Appendix C

Assigned uncertainties[1]

This Appendix describes the origin of the uncertainty estimates that are given in the TDB tables of selected data. The original text in [1992GRE/FUG] has been retained in [1995SIL/BID], [1999RAR/RAN] and [2001LEM/FUG], except for some minor changes. Because of the importance of the uncertainty estimates, the present review offers a more comprehensive description of the procedures used.

C.1 The general problem

The focus of this section is on the uncertainty estimates of equilibria in solution, where the key problem is analytical, *i.e.*, the determination of the stoichiometric composition and equilibrium constants of complexes that are in rapid equilibrium with one another. We can formulate analyses of the experimental data in the following way: From N measurements, y_i, of the variable y we would like to determine a set of n equilibrium constants k_r, $r = 1, 2,\ldots, n$, assuming that we know the functional relationship:

$$y = f(k_1, k_2, \ldots k_r \ldots k_n; a_1, a_2, \ldots) \tag{C.1}$$

where a_1, a_2 , *etc.* are quantities that can be varied but whose values (a_{1i}; a_{2i}; *etc.*) are assumed to be known accurately in each experiment from the data sets (y_i, a_{1i}, $a_{2i}, \ldots$), $i = 1, 2, \ldots N$. The functional relationship (C.1) is obtained from the chemical model proposed and in general several different models have to be tested before the "best" one is selected. Details of the procedures are given in Rossotti and Rossotti [1961ROS/ROS].

When selecting the functional relationship (C.1) and determining the set of equilibrium constants that best describes the experiments one often uses a least-squares method. Within this method, the "best" description is the one that will minimise the residual sum of squares, U:

$$U = \sum_i w_i \left[y_i - f(k_1 \ldots k_n; a_{1i}, a_{2i} \ldots) \right]^2 \tag{C.2}$$

where w_i is the weight of each experimental measurement y_i.

---

[1] This Appendix essentially contains the text of the TDB-3 Guideline, [1999WAN/OST], earlier versions of which have been printed in all the previous NEA TDB reviews. Because of its importance in the selection of data and to guide the users of the values in Chapters III and IV the text is reproduced here with minor revisions.

The minimum of the function (C.2) is obtained by solving a set of normal equations:

$$\frac{\partial U}{\partial k_r} = 0, \, r = 1,.....n \tag{C.3}$$

A "true" minimum is only obtained if:

- the functional relationship (C.1) is correct, *i.e.*, if the chemical model is correct.

- all errors are random errors in the variable *y*, in particular there are no systematic errors.

- the random errors in *y* follow a Gaussian (normal) distribution.

- the weight $w_i(y_i, \, a_{1i}, \, a_{2i}....)$ of an experimental determination is an exact measure of its inherent accuracy.

To ascertain that the first condition is fulfilled requires chemical insight, such as information of the coordination geometry, relative affinity between metal ions and various donor atoms, *etc*. It is particularly important to test if the chemical equilibrium constants of complexes that occur in small amounts are chemically reasonable. Too many experimentalists seem to look upon the least-squares refinement of experimental data more as an exercise in applied mathematics than as a chemical venture. One of the tasks in the review of the literature is to check this point. An erroneous chemical model is one of the more serious type of systematic error.

The experimentalist usually selects the variable that he/she finds most appropriate to fulfill the second condition. If the estimated errors in a_{1i}, a_{2i} … are smaller than the error in y_i, the second condition is reasonably well fulfilled. The choice of the error-carrying variable is a matter of choice based on experience, but one must be aware that it has implications, especially in the estimated uncertainty.

The presence of systematic errors is, potentially, the most important source of uncertainty. There is no possibility to handle systematic errors using statistics; statistical methods may indicate their presence, no more. Systematic errors in the chemical model have been mentioned. In addition there may be systematic errors in the methods used. By comparing experimental data obtained with different experimental methods one can obtain an indication of the presence and magnitude of such errors. The systematic errors of this type are accounted for both in the review of the literature and when taking the average of data obtained with different experimental methods. This type of systematic error does not seem to affect the selected data very much, as judged by the usually very good agreement between the equilibrium data obtained using spectroscopic, potentiometric and solubility methods.

The electrode calibration, especially the conversion between measured pH and $-\log_{10}[H^+]$ is an important source of systematic error. The reviewers have when possible corrected this error, as seen in many instances in Appendix A.

The assumption of a normal distribution of the random errors is a choice made in the absence of better alternatives.

Finally, a comment on the weights used in least-squares refinements; this is important because it influences the uncertainty estimate of the equilibrium constants. The weights of individual experimental points can be obtained by repeating the experiment several times and then calculating the average and standard deviation of these data. This procedure is rarely used, instead most experimentalists seem to use unit weight when making a least-squares analysis of their data. However, also in this case there is a weighting of the data by the number of experimental determinations in the parameter range where the different complexes are formed. In order to have comparable uncertainty estimates for the different complexes, one should try to have the same number of experimental data points in the concentration ranges where each of these complexes is predominant; a procedure very rarely used.

As indicated above, the assignment of uncertainties to equilibrium constants is not a straightforward procedure and it is complicated further when there is lack of primary experimental data. The uncertainty estimates given for the individual equilibrium constants reported by the authors and for some cases re-estimated by this review are given in the tables of this and previous reviews. The procedure used to obtain these estimates is given in the original publications and in the Appendix A discussions. However, this uncertainty is still a subjective estimate and to a large extent based on "expert judgment".

