TCIII-2019-09E

ThermoChimie

ThermoChimie Benchmarking Workshop Report

Edited by A.J. Fuller (Galson Sciences Ltd)

December 2019

The ThermoChimie database was first developed in 1995 by Andra, the French national radioactive waste management agency. They have since been joined by Radioactive Waste Management (RWM) from the UK, and ONDRAF/NIRAS from Belgium.

ThermoChimie provides an accurate and consistent set of data specifically chosen for use in modelling the behaviour of radionuclides in waste packages, engineered barriers, and both the near surface and deep geosphere. The database can be used to model the speciation and solubility of a wide range of stable and radioactive elements, organics, and solid phases including cements, clay minerals and degradation products (such as zeolites). The database is suitable for use within the range of conditions expected in both near-surface and geological disposal facilities: pH 6-14, ionic strength up to SIT, Eh within the stability fields of water, and temperatures from 15 to 80°C.

ThermoChimie is intended for use across the radioactive waste management community, to support repository performance assessment, research and development activities and decision making. To maximise their utility the data are therefore provided in formats suitable for use with common geochemical modelling codes. The database can be viewed and downloaded from the project website: https://www.thermochimie-tdb.com/, where additional information and supporting documents are also available.

This document provides a summary of the expert opinions gathered at the ThermoChimie Benchmarking workshop held on the 15th October 2019 in Manchester, UK.

Version 1 – 16 December 2019
Authored by: F. Bok, S. Salah, T. Heath, L. De Windt, K. Spahiu and D. Reed

Edited by: A.J. Fuller (Galson Sciences Ltd)

Approved by: The ThermoChimie steering committee: B. Madé (Andra), R. Hibberd, (RWM), and S. Brassinnes (ONDRAF/NIRAS)

Contents

ThermoChimie Benchmarking Workshop Report

1 Introduction

ThermoChimie is a thermodynamic database initially created and developed by Andra (French National Radioactive Waste Management Agency), and under development for more than twenty years (since 1995). In October 2014, Radioactive Waste Management Limited (UK) joined the project and the ThermoChimie consortium was formed. In March 2018, ONDRAF/NIRAS (National Agency for Radioactive Waste Management, Belgium) also joined the ThermoChimie consortium.

The planned ThermoChimie (TC) benchmarking exercises aim to evaluate the performance of the ThermoChimie database when compared to other high-quality internationally recognised thermodynamic databases and test it against well-selected and high-quality experimental datasets. This should allow the strengths and weaknesses of the ThermoChimie database to be identified. The outputs of the subsequent benchmarking exercises will be used by the ThermoChimie steering committee to define the future work programme and allow further improvements to be made to the ThermoChimie database so that it better meets the consortium members' needs.

In practice, the benchmarking exercises will focus on testing a given set of geochemical systems of relevance to the geological disposal of radioactive waste (and possibly other geochemical systems for which the ThermoChimie database might be fit for purpose). The evolution of the geological disposal system involves numerous processes (e.g., speciation, solubility, sorption, complexation) and influencing factors (e.g., temperature, various ligands, chemical disturbances). The benchmarking exercises may also consider how temperature, redox potential and ionic strength (and ion interactions theories) affect elemental speciation and related solubility values.

Opinions have been sought from a range of experts, and this document summarises the outcomes of a workshop organised by the ThermoChimie consortium held in Manchester on 15th October 2019 with invited experts in the field of thermodynamic database development and geochemical modelling. Prior to the workshop, the expert opinions were collected and shared among the attendees to facilitate the discussion at the ThermoChimie benchmarking workshop. These opinions respond to a series of questions posed by the ThermoChimie Steering Committee:

1. In your opinion, what are the possible uses of a benchmarking exercise making use of thermodynamic databases? Please, illustrate this with a set of examples of previous benchmarking exercises you have been involved in.

2. In your opinion, what are the most relevant types of modelling for the ThermoChimie benchmarking exercise? In other words, should we only model chemical process at equilibrium (which ones?) or do you think we also need to tackle close to equilibrium processes and/or reactive transport?

3. In your opinion, which geochemical systems should be considered in this benchmarking exercise? In other words, what are the geochemical systems most likely to reveal the strengths and limitations of the ThermoChimie database?

4. Should the benchmarking exercises be based on modelling results only or should the benchmarking refer to a selected set of well-defined experimental and/or natural (e.g., groundwater composition…) datasets as in the case of validation exercises?

5. Which other high-quality databases should be considered for the benchmarking exercise? Which codes should be used for the benchmarking exercise?

6. Should the uncertainties provided in the database be used for benchmarking?

7. In your opinion, what are the essential points of focus that need to be covered for the benchmarking exercise to usefully evaluate thermodynamic databases?

The ThermoChimie steering committee will use this synthesis of the experts' recommendations in conjunction with the needs of the waste management organisations, to define future benchmarking exercises. This will ensure that best practice is followed when evaluating the ThermoChimie database.

2 Expert Opinions

This section contains written responses from the experts to the numbered questions, shown in bold.

2.1 Frank Bok

1. What are the possible uses of a benchmarking exercise making use of thermodynamic databases?

Besides the direct comparison against other databases, other useful information can be obtained from benchmarking calculations:

1) The data quality itself (especially with respect to data from chemical analogues or estimates) through direct comparison with experimental data.
2) Validity of internal calculation routines as well as data export parsers.
3) Behaviour of the data when using different geochemical codes (including information on numerical stability that can be affected mostly by highly charged species or complex chains of redox couples).
4) Identification of missing data, typos, errors or consistency problems.
5) Consequences of data updates, removals and additions (especially unplanned side effects).

In addition, data-independent effects can also be identified, for example necessary adjustments of export parsers to new versions of the supported geochemical codes.

2. What are the most relevant types of modelling for the ThermoChimie benchmarking exercise?

Since ThermoChimie is a compilation of thermodynamic data (as well as all the other databases mentioned in question 5), the modelling should focus on chemical processes at equilibrium. It might be useful to also include examples containing pseudo-equilibria such as the solubility of freshly precipitated amorphous mineral phases, containing radionuclides – many thermodynamic databases contain data for such phases because of their expected relevance for radionuclide migration/retention.

Reactive transport scenarios are often too complex to identify the source of deviations between modelling results and experimental data. Furthermore, reactive transport calculations contain processes that are not considered in (many of) the thermodynamic databases (e.g. sorption, ion exchange, microbial activity, system heterogeneities …).

3. Which geochemical systems should be considered in this benchmarking exercise?

Taking into account the different purposes of the databases used in the benchmarking exercise, well-defined systems should be chosen. Proposed examples could be:

1) Solubility of radionuclide phases under given conditions of pH, Eh, ionic strength, pCO_2, etc. Points of phase transition are to be preferred because they are often well-defined and allow several thermodynamic values and interaction parameters to be checked simultaneously with one calculation.

2) Material corrosion (steel) or alteration (clay minerals) in a given solution. Here solutions with higher ionic strength would be an ideal test of the used activity model (SIT) and the chosen thermodynamic data to reveal the strengths and limitations of the ThermoChimie database. This might be also of interest because pore waters with high ionic strengths (up to 3 molal) have been measured in certain clay formations.
3) To show the strengths and limitations of ThermoChimie, a chemical scenario should be modelled using ThermoChimie's SIT approach in comparison with other activity models (e.g. Pitzer). The ranges of ionic strength over which these two approaches are valid overlaps (SIT: up to 3–5 molal, Pitzer: 3–15 molal), so this could reveal possibilities to mutually close data gaps in both approaches.

4. Should the benchmarking exercises be based on modelling results only?

No, if benchmarking is conducted without including any comparison between modelling results and experimental data, only very limited statements could be made about the quality of the data in the ThermoChimie database in comparison with other databases.

High-quality data from well-defined experimental scenarios should come from peer-reviewed publications and be evaluated by an expert before comparing it with modelling results.

5. Which other high-quality databases / codes should be considered for the benchmarking exercise?

Besides the ThermoChimie database, the following thermodynamic databases are actively maintained and should be considered in the benchmarking exercise: THEREDA[1], OECD/NEA Thermochemical Database[2], PSI/Nagra Chemical Thermodynamic Database[3], JAEA Thermodynamic DataBase[4] and Thermoddem Geochemical Database[5]. The majority of these databases provide data for the SIT activity model, but different activity models and chemical speciation should be taken into account as well as the original purpose of these databases. Due to difference in the activity models and database purposes, only limited general statements can be made about the overall quality of the individual databases.

Concerning the tailored parameter files provided by the individual database projects for the various geochemical codes, the benchmark exercise should be performed using PHREEQC[6] or, with a few exceptions, Geochemist's Workbench[7].

6. Should the uncertainties provided in the database be used for benchmarking?

It is difficult to include uncertainty data in geochemical modelling of more complex systems, as the types of uncertainty are often completely different (confidence interval, standard derivation, variance, experimental / analytical parameter, estimations or unclear ranges). The values have to first be unified to make them comparable. In addition, the authors of the primary data often understate the uncertainty.

Furthermore, consideration of the uncertainties on the thermodynamic data or interaction parameters during the geochemical calculation is not natively supported by any code. Code-coupling with some Monte-Carlo-like software and statistical

evaluation is possible but can become quite complicated for systems containing many species' log$K°$ values including interaction parameters (SIT, Pitzer). A more effective approach might be the comparison of the modelling results with quality-assured (peer-reviewed) experimental data and consideration of their uncertainty information.

7. What are the essential points of focus that need to be covered for the benchmarking exercise to usefully evaluate thermodynamic databases?

The focus of the benchmarking activity should be on the numerical reproduction of quality-assured experimental data for specific chemical scenarios. Ideally, these experimental data should not be those used to obtain the thermodynamic parameters in ThermoChimie. A large number of smaller, well-defined chemical scenarios (invariant points, phase transitions, etc.) can be used to test a wide range of data. Possible deviations in the obtained calculation results away from known experimental data allow conclusions to be drawn about the reason for these differences. It is not as easy to draw these conclusions using a complex reactive transport calculation.

A further objective of the benchmarking exercises should be to identify the species in the databases that are responsible for the differences between the measured values and the modelled results. Moreover, great emphasis should be put on quality assurance. This includes automated calculation routines, complete and open documentation and an audit.

References

Hyperlinks were accessed on September 26th, 2019.

1. THEREDA: https://www.thereda.de
2. OECD/NEA Thermochemical Database: https://www.oecd-nea.org/dbtdb/tdbdata/
3. PSI/Nagra Chemical Thermodynamic Database: https://www.psi.ch/en/les/database
4. JAEA Thermodynamic DataBase: https://migrationdb.jaea.go.jp/cgi-bin/db_menu.cgi?title=TDB&ej=1
5. Thermoddem: http://thermoddem.brgm.fr
6. PHREEQC: https://www.usgs.gov/software/phreeqc-version-3
7. Geochemist's Workbench: http://www.gwb.com

2.2 Sonia Salah

1. **In your opinion, what are the possible uses of a benchmarking exercise making use of thermodynamic databases? Please illustrate this with a set of examples of previous benchmarking exercises you have been involved in.**

   In our opinion, there are two main uses of benchmarking exercises:

   1) to compare the ability of different TDBs (thermodynamic databases) to adequately (and accurately) model key processes (with or without kinetics) and reactions (i.e. aqueous & surface complexation, acid-base & redox reactions, cation-exchange reactions, precipitation & dissolution reactions, etc.….) defined in the conceptual model of the system under consideration.

   2) to determine the behaviour of radionuclides in waste packages and disposal systems over the long-term using reactive transport/flow-through calculations.

   Examples:

   - Benchmarking TDB using speciation/solubility calculations with GWB/MOLDATA – Sensitivity analysis; L. Wang (2016): Other used TDB's: THEREDA and ThermoChimie v.9b.0.

     Aim: Assess the influence of different water compositions and chemical conditions on radionuclide (i.e. Am, Eu, Mo, Np, Pu, U) speciation & solubility. Benchmarking involved conducting speciation calculations as a function of changing pH, DIC and ionic strengths, and plotting Pourbaix diagrams (Eh-pH) and solubility curves as a function of changing pH, DIC and ionic strengths.

   - Scientific review (organised by Andra) of the state-of-the-art (of ThermoChimie v.9) on organic thermodynamics; C. Bruggeman & S. Salah (2015).

     Aim/main questions: Assess which would be the most sensitive organic species with respect to their ligand properties under cementitious and under natural groundwater conditions (in clay & crystalline systems). State of knowledge about the stability of organic complexes, and which accurate estimation methods could be used to obtain lacking data? Assessment of the most relevant organic systems/conditions to be studied experimentally?

2. **In your opinion, what are the most relevant types of modelling for the ThermoChimie benchmarking exercise? In other words, should we only model chemical process at equilibrium (which ones?) or do you think we also need to tackle close to equilibrium processes and/or reactive transport?**

   The most relevant calculations to test the performance of one or more TDB's are considered to be thermodynamic equilibrium calculations, such as: speciation & solubility calculations at 25°C under different chemical conditions (including organic ligands/species and higher temperatures), plotting predominance diagrams to check phase relations as a function of temperature or different activity ratios, and plotting activity diagrams. Other calculations could comprise:

   - testing the impact of different activity formulism's in systems with ionic strengths > 0.3 M,
   - Rn-sorption, which strongly depends on speciation,

- Calculating gas solubilities at 25°C, as well as at higher temperatures,
- Redox reactions/disequilibrium,
- Formation of solid solutions.

We would suggest to perform these types of calculations, using the same code to enable discrimination of the "real" differences related to the TDB and not to the code (i.e. the mathematical/numerical model implementation).

Reactive transport calculations are considered to be less relevant, as they are/were used (in the past) to "test" different codes and transport simulation capabilities rather than the TDB quality. In most cases reactive transport calculations are performed to test the robustness, accuracy, stability and efficiency of different codes, as well as to validate the implemented numerical model(s) (e.g. coupling and discretisation schemes, algorithms and iterations, etc.). They may however be useful tools to validate different thermodynamic data file formats, as they are code-specific and not TDB-specific. Besides this, they may reveal how minor differences observed in equilibrium calculations may evolve over longer time periods (be cumulative or even out).

3. **In your opinion, which geochemical systems should be considered in this benchmarking exercise? In other words, what are the geochemical systems most likely to reveal the strengths and limitations of the TC database?**

   - Clay systems, backfill/containment materials (e.g. bentonite: Eh-pH control by accessory minerals, such as pyrite, calcium carbonate, gypsum, and quartz),
   - High pH (cement/concrete), higher temperature (near-field), higher ionic strength systems & their combinations,
   - Systems with steep chemical (pH, Eh) & concentration gradients, e.g. clay/cement interfaces,
   - Systems involving changing redox conditions (due to e.g. microbial activity or corrosion processes).

4. **Should the benchmarking exercises be based on modelling results only or should the benchmarking refer to a selected set of well-defined experimental and/or natural (e.g., groundwater composition…) datasets as in the case of validation exercises?**

   It is definitely a good idea for the benchmarking to reference to well defined experimental and/or natural datasets/analogues. In our experience, verification of the thermodynamic data in ThermoChimie (and definitely for the NEATDB) already involves comparison to experimental data (e.g. solubility experiments). Such comparisons are however not always straightforward, as thermodynamic properties only apply to phases of defined composition (e.g. amount of bound water). Often variation in composition (e.g. C-S-H, clays) and/or structure (e.g. zeolites), as well as the presence of impurities may lead to poor (or even bad) reproducibility of experimental and/or natural datasets.

5. Which other high-quality databases should be considered for the benchmarking exercise? Which codes should be used for the benchmarking exercise?

Other TDB's to be considered:

- PSI-NAGRA (Thoenen et al., 2014; Hummel et al, 2002)
- MOLDATA (Wang et al., 2010)
- JAEA (Kitamura et al. (2014)
- Yucca Mountain TDB (Wolery and Jove-Colon, 2007; Johnson et al., 1992; Oelkers et al., 2009)

Less relevant:

- THERMODDEM (Blanc et al., 2012), as part of ThermoChimie or similar (but smaller)
- HATCHES (Cross and Ewart, 1991)
- THEREDA (Moog et al., 2015; Gester et al., 2009), as mainly applicable to higher ionic strength systems and database is quite "small".

Codes: PHREEQC, Geochemist's workbench

6. Should the uncertainties provided in the database be used for benchmarking?

Indeed, it would be nice to include uncertainties in the benchmarking, but that seems almost impossible/very unrealistic. TDBs don't (i.e. the electronic format) include the uncertainties, and most of the common codes cannot include all uncertainties when performing calculations.

7. In your opinion, what are the essential points of focus that need to be covered for the benchmarking exercise to usefully evaluate thermodynamic databases?

- Internal consistency and completeness of database
- Traceability/documentation of data source and selection
- Data gaps/weaknesses
- Used estimation and extrapolation methods
- Available database formats (implementation of different activity models)

2.3 Tim Heath

1. **In your opinion, what are the possible uses of a benchmarking exercise making use of thermodynamic databases? Please illustrate this with a set of examples of previous benchmarking exercises you have been involved in.**

   Some possible applications of benchmarking activities are:

   - Predicting and correcting groundwater compositions and the effect of mixing groundwaters with each other or other porewaters (e.g. the thermodynamic benchmarking exercise under the Grimsel CFM project).

   - Predicting radionuclide solubility and speciation in specific waters for specific conditions (e.g. the thermodynamic benchmarking exercise under the Grimsel CFM project; the database comparison of ThermoChimie with HATCHES for RWM).

   - Testing of radionuclide sorption and uptake models. But note that this requires some sorption model specification that is not included in most thermodynamic databases (e.g. the NEA Sorption Forum project).

   - Predicting the evolution of key engineered barrier materials, including their interaction with groundwater and, potentially, waste components. These should include cementitious materials and bentonite. (the database comparison of ThermoChimie with HATCHES performed for RWM before joining the ThermoChimie consortium).

   - Testing of the use of uncertainty values for data in thermodynamic databases for use in the propagation of uncertainty to predict pH values, water compositions and radionuclide solubility (e.g. the thermodynamic benchmarking exercise under the Grimsel CFM project).

2. **In your opinion, what are the most relevant types of modelling for the ThermoChimie benchmarking exercise? In other words, should we only model chemical process at equilibrium (which ones?) or do you think we also need to tackle close to equilibrium processes and/or reactive transport?**

   I consider the most relevant types of modelling for the ThermoChimie benchmarking exercise to be:

   - Prediction of the saturation state of specified groundwaters with respect to relevant mineral phases, and the effects of mixing groundwaters

   - Radionuclide chemistry, including:

     - Speciation in selected groundwaters and near-field porewaters

     - Prediction of redox boundaries for key radionuclide oxidation state transitions

- Effect of important organic complexants on radionuclide speciation and solubility

- But probably not including radionuclide sorption because the mechanisms and data for sorbed species are not part of the database so it is not a good test of the database

- Cement mineral assemblage and bentonite performance, including:

  - Prediction of the initial mineralogy and porewater composition from a specified formulation

  - Prediction of the evolution of the mineral assemblage and conditioned porewater composition due to interactions with groundwater (and possibly waste components)

- Potentially consider the effect of ionic strength and temperature on a selection of the above examples.

- Consider probabilistic modelling to assess the use of uncertainty values taken from the database

Regarding the potential consideration of tackling "close to equilibrium processes and/or reactive transport", I think this needs careful consideration if it implies the inclusion of kinetic reactions. At the moment ThermoChimie is purely a thermodynamic database and includes no kinetic data. If benchmarking exercises involving kinetic reactions, as key components, are developed, these will not be a good test of the database as the results will be strongly dependent on kinetic data supplied as part of the exercise, but not part of the database. If consideration is to be given to including some kinetic reactions as a part of the database, this would require careful and separate consideration and discussion. If the object of the exercises is to compare the effects of data differences between databases, then keeping the exercises simple (based on equilibrium chemistry and no or very simple transport) will aid the comparison of results and identification of the most important differences.

3. **In your opinion, which geochemical systems should be considered in this benchmarking exercise? In other words, what are the geochemical systems most likely to reveal the strengths and limitations of the TC database?**

   The systems above in the answer to the previous question should be considered. This might include variation of the ionic strength of groundwaters and porewaters within the defined range for which ThermoChimie has been developed. But it would exclude evaporite-based scenarios with very high ionic strengths. The effect of elevated temperature on porewaters associated with bentonite materials as well as clay and hard rock mineral phases might also be considered within the temperature limits of ThermoChimie.

4. **Should the benchmarking exercises be based on modelling results only or should the benchmarking refer to a selected set of well-defined experimental and/or natural (e.g., groundwater composition…) datasets as in the case of validation exercises?**

I consider both approaches to be valid. But it should be made clear for each exercise whether the objective is to test ThermoChimie against the corresponding datasets in other databases, or whether it is to test ThermoChimie against its ability to predict or explain observed data. This is because it would be wise to avoid situations where differences in results are due to the combined effects of different modelling decisions and different thermodynamic databases, but separation of the two effects is difficult.

5. **Which other high-quality databases should be considered for the benchmarking exercise? Which codes should be used for the benchmarking exercise?**

Other Databases

- Nagra/PSI (plus CEMDATA where relevant)
- PHREEQC standard database (geochemical data)
- Nuclear Energy Agency thermodynamic database NEA/TDB (radionuclides and supporting data only)
- LLNL database (derived from databases for EQ3/6 and Geochemist's Workbench)
- THERMODDEM database; Pitzer.DAT (available with PHREEQC); databases developed for use at high ionic strengths but also applicable at lower values.
- (others MINTEQ, WATEQ4F, JAEA Thermodynamic database)

Codes

- PHREEQC
- ToughReact
- PFLOTRAN
- Geochemist Workbench

6. **Should the uncertainties provided in the database be used for benchmarking?**

This would certainly be useful, particularly in the case of the UK approach to the treatment of uncertainty. It would provide consistency with the assessment modelling approach based on random sampling of input parameter values and probabilistic calculations to propagate the uncertainty to the output results.

7. In your opinion, what are the essential points of focus that need to be covered for the benchmarking exercise to usefully evaluate thermodynamic databases?

Testing of datasets for key complexants with selected radionuclides under alkaline cementitious conditions, particularly for systems recently studied in high quality work, and for any planned experiments for subsequent comparison with experimental results.

Testing of the ThermoChimie data for cementitious mineral phases: this area of ThermoChimie has not been reviewed in recent years and is a long way behind the state-of-the-art database (i.e. CEMDATA)

Testing predictions of bentonite behaviour at higher temperatures.

2.4 Laurent De Windt

1. Possible uses of a benchmarking exercise making use of thermodynamic databases

The first goal of the benchmarking exercise is to compare results output from at least 3 international thermodynamic databases (TDB) for systems with multiple interacting components. This exercise will evaluate the completeness of the dataset, assess its quantitative results, and check its robustness and internal consistency (see Q7 for details).

The benchmarking study should be of significant interest to the scientific and/or disposal safety communities and draw the attention of the wider scientific community to the TC database. In this context, some points of comparison with experimental or natural data from complex systems can be an effective "benchmark", but careful account has to be taken of the full experimental conditions.

The benchmarking could be performed as a set of targeted scientific papers grouped into a dedicated volume. This approach is currently followed when benchmarking reactive transport models (RTM).

Some examples of similar benchmarking exercises include:

- RTM benchmarking of cement/clay interfaces, focusing on the TDB and solid phases,
- sensitivity analysis on the impact of redox potential on the speciation of actinides in clay (COx) groundwater and steel environment,
- the ThermoBridge internal TC benchmark (partly using our geochemical code CHESS) and some recent database benchmarks from the literature.

2. Most relevant types of modelling for the TC benchmarking exercise

Most of the exercises can be done with a geochemical code at thermodynamic equilibrium. This would clearly focus on the core data in TC, but also minimize numerical uncertainties brought about by modelling kinetics or RTM. Titration models (covering a range of pH, Eh or species concentrations) are useful in addition to speciation calculations of a single solution. The mixing of two geochemical systems is less relevant than RTM. Activity-activity diagrams can be used to highlight differences between TDBs when modelling the same system.

RTMs in a simple configuration (1D regular mesh) are helpful for complex binary systems, such as cement/clay interfaces. The spatial distribution better discriminates between the multiple reactive fronts and ranges of aqueous concentrations. The combination of TC and RTM is also positive from a communication point of view.

The issue of solid solutions vs. discrete phases is essential for the cement phases (C-S-H, AFm…) and, to a lesser extent, the cationic end members of clay phases (Na- vs. Mg-smectite…) vs. cation exchange models.

(Near-equilibrium modelling can smooth sharp transitions between mineral phases of similar formation constants, useful while comparing TDB, but saturation indices bring the same information).

