Strategy for obtaining consensus solubility values with measurement uncertainties at given temperatures from published data

W. Earle Waghorne
University College Dublin
D. Brynn Hibbert
b.hibbert@unsw.edu.au

UNSW Sydney

Research Article

Keywords: Solubility, Measurement uncertainty, Consensus values, DerSimonian-Laird

Posted Date: June 27th, 2025

DOI: https://doi.org/10.21203/rs.3.rs-6772711/v1

Additional Declarations: No competing interests reported.

Version of Record: A version of this preprint was published at Journal of Solution Chemistry on October 16th, 2025. See the published version at https://doi.org/10.1007/s10953-025-01519-3.

Strategy for obtaining consensus solubility values with measurement uncertainties at given temperatures from published data

W. Earle Waghorne, University College Dublin, Dublin, Ireland. ORCID: 0000-0002-5271-0392

D. Brynn Hibbert*, UNSW, Sydney, Australia. ORCID: 0000-0001-9210-2941

* Corresponding Author: b.hibbert@unsw.edu.au

Keywords: Solubility; Measurement uncertainty; Consensus values; DerSimonian-Laird;

Abstract

A strategy for determining consensus solubilities with measurement uncertainties at common temperatures from data reported at varying temperatures is presented. The approach involves fitting the individual data sets to an appropriate equation of solubility (or log solubility) as a function of temperature. This allows estimation of the standard uncertainty of each data point and, if required, interpolated values of the solubility, with uncertainties, at the temperatures of interest. Between-laboratory uncertainty is assessed by the DerSimonian−Laird procedure.

The strategy is applicable to any measured quantity where its variation with experimental variables can be represented by a robust equation.

1 Introduction

When tasked with critically evaluating published data it is rare to have well-described results with a properly-constructed uncertainty budget that will allow the evaluator to decide whether, and then how, to include the data in the consensus value. A recent guide [1] sets out the principles of data evaluation, which accepts that a desire for perfection would not leave much data to evaluate. However, by using their skill, evaluators can fill in some gaps to produce the best possible consensus that can be used by chemists. The task is, therefore, to provide a best estimate of the value of the target quantity with a reasonable standard measurement uncertainty, and a description of how this was arrived at.

In this paper we provide an approach for dealing with solubility data from a number of reports that measure solubility as a function of temperature, but without estimates of measurement uncertainty. First, the data from each report are fitted to an appropriate equation (van't Hoff or Apelblat for example) which then yields the solubility at any desired temperature. The measurement uncertainties of each reported set of solubilities are obtained from the statistics of the linear regression. For a particular temperature the consensus value and its standard uncertainty are calculated from fitted values weighted by their uncertainties. Between-laboratory uncertainty is then included by the DerSimonian–Laird method.

We discuss approaches to decide which equations are most appropriate for fitting the data and the implications of the use of the data to obtain consensus results with associated standard uncertainties.

Two systems have been chosen to illustrate the method and highlight issues, the solubility of hexanedioic acid (adipic acid) in ethanoic acid (acetic acid) and in propanone.

2 Terms and Symbols

The language of metrology can be somewhat obscure and confusing. Here we describe, for the practicing chemist, terms and symbols used in this paper. We make *measurements*[1] to obtain the *value of a quantity* (the value of the *measurand*). A *measurement result* is considered to be our best estimate of the *true value* of the measurand and consists of a number, a *unit*, and an expression of *uncertainty*. Measurement uncertainty has recently been defined as “doubt about the true value of the measurand that remains after making a measurement”. It may be quantified in a number of ways, most commonly as a single value called the *standard uncertainty* which has the symbol u. This is a standard deviation of a distribution (often taken as a normal distribution) that represents the possible spread of values of the measurand [2].

A normal distribution has mean μ and standard deviation σ. Estimates of these quantities may be written $\hat{\mu}$ and $\hat{\sigma}$. (Note that a hat ˆ above a symbol indicates an estimated value but is often omitted). For a set of real data, the mean (average) is written $\bar{x}$, where x is the symbol for the quantity of interest, and its standard deviation, calculated from the set of data, is s_x. The symbol s is also used for the standard error of coefficients of a regression (s_a and s_b), and $s_{y/x}$ for the standard error of the regression. It is common to express the uncertainty as an *expanded uncertainty* (symbol U) which is the *combined standard uncertainty*[2] multiplied by a constant, where the interval $x \pm U$, is a region in which the true value of the measurand x is expected to lie with a stated probability. For a normal (Gaussian) distribution ± $2u$ gives about a 95% probability and ± $3u$ gives about a 99.7% probability. Note that ± U is not the confidence interval calculated from the standard deviation of a number of repeats. The standard uncertainty should include this repeatability, but also wider random effects and systematic errors should be included in a full *uncertainty budget*. In the absence of sufficient information to construct such an uncertainty budget in a

---

[1] In this section, on first occurrence, terms in italics indicate terms defined in the International Vocabulary of Metrology (VIM) [3] and are hyperlinked to definitions in the on-line version (https://jcgm.bipm.org/vim/en/).

