

Metrological View of Probabilistic Modelling of Trace Elements Behaviour in Aqueous Environments

Günther Meinrath

RER Consultants, Schiessstattweg 3a, D-94032 Passau

Technische Universität Bergakademie Freiberg, Institut für Geologie, D-09596 Freiberg

Technische Universität Bergakademie Freiberg, Institut für Anorganische Chemie, D-09596 Freiberg

Numerical modeling of trace component in waters is discussed under a metrological point of view. Some basic metrological concepts are presented and justified. The current problems of analytical chemistry in adopting metrological concepts are illustrated and the consequences for thermodynamic data collections outlined. The importance of meaningful measurement uncertainties for thermodynamic quantities are demonstrated by applying the LJUNGSKILE code for investigation of uncertainties in chemical speciation to some aspects of acid mine drainage.

1 Introduction

The investigation of trace element behaviour in natural aqueous systems is an important sector of environmental study. The distribution and fluxes of minor elements may serve as indicators for physicochemical and biological processes in nature. The enormous complexity of most natural aqueous systems requires sensitive and selective analytical techniques for qualitative and quantitative determination of the constituent(s) of interest. Modern instrumental analysis has a broad range of techniques at hand from single atom detection to coupled speciation techniques for inorganic and organic materials analysis including biosubstances. The versatility of the instruments is enhanced by powerful chemometric data analysis techniques. Bioanalytical methods, i.e. immunoassays and biosensors, further extend the toolbox of analytical chemistry. The enormous progress in computing power is one of the major driving forces behind these developments. Impressive as the progress is, there are considerable caveats and warnings that, however, have their difficulties to reach the attention of those involved in the application of analytical methods and techniques.

The quantitative description of trace element behaviour is intimately connected with analytical chemistry. Either the site specific chemical input data of a model have been obtained by chemical analysis or the geochemical database has been obtained from chemical thermodynamic studies

which deeply rely on analytical techniques. It should be reminded that the thermodynamic constants describing ion interaction in aqueous solutions result from the quantitative determination of often very sensitive solution equilibria.

For the applicant of chemical information, e.g. thermodynamic or site-specific data in reactive transport modeling, follows that the reliability of chemical data is much less than anticipated. Since quality control measures for chemical data came under discussion only recently no straightforward recommendation how to use these chemical data can be given.

2 Some Metrology Basics

Metrology is the science of measurement. A measurement is a comparison. Since there are no comparisons without imperfectly known errors, the significance which one can give to the measurement must take account of this uncertainty. Two independent measurements can be comparable only if they are both related to the same reference. Because measurements play an important role in daily life, the importance of such fundamental references has been recognised early. Today a system of seven fundamental units (metre, kilogram, second, ampere, kelvin, mole and candela) form the basis of système international d'unités (SI). The SI is backed by international treaties (convention du mètre) and subsequent international agreements, e.g. the reconnaissance mutuelle (Mutual Recognition Agree-

ment; MRA). Ideally, all comparisons should be made to these basic units, thus warranting traceability. Traceability is the property of the result of a measurement or the value of a standard whereby it can be related to stated references, usually national or international standards, through an unbroken chain of comparisons all having stated uncertainties (VIM 1994). Practically, there exists a hierarchy of standards running from the basic units as realized by the national metrological institutions (NMI) down to the working standard materials applied in the laboratories of research institutes and industry. The bureau international des poids and mesures (BIPM) near Paris/France warrants the comparability of the realisations in different NMI. Thus the very basic uncertainty present in each laboratory measurement results from the comparisons of the different references with the higher standard (national standard, calibration standard, working standard; WIELGOSZ 2002).