3. Geochemical systems of relevance for the TC benchmarking exercise

The speciation of several key radioactive elements in the porewater of deep clay formations is relevant (e.g. the COx porewater, at midway between diluted and saline clay chemistry). The selection of elements has to balance remaining uncertainty on their speciation with interest for the safety of underground (and subsurface?) disposals; for instance, U, Pu, and Se. Speciation focuses on aqueous chemistry and the saturation indices of all solid phases in equilibrium with a selected set of solid phases relevant for solubility limits.

The effect of the redox potential on speciation at a fixed pH can be addressed, e.g. the competition between the ternary Ca(Mg)-CO_3-U complex and the U(VI)/U(IV) redox couple (to simulate a transient stage from oxic to anoxic conditions). Similarly, the sensitivity of speciation to a temperature decrease is relevant due to the temperature changes expected in disposal facilities, but also the large set of TC data over 10 – 90 °C. The effect of salinity (ionic strength) is maybe less essential for the TC benchmarking exercises since it depends more on the activity model used than the TDB itself. The question of phosphates or hydrophilic organic species (e.g. carboxylic acids) can be debated.

Binary cement/clay and iron/clay are key systems in geological disposal that can be assessed using RTMs (see Q2). Cement/clay is maybe the easier system to benchmark since cement phases are mostly under thermodynamic equilibrium, there is a large amount of constant data available for the cement phases and cement-based materials are ubiquitous in disposal systems. The clay phases are also well represented in TC. The literature on RTM benchmarking has to be taken into account to avoid replication. Corrosion products of iron in a natural water at different temperatures can be used as an alternative to the iron/clay interface (although metastable phases can coexist for kinetic reasons).

(A benchmark exercise coupling radionuclide speciation and engineered barrier evolution will illustrate that TC is able to simultaneously tackle these two aspects of performance and safety assessment; although sorption cannot be considered in TC).

4. Validation through modelling results or well-defined experimental/natural datasets

The validation of thermodynamic data on experimental data is more the job of the NEA expert groups than the TC benchmarking one. I would put a larger focus on comparing the consistency and completeness of TC against other databases, but one or two real systems can be added to the process.

The validation of TC calculations on real systems is useful for large multi-component systems or for recent issues such as ternary actinide complexes. Furthermore, comparison with complex real systems brings confidence and visibility to TC.

5. Databases and codes for the TC benchmarking exercise

TC is very powerful for computing radionuclide chemistry as a constant effort has been made to compile most of the data from the NEA “blue books”. TC is also well capable of simulating the evolution of the engineered and geological barriers (clay phases,

corrosion products, cement phases). I am not sure than the other TDBs cover these aspects so well (it would perhaps be necessary to select a different set of TDBs for some exercises).

CEMDATA2018 is essential for cement phases but also highly specialised. The NAGRA/CEMDATA is a more ubiquitous choice. A non-European TDB would be welcome, such as the LLNL TDB (or the JAEA TDB provided a Japanese colleague is involved in the benchmarking process). (I do not know how different the NEA and MOLDATA TDB are from TC, THERADA seems to focus on highly saline environments.)

The simplest way is to use the PhreeqC code since it supports most TDB and has several activity-correction models. This approach minimizes the numerical discrepancies among codes or the possibility of introducing errors while extracting the data for each code. However, the chance of realising these two risks is rather low and other codes (such as CHESS) can be used if it eases the management of the project.

6. Uncertainties in the database for the TC benchmarking exercise

It is very important than the TC web site and documentation provide an estimation of the uncertainty for each thermodynamic constant (but too much information in the database itself makes the file less readable).

(Uncertainty propagation for a multi-component system could be interesting but is maybe out of the scope of TC benchmarking. I do not have experience on probabilistic algorithms used in safety analysis.)

7. Essential points of focus for benchmarking to evaluate thermodynamic databases

- To use multi (interacting) component systems and elements of interest to both the scientific community and safety assessors; one strength of TC is its ability to combine multibarrier materials and radionuclide speciation;
- to evaluate the completeness of the dataset (lack of essential reactions or secondary phases) for the behaviour or an element (e.g. Pu) or solid phase (e.g. corrosion products) in multi-component, complex, and realistic systems;
- to assess its quantitative results (solubility, relative proportion of aqueous complexes, Mg/Si ratio in M-S-H…);
- to check the robustness and internal consistency of the database when changing a key parameter, e.g. the redox potential or the temperature.

The use of an RTM with a simple spatial configuration may be useful for evaluating the mineralogical evolution at interfaces. Comparison with a few complex real systems brings confidence and visibility to TC. At the end of the process, TC benchmarking can lead to a set of targeted scientific papers.

2.5 Kastriot Spahiu

1. **In your opinion, what are the possible uses of a benchmarking exercise making use of thermodynamic databases? Please illustrate this with a set of examples of previous benchmarking exercises you have been involved in.**

   I found in Wikipedia that "Benchmarking is the practice of comparing business processes and performance metrics to industry bests and best practices from other companies". The term used back in the 80s and 90s was comparison of databases used in geochemical modelling. In that case the same geochemical code was used with various databases. I have participated in the later stages of the EU-project Chemval [1, 2]. Another example of database comparison is a White Report[3] sent to NEA-TDB by Tom Wollery in 2005, pointing out discrepancies between the NEA-TDB and the NIST and other databases which were used for the EQ 3/6 database (it was considered quality assured for the Yucca Mountain project). Other examples may be found in a publication by B. Merkel –Thermodynamic data dilemma [4], where differences in U(VI) speciation predicted by various databases and calculation codes are reported. Emren et al [5] have compared solubilities of $Pu(OH)_4$ calculated by four different modellers and found a large influence of the modeller in the calculation results.

2. **In your opinion, what are the most relevant types of modelling for the ThermoChimie benchmarking exercise? In other words, should we only model chemical process at equilibrium (which ones?) or do you think we also need to tackle close to equilibrium processes and/or reactive transport?**

   It is my understanding that Thermochimie is a thermodynamic database thus it contains thermodynamic data at standard state for all possible species and can be used mainly to calculate chemical equilibrium data, such as chemical speciation in aqueous systems with several components including solid phases. Of course, this does not exclude its use in estimating e.g. Gibbs energies of formation of intermediary states and thus estimating energy barriers in kinetic studies or its use to calculate speciation when surface complexation modelling is used for sorption.

   I struggle to grasp its use in close to equilibrium processes (the dissolution rate of a solid decreases as you approach equilibrium, but this is a kinetic and not thermodynamic issue) while I see no problem with its use to calculate chemical speciation at equilibrium in each cell considered in a reactive transport model. Aqueous complexation or redox reactions are usually fast enough to reach equilibrium in each calculation cell, but exceptions to this rule cannot be excluded.

3. **In your opinion, which geochemical systems should be considered in this benchmarking exercise? In other words, what are the geochemical systems most likely to reveal the strengths and limitations of the TC database?**

   My understanding is that a thermodynamic database can be used with an appropriate geochemical calculation code to estimate the equilibrium concentrations of all species and the solubilities inside e.g. a damaged high-level waste container or in a storage room of a LILW repository, by making the appropriate assumptions. In SKB´s LILW case, solubility limits are seldom reached, especially if one accounts for sorption equilibria, which are usually faster. Besides its use in reactive transport, I don't think calculation of radionuclide solubilities is relevant in the far field. It is my understanding

that Andra had the ambition to calculate clay and cement mineral equilibria, but I am not well acquainted with the progress in this field. On the other hand, this is an important point which also affects the calculation of radionuclide solubilities, as discussed e.g. by Emren et al. [5].

4. **Should the benchmarking exercises be based on modelling results only or should the benchmarking refer to a selected set of well-defined experimental and/or natural (e.g., groundwater composition…) datasets as in the case of validation exercises?**

   I would say both would be valuable, even though I am not an expert in this field. In connection with this task, I read a couple of very good articles about model validation (Sheng et al. [6], Nordstrom [7]). In my opinion, these studies (especially Sheng et al. [6]) give a useful summary of how a benchmarking or validation exercise should be carried out and they also discuss the importance of choosing appropriate experimental data for comparison.

5. **Which other high-quality databases should be considered for the benchmarking exercise? Which codes should be used for the benchmarking exercise?**

   There are a number of databases that have a good reputation in our field, such as the one used by Nagra, Switzerland or JAEA, Japan and consideration should be given other ones too, such as database of Eq 3/6 used in US or Thereda in Germany. The results of the benchmarking exercise will depend on the database and the calculation code used. I would be very careful with the choice of codes, mainly concerning the approach they use for ionic strength corrections. Many years ago, I was using Phreeqe to calculate the speciation in a groundwater containing high NaCl concentrations and U(VI) and I found mainly hydrolysed U(VI) instead of the expected carbonate complexes. I was forced to take away the NaHCO3 complex from the database to get the correct speciation. I quickly checked the Phreeqe SIT database (it is mentioned in Giffaut et al.[8] that weak complexes are taken away) and I found both epsilons for Cl, Eu, and NO3 as well as stability constants for EuNO3 and $EuCl^{2+}$ complexes. In this case the epsilon becomes erroneous, because it is determined in concentrated $EuCl_3$ or $EuNO_3$ solutions neglecting the complexation and assuming e.g. fully dissociated $EuCl_3$.

6. **Should the uncertainties provided in the database be used for benchmarking?**

   I have tried for more than 20 years to find out how this can be done for NEA-TDB, but to my knowledge there is as yet no geochemical code that can make use of the uncertainties in the constants. From contacts with mathematician groups I understood that the problem is a complicated one. To use Monte Carlo methods would be a way forward, if the uncertainties would be independent from each other-unhappily this is not the case for e.g. carbonate, bicarbonate and pH. I would be delighted if you find a way to solve this problem.

7. **In your opinion, what are the essential points of focus that need to be covered for the benchmarking exercise to usefully evaluate thermodynamic databases?**

   I would check other rather important issues concerning a high-quality database together or before starting with benchmarking. As I have understood, the core of the data for

most radionuclides in ThermoChimie is NEA-TDB [8]. In the case of these data, the chair of the review team has the responsibility to check the internal consistency of the data. This is not trivial in the case of the procedure used in NEA-TDB, because each constant is chosen individually based on scrutinizing the published literature. In the case of NBS-tables all accepted experimental data were fed into a computer program which selected the values assuring the closure of all thermodynamic cycles possible and the minimisation of the Gibbs energy. In this way it selected the individual values to be included in the database. In this case consistency is assured from the procedure of selection; on the other hand, the coupling between the values selected and those fed to the program from the beginning is not so strong. It was enough to include a single erroneous value for heat of solution of P_2O_5 in the phosphate cycle [9] to get large discrepancies for all phosphate species.

ThermoChimie has been extended through the inclusion of several constants from e.g. Smith & Martell or other database sources especially for metals or ligands not included in NEA-TDB. It is imperative to carry out a consistency check, which would reveal also potential typos or otherwise erroneous values.

Other issues to be checked concern the completeness of the database. Thus, for example sulphides and phosphates have many very insoluble solids, making it difficult to investigate the formation of their complexes. There is an excellent example by Thoenen [10] that demonstrates how the lack of values for nickel sulphide complexes results in very low solubilities for the solid nickel sulphide. These issues are not covered in NEA-TDB, i.e. there is no requirement for completeness, and just a review of published values is carried out. Even a simple comparison of the values of the constants in ThermoChimie with other databases may reveal issues that need to be checked, that this may potentially improve the database.

References.

1. D. Read, T. W. Broyd, Chemval project. Report on stage 1: verification of speciation models EUR Report 12237 EN, Brussels, 1989. D. Read, ed. Chemval project. Report on stage 2: Application of speciation models to laboratory and field datasets, EUR Report 13124 EN, Brussels 1990.

2. M.R. Chandratillake, G. W. Newton, V.J. Robinson, Comparison of thermodynamic databases used in geochemical modelling, Topical report EUR 11891 EN, Brussels 1988.

3. T. Wollery, Discrepancies in key thermodynamic data for the fundamental aqueous spaecies, White paper, 2005.

4. B. Merkel, Thermodynamic data dilemma, In The new uranium mining boom, Springer, Berlin, Heidelberg 2012, pp.627-633.

5. Emren, R. Arthur, P.D. Glynn, J. McMurry, The modeller´s influence on calculated solubilities for performance assessment at the Äspö hard-rock laboratory, MRS Symp. Proc. 556(1999) p. 559-566.

6. G. Sheng, M.S. Elzas, T.I. Ören, B.T. Cronhjort, Model validation: a systemic and systematic approach, Reliability Engineering and System Safety 42(1993) pp. 247-259.

7. D.K. Nordstrom, Models, validation and applied geochemistry: Issues in science, communication and philosophy, Appl. Geochem. 27 (2012) pp. 1899-1919.

8. E. Giffaut, M. Grieve, Ph. Blanc, Ph. Vieillard, E. Colas, H. Gailhanou, S. Gaboreau, N. Marty, B. Made, L. Duro, Andra thermodynamic database for performance assessment: Themochimie, Appl. Geochem. 49 (2014) pp. 225-236.

9. J. A. Rard,,T.J. Wolery,, The standard chemical-thermodynamic properties of phosphorus and some of its key compounds and aqueous species: An evaluation of differences between the previous recommendations of NBS/NIST and CODATA. J. Soln. Chem., 36, (2007), 1585–1599

10. T. Thoenen, Pitfalls in the use of solubility limits for radioactive waste disposal: the case of nickel in sulphidic groundwaters, Nucl. Techn. 126 (1999) pp. 75-86.

2.6 Don Reed

The process of validating models through benchmarking activities is a critical and needed step in the development of a credible and defensible safety case in a repository concept. This accomplishes two important objectives: First, it confirms that the models can correctly predict the outcome of results when many things are well understood and defined (e.g., the results for simplified well controlled studies) and second, it challenges the understanding and applicability of the model to predict the geochemical conditions and source term description in the real-system application of a repository concept. These altogether will build public confidence and regulator acceptance of the repository safety case.

The Waste Isolation Pilot Plant (WIPP), which has been an active TRU waste repository since its licensing in 1998, has utilized various forms of benchmarking in its model and database development and application throughout its operational history. These have had a limited amount of success and could be done more systematically. These studies, as discussed in more detail below, have led to the discovery of missed dependencies and provide a measure of conservatism in the modelling approaches and databases being utilized.

1. **What are the possible uses of a benchmarking exercise making use of thermodynamic databases?**

Benchmarking exercises of a model/database have primarily two important uses:

a. To confirm the correctness of the choices, dependencies and selected data in the database. This can be done by model to model predictive comparisons of the same well-defined system. They are best done in conjunction with a well-defined experiment where there are both model-model comparisons and model-experiment comparisons.

In the WIPP project we have been working with the THEREDA model/database to make predictive comparisons with the WIPP FMT model. These models, although both based on an application of the Pitzer approach, were developed completely independently and at different times meaning they are based on different experimental results. These comparisons have shown that there are fairly good agreements with most of the brine components (e.g., oceanic salt components) with the exceptions of Mg and Ca. Additionally, they showed some significant discrepancies in the dependencies and prediction of key actinide solubilities. These exercises confirm what is working well, identify data gaps or areas where more studies are needed, and identify potential errors in implementation.

b. To make sure that all key dependencies and critical relationships are identified.

Complex brine experiments, designed to challenge the WIPP model, led to an understanding of the key roles played by the lesser components of the brine (borate, sulfate and bromide). Borate complexation of the trivalent actinides, which was missing in our models, was discovered. Bromide was found to have a key role in the radiolysis of brine systems in that hypobromite was preferentially formed over hypochlorite. Sulfate phases were identified as solubility-controlling at the lower pHs investigated.

2. **What are the most relevant types of modeling for ThermoChimie benchmarking?**

   ThermoChimie should model two types of systems. First well-defined (simplified) systems that are also likely to be at/near equilibrium. These model results can be also compared to those obtained by similar databases. Second, the model should be challenged by real-system applications where complete equilibrium is not expected or predicted. This latter exercise will be the likely case for moving forward in defending a safety case for a specific repository application.

3. **Which geochemical systems should be considered for model strengths and limitations?**

   The limitations of the model will be best evaluated by modelling well-controlled complex (e.g., simulated groundwater) experiments. This will establish if the proper relationships are fully accounted for in the model. In this sense, it will identify possible gaps in the database. Examples of this approach are the modelling of simulated groundwater experiments and real-system application to natural analogues or site-specific (for the repository) conditions/groundwaters.

   The strengths of the model will likely be established in simplified system studies that are expected to be well behaved.

4. **Should the benchmarking exercises be based on modelling results only or should the benchmarking refer to a selected set of well-defined experimental and/or natural (e.g., groundwater composition…) datasets as in the case of validation exercises?**

   My personal opinion is that you need to do both.

   There should be a systematic comparison with other models for known well defined systems that are fully addressed by both models – this will build credibility for the approach and data content of the model.

   Additionally (see answer to question 3), the most important challenge to the database is its ability to model the natural or well-defined complex systems as this would test the completeness and adequacy of the proposed database and models. This issue of completeness, or in the absence of this, the conservatism of the modelling will be a very critical factor into the regulatory process for repository licensing.

5. **Which other high-quality databases should be considered for the benchmarking exercise? Which codes should be used for the benchmarking exercise?**

   This question is away from my area of expertise. As to the high ionic-strength applications, THEREDA is a well-documented model that is “complete” in some areas – this could be a good check on this aspect of the ThermoChimie database and model.

6. **Should the uncertainties provided in the database be used for benchmarking?**

   It is important to find some way to address uncertainties. A benefit of doing this within ThermoChimie would be that it would help identify weakness (hence prioritize fixes to the database) and it would provide a more realistic sense of what levels of uncertatity can be sensibly supported by current data. There are far too many modelling

calculations within the field that provide levels of precision that far exceed what is available experimentally (e.g. six significant figure calculations of pH and solubility). That being said, this will miss what might be the most critical source of uncertainty, which are the data gaps or missing processes/relationships in the model. The database sources of uncertainty are a good and perhaps critical input into the overall uncertainty assessment, but this will not be sufficient and needs to be considered in the broader context of how uncertainties are handled within the safety case.

7. **In your opinion, what are the essential points of focus that need to be covered for the benchmarking exercise to usefully evaluate thermodynamic databases?**

   It is important to show that the database works well for the systems we understand very well. But is also important to show that they account for or bound the key processes that will define the source term in repository applications. The latter issue is the most important as it is key to the defensibility of the safety case.

3 Workshop Summary

This section summarises the main points from the presentations and discussions at the ThermoChimie workshop. It is organised around themes, which loosely address the questions outlined in the introduction to this report. Except where a single contributor made a specific point, the experts' opinions have been summarised together.

3.1 Aims and possible uses of benchmarking exercises

The most commonly suggested aim of the ThermoChimie benchmarking exercise was that it should assess the strengths and weaknesses of the database. The exercise should test the ability of ThermoChimie to accurately reproduce experimental results and compare its results to those given by other databases.

In addition to this overall aim, a number of possible additional uses for the benchmarking exercise were suggested:

- Identify gaps in the database and spot any errors.
- Test how the database performs with different geochemical codes.
- Raise the profile of ThermoChimie with the wider scientific community (though publication of the benchmarking results).

Kastriot Spahiu suggested that to give it maximum credibility the ThermoChimie benchmarking exercise should follow similar approaches to those taken by other internationally recognised databases.

Don Reed stressed that the aim of any geochemical modelling should be to underpin the post-closure safety case. Therefore, any arguments made from geochemical models should be well supported and defensible. Benchmarking exercises can help in this aim by building confidence in the ThermoChimie database and the geochemical models in which it is used.

3.2 Potential benchmarking models – equilibrium, kinetic and reactive transport

The experts agreed that given the nature of the ThermoChimie database, which contains thermodynamic data, any benchmarking exercises should focus primarily on equilibrium reactions, rather than kinetics. A simple 1D reactive transport model could provide a useful benchmark, however some of the experts advised against including a transport model as it would introduce excessive complexity and uncertainty that may be difficult to interpret (e.g. are errors due to values in the database or other parameters, such as diffusivity).

Some specific models that could be run for the benchmarking were:

- Solubility calculations for radionuclides and chemotoxics (including speciation when experimental data are available for discrimination). U was suggested as a good candidate species as there is a lot of data available, but its chemistry is complex.

- Sorption (not sorption model development, but e.g. sorption reduction factors based on speciation).
- Redox reactions.
- Organic complexation.
- Stability and evolution of engineered barrier systems (clay and cement based), including steep gradients such as at the interface between two materials/components.
- A high ionic strength model comparing the SIT and Pitzer approaches.

It was suggested that any benchmarking models should follow the "KISS" principle (keep it simple stupid), beginning with simple systems with two components at equilibrium. Complexity could then be added step-by-step to model more 'realistic' systems. Examples of increased complexity included, modelling at high temperature, at higher ionic strength, and with organics present. It was also suggested that redox processes and solid-solutions could be included.

Don Reed demonstrated with examples from the WIPP that modelling complex systems can be a good test of how well a model performs, and may highlight gaps in the database (for example missing phases). When evaluating these models of real systems, it is important to make sure any disagreement between model and observation is coming from the model rather than from a lack of understanding of the system.

3.3 Geochemical systems

A large number of geochemical systems were suggested for use as benchmark examples. It was generally agreed that the focus should be on conditions that were of most relevance to the safety case for a geological disposal facility, including in clay rocks, crystalline rocks and cementitious systems. A suggestion was made, however, that it might be interesting to benchmark the database in the extremes of these systems, for example at high pH, ionic strength or temperature. ThermoChimie is also used for work on heavy metals and contaminated land, but these are beyond the designed remit of the database so of less importance for the benchmarking exercises.

As part of the benchmarking is likely to involve comparison between different databases, it is important to ensure that all of the databases are capable of modelling the chosen systems. This is not always the case as the species present in the databases and the conditions over which they are valid (T, P, etc.) vary. So, care should be taken when selecting the test systems. However, it is important not to cherry-pick systems where ThermoChimie performs better than other databases. In fact, some experts suggested inverse cherry-picking of systems where ThermoChimie is known to be less complete than other databases.

3.4 Suitable modelling and experimental studies

It was repeatedly highlighted that it is important to benchmark against high-quality experimental data, rather than simply comparing modelling results. However, when experimental data is used, this should be reviewed and selected by experts as there is a

lot of bad data out there! However, high-quality data that has already been used to build the database should be avoided as the use of these data sources might lead to self-reinforcing results that undermine the usefulness of the benchmarking exercise.

Ideally, experimental data used for the benchmarking exercise should come from well constrained systems. In practice this means using laboratory experiments, where all the variables can be controlled, rather than observations from natural systems where there are a lot of unknowns.

3.5 Other databases and codes

It was generally agreed that a good benchmarking exercise would be to model the same geochemical systems using both ThermoChimie and a range of other high-quality thermodynamic databases. The results of this modelling could then reveal gaps in ThermoChimie that other databases fill. A number of different geochemical databases were suggested, but most experts recommended those that were internally consistent, including:

- The NAGRA-PSI database.
- The JAEA database, as a good non-European option. This database would be particularly suitable for modelling the actinides.
- The THEREDA database for calculations at higher ionic strength. This is a Pitzer database so would allow comparison of results using this approach against the SIT approach used in ThermoChimie.
- The CEMDATA database for modelling cement phases, many of which are currently absent from ThermoChimie. It would be interesting to see if the additional phases make any significant difference to the modelling outputs.

The LLNL and Yucca Mountain databases should be avoided as it is unclear how they are updated and may be out of date. The NEA-TDB database and Thermoddem share many values with ThermoChimie so are less useful comparators. Since RWM has now joined the ThermoChimie consortium, their old Hatches database is no longer being actively updated or supported.

Running speciation calculations with different databases can be problematic as different databases may include different phases and/or different aqueous species (especially where different activity models are contained in the databases under consideration).

The experts recommended that PHREEQC should be used as the main code for the benchmarking exercises as it is freely available and has a robust feature set. Any comparative codes should include similar functionality. Frank Bok also highlighted that when running models, it was best to use copies of databases that come with the code as they have been verified as functional. There can sometimes be issues in syntax when importing a new database to use with a code.

Tim Heath suggested that PFLOTRAN could be used for the benchmarking if a reactive transport model was used. However, ThermoChimie is not currently available for this code, so it would first need to be extracted into a compatible format.

3.6 Uncertainties

The experts agreed that it would be good to try and incorporate uncertainties from the database into the benchmarking models. However, the exact approach to do this was not agreed. This may also be a difficult task as many of the common geochemical codes that would be used to run the benchmark models do not have any way to calculate uncertainties. Indeed, Laurent De Windt made an appeal that uncertainties should not be included in the database files, formatted for use in the codes, as this would just make them more difficult to read, without any real benefit. Instead details of uncertainties should just be provided through the website and supporting documents.