C.2   Uncertainty estimates in the selected thermodynamic data

The uncertainty estimate in the selected thermodynamic data is based on the uncertainty of the individual equilibrium constants or other thermodynamic data, calculated as described in the following sections. A weighted average of the individual $\log_{10}K$ values is calculated using the estimated uncertainty of the individual experimental values to assign its weight. The uncertainty in this average is then calculated using the formulae given in the following text. This uncertainty depends on the number of experimental data points – for N data point with the same estimated uncertainty, σ, the uncertainty in the average is $\sigma/\sqrt{N}$. The average and the associated uncertainty reported in the tables of selected data are reported with many more digits than justified only in order to allow the users to back-track the calculations. The reported uncertainty is much smaller than the estimated experimental uncertainty and the users of the tables should look at the discussion of the selected constants in order to get a better estimate of the uncertainty in an experimental determination using a specific method.

One of the objectives of the NEA Thermochemical Data Base (TDB) project is to provide an idea of the uncertainties associated with the data selected in this review. As a rule, the uncertainties define the range within which the corresponding data can be reproduced with a probability of 95% at any place and by any appropriate method. In many cases, the statistical treatment is limited or impossible due to the availability of only one or few data points. A particular problem has to be solved when significant discrepancies occur between different source data. This appendix outlines the statistical procedures, which were used for fundamentally different problems, and explains the philosophy used in this review when statistics were inapplicable. These rules are followed consistently throughout the series of reviews within the TDB Project. Four fundamentally different cases are considered:

1. One source datum available

2. Two or more independent source data available

3. Several data available at different ionic strengths

4. Data at non-standard conditions: Procedures for data correction and recalculation.

C.3   One source datum

The assignment of an uncertainty to a selected value that is based on only one experimental source is a highly subjective procedure. In some cases, the number of data points, on which the selected value is based, allows the use of the "root mean square" [1982TAY] deviation of the data points, X_i, to describe the standard deviation, s_X, associated with the average, $\overline{X}$:

$$s_X = \sqrt{\frac{1}{N-1} \sum_{i=1}^{N} (X_i - \overline{X})^2} \qquad (C.4)$$

The standard deviation, s_X, is thus calculated from the dispersion of the equally weighted data points, X_i, around the average $\overline{X}$, and the probability is 95% that an X_i is within $\overline{X} \pm 1.96\, s_X$, see Taylor [1982TAY] (pp. 244-245). The standard deviation, s_X, is a measure of the precision of the experiment and does not include any systematic errors.

Many authors report standard deviations, s_X, calculated with Eq. (C.4) (but often not multiplied by 1.96), but these do not represent the quality of the reported values in absolute terms. Therefore, it is thus important not to confuse the standard deviation, s_X, with the uncertainty, σ. The latter reflects the reliability and reproducibility of an experimental value and also includes all kinds of systematic errors, s_j, that may be involved. The uncertainty, σ, can be calculated with Eq. (C.5), assuming that the systematic errors are independent.

$$\sigma_X = \sqrt{s_X^2 + \sum_j (s_j^2)} \tag{C.5}$$

The estimation of the systematic errors s_j (which, of course, have to relate to $\overline{X}$ and be expressed in the same units) can only be made by a person who is familiar with the experimental method. The uncertainty, σ, has to correspond to the 95% confidence level preferred in this review. It should be noted that for all the corrections and recalculations made (*e.g.*, temperature or ionic strength corrections) the rules of the propagation of errors have to be followed, as outlined in Section C.6.2.

More often, the determination of s_X is impossible because either only one or two data points are available, or the authors did not report the individual values. The uncertainty σ in the resulting value can still be estimated using Eq. (C.5) assuming that s_X^2 is much smaller than $\sum_j (s_j^2)$, which is usually the case anyway.

C.4   Two or more independent source data

Frequently, two or more experimental data sources are available, reporting experimental determinations of the desired thermodynamic data. In general, the quality of these determinations varies widely, and the data have to be weighted accordingly for the calculation of the mean. Instead of assigning weight factors, the individual source data, X_i, are provided with an uncertainty, σ_i, that also includes all systematic errors and represents the 95% confidence level, as described in Section C.3. The weighted mean $\overline{X}$ and its uncertainty, $\sigma_{\overline{X}}$, are then calculated according to Eqs. (C.6) and (C.7).

$$\overline{X} \equiv \frac{\sum_{i=1}^{N} \left(\frac{X_i}{\sigma_i^2} \right)}{\sum_{i=1}^{N} \left(\frac{1}{\sigma_i^2} \right)} \tag{C.6}$$

$$\sigma_{\overline{X}} = \sqrt{\frac{1}{\sum_{i=1}^{N} \left(\frac{1}{\sigma_i^2} \right)}} \tag{C.7}$$

Eqs. (C.6) and (C.7) may only be used if all the X_i belong to the same parent distribution. If there are serious discrepancies among the X_i, one proceeds as described below under Section C.4.1. It can be seen from Eq. (C.7) that $\sigma_{\overline{X}}$ is directly dependent on the absolute magnitude of the σ_i values, and not on the dispersion of the data points around the mean. This is reasonable because there are no discrepancies among the X_i, and because the σ_i values already represent the 95% confidence level. The selected uncertainty, $\sigma_{\overline{X}}$, will therefore also represent the 95% confidence level.

