



ELSEVIER

doi:10.1016/j.gca.2004.09.033

## The effects of database parameter uncertainty on uranium(VI) equilibrium calculations

FRANK H. DENISON and JACQUELINE GARNIER-LAPLACE\*

Laboratoire de Radioécologie et Ecotoxicologie, Institut de Radioprotection et Sûreté Nucléaire, CE Cadarache, Bât. 186, BP 3, 13115 St-Paul-Lez-Durance cedex, France

(Received March 8, 2004; accepted in revised form September 30, 2004)

**Abstract**—The propagation of database parameter uncertainty has been assessed for aqueous and mineral equilibrium calculations of uranium by Monte Carlo and quasi-Monte Carlo simulations in simple inorganic solution compositions. The concentration output distributions of individual chemical species varies greatly depending on the solution composition modelled. The relative uncertainty for a particular species is generally reduced in regions of solution composition for which it is predicted to be dominant, due to the asymptotic behaviour imposed by the mass balance constraint where the species concentration approaches the total element concentration. The relative uncertainties of minor species, in regions where another species comprising one or several of the same components is predicted to be dominant with a high probability, also appear to be reduced slightly. Composition regions where two or several species are equally important tend to produce elevated uncertainties for related minor species, although the uncertainties of the major species themselves tend to be reduced. The non-linear behaviour of the equilibrium systems can lead to asymmetric or bimodal output distributions; this is particularly evident close to equivalence points or solubility boundaries. Relatively conservative estimates of input uncertainty can result in considerable output uncertainty due to both the complexity of uranium solution chemistry and the system interdependencies. The results of this study suggest that for some modelling scenarios, “classical” speciation calculations based on mean value estimates of the thermodynamic values may result in predictions of a relatively low probability compared to an approach that considers the effects of uncertainty propagation. Copyright © 2005 Elsevier Ltd

### 1. INTRODUCTION

Aqueous speciation modelling is a widely used interdisciplinary activity encompassing, amongst others, the fields of numerical mathematics, chemistry, geochemistry, hydrology and ecotoxicology. Speciation studies, both analytical and modelling approaches, are central to modern geochemistry and other fields such as ecotoxicology. They have been successfully applied to a wide range of systems to elucidate the processes underpinning water quality, transport/retention phenomena (Van der Lee and De Windt, 2001) and bioavailability or toxicity of contaminants to the biota (Paquin et al., 2002). Although analytical techniques to determine speciation are improving, it is often not practicable or possible to directly measure the activities of individual solution species, particularly at the low concentrations of environmental interest. Therefore predictive geochemical speciation models are employed to estimate the distribution of the total metal concentration through its various possible species (i.e., solution and surface complexes or mineral phases).

Equilibrium solution speciation has been extensively studied, and a robust comprehensive theoretical framework based on equilibrium thermodynamic principles has been developed. As for all mathematical models the predictive ability is limited by both the conceptual uncertainties in the model formulation and the quality of the data supplied to the model, including both the modelling scenario and the model constants (Ekberg, 1999). Within the context of geochemical modelling to calculate the distribution of the various chemical species of an element in the

aqueous phase, conceptual uncertainties such as the inclusion or omission of chemical species and the veracity of the chosen chemical models (e.g., for activity-concentration relationships and the assumption of system equilibrium) will generally result in a systematic bias, although the vector of bias may vary depending on the particular modelling scenario. In addition to the potential uncertainty in the structural model of the system of interest, there is a degree of uncertainty inherent in the analytical measurement of the required thermodynamic data that will result in uncertainty in the model predictions. Obviously, all of these statements are also applicable to the modelling of the distribution of an element between solid and liquid phases.

Quality assessment should be an integral part of the application of computer modelling to environmental problems and a number of authors have underlined the significant variations in existing thermodynamic data (OECD-NEA, 1996; Unsworth and Jones, 2002). However with a few exceptions (Cabaniss, 1999; Ekberg, 1999; Smith et al., 1999; Nitzsche et al., 2000; Tebes-Stevens et al., 2001; Criscenti et al., 1996), systematic approaches to investigate the effects of this inherent uncertainty have not been applied. A variety of speciation codes are widely available which apply one of two distinct, but thermodynamically related, techniques to calculate equilibria in aqueous systems: Gibbs free energy minimisation methods or the more commonly applied equilibrium constant method. Both of these approaches require a reliable and consistent database of accurate thermodynamic values appropriate to the domain of application to be provided to the model in some form. Databases that are sufficiently coherent to be applied to a wide range of different systems are large, typically containing data for several thousand chemical species, which creates an obvious potential

\* Author to whom correspondence should be addressed (jacqueline.garnier-laplace@irsn.fr).

for the introduction of errors and uncertainty. It has long been recognised that considerable differences exist between different database compilations, and intercomparison exercises have frequently resulted in widely differing results. These differences are sometimes the result of errors introduced by data reduction or conversion (Serkiz et al., 1996) but more generally due to a lack of consistency in the compilation. Literature values frequently give small confidence limits, if any are given, however when a number of determinations are compared, much greater differences are found than these intervals would indicate. This variability arises from a number of causes including inconsistent treatment of the chemical system and systematic errors of the applied method. Additionally, as noted by Nitzsche et al. (2000), the requirements for accuracy and precision of data in speciation modelling databases are normally much higher than those that were demanded for the original application.

It is obvious that there is inherent uncertainty in the application of speciation modelling, even in conditions where the conceptual model formulation is ideal. Normally, modelling is performed using only single values of the thermodynamic data contained in the database. This approach does not permit any estimation of the uncertainty associated with the model predictions. Ideally, each thermodynamic value may be characterised by its expectation value and an assigned uncertainty depending on the extent and quality of the available data. The propagation of these input uncertainty distributions can lead to very different species concentration output distributions (Cabaniss, 1997). Consideration of this uncertainty propagation should be integral to the application of speciation modelling to environmental studies, including both the interpretation of acquired data and the predictive modelling for transport, bioavailability or toxicity studies for the studied element.

The objective of this work is to investigate the effect of database parameter uncertainty, using realistic estimates for the expectation and uncertainty values, on the output concentration distributions for the aqueous speciation modelling of uranium in very simple inorganic solution compositions. Uranium was selected as the element of interest due to the existence of generally high quality thermodynamic data resulting from the OECD-NEA review work and subsequent studies and its relatively complex solution chemistry to highlight the potential for uncertainty propagation. The calculations were performed using Monte Carlo (MC) and quasi-Monte Carlo (QMC) methods, assuming mutually independent gaussian distributions for the parameter uncertainty. Although derivative methods of predicting uncertainty exist (Cabaniss, 1997) and are considerably faster than Monte Carlo methods, the requirement of linear combination of the parameter uncertainties was not satisfied for the speciation of uranium, and so this method was not applied. The number of runs required to provide acceptably small parameter estimate errors was investigated for a number of descriptive parameters. The potential of the different sampling strategies of quasi-Monte Carlo methods to improve efficiency over the Monte Carlo method was also assessed. Finally, recommendations are provided for the proper use of chemical speciation models, considering the effects of uncertain thermodynamic data.

## 2. METHODOLOGY

The modelling scenarios selected for the study were deliberately very simple with restricted ranges of solution compositions. The effects of varying the partial pressure of CO<sub>2</sub> (0–10<sup>-2</sup> atm) and the total concentration of phosphate (0–10<sup>-3</sup> mol dm<sup>-3</sup>) were both investigated in the pH range 4 to 9. The presence of a non-complexing electrolyte of 10<sup>-2</sup> mol dm<sup>-3</sup> was included in all scenarios to maintain a constant ionic strength.