Tim Heath presented the most detailed discussion of uncertainties and suggested a probabilistic approach, producing PDFs for the values. This involved running repeat models while varying the input values, using Monte Carlo analysis to find the uncertainty range. Frank Bok suggested a similar approach.

It was generally agreed that that any benchmarking of uncertainty should only be undertaken for simple systems. Attempts to including uncertainty analysis in complex systems (such as solid solutions or transport models) could introduced too much complexity and lead to unnecessary inflation of errors due to unaccounted for parameter correlation or double counting (e.g. inclusion of error for the same element as both a solid and aqueous species).

Don Reed stressed that if uncertainty analysis is not carefully conducted, the resulting values can be unrealistically, and unhelpfully large. Therefore, it is important to ensure that where uncertainties are included, they are developed from reliable data sources, following a robust methodology (such as Monte Carlo analysis).

3.7 Points of focus for the benchmarking exercises

In addition to addressing the specific questions, the experts made a number of general points regarding the focus of the benchmarking exercises and things to consider when performing the benchmarking work:

- The benchmarking exercise could be automated to limit user error and speed up the process – e.g. generate a python script to run all the tests.
- The benchmarking process should be well documented, and the results should be published. Publishing in a peer reviewed journal could both raise the profile of the database and increase the credibility of the benchmarking work.
- Phase diagrams could be used as a good way to show the results, particularly when comparing between databases.
- Different modellers can model the same system and get different results depending on the approach taken. This can either be accounted for and used as a benchmarking test, giving the modellers freedom of approach, or the model could be quite prescriptive in its approach to remove this variable.

- Models need to be defensible and explainable, even if they are not truly accurate descriptions of the system. If performed correctly, benchmarking exercises can build confidence in the models used to support the safety case.

Appendix A – Presentation Slides

A.1 Steering Committee Introduction

ThermoChimie

"BENCHMARK" WORKSHOP

1. R. Hibberd – RWM
2. B. Madé – Andra
3. S. Brassinnes – O/N

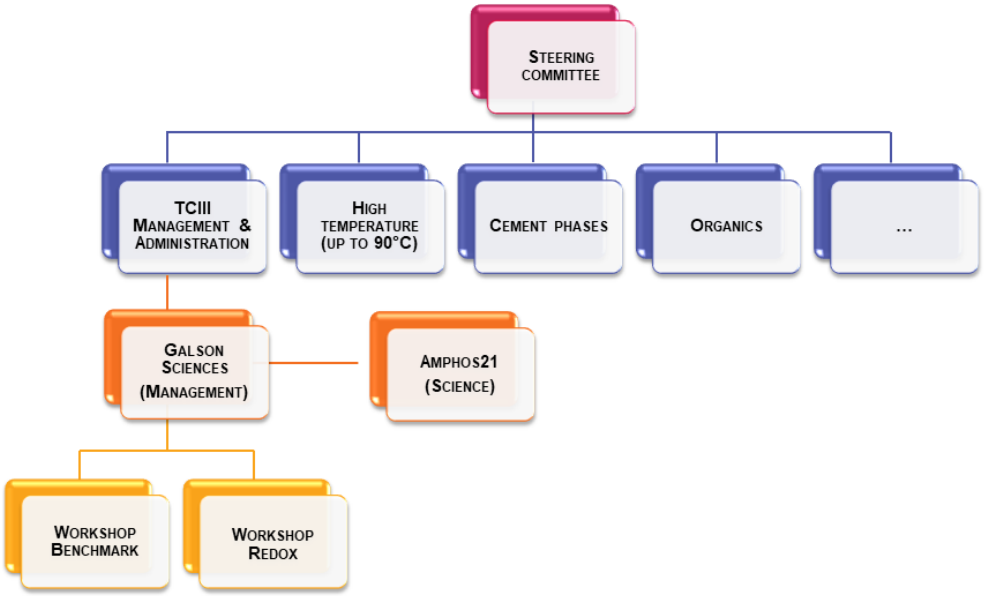
ANDRA ONDRAF/NIRAS Radioactive Waste Management *TC XXX workshop, Manchester, 15/16th October 2019*

HISTORY

- TC, since 1995, project inception at Andra
  - ❖ Data sets review/acquisition (1995-1999, *A21*)
  - ❖ Clay phases and radionuclides (since 2000, *A21-BRGM*)
  - ❖ Cement phases and chemotoxic elements (since 2010, *A21-BRGM)*
  - ❖ Organic molecules and complexes (Rn, metals, cations…) (since 2012, *A21*)
  - ➢ version 8 TDB (2014)
- TCII, August 2014-March 2018, the TC consortium with RWM was established
  - ❖ Organics, Temperature (up 90°C), Saline environment (SIT) topics (since 2015, *A21-BRGM-Galson*)
  - ➢ version 9 TDB (10/2015)
- TCIII, April 2018 – April 2023, O/N joins the TC consortium
  - ➢ version 10A TDB (06/2019)

ThermoChimie *TC XXX workshop, Manchester, 15/16th October 2019*

TCIII ORGANIZATIONAL ASPECTS



Workshops goal

- Gather expert advice on
  - ❖ Thermodynamic Database (TDB) Benchmarking
  - ❖ Redox (& *kinetics, microbiology effect*) in geochemical modelling
- To help us with the identification of
  - ❖ Do & Don't do
  - ❖ Points of attention to be carefull about

→ So that, we can establish the related TC3 further activities

! To be communicated to user community @ TC special day next to Clay Conference 2020 (Nancy).

Uses of thermodynamic databases in radioactive waste management

- To test scientific understanding of processes and the results from experimental work
- To bound the chemical evolution of the various disposal system components
- To perform scoping calculations supporting Performance Assessment assumptions (such as solubility limits)

Objectives of Benchmarking

- To evaluate the performance of TC database in comparison with other international high-quality databases – identify areas for improvement
  - ❑ Good performance is the ability to accurately and completely model all of the processes in a geochemical system
- To test the database on a set of geochemical conditions relevant to cement-based deep geological disposal facility in a clay and crystalline rock
- To determine the effect of temperature and ionic strength on speciation and solubility of elements

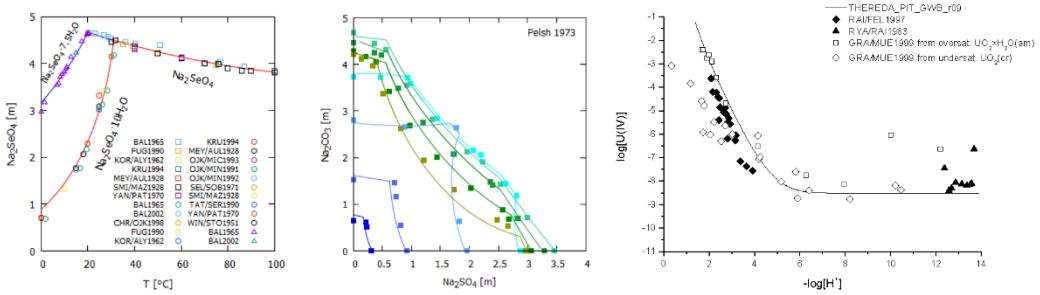
Questions for Experts

- Uses of a benchmarking exercise using thermodynamic databases?
- Most relevant types of modelling? (just equilibrium, near equilibrium, reactive transport?)
- Which geochemical systems should be considered? Which are most likely to strengths and limitations?
- Based only on modelling results? Or should we use a well-defined experimental dataset? (E.g. validation exercises.)
- Which databases and codes should be used for comparison?
- Should uncertainties be included?
- What are the essential points for a benchmarking exercise in order to be useful?

Attendees and Roles

- **Benoît Madé** (Andra, France) – ThermoChimie Steering Committee
- **Stéphane Brassinnes** (ONDRAF/NIRAS, Belgium) – ThermoChimie Steering Committee
- **Rosie Hibberd** (RWM, UK) – ThermoChimie Steering Committee
- **Lara Duro** (Amphos21, Spain) – Facilitator
- **Eli Colas** (Amphos21, Spain) – ThermoChimie Development Expert
- **Adam Fuller** (Galson Sciences Ltd.) – Project Lead / Secretariat
- **Frank Bok** (HZDR, Germany) – Independent Expert
- **Sonia Salah** (SCK-CEN, Belgium) – Independent Expert
- **Tim Heath** (Wood, UK) – Independent Expert
- **Laurent De Windt** (Mines ParisTech, France) – Independent Expert
- **Kastroit Spahiu** (SKB, Sweden) – Independent Expert
- **Don Reed** (LANL, USA) – Independent Expert
- **Benoit Cochepin** (Andra, France) - Expert on PA modelling

A.2 Frank Bok



ThermoChimie Benchmarking Workshop

Frank Bok, Helge C. Moog

Manchester, October 15th 2019

DR. FRANK BOK

HZDR
HELMHOLTZ ZENTRUM DRESDEN ROSSENDORF

Research associate at the Helmholtz-Zentrum Dresden-Rossendorf (Institute for Resource Ecology)

- Database Development:
  - **THEREDA** (Thermodynamic Reference Database) (https://www.thereda.de)
  - **RES³T** (Rossendorf Expert System for Surface and Sorption Thermodynamics) (https://www.hzdr.de/res3t)
- Thermodynamic Modelling:
  - Geochemist's Workbench, PHREEQC, EQ3/6, GEMS
- Modelling lectures at the Technical University Bergakademie Freiberg and the Friedrich-Schiller-University Jena

Uses of a benchmarking exercise

1. What are the possible uses of a benchmarking exercise making use of thermodynamic databases?

- Data quality (compared to experimental data)
- Identification of data gaps / bad data
- Identification of typos, errors
- Database comparison
- Consequences of data updates, removals and additions (especially unplanned side effects)
- Correctness of internal calculation routines / consistency
- Correctness of data export to geochemical codes formats
- Behaviour using different geochemical codes
- Numerical behaviour using the data in complex systems:
  - Relevant species ⟶ numerical stability (highly charged species)
  - Treatment of redox species / redox couple chains
- Publication, Promotion...

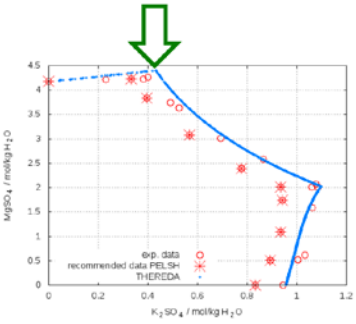
Dr. Frank Bok | f.bok.hzdr.de

Uses of a benchmarking exercise

1. What are the possible uses of a benchmarking exercise making use of thermodynamic databases?

1. **Manual test calculations:**
   - Manually by members of the THEREDA team
   - Repeated after data updates / upgrades / new version of geochemical codes / new experimental data
   - All supported codes (if possible): PHREEQC, GWB, ChemApp, ~~EQ3/6~~, *GEMS, ToughReact*)
   - Control of data input, internal recalculations, data export
   - Single point calculations of defined systems

Chemical scenario	Invariant point: Hexahydrite & Leonite @ T = 323.15 K				
Code	K(tot)	Mg(tot)	S(tot)	$a(H_2O)$	pH
	mol/kg_{H_2O}			-	-
PHREEQC	0.869	4.412	4.847	0.81730	5.708
Geochemist's Workbench	0.859	4.376	4.805	0.81992	5.713
ChemApp	0.868	4.411	4.845	0.81734	5.705



   - Fully documented (incl. input syntax and files)
   - Results and code comparison

Dr. Frank Bok | f.bok.hzdr.de

Uses of a benchmarking exercise

1. What are the possible uses of a benchmarking exercise making use of thermodynamic databases?

2. **Automated test calculations:**
   - Python-trigger set of calculations (at present 201 test cases)
   - Results automatically compared to prior results and (before data releases) stored in DB
   - Will become part of our quality assurance / data release scheme
   - Check data input, internal recalculations, data export, ...

   - Single point calculations of defined systems (Solubility, water activity, speciation, ...)
   - In principle also capable of comparison with other TDBs

Dr. Frank Bok | f.bok.hzdr.de

Uses of a benchmarking exercise

1. What are the possible uses of a benchmarking exercise making use of thermodynamic databases?

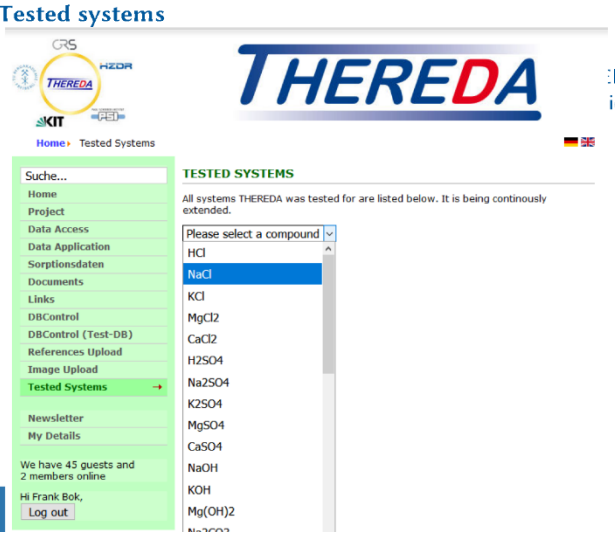
3. **Tested systems**
   - Graphical output for users via website (in preparation!)
   - We show possibilities (and sometimes discrepancies) of THEREDA,
   - We cannot/do not want to make any statement about the applicability to systems that do not appear in the list.

Dr. Frank Bok | f.bok.hzdr.de

Uses of a benchmarking exercise

1. What are the possible uses of a benchmarking exercise making use of thermodynamic databases?

3. Tested systems

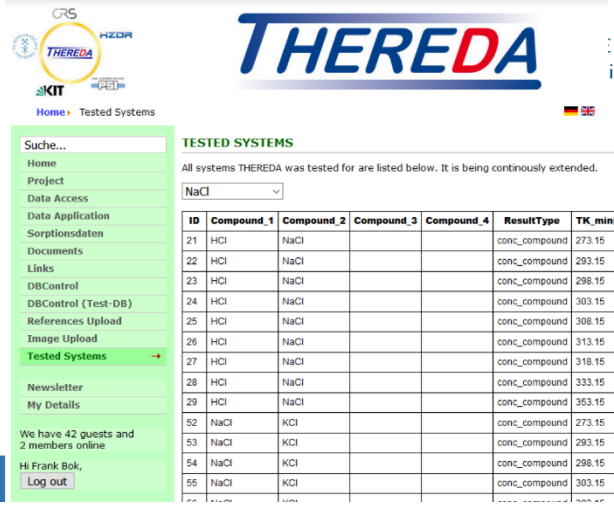


:DA,
icability

Uses of a benchmarking exercise

1. What are the possible uses of a benchmarking exercise making use of thermodynamic databases?

3. Tested systems



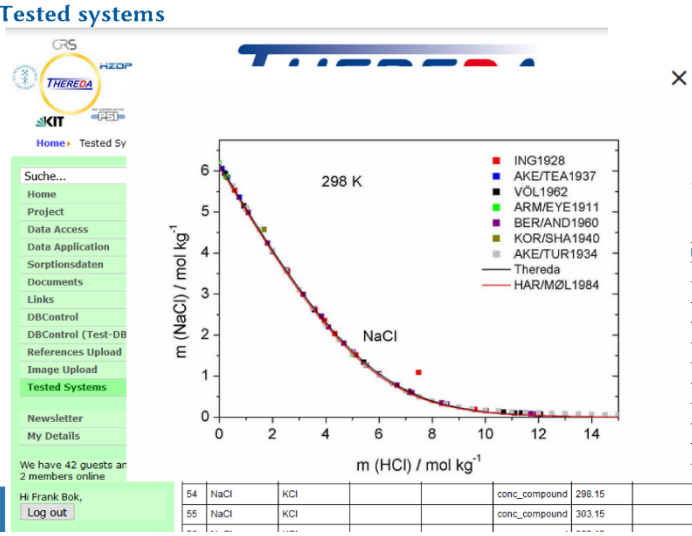
:DA,
icability

ID	Compound_1	Compound_2	Compound_3	Compound_4	ResultType	TK_min[K]	TK_max[K]	pcH_min	pcH_max	Diagram	description
21	HCl	NaCl			conc_compound	273.15				0/sol	
22	HCl	NaCl			conc_compound	293.15				20/sol	
23	HCl	NaCl			conc_compound	298.15				25/sol	
24	HCl	NaCl			conc_compound	303.15				30/sol	
25	HCl	NaCl			conc_compound	308.15				35/sol	
26	HCl	NaCl			conc_compound	313.15				40/sol	
27	HCl	NaCl			conc_compound	318.15				45/sol	
28	HCl	NaCl			conc_compound	333.15				60/sol	
29	HCl	NaCl			conc_compound	353.15				80/sol	
52	NaCl	KCl			conc_compound	273.15				0/sol	
53	NaCl	KCl			conc_compound	293.15				20/sol	
54	NaCl	KCl			conc_compound	298.15				25/sol	
55	NaCl	KCl			conc_compound	303.15				30/sol	

Uses of a benchmarking exercise

1. What are the possible uses of a benchmarking exercise making use of thermodynamic databases?

3. Tested systems



pcH_min	pcH_max	Diagram	description
		0/sol	
		20/sol	
		25/sol	
		30/sol	
		35/sol	
		40/sol	
		45/sol	
		60/sol	
		80/sol	
		0/sol	
		20/sol	
		25/sol	
		30/sol	

54	NaCl	KCl			conc_compound	298.15	
55	NaCl	KCl			conc_compound	303.15	

Page 9

Uses of a benchmarking exercise

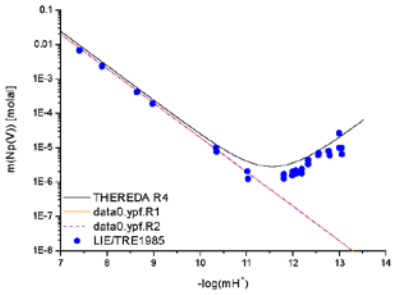
1. What are the possible uses of a benchmarking exercise making use of thermodynamic databases?

4. Against other (PITZER-) databases

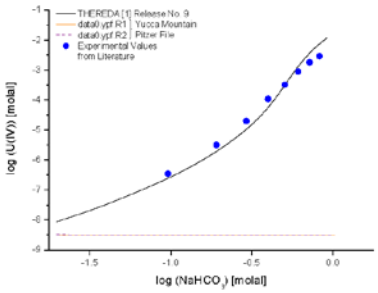
→ Graphical output

→ Presented at conference talks or posters

Solubility of $NpO_2(OH)(am)$ in 1 m NaCl solution



Solubility of $U(OH)_4(am)$ in $NaHCO_3$ solution



Be careful when interpreting the results!

Page 10

Dr. Frank Bok | f.bok.hzdr.de

Relevant types of modelling for benchmark

2. What are the most relevant types of modelling for the ThermoChimie benchmarking exercise?

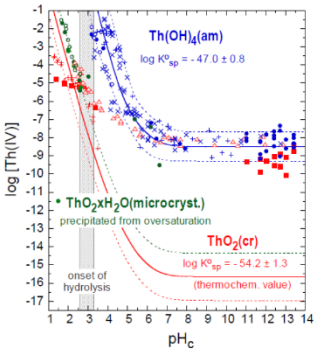
- Prerequisite for any benchmark:
  - Experimental data for combinations of variables and results
- Results:
  - Solubility / saturation indices
  - Osmotic coefficient / water activity
  - Vapor pressure
  - Activity coefficients
  - ~~Speciation~~
- Variables:
  - pH
  - Eh
  - $p(CO_2)$
  - T
  - IS
  - $m(NaCl)$, $m(CaCl_2)$

Dr. Frank Bok | f.bok.hzdr.de

Relevant types of modelling for benchmark

2. What are the most relevant types of modelling for the ThermoChimie benchmarking exercise?

- Equilibrium examples are to be preferred!
- Non-equilibrium examples involve kinetics (not in TDBs)
  → Results may not be driven by TDB's!
- Pseudo-equilibria of amorphous phases might be ok...
- Reactive transport experiments are way too complex and involve:
  - Unknown or limited known parameters (ground water flow, porosity, heterogeneities, ...)
  - Processes not covered by TDB (Sorption, Ion exchange, Colloids, Microbial activity, ...)

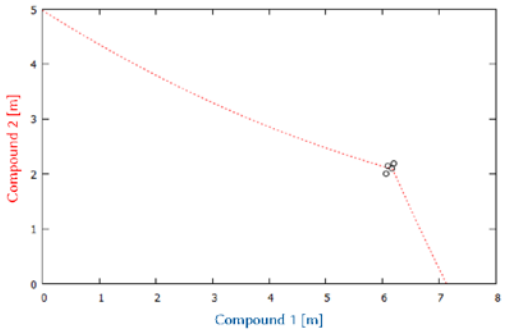


Dr. Frank Bok | f.bok.hzdr.de

Relevant types of modelling for benchmark

2. What are the most relevant types of modelling for the ThermoChimie benchmarking exercise?

- KISS principle (**K**eep **I**t **S**mall & **S**imple)
- Simple scenarios, *e.g.* invariant points / phase transitions:
  - Chemically well-defined
  - Lots of data in literature
  - Easy to model
  - Give access to a lot of useful information (for such a small scenario)

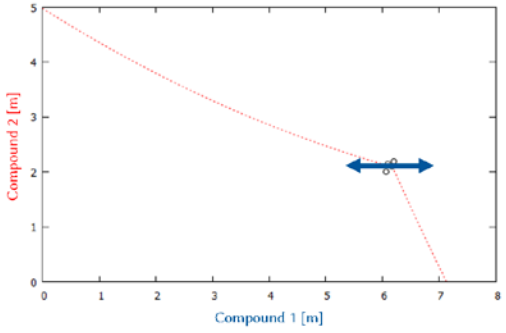


Dr. Frank Bok | f.bok.hzdr.de

Relevant types of modelling for benchmark

2. What are the most relevant types of modelling for the ThermoChimie benchmarking exercise?

- KISS principle (**K**eep **I**t **S**mall & **S**imple)
- Simple scenarios, *e.g.* invariant points / phase transitions:
  - Chemically well-defined
  - Lots of data in literature
  - Easy to model
  - Give access to a lot of useful information (for such a small scenario)

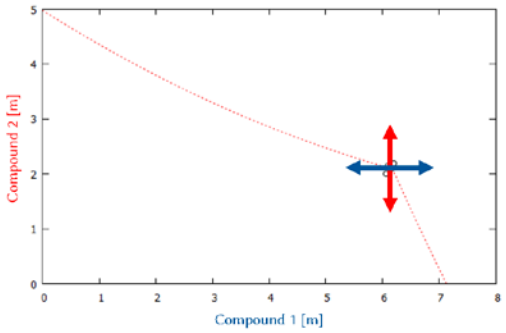


Dr. Frank Bok | f.bok.hzdr.de

Relevant types of modelling for benchmark

2. What are the most relevant types of modelling for the ThermoChimie benchmarking exercise?

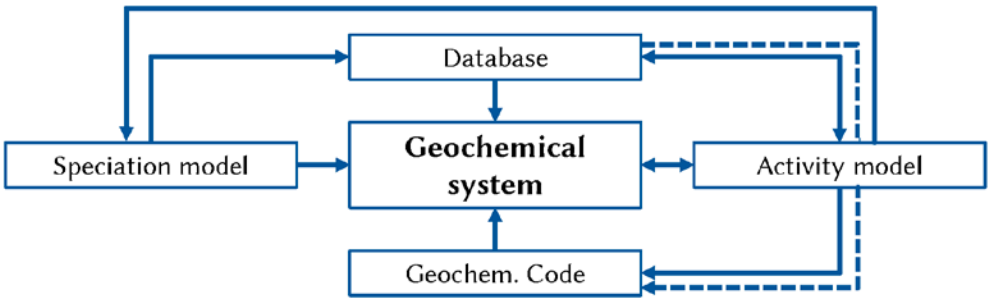
- KISS principle (**K**eep **I**t **S**mall & **S**imple)
- Simple scenarios, *e.g.* invariant points / phase transitions:
  - Chemically well-defined
  - Lots of data in literature
  - Easy to model
  - Give access to a lot of useful information (for such a small scenario)



Dr. Frank Bok | f.bok.hzdr.de

Which geochemical systems?

3. Which geochemical systems should be considered in this benchmarking exercise?
(geochemical systems most likely to reveal the strengths and limitations of the ThermoChimie database)



- Redox behaviour
- Temperature functions / extrapolations

Dr. Frank Bok | f.bok.hzdr.de

Which geochemical systems?