In cases where all the uncertainties are equal, $\sigma_i = \sigma$, Eqs. (C.6) and (C.7) reduce to Eqs. (C.8) and (C.9).

$$\overline{X} = \frac{1}{N}\sum_{i=1}^{N} X_i \qquad (C.8)$$

$$\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{N}} \qquad (C.9)$$

Example C.1:

Five data sources report values for the thermodynamic quantity, X. The reviewer has assigned uncertainties that represent the 95% confidence level as described in Section C.3.

i	X_i	σ_i
1	25.3	0.5
2	26.1	0.4
3	26.0	0.5
4	24.85	0.25
5	25.0	0.6

According to Eqs.(C.6) and (C.7), the following result is obtained:

$$\overline{X} = (25.3 \pm 0.2).$$

The calculated uncertainty, $\sigma_{\overline{X}} = 0.2$, appears relatively small, but is statistically correct, as the values are assumed to follow a Gaussian distribution. As a consequence of Eq. (C.7), $\sigma_{\overline{X}}$ will always come out smaller than the smallest σ_i. Assuming $\sigma_4 = 0.10$ instead of 0.25 would yield $\overline{X} = (25.0 \pm 0.1)$ and $\sigma_4 = 0.60$ would result in $\overline{X} = (25.6 \pm 0.2)$. In fact, the values $(X_i \pm \sigma_i)$ in this example are at the limit of consistency, *i.e.*, the range $(X_4 \pm \sigma_4)$ does not overlap with the ranges $(X_2 \pm \sigma_2)$ and $(X_3 \pm \sigma_3)$. There might be a better way to solve this problem. Three possible choices seem more reasonable:

i.   The uncertainties, σ_i, are reassigned because they appear too optimistic after further consideration. Some assessments may have to be reconsidered and the uncertainties reassigned. For example, multiplying all the σ_i by 2 would yield $\overline{X} = (25.3 \pm 0.3)$.

ii. If reconsideration of the previous assessments gives no evidence for reassigning the X_i and σ_i (95% confidence level) values listed above, the statistical conclusion will be that all the X_i do not belong to the same parent distribution and cannot therefore be treated in the same group (*cf.* item iii below for a non-statistical explanation). The values for $i = 1$, 4 and 5 might be considered as belonging to Group A and the values for $i = 2$ and 3 to Group B. The weighted average of the values in Group A is X_A ($i = 1, 4, 5) = (24.95 \pm 0.21)$ and of those in Group B, X_B ($i = 2, 3) = (26.06 \pm 0.31)$, the second digit after the decimal point being carried over to avoid loss of information. The selected value is now determined as

described below under "Discrepancies" (Section C.4.1, Case I). X_A and X_B are averaged (straight average, there is no reason for giving X_A a larger weight than X_B), and $\sigma_{\bar{X}}$ is chosen in such a way that it covers the complete ranges of expectancy of X_A and X_B. The selected value is then $\bar{X} = (25.5 \pm 0.9)$.

iii. Another explanation could be that unidentified systematic errors are associated with some values. If this seems likely to be the case, there is no reason for splitting the values up into two groups. The correct way of proceeding would be to calculate the unweighted average of all the five points and assign an uncertainty that covers the whole range of expectancy of the five values. The resulting value is then $\bar{X} = (25.45 \pm 1.05)$, which is rounded according to the rules in Section C.6.3 to $\bar{X} = (25.4 \pm 1.1)$.

C.4.1   Discrepancies

Two data are called discrepant if they differ significantly, *i.e.*, their uncertainty ranges do not overlap. In this context, two cases of discrepancies are considered. Case I: Two significantly different source data are available. Case II: Several, mostly consistent source data are available, one of them being significantly different, *i.e.*, an "outlier".

Case I. Two discrepant data: This is a particularly difficult case because the number of data points is obviously insufficient to allow the preference of one of the two values. If there is absolutely no way of discarding one of the two values and selecting the other, the only solution is to average the two source data in order to obtain the selected value, because the underlying reason for the discrepancy must be unrecognised systematic errors. There is no point in calculating a weighted average, even if the two source data have been given different uncertainties, because there is obviously too little information to give even only limited preference to one of the values. The uncertainty, $\sigma_{\bar{X}}$, assigned to the selected mean, $\bar{X}$, has to cover the range of expectation of both source data, X_1, X_2, as shown in Eq.(C.10),

$$\sigma_{\bar{X}} = \left| X_i - \bar{X} \right| + \sigma_{max} \qquad (C.10)$$

where $i = 1, 2$, and σ_{max} is the larger of the two uncertainties σ_i, see Example C.1.ii and Example C.2.