To assess the effects of database uncertainty on model output, realistic values for all the database constants and their uncertainties are required. To realise this, a comprehensive database was compiled, based mainly on the OECD-NEA data reviews (Grenthe et al., 1992) and modified where more recent data have become available. New literature data were added following the NEA recommended procedures as closely as possible and the equilibrium constants were reduced to standard molar Gibbs energy of formation values for the calculation of uncertainty values. To eliminate error amplification due to this data reduction the uncertainty values were calculated relative to a set of basis species adapted to the solution compositions rather than the NEA elemental reference states. The thermodynamic data together with the uncertainties as single standard deviation values are given in Table 1.

The chemical equilibrium program CHESS (Van der Lee, 1998) was used to perform all speciation calculations using the truncated-Davies model of activity-correction, the modified Newton-Raphson error criterion set to 10<sup>-10</sup> and suppressing the formation of mineral phases. To study the uncertainty propagation Monte Carlo simulations of 10,000 runs were performed, each run drawing input thermodynamic data from independent normal distributions of given mean value and standard deviation for each equilibrium equation. For the Monte Carlo simulations the input thermodynamic data were generated using the Box-Muller (Box and Muller, 1958) algorithm, supplied with uniform variates by the "Mersenne Prime Twister" random number generator MT19937 of Matsumoto and Nishimura (1998). For the quasi-Monte Carlo simulations the samples were generated from a Sobol (1967) low discrepancy sequence point set, transformed to gaussian variates by Moro's (1995) algorithm. Simulations performed using samples generated from low discrepancy sequences are theoretically more efficient than simulations using random sampling, i.e., the sample size required to obtain a sufficiently good representation of the output distribution is theoretically smaller. A program was written in C using functions from the GNU-GSL (Galassi, 2002) to provide input data files, extract the required data from the output files and perform statistical analyses on the output distributions.

The probabilistic errors of the parameter estimations were calculated by a jackknife estimate of variance (Efron and Tibshirani, 1993)  $\sigma^2$  given by

$$\sigma^2 = \frac{n-1}{n} \sum_{i=1}^n (p_i - \bar{p})^2$$

where  $n$  is the number of data,  $\bar{p}$  is the estimated parameter and  $p_i$  is the parameter recalculated  $n$  times omitting the  $i$ th datum.

## 3. RESULTS AND DISCUSSION

### 3.1. Evaluation of Simulation Input Parameter Distributions

The generated input parameter distributions for both the MC and QMC methods were evaluated by the Kolmogorov-Smirnov test (Knuth, 1981) and were acceptably gaussian for  $n = 100$  or greater. The maximum difference between the distributions' mean and the assigned mean values of the equilibrium reactions given in Table 1 was ~2% and 0.8% for  $n = 100$ , dropping to 0.2% and 0.02% for  $n = 10^4$  for the MC and the QMC methods respectively.

Table 1. Selected thermodynamic data (mean log equilibrium constants and their associated standard deviation values) for the system H<sub>2</sub>O-UO<sub>2</sub><sup>2+</sup>-CO<sub>2</sub>-PO<sub>4</sub><sup>3-</sup> at *T* = 0 and 25°C.