3. Which geochemical systems should be considered in this benchmarking exercise?
(geochemical systems most likely to reveal the strengths and limitations of the ThermoChimie database)

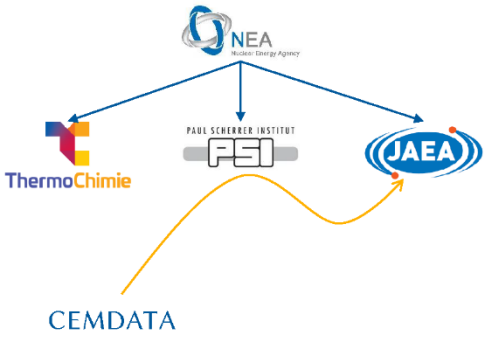
- Choose elemental setup that you consider to be critical in your safety-case.
- Example scenarios for ThermoChimie:
  - Solubility of Radionuclides (*e.g.* medium half-life nuclides behaviour)
  - Stability of cementitious materials / clay phases (or other host rock mineral phases)
  - Stability of containment material (copper / steel)
- High ionic strength example:
  - Some clay rock pore waters have ionic strength ~3 molal
  - Good to reveal strengths and limitations
  - Comparison of (ThermoChimie's) SIT vs. PITZER model

Dr. Frank Bok | f.bok.hzdr.de

Which geochemical systems?

3. Which geochemical systems should be considered in this benchmarking exercise?
(geochemical systems most likely to reveal the strengths and limitations of the ThermoChimie database)

- Data overlap in different TDB's! Benchmarks will have similar results there.



Dr. Frank Bok | f.bok.hzdr.de

Which geochemical systems?

4. Should the benchmarking exercises be based on modelling results only?

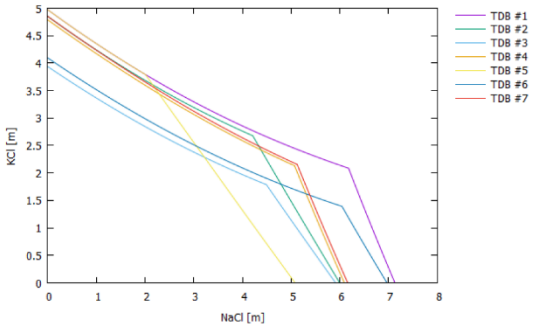
No!

Dr. Frank Bok | f.bok.hzdr.de

Which geochemical systems?

4. Should the benchmarking exercises be based on modelling results only?

- Which TDB would you judge to be correct?

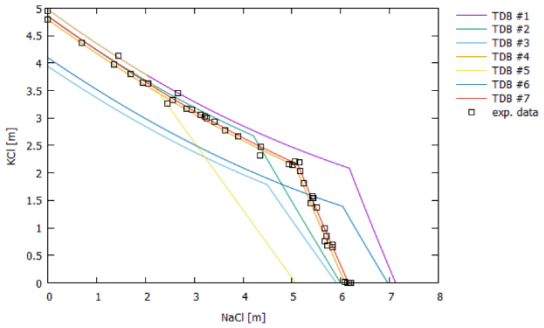


Dr. Frank Bok | f.bok.hzdr.de

Which geochemical systems?

4. Should the benchmarking exercises be based on modelling results only?

- Which TDB would you judge to be correct?



Which geochemical systems?

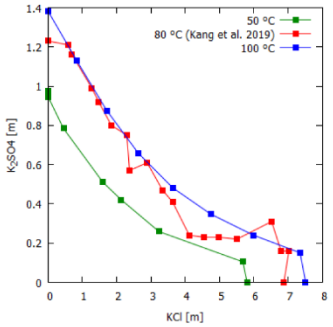
4. Should the benchmarking refer to a selected set of **well-defined experimental** and/or **natural (*e.g.*, groundwater composition...) datasets** as in the case of validation exercises?

Yes!

Which geochemical systems?

4. Should the benchmarking refer to a selected set of **well-defined experimental** and/or **natural (*e.g.*, groundwater composition…) datasets** as in the case of validation exercises?

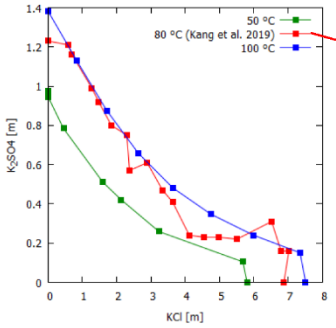
- Quality assured experimental data (peer reviewed) and chosen by an expert!
- It's a mess out there!



Which geochemical systems?

4. Should the benchmarking refer to a selected set of **well-defined experimental** and/or **natural (*e.g.*, groundwater composition…) datasets** as in the case of validation exercises?

- Quality assured experimental data (peer reviewed) and chosen by an expert!
- It's a mess out there!



Journal of chemical & engineering data

Benchmark against other databases

5. Which other high-quality databases should be considered for the benchmarking exercise?

- THEREDA
  (https://www.thereda.de)
- OECD/NEA Thermochemical Database
  (https://www.oecd-nea.org/dbtdb/tdbdata/)
- PSI/Nagra Chemical Thermodynamic Database
  (https://www.psi.ch/en/les/database)
- JAEA Thermodynamic DataBase
  (https://migrationdb.jaea.go.jp/cgi-bin/db_menu.cgi?title=TDB&ej=1)
- Thermoddem
  (http://thermoddem.brgm.fr/)
- LLNL / LLNL V8 R6+ (Gembochs)
- Pitzer.dat (as delivered with PHREEQC)

Benchmark against other databases

5. Which codes should be used for the benchmarking exercise?

- Codes supported by the different database projects:

Database	PHREEQC	GWB	Spana	Chess	Crunch	Toughreact	GEMS-PSI	ChemApp	EQ3/6
ThermoChimie	✓	✓	✓	✓	✓	✓			
THEREDA	✓	✓				(✓)	(✓)	✓	✓
OECD/NEA TDB	✓								
PSI/Nagra	✓	(✓)					✓		
JAEA	✓	✓							
Thermoddem	✓	✓		✓	✓	✓			
LLNL	✓	✓							✓
Pitzer.dat	✓	✓							

- ➢ Go for natively supported codes to avoid transformation errors.
- ➢ Check data exporter and internal calculation routines.

Benchmark against other databases

5. Which codes should be used for the benchmarking exercise?

- Activity models:

Database	PHREEQC	EDH / B-dot	Davies	SIT	PITZER
ThermoChimie	X	✓	✓	✓	
THEREDA	X				✓
OECD/NEA TDB	X			✓	
PSI/Nagra	X		✓	(✓)[1]	
JAEA	X	✓		✓	
Thermoddem	X	✓			
LLNL (Gembochs)	X	✓			

1) SIT parameters ε not included in tailored file, tables in reports only.

- Chemical scenario that can be modelled with all databases/codes?
- Nomenclature in different TBDs! Watch for automated scripts!
- Different speciation models!

Dr. Frank Bok | f.bok.hzdr.de

Benchmark against other databases

5. Which other high-quality databases should be considered for the benchmarking exercise?

Dr. Frank Bok | f.bok.hzdr.de

Uncertainties

6. Should the uncertainties provided in the database be used for benchmarking?

- Given uncertainty information is often questionable...
  - Confidence interval
  - Standard derivation
  - Variance
  - Experimental / analytical parameter
  - Guestimations
  - Unclear "range" (x ± y)
  - Mixtures of the above
  - Not given
- Very often given too small
- Dependent / independent uncertainty information
- What about chemical analogues?
- Codes (PHREEQC) alone cannot handle uncertainties ⟶ code coupling
- Parameter explosion!

Dr. Frank Bok | f.bok.hzdr.de

Uncertainties

6. Should the uncertainties provided in the database be used for benchmarking?

- Example calculation
  (*e.g.* solubility of schoepite as function of pH at given ionic strength)
  - Vary all logK and all ε randomly within their uncertainty range:
    a) One value changes all other stay, or
    b) All values change at once.
  - Repeat until statistically significant.
  - Compare to experimental data.

- Extremely useful idea but an enormous amount of work if not cut down to a small example!
- Better: refer to experimental uncertainties (with good quality)!

Dr. Frank Bok | f.bok.hzdr.de

What are the essential points of focus?

7. What are the essential points of focus that need to be covered for the benchmarking exercise to usefully evaluate thermodynamic databases?

- Reproduction of high-quality experimental data of well-defined chemical scenarios (that was not used for generating ThermoChimie's data)
- Large number of different simple test calculations
- Comparison with other databases
- Identify reasons for the differences
- Quality assurance (automation, documentation & audit)

Dr. Frank Bok | f.bok.hzdr.de

What are the essential points of focus?

7. What are the essential points of focus that need to be covered for the benchmarking exercise to usefully evaluate thermodynamic databases?

- Does ThermoChimie's SIT correctly reflect results at elevated ionic strength?
  - ➢ Comparison with PITZER database (then go for solubility experiments only)
  - ➢ Possibilities to close data gaps mutually

Dr. Frank Bok | f.bok.hzdr.de

Thank you!

A.3 Sonia Salah

SCK•CEN

STUDIECENTRUM VOOR KERNENERGIE
CENTRE D'ETUDE DE L'ENERGIE NUCLEAIRE

ThermoChimie Benchmarking Workshop

Sonia Salah

Manchester, 15/10/2019

1

Q1 **What are the possible uses of a benchmarking exercise making use of thermodynamic databases?**

4 different types/uses of benchmarking exercises

Thermodynamic — Does the TDB fit the purpose ?
- Presence of most relevant solid phases, aq. species and gases included
- Implementation of different activity models
- Availability of different TDB formats
- Traceability of data source & selection criteria
- Identification of data gaps/weaknesses

Conceptual — Is the real system represented adequately ?
- Have the key physical and geochemical processes been identified ?
- Has the conceptual model been simplified as much as possible, but retains enough complexity to adequately reproduce the system behavior ?

Qualitative model

Mathematical — Are the key processes adequately described ?
- Have the governing equations & parameters (e.g. Fick's law, Darcy's law, etc.) been chosen well ?
- Chem. Equilibrium model: LMA versus GEM
- Have boundary & initial conditions been set properly ?

Quantitative model

Numerical — Is the mathematical model correctly solved ?
- Test computer codes for accuracy, robustness stability & efficiency
- Iterations & algorithms → convergence
- Discretization schemes (e.g. FEM, FVM, FDM)
- Proper coupeling of different processes, (e.g. chem. reactions & mass transport processes) ?

Compare performance against others & reproducibility of results

2

Q1: Continued…. Please illustrate this with a set of examples of previous benchmarking exercises you have been involved in

Benchmarking TDB on speciation/solubility calculations with GWB/MOLDATA – Sensitivity analysis;
L. Wang (2016)
Compare different TDB's: MOLDATA, THEREDA and ThermoChimie v.9b to:
Assess the influence of different chem. conditions on Rn speciation & solubility calculations
Elements: Am, Eu, Mo, Np, Pu, U
→ Rn-speciation as function of changing pH, DIC and ionic strength
→ Pourbaix diagrams: Eh-pH diagrams
→ Solubility diagrams as function of changing pH, DIC and ionic strength

Scientific review of ThermoChimie v.9b on the state-of-the-art of organic thermodynamics;
C. Bruggeman & S. Salah (2015)
→ Which should be the most sensitive organic species with respect to their ligand properties under cementitious and under natural cond. (clay & crystalline ones) ?
→ State of knowledge about stability of org. complexes
→ What are accurate estimation methods to obtain lacking data ?
→ Most relevant organic systems/conditions to be studied experimentally?

Benchmarking the implementation of CEMDATA07 database (Matschei et al., 2007; Lothenbach et al., 2008) using PHREEQC and the GEM-Selektor);
D. Jacques et al. (2012)
→ Three types of chemical reactions were simulated (1/ carbonation, 2/sulphate attack, 3/decalcification/leaching) relevant to the assessment of long-term cement and concrete durability
→ Two approaches were compared to calculate thermodynamic eq. at higher T (LMA versus GEM)

3

Q1 Examples → continued

Benchmarking of cement degradation in cracked concrete due to chemical leaching; J. Perko et al. (2013)
Codes: MIN3P, OGS-GEM, Orchestra, Comsol-PHREEQC, HYTEC
→ Geared towards mathematical & numerical model benchmarking
Problem definition with gradual approach (5 case studies with increasing complexity)
→ Different solution approaches of thermodynamic system (LMA versus GEM)
→ Different transport solution methods (FD, FVM, FEM)
→ Different operator splitting techniques (SNIA, DSA)

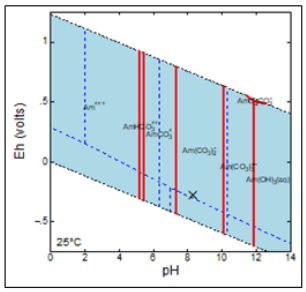
Benchmarking the ability of reactive transport codes to model cement/clay interactions; Marty et al. (2015)
→ PHREEQC2, iPHREEQC3 with external transport module, THOUGHREACT, CRUNCH, HYTEC, MIN3P, ORCHESTRA
→ THERMODDEM (Blanc et al., 2012)
→ Design setup for ILLW disposal cell plug with clay/concrete interface (reactivity over 100.000 years)

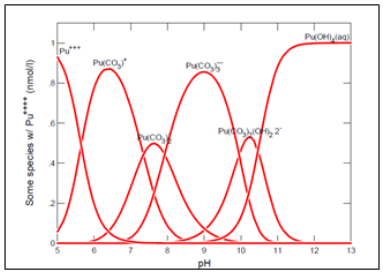
Objectives:
1) Make sure that all results obtained by the various codes agree in predicting the same mineralogical and chemical changes considering (i) steep pH and Eh gradients, (ii) highly complex mineralogies (considering local equilibrium and reaction kinetics), and (iii) the same mesh and transport parameters.
2) Improve confidence in long-term modeling

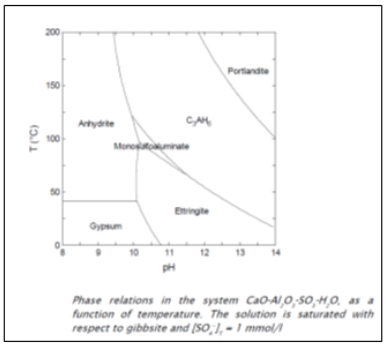
4

Q2 **What are the most relevant types of modelling for the ThermoChimie benchmarking exercise? Should we only model chemical processes at equilibrium (which ones?) or do you think we also need to tackle close to equilibrium processes and/or reactive transport?**

- Thermodynamic equilibrium calculations, such as e.g.:
  - → Speciation & solubility calculations at 25°C under different chemical conditions (higher temp. cond. could be interesting, including organic ligands)
  - → Calculation of saturation states of disposal relevant minerals
  - → Plotting predominance diagrams: check phase relations as function of temp. or different act. ratios
  - → Plotting activity diagrams







Phase relations in the system CaO-Al_2O_3-SO_3-H_2O, as a function of temperature. The solution is saturated with respect to gibbsite and $[SO_4^{2-}]_t$ = 1 mmol/l

5 © SCK•CEN, 2019

Q2 **Continued**

- Test impact of activity formulation in systems with ionic strengths > 0.3 m (Marty et al., 2015)
- Sorption modeling, as sorption behavior strongly depends on speciation
- Calculating gas solubilitites at 25°C and higher p-T conditions
- Testing effect of redox on speciation and solubility
- Formation of solid solutions

6 © SCK•CEN, 2019

Q2 Continued

- **Reactive transport calculations are considered to be less relevant, as they:**

→ Test rather the code capabilities than the TDB quality

→ Test transport simulation capabilitites (reactive mass transport processes)

→ Ability to "process" complex geochemistry

→ Check effect of different activity formulations and coupeling schemes

→ How geometry of the system under consideration is best represented (grid, mesh, 1D, 2D, radial, etc.),

→ How variation in total cation exchanger is calculated (different approaches),

→ How to effectively adapt time steps to reach correct numerical convergence

→ Implementation of kinetics → TST formulation, rate laws, test effect of kinetic rates on numerical capability

But still interesting, if kept simple .....☺

Q3

Which geochemical systems should be considered in this benchmarking exercise? In other words, what are the geochemical systems most likely to reveal the strengths and limitations of the TC database?

Disposal systems

- Clay (e.g. COX, BC, OPA) systems, backfill/containment materials (e.g. bentonite), cement/concrete systems
- Higher pH (cement/concrete), higher temperature (NF), high(er) ionic strength systems (e.g. salt repositories, seawater intrusion scenarios) & different combinations
  → *within range for which TDB was developed*
- Systems involving changing redox conditions (microbial activity, corrosion processes)
- Systems/processes comprising organic species (cellulose degradation → ISA)
- Systems with steep chemical (pH, Eh) & concentration gradients → e.g. clay/cement interfaces (RTM)

Other:

- Systems relevant for environmental applications: e.g. Remediation of contaminated sites (wrt heavy metals)
- Toxic metal behavior in industrial liquid effluents
- Sites of former mining activities (e.g. AMD sites)

Q4 **Should the benchmarking exercises be based on modelling results only or should the benchmarking refer to a selected set of well-defined experimental and/or natural (e.g., groundwater composition…) datasets as in the case of validation exercises?**

- Definitely good idea, if benchmarking refers to well defined exp. and/or natural datasets/analogues

  → Anyhow verification of thermodynamic data comprised in ThermoChimie is generally done via comparison to experimental data (e.g. solubility exp.), if possible.

Drawback: thermodynamic properties only apply to phases of determined composition (e.g. amount of bound water). The latter is however often not exactly known/given in exp. studies and may lead to erroneous calc., due to composition variation (e.g. of zeolites), different structures (e.g. of zeolitic framework), or degree of hydration (clays, zeolites, C-S-H), or presence of impurities.

Q5 **Which other high-quality databases should be considered for the benchmarking exercise? Which codes should be used for the benchmarking exercise?**

Suggested TDB's (except NEA)

- PSI-NAGRA (Thoenen et al., 2014; Hummel et al, 2002)
- JAEA (Kitamura et al. (2014)
- Yucca Mountain TDB (Wolery and Jove-Colon, Johnson et al., 1992, Oelkers et al., 2009)
- MOLDATA (Wang et al., 2010)

Suggested codes

- PHREEQC
- GWB

Less relevant:

- THERMODDEM (Blanc et al., 2012), as it is part of ThermoChimie or similar (but smaller)
- THEREDA (Moog et al., 2015; Gester et al., 2009), as mainly applicable to higher ionic strength systems and database "quite small"
- HATCHES (Cross and Ewart, 1991)

Q6 **Should the uncertainties provided in the database be used for benchmarking?**

- Indeed, it would be nice to include uncertainties in the benchmarking, but this seems almost impossible…., as they are generally not included in the electronic formats of the TDBs.

But, according to our knowledge none of the common computer/modeling codes up-to-now is able to include uncertainties of all species/complexes, solids and gases (e.g. in a solubility calculation) → eventually MC calculations could…..?

Q7 **What are the essential points of focus that need to be covered for the benchmarking exercise to usefully evaluate thermodynamic databases?**

- Completeness of database
- Traceability/documentation of data source and selection
- Data gaps/weaknesses
- Database formats
- Incorporation of uncertainties

- Internal consistency
- Estimation methods, extrapolations

References

Blanc, P., Lassin, A., Piantone, P., Azaroual, M., Jacquemet, N., Fabbri, A., Gaucher, E.C. (2012): Thermoddem: A geochemical database focused on low temperature water/rock interactions and waste materials. Appl. Geochem. 27(10), 2107–2116 (2012).

Cross, J.E., Ewart, F.T. (1991): HATCHES e a Thermodynamic Database and Management System. Radiochim. Acta 52, 421-422.

Gester, S., Altmaier, M., Brendler, V., Hagemann, S., Herbert, H.-J., Marquardt, C., Moog, H. C., Neck, V., Richter, A., Voigt, W., Wilhelm, S. (2009): THEREDA - The Thermodynamic Reference Database for Nuclear Waste Disposal in Germany. WM2009 Conference, March 1-5, Phoenix, AZ.

Hummel, W., Berner, U., Curti, E., Pearson, F.J., Thoenen, T. (2002): Nagra/PSI Chemical Thermodynamic Data base 01/01. Radiochim. Acta 90, 805-819.

Idiart, A.; Garçia, D.; and Grivé, M. (2013): Benchmark of reactive transport codes using the ThermoChimie Database. Poster presentation at the 14th Int. Conf. on the chemistry and Migration behavior of Actinides and Fission products in the geosphere, 8-13th September 2013, Brighton, UK.

Jacques, D.; Wang, L.; Martens, E.; and Mallants, D. (2014): Benchmarking the CEMDATA07 database to model chemical degradation of concrete using GEMS and PHREEQC. NEA//RWM/R(2012)3. Appendix D: Workshop posters.

Johnson, J.W., Oelkers, E.H., Helgeson, H.C. (1992): SUPCRT92: a Software Package for Calculating the Standard Molal Thermodynamic Properties of Minerals, Gases, Aqueous Species, and Reactions from 1 to 5000 bar and 0 to 1000°C. Comp. Geosci. 18, 899-947.

Kitamura, A., Doi, R., Yoshida, Y. (2014). Update of JAEA-TDB: Update of Thermodynamic Data for Palladium and Tin, Refinement of Thermodynamic Data for Protactinium, and Preparation of PHREEQC Database for Use of the Brønsted-Guggenheim-Scatchard Model. JAEA-Data/Code 2014-009.

Lothenbach, B., Matschei, T., Möschner, G., Glasser, F. (2008): Thermodynamic modelling of the effect of temperature on the hydration and porosity of Portland cement. Cem Conc Res, 38(1), 1-18.

Marty, N.C.M., Bildstein, O., Blanc, P., Claret, F., Cochepin, B., Gaucher, E.C., Jacques, D., Lartigue J-E.,,Liu, S., Mayer, K.U., Meussen, C.L., Munier, I., Pointeau, Su, D., Steefel, C.I. (2015): Benchmarks for Multicomponent reactive transport across a cement/clay interface. Comput. Geosci. DOI 10.1007/s10596-014-9463-6

Matschei, T., Lothenbach, B., Glasser, F. (2007): Thermodynamic properties of Portland cement hydrates in the system CaO-Al2O3-SiO2-CaSO4-CaCO3-H2O. Cem. Concr. Res., 37, 1379-1410.

Moog, H.C., Bok, F., Marquardt, C.M., Brendler, V. (2015): Disposal of NuclearWaste in Host Rock Formations Featuring High-saline Solutions e Implementation of aThermodynamic Reference Database (THEREDA). Appl. Geochem 55, 72-84.

Oelkers, E.H., Benezeth, P., Pokrovski, G.S. (2009): Thermodynamic databases for water-rock interaction. Rev. Mineral. Geochem., 70, 1-46.

Thoenen, T., Hummel, W., Berner, U., Curti, E. (2014): The PSI/Nagra Chemical Thermodynamic Database 12/07 PSI Bericht Nr. 14-04.

Wang, L. (2016): Benchmarking TDB on speciation /solubility calculations with GWB code and MOLDATA – Task N° 3 – Sensitivity analysis. Technical report, SCK•CEN/21572453, SCK•CEN T-0041.

Wang, L., Salah, S., De Soete, H. 2010. MOLDATA: A Thermochemical Database for Phenomenological and Safety Assessment Studies for Disposal of Radioactive Waste in Belgium. Proceedings of Clays in Natural & Engineered Barriers for Radioactive Waste Confinement - 4th international meeting - March 2010 - Nantes, France., 849-850.

Wolery, T.J., Jove-Colon, C.F. (2007): Qualification of Thermodynamic Data for Geochemical Modeling of Mineral-Water Interactions in Dilute Systems. Technical report, US DoE ANL-WIS-GS-000003 REV 01.

SCK•CEN
Studiecentrum voor Kernenergie
Centre d'Etude de l'Energie Nucléaire
Belgian Nuclear Research Centre

Stichting van Openbaar Nut
Fondation d'Utilité Publique
Foundation of Public Utility

Registered Office: Avenue Herrmann-Debrouxlaan 40 – BE-1160 BRUSSELS
Operational Office: Boeretang 200 – BE-2400 MOL

ThermoChimie is a thermodynamic database initially created and developed by Andra (French National Radioactive Waste Management Agency), for more than twenty years (1995). In October 2014, Radioactive Waste Management Limited (NDA, UK) joined the project and the "ThermoChimie consortium" was formed. In March 2018, Ondraf/Niras (National Agency for Radioactive Waste Management, Belgium) also joined the "ThermoChimie consortium".