Example C.2:

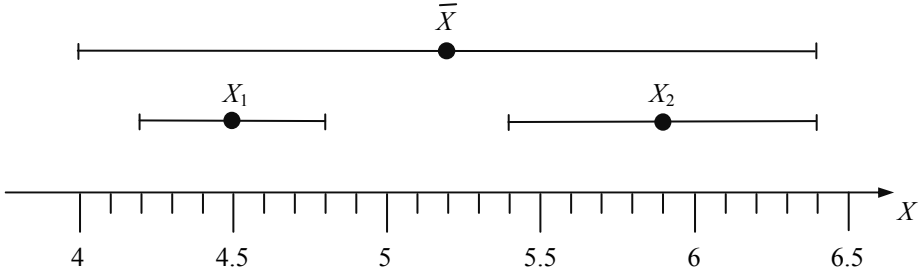The following credible source data are given:

$$X_1 = (4.5 \pm 0.3)$$

$$X_2 = (5.9 \pm 0.5).$$

The uncertainties have been assigned by the reviewer. Both experimental methods are satisfactory and there is no justification to discard one of the data. The selected value is then:

$$\bar{X} = (5.2 \pm 1.2).$$

Figure C-1: Illustration for Example C.2

Case II. Outliers: This problem can often be solved by either discarding the outlying data point, or by providing it with a large uncertainty to lower its weight. If, however, the outlying value is considered to be of high quality and there is no reason to discard all the other data, this case is treated in a way similar to Case I. Example C.3 illustrates the procedure.

Example C.3:

The following data points are available. The reviewer has assigned the uncertainties and sees no justification for any change.

i	X_i	σ_i
1	4.45	0.35
2	5.9	0.5
3	5.7	0.4
4	6.0	0.6
5	5.2	0.4

　　　　There are two data sets that, statistically, belong to different parent distributions, A and B. According to Eqs. (C.6) and (C.7), the following average values are found for the two groups: $X_A(i =1) = (4.45 \pm 0.35)$ and $X_B(i = 2, 3, 4, 5) = (5.62 \pm 0.23)$. The selected value will be the straight average of X_A and X_B, analogous to Example C.1:

$$\bar{X} = (5.0 \pm 0.9).$$

C.5   Several data at different ionic strengths

The extrapolation procedure for aqueous equilibria used in this review is the specific ion interaction model outlined in Appendix B. The objective of this review is to provide selected data sets at standard conditions, *i.e.,* among others, at infinite dilution for aqueous species. Equilibrium constants determined at different ionic strengths can, according to the specific ion interaction equations, be extrapolated to $I = 0$ with a linear regression model, yielding as the intercept the desired equilibrium constant at $I = 0$, and as the slope the stoichiometric sum of the ion interaction coefficients, $\Delta\varepsilon$. The ion interaction coefficient of the target species can usually be extracted from $\Delta\varepsilon$ and would be listed in the corresponding table of Appendix B.

The available source data may sometimes be sparse or may not cover a sufficient range of ionic strengths to allow a proper linear regression. In this case, the correction to $I = 0$ should be carried out according to the procedure described in Section C.6.1.

If sufficient data are available at different ionic strengths and in the same inert salt medium, a weighted linear regression will be the appropriate way to obtain both the constant at $I = 0$, $\bar{X}°$, and $\Delta\varepsilon$. The first step is the conversion of the ionic strength from the frequently used molar (mol·dm^{-3}, M) to the molal (mol·kg^{-1}, m) scale, as described in Section II.2. The second step is the assignment of an uncertainty, σ_i, to each data point X_i at the molality, $m_{k,i}$, according to the rules described in Section C.3. A large number of commercial and public domain computer programs and routines exist for weighted linear regressions. The subroutine published by Bevington [1969BEV] (pp.104 − 105) has been used for the calculations in the examples of this appendix. Eqs. (C.11) through (C.15) present the equations that are used for the calculation of the intercept $\bar{X}°$ and the slope $-\Delta\varepsilon$:

$$\bar{X}° = \frac{1}{\Delta}\left(\sum_{i=1}^{N}\frac{m_{k,i}^2}{\sigma_i^2}\sum_{i=1}^{N}\frac{X_i}{\sigma_i^2} - \sum_{i=1}^{N}\frac{m_{k,i}}{\sigma_i^2}\sum_{i=1}^{N}\frac{m_{k,i}X_i}{\sigma_i^2}\right) \tag{C.11}$$

$$-\Delta\varepsilon = \frac{1}{\Delta}\left(\sum_{i=1}^{N}\frac{1}{\sigma_i^2}\sum_{i=1}^{N}\frac{m_{k,i}X_i}{\sigma_i^2} - \sum_{i=1}^{N}\frac{m_{k,i}}{\sigma_i^2}\sum_{i=1}^{N}\frac{X_i}{\sigma_i^2}\right) \tag{C.12}$$

$$\sigma_{\bar{X}°} = \sqrt{\frac{1}{\Delta}\sum_{i=1}^{N}\frac{m_{k,i}^2}{\sigma_i^2}} \tag{C.13}$$

$$\sigma_{\Delta\varepsilon} = \sqrt{\frac{1}{\Delta}\sum_{i=1}^{N}\frac{1}{\sigma_i^2}} \tag{C.14}$$

where $\quad\quad\quad \Delta = \sum_{i=1}^{N}\frac{1}{\sigma_i^2}\sum_{i=1}^{N}\frac{m_{k,i}^2}{\sigma_i^2} - \left(\sum_{i=1}^{N}\frac{m_{k,i}}{\sigma_i^2}\right)^2. \tag{C.15}$

In this way, the uncertainties, σ_i, are not only used for the weighting of the data in Eqs. (C.11) and (C.12), but also for the calculation of the uncertainties, $\sigma_{\bar{X}^\circ}$ and $\sigma_{\Delta\varepsilon}$, in Eqs. (C.13) and (C.14). If the σ_i represents the 95% confidence level, $\sigma_{\bar{X}^\circ}$ and $\sigma_{\Delta\varepsilon}$ will also do so. In other words, the uncertainties of the intercept and the slope do not depend on the dispersion of the data points around the straight line, but rather directly on their absolute uncertainties, σ_i.

