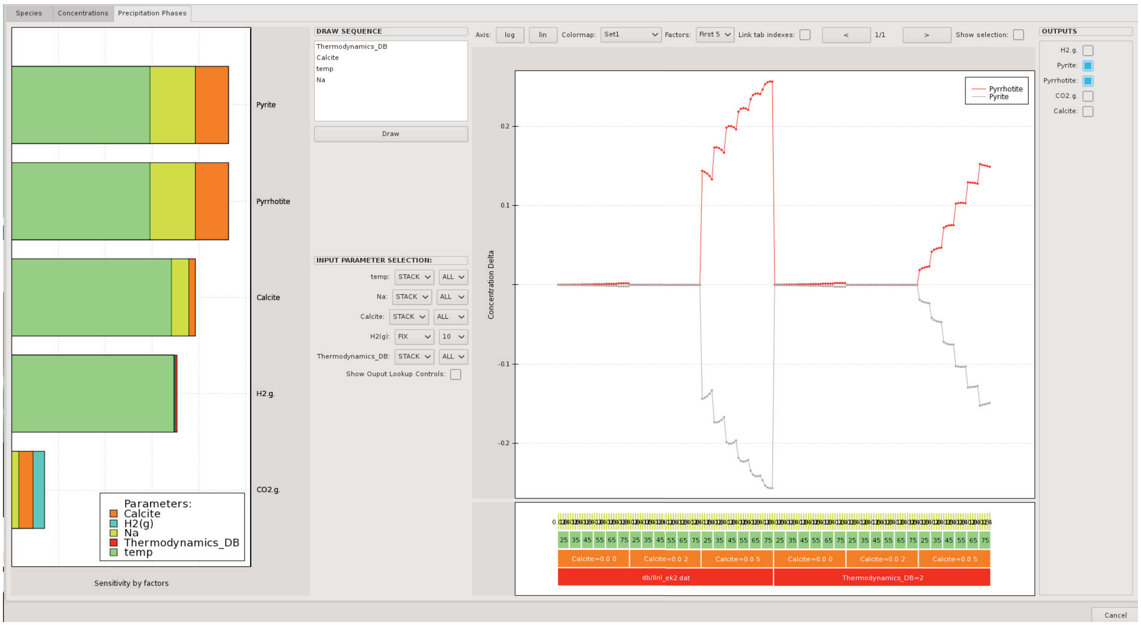
Mineral parametrization

Barite	Calcite	Albite	Dolomite
□ 0.0 0	☑ 0.0 0	□ 0.0 0	☑ 0.0 0
□ 0.0 1	☑ 0.0 1	□ 0.0 1	☑ 0.0 1
□ Excl	□ Excl	□ Excl	☑ Excl
0,88,156)	55,104,0)	0,88,156)	(128,0,0)

Set up with 9000 parameter combinations

Fig. 3: The GCex component for simulation setup and data import/export. The user can visually control range and sampling density for each input parameter, the different thermodynamical databases, the options for parallel computations and some further graphical parameters.

The core graphical component of GCex is the interactive visual exploration GUI (figure 4). It has two essential elements: a *guidance* component (the left region of the tool's window) and the SPR encoding (the right region). Upon import of the input/output data, a sensitivity analysis of standardized regression coefficients [5] is computed. Its results are depicted in a *sensitivity overview visualization* (the left panel of figure (4), which shows through a stacked bar chart [6] the influence of each considered input parameter on each of the output variables. These sensitivities are furthermore used to compute *relationship groups*, which highlight potentially interesting regions of input-output data, thus constituting a starting point for further exploration of the model. On the top of the window, the user finds the control buttons to cycle through the precomputed groups. In our experience, this guidance component makes it very efficient to start the analysis of the data, however it can not replace the expert knowledge of the modeler, since it only captures numerically significant relationships. Again, it is up to the modeler to focus on rather physically meaningful relationships, which is ensured by the interactive design of GCex.



DRAW SEQUENCE: Thermodynamics_DB, Calcite, temp, Na

INPUT PARAMETER SELECTION:
temp: STACK ALL
Na: STACK ALL
Calcite: STACK ALL
H2(g): FIX 10
Thermodynamics_DB: STACK ALL
Show Ouput Lookup Controls: □

Axis: log lin Colormap: Set1 Factors: First 5 Link tab indexes: □ Show selection: □

OUTPUTS
H2.g. □
Pyrite: ☑
Pyrrhotite: ☑
CO2.g. □
Calcite: □

Fig. 4: The core GCex GUI component for interactive exploration of model outputs. On the left is placed a tornado plot for automatic ranking of sensitivities, provided as guidance for the user; on the right the SPR window surrounded by widgets for filtering inputs or outputs and control the ordering of the hierarchical axes. The hierarchy of the nested axes is represented at the bottom of the SPR window.

The main model exploration happens in the SPR panel of the tool. Control widgets to filter the input and output side of the SPR, as well as manually adjust the input hierarchy complete the graphical interface.