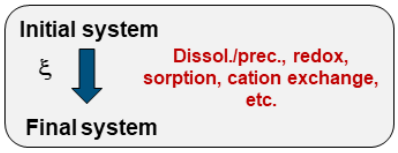
In waste management, geochemical modelling is used in support of the assessment of radionuclide and non-radiological pollutant behaviour in a range of scenarios, such as within radioactive waste packages and geological disposal facilities, through the geosphere, and in legacy contaminated land. This can be in support of repository performance assessments, research activities (such as modelling experiments), or decisions about waste conditioning, reprocessing, and disposability. However, an accurate, consistent and complete thermodynamic data set is required for these models to be meaningful.

ThermoChimie is designed to be applied over the 6 - 14 pH range at temperatures below 80°C and in systems with an Eh in the range -0.5V to +0.5V since these are the conditions generally expected within radioactive waste repositories. ThermoChimie provides robust thermodynamic data for a wide range of radionuclides and non-radiological pollutants, as well as major components expected within a geological disposal facility, including constituent host-rock mineral phases, bentonites, cements, and their evolving secondary phases. However, the database can be applied to other systems within the water stability domain. These thermodynamic data are mainly derived from comprehensive, active literature studies and are supplemented by an experimental program when required.

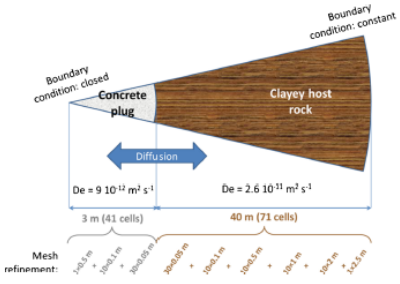
Q1 **What are the possible uses of a benchmarking exercise making use of thermodynamic databases?**

Common approaches:

- ❑ **Process / reaction path simulations (with or without kinetics)**



- ❑ **Performance assessment calculations**
  To simulate Rn behavior in waste packages and disposal facilities
  → Reactive transport/flow-through modeling



Marty et al. (2015)

17 © SCK•CEN, 2019

A.4 Tim Heath

wood.

ThermoChimie Benchmarking Workshop

Tim Heath
Manchester, 15 October 2019

woodplc.com

1. Possible uses of Benchmarking

- Main objective:
  - to test the key datasets in ThermoChimie against other databases or experimental data
- Benchmarking exercises may be used:
  - to test between different thermodynamic datasets or databases (data validation)
  - to test between model development approaches and user decisions (model/data validation)
  - to test, against measured data, whether ThermoChimie + appropriate model can explain relevant complex systems (model/data validation)
  - to test process implementation and data application between different geochemical programs (program verification)

1. Possible uses of Benchmarking for TDB

- Some possible applications:
  - predicting groundwater compositions, water mixing and conditioning (e.g. Grimsel CFM benchmarking)
  - predicting radionuclide solubility and speciation (e.g. Grimsel CFM, ThermoChimie/HATCHES comparison)
  - testing sorption models – but be clear on objective (e.g. Chemval, NEA Sorption Forum, ThermoChimie/HATCHES)
  - evolution of engineered barrier materials (e.g. NFCM)
  - standard test cases for new databases versions
  - testing of the use of uncertainty values in thermodynamic data (e.g. Grimsel CFM)

1. RWM Grimsel CFM – mixing of waters

Effect of modelling decisions on predicted Eh of mixed waters

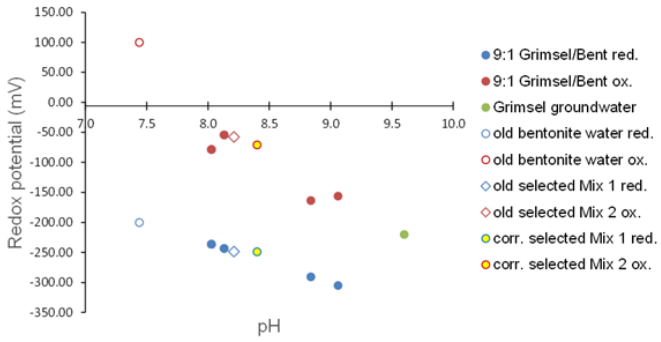


Figure 2 Redox potential as a function of pH in the original waters and mixed waters; (results for mixed water compositions used in subsequent calculations are shown as "corr. selected Mix 1 red" and "corr. selected Mix 2 ox").

1. RWM Grimsel CFM solubility benchmarking

1. Speciation calculation 1		
Bentonite porewater, reducing		
Species	Fraction of total diss. Tc [%]	Uncertainty
1. TcO(OH)2	99.3%	0.2%
2. Tc(OH)2CO3	0.4%	0.2%
3. Tc(OH)3CO3-	0.1%	0.1%

2. Speciation calculation 2		
Bentonite porewater, reducing		
Assumed solid phase	TcO2:1.63H2O(s)	Uncertainty
Total dissolved Tc [mol/l]	4.02E-09	
log10(Total dissolved Tc/ mol l-1)	*-8.39*	*0.24*
Species	Fraction of total diss. Tc [%]	
1. TcO(OH)2	99.3%	0.2%
2. Tc(OH)2CO3	0.4%	0.2%
3. Tc(OH)3CO3-	0.1%	0.1%
4. (TcO)(OH)3-	0.1%	0.1%

1. Speciation calculation 1		
Bentonite porewater, oxidising		
Species	Fraction of total diss. Tc [%]	Uncertainty
1. TcO4-	100.0%	0.0%

2. Speciation calculation 2		
Bentonite porewater, oxidising		
Assumed solid phase	TcO2:1.63H2O(s)	Uncertainty
Total dissolved Tc [mol/l]	4.93E-03	
log10(Total dissolved Tc/ mol l-1)	*-2.30*	*0.99*
Species	Fraction of total diss. Tc [%]	
1. TcO4-	100.0%	0.0%

1. Speciation calculation 1		
Grimsel groundwater		
Species	Fraction of total diss. Tc [%]	Uncertainty
1. TcO(OH)2	87.9%	11.3%
2. (TcO)(OH)3-	7.5%	6.4%
3. TcO4-	4.4%	10.4%
4. Tc(OH)3CO3-	0.2%	0.2%

2. Speciation calculation 2		
Grimsel groundwater		
Assumed solid phase	TcO2:1.63H2O(s)	Uncertainty
Total dissolved Tc [mol/l]	4.28E-09	
log10(Total dissolved Tc/ mol l-1)	*-8.33*	*0.26*
Species	Fraction of total diss. Tc [%]	
1. TcO(OH)2	86.9%	12.4%
2. (TcO)(OH)3-	7.9%	6.1%
3. TcO4-	4.9%	11.8%
4. Tc(OH)3CO3-	0.2%	0.1%

1. Speciation calculation 1		
Mixed water, reducing		
Species	Fraction of total diss. Tc [%]	Uncertainty
1. TcO(OH)2	99.2%	0.7%
2. (TcO)(OH)3-	0.7%	0.7%
3. Tc(OH)3CO3-	0.1%	0.1%

2. Speciation calculation 2		
Mixed water, reducing		
Assumed solid phase	TcO2:1.63H2O(s)	Uncertainty
Total dissolved Tc [mol/l]	4.01E-09	
log10(Total dissolved Tc/ mol l-1)	*-8.39*	*0.24*
Species	Fraction of total diss. Tc [%]	
1. TcO(OH)2	99.1%	0.6%
2. (TcO)(OH)3-	0.7%	0.6%
3. Tc(OH)3CO3-	0.1%	0.1%

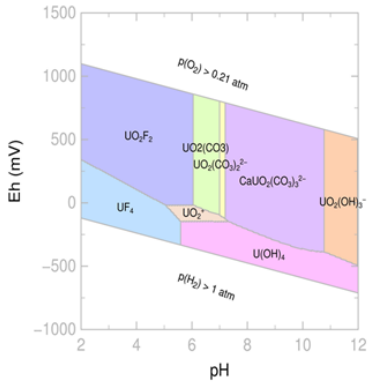
1. Speciation calculation 1		
Mixed water, oxidising		
Species	Fraction of total diss. Tc [%]	Uncertainty
1. TcO4-	65.2%	31.5%
2. TcO(OH)2	34.5%	31.3%
3. (TcO)(OH)3-	0.2%	0.3%

2. Speciation calculation 2		
Mixed water, oxidising		
Assumed solid phase	TcO2:1.63H2O(s)	Uncertainty
Total dissolved Tc [mol/l]	1.35E-08	
log10(Total dissolved Tc/ mol l-1)	*-7.62*	*0.74*
Species	Fraction of total diss. Tc [%]	
1. TcO4-	65.0%	31.0%
2. TcO(OH)2	34.7%	30.7%
3. (TcO)(OH)3-	0.2%	0.3%

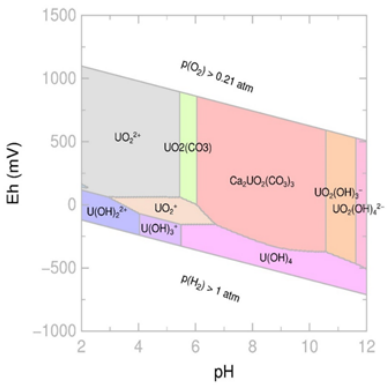
1. RWM Grimsel CFM: radionuclide redox chemistry

Uranium speciation in:

Grimsel groundwater



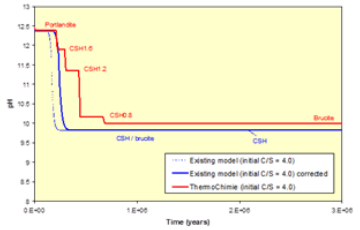
bentonite water



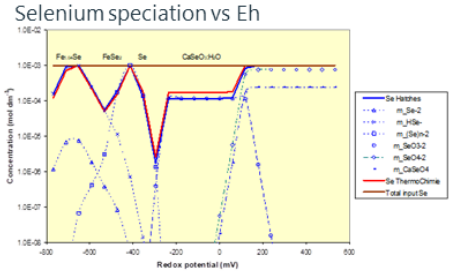
Graphical comparisons are useful in comparing benchmark exercise results (as well as tabulated)

1. RWM ThermoChimie review (vs HATCHES)

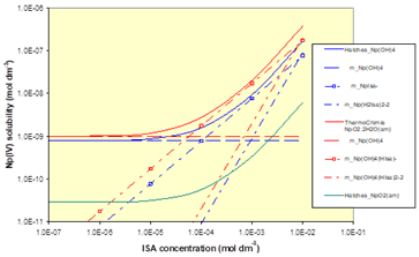
Backfill evolution



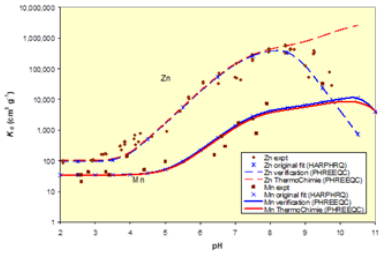
Selenium speciation vs Eh



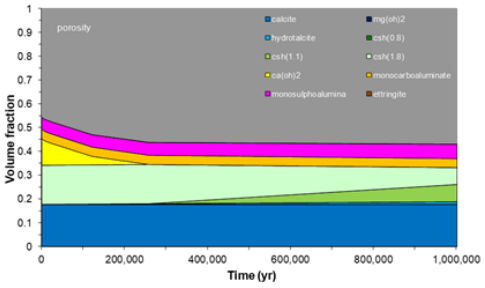
Neptunium solubility vs [ISA]

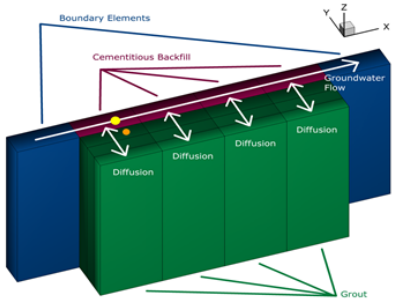


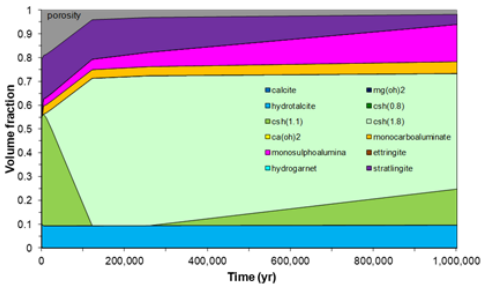
Zn, Mn sorption to montmorillonite vs pH

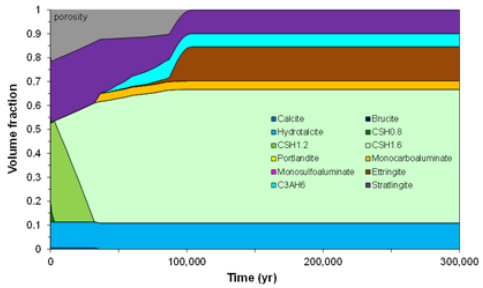


1. RWM: Near-field Component Model



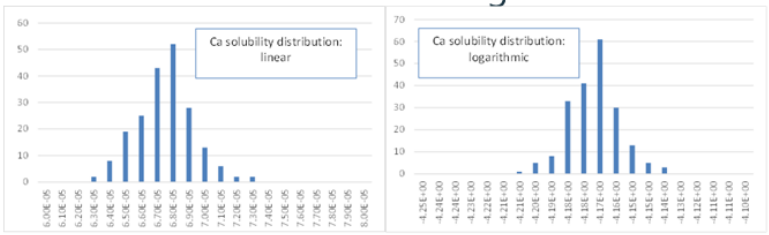




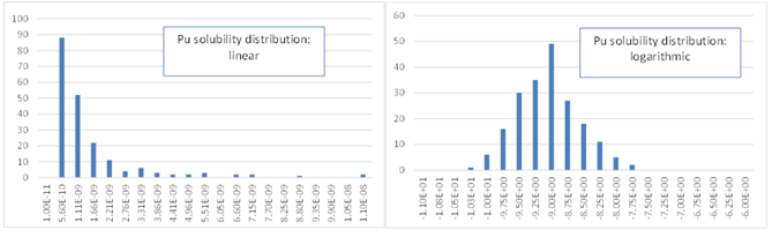


1. RWM Grimsel - predicted solubility distributions

linear logarithmic



Calcium narrower uncertainty in input TD data



Plutonium higher uncertainty in input TD data

2. Types of modelling: equilibrium, "close to" or R-T

- ThermoChimie is thermodynamic database so the main emphasis should be on equilibrium conditions and thermodynamic data
  - some amorphous solids included, e.g. UO_2:2H_2O(am) (their definitions in ThermoChimie could be improved)
- "Close to" equilibrium is governed by kinetics
- Reactive-transport models are generally complex
  - used in exercises to compare user decisions on model development
  - not ideal for testing thermodynamic datasets
  - but maybe necessary for testing understanding of complex systems

2. Types of modelling: equilibrium, "close to" or R-T

- Is the question really about whether relevant kinetic data should be included in ThermoChimie?
- Could define kinetic parameters for use in common systems under ThermoChimie project, but:
  - better to consider this question explicitly (if at all)
  - better to keep separate kinetic data system and combine at modelling stage?
  - would need:
    - careful definition of parameters
    - recording of ranges of applicable conditions
  - may need
    - a reference thermodynamic database for defining some kinetic data (e.g. for mineral dissolution)

3. Geochemical systems

- Application: systems relevant to radwaste disposal:
  - groundwater chemistry and mineral interactions
  - radionuclide chemistry
    - redox boundaries, solubility, (sorption), effect of organics
    - await NEA/TDB update publication for relevant radionuclides?
  - EBS porewater chemistry and mixing
  - EBS materials and their evolution
    - **cementitious materials**
    - bentonite
  - container materials and their evolution

3. Geochemical systems

- Consider the effects of:
  - ionic strength on groundwater, porewater or radionuclide chemistry
  - temperature, e.g. on bentonite close to HLW/SF container
  - uncertainty, e.g. on radionuclide solubility

3. Cementitious systems

- Important area for ILW safety cases
  - initial compositions of cementitious backfills and encapsulation grouts
  - conditioning of porewater and effects on radionuclide solubility and sorption
  - backfill evolution due to reactions with waste and groundwater components
- Current version of ThermoChimie:
  - has a rather limited dataset
  - C-S-H representation by just three C/S ratio solids
  - should be considered for review and update?
- Benchmarking exercises could be a good starting point for a data review

3. Cementitious systems

- Significant data developments since data set last considered, including:
  - EMPA-led work and associated thermodynamic data developed in CEMDATA (Lothenbach et al. C&CR 2018)
  - ThermoChimie-related experimental work (Roosz et al. Appl. Geochem. 2018)
  - NEA/TDB state-of-the-art review?

3. Cementitious systems

- Suggest it is a high priority for a benchmarking exercise; this might include:
  - prediction of initial backfill mineralogy and porewater
  - evolution of mineral assemblage and porewater, including effect of:
    - groundwater interactions
    - waste interactions
- Exercise should include comparison with a range of databases, but particularly:
  - Latest version of CEMDATA with Nagra/PSI aqueous database
  - ThermoChimie "plus" (updated for Roosz et al 2018 data)

4. Benchmarking basis: modelling results or experimental/ natural data

- **Both approaches valid**
  - need to be clear for each exercise what the objectives are and what is being tested
  - try to avoid exercises where effects of modelling decisions and different databases are difficult to separate
- **Modelling testing:**
  - useful for testing ThermoChimie against other databases
  - can just be testing whether database is right or wrong in the same ways as other databases
  - but useful where other databases are known to have more developed or validated datasets
- **Experimental or natural system data testing**
  - good for validation of ThermoChimie against real data
  - will also depend on user modelling decisions
  - identify planned experimental studies and predict before measurement?

5. Which other databases and programs?

- Databases

  Nagra/PSI + CEMDATA
  NEA/TDB (radionuclides)
  JAEA database
  THERMODDEM
  PHREEQC standard database (geochem.)
  LLNL database
  THEREDA

- Programs

  PHREEQC
  ToughReact
  PFLOTRAN
  Geochemist Workbench

- Need to select appropriate databases for particular system/ exercises
- Probably a topic for general discussion

6. Should uncertainties be included?

- Yes, I suggest that this would be useful
  - particularly, from a UK perspective in supporting a probabilistic approach to performance assessment calculations
  - but also more generally for investigations of uncertainty
- Some examples of application on UK programme include:
  - radionuclides solubilities (Grimsel CFM)
  - radionuclide release from a cementitious near-field (Near-field component model)
  - prediction of uncertainty in radionuclide input parameters for assessment calculations (ongoing work)

6. Should uncertainties be included?

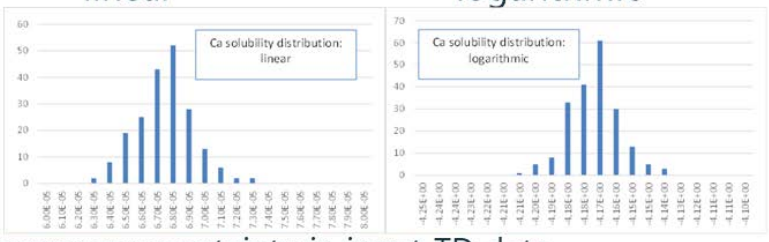
- What would such an exercise look like?
  - based on probabilistic chemical modelling:
    - random sampling of input parameter values
      - primarily thermodynamic data
      - others such as initial concentrations
    - from normal distributions defined by uncertainty value
    - implies assumption that parameters are uncorrelated
  - Monte Carlo approach (100s or 1,000s of calculation)
    - each calculation based on separate set of sampled values
    - allows analysis of sensitivity of results to each parameter and output parameter correlations

6. Should uncertainties be included?

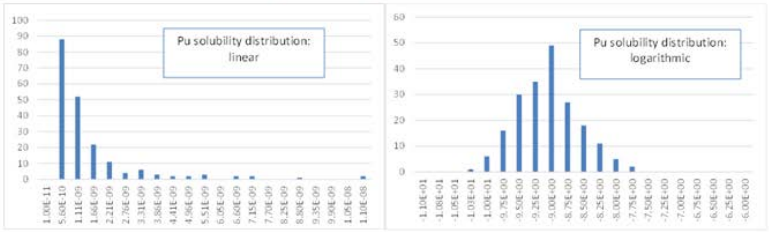
- What systems should it be applied to?
  - in principle could be any relevant geochemical system
  - suggest selecting simple and relevant calculations e.g.
    - prediction of porewater compositions
    - solubility of selected radionuclides
- Issues to consider:
  - thermodynamic parameter correlations
  - unnecessary error inflation

6. Predicted solubility distributions

linear logarithmic

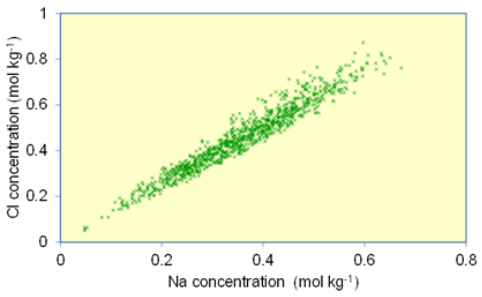


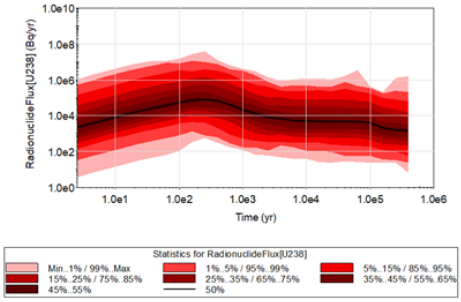
Calcium narrower uncertainty in input TD data

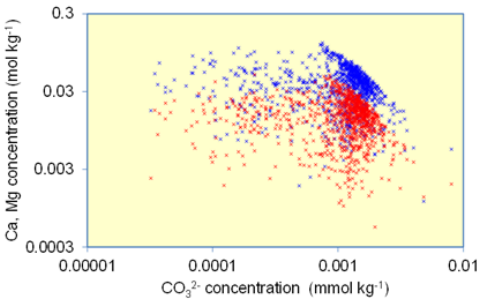


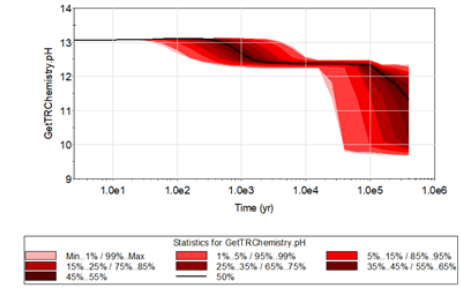
Plutonium higher uncertainty in input TD data

6. NFCM: groundwater uncertainty and U release









6. RWM: testing inputs for assessment calculations

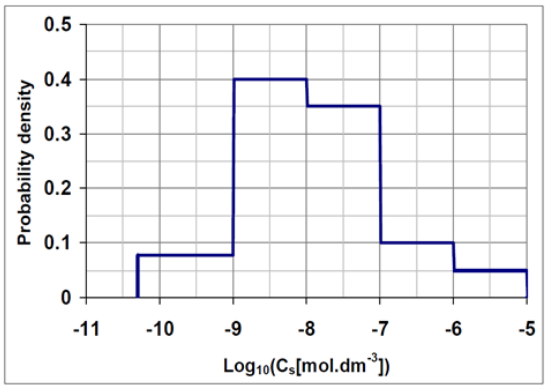


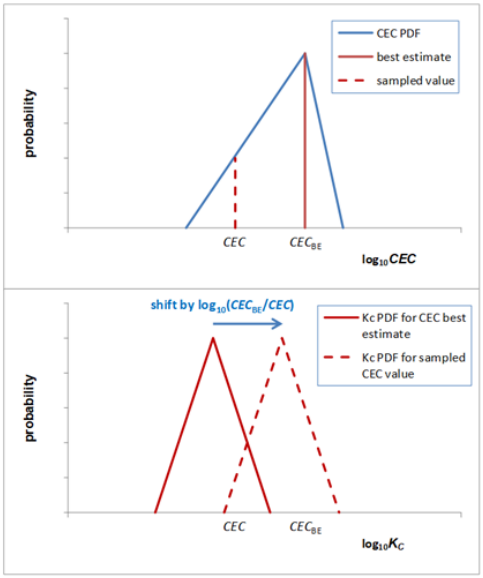
Figure 2. Probability density function corresponding to the elicited cumulative distribution function for the solubility C_s of uranium(IV) for pH in the range 12.3 to 13.5

UK approach includes uncertainty distributions (PDFs) for input parameters

- determined by formal expert elicitation
- probabilistic chemical modelling allows independent testing of elicited PDFs

Example from Serco Report SA/ENV/0920 Issue 3, 2007

6. Parameter correlations in probabilistic modelling



Example of simple inverse correlation between concentration of uptake sites (CEC) and selectivity coefficients (K_c)

6. Error inflation

- "To-and-fro" data conversions associated with database entry and application in geochemical programs typically lead to unnecessary error inflation
- For example, becomes an issue for radionuclide solubility dissolution equilibria:
  - if determined from data at high pH (e.g. with higher aqueous hydrolysis products), but
  - converted to form involving simple metal ion in database, and
  - reapplied at high pH in geochemical model

7. Essential points - general approach

- At exercise definition stage: careful consideration of objective and form of results for comparison
  - Thermodynamic data set testing should be main objective (in preference to model decisions, program verification)
  - Where exercise is purely to test thermodynamic data, use a standard program (PHREEQC?) and fully define model
    - no modelling decisions at implementation - just change database
  - Where model decisions or program verification is part of the exercise, ensure key effects can be separated:
    - define standard database (ThermoChimie) and program (PHREEQC?) to be used as a baseline for all cases with variant calculations for other databases or programs

7. Essential points – priority systems

- Cementitious systems in context of ILW disposal concepts
  - including evolution of mineral assemblage and porewater
  - effect of groundwater composition
  - include CEMDATA and "ThermoChimie plus" comparisons
- Bentonite systems in context of HLW/spent fuel disposal concepts
  - including evolution of mineral assemblage and porewater
  - effect of temperature, corrosion products
- Use of uncertainty values
  - e.g. in porewater composition prediction
  - radionuclide solubility

7. Essential points – priority systems

- Radionuclide chemistry:
  - solubility and speciation
  - organic complexation
  - comparison of how gaps in NEA/TDB are filled in different databases
  - await update publication from NEA/TDB (and ThermoChimie update) before implementing?

