The Role of Metrology and Statistics in Making Geochemical Modeling Defensible

G. Meinrath[1], B. Merkel[2] , P. Schneider[3], B. Delakowitz[4]

[1] Schießstattweg 3A, 94032 Passau
[2] TU Bergakademie Freiberg, Institute for Geology, 09599 Freiberg
[3] Hydropiewak GmbH, Oberfrohnaer Str. 84 D-09117 Chemnitz
[4] Hochschule Zittau/Gölitz, Postfach 261, 02755 Zittau

Abstract. Geochemical transport modeling of contaminants in the aquatic environment is an important tool supporting decision-making on both the technical and political level. There is a specific meaning for modeling emissions of radioactive contaminants from uranium mining and processing sites as well as from nuclear installations and waste repositories due to the longevity and toxicity of some radionuclides. The complexity of issues studied by applying reactive transport modeling is steadily growing with the development of hard- and software. Reliability of model output is of crucial importance. Reliability of a model output is intimately related to the quality of the model's input data. Thus, a quantitative measure for data quality and quality assurance becomes essential in addition to transparency and documentation.

Introduction

Geochemical modeling is a genuinely interdisciplinary field. Hydrogeology, chemistry, geochemistry, water chemistry, rheology, information science, numerical mathematics and others more contribute to its development. All contributing factors cause uncertainties in the total process. In the present discussion, concentration will be on the contributions from chemistry. The past two decades of impetuous growth in CPU clock speed and digital storage capacity have raised expectations to model and to predict complex scenarios on large time scales. Modeling codes of considerable complexity have been created raising the expectation that predictions become the more reliable the more data and details can be included into the model (Voss 1998; Prado et al. 1999; Ekberg 2000). It is meanwhile understood that this expectation has ignored several limitations in the accuracy and precision of both the input data to a model and the modeling algorithms themselves as well as our limitations in the perception of the complex processes, features and events that in total make up the real world surrounding us. Hence, information is necessary to quantify the limitations of our understanding and insight into natural processes and the rules governing them.

Thermodynamic data as results of analytical measurements

Prediction of future processes on basis of quantitative information requires a causal relationship between causes and their effects. The relationship not necessarily needs to be strictly deterministic. To take an example, metabolic processes are quite difficult to take into account. Geochemical reactive transport modeling is commonly based on thermodynamic data of those reactions considered to be relevant. During the past century a larger amount of thermodynamic data of metal ions in a wide range of aqueous solutions has been forwarded. These data are commonly collated into collections, either in printed form, on-line data base or code-specific thermodynamic data base for speciation codes (Hefter 1979; May and Murray 1991; Parkhurst 1995).

The fundamental relationship Eq. 1 links chemical reactions with thermodynamics:

$$\ln K_r' = -\frac{\Delta G_r'}{RT} \tag{1}$$

K_r' : formation constant of a reaction r
$\Delta G_r'$: Gibbs' free energy involved in reaction r
R : gas constant
T : absolute temperature

Thus, formation constant K_r' may be understood as a constant of nature under given conditions specified by the superscript ('). There are numerous ways of evaluating either $\Delta G_r'$ or K_r'. Thus, the formation constant K_r' represents an energy difference. This energy difference manifests itself in often quite sensitive equilibria of different chemical forms of the constituents of interest. The evaluation of the energy almost always must rely on analytical chemical data of the concentrations of these different chemical forms. Almost always a certain amount of the chemical forms cannot be analyzed by direct measurement but has to be estimated on basis of mass-balance considerations and other general assumptions.

Since thermodynamic data are a basis of predictive reactive transport modeling and obtained by analytical measurements, these data are affected by measurement uncertainty. Measurement results without statements of reliability (i.e. uncertainty) should not be taken seriously as soon as important decisions are based on these data. The major reason is that such data cannot be trusted because the analyst forwarding these data has not considered whether his approach and procedure is itself trustworthy (de Bièvre 1998).