| Reaction   | Log <i>K</i> | $\sigma$ | References |
|--|--------------|----------|------------|
| 1 H <sub>2</sub> O = 1 OH <sup>-</sup> + 1 H <sup>+</sup>  | -14.00       | 0.005    | a          |
| 1 HCO <sub>3</sub> <sup>-</sup> + 1 H <sup>+</sup> = 1 CO <sub>2</sub> (g) + 1 H <sub>2</sub> O  | 7.83         | 0.01     | a          |
| 1 HCO <sub>3</sub> <sup>-</sup> + 1 H <sup>+</sup> = 1 CO <sub>2</sub> (aq) + 1 H <sub>2</sub> O   | 6.35         | 0.026    | a          |
| 1 HCO <sub>3</sub> <sup>-</sup> = 1 CO <sub>3</sub> <sup>2-</sup> + 1 H <sup>+</sup>   | -10.33       | 0.036    | a          |
| 1 HPO <sub>4</sub> <sup>2-</sup> = 1 PO <sub>4</sub> <sup>3-</sup> + 1 H <sup>+</sup>  | -12.35       | 0.14     | a          |
| 1 HPO <sub>4</sub> <sup>2-</sup> + 1 H <sup>+</sup> = 1 H <sub>2</sub> PO <sub>4</sub> <sup>-</sup>  | 7.21         | 0.14     | a          |
| 1 HPO <sub>4</sub> <sup>2-</sup> + 2 H <sup>+</sup> = 1 H <sub>3</sub> PO <sub>4</sub> (aq)  | 9.35         | 0.14     | a          |
| 1 UO <sub>2</sub> <sup>2+</sup> + 1 H <sub>2</sub> O = UO <sub>2</sub> OH <sup>+</sup> + 1 H <sup>+</sup>  | -5.36        | 0.22     | a,b,c      |
| 1 UO <sub>2</sub> <sup>2+</sup> + 2 H <sub>2</sub> O = UO <sub>2</sub> (OH) <sub>2</sub> (aq) + 2 H <sup>+</sup>   | -11.75       | 0.46     | b,d,e,f    |
| 1 UO <sub>2</sub> <sup>2+</sup> + 3 H <sub>2</sub> O = UO <sub>2</sub> (OH) <sub>3</sub> <sup>-</sup> + 3 H <sup>+</sup>   | -19.6        | 0.46     | a,g        |
| 1 UO <sub>2</sub> <sup>2+</sup> + 4 H <sub>2</sub> O = UO <sub>2</sub> (OH) <sub>4</sub> <sup>2-</sup> + 4 H <sup>+</sup>  | -34.23       | 1.7      | a,h        |
| 2 UO <sub>2</sub> <sup>2+</sup> + 1 H <sub>2</sub> O = 1 (UO <sub>2</sub> ) <sub>2</sub> OH <sup>3+</sup> + 1 H <sup>+</sup>   | -2.7         | 0.60     | a          |
| 2 UO <sub>2</sub> <sup>2+</sup> + 2 H <sub>2</sub> O = (UO <sub>2</sub> ) <sub>2</sub> (OH) <sub>2</sub> <sup>2+</sup> + 2 H <sup>+</sup>  | -5.7         | 0.29     | a,c,i,j    |
| 3 UO <sub>2</sub> <sup>2+</sup> + 4 H <sub>2</sub> O = 1 (UO <sub>2</sub> ) <sub>3</sub> (OH) <sub>4</sub> <sup>2+</sup> + 4 H <sup>+</sup>  | -11.9        | 0.50     | a          |
| 3 UO <sub>2</sub> <sup>2+</sup> + 5 H <sub>2</sub> O = (UO <sub>2</sub> ) <sub>3</sub> (OH) <sub>5</sub> <sup>+</sup> + 5 H <sup>+</sup>   | -15.59       | 0.40     | a,c,j      |
| 3 UO <sub>2</sub> <sup>2+</sup> + 7 H <sub>2</sub> O = (UO <sub>2</sub> ) <sub>3</sub> (OH) <sub>7</sub> <sup>-</sup> + 7 H <sup>+</sup>   | -30.18       | 2.0      | a,c,g,j    |
| 4 UO <sub>2</sub> <sup>2+</sup> + 7 H <sub>2</sub> O = 1 (UO <sub>2</sub> ) <sub>4</sub> (OH) <sub>7</sub> <sup>+</sup> + 7 H <sup>+</sup>   | -21.9        | 0.81     | a          |
| 1 UO <sub>2</sub> <sup>2+</sup> + 1 HCO <sub>3</sub> <sup>-</sup> = 1 UO <sub>2</sub> CO <sub>3</sub> (aq) + 1 H <sup>+</sup>  | -0.65        | 0.16     | a          |
| 1 UO <sub>2</sub> <sup>2+</sup> + 2 HCO <sub>3</sub> <sup>-</sup> = 1 UO <sub>2</sub> (CO <sub>3</sub> ) <sub>2</sub> <sup>2-</sup> + 2 H <sup>+</sup>   | -3.71        | 0.18     | a          |
| 1 UO <sub>2</sub> <sup>2+</sup> + 3 HCO <sub>3</sub> <sup>-</sup> = 1 UO <sub>2</sub> (CO <sub>3</sub> ) <sub>3</sub> <sup>4-</sup> + 3 H <sup>+</sup>   | -9.38        | 0.19     | a          |
| 3 UO <sub>2</sub> <sup>2+</sup> + 6 HCO <sub>3</sub> <sup>-</sup> = 1 (UO <sub>2</sub> ) <sub>3</sub> (CO <sub>3</sub> ) <sub>6</sub> <sup>6-</sup> + 6 H <sup>+</sup>   | -7.96        | 0.72     | a          |
| 2 UO <sub>2</sub> <sup>2+</sup> + 1 HCO <sub>3</sub> <sup>-</sup> + 3 H <sub>2</sub> O = 1 (UO <sub>2</sub> ) <sub>2</sub> CO <sub>3</sub> (OH) <sub>3</sub> <sup>-</sup> + 4 H <sup>+</sup>                       | -11.18       | 0.40     | a          |
| 3 UO <sub>2</sub> <sup>2+</sup> + 1 HCO <sub>3</sub> <sup>-</sup> + 3 H <sub>2</sub> O = 1 (UO <sub>2</sub> ) <sub>3</sub> O(OH) <sub>2</sub> (HCO <sub>3</sub> ) <sup>+</sup> + 4 H <sup>+</sup>                  | -9.67        | 0.54     | a          |
| 11 UO <sub>2</sub> <sup>2+</sup> + 6 HCO <sub>3</sub> <sup>-</sup> + 12 H <sub>2</sub> O = 1 (UO <sub>2</sub> ) <sub>11</sub> (CO <sub>3</sub> ) <sub>6</sub> (OH) <sub>12</sub> <sup>2-</sup> + 18 H <sup>+</sup> | -25.54       | 2.0      | a          |
| 1 UO <sub>2</sub> <sup>2+</sup> + 1 HPO <sub>4</sub> <sup>2-</sup> = 1 UO <sub>2</sub> PO <sub>4</sub> <sup>-</sup> + 1 H <sup>+</sup>   | 0.88         | 0.22     | a          |
| 1 UO <sub>2</sub> <sup>2+</sup> + 1 HPO <sub>4</sub> <sup>2-</sup> = UO <sub>2</sub> HPO <sub>4</sub> (aq)   | 7.3          | 0.24     | a,g,k      |
| 1 UO <sub>2</sub> <sup>2+</sup> + 1 HPO <sub>4</sub> <sup>2-</sup> + 1 H <sup>+</sup> = UO <sub>2</sub> H <sub>2</sub> PO <sub>4</sub> <sup>+</sup>  | 10.43        | 0.34     | a,k,l      |
| 1 UO <sub>2</sub> <sup>2+</sup> + 1 HPO <sub>4</sub> <sup>2-</sup> + 2 H <sup>+</sup> = 1 UO <sub>2</sub> H <sub>3</sub> PO <sub>4</sub> <sup>2+</sup>   | 10.11        | 0.22     | a          |
| 1 UO <sub>2</sub> <sup>2+</sup> + 2 HPO <sub>4</sub> <sup>2-</sup> + 2 H <sup>+</sup> = UO <sub>2</sub> (H <sub>2</sub> PO <sub>4</sub> ) <sub>2</sub> (aq)  | 20.2         | 0.88     | a,k,l      |
| 1 UO <sub>2</sub> <sup>2+</sup> + 2 HPO <sub>4</sub> <sup>2-</sup> + 3 H <sup>+</sup> = 1 UO <sub>2</sub> (H <sub>2</sub> PO <sub>4</sub> )(H <sub>3</sub> PO <sub>4</sub> ) <sup>+</sup>                          | 20.35        | 0.33     | a          |
| 1 UO <sub>2</sub> <sup>2+</sup> + 3 H <sub>2</sub> O = 1 UO <sub>3</sub> · 2H <sub>2</sub> O <sub>(s)</sub> + 2 H <sup>+</sup>   | -4.81        | 0.15     | a          |

<sup>a</sup> Grenthe et al. (1992).

<sup>b</sup> Choppin and Mathur (1991).

<sup>c</sup> De Stefano et al. (2002).

<sup>d</sup> Silva (1992).

<sup>e</sup> Lemire and Tremaine (1980).

<sup>f</sup> Grenthe et al. (1995).

<sup>g</sup> Sandino and Bruno (1992).

<sup>h</sup> Baston et al. (1993).

<sup>i</sup> Meinrath (1998).

<sup>j</sup> Palmer and Nguyen-Trung (1995).

<sup>k</sup> Brendler et al. (1996).

<sup>l</sup> Mathur (1991).

### 3.2. Probabilistic Error of the Parameter Estimates for the Uranium Species Output Distributions

The value of *n* that was required to obtain acceptable estimates of the output distributions depended on: the solution composition, the output species and the parameter(s) of interest. The coefficient of variance of the jackknife  $\sigma$  estimates for a number of different statistical parameters was calculated for different *n* values between 10<sup>2</sup> and 10<sup>4</sup> for both the MC and QMC methods, for various different solution compositions. Some representative results of the jackknife  $\sigma$  estimates are shown in Figure 1, for a solution composition of: total uranium concentration 10<sup>-6</sup> mol dm<sup>-3</sup>, pH 6, pCO<sub>2</sub> 3.10<sup>-4</sup> atm in a hypothetically non-complexing electrolyte of 10<sup>-2</sup> mol dm<sup>-3</sup>. The coefficient of variance of the MC and QMC estimates of the mean values for the output distributions of the species UO<sub>2</sub><sup>2+</sup>, UO<sub>2</sub>OH<sup>+</sup>, UO<sub>2</sub>(OH)<sub>2</sub>(aq), UO<sub>2</sub>(OH)<sub>3</sub><sup>-</sup> and UO<sub>2</sub>(CO<sub>3</sub>)<sub>2</sub><sup>2-</sup>

are shown. The results show the expected reduction in the probabilistic error of the MC mean estimates as the value of *n* increases; the error has probabilistic order  $O(n^{-1/2})$ , i.e., the probabilistic error reduces proportionally to  $n^{-1/2}$  as the sample size, *n*, increases. No improvement in the error bound was found by using the QMC instead of the MC method in this range of *n*, although the convergence of the error bound was often more regular. The theoretical advantage of using a low-discrepancy sequence for the QMC method is that the asymptotic error bound rate (as  $n \rightarrow \infty$ ) is superior to that of the MC method for a fixed dimension, *s*, if *n* is sufficiently large i.e.,  $O(n^{-1} \log^s n)$ . However, for reasonably large values of *s*, the value of *n* required to ensure that the QMC error bound is smaller than that of the MC can be very large and hence of little practical advantage (L'Ecuyer and Lemieux, 2002) as appears to be the case here. The error bounds of other statistical parameters of the