The SI is strongly directed to the physical world. Chemical measurements have become a subject of metrology only in 1972 with the adoption of the mol as a SI unit. The quantity of which the mol is the basic unit is 'amount of substance'. Most chemists will never have heard about this quantity neither use this term during their professional lifetime. The quantity of which mol is the unit is still under considerable discussion. To take an example, the term 'psammity' was proposed (KELL 1977) but not adopted. Another suggestion is 'Dalton'. One of the major problems is the fact that it is not possible to state the mass of an atom in terms of the number that appears on the periodic table and in SI units. There is no feasible method to determine the mass of a single atom with sufficient precision to define $1 \text{ u} = m(^{12}\text{C})/12 = 1.\text{xyz}... \cdot 10^{-27} \text{ kg}$. One is forced to use 'a non-SI unit accepted for use with SI, the 'unified atomic mass unit' (u) defined as $1 \text{ u} = m(^{12}\text{C})/12$ or to resort to the 'relative atomic mass' nomenclature. In addition, chemists measure concentration in units of mole per liter. Liter is a derived unit. In terms of SI 'the word liter may be employed as a special name for the cubic decimeter...(but that) liter should not be employed (to describe) high accuracy volume measurements'. Thus the status of the liter is ambiguous and, perhaps, even confusing. Worse, some textbooks use M to represent the unit mol/L. Sometimes M is termed molarity, sometimes molar concentration. IUPAC (International Union of Pure and Applied Chemistry) holds position that the adjective molar, when applied to

extensive quantities, signifies division of that quantity by chemical amount (n) expressed in moles. The quantity expressed in mol/L is to be called 'amount of substance concentration'. However, the Clinical Chemistry Division of IUPAC suggests 'substance concentration'. Instead of M, the IUPAC recommends the use of 'c' to designate mol/L.

3 Accuracy in Analytical Chemistry

In an editorial (DE BIÈVRE 1999), Prof. de Bièvre described the presentation of an ICP-mass spectrometer. The device came with a software transforming the commonly single line in the mass spectrum into a three-dimensional graphics. The three-dimensional graphics made it possible for the analyst to vary the measurement parameters during a measurement and to monitor the consequences of his changes real-time in the graphics display. The manufacturer made considerable efforts to communicate how to optimize by this way the repeatability by using the relative standard deviation of the three-dimensional signal as a criterion. The manufacturer demonstrated this procedure using several lines as examples selling to the audience that the best value for the measured quantity would be the value with the best repeatability. To reach this goal there was no hesitation to modify the the measurement parameters while going from one isotope to the next. As a consequence, different isotopes in the same sample were measured under different measurement conditions. The results, however, were finally combined. The optimisation of signal heights on basis of repeatabilities is irritating. De Bièvre however described that most potential clients felt thrilled by the idea to operate such a machine in the way demonstrated by the manufacturer.

It would be short-sighted to blame the manufacturer for this style of presentation of a complicated measurement instrument. Manufacturers are rarely scientists but merchants. They want to sell – to do so they are oriented to the requests and desires of their clients. The clients have to make clear to the manufacturers that they are only willing to buy equipment that is easy to handle, transparent in its construction, aiming at repeatable measurement results under comparable conditions of measurement – instead of transferring the experience to the manufacturers that black boxes including complicated data treatment gadgets sell nicely.

The dilemma described by de Bièvre isn't a recent one (CHALMERS 1993). In the early times of analytical chemistry the need to differentiate between precision, accuracy and repeatability has been emphasised eagerly. "Das Ziel eines Analytikers darf nicht sein, Resultate zu erzielen die absolut genau sind – welche meiner Überzeugung nach nur durch Zufall erreicht werden können – sondern dem genauen Werte so nahe zu kommen wie es chemische Analytik möglich machen kann" (Berzelius 1779 – 1848) and "Ein Chemiker, der nicht einen Eid darauf ablegen würde, dass seine Arbeit sorgfältig und genau ist sollte seine Ergebnisse niemals an die Öffentlichkeit geben, denn in dem Falle wären diese Ergebnisse von zerstörender Wirkung, nicht nur für ihn selbst, sondern für die gesamte Wissenschaft" (Fresenius 1818 – 1897). The Australian metrologist G. Price calls the procedure described above 'a warranty for the absence of accuracy' (PRICE 2000). There is obviously a problem in analytical chemistry that has gone unnoticed for a considerable time. The enthusiasm with reference to the new possibilities resulting from the instrumental revolution in analytical chemistry has left out of sight the necessity to understand the importance of the accuracy of a value from analytical measurement.

4 Quality of Thermodynamic Data Collections

There is no need for analytical chemistry to feel shame for its past achievements during the past decades. The limits have been pushed enormously. However, the strength of analytical chemistry is not necessarily the quantification of amounts of substances but the proof of identity of molecules. It is almost routine nowadays to analyse the stereochemical structure of a natural compound unambiguously, to synthesise this compound and to prove that natural and synthetic material are identical. There is no need to question analytical chemistry in general. The user of thermodynamic information must keep in mind that determination of thermodynamic data is a complicated and extremely time-consuming process. Before the advent of high-speed desktop computers, thermodynamic data has never been determined with such far-reaching applications in mind as reactive geochemical transport modeling.