Measurements in chemistry, responsibility and metrology

Chemistry has performed and interpreted analytical measurements for some centuries. Great chemists have moved chemical science by personal experience, hard work, talent and a considerable portion of try-and-error into the 21st century. The science of analytical chemistry has developed over this period and provided remarkable successes. It is well possible to enjoy the results of these successes as interesting and entertaining, as a kind of social activity. However, the successes of analytical chemistry include an important factor: responsibility. In forwarding a result of analytical measurement, the analyst takes responsibility for the conclusions drawn on basis of this measurement result. Chemical measurements of various kinds are playing a rapidly expanding role in modern society and increasingly form the basis of important decisions. Chemical measurements form a basis for agreement in situations of discordant interests and even distrust. Thus, the role of analytical chemistry vastly exceeds just satisfying scientific curiosity. Global trade, food safety, clinical chemistry, consumer protection and product quality control are only a few fields where important decisions are based on chemical measurements. The consequences from the results of these measurements can affect the lives of many people - not only today but also in the future. Hence, the science of measurement has become a national sovereignty, institutionalized in the national metrological institutes (NMI's). The science of measurement is subject to international treaties like the Meter Convention (1875) as well as the establishment of the International Organization of Legal Metrology (1955). Hence, metrology is not about esoteric, academic measurements of the highest precision performed in an ivory tower at universities or the NMIs. Metrological rules are driven by the discordant interests of the global trade partners where high quality products should be imported cheaply (if at all) while the own products should be exported with high profits (at best without any limitations). It is evident that promotion of science has no priority in the development of metrology. It is, on the other hand, evident that modern science likewise has developed into a business where competition for funding and publication follows rules not completely different from economic markets. It even may be suspected that political systems could have encouraged experimenters to obtain data to be favorable for a certain politically desirable decision. Measurement values complying with metrological rules in a stable metrological framework will keep their significance over long time periods. Metrology therefore is an important element of sustainability.

Comparability of measurements - the international framework

In order to achieve comparability of measurements on a world-wide level, not only objective criteria and an international agreed system of references must exist. An infrastructure of closely cooperating institutions must exist that ensures and controls the conformity of measurements with criteria and points of reference. The System of Units (système international; SI) represents implements this basis, and by use of traceable measurements provides an international infrastructure for comparable measurements. The institutional infrastructure is realized by the Meter Convention. Even closer ties were formed in 1999 with the Mutual Recognition Agreement (MRA) signed by the directors of 38 member states of the Meter Convention. The objective of the MRA are to: establish the degree of equivalence of measurement standards by NMIs; to provide for the mutual recognition of calibration and measurement certificates issued by NMIs; thereby to provide governments and other parties with a secure technical foundation for wider agreement related to international trade, commerce and regulatory affaires (Wielgosz 2002).

Deficits of existing thermodynamic data

Existing deficits in most contemporary chemical measurements have been discussed elsewhere in some more detail (Chalmers 1993; Thompson 1994; de Bievre 1997; Meinrath and Spitzer 2000). Likewise, fundamental principles of metrology for thermodynamic data as well as concepts for obtaining and applying thermodynamic data by metrological rules have been forwarded (Meinrath et al. 2000a; May and Murray 2001; Ekberg et al. 2002; Meinrath and May 2002).

Serious discrepancies between published thermodynamic parameters of chemical reactions are well-known. Since there are many different causes of these problems (such as experimental error, inadequate theory, and carelessness), they can be very difficult to pinpoint and to eliminate. The situation is made worse because many thermodynamic data persisting in the scientific literature stem from values that are later corrected or become experimentally superseded (May and Murray 2001).

The existing deficits in thermodynamic data of reactions relevant for reactive geochemical transport modeling can be separated in three groups:

- Documentation and communication deficits
- Competence deficits
- Consistency deficits

Documentation and communication deficits mainly limit a critical reassessment of thermodynamic data. Thus, despite the fact that many collections of thermodynamic data are termed 'critical', there is no support for such a claim other than the expectation that the reviewer(s) recognize(s) a 'good' data if he/they see(s) it. The almost complete lack of statistical assessment (including essential issues like the detection and treatment of extraneous data, the discussion of the optimization criterion and the reliability of auxiliary data) of thermodynamic data precludes any post-evaluation. At best, obvious severe errors in the experimental conditions and evaluation of the data can be indicated.

Competence deficits direct to the experimenter performing the experimental work. Not seldom, these are comparatively unexperienced novices to the science of chemical analysis performing the measurements as a part of their academic education.

Next to experimental errors, carelessness and inadequate theory the -sometimes undocumented- use of auxiliary data, the use of different calibration methods and differences in the calibration standards and sometimes even the different equipment generates the well-known differences in chemical thermodynamic data obtained by the nominally same experimental method for the same reaction. Inadequacies in the theory of electrolyte solutions has resulted in a larger number of procedures for extrapolation of experimentally determined formation constants of a reaction to standard conditions, thus adding to the inconsistency of available thermodynamic data.

The role of metrology and statistics

Evaluation and documentation of experimental data serving as a basis of important decisions (= decisions affecting other people) must allow a reassessment by other people. This situation is most obvious if the decision is questioned in a law court. Such a reassessment needs several elements. A primary requirement is a precisely defined terminology. There is an essential difference in using well-defined terms like reproducibility and repeatability (VIM 1994) or undefined terms like validation and standard (De Bievre 1996; 1997a). To provide definitions and to set up criteria for their use is the task of metrology. Metrology is the science of measurement. Thus, it deals with values resulting from measurements.