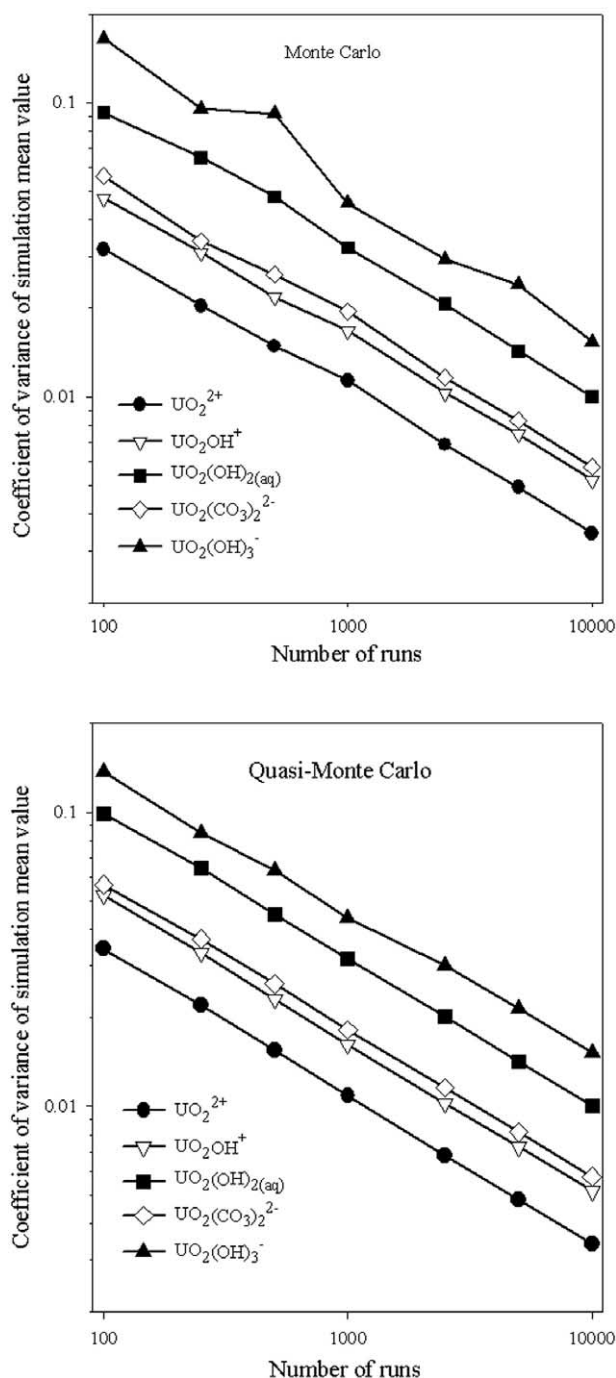


Fig. 1. Coefficient of variance of distribution mean estimates by the MC and QMC methods for selected output species.  $[\text{UO}_2^{2+}]_T = 10^{-6}$  mol dm $^{-3}$ , pH = 6,  $p\text{CO}_2 = 3.10^{-4}$  atm.

distributions, such as the standard deviation, higher moments or quantiles are greater than those of the mean values, sometimes very much so. The value of  $n$  required to obtain acceptable estimates of some parameters for some (usually minor) species can be prohibitively large. However in the case of  $\text{UO}_2^{2+}$  for the conditions investigated (pH 5–8,  $p\text{CO}_2$  0– $10^{-2}$  atm,  $[\text{UO}_2]_T = 10^{-6}$  mol dm $^{-3}$ ), the maximum error bound of the jackknife estimates of standard

deviation for the mean, median, standard deviation and a number of arbitrary quantiles was  $\sim 10\%$  of the parameter value for  $n = 10^3$ , dropping to  $\sim 2\%$  for  $n = 10^4$ . This demonstrates the principal limitation of the MC methods, i.e., it is not possible to determine a priori the value of  $n$  required to provide acceptable estimates of the output distributions, additionally the value of  $n$  can vary greatly depending on the particular modelling scenario and output species of interest. In cases where the value of  $n$  required is large, alternative strategies to improve the sampling efficiency of the parameter space could be advantageous.

### 3.3. Concentration Output Distributions

Just as the concentrations of individual solution species of uranium can change dramatically for quite small changes in the solution composition, so can the forms of the output distributions. The output distributions are generally not gaussian (or log gaussian) and the forms of the distributions (dispersion, skewness, kurtosis) can change significantly on changing the solution composition. This is best shown by the skewness values, a measure of the distribution's asymmetry calculated from the third moment of the data. Symmetrical distributions have a small skewness value, whilst asymmetric distributions have large either positive or negative values. Figure 2 shows MC estimates of the skewness values of the output distributions of four species as functions of pH and for several different values of  $p\text{CO}_2$ . One consequence of these asymmetric distributions is that summarising the distributions using parametric statistical functions can be misleading. Additionally, the distribution mean value estimates are often significantly different from the mean-value database calculations. This suggests that "classical" speciation calculations based on mean value estimates of the thermodynamic values may result in predictions of a relatively low probability compared to an approach that considers the effects of uncertainty propagation.

### 3.4. The Effect of Solution Composition on the Output Distributions

To investigate the effects of changing solution composition on the concentration output distributions and the relative uncertainties of model predictions, three parameters that strongly influence the speciation of uranium were selected and varied within realistic ranges. All calculations were performed using the MC method with  $n = 10^4$ , total uranyl concentration was fixed at  $10^{-6}$  mol dm $^{-3}$ , the ionic strength was maintained constant by including a hypothetically non-complexing electrolyte of  $10^{-2}$  mol dm $^{-3}$  and the formation of solid phases was suppressed. Typical output distributions are shown for selected species at several pH values and a  $\text{CO}_2$  partial pressure of  $3.10^{-4}$  atm in Figure 3, the areas under the probability density curves are unity. The distributions for each species can vary considerably between different pH values, and the concentration distributions for some species can cover several orders of magnitude. To summarise the effect of database uncertainty on the relative uncertainty of output concentrations as functions of varying solution composition, interquantile ranges normalised with respect to the distribution mean were calculated. Figure 4 shows the effect of varying pH and  $p\text{CO}_2$  on the interdecile