The fundamental reason for the collection of thermodynamic databases is the relationship

$$\Delta G^{\circ}_R = -RT \ln K^{\circ}. \quad (1)$$

ΔG°_R is the Gibbs free energy available in a reaction under standard conditions (commonly 298.15 K and 1013 hPa), while K° is the equilibrium constant of the reaction under the same conditions. R is the fundamental gas constant and T the absolute temperature in kelvin. Eq. (1) relates the measurable quantity K° to the fundamental laws of thermodynamics. Hence, a constant K° determined under one condition is valid independently from time and space. In other words, K° is a fundamental law of nature itself. Hence, it is possible to measure a constant K° under convenient conditions in a laboratory and apply the result in a geochemical calculation, e.g. at much lower concentrations and media of vastly different chemical composition. For other processes with potential relevance in natural aqueous systems, e.g. sorption and kinetically controlled processes, no equivalent to eq. (1) exists that would warrant a comparable justification for data transferability. On the contrary, both processes are known to depend sensitively on physico-chemical conditions.

For a long time, eq. (1) has been considered as a justification for the collation of thermodynamical data into so-called 'databases' (better: data collections). Such databases are abundant and may differ considerably. Especially, they do not provide any indication of the reliability of a numerical value, e.g. by stating an uncertainty. In most but the most important cases (e.g. dissociation constants of CO_2 in aqueous solutions, solubility of CaCO_3 or the autoprotolysis constant of H_2O) the amount of data is too small for a proper statistical treatment. In addition, evaluation process of thermodynamic data have been much too complicated even for an approximate statistical analysis of the measurement process. Hence, the main justification of a collection of a thermodynamic data is the deep belief of the compilers to recognise 'good' data if seeing it... (MEINRATH & LIS 2002).

Nowadays the situation has improved fundamentally. ISO has issued the Guide to the Expression of Uncertainty in Measurement (GUM 1993). This guide responds to the urgent demand of general rules for the assessment of data reliability valid in all fields of technology and science. The major NMI have established departments for 'Metrology in Chemistry'. EURACHEM, the organisation of European analytical laboratories, has issued a guideline for applying the GUM rules to analytical chemistry (EURACHEM 2000). Modern statistical tools, e.g. computer-intensive methods of statistical analysis like the bootstrap

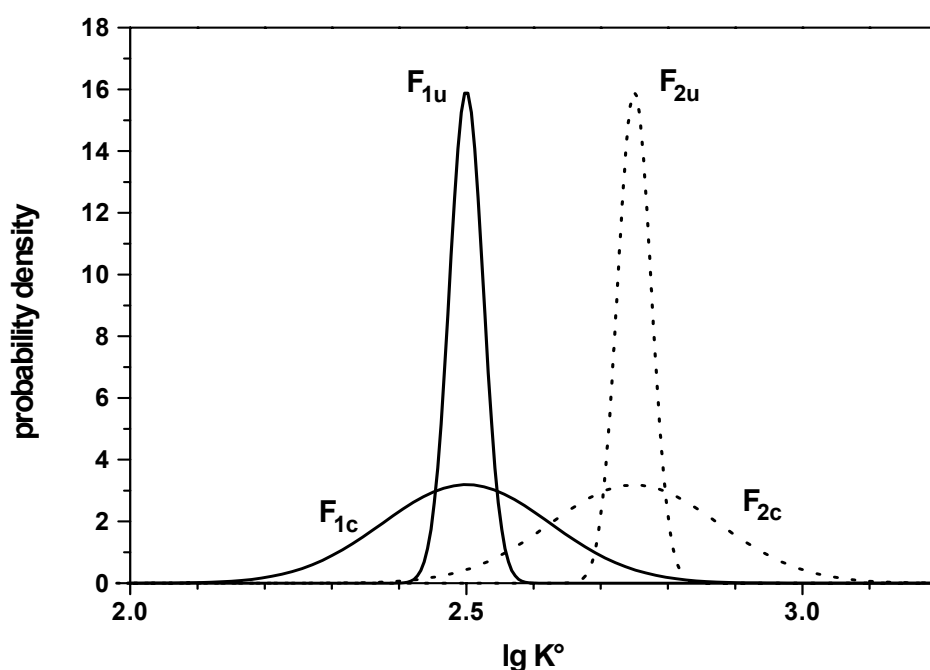


Figure 1: Two researchers (F_1 and F_2) study the same chemical system and report thermodynamic constant $\lg K^\circ$. Both assume a measurement uncertainty of ± 0.05 (.68 percentile) (researcher 1: F_{1u} ; researcher 2: F_{2u}). The probability distributions F_{1u} and F_{2u} do not overlap and