2.3. Some technical details

GCex has been built using the open source, multi-platform and freely available languages python (www.python.org), mainly for the graphical part, and R (www.r-project.org) for numerical calculations and data storage. Several third party packages have been employed for this project, and must be installed for GCex to work. From the python side, the Entought Python Distribution is necessary for its component TraitsUI and Chaco, which take care respectively of the dynamic building of interface widgets at run-time and for interactive 2D plotting capabilities. On the R side, the `Rphree` interface to the geochemical program `PHREEQC` (http://rphree.r-forge.r-project.org/) allows the online calculation of geochemical models. Third party R extension packages such as `data.table` and `plyr` enable fast in-memory database for interactively slicing the input-output data and, on supported platforms, the parallelised dispatching of simulations on multiple core CPUs. The latter feature has been built into the graphical interface and is therefore completely transparent to the user.

The simulation setup tool (figure 3) allows the definition of ranges of variation for input parameters and a sampling density for each of them; for mineral at equilibrium, one of three states are possible (as primary, with initial positive concentration; as secondary, with initial zero concentration; or completely absent from the simulations). Due to the large calculation time associated with kinetic simulations, these are currently not implemented in the simulation setup tool. The tool generates a regular sampling grid of the defined multivariate input space: the SPR visualization is best suited for sparse input/output spaces (meaning with no more than 10 levels for each input). For more sophisticated sampling plans, such as local refinements or latin hypercube, the user can easily program it in the R environment and use the provided set of utility functions to quickly generate the combinations table, the input scripts for `PHREEQC` simulations, let them run online or in batch jobs and load the results back for the visualization. In this fashion, virtually any kind of analysis or arbitrary sampling are possible within GCex with a limited amount of programming.

GCex has been designed to deal with large ensembles of simulations. We routinely deal with datasets in the order of few tens of thousand `PHREEQC` simulations with common multicore workstations or even laptops. It is particularly suited, but in principle not limited, to geochemical simulations at equilibrium, which are not much computationally expensive to run.

GCex is currently developed and tested under unix/linux platforms; however, being all of its components and building blocks platform-independent, the porting to MS windows operating systems should be straightforward and will be provided in future.

3. Application example: pyrite-pyrrothite reaction

Excess wind energy stored underground in form of H_2 is discussed as one of possible options for meeting environmental and energetical targets in the near future. The feasibility assessment of such a technology, for regions where no salt caverns can be build, include storage of H_2 in porous media. As consequence, possible fluid-fluid and fluid-rock interactions of hydrogen with minerals and formation fluids have been investigated, also by means of geochemical models.

Truche et al. [7] discuss a reaction involving the injected H_2 potentially occurring at low temperatures: the reductive dissolution of pyrite with formation of pyrrothite, resulting in damage to the host rock and energy and economic loss by consumption of H_2. In their study they investigate three possible scenarios, namely if the H_2-pyrite-pyrrhotite system is unbuffered or it is maintained in alkaline conditions by the presence of calcium carbonate (calcite) or in acidic conditions by large partial pressure of CO_2.

The GCex application example visualized in figure 5 replicates and extends the above study, exploring under which conditions this economically unfavorable reaction is thermodynamically favorable. Included in the analysis are: temperature from 25 to 75 °C; partial pressure of gaseous H_2 from 1 to 10 atm; background salinity (pure NaCl) from 0.01 mol/kgw to 4; the three scenarios, unbuffered system, with calcite (alkaline), with a significant partial pressure of CO_2 (acidic). We have also added the comparison of different thermodynamic databases, namely the `llnl.dat` and the `thermoddem` databases. A total of 1800 `PHREEQC` simulations constitute the dataset for this study, the most meaningful subset of which is visualized in figure 5.

The SPR encoding of the results gives at once a very clear information: in acidic environments the reaction is negligible, whereas under alkaline conditions it is greatly enhanced with respect to the unbuffered system. The reaction is also enhanced at higher temperatures (green axis). The hierarchically last of the represented inputs is the background NaCl salinity, and here again an important fact is captured by the SPR visualisation: for the llnl database, at low temperature an increasing background salinity acts as inhibitor for the reaction, while it reverts this behavior at higher temperature. For the thermoddem database the effect of background salinity is less evident but goes in the opposite direction: at low temperature, increasing the salinity increases the pyrrhotite reaction, and at high temperature the opposite happens. The comparison of different thermodynamical databases also points out that the corresponding calculated speciation, in particular concerning redox reactions and activity coefficients, is of the utmost importance for quantitative predictions. The partial pressure of H_2 plays a secondary role in this context, since the system's behavior stays qualitatively constant in the investigated pressure range.

Through GCex and the SPR encoding, an analysis of this kind is matter of minutes. The modeler gets valuable insights from the overview of model results and from the clear visualisation of the relationships between varying parameters and model outcomes.