7. Suggestions for uncertainties in ThermoChimie

- Ensure a single, clear definition of uncertainty for ThermoChimie in user documentation and website
  - including statement that it is consistent with NEA/TDB
  - identify and fill missing uncertainty values
  - avoid over-estimation of uncertainties (can grossly distort outputs of probabilistic calculations
  - for key radionuclide solubility limiting phases consider:
    - re-definition of dissolution equilibria and uncertainty assignment, or
    - extra analysis and database fields

    to allow user to avoid error inflation

Thank you for your attention

A.5 Laurent De Windt

ThermoChimie Benchmarking Workshop

Manchester, 15 – 16 October 2019

Dr. Laurent De Windt
Centre de Géosciences, Fontainebleau (France)

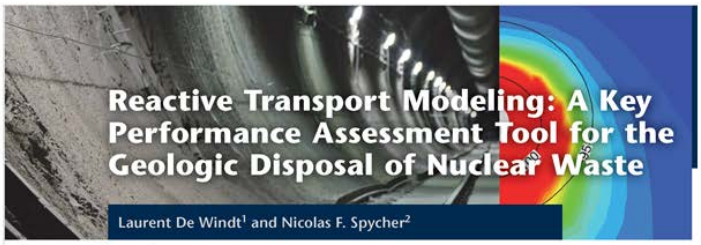
laurent.dewindt@mines-paristech.fr

MINES ParisTech

1

ThermoChimie Benchmarking Workshop

Manchester, 15 – 16 October 2019



Elements (2019)

$Pb_3(H(UO_2)_3O_2(PO_4)_2)_2 \cdot 12H_2O$

Dr. Laurent De Windt
Centre de Géosciences, Fontainebleau (France)

laurent.dewindt@mines-paristech.fr

MINES ParisTech

2

Q1 – 7. Main messages

- Emphasize on
  - **step-by-step => multi component complex** systems (e.g. set of RN in the full chemistry of a pore water)
  - **redox titration/transient stage** (e.g. actinide speciation in subsurface systems, or transient stage in deep disposal)
  - **temperature titration/decrease** (e.g. clay/cement & clay/metal systems of deep disposals)
- Mostly **thermodynamic equilibrium**
  - maybe **simple RTM** to discriminate multiple reaction fronts (or activity – activity diagrams)
  - Experimental/natural data bring a positive support
    - but careful account of the full experimental conditions

3

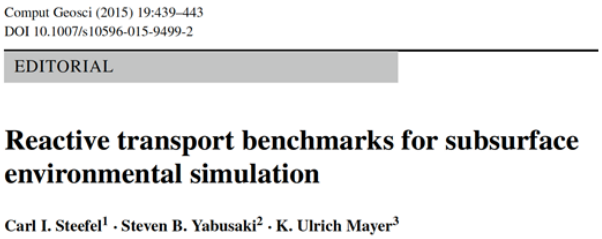
Q1 & 7. Uses and key points of TDB benchmarking

- The **main goal** is to **compare** results **output**
- **Multi-component**, interacting and realistic systems to assess of
  - **completeness** (no lack of essential reactions or secondary phases) for the **speciation** of an **element** (Pu…) or **solid phase** (corrosion products…)
  - **quantitative** results (solubility, aqueous complexes, Mg/Si ratio in M-S-H…)
  - **robustness and internal consistency** when changing a key parameter, e.g. the **redox potential, temperature…**
- **Supported by** some points of comparison with **experimental or natural data** from complex systems

4

Q1 & 7. Uses and key points of TDB benchmark (cont.)

- **Significant interests** to
  - safety and performance of radwaste disposals
  - draw the attention of the wider scientific community ("Migration in the subsurface") to the TC database
- A set of targeted **scientific papers**, as done for benchmarking RTMs


Comput Geosci (2015) 19:439–443
DOI 10.1007/s10596-015-9499-2

EDITORIAL

Reactive transport benchmarks for subsurface environmental simulation

Carl I. Steefel[1] · Steven B. Yabusaki[2] · K. Ulrich Mayer[3]

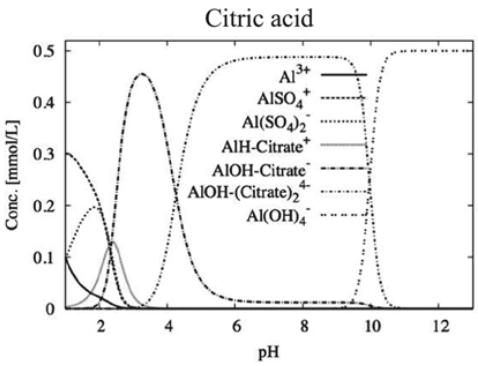
5

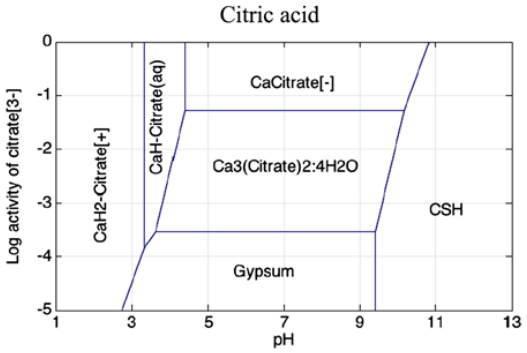
Q2. Types of modeling

- **Core = chemical thermodynamic equilibrium**
  - **Activity-activity diagrams** to highlight differences between TDBs
  - **Titration models** (range of pH, Eh or species concentrations) in addition to speciation calculations of a single solution
  - *(to minimize numerical uncertainties brought about by modeling kinetics or RTM)*
- **Some RTMs** in a simple configuration (1D regular mesh) for complex **binary systems** (cement/clay interfaces)
  - Instead of mixing of two geochemical systems
  - spatial discrimination between **multiple reactive fronts** & ranges of aqueous concentrations
  - TC + RTM is positive from a communication view point

6

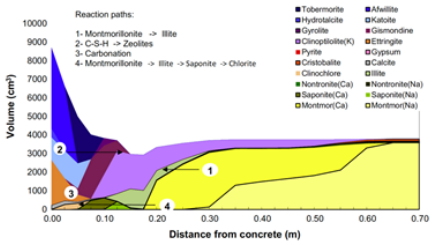
Q2. Types of modeling





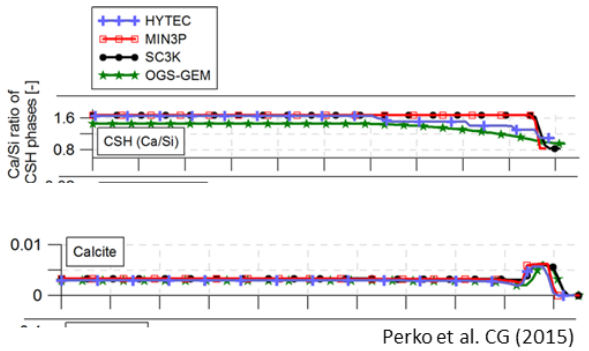
De Windt et al. CCR (2015)

Q2. Types of modeling



Blanc et al. AG (2015)

Case no.	Chemical system	Update of transport properties	Transport mode
0	Conservative tracer	No update	Diffusive
1	Simple	No update	Diffusive
2	Simple	No update	Advective
3	Simple	Update	Advective[a]
4	Complex	Update	Advective[a]
5	Complex + contaminant	Update	Advective[a]



Perko et al. CG (2015)

Q2. Types of modeling (complement)

- **Solid solutions** vs. discrete phases
  - essential for the **cement phases** (C-S-H, AFm…)
  - *(the cationic end members of clay phases (Na- vs. Mg-smectite…) vs. cation exchange models)*

- *(Near-equilibrium modeling*
  - *can smooth sharp transitions between mineral phases of similar formation constants, useful while comparing TDB,*
  - *but saturation indices bring the same information).*

9

Q3. Geochemical systems of relevance

RN speciation

- **Scopus TC** => radionuclide (+++) As, Cr Cm, Th, U Mo, Ni, Ra, Se

  => cement phases & zeolites (++), clay phases (+)

- **Speciation** of several key **radioactive elements** in natural waters
  - deep geologic (and subsurface?) environments
    - *(e.g. the COx porewater, at midway between diluted and saline clay chemistry)*
  - balance between remaining uncertainty and interest for the safety; e.g. U, Pu, and Se, ?
- Aqueous chemistry & saturation indices of all solid phases
- Equilibrium with a selected set of solid phases relevant for solubility limits

10

Q3. Geochemical systems of relevance

RN speciation (cont.)

- Effect of the **redox potential titration** on speciation at a fix pH
  - e.g. competition between ternary Ca(Mg)-CO_3-U complex and the U(VI)/U(IV) redox couple but also <u>for other actinides</u>
  - to simulate a transient stage from oxic to anoxic conditions
- Effect of a temperature decrease
  - progressive decreasing of heat release in disposal facilities,
  - large set of TC data over 10 – 90 °C
- *(Effect of salinity (ionic strength) less essential for the TC benchmark since it depends more on the activity model used than the TDB itself)*
- Phosphates or hydrophilic organic species (e.g. carboxylic acids)?

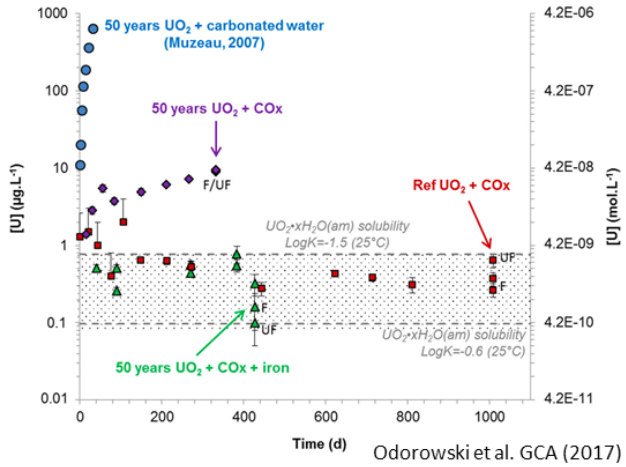
11

Q3. Geochemical systems of relevance

RN speciation (cont.)

	PAC1002
pH	7.2
Eh (mV)	-177
Ionic strength (m)	0.087
Concentration (M)*	
TIC	$2.1 \cdot 10^{-3}$
Ca	$8.0 \cdot 10^{-3}$
Na	$4.3 \cdot 10^{-2}$
Mg	$5.0 \cdot 10^{-3}$
K	$1.0 \cdot 10^{-3}$
Fe**	$<1 \cdot 10^{-3}$
Al	--
Cl	$4.1 \cdot 10^{-2}$
SO_4	$1.5 \cdot 10^{-2}$
S^{2-}	--
Sr**	$<1 \cdot 10^{-3}$
Si**	$<1 \cdot 10^{-3}$

*Considered as total concentration
** Upper limit values

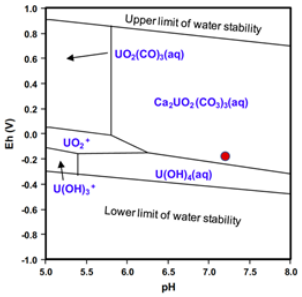


Odorowski et al. GCA (2017)

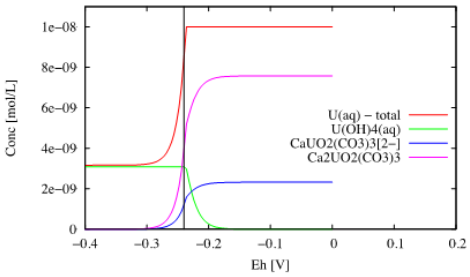
12

Q3. Geochemical systems of relevance

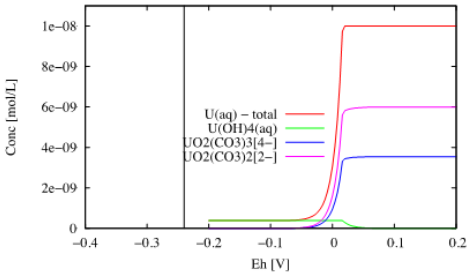
RN speciation (cont.)



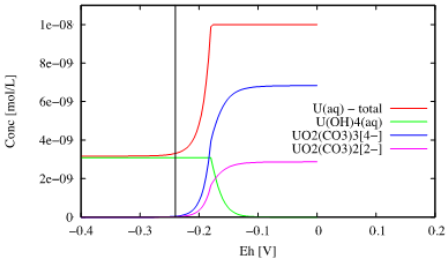
TC



EQ3/6 – old NEA



TC without Ca-UO_2-CO_3



pH = 7.1, Fe(III)/Fe(II) : Eh = -0.24 V

Q3. Geochemical systems of relevance

RN speciation (cont.)

Table 2
Calculated solubility for different radionuclide solid phases under Callovo-Oxfordian conditions, at 25°C and 80 °C.

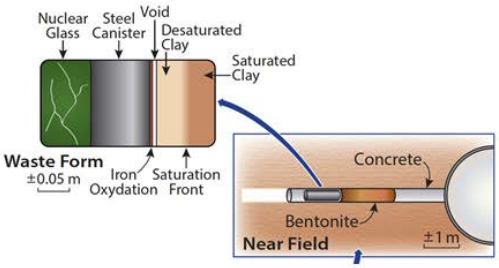
Element	25 °C		80 °C	
	Solid phase	Solubility (m)	Solid phase	Solubility (m)
Am	$Am(CO_3)(OH)_{(am)}$	$6 \cdot 10^{-6}$	$Am(CO_3)(OH)_{(cr)}$	$1 \cdot 10^{-6}$
Th	$ThO_{2(aged)}$[a]	$1 \cdot 10^{-8}$	$ThO_{2(aged)}$	$6 \cdot 10^{-9}$
U	$UO_2{:}2H_2O_{(am)}$[a]	$3 \cdot 10^{-6}$	$UO_2{:}2H_2O_{(am)}$[a]	$3 \cdot 10^{-5}$
Np	$NpO_2{:}2H_2O_{(am)}$	$1 \cdot 10^{-9}$	$NpO_2{:}2H_2O_{(am)}$	$4 \cdot 10^{-9}$
Pb	$Pb(CO_3)_{(s)}$	$8 \cdot 10^{-7}$	$Pb(CO_3)_{(s)}$	$7 \cdot 10^{-6}$
Cd	$Cd(CO_3)_{(s)}$	$4 \cdot 10^{-6}$	$Cd(CO_3)_{(s)}$	$1 \cdot 10^{-5}$
Nb	$Nb_2O_{5(s)}$	$1 \cdot 10^{-7}$	$Nb_2O_{5(s)}$	$5 \cdot 10^{-8}$
Mo	$MoO_{2(s)}$	$2 \cdot 10^{-7}$	$MoO_{2(s)}$	$1 \cdot 10^{-6}$

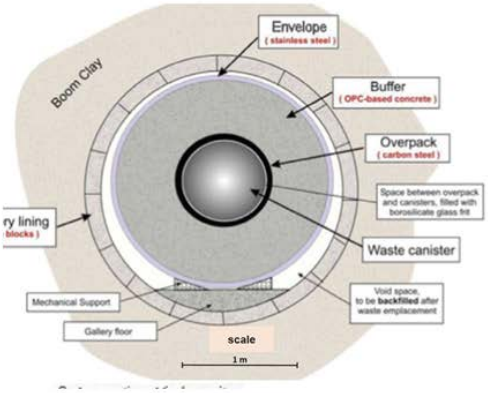
Grivé et al. AG (2015)

Q3. Geochemical systems of relevance

Binary complex interfaces

- **Cement/clay & iron/clay** are key systems in geological disposal
  - by using RTMs (see Q2)





15

Q3. Geochemical systems of relevance

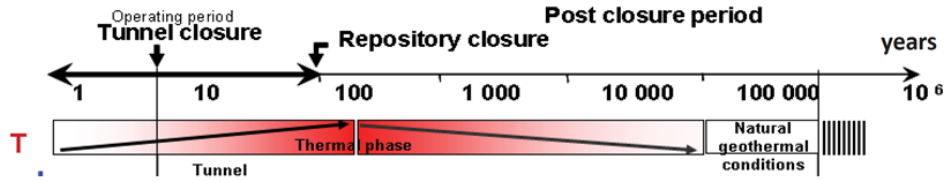
Binary complex interfaces

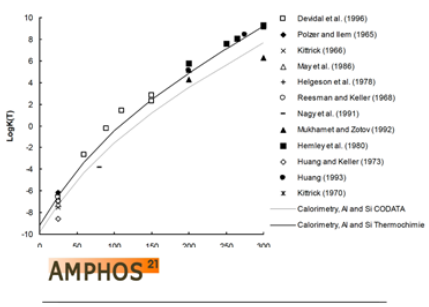
- **Cement/clay & iron/clay** are key systems in geological disposal
  - by using RTMs (see Q2)
- **Cement/clay** is maybe the easiest system
  - cement phases are mostly under thermodynamic equilibrium
  - large amount of constant data for the cement phases
  - the clay phases are also well represented in TC
  - beyond CEM I?
- **Corrosion products of iron** (+ Cu, Mg?) in a natural water at different temperatures
  - as an alternative to the iron/clay interface
  - (although metastable phases can coexist for kinetic reasons)
- **Avoid replication** with previous RTM benchmarking

16

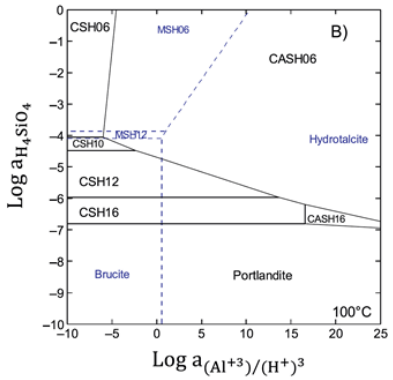
Q3. Geochemical systems of relevance

Binary complex interfaces (cement/clay)





ThermoBridge
Task 1.
Validation of the extracted databases using test cases

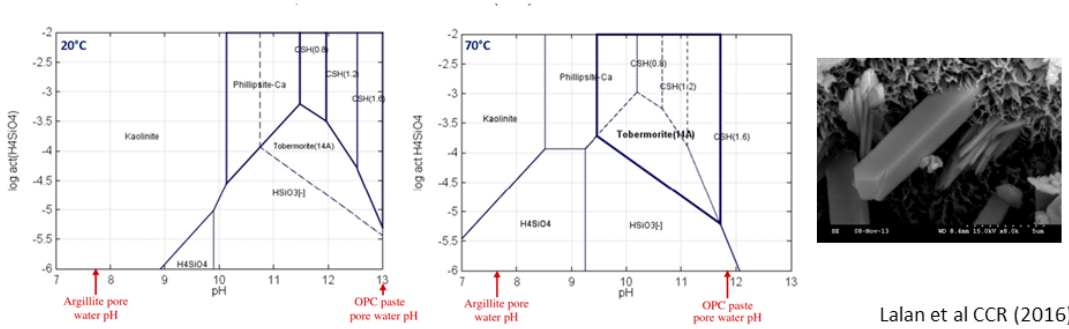


Rooz et al. AG (2018)

Q3. Geochemical systems of relevance

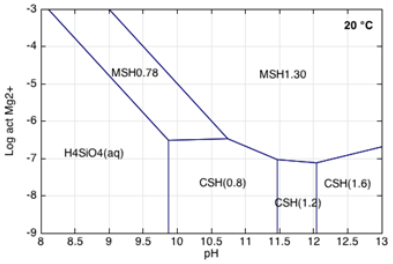
Binary complex interfaces (cement/clay)

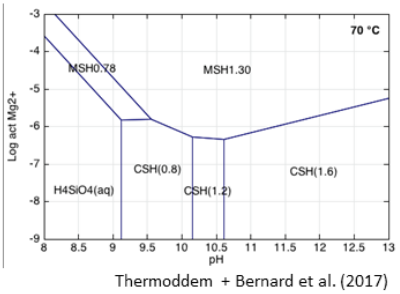
Ordinary Portland cement



Lalan et al CCR (2016)

Alternative cements to the Portland one

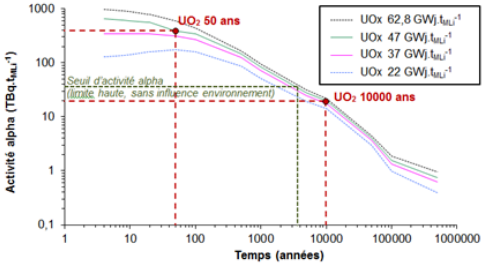


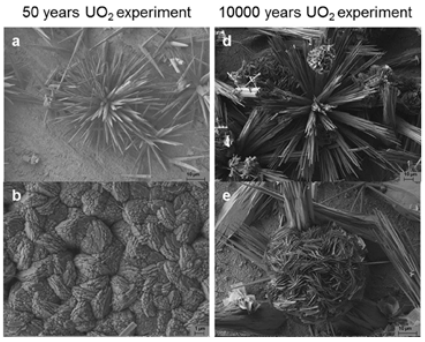


Thermoddem + Bernard et al. (2017)

Q3. Geochemical systems of relevance

Binary complex interfaces (corrosion products)





Goethite	α-FeOOH	Fe(III)
Lépidocrocite	γ-FeOOH	Fe(III)
Akaganéite	β-FeOOH	Fe(III)
Hématite	α-Fe_2O_3	Fe(III)
Maghémite	γ-Fe_2O_3	Fe(III)
Ferrihydrite	$(Fe)_2O_3 \cdot 0.5H_2O$	Fe(III)
Magnétite	Fe_3O_4	Fe(II)-Fe(III)
Rouilles vertes	$Fe^{II}_2Fe^{III}O_x(OH)_y$	Fe(II)-Fe(III)
Greenalite	$Fe^{II}_{2.3}Fe^{III}_{0.5}Si_{2.2}O_5(OH)_{3.3}$	Fe(II)-Fe(III)
Cronstedtite	$Fe^{II}_2Fe^{III}_2SiO_5(OH)_4$	Fe(II)-Fe(III)
Hydroxychlorure	β-$Fe_2(OH)_3Cl$	Fe(II)
Chukanovite	$Fe_2(OH)_2CO_3$	Fe(II)
Sidérite	$FeCO_3$	Fe(II)
Pyrite	FeS_2	Fe(II)
Chlorite	$(Fe,Mg,Al)_6(Si,Al)_4O_{10}(OH)_8$	Fe(II)-Fe(III)

Odorowski et al. GCA (2017) 19

Q3. Geochemical systems of relevance

Speciation & binary complex interfaces

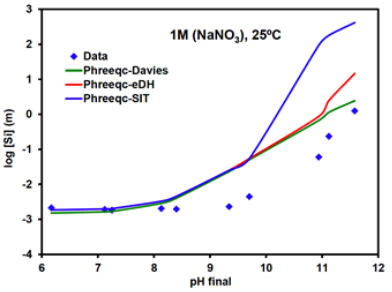
- A benchmark exercise **coupling radionuclide speciation & engineered barrier** evolution
  - will illustrate that TC is able to simultaneously tackle these two aspects of performance and safety assessment
  - although sorption cannot be considered in TC