(EFRON & TIBSHIRANI 1997) allow a statistical analysis of the analytical processes from which thermodynamic data are derived (MEINRATH 2001; MEINRATH & LIS 2002; MEINRATH, EKBERG & STRÖMBERG 2003). The procedure to establish a complete measurement uncertainty budget of a thermodynamic value are rather complicated. A completely satisfactory uncertainty budget also needs to include measurement uncertainty of auxiliary data, notably the pH in case of pH-dependent reactions. Assessing the measurement uncertainty of pH is still work in progress (SPITZER & WERNER 2002; MEINRATH & SPITZER 2000). There is still some way to walk.

4.1 Both overinterpretation and underestimation of data reliability invite problems

There is no safe side to walk when it comes to assessment of data quality of thermodynamic data. There are few systems where conditions can be found to obtain information on one species uncorrelated with the other components of the system. Hence, chemical systems are prone

to correlation at different levels (MEINRATH 2000). A fundamental property of each chemical system is the number of species involved in the equilibrium under study. This question must be answered mostly from a limited number of data points more often than not correlated and nonlinear. Overinterpretation of accuracy and precision leaves unexplained variance and causes the introduction of non-existent species. Underestimation risks to ignore relevant but less dominant species. A similar situation exists for the values of the thermodynamic constants themselves. Figures 1/2 represent the both situations schematically. Two researchers are assumed to have determined a formation constant in the same system under comparable conditions. For sake of simplicity it is assumed that the results are free from bias and the uncertainty follows normal distributions. In figure 1, the total uncertainty budget is underestimated constants. The both distribution curves do not overlap – the discrepancy is significant. Tedious redeterminations will be necessary. In figure 2, the situation is reversed; the researchers underestimate the total uncertainty budget. Their distributions overlap significantly indicating agreement in the results.

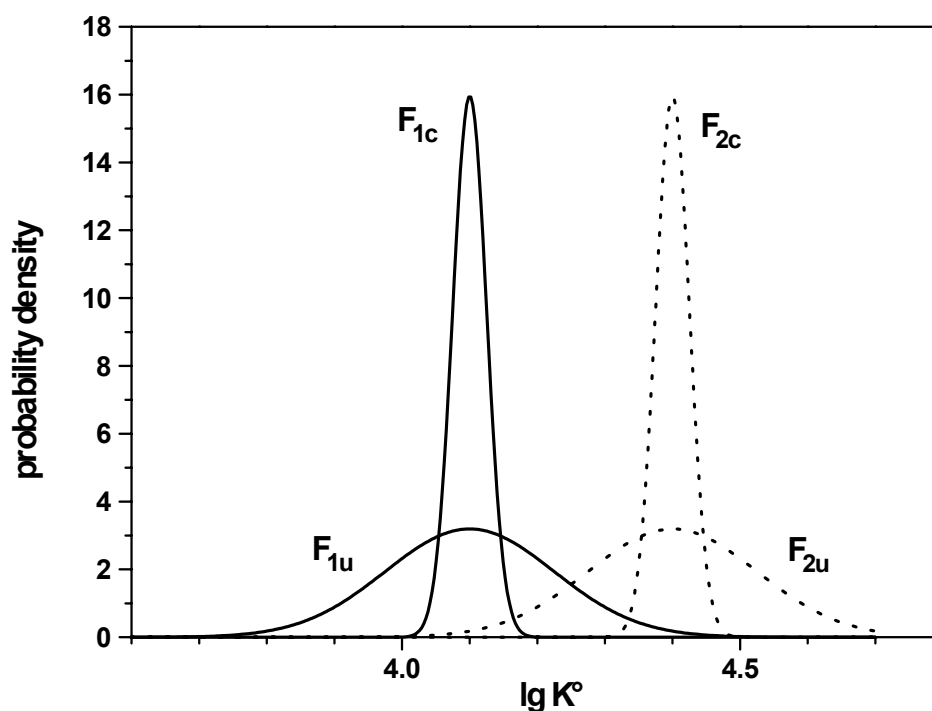


Figure 2: The reverse situation to figure 1 is illustrated. Both researchers overestimate the complete uncertainty budget of their measurements. They conclude reasonable agreement. If, however, the curves F_{1c} and F_{2c} represent the realistic total uncertainty budget, the system is likely to be incorrectly interpreted.