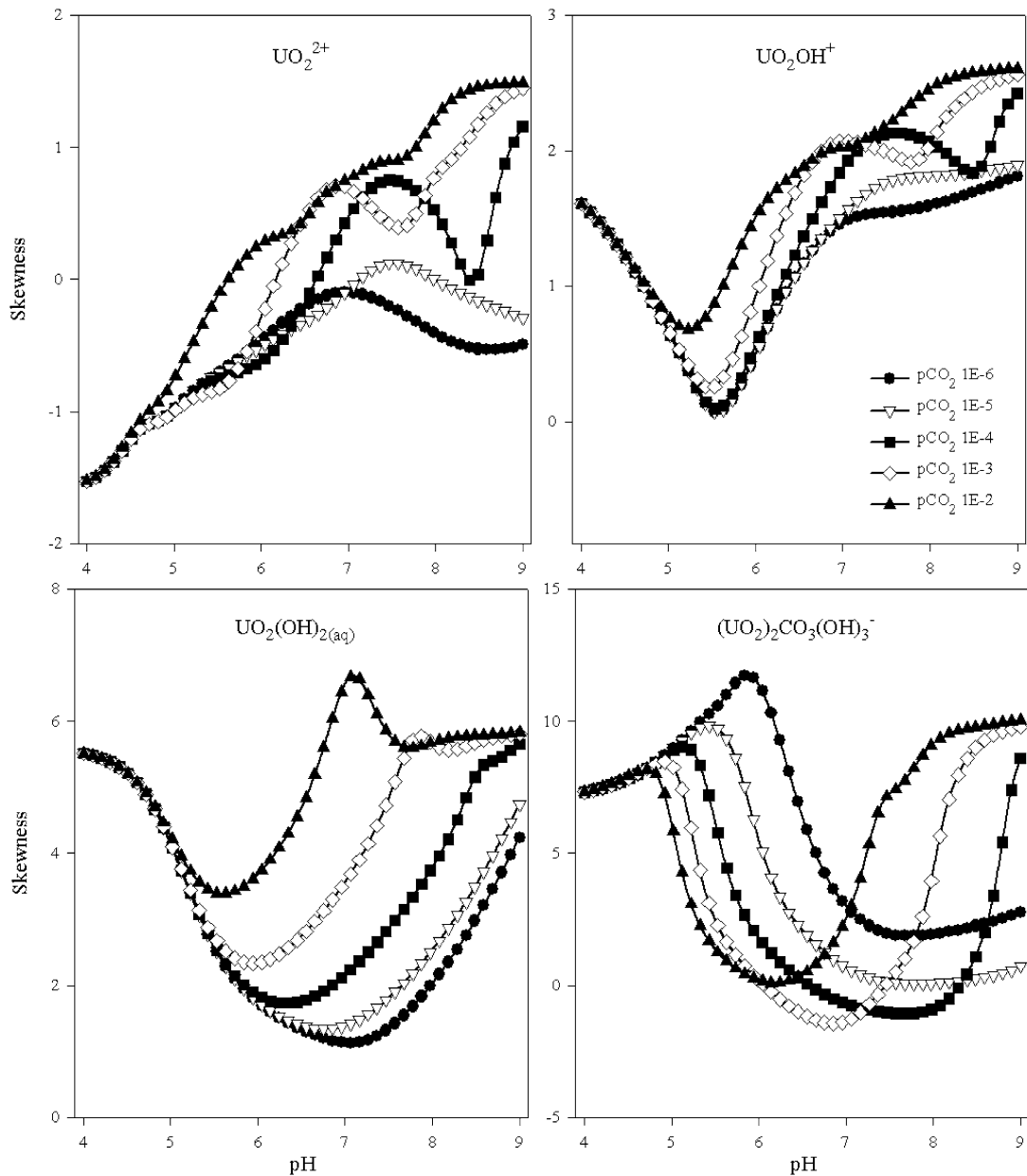


Fig. 2. Skewness of selected output distributions as a function of pH and at several CO<sub>2</sub> partial pressures.

(Q0.1–Q0.9) distribution intervals as a percentage of the mean for the species  $\text{UO}_2^{2+}$  and  $\text{UO}_2(\text{CO}_3)_2^{2-}$ . Figure 5 shows the effect of varying pH and phosphate concentration on the species  $\text{UO}_2^{2+}$  and  $\text{UO}_2\text{HPO}_4(\text{aq})$ . In all cases, the relative uncertainty varies considerably as a function of the solution composition. In general, in solution compositions for which the mean-value database predicts a species to be dominant, the output uncertainty of that species is relatively low compared to other composition regions. This can be seen for the species  $\text{UO}_2^{2+}$  at low pH and phosphate concentration, the species  $\text{UO}_2\text{HPO}_4(\text{aq})$  at moderately high phosphate concentrations and low pH and also for the local minima of the species  $\text{UO}_2(\text{CO}_3)_2^{2-}$ . The dominant species predicted by the mean-value database for the region of pH values between 7 and 8 and

pCO<sub>2</sub> values between 10<sup>-4</sup> and 10<sup>-2</sup> atm changes from the species  $(\text{UO}_2)_2\text{CO}_3(\text{OH})_3^-$  to  $\text{UO}_2(\text{CO}_3)_2^{2-}$  to  $\text{UO}_2(\text{CO}_3)_3^{4-}$  on increasing pH and pCO<sub>2</sub>; the relative uncertainties of all three of these species exhibit local minima in the regions where they are predicted to be dominant with a high probability. The relative uncertainties of minor species, in regions where another species comprising one or several of the same components is predicted to be dominant with a high degree of certainty, also appear to be reduced slightly (e.g., the behaviour of the relative uncertainty of  $\text{UO}_2^{2+}$  in the region of pH and pCO<sub>2</sub> where the dominant uranyl species changes). Similarly, equivalence points between two species dominant in different regions tend to produce elevated uncertainties for related minor species, although the uncertainties of the major species them-