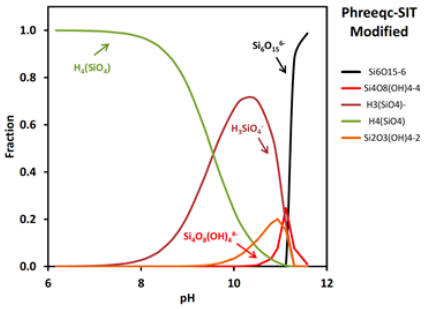
20

Q3. Geochemical systems of relevance

SIT & silica solubility

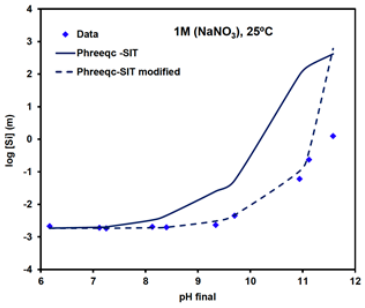
$\log(\gamma_i) = -Az_i^2\sqrt{I_m}$	DEBYE-HÜCKEL
$\log(\gamma_i) = \frac{-Az_i^2\sqrt{I_m}}{1 + Ba_i\sqrt{I_m}}$	EXTENDED DEBYE-HÜCKEL
$\log(\gamma_i) = -Az_i^2\left(\frac{\sqrt{I_m}}{1+\sqrt{I_m}} - 0.3I_m\right)$	DAVIES
$\log(\gamma_i) = -z_i^2\left(\frac{A\sqrt{I_m}}{1+b\sqrt{I_m}} + \frac{2}{b}\ln\left(1+b\sqrt{I_m}\right)\right) + \sum_c\sum_a m_c m_a B'_{ca} + \sum_c\sum_{c'} m_c m_{c'} \phi'_{cc'} + \sum_a\sum_{a'} m_a m_{a'} \phi'_{aa'} + \sum_c m_c(2B_{ca} + ZC_{ca}) + \sum_a m_a(2\phi_{ca} + \sum_c m_c \psi_{cca}) + \sum_c\sum_a m_c m_a \psi_{aca} + \lvert z_i\rvert\sum_c\sum_a m_c m_a C_{ca} + 2\sum_n m_n \lambda_{ni}$	PITZER
$\log(\gamma_i) = -z_i^2\left(\frac{A\sqrt{I_m}}{1+Ba_i\sqrt{I_m}}\right) + \sum_k \varepsilon(i,k,I_m)m_k$	SIT





AMPHOS 21

ThermoBridge
Task 1.
Validation of the extracted databases using test cases



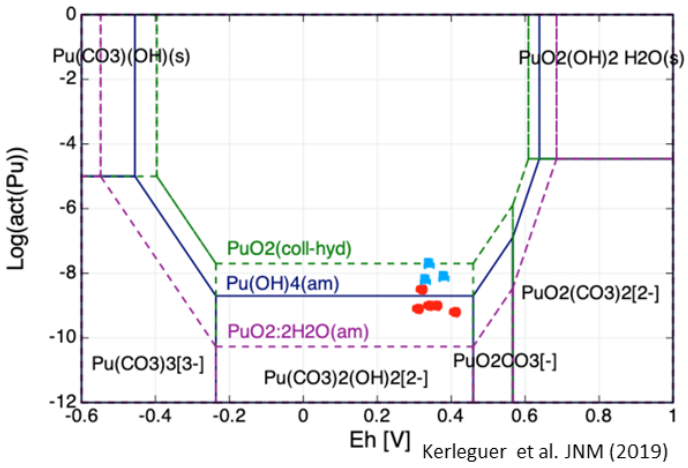
21

Q4. Experimental/natural datasets

- Validation on experimental data = job of the NEA expert groups
- Some points of comparison with one or two experimental/natural data from complex systems
  - comparison with a few complex real systems **brings confidence and visibility** to TC
  - maybe in “background” to classical TDB benchmarking
  - **careful account** has to be taken of the full experimental **conditions** (e.g. amorphous phases?)

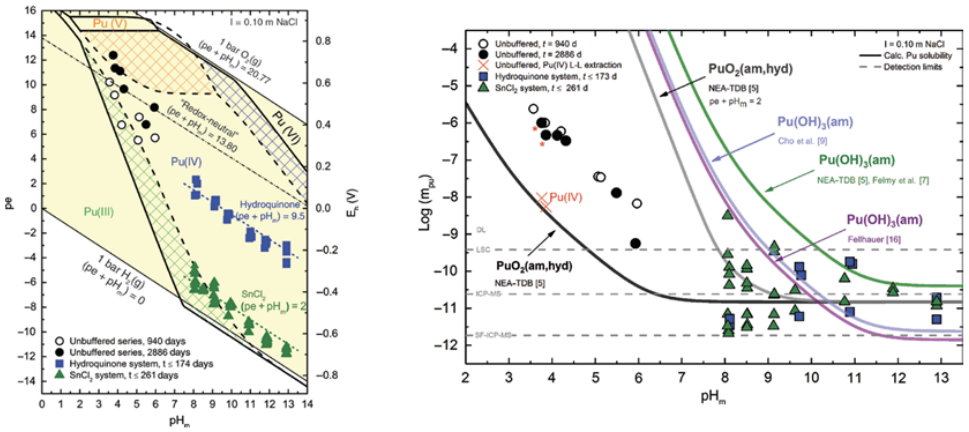
22

Q4. Experimental/natural datasets



Kerleguer et al. JNM (2019)

Q4. Experimental/natural datasets



Tasi et al., RCA 2018
Redox behavior and solubility of plutonium under alkaline, reducing conditions

Q5a. International TDB for benchmarking

- The strength of **TC** is its ability to be relevant for **both radionuclide speciation and multibarrier materials**
  - *(maybe necessary to select a different set of TDB according to the exercises)*
- CEMDATA2018 essential for cement phases but highly specialized
  - => **NAGRA/CEMDATA** can be a more ubiquitous choice
- A non-European TDB, e.g. the **LLNL TDB** (or the JAEA TDB ?)
- *(I do not know how different the NEA and MOLDATA TDB are from TC, THERADA seems to focus on highly saline environments)*
- *(Thermoddem has many common data with TC)*

25

Q5a. International TDB for benchmarking

Cement and Concrete Research 115 (2019) 472–506

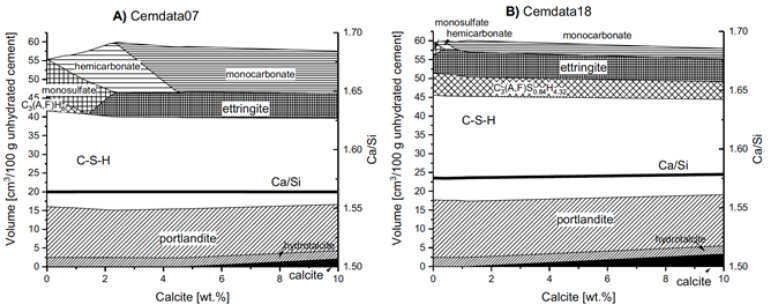
Contents lists available at ScienceDirect

Cement and Concrete Research

journal homepage: www.elsevier.com/locate/cemconres

Cemdata18: A chemical thermodynamic database for hydrated Portland cements and alkali-activated materials

Barbara Lothenbach[a,*], Dmitrii A. Kulik[b], Thomas Matschei[c], Magdalena Balonis[d], Luis Baquerizo[e], Belay Dilnesa[f], George D. Miron[b], Rupert J. Myers[g,1]



26

Q5a. International TDB for benchmarking

Journal of Geochemical Exploration 204 (2019) 33–42

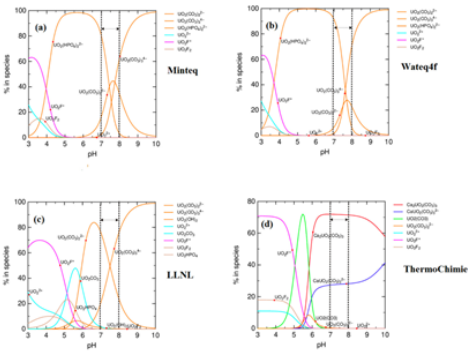
Contents lists available at ScienceDirect

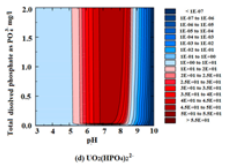
Journal of Geochemical Exploration

journal homepage: www.elsevier.com/locate/gexplo

Effect of thermodynamic database selection on the estimated aqueous uranium speciation

Xinyu Wang[a,b,c], Zeming Shi[b,c,*], David G. Kinniburgh[d], Laishi Zhao[c], Shijun Ni[a], Ruilin Wang[a], Yun Hou[a], Ke Cheng[a], Bocheng Zhu[a]





Applied Geochemistry 100 (2019) 213–222

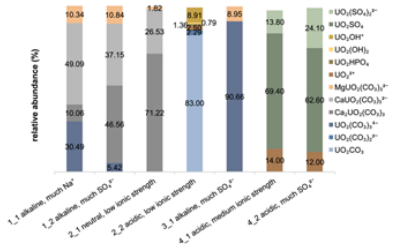
Contents lists available at ScienceDirect

Applied Geochemistry

journal homepage: www.elsevier.com/locate/apgeochem

Speciation of uranium: Compilation of a thermodynamic database and its experimental evaluation using different analytical techniques

Elena L. Mühr-Ebert[b,*], Frank Wagner[b], Clemens Walther[a]



Q5a. International TDB for benchmarking

Hörbrand et al. Geotherm Energy (2018) 6:20
https://doi.org/10.1186/s40517-018-0106-3

Geothermal Energy

RESEARCH Open Access

Validation of hydrogeochemical databases for problems in deep geothermal energy

Thorsten Hörbrand[1,2*], Thomas Baumann[2] and Helge C. Moog[3]

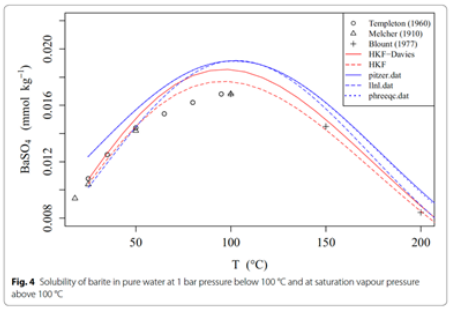


Fig. 4 Solubility of barite in pure water at 1 bar pressure below 100 °C and at saturation vapour pressure above 100 °C

Table 1 Summary of model approaches used in the present work

Code	PHREEQC			ChemApp
Parameter file	phreeqc.dat	llnl.dat	pitzer.dat	slop16.dat
Source of thermodynamic data	Unknown	Mainly: slop	Unknown	slop16
Activity model	Extended Debye–Hückel equation		Pitzer equations	Davies equation
Approx. ionic strength limits of the activity model (mol/kg)	0.1–3[a]	0.1–3[a]	6[a]	0.1–0.5[a]
Pressure dependence	HKFmoRR	None	HKFmoRR	HKF
Temperature dependence	Fitted polynomial equation or two-term extrapolation [cf. Eqs. (1) and (2)]			HKF

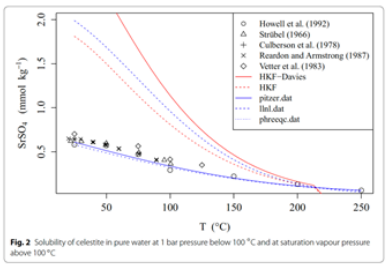


Fig. 2 Solubility of celestite in pure water at 1 bar pressure below 100 °C and at saturation vapour pressure above 100 °C

Q5a. International TDB for benchmarking

Table 1
NEA TDB reviews containing selection of thermochemical data.

Publication	Year	Authors
Chemical Thermodynamics of Uranium	1992	I. Grenthe, J. Fuger, R. J. M. Konings, R. Lemire, A. B. Muller, C. Nguyen-Trung, H. Wanner
Chemical Thermodynamics of Americium	1995	R. J. Silva, G. Bidoglio, M. H. Rand, P. B. Robouch, H. Wanner, I. Puigdomenech
Chemical Thermodynamics of Technetium	1999	J. A. Rard, M. H. Rand, G. Anderegg, H. Wanner
Chemical Thermodynamics of Neptunium and Plutonium	2001	R. Lemire, J. Fuger, H. Nitsche, P. Potter, M. H. Rand, J. Rydberg, K. Spahiu, J. C. Sullivan, W. J. Ullman, P. Vitorge, H. Wanner
Update on the Chemical Thermodynamics of Uranium, Neptunium, Plutonium, Americium and Technetium	2003	R. Guillaumont, T. Fanghänel, J. Fuger, I. Grenthe, V. Neck, D. A. Palmer, M. H. Rand
Chemical Thermodynamics of Nickel	2005	H. Gamsjäger, J. Bugajski, T. Gajda, R. Lemire, W. Preis
Chemical Thermodynamics of Selenium	2005	Å. Olin, B. Noläng, L.-O. Öhman, E. Osadchii, E. Rosén
Chemical Thermodynamics of Zirconium	2005	P. L. Brown, E. Curti, B. Grambow, C. Ekberg
Chemical Thermodynamics of Compounds and Complexes of U, Np, Pu, Am, Tc, Se, Ni and Zr with Selected Organic Ligands	2005	W. Hummel, G. Anderegg, I. Puigdomenech, L. Rao, O. Tochiyama
Chemical Thermodynamics of Thorium	2009	M. Rand, J. Fuger, I. Grenthe, V. Neck, D. Rai
Chemical Thermodynamics of Tin	2012	H. Gamsjäger, T. Gajda, J. Sangster, S. K. Saxena, W. Voigt
Chemical Thermodynamics of Iron, Part 1	2013	R. Lemire, U. Berner, C. Musikas, D. A. Palmer, P. Taylor, O. Tochiyama

Table 3
Basic characteristics of NEA-influenced thermodynamic databases.

Database	Link	Data source included	NEA TDB guidelines used	Software compatibility
NEA TDB	www.oecd-nea.org/dbtdb/			
Yucca mountain	-	Yes	No	EQ3/6, PHREEQC
THEREDA	www.thereda.de	Yes	Yes	EQ3/6, PHREEQC, ChemApp, Geochemist's Workbench, JSON
PSI-NAGRA	www.psi.ch/les/database	Yes	In most cases	PHREEQC
JAEA	http://migrationdb.jaea.go.jp/	Yes	Yes	EQ3/6, PHREEQC, Geochemist's Workbench
HATCHES	www.hatches-database.com/	Yes	Yes	PHREEQC, PICKER/PHREEQE, EQ3/6
MOLDATA	-	Yes	No	PHREEQC, Geochemist's Workbench
THERMOCHIMIE	www.thermochimie-tdb.com/	Yes	Yes	PHREEQC, Crunch, Chess, Toughreact

29

Q5b. Codes for benchmarking

- **PHREEQC** code
  - interfaced with most TDB,
  - several activity-correction models
  - can do simple RTM
  - *(this minimizes numerical discrepancies among codes or possible errors while extracting the data for each code)*

- Other codes (such as CHESS) if it eases project management
- If RTMs: CrunchFlow, HYTEC, ThoughReact…

30

Q6. Uncertainties in the database

- *(I am not familiar with the topic)*
- Very **important** than the TC web site and **documentation** provide an estimation of the uncertainty for each thermodynamic constant
  - but too much information in the database itself makes the file less readable
- **Uncertainty propagation** for a **multi-component** system
  - but maybe out of the scope of TC benchmarking?

31

Q1 – 7. Main messages

- Emphasize on
  - **step-by-step => multi component complex** (e.g. set of RN in the full chemistry of a pore water)
  - **redox titration/transient stage** (e.g. actinide speciation in subsurface systems, or transient stage in deep disposal)
  - **temperature titration/decrease** (e.g. clay/cement & clay/metal systems of deep disposals)
- Mostly **thermodynamic equilibrium**
  - maybe **simple RTM** to discriminate multiple reaction fronts (or activity – activity diagrams)
  - Experimental/natural data bring a positive support
    - but careful account of the full experimental conditions

32

A.6 Kastriot Spahiu

Benchmarking and validation of Thermochimie.

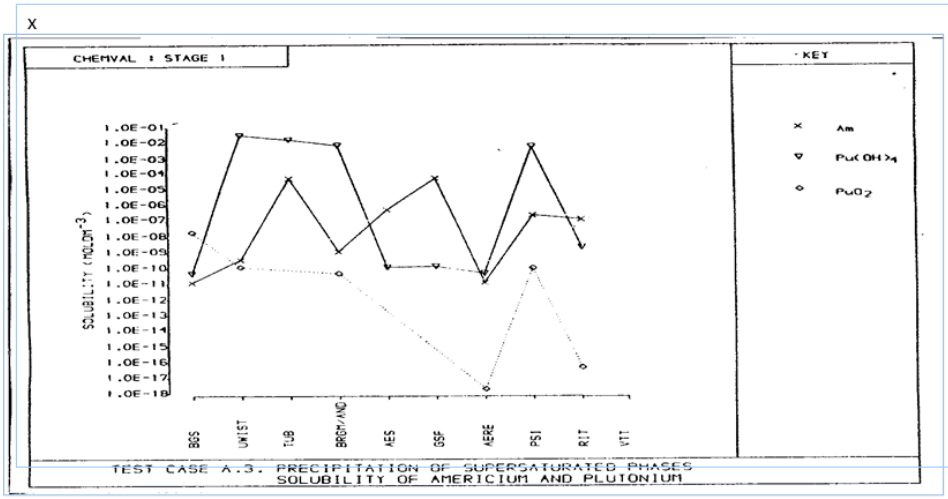
Kastriot Spahiu
Chalmers & SKB

Answers to the questionnaire.

- Wikipedia: "Benchmarking is the practice of comparing business processes and performance metrics to industry bests and best practices from other companies"
- "Based on a compilation of 308 references, Balci&Sargent state that the following 16 items are commonly encountered and used interchangeably in various disciplines: *acceptability, accuracy, analysis, assessment, calibration, certification, confidence, credibility, evaluation, performance, qualification, quality assurance, testing validation and verification*." Sheng et al. 1993.
- Examples of benchmarking exercises: Chemval 1 and 2, T. Wollery White paper, B. Merkel on speciation of U(VI) in groundwaters and A. Emrén et al. on solubility of $Pu(OH)_4(s)$ in granitic groundwaters in the following slides.
- Thermochimie is a thermodynamic database and should be tested how well it performs in equilibrium calculations used to determine the speciation and solubilities of radionuclides in relevant repository scenarios. The use in a reactive transport model should include a check of the kinetics of the most important complexation, redox, sorption or other reactions.
- Benchmarking can be carried out comparing the modelling results obtained with Thermochimie and another database (Nagra-PSI, JAEA,EQ 3/6 or CWB databases, or by comparing the modelling results with field groundwater data. In the second case, data from experiments using radionuclides have to be selected by the interested organizations in case one wants to test radionuclide speciation. An alternative is modelling field data from natural analogues.

Some examples of benchmarking or validation, not specific for databases.

Validation exercise at Chemval 1



(*From CEC Report EUR 12237 en (1989): CHEMVAL Project Stage 1: Verification of speciation models*)

Tom Wollery´s White Report 2005.

- In 2005 at NEA-TDB a white report by Tom Wollery was delivered, pointing out discrepancies between SUPCRT92 or NIST databases (applied in EQ3/6 database) and the CODATA 89 used at NEA.
- Of course we took it seriously and checked the phosphoric acid dissociation constants. They were almost equal in all databases, meaning that the discrepancies were due to calorimetric data.
- As it was discovered later by Joe Rard, the whole set of phosphate data suffered from large errors because of the inclusion of a single erroneous value (heat of combustion and $\Delta H_f°P_4O_{10}$ from Holmes, 1962) used in the process of data selection for the NIST database.

Table 6. Comparison of 298.15K values of the standard Gibbs energies of formation ($\Delta G°_f$ J/mol) of key aqueous species, SUPCRT92 (as used by Steinborn et al., 2003; see text) vs. CODATA89 (Cox et al., 1989). Species shown match those in previous tables. Discrepancies of magnitude greater than 1000 J/mol are highlighted in red; lesser discrepancies with magnitude greater than 500 J/mol are highlighted in pink.

Species	CODATA89	SUPCRT92	Discrepancy
Ag+	77096.0	77098.6	-2.5
Al+++	-491507.0	-483708.1	-7798.9
Br-	-103850.5	-104056.1	205.6
CO3--	-527899.8	-527983.1	83.4
Ca++	-552806.2	-552790.1	-16.1
Cd++	-77732.8	-77655.0	-77.7
Cl-	-131218.4	-131289.7	71.4
ClO4-	-7890.5	-8535.4	644.8
Cs+	-291455.4	-291666.6	211.2
Cu++	65040.1	65584.2	-544.1
F-	-281523.3	-281750.6	227.2
H+	0.0	0.0	0.0
HCO3-	-586845.2	-586939.9	94.7
HPO4--	-1095985.5	-1089137.0	-6848.5
HS-	12243.1	11966.2	276.9
HSO4-	-755315.1	-755755.9	440.8
H2PO4-	-1137152.4	-1130265.8	-6886.6
Hg++	164667.4	164682.2	-14.8
Hg2++	153566.5	153594.6	-28.1
I-	-51723.4	-51923.4	200.1
K+	-282509.6	-282461.8	-47.7
Li+	-292918.3	-292599.7	-318.7
Mg++	-455375.1	-453984.9	-1390.2
NH4+	-79397.9	-79454.2	56.3
NO3-	-110794.3	-110905.3	111.0
Na+	-261952.9	-261880.7	-72.1
OH-	-157219.9	-157297.5	77.6
Pb++	-24237.9	-23890.6	-347.3
Rb+	-284008.9	-283675.2	-333.7
SO4--	-744004.5	-744459.1	454.6
Sn++	-27623.8	-27488.9	-134.9
UO2++	-952550.7	-952613.1	62.4
Zn++	-147203.4	-147276.8	73.4

B. Merkel, 2012-Phreeqe C calculations with various TDB

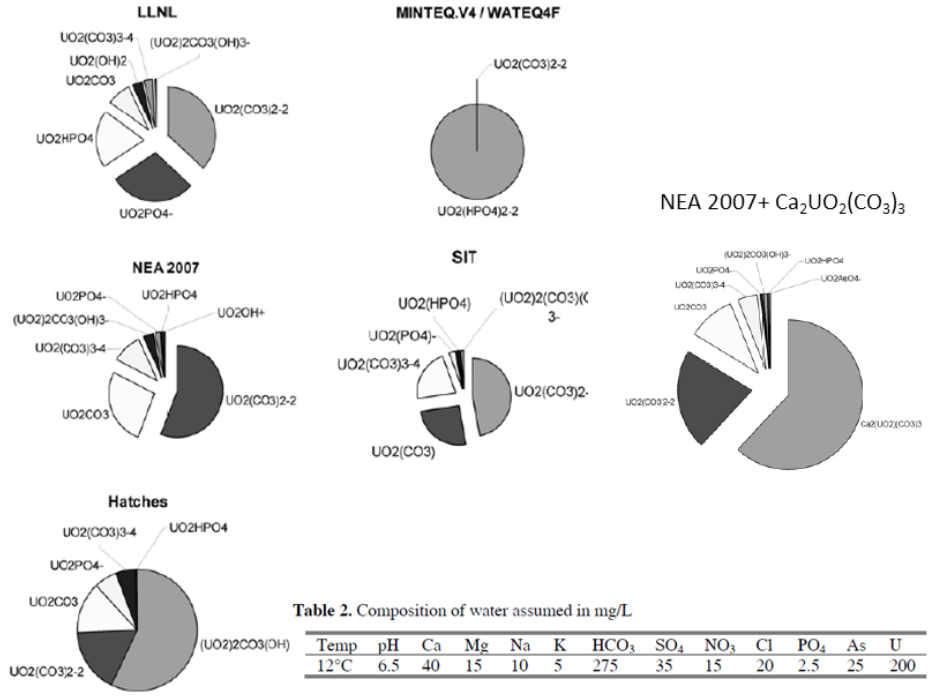
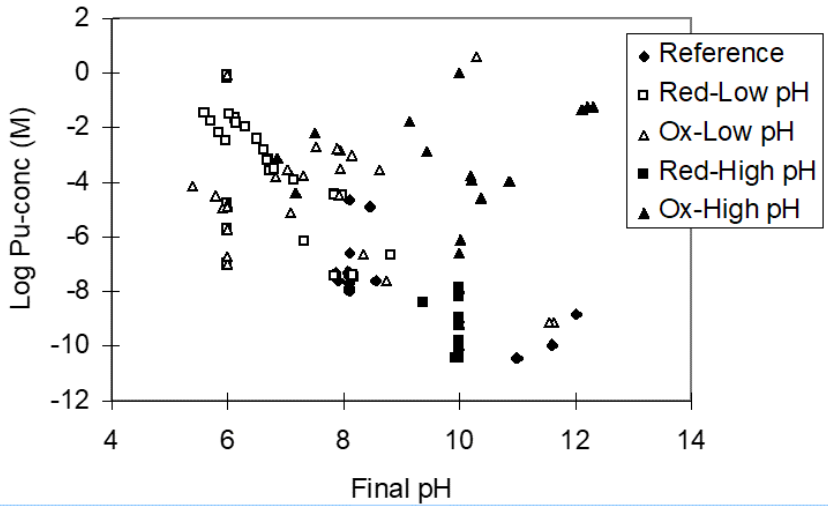


Table 2. Composition of water assumed in mg/L

Temp	pH	Ca	Mg	Na	K	HCO_3	SO_4	NO_3	Cl	PO_4	As	U
12°C	6.5	40	15	10	5	275	35	15	20	2.5	25	200

Calculated solubilities of $Pu(OH)_4$ by 4 modelers (Emrén et al. 1999)

The same version of Hatches used by 4 modelers. Equilibration with fracture filling minerals used in most cases. Besides the base case (pH=8.3, pe=-4.37) calculations at low pH=6 and high pH=10 , under reducing (pe=-4.37) or oxidizing conditions (pe=15) were also modelled.