However, in such a situation experimental variance is incorrectly attributed to the species with the formation constant $\lg K^\circ$ that should be attributed to another, probably unnoticed, species.

As a matter of fact there isn't any data collection where these questions have been addressed or even answered. The assessment of uncertainties in chemical measurement is a recent activity. Since progress mostly is understood in doing things better and better, a field of research where the limitations of chemical data (not seldom

highly acclaimed) are investigated does not expect to receive much focus. Its competences require such unpleasant fields as statistics (seen from most chemists as a 'cure' (THOMPSON 1993) and scientific computing. Considering that most chemists never have attended a statistical course beyond the most rudimentary introduction of t-tests and, perhaps, ANOVA, enthusiasm for metrology in chemistry should be expected to be moderate.

Table 1: Selected hydrolysis constants of Fe(III) from different compilations.

$\text{Fe}(\text{OH})^{2+}$	$\text{Fe}(\text{OH})_2^+$	$\text{Fe}(\text{OH})_3^\circ$	$\text{Fe}_2(\text{OH})_2^{4+}$	$\text{Fe}_3(\text{OH})_4^{5+}$	$\text{Fe}(\text{OH})_4^-$	$\text{Fe}(\text{OH})_{3(s)}$	source
61 ^{a)}	9 ^{a)}	2 ^{a)}	20 ^{a)}	5 ^{a)}	3 ^{a)}		JESS
-2.19	-5.67	-12.0	-	-	-21.6		B/M
-2.5	-6.5	-12.0	-3.0	-6.1	-21.6		CIV
-2.19	-5.67	-12.56	-2.95	-6.3	-		PQC
-2.19	-4.59	-	-2.85	-6.3	-		NIST
-2.05	-6.35	-13.45	-2.90	-6.04	-21.43	3.20	JESS ^{b)}

^{a)} number of literature references in the JESS database for a particular species; ^{b)} the weights given in JESS database indicate that only FeOH^{2+} can be considered to be well characterised. B/M: BAES & MESMER 1976; JESS: Joint Expert Speciation System: MAY & MURRAY 1991; 2000; NIST: MARTELL et al. 1993; PQC: PHREEQC Default Database (PARKHURST 1995); CIV: CHEMVAL (1992)

Table 2: Uncertainties applied in calculation of the speciation diagram Fig. 4.

	uncertainty level I	uncertainty level II
FeOH^{2+}	± 0.28	± 1.0
$\text{Fe}(\text{OH})_2^+$	± 0.22	± 0.5
$\text{Fe}(\text{OH})_3^\circ$	± 0.23	± 0.5
$\text{Fe}(\text{OH})_4^-$	± 0.23	± 0.5
$\text{Fe}_2(\text{OH})_2^{2+}$	± 0.2	± 0.5
$\text{Fe}_3(\text{OH})_4^{2+}$	± 0.2	± 0.5
$\text{Fe}(\text{OH})_{3(s)}$	± 0.23	± 0.5

5 Making uncertainties visible: the Ljungskile code

The total uncertainty budget of a complicated quantity like a thermodynamic constant is a somewhat abstract concept. Does it really matter whether a constant is given as, say, $\lg K^\circ = 4.3$ or, say $\lg K^\circ = 4.3 \pm 0.15$? As a matter of fact, there are only very few examples available in literature where the importance of uncertainty in thermodynamic constants has been discussed. Because geochemical modeling involves usually a large number of species, to take uncertainty in

each species into account may become a task difficult to be tackled. As an illustrating example of uncertainty propagation, the probabilistic speciation code Ljungskile will be applied using Fe(III) at trace concentration levels as an example (ÖDEGAARD-JENSEN, MEINRATH & EKBERG 2003).