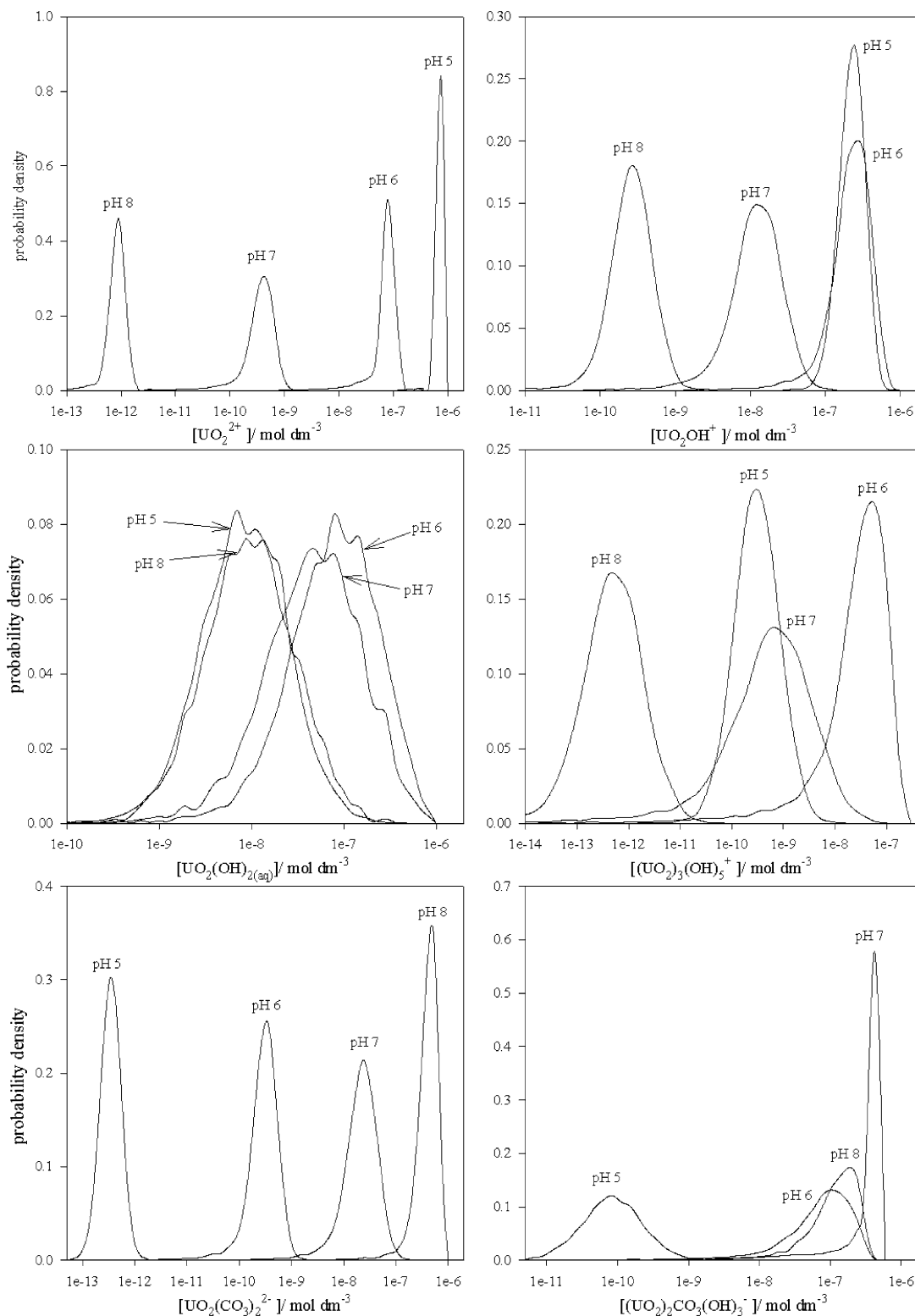


Fig. 3. Probability density functions of concentration output distributions from  $n = 10^4$  Monte Carlo simulations at four pH values and a  $p\text{CO}_2$  of  $3.10^{-4}$  atm.

selves tend to be reduced. This type of behaviour has been observed before, for example simulations of base titration's of simple acids performed by Cabaniss (1997) produced large

uncertainties in the pH output distributions near to the equivalence points, the distributions being bimodal with minima at or close to the equivalence points.

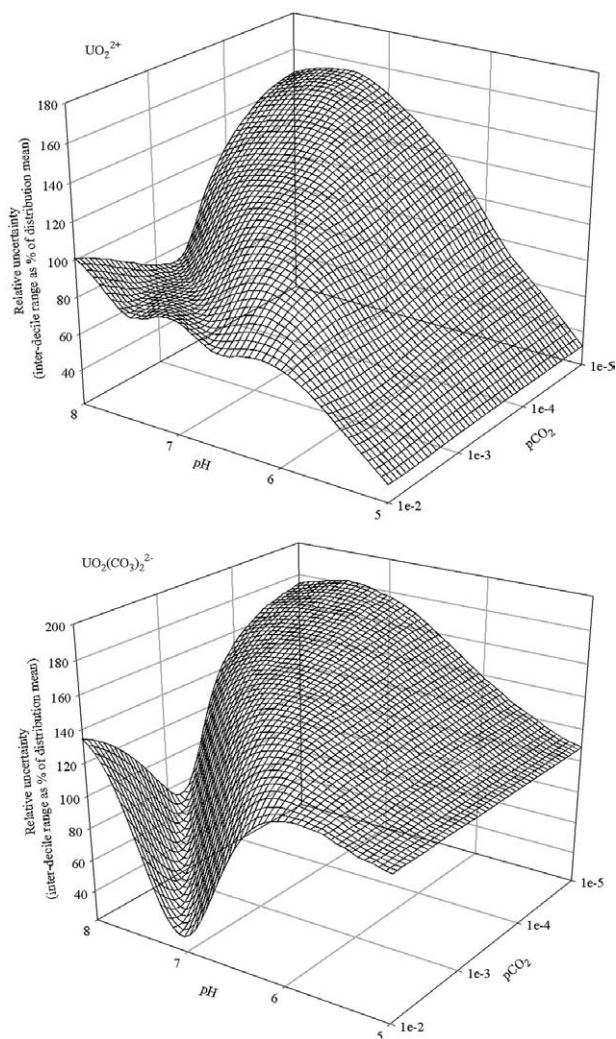


Fig. 4. Relative uncertainty of output concentrations of selected species as a function of pH and  $p\text{CO}_2$ .

### 3.5. Effect on Solubility Equilibria

The effect of database uncertainty on calculations involving equilibrium with a solid phase was investigated, by performing simulations involving equilibrium with  $\text{UO}_3 \cdot 2\text{H}_2\text{O}(\text{s})$  (schoepite). Total uranyl concentration was fixed at  $10^{-2} \text{ mol dm}^{-3}$  and a range of solution compositions varying both pH and  $p\text{CO}_2$  were modelled. Calculations using the mean-value database indicated that an equivalence point between the solid and aqueous uranium fractions, at atmospheric  $p\text{CO}_2$ , is located at a pH value of  $\sim 8.7$ . As stated earlier, composition regions near to equivalence points are often sensitive to parameter uncertainty and this was again found to be the case. Because the total uranyl concentration was constrained, the output distributions of the soluble uranium fraction close to the equivalence point are bimodal, as the concentration output distribution traverses the total concentration constraint. Figure 6 shows the output distributions of the soluble uranium fraction at pH 8.7 and three  $\text{CO}_2$  partial pressures close to the equivalence point. The distributions cover over an order of magnitude in concen-

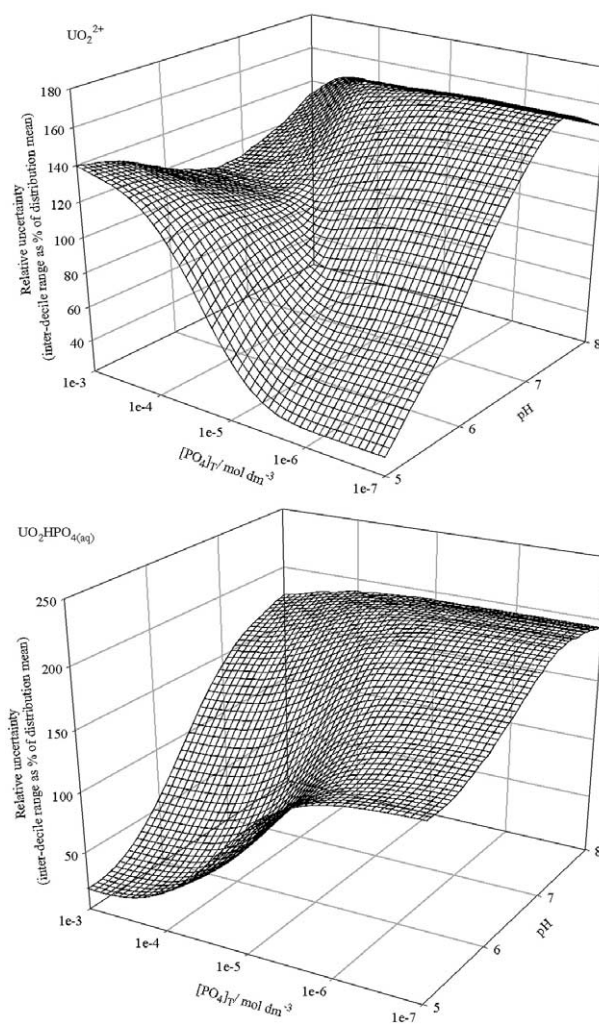


Fig. 5. Relative uncertainty of output concentrations of selected species as a function of pH and phosphate concentration.

tration and obviously predictions of dissolved uranium concentrations under these sensitive conditions would be quite uncertain. A contour plot of the probability densities as a function of the pH at atmospheric  $p\text{CO}_2$  is shown in Figure 7, again the soluble uranium concentration distributions are broadly spread over a large range of pH values with bimodal concentration distributions in some regions of solution composition. A discontinuity at elevated pH near to the dissolution boundary can be clearly seen.