Example C.4:

Ten independent determinations of the equilibrium constant, $\log_{10} {}^*\beta$, for the reaction:

$$UO_2^{2+} + HF(aq) \rightleftharpoons UO_2F^+ + H^+ \tag{C.16}$$

are available in $HClO_4/NaClO_4$ media at different ionic strengths. Uncertainties that represent the 95% confidence level have been assigned by the reviewer. A weighted linear regression, ($\log_{10} {}^*\beta + 2D$) *vs.* m_k, according to the formula, $\log_{10} {}^*\beta$ (C.16) + $2D = \log_{10} {}^*\beta^\circ$ (C.16) $- \Delta\varepsilon\, m_k$, will yield the correct values for the intercept, $\log_{10} {}^*\beta^\circ$ (C.16), and the slope, $\Delta\varepsilon$. In this case, m_k corresponds to the molality of ClO_4^-. D is the Debye-Hückel term, *cf.* Appendix B.

i	$m_{ClO_4^-,i}$	$\log_{10} {}^*\beta + 2D$	σ_i
1	0.05	1.88	0.10
2	0.25	1.86	0.10
3	0.51	1.73	0.10
4	1.05	1.84	0.10
5	2.21	1.88	0.10
6	0.52	1.89	0.11
7	1.09	1.93	0.11
8	2.32	1.78	0.11
9	2.21	2.03	0.10
10	4.95	2.00	0.32

The results of the linear regression are:

$$\text{intercept} = (1.837 \pm 0.054) = \log_{10} {}^*\beta^\circ \text{ (C.16)}$$

$$\text{slope} = (0.029 \pm 0.036) = -\Delta\varepsilon$$

Calculation of the ion interaction coefficient $\varepsilon(UO_2F^+, ClO_4^-) = \Delta\varepsilon + \varepsilon(UO_2^{2+}, ClO_4^-) - \varepsilon(H^+, ClO_4^-)$: from $\varepsilon(UO_2^{2+}, ClO_4^-) = (0.46 \pm 0.03)$ kg·mol^{-1}, $\varepsilon(H^+, ClO_4^-) = (0.14 \pm 0.02)$ kg·mol^{-1} (see Appendix B) and the slope of the linear

regression, $\Delta\varepsilon = -(0.03 \pm 0.04)$ kg·mol^{-1}, it follows that $\varepsilon(UO_2F^+, ClO_4^-) = (0.29 \pm 0.05)$ kg·mol^{-1}. Note that the uncertainty ($\pm 0 05$) kg·mol^{-1} is obtained based on the rules of error propagation as described in Section C.6.2:

$$\sigma = \sqrt{(0.04)^2 + (0.03)^2 + (0.02)^2}$$

The resulting selected values are thus:

$$\log_{10} {}^*\beta° (C.16) = (1.84 \pm 0.05)$$

$$\varepsilon(UO_2F^+, ClO_4^-) = (0.29 \pm 0.05) \text{ kg·mol}^{-1}.$$

C.5.1 Discrepancies or insufficient number of data points

Discrepancies are principally treated as described in Section C.4. Again, two cases can be defined. Case I: Only two data points are available. Case II: An "outlier" cannot be discarded. If only one data point is available, the procedure for correction to zero ionic strength outlined in Section C.6 should be followed.

Case I. Too few molalities: If only two source data are available, there will be no straightforward way to decide whether or not these two data points belong to the same parent distribution unless either the slope of the straight line is known or the two data refer to the same ionic strength. Drawing a straight line right through the two data points is an inappropriate procedure because all the errors associated with the two source data would accumulate and may lead to highly erroneous values of $\log_{10} K°$ and $\Delta\varepsilon$. In this case, an ion interaction coefficient for the key species in the reaction in question may be selected by analogy (charge is the most important parameter), and a straight line with the slope $\Delta\varepsilon$ as calculated may then be drawn through each data point. If there is no reason to discard one of the two data points based on the quality of the underlying experiment, the selected value will be the unweighted average of the two standard state data point obtained by this procedure, and its uncertainty must cover the entire range of expectancy of the two values, analogous to Case I in Section C.4. It should be mentioned that the ranges of expectancy of the corrected values at $I = 0$ are given by their uncertainties, which are based on the uncertainties of the source data at $I \neq 0$ and the uncertainty in the slope of the straight line. The latter uncertainty is not an estimate, but is calculated from the uncertainties in the ion interaction coefficients involved, according to the rules of error propagation outlined in Section C.6.2. The ion interaction coefficients estimated by analogy are listed in the table of selected ion interaction coefficients (Appendix B), but they are flagged as estimates.