4. Conclusion

The Stacked Parameter Relation (SPR) encoding allows seeing the responses of multiple model outputs to multiple input parameters over their entire ranges. This has an excellent synthesis power when exploring broad ranges of values for many parameters. It is an effective way to display multivariate sensitivities or, in a more general way, input/output relationship. Although particularly suited to equilibrium geochemical models, where the input parameters are a mix of boolean, discrete and continuous variables, it is a fully general visualization method, completely model-agnostic,

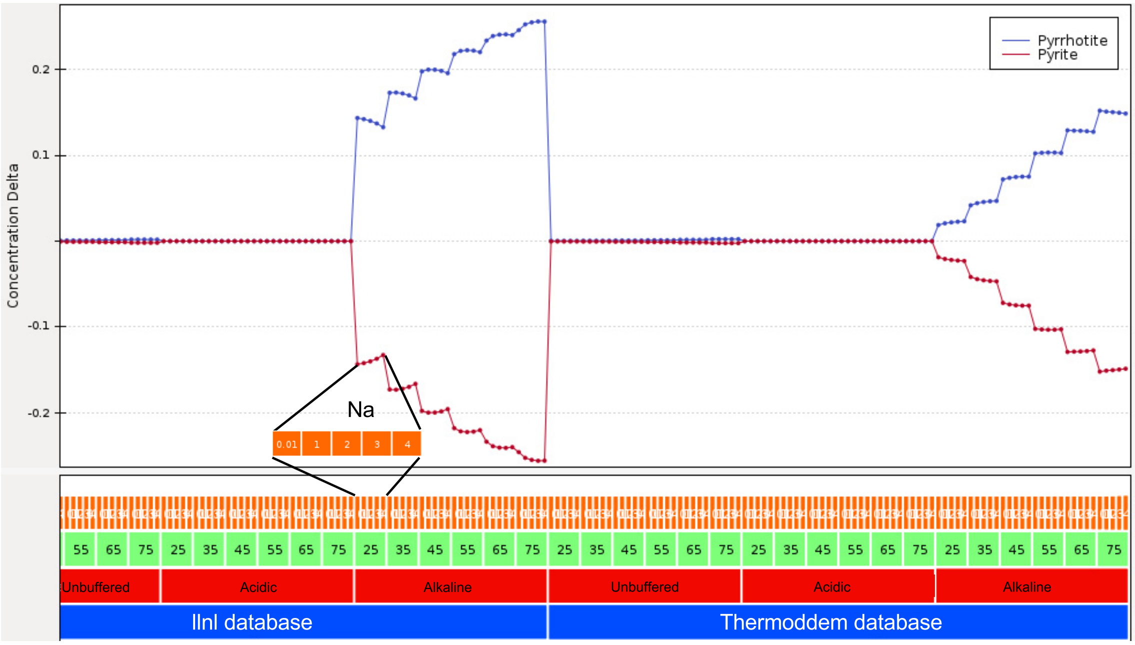


Fig. 5: SPR overview of the analysis of H_2/pyrite/pyrrhotite reaction for different databases, under unbuffered/acidic/alkaline conditions, from 25 to 75 °C and from 0.01 to 4 molal NaCl background salinity.

and can be applied to arbitrary models in every domain. It is best suited for sparse samplings of parameter spaces, and scales well up to 4 or 5 inputs visualised at the same time.

GCex is a graphical tool implementing SPR as well as the complete toolchain for running combinatorial analysis of geochemical models through the PHREEQC geochemical reaction and speciation program. It is designed for flexibility, efficiency and modularity, and scales well up to several tens of thousands simulations on common workstations. It supports the exploration of model results and the visual interpretation of sensitivity and uncertainty analysis. Therefore, it deepens the understanding of model's behavior through Visual Analytics, in an intuitive and productive way.

For further information, examples, documentations and download of GCex the interested reader is referred to the project page www.gfz-potsdam.de/en/section/geoinformatics/projects/gcex/

References

[1] De Lucia M, Kühn M. Coupling R and PHREEQC: Efficient Programming of Geochemical Models, Energy Procedia 2013;40:464-471, doi:10.1016/j.egypro.2013.08.053.

[2] Dethlefsen F, Haase C, Ebert M, Dahmke A. Uncertainties of geochemical modeling during CO_2 sequestration applying batch equilibrium calculations, Environ Earth Sci 2011;65:4,pp 1105-1117, doi:10.1007/s12665-011-1360-x.

[3] Parkhurst D, Appelo C. Users guide to PHREEQC (version 2). Tech. rep., U.S. Geological Survey, 1999.

[4] Mihalisin T, Timlin J, Schwegler J. Visualization and analysis of multi-variate data: a technique for all fields. In: IEEE Computer Society Press, Proceedings of the 2nd conference on Visualization, pp. 171-178, 1991.

[5] Saltelli A, Ratto M, Tarantola S, Campolongo F. Sensitivity analysis for chemical models. Chem Rev 2005;105:2811-2828.

[6] Gratzl S, Lex A, Gehlenborg N, Pfister H, Streit M. Lineup: Visual analysis of multi-attribute rankings. IEEE Transactions on Visualization and Computer Graphics 2013;19(12), 2277-2286.

[7] Truche L, Jodin-Caumon MC, Lerouge C, Berger G, Mosser-Ruck R, Giffaut E, Michau N. Sulphide mineral reactions in clay-rich rock induced by high hydrogen pressure. Application to disturbed or natural settings up to 250 C and 30 bar. J Chem Geo 2013;351:217-228, doi:10.1016/j.chemgeo.2013.05.025.