### 3.6. Processing Time

The principal limitation of the Monte Carlo and quasi-Monte Carlo methods is the time required to perform the simulations. The choice of the program CHESS was decided by these concerns, due to its superior computational speed compared to other available programs. The method of coupling the program written to perform the simulations and calculate the statistical parameters of the output distributions with CHESS was decided by the ease of development and was not very efficient, however the processing time required for the presented simulations was

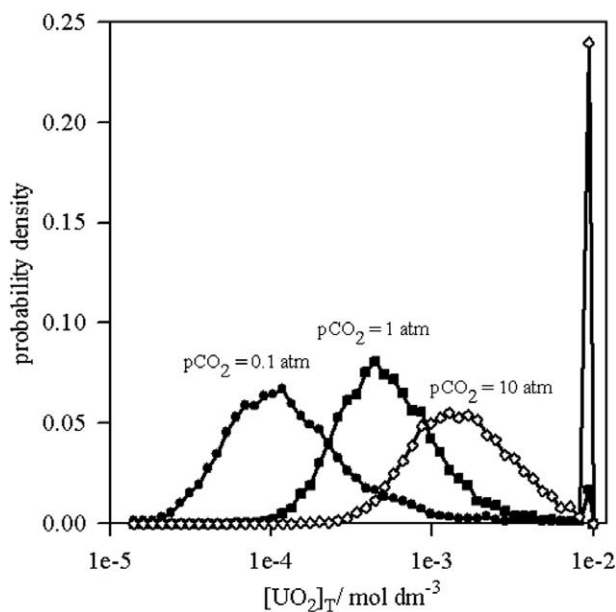


Fig. 6. Probability density functions of the total aqueous uranium fraction in equilibrium with schoepite at pH 8.7 and three different  $\text{CO}_2$  partial pressures.

not prohibitive. For example a simulation of  $n = 10^4$  samples took  $\sim 5$  s using an AMD Athlon 1700+ processor. There is considerable scope to reduce processing time by either improving the method of coupling the two programs, or integrating the Monte Carlo and speciation codes. However the application of the Monte Carlo approach to some speciation modelling applications, such as reactive transport modelling, will obviously be constrained by the processing time. Attempts to reduce the

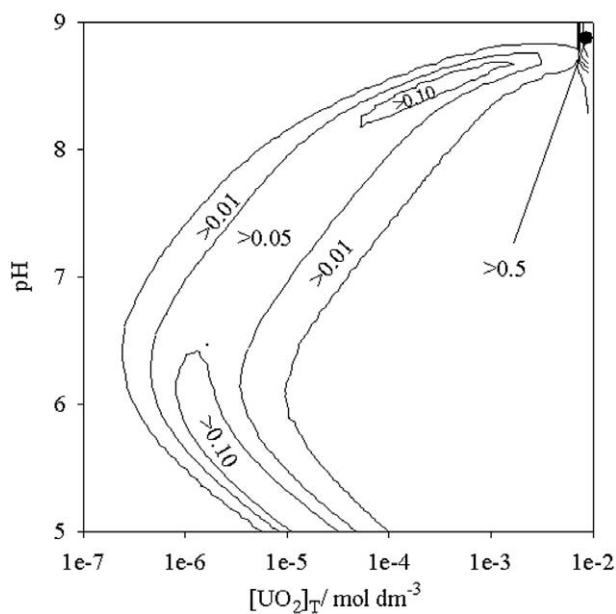


Fig. 7. Contour plot of the probability density function of the total aqueous uranium fraction in equilibrium with schoepite as a function of pH at a  $\text{CO}_2$  partial pressure of 1 atm.

number of runs required to obtain acceptable parameter estimates were not successful for the considered scenarios. The use of a low discrepancy sequence and also Latin Hypercube sampling (McKay et al., 1979), results not presented, did not reduce the error of the parameter estimates compared to the Monte Carlo method, however these different sampling strategies may be more successful when applied to different scenarios. If the modelled system is suitable then the derivative method of estimating the uncertainty proposed by Cabaniss (1997) would be preferred due to the greatly reduced processing requirements. Large ranges of input conditions used together with the randomised parameter values can lead to convergence problems in the modified Newton-Raphson algorithm of CHES. Judicious selection of the scenario basis species was sufficient to overcome this problem for the scenarios presented.

#### 4. CONCLUSIONS

Uncertainty is inherent in the application of geochemical speciation modelling and this is especially the case where there is uncertainty regarding the model formulation, for example the chemical equilibrium model, the assumption of equilibrium conditions and the models used to describe concentration-activity relationships as well as the model constants. Only this second aspect has been investigated in this study, and it is clear that there is the potential for very considerable uncertainty in model predictions from this alone. Model output is only as reliable as the input data and the magnitude and consequences of this uncertainty is specific to the modelling scenario. Due to the very large number of potential applications it is difficult to generalise what the impact of the uncertainty will be. The results of this study illustrate the importance of very high quality thermodynamic data to minimise the effects of uncertainty propagation.

This study is, by necessity, of limited scope and based on a number of simplifying assumptions. The mean and uncertainty values assigned to the thermodynamic parameters are obviously of fundamental importance to the results, as is the assumption of independent gaussian distributions. For a truly coherent database compilation this assumption of mutual independence will be invalid for certain combinations of interdependent parameters, for example the successive hydrolysis products of uranium. However establishing the correlation relationships between different input parameters would be a very significant task. The quality of available databases is highly variable, and there are few that approach the ideal of being both internally consistent and sufficiently coherent to be applied to a wide range of different systems. Hence, although the assumption of mutual independence may be theoretically limited, it is probably justified for many databases in routine use. The issue of uncertainty needs to be considered in all stages of the modelling process to enable valid estimates of a model's predictive ability for a particular scenario, from the experiments performed to determine the chemical equilibrium model and parameter constants to the application of the model. The application of uncertainty and sensitivity analyses to the interpretation of experimental data would serve a number of useful purposes. These include a more robust probabilistic interpretation of the chemical system (and avoidance of the temptation to obtain perfect model fits with the particular database employed by proposing increasingly elaborate chemical systems). Obvi-



ously physical evidence for the existence of the proposed species is always to be preferred. Sensitivity analysis can lead to the establishment of interparameter relationships, which can subsequently be used to both improve uncertainty estimates for the model applications, and also facilitate the maintenance of database consistency. Uncertainty and sensitivity analysis of modelling scenarios can provide information about which parameters are the most sensitive for a particular scenario and allow better targeting of data gathering requirements, although frequently the most sensitive parameters are intuitive with a good understanding of the chemical system.

The focus of this work has been the uncertainty of uranium equilibrium calculations, which was motivated by its environmental relevance and the availability of high quality data reviews. However the aqueous speciation of uranium is complex compared to many other elements, undergoing very large changes over relatively small ranges of solution composition with a considerable number of species that may be significant for various ranges of solution composition. The complexity of uranium aqueous speciation leads to elevated propagation of uncertainties due to the interdependencies of the system (or conversely to a higher requirement for input parameter precision to obtain an acceptable level of output uncertainty). Less complex systems can be expected to result in reduced propagation of input uncertainty, although in particular composition regions, for example near to equivalence points or solubility boundaries, relatively elevated output uncertainties can still be expected.