Case II. Outliers and inconsistent data sets: This case includes situations where it is difficult to decide whether or not a large number of points belong to the same parent distribution. There is no general rule on how to solve this problem, and decisions are left to the judgment of the reviewer. For example, if eight data points follow a straight line reasonably well and two lie way out, it may be justified to discard the "out-

liers". If, however, the eight points are scattered considerably and two points are just a bit further out, one can probably not consider them as "outliers". It depends on the particular case and on the judgment of the reviewer whether it is reasonable to increase the uncertainties of the data to reach consistency, or whether the slope, $\Delta\varepsilon$, of the straight line should be estimated by analogy.

Example C.5:

Six reliable determinations of the equilibrium constant, $\log_{10}\beta$, of the reaction:

$$UO_2^{2+} + SCN^- \rightleftharpoons UO_2SCN^+ \tag{C.17}$$

are available in different electrolyte media:

$$I_c = 0.1 \text{ M (KNO}_3)\qquad \log_{10}\beta \text{ (C.17)} = (1.19 \pm 0.03)$$

$$I_c = 0.33 \text{ M (KNO}_3)\qquad \log_{10}\beta \text{ (C.17)} = (0.90 \pm 0.10)$$

$$I_c = 1.0 \text{ M (NaClO}_4)\qquad \log_{10}\beta \text{ (C.17)} = (0.75 \pm 0.03)$$

$$I_c = 1.0 \text{ M (NaClO}_4)\qquad \log_{10}\beta \text{ (C.17)} = (0.76 \pm 0.03)$$

$$I_c = 1.0 \text{ M (NaClO}_4)\qquad \log_{10}\beta \text{ (C.17)} = (0.93 \pm 0.03)$$

$$I_c = 2.5 \text{ M (NaNO}_3)\qquad \log_{10}\beta \text{ (C.17)} = (0.72 \pm 0.03)$$

The uncertainties are assumed to represent the 95% confidence level. From the values at $I_c = 1$ M, it can be seen that there is a lack of consistency in the data, and that a linear regression similar to that shown in Example C.4 would be inappropriate. Instead, the use of $\Delta\varepsilon$ values from reactions of the same charge type is encouraged. Analogies with $\Delta\varepsilon$ are more reliable than analogies with single ε values due to canceling effects. For the same reason, the dependency of $\Delta\varepsilon$ on the type of electrolyte is often smaller than for single ε values.

A reaction of the same charge type as Reaction (C.17), and for which $\Delta\varepsilon$ is well known, is:

$$UO_2^{2+} + Cl^- \rightleftharpoons UO_2Cl^+. \tag{C.18}$$

The value of $\Delta\varepsilon\,(\text{C.18}) = -(0.25 \pm 0.02)$ kg $\cdot$ mol^{-1} was obtained from a linear regression using 16 experimental values between $I_c = 0$ 1 M and $I_c = 3$ M Na(Cl,ClO$_4$) [1992GRE/FUG]. It is thus assumed that:

$$\Delta\varepsilon\,(\text{C.17}) = \Delta\varepsilon\,(\text{C.18}) = -(0.25 \pm 0.02) \text{ kg} \cdot \text{mol}^{-1}.$$

The correction of $\log_{10}\beta$ (C.17) to $I_c = 0$ is done using the specific ion interaction equation, *cf.* TDB-2, which uses molal units:

$$\log_{10}\beta + 4D = \log_{10}\beta^\circ - \Delta\varepsilon\,I_m. \tag{C.19}$$

D is the Debye-Hückel term and I_m the ionic strength converted to molal units by using the conversion factors listed in Table II-5. The following list gives the details

of this calculation. The resulting uncertainties in $\log_{10}\beta$ are obtained based on the rules of error propagation as described in Section C.6.2.

Table C-1: Details of the calculation of equilibrium constant corrected to $I = 0$, using (C.19).

I_m	electrolyte	$\log_{10}\beta$	$4D$	$\Delta\varepsilon\, I_m$	$\log_{10}\beta°$
0.101	KNO$_3$	(1.19 ± 0.03)	0.438	-0.025	$(1.68 \pm 0.03)^a$
0.335	KNO$_3$	(0.90 ± 0.10)	0.617	-0.084	$(1.65 \pm 0.10)^a$
1.050	NaClO$_4$	(0.75 ± 0.03)	0.822	-0.263	(1.31 ± 0.04)
1.050	NaClO$_4$	(0.76 ± 0.03)	0.822	-0.263	(1.32 ± 0.04)
1.050	NaClO$_4$	(0.93 ± 0.03)	0.822	-0.263	(1.49 ± 0.04)
2.714	NaNO$_3$	(0.72 ± 0.03)	0.968	-0.679	$(1.82 \pm 0.13)^a$

a: These values were corrected for the formation of the nitrate complex, $UO_2NO_3^+$, by using $\log_{10}K(UO_2NO_3^+) = (0.30 \pm 0.15)$ [1992GRE/FUG].