Iron is an important element in the environment. The Fe(II)/Fe(III) redox equilibrium in combination with the considerable differences in chemical behaviour of the both states causes mineral formation, alteration and dissolution. Iron is a limiting element for biochemical processes (MARTIN & FITZWATER 1988). Iron affects the

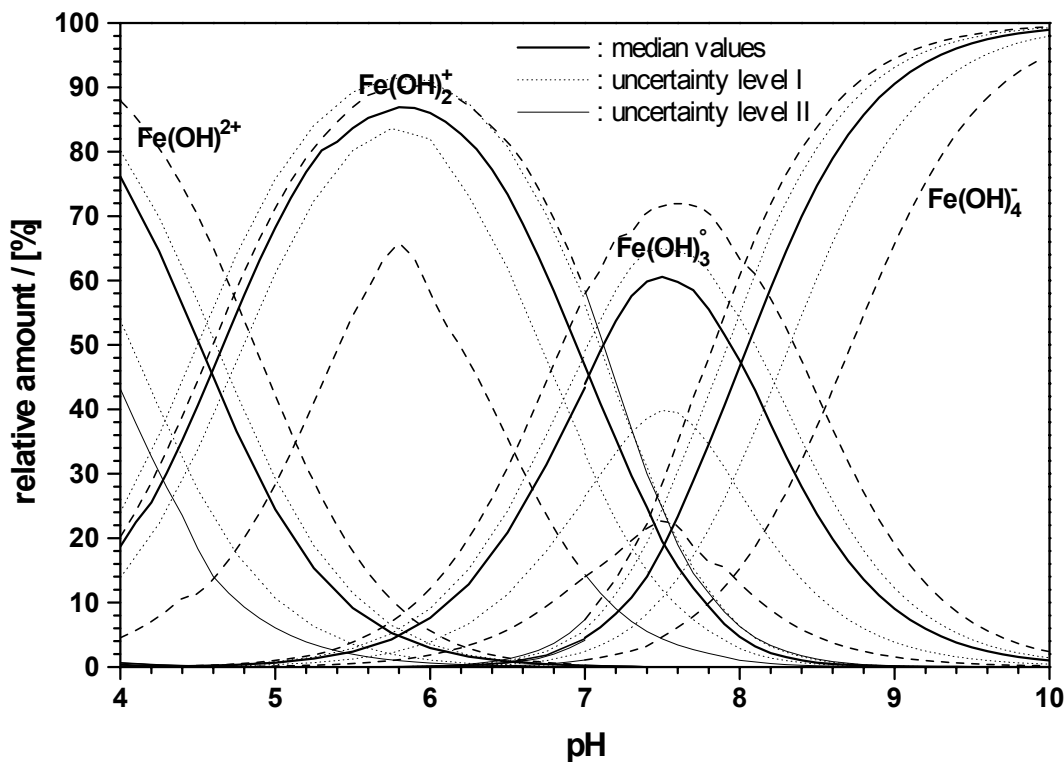


Figure 3: Probable composition of an aqueous solution in equilibrium with $\text{Fe}(\text{OH})_3(s)$ and very low ionic strength ($I < 10^{-3} \text{ mol L}^{-1}$) (LJUNGSKILE simulation). The Fe(III) hydrolysis species are taken into account with formation constants given in last row of Table 1 and two uncertainty levels given in Table 2.

rates of redox processes and the redox state of natural waters (MILLERO, YAO & AICHER 1995). In acid mine drainage (AMD) iron plays an important role as a source of hydronium ions, e.g. from acid producing sulfide minerals. The study of AMD is an important field of research, however, it is often not sufficiently appreciated that the thermodynamic data basis for a modeling of the complex processes in AMD is rather poor. After compilation of the vast majority of reports on the thermodynamics of Fe(III) hydrolysis publicly available, MAY AND MURRAY (1991) conclude that only the formation constant for FeOH^{2+} is well characterised. It is also apparent from Table 1 that the elicitations in the compilations are not independent of each other.