The requirement of traceable values results both from the request for comparability with related data from sources located at other places but even more from the possible impact of current day decisions on future societies, their socio-economic development and their competitiveness under their living conditions. Once a future society has the impression that the data on which a past decision with long-term effect has been based on (i.e. the construction of a repository for nuclear wastes) have not been interpreted fairly, the affected area may experience drastic losses i.e. in the economic competition, loss of population. The affected society even may have to invest a considerable amount of its GNP to reverse such a past decision.

Statistics is an essential part of metrology. Statistics is searching for patterns in data and tries to find characteristics in groups of data. The major role of statistics is to transfer the message to other people that the data has been treated fairly (Efron 1981). Most people are not natural-born statisticians. Left to their own, human beings are not very good at picking out patterns from a sea of noisy data. To put it the other way, we are all to good at picking out non-existent patterns that happen to suit our purposes and/or prejudices (Efron and Tibshirani 1993). In the past, a major task for statisticians has been to find mathematical approximations in analyzing a given data set in order to make the statistical task tractable. Thus, parametric statistics on basis of some distributions, e.g. normal, Poisson or Student distributions has been common. Nowadays, modern computer-based statistics allow focusing on the question of interest instead of mathematical tractability (Efron and Tibshirani 1991). Complex datasets commonly occurring in chemistry, including chemical thermodynamics, can be treated (Meinrath et al. 2000a; Meinrath 2000; Ekberg et al. 2002a; Meinrath and Lis 2002).

International requirements for comparable data

From the preceding discussion, it is clear that existing thermodynamic data cannot serve as a reliable basis for predictive geochemical modeling because neither the documentation nor the evaluation procedures of individual data are sufficient. A simple value, in some cases even associated with a statement of uncertainty of unknown meaning (standard deviation? marginal confidence interval? Uncertainty limit? Including correlation effects? Considering uncertainty contributions from auxiliary parameters?) is not sufficient for quality assessment.

The enormous and still increasing relevance of chemical analytical data in many areas of daily life has created a demand for a world-wide system of references. The establishment of such a reference system is not without its own difficulties and quite fundamental problems are discussed among metrology scholars. But these aspects are beyond the scope of this discussion. A basis of the international metrological is the document 'Guide to the Expression of Uncertainty in Measurement' (ISO 1993). A major element in this document is the total uncertainty budget as an essential prerequisite for comparability of measurement results inside and outside chemistry. The total uncertainty budget requires the quantitative estimation and inclusion of <u>all</u> contributions of uncertainty to a measurement result. For chemical thermodynamic data, this requirement poses quite a task to the experimenter. It also explains why existing thermodynamic data cannot be comparable. There is no argument what-so-ever that allows to extract reliable values from the heap of doubtful data available in current literature.

It must be stated that at least for the thermodynamic data obtained in pre-1980ies, geochemical modeling was an application out of the scope for an application of chemical thermodynamic data. However, by ignoring important uncertainty contributions in the evaluation of thermodynamic data, the experimenters have invited their own problems. By underestimating the variability of the data forwarded as a result of a metrologically insufficient evaluation process, the differences between data from different sources became significant. Thus they had to be considered as discrepancies (de Bièvre 1996a). The total uncertainty budget, required by metrology as an essential part of the traceability concept, will in most cases lead to larger uncertainties in chemical thermodynamic data. However, some apparent discrepancies will vanish and the real problems will stick out. Thus, the risk to stick with problems caused by seemingly discrepant data will be reduced. Consequently, action to resolve the actual problems will become more efficient. The evaluation of a total uncertainty budget for complex situations is still an open field of discussion. It is clear that such protocols must serve metrological goals: traceability to reference standards and comparability by a reasonable and realistic estimate of a complete uncertainty budget for a given value. Both overinterpretation and underestimation must be avoided. There is no use to forward a value where a small uncertainty is assigned just in order to stick out among a larger number of preceding publications. Since we do not know the motivations having caused the reported reliability estimates in available literature (in the absence of the comparability goal), discarding of these data is possible only in case of obvious errors (Meinrath et al. 1999).

References

Chalmers R.A. (1993) Space age analysis. Talanta 40: 121 - 126.

De Bièvre P. (1996) Where pragmatism and clarity go on collision course. Accred. Qual. Assur. 1: 159.

De Bièvre P. (1996a) Do measurement scientists sometimes invite their own problems? Accred. Qual. Assur. 1: 91.

De Bièvre P. (1997) Measurement results without statements of reliability should not be taken seriously. Accred. Qual. Assur. 2: 269.