[2] ‘Combined standard uncertainty’ is used to describe a measurement uncertainty obtained from a number of individual standard uncertainties in a measurement model. Its use is likely to be discontinued.

single laboratory, the consensus method outlined below does allow laboratories' systematic effects to be accounted for.

3 Experimental Data

Experimentally, the solubilities in the examples chosen were measured using two approaches: (1) measurement of the composition of a solution saturated at a specific temperature by gravimetry with evaporation of the solvent or by analysis of the saturated solution by acid-base titration, liquid chromatography, spectrophotometry or heat flow calorimetry or (2) measurement of the saturation temperature of a system of known composition by turbidometry with laser detection or differential scanning calorimetry. Results from all of these methods are generally in good agreement.

A complicating factor in evaluating these solubilities is that measurements of the saturation temperature provide solubilities at essentially random temperatures, so that comparison of solubilities at a particular temperature usually requires interpolation via regression analysis.

A second complication is that the authors provide no uncertainty budgets for the measured amount fraction solubilities. Rather, differences between the experimental values and those recovered from fitting the data to one or more model equations are reported. The lack of a properly formed uncertainty in the published work may be at least partly overcome for the consensus value by combining the within-laboratory random errors from the fits of the linear regressions of the data sets considered, and then adding an estimate of the between laboratory error.

4 Theory

4.1 Choosing data sets for consensus analysis

The evaluator will use their best judgement to triage the available published data [1]. If a data set is considered for inclusion in the consensus, then it is recommended that any calculated solubilities are interpolations within the temperature range of the set. Extrapolating outside the fitted data can be dangerous and should only be contemplated when the evaluator believes the linear relation still holds with an acceptable uncertainty.

4.2 Regression on individual data sets

Fitting experimental data to a measurement model both allows interpolation of solubilities to a common temperature and provides estimates of the random uncertainty in the experimental values.

Two equations, the van't Hoff equation:

$$\ln S = A + B/T \qquad (1)$$

and the Apelblat equation:

$$\ln S = A + B/T + C \ln T \qquad (2)$$

are commonly used to represent the temperature dependence of the solubilities of solids, where *S* is the amount fraction of the solute in a saturated solution (solubility) and *T* the thermodynamic temperature. *A*, *B* and, for Eq. 2, *C*, are coefficients of the respective models. (Note *S* is used here for solubility to avoid confusion with *x* being the independent variable in a regression).

Mathematically, Eq. 1 is simply Eq. 2 in the case where *C* is set to zero. The decision of whether to treat the data using Eq. 1 or Eq.2 was based on fitting the data to Eq. 2 and then testing the statistical significance of the coefficient *C* with a two-tail Student-t test on *C* divided by its standard error: C/s_C with $n - 3$ degrees of freedom. The null hypothesis is that the coefficient is not significantly different from zero and a probability of this hypothesis of greater than 0.05 indicates that the use of the additional term is not justified. In this case Eq. 1 equation is preferred.

The van't Hoff equation (Eq. 1) is a simple linear model of the form $Y = a + b X$, where, in this discussion, $Y = \ln S$ and $X = 1/T$. The uncertainty budgets for the temperatures of the measurements range from 0.05 K [3] to 0.13 K [4], which, given the temperature dependences of the solubilities, would affect the measured ln *S* in the fourth significant figure, well below the experimental uncertainty; so that the uncertainty in the temperature was neglected in analyses of the solubilities. This allows the use of a classical least squares approach to fitting the data, which assumes only random, homoscedastic[3]error in *Y*, i.e. in the measured solubility.

We can usefully rewrite Eq. 1, as

$$\ln S = A' + B'\left((1/T) - \overline{(1/T)}\right) \tag{3}$$

where *A'* estimates $\ln \bar{S}$ and $\ln \bar{S}$ and $\overline{(1/T)}$ are the means of ln *S* and 1/*T* respectively. This sidesteps the correlation between the coefficients *A* and *B*, leaving the uncorrelated coefficients *A'* and *B'*. Recall that in the classic least squares model there is no uncertainty in *X* and therefore no uncertainty in $\bar{X}$. From the Gaussian approach to estimating measurement uncertainty [2], the standard uncertainty for ln *S* in Eq. 3

$$u_{\ln S} = \sqrt{s_{A'}^2 + s_{B'}^2\left((1/T) - \overline{(1/T)}\right)^2} \tag{4}$$

where $s_{A'}$ and $s_{B'}$ are the standard errors of the regression coefficients. Equation 4 is derived in Appendix I.