When using mean values of a thermodynamic constant for each species, speciation codes to solve the respective systems of equations are available. But propagating probability density curves through a speciation code is a task running on a different level of complexity. However, in combining non-parametric statistics, efficient algorithms and high-speed computers feasible ways to solve the problem can be found. Such an approach has been implemented in the code LJUNGSKILE 1.0, where stratified sampling strategies are combined with the speciation code PHREEQC to evaluate estimates for speciation variability as a result of uncertainties in the formation constants. In the present implementa-

tion, the species and the uncertainties must be explicitly specified by the user. A solid phase can be selected but an uncertainty for the solubility product cannot yet be specified. Due to the absence of meaningful uncertainties in literature, the uncertainties must be estimated. A calculation for the Fe(III)-H₂O system at $(\text{Fe})_{\text{total}} = 10^{-8} \text{ mol L}^{-1}$ is shown in fig. 3 using two uncertainty levels (.68 percentiles).

It is obvious that the polynuclear species are absent. Provided, the formation constants are not grossly underestimated, they do not play a role in solid-aqueous phase equilibria of Fe(III) at very low ionic strengths. It is also obvious that reporting relative species amounts to the percent level is inadequate. The uncertainties are in the order of ten percent. Both methods advice care in the statement which species is dominating in the region pH 6.5 to pH 8.5. This pH region is typical for natural aqueous systems. Independent whether uncertainty ranges are symmetric or asymmetric, the uncertainties in the relative composition of the solution are in the order 10¹ %. Considering the abundance of mean value-based speciation calculations and diagrams in the literature with figures given to the tenth of a promille and even beyond, it is obvious that in the future diagrams and calculations without stated uncertainty should not be acceptable because the risk of over- and misinterpretation is high.

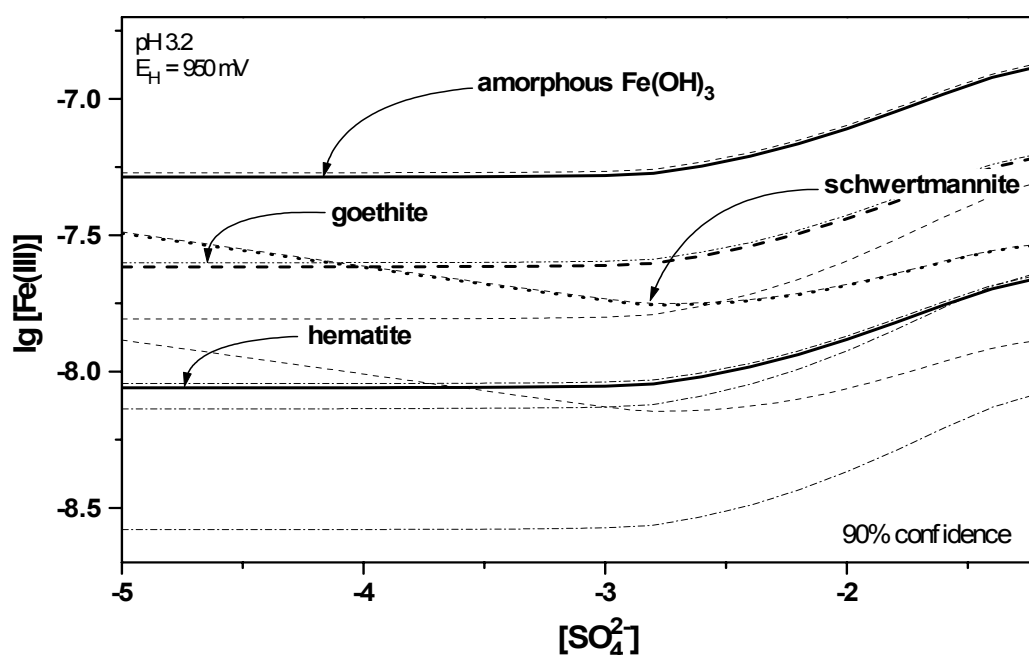


Figure 4: Solubility of amorphous $\text{Fe}(\text{OH})_3$ ($\lg K_{\text{sp}} = 0.89$), goethite ($\lg K_{\text{sp}} = 0.5$), hematite ($\lg K_{\text{sp}} = 0.00$) and schwertmannite ($\text{Fe}_8\text{O}_8(\text{OH})_6\text{SO}_4$; $\lg K_{\text{sp}} = -8$) as a function of the sulfate concentration. Dashed lines give 90% confidence limits.