*Acknowledgments*—The authors gratefully acknowledge three anonymous reviewers and Dr. Eric Oelkers for their constructive comments.

*Associate editor:* E. H. Oelkers

## REFERENCES

- Baston G. M. N., Brownsword M., Cross J. E., Hobley J., Moreton A. D., Smith-Briggs J. K., and Thomason H. P. (1993) *The Solubility of Uranium in Cementitious Near-Field Chemical Conditions*. United Kingdom Atomic Energy Authority, Oxford.
- Box G. E. P. and Muller M. E. (1958) A note on the generation of random normal deviates. *Ann. Math. Stat.* **29**, 610–611.
- Brendler V., Geipel G., Bernhard G., and Nitsche H. (1996) Complexation in the system  $\text{UO}_2[2+]/\text{PO}_4[3-]/\text{OH}[-]$ : Potentiometric and spectroscopic investigations at very low ionic strengths. *Radiochim. Acta* **74**, 75–80.
- Cabaniss S. E. (1997) Propagation of uncertainty in aqueous equilibrium calculations: Non-gaussian output distribution. *Anal. Chem.* **69**, 3658–3664.
- Cabaniss S. E. (1999) Uncertainty propagation in geochemical calculations: Non-linearity in solubility equilibria. *Appl. Geochem.* **14**, 255–262.
- Choppin G. R. and Mathur J. N. (1991) Hydrolysis of actinyl(VI) cations. *Radiochim. Acta* **52/53**, 25–28.
- Criscenti L. J., Laniak G. F., and Erikson R. L. (1996) Propagation of uncertainty through geochemical code calculations. *Geochim. Cosmochim. Acta* **60**, 19, 3551–3568.
- De Stefano C., Gianguzza A., Leggio T., and Sammartano S. (2002) Dependence on ionic strength of the hydrolysis constants for dioxouranium(VI) in  $\text{NaCl}(\text{aq})$  and  $\text{NaNO}_3(\text{aq})$ , at  $\text{pH} < 6$  and  $t = 25^\circ\text{C}$ . *J. Chem. Eng. Data* **47**, 533–538.
- Efron B. and Tibshirani R. J. (1993) *An Introduction to the Bootstrap. Monographs on Statistics and Applied Probability*, Vol. 57. Chapman & Hall, London, UK.
- Ekberg C. (1999) Sensitivity analysis and simulation uncertainties in predictive geochemical modelling. Freiberg On-Line Geoscience. Available at: <http://www.geo.tu-freiberg.de/fog>.
- Galassi M. (2002) GNU Scientific Library. Free Software Foundation, Cambridge, MA.
- Grenthe I., Fuger J., Konings R. J. M., Lemire R. J., Muller A. B., Nguyen-Trung C., and Wanner H. (1992) *Chemical Thermodynamics of Uranium*. North-Holland, Amsterdam, the Netherlands.
- Grenthe I., Puigdomenech I., Sandino M. C. A. and Rand M. H. (1995) Corrections to the uranium NEA-TDB review. In *Chemical Thermodynamics of Americium* (eds. R. J. Silva et al.), pp. 347–374. North-Holland, Amsterdam, the Netherlands.
- Knuth D. E. (1981) *Seminumerical Algorithms*. 2nd ed. *The Art of Computer Programming*, Vol. 2. Addison-Wesley, Reading, MA.
- L'Ecuyer P. and Lemieux C. (2002) Recent advances in randomized quasi-Monte Carlo methods. In *Modelling Uncertainty: An Examination of Its Theory, Methods, and Applications* (eds. M. Dror, P. L'Ecuyer and F. Szidarovszki), pp. 419–474. Kluwer Academic, Amsterdam, the Netherlands.
- Lemire R. J. and Tremaine P. R. (1980) Uranium and plutonium equilibria in aqueous solutions to  $200^\circ\text{C}$ . *J. Chem. Eng. Data* **25**, 361–370.
- Mathur J. N. (1991) Complexation and thermodynamics of the uranyl ion with phosphate. *Polyhedron* **10**, 47.
- Matsumoto M. and Nishimura T. (1998) Mersenne Twister. A 623-dimensionally equidistributed uniform pseudo-random number generator. *ACM Trans. Model. Comput. Simul.*
- McKay M. D., Conover W. J., and Beckman R. J. (1979) A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics* **21**, 239–245.
- Meinrath G. (1998) Direct spectroscopic speciation of schoepite phase equilibria. *J. Radioanal. Nucl. Chem.* **232**, 1–2, 179–188.
- Moro B. (1995) “The full Monte.” In *Over the Rainbow: Developments in Exotic Options and Complex Swaps* (ed. R. Jarrow). RISK Publications.
- Nitzsche O., Meinrath G., and Merkel B. (2000) Database uncertainty as a limiting factor in reactive transport prognosis. *J. Contam. Hydrol.* **44**, 223–237.
- OECD-NEA (1996) *Survey of Thermodynamic and Kinetic Databases*. Organisation for Economic Co-operation and Development, Nuclear Energy Authority, Paris.
- Palmer D. A. and Nguyen-Trung C. (1995) Aqueous uranyl complexes. 3. Potentiometric measurements of the hydrolysis of uranyl(VI) ion at  $25^\circ\text{C}$ . *J. Solut. Chem.* **24**, 1281–1291.
- Paquin P. et al. (2002) The biotic ligand model: A historical overview. *Compar. Biochem. Physiol. C: Toxicol. Pharmacol.* **133**, 1–2, 3–35.
- Sandino A. and Bruno J. (1992) The solubility of  $\text{UO}_2(\text{PO}_4)_2 \cdot 4\text{H}_2\text{O}(\text{s})$  and the formation of U(VI) phosphate complexes: Their influence in uranium speciation in natural waters. *Geochim. Cosmochim. Acta* **56**, 4135.
- Serkiz S. M. et al. (1996) Correcting errors in the thermodynamic database for the equilibrium speciation model MINTEQA2. *Wat. Res.* **30**, 8, 1930–1933.
- Silva R. J. (1992) Mechanisms for the retardation of uranium(VI) migration. *Mat. Res. Cos. Symp. Proc.* **257**, 323.
- Smith D. S., Adams N. W. H., and Kramer J. R. (1999) Resolving uncertainty in chemical speciation determinations. *Geochim. Cosmochim. Acta* **63**, 19/20, 3337–3347.
- Sobol I. M. (1967) The distribution of points in a cube and the approximate evaluation of integrals. *USSR Comput. Math. Math. Phys.* **7**, 86–112.
- Tebes-Stevens C. L., Espinoza F., and Valocchi A. J. (2001) Evaluating the sensitivity of a subsurface multicomponent reactive transport model with respect to transport and reaction parameters. *J. Contam. Hydrol.* **52**, 3–27.
- Unsworth E. R. and Jones P. (2002) The effect of thermodynamic data on computer model predictions of uranium speciation in natural water systems. *J. Environ. Monit.* **4**, 4, 528–532.
- Van der Lee J. (1998) *Thermodynamic and Mathematical Concepts of Chem. École des Mines de Paris*, Paris.
- Van der Lee J. and De Windt L. (2001) Present state and future directions of modeling of geochemistry in hydrogeological systems. *J. Contam. Hydrol.* **47**, 2–4, 265–282.