As was expected, the resulting values, $\log_{10}\beta°$, are inconsistent and have therefore to be treated as described in Case I of Section C.4. That is, the selected value will be the unweighted average of $\log_{10}\beta°$, and its uncertainty will cover the entire range of expectancy of the six values. A weighted average would only be justified if the six values of $\log_{10}\beta°$ were consistent. The result is:

$$\log_{10}\beta° = (1.56 \pm 0.39).$$

C.6   Procedures for data handling

C.6.1   Correction to zero ionic strength

The correction of experimental data to zero ionic strength is necessary in all cases where a linear regression is impossible or appears inappropriate. The method used throughout the review is the specific ion interaction equations described in detail in Appendix B. Two variables are needed for this correction, and both have to be provided with an uncertainty at the 95% confidence level: the experimental source value, $\log_{10}K$ or $\log_{10}\beta$, and the stoichiometric sum of the ion interaction coefficients, $\Delta\varepsilon$. The ion interaction coefficients (see Tables B-4, B-5, B-6 and B-7 of Appendix B) required to calculate $\Delta\varepsilon$ may not all be known. Missing values therefore need to be estimated. It is recalled that the electric charge has the most significant influence on the magnitude of the ion interaction coefficients, and that it is in general more reliable to estimate $\Delta\varepsilon$ from known reactions of the same charge type, rather than to estimate single ε values. The uncertainty of the corrected value at $I = 0$ is calculated by taking into account the

propagation of errors, as described below. It should be noted that the ionic strength is frequently given in moles per dm^3 of solution (molar, M) and has to be converted to moles per kg H$_2$O (molal, m), as the model requires. Conversion factors for the most common inert salts are given in Table II.5.

Example C.6:

For the equilibrium constant of the reaction:

$$M^{3+} + 2 H_2O(l) \rightleftharpoons M(OH)_2^+ + 2 H^+ , \qquad (C.20)$$
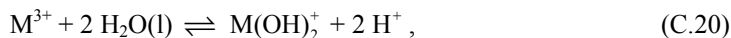
only one credible determination in 3 M NaClO$_4$ solution is known to be, $\log_{10} {}^*\beta$ (C.20) $= -6.31$, to which an uncertainty of ± 0.12 has been assigned. The ion interaction coefficients are as follows:

$$\varepsilon(M^{3+},\ ClO_4^-) = (0.56 \pm 0.03)\ kg\cdot mol^{-1},$$

$$\varepsilon(M(OH)_2^+, ClO_4^-) = (0.26 \pm 0.11)\ kg\cdot mol^{-1},$$

$$\varepsilon(H^+, ClO_4^-) = (0.14 \pm 0.02)\ kg\cdot mol^{-1}.$$

The values of $\Delta\varepsilon$ and $\sigma_{\Delta\varepsilon}$ can be obtained readily (*cf.* Eq. (C.22)):

$$\Delta\varepsilon = \varepsilon(M(OH)_2^+, ClO_4^-) + 2\varepsilon(H^+,\ ClO_4^-) - \varepsilon(M^{3+},\ ClO_4^-) = -0.22\ kg\cdot mol^{-1},$$

$$\sigma_{\Delta\varepsilon} = \sqrt{(0.11)^2 + (2\times 0.02)^2 + (0.03)^2} = 0.12\ kg\cdot mol^{-1}.$$

The two variables are thus:

$$\log_{10} {}^*\beta\ (C.20) = -(6.31 \pm 0.12),$$

$$\Delta\varepsilon = -(0.02 \pm 0.12)\ kg\cdot mol^{-1}.$$

According to the specific ion interaction model the following equation is used to correct for ionic strength for the reaction considered here:

$$\log_{10} {}^*\beta\ (C.20) + 6D = \log_{10} {}^*\beta^\circ\ (C.20) - \Delta\varepsilon\, m_{ClO_4^-}$$

D is the Debye-Hückel term:

$$D = \frac{0.509\sqrt{I_m}}{(1+1.5\sqrt{I_m})}.$$

The ionic strength, I_m, and the molality, $m_{ClO_4^-}$ ($I_m \approx m_{ClO_4^-}$), have to be expressed in molal units, 3 M NaClO$_4$ corresponding to 3.5 m NaClO$_4$ (see Section II.2), giving $D = 0.25$. This results in:

$$\log_{10} {}^*\beta^\circ\ (C.20) = -4.88.$$

The uncertainty in $\log_{10} {}^*\beta^\circ$ is calculated from the uncertainties in $\log_{10} {}^*\beta$ and $\Delta\varepsilon$ (*cf.* Eq. (C.22)):

$$\sigma_{\log_{10}{}^*\beta^\circ} = \sqrt{\sigma^2_{\log_{10}{}^*\beta} + (m_{ClO_4^-}\sigma_{\Delta\varepsilon})^2} = \sqrt{(0.12)^2 + (3.5\times0.12)^2} = 0.44$$

The selected, rounded value is:

$$\log_{10}{}^*\beta^\circ \text{ (C.20)} = -(4.9 \pm 0.4).$$

C.6.2 Propagation of errors

Whenever data are converted or recalculated, or other algebraic manipulations are performed that involve uncertainties, the propagation of these uncertainties has to be taken into account in a correct way. A clear outline of the propagation of errors is given by Bevington [1969BEV]. A simplified form of the general formula for error propagation is given by Eq.(C.21), supposing that X is a function of $Y_1, Y_2,\ldots,Y_N$.

$$\sigma^2_X = \sum_{i=1}^{N}\left(\frac{\partial X}{\partial Y_i}\sigma_{Y_i}\right)^2 \tag{C.21}$$

Eq. (C.21) can be used only if the variables, $Y_1, Y_2,\ldots,Y_N$, are independent or if their uncertainties are small, *i.e.*, the covariances can be disregarded. One of these two assumptions can almost always be made in chemical thermodynamics, and Eq. (C.21) can thus almost universally be used in this review. Eqs. (C.22) through (C.26) present explicit formulas for a number of frequently encountered algebraic expressions, where c, c_1, c_2 are constants.