De Bièvre P. (1997a) Validation, valid measurement and more about validity. Accred. Qual. Assur. 2: 55.

De Bièvre P. (1998) Uncertainty assessment is an evaluation process. Accred. Qual. Assur. 3: 391.

Efron B. (1981) Nonparametric estimates of standard error: the jackknife, the bootstrap and other methods. 68: 589 - 599.

Efron B, Tibshirani R. (1991) Statistical data analysis in the computer age. Science 253: 390 - 395.

Efron B, Tibshirani R.J. (1993) An introduction to the bootstrap. Monographs on statistics and applied probability 57. Chapman & Hall. London/UK.

Ekberg C., Meinrath G., Strömberg B. (2002) A retraceable method to assess uncertainties in solubility estimations exemplified by a few Americium solids. submitted.

Ekberg C, Börjesson S, Emrén A.T., Samuelsson A (2000a) MINVAR and UNCCON, computer programs for uncertainty analysis of solubility calculations in geological systems. Comput. Geosci. 26: 219 - 226.

Ekberg C, Ödegaard-Jensen A., Meinrath G. (2002) LJUNGSKILE - a computer program for investigation of uncertainties in chemical speciation. REPORT SKI-R--02/xxx. Statens Kärnkraftinspektion. Stockholm/S.

Hefter G (1979) Critical survey of stability constants and related thermodynamic data of fluoride complexes in aqueous solution. IUPAC. Pergamon Press/Oxford.

ISO (1993) Guide to the expression of uncertainty in measurement. International Standard Organistation. Geneva/CH.
May P.M., Murray K. (1991) JESS, a joint expert speciation system I: raison d'être. Talanta 38: 1409 - 1417.
May P.M., Murray K. (2001) Database of chemical reactions designed to achieve thermodynamic consistency automatically. J. Chem. Eng. Data 46: 1035 - 1040.
Meinrath G (2000) Computer-intensive methods for uncertainty estimation in complex situations. Chemom. Intell. Lab. Syst. 51: 175 - 187.
Meinrath G. (2000a) Comparability of thermodynamic data - a metrological point of view. Fresenius J. Anal. Chem. 368: 574 - 584.
Meinrath G (2001) Measurement uncertainty of thermodynamic data. Fresenius J. Anal. Chem. 369: 690 - 697.
Meinrath G., Spitzer P. (2000) Uncertainties in determination of pH. Mikrochim. Acta 135: 155 - 168.
Meinrath G, Lis S (2002) Application of cause-and-effect diagrams to the interpretation of UV-Vis spectroscopic data. Anal. Bioanal. Chem. 372: 333 - 340.
Meinrath G., May P.M. (2002) Thermodynamic Prediction in the mine water environment. Mine Water Environ. 21: 24 - 35.
Meinrath G., Kato Y., Kimura T., Yoshida Z. (1999) Comparative analysis of actinide(VI) carbonate complexation by Monte Carlo resampling methods. Radiochim. Acta 84: 21 - 29.
Meinrath G, Hurst S, Gatzweiler R (2000) Aggravation of licensing procedures by doubtful thermodynamic data. Fresenius J. Anal. Chem. 368: 561 - 566.
Meinrath G, Ekberg C, Landgren A, Liljenzin JO (2000a) Assessment of Uncertainty in Parameter Evaluation and Prediction. Talanta 51: 231 - 246.
Nitzsche O, Meinrath G, Merkel B (2000) Database Uncertainty as a limiting factor in reactive transport prognosis. J. Contam. Hydrol. 44: 223 - 237.
Parkhurst D.L. (1995) User's guide to PHREEQC-a computer program for speciation, reaction, reaction-path, advective transport, and inverse geochemical calculations. Water Resources Investigations Report 95-4277. USGS Lakewood/USA.
Prado P., Draper D., Saltelli S.A., Pereira A., Mendes B., Eguilior S., Cheal R., Tarantola S. (1999) Gesamac: conceptual and computational tools to tackle the long-term risk from nuclear waste disposal in the geosphere. REPORT EUR 19113. European Communities, Luxembourg.
Thompson M. (1994) Statistics - the curse of the analytical classes. Analyst 119: 127N.
VIM (1994) Intenational Vocabulary of Basic and General terms in Metrology. Deutsches Institut für Normung e.V. Beuth Verlag Berlin/FRG
Voss C.I. (1998) Hydrogeol J 6: A4 - A6.
Wielgosz R.I. (2002) International Comparability of chemical measurements. In: Importance of traceable pH measurements in science and technology (Spitzer, Meinrath, eds.). REPORT PTB (xx-2002) Physikalisch-Technische Bundesanstalt Braunschweig, Braunschweig/FRG.