Figure 4 illustrates another consequence of measurement uncertainty in trace element modeling. Processes in aqueous systems are strongly affected by mineral transformations and alterations. It has occurred that schwertmannite compounds are present in AMD waters. Schwertmannites are sulfate containing iron oxyhydrates. Their formation in high-sulfate AMD waters generates about 2.7 H⁺ ions per formula unit Fe(III). However, the stability region of schwertmannites is badly understood (SCHWERTMANN, BIGHAM & MURAD 1995). Considering only the mean values in the simulation fig. 4, it may be concluded that hematite is the most stable among the four phases considered while schwertmannite is more stable than amorphous Fe(OH)₃ at sulfate concentrations above 10–5 mol L⁻¹. The LJUNGSKILE simulation, however, explains that given the uncertainties in the formation constants of the Fe(III) hydrolysis species (uncertainty level I) no unambiguous statement can be made about the stability of schwertmannite Fe₈O₈(OH)₆SO₄ vs. either amorphous Fe(OH)₃, goethite or hematite – except perhaps that at low sulfate concentrations hematite may be more stable than schwertmannite. Taking into account the difficulty to assess solubility products of highly insoluble compounds like Fe(III) hydrates and oxyhydrates, the variability of chemical composition of these compounds and the variability of compositions (Schwertmann & Cornell 1991), the prediction becomes even more questionable. These latter uncertainties are not yet included in the uncertainty intervals in fig. 4.

References

- CHALMERS, R.A. (1993): Space Age Analysis. *Talanta* 40: 121 – 126.
- DE BIÈVRE, P. (1991): Reproducibility is never a proof of accuracy... *Accred. Qual. Assur* 4: 387
- EKBERG, C., ÖDEGAARD-JENSEN, A. & MEINRATH, G. (2003): LJUNGSKILE 1.0: A Computer Program for Investigation of Uncertainties in Chemical Speciation. SKI REPORT 2003:03. Swedish Nuclear Power Inspektorat. Stockholm/S 2003.
- EURACHEM (2000): Quantifying Uncertainty in Chemical Measurement (Ellison, Rösslein, Williams, eds.) EURACHEM.
- GUM (1993): Guide to the Expression of Uncertainty in Measurement. ISO Genève/CH.
- KELL, G.S. (1977): *Nature* 267: 665 – 667.
- MARTIN, J.H. & FITZWATER, S.E. (1988): Iron deficiency limits phytoplankton growth in the north-east Pacific subarctic. *Nature* 331: 341 – 343.
- MAY, P.M. & MURRAY, K. (1991): JESS – a joint expert speciation system I. *Raison d'être. Talanta* 38: 1409 – 1417
- MEINRATH, G. (2000): Comparability of thermodynamic data – a metrological point of view. *Fresenius J. Anal. Chem.* 368: 574 – 584.
- MEINRATH, G. & SPITZER P. (2000): Uncertainties in Determination of pH. *Mikrochim. Acta* 135: 155 – 168.
- MEINRATH, G. (2001): Measurement uncertainty of thermodynamic data. *Fresenius J. Anal. Chem.* 369: 690 – 697
- 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., EKBERG, C. & STRÖMBERG, B. (2003): A retraceable method to assess uncertainties in solubility estimations exemplified by a few americium solids. 35: 55 – 66.
- MILLERO, F.J., YAO, W. & AICHER, J. (1995): The speciation of Fe(II) and Fe(III) in natural waters. *Mar. Chem.* 50 (1995) 21 – 39.
- PRICE, G. (2000): Pragmatic philosophies of measurement in chemistry? *Chem. Australia*. October 2000: 38 – 40
- SCHWERTMANN, U., BIGHAM, J.M. & MURAD, E. (1995): The first occurrence of schwertmannite in a natural stream environment. *Eur. J. Mineralogy* 7: 547 – 552.
- SCHWERTMANN, U. & CORNELL, R.M. (1991): *Iron Oxides in the Laboratory*. VCH Verlagsgesellschaft Weinheim/FRG
- SPITZER, P. & WERNER, B. (2002): Improved reliability of pH Measurement. *Anal. Bioanal. Chem.* 374: 787 – 795.
- THOMPSON, M. (1993): Statistics – the curse of analytical classes. *Analyst* 119: 127N
- VIM (1994): *International Vocabulary of Basic and General Terms in Metrology*. Beuth Verlag Berlin/FRG.
- WIELGOSZ, R.I. (2002): International comparability of chemical measurement results. *Anal. Bioanal. Chem.* 374: 767 – 771