$$X = c_1 Y_1 \pm c_2 Y_2 : \qquad\qquad \sigma^2_X = (c_1\sigma_{Y_1})^2 + (c_2\sigma_{Y_2})^2 \tag{C.22}$$

$$X = \pm c Y_1 Y_2 \text{ and } X = \pm\frac{cY_1}{Y_2} : \qquad \left(\frac{\sigma_X}{X}\right)^2 = \left(\frac{\sigma_{Y_1}}{Y_1}\right)^2 + \left(\frac{\sigma_{Y_2}}{Y_2}\right)^2 \tag{C.23}$$

$$X = c_1 Y^{\pm c_2} : \qquad\qquad \frac{\sigma_X}{X} = c_2\frac{\sigma_Y}{Y} \tag{C.24}$$

$$X = c_1 e^{\pm c_2 Y} : \qquad\qquad \frac{\sigma_X}{X} = c_2\sigma_Y \tag{C.25}$$

$$X = c_1 \ln(\pm c_2 Y) : \qquad\qquad \sigma_X = c_1\frac{\sigma_Y}{Y} \tag{C.26}$$

Example C.7:

A few simple calculations illustrate how these formulas are used. The values have not been rounded.

Eq. (C.22) : $\Delta_r G_m = 2\cdot[-(277.4 \pm 4.9)] \text{ kJ·mol}^{-1} - [-(467.3 \pm 6.2)] \text{ kJ·mol}^{-1}$

$$= -(87.5 \pm 11.6) \text{ kJ·mol}^{-1}.$$

Eq. (C.23) : $K = \dfrac{(0.038\pm0.002)}{(0.0047\pm0.0005)} = (8.09\pm0.96)$

Eq. (C.24) : $K = 4 \cdot (3.75 \pm 0.12)^3 = (210.9 \pm 20.3)$

Eq. (C.25) : $K^\circ = e^{\frac{-\Delta_r G_m^\circ}{RT}}$; $\Delta_r G_m^\circ = -(2.7 \pm 0.3) \text{ kJ} \cdot \text{mol}^{-1}$

$$R = 8.3145 \text{ J} \cdot \text{K}^{-1} \cdot \text{mol}^{-1}$$

$$T = 298.15 \text{ K}$$

$$K^\circ = (2.97 \pm 0.36).$$

Note that powers of 10 have to be reduced to powers of e, *i.e.*, the variable has to be multiplied by $\ln(10)$, *e.g.*,

$\log_{10} K = (2.45 \pm 0.10); \; K = 10^{\log_{10} K} = e^{(\ln(10) \cdot \log_{10} K)} = (282 \pm 65).$

Eq. (C.26) : $\Delta_r G_m^\circ = -RT \ln K^\circ$; $K^\circ = (8.2 \pm 1.2) \times 10^6$

$$R = 8.3145 \text{ J} \cdot \text{K}^{-1} \cdot \text{mol}^{-1}$$

$$T = 298.15 \text{ K}$$

$\Delta_r G_m^\circ = -(39.46 \pm 0.36) \text{ kJ} \cdot \text{mol}^{-1}$

$\ln K^\circ = (15.92 \pm 0.15)$

$\log_{10} K^\circ = \ln K^\circ / \ln(10) = (6.91 \pm 0.06).$

Again, it can be seen that the uncertainty in $\log_{10} K^\circ$ cannot be the same as in $\ln K^\circ$. The constant conversion factor of $\ln(10) = 2.303$ is also to be applied to the uncertainty.

C.6.3   Rounding

The standard rules to be used for rounding are:

1. When the digit following the last digit to be retained is less than 5, the last digit retained is kept unchanged.

2. When the digit following the last digit to be retained is greater than 5, the last digit retained is increased by 1.

3. When the digit following the last digit to be retained is 5 and

   a) there are no digits (or only zeroes) beyond the 5, an odd digit in the last place to be retained is increased by 1 while an even digit is kept unchanged.

   b) other non-zero digits follow, the last digit to be retained is increased by 1, whether odd or even.

This procedure avoids introducing a systematic error from always dropping or not dropping a 5 after the last digit retained.

When adding or subtracting, the result is rounded to the number of decimal places (not significant digits) in the term with the least number of places. In multiplication and division, the results are rounded to the number of significant digits in the term with the least number of significant digits.

In general, all operations are carried out in full, and only the final results are rounded, in order to avoid the loss of information from repeated rounding. For this reason, several additional digits are carried in all calculations until the final selected data set is developed, and only then are data rounded.

C.6.4   Significant digits

The uncertainty of a value basically defines the number of significant digits a value should be given.

Example: (3.478 ± 0.008)

(3.48 ± 0.01)

(2.8 ± 0.4)

(10 ± 1)

(105 ± 20).

In the case of auxiliary data or values that are used for later calculations, it is often inconvenient to round to the last significant digit. In the value (4.85 ± 0.26), for example, the "5" is close to being significant and should be carried along a recalculation path in order to avoid loss of information. In particular cases, where the rounding to significant digits could lead to slight internal inconsistencies, digits with no significant meaning in absolute terms are nevertheless retained. The uncertainty of a selected value always contains the same number of digits after the decimal point as the value itself.