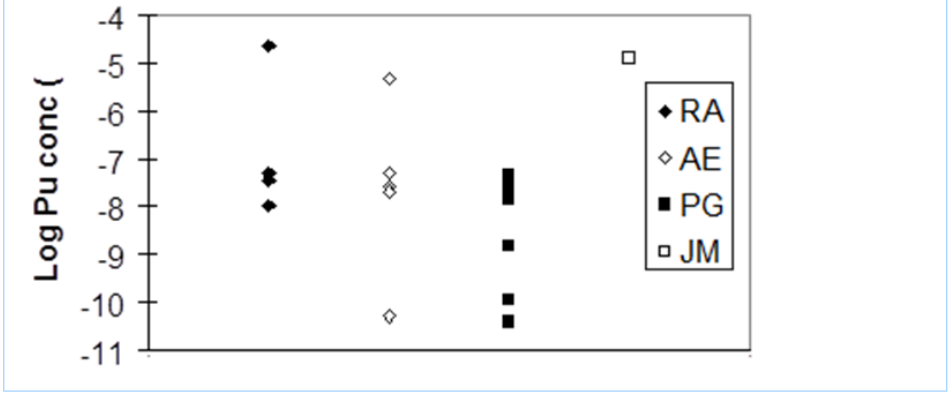


Calculation for the base case, pH=8.3, pe= -4,13

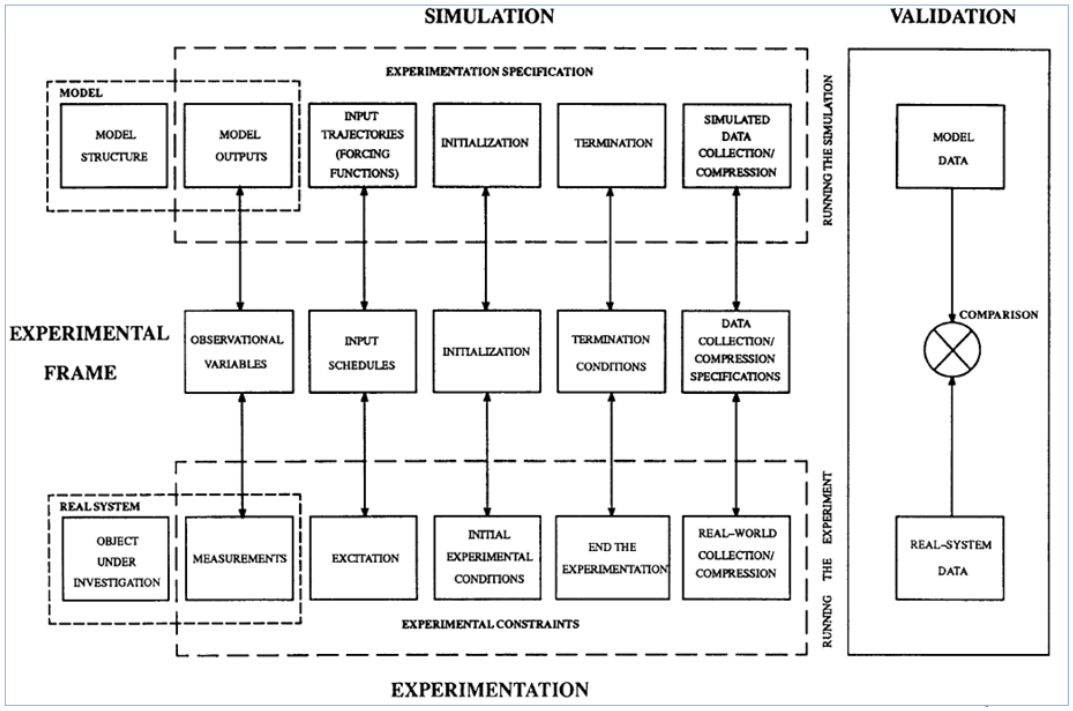
RA-Randy Arthur, QuantiSci
AE-Allan Emrén, Chalmers
PG-Pierre Glynn, USGS
JM-Jude McMurry, AECL

Groundwater composition

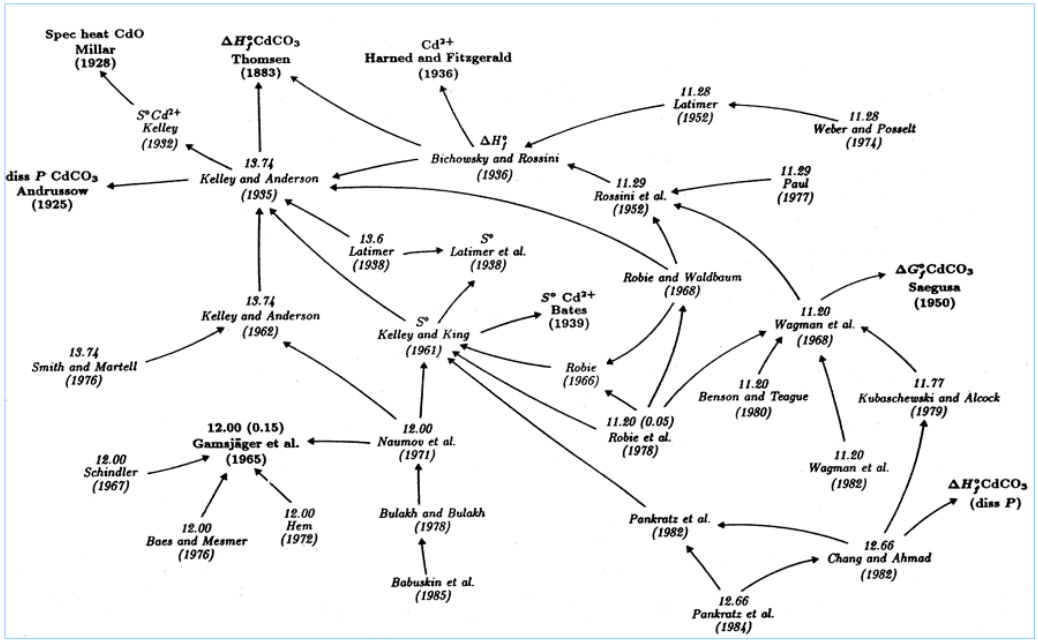
Element	Conc (M)
Ca	4.72E-02
Mg	1.73E-03
Na	9.13E-02
K	2.07E-04
Fe	4.37E-06
Mn	5.28E-06
Si	1.46E-04
Cl	1.81E-01
HCO3- (tot)	1.64E-04
S (tot)	5.83E-03
F	7.90E-05



Concept of experimental frame (Ören & Ziegler 1979)

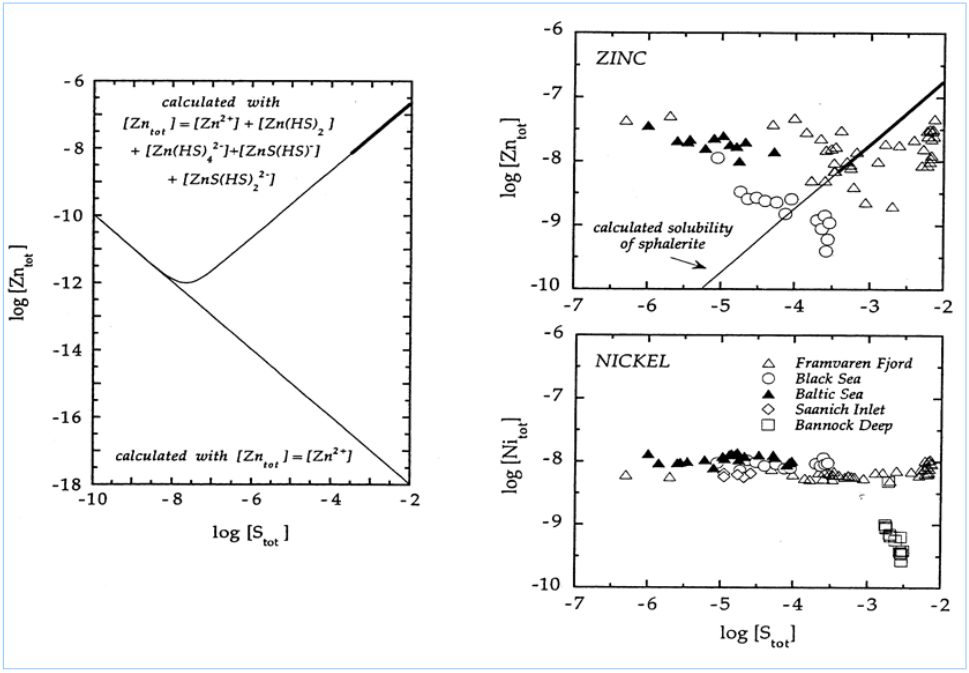


"Genealogical tree" of data on the solubility of $CdCO_3$(s) [Stipp et al. 1993]



Stipp, S., Parks G., Nordström K., Leckie J., Geochim. Cosmochim. Acta 57(1993)2699-2713.

A complete TDB-NiS and ZnS solubilities [Thoenen 1999]



Benchmarking and validation

- The geochemical code to be used should be chosen based on the version of the database used together with it. A SIT or Pitzer version should not be used together with constants for weak complexes. Vice versa, it has to be checked how standard Phreeqe C solves the problem of activity coefficients for e.g. Eu(III) ion in a chloride solution: does it assign a complexation constant internally or uses the one in the database.
- The careful analysis of the benchmarking results is important to localize problems and correct them or to reveal the advantages of the database.
- A benchmarking exercise for a TDB should be carried out with the same calculation code and using various other TDB in order to be able to carry out the correct analysis of the results.
- The experimental or modeling data should concern bedrock of interest for the organizations financing the development of the database. I see no problem if the modeling of highly saline waters as those of e.g. WIPP has flaws; on the other hand the correct modelling of test cases concerning clay or granite ground waters combined with high pH concrete waters should be correct.
- Also the correct referencing of the source data in the database should be given special attention-an example of the "jungle" of cross-referencing of thermodynamic data is given by Stipp et al. 1993.
- The internal consistency of the thermodynamic data (ΔG, ΔH, ΔS) should be checked and issues concerning the completeness of the database should also be discussed. Issues of the completeness of the database can be solved only by expert judgement, while the estimation of missing data should be based in methods discussed e.g. in the corresponding chapter of "Modelling in aqueous solutions".

Thank you!

Additional issues

- Discussion with IP reveals two minor errors in Amphos 21 TDB (not clear if they are in TC also) concerning Ag and Ra databases. Aina B. pointed out several differences in numerical values of constants between Chimera (Medusa TDB) and Thermochimie-not checked in detail but the material can be made available to the Thermochimie team.
- In Giffaut et al. 2014 accuracy, completeness and traceability are cited as main properties of Thermochimie. Development of database connected with experimental program on clay minerals, cementitious phases, actinides and fission products.
- Completeness cited through inclusion of complexes with constants having large uncertainty. Accuracy checked with tests of modelling exercises. The general process of data selection ensures consistency-this is not clear to me. Includes molar volumes to calculate porosity of precipitates during transformations.
- Data selected based on: previous TDB compilations, open literature, experimental program, estimations.
- Data for major elements of host rock and EBS: 350 minerals contained, ca 100 phylosilicates .
- Near field perturbations and kinetic modelling-not clear how this is achieved. Does Thermochimie contain rate constants?
- Personal example of noble metal solution concentrations (Mo, Ag, Pd, Pt, Rh, Ir. Os) around 10^{-10} M which does not correspond to the solubility of the corresponding oxides(higher) or that of the metal under reducing conditions(much lower).

A.7 Don Reed

UNCLASSIFIED

ThermoChimie Benchmarking Workshop Examples from the WIPP Project

Donald T. Reed

Team Leader: Actinide Chemistry and Repository Science Project
Repository Science and Operations
Los Alamos National Laboratory, Carlsbad Operations

UNCLASSIFIED

Overview

- My perspectives and point of view
- Role and importance of benchmarking studies
  - Example of complex brine system studies
  - Example of amorphous vs. crystalline An(III) phases
  - Example of regulatory "reality" vs. expected repository processes
- Key geochemical systems
- Role of uncertainties
- Most critical focus benchmarking exercise

Los Alamos NATIONAL LABORATORY EST.1943
UNCLASSIFIED
Operated by the Los Alamos National Security, LLC for the DOE/NNSA
NNSA

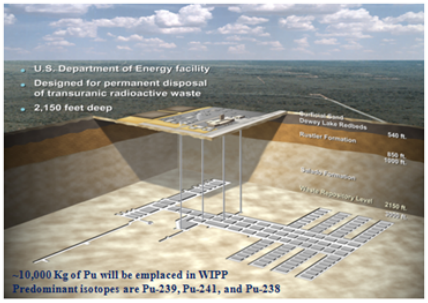
Safety Case for a Nuclear Repository

Key Concepts for the Geologic Disposal of Nuclear Waste

- Geologic isolation
- Favorable thermodynamics
  - Reducing conditions
  - Reactive redox control
- Cost is an issue
- Favorable Local politics

Microbial activity can influence both the near-field and far-field in repository performance

WIPP TRU Waste Repository



My Background/Experience

- Little/no direct experience with ThermoChimie
- Focus on brine systems
- ~ 20 years of operating a repository where the safety case is constantly challenged by the regulatory process

UNCLASSIFIED

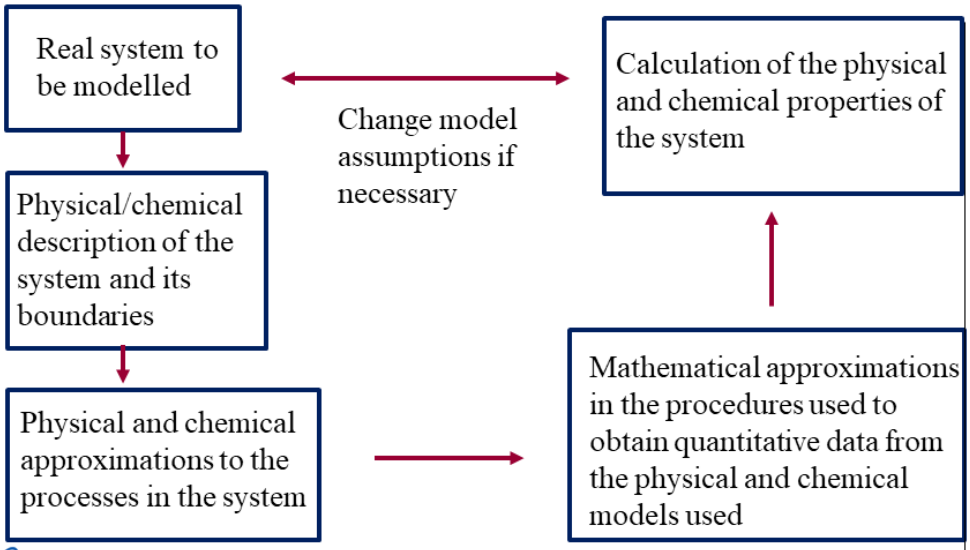
Are Current Thermodynamic Data Bases Sufficient?

Current repository and site specific thermodynamic data applications rely on estimations, approximations and simplifications done in a way to maintain conservatism with the model.

- ❑ Simplification is easier to implement and more explainable (therefore better defended) in a regulatory environment
- ❑ Circumvents data quality issues (that prevent its inclusion in the NEA databases)
- ❑ Account for site-specific gaps in data
- ❑ There are specific processes that are difficult to describe in thermodynamic terms

UNCLASSIFIED

Comparison of the properties of the real system with those of the model



UNCLASSIFIED

Comparison of Laboratory vs. Real Systems NEA view

Laboratory system

- Well defined, with few variables defining the system and its surroundings.
- Good control of system variables.
- The laboratory system is idealized and used to obtain quantitative information that provides the basis for the testing of different scientific theories

Provide mechanistic insight. The high accuracy of the quantitative information can rarely be exploited in the descriptions of real systems

Use model data to obtain ideas for new laboratory experiments that might provide new information on mechanisms

System in nature

- Often ill-defined and requires models with many variables for its description.
- In general no control of system variables.
- The system features may often be described equally well by different models.
- The system properties may give ideas for laboratory studies aimed at identifying new phenomena and to develop the existing system theoretical framework.

Example 1: Role and Importance of Complex Brine Studies

- WIPP has always had a parallel simplified and complex brine experimental approach:
  - Simplified systems: develop/measure thermodynamic data
  - Complex (simulated) and real-system brines: confirm or validate model predictions
  - There are established differences between simplified, simulated/complex and real-system studies
- Simulated/complex groundwater studies are a critical compliment to simplified studies in that they help identify key processes or interactions that may be missing from the model.

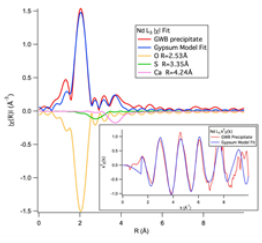
Benefits of Site-Specific Data in Simulated Brine Systems

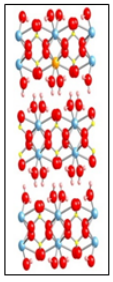
- **Key dependencies are established: broad range pH response (e.g., amphotericity), absence/presence of phases/dependencies**
- **Minor Brine Constituents Can Influence the Actinide Chemistry (argument for simulated vs. simplified brine approach)**

Borate: An (III, V and VI) complexation

Bromide: Key in radiolytic pathways, Br substituted green rust formation observed

Sulfate:





Nd substituted Gypsum

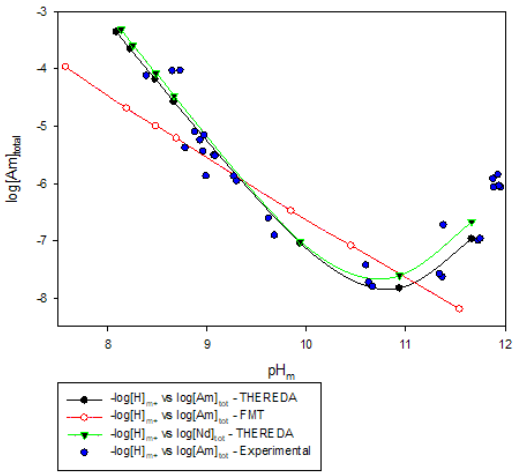
Example 2: Amorphous vs. Crystalline An(III) Phases in Solubility Calculations

- Thermodynamic data is most often linked to crystalline phase formation, although the NEA-TDB recognizes some defined amorphous phases
- Laboratory experiments do not always establish the formation of crystalline phases in simplified systems
- In complex systems, amorphous phases are often the rule rather than the exception
- Defensibility (driven by the regulatory side of the safety case argument), will require the higher-solubility realization

Model Fit Comparison: Thereda vs. FMT (WIPP) model

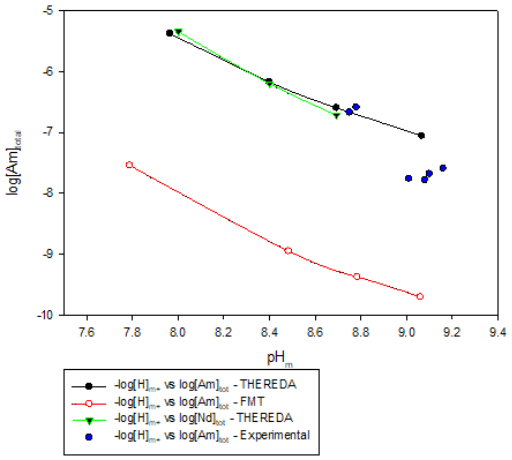
3.86 m $CaCl_2$

- Wrong dependency noted
- Agreement, however, at the "predicted" repository pH (~ 9.4)



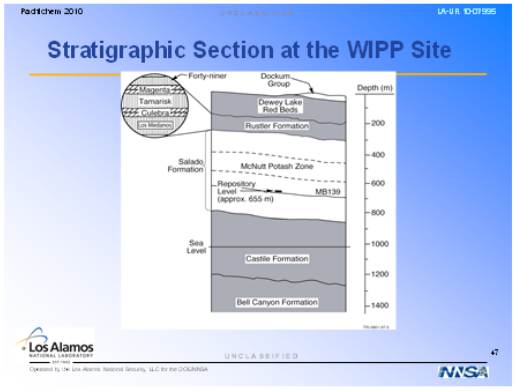
Model Fit Comparison: Thereda vs. FMT (WIPP) model

- Choice of amorphous or crystalline frame of reference leads to large discrepancy
- For high I systems, this will also distort ionic strength corrections used



Regulatory Reality vs. Predicted Repository Behavior

- The regulatory process will impose a repository system "reality" that could be outside of what is realistically predicted:
  - This is driven by defensibility and conservatism arguments
  - Often/usually there are unintended consequences
  - This can/will drive model development and research focus



Impacts of the "Microbial Assumption"

To suppress methanogenesis as an option in the repository (which impacts the availability of biogenic carbon dioxide) – the project was directed to assume infinite availability of sulfate (so also calcium) from the interbed anhydrite:

- It has been since shown that methanogenesis was never an option for WIPP conditions (so this assumption is not needed)
- This effectively pumps a lot of sulfate and calcium into the repository with unintended consequences:
  - Brine chemistry becomes distorted
  - Sulfate leads to significantly higher pH
  - Calcium leads to strong competition with organic/inorganic ligand to impact actinide solubilities
  - Solution carbonate levels are decreased
- Significant shift in model focus and research was required:
  - Calcium complexation data as f(I) with organics
  - Solid phase focus on oxyhydroxides away from carbonates

Key Geochemical systems and benchmarking focus

Geochemical modeling goal in repository applications - Provide a defensible and explainable (not necessarily accurate) description of the radionuclide source term:

- Ideally it will exhibit the right dependency on key parameters as repository conditions change
- It should account for identifiable processes and interactions
- Strategically simpler speciation schemes are easier to defend/explain

Focus of benchmarking exercises should be varied:

- Model-specific exercises with simple well defined systems to establish accuracy (e.g. comparisons with Thereda on I dependencies; model predictions of experimental results in simplified systems)
- Challenge the completeness of the model by predicting speciation/solubility in complex/real systems applications.

Role of Uncertainty: WIPP Example

- For the WIPP model, we do not calculate uncertainty based on the inherent uncertainty in the database:
  - This is actually a limitation as it leads to unrealistic predictions (e.g., six significant figure pH values, etc.) and understanding
  - There should be advantages to ThermoChimie if this capability is present
- Uncertainty is addressed by sampling within a distribution assigned to each solubility that defines the source term
  - This is done by the application of a selection criteria to all published data for a broad range of possible repository conditions – this in practice is a very subjective process that is not satisfactory
  - Uncertainties assigned in this way are very large (in many cases not realistic)
  - The underlying driver for this approach is that the key contributor to uncertainty is not what is in the data base, but what is perhaps missed

Recommendations and Final Thoughts

- Benchmarking is a critical aspect of model development to assure the public/regulator that you have a good understanding of the chemistry and associated dependencies that impact or could impact your source term for the range of expected repository conditions
- In my opinion, the greatest value are challenges to the model predictions by performing site-specific complex-system studies and/or predicting known/established site-specific chemistry. These will validate the robustness of the model or identify key gaps or dependencies that are still missed.
- There is value, but much less value to me, in model-model predictive comparisons. These are needed to show that the mechanics of the modeling approach work, but they will not fully identify key gaps and limitations of the model