---

[3] A homoscedastic data set is one in which the variance of the data is constant and independent of the value of *X*. Data sets for which this is untrue are heteroscedastic. Homoscedasticity is assumed in unweighted least squares analyses.

The Apelblat equation (Eq. 2) is also linear (in $1/T$ and ln T) but the addition of another correlated coefficient requires the use of the covariance matrix. The standard uncertainty of ln S for the Apelblat equation is

$$u_{\ln S} = \sqrt{s_A^2 + s_B^2/T^2 + s_C^2(\ln T)^2 + 2\, s_{AB}/T + 2 s_{AC} \ln T + 2 s_{BC} \ln T/T} \tag{5}$$

where s_{AB}, s_{AC}, and s_{BC}, are the covariances between the coefficients A, B and C. and are recovered as elements of the covariance matrix of the coefficients. Note that the diagonal elements of the covariance matrix are the squares of the standard errors of the coefficients. Equation 5 is also derived in Appendix I.

Fitting to either Eq. 1 or Eq. 2 may be performed in Microsoft Excel, with the function LINEST providing the standard errors (s_A, s_B, $s_{y/x}$). The covariance matrix for Eq. 5 may also be calculated using spreadsheet matrix functions (See Supplementary Material).

4.2.1 Data at the same temperature

For the case where each data set includes a measurement at the desired temperature for which a consensus is required, it is still necessary to perform a regression analysis to obtain an estimate of the standard uncertainty (in the absence of any other information). Because of the assumptions of classical linear regression (random, homoscedastic error in y only), the standard uncertainty of a single y value is estimated by the standard error of the regression, $s_{y/x}$. Given the value of ln S and taking $u = s_{y/x}$, from each data set, the consensus value may be calculated by the method described below. See Supplementary Material.

4.2.2 Consensus value

Fitting the data from a particular study allows calculation of the solubility and its uncertainty at any temperature. However, simply averaging the interpolated values of ln S from n different studies, even weighted by the individual uncertainties (usually as $1/u_{\ln S}^2$), may give an uncertainty of the consensus value that is too small because it ignores potentially large between-laboratory uncertainty, so called “dark uncertainty”[5]. This arises from different systematic errors in the results from the laboratories, that are not picked up in regression of the individual laboratory’s data. That is, within each laboratory the results might appear internally consistent but all the data might suffer from the same non-random error. This extra uncertainty manifests as values from the different laboratories that are separated by greater differences than the individual uncertainties imply (i.e., the error bars do not overlap). The between-laboratory uncertainty may be estimated by the method of DerSimonian and Laird [6], which adds a single random between-laboratory error to the measurement model.

The laboratory random effects model of DerSimonian and Laird represents the standard uncertainty of a result from a laboratory (data set) as the combination of the uncertainty from the fitted data discussed above ($u_{\ln S}$) and an estimate of the between-laboratory uncertainty τ. It is this combination that gives the new weight, w^*. The weight for laboratory j is

$$w_j^* = 1/\left(\tau^2 + \left(u_{\ln S_j}\right)^2\right) \tag{6}$$

τ is obtained by the method of moments (see [6] and Appendix II). The presence of τ^2 in the weight mitigates the effect of one very small laboratory measurement uncertainty.

The DerSimonian-Laird consensus value ($\ln S^*$) is then a weighted average of results from the *N* laboratories.

$$\ln S^* = \frac{\sum_{j=1}^{N} w_j^* (\ln S)_j}{\sum_{j=1}^{N} w_j^*} \tag{1}$$

The standard uncertainty ($u_{\ln S^*}$) of the consensus value is

$$u_{\ln S^*} = \sqrt{\frac{1}{\sum_{j=1}^{N} w_j^*}} \tag{2}$$

The approach to obtain a consensus value at a given temperature is therefore:

1. Consider all data and determine which sets can be used.
2. Obtain the fitted equation of choice (van't Hoff Eq. 1, Apelblat Eq. 2) for the data sets.
3. For each valid set, *j*, using the fitted equation, calculate the solubility $(\ln S)_j$ and its standard uncertainty $u_{\ln S_j}$ at the desired temperature.
4. Calculate the consensus mean (Eq. 7) and its standard uncertainty (Eq. 8) by the DerSimonian Laird procedure (Appendix II).

A step-by-step method that may be implemented in a spreadsheet to obtain consensus values with uncertainties by the DerSimonian Laird approach is given in Appendix II. Worked examples in Microsoft Excel are in the Supplementary Material.

5 Examples

Two examples of the proposed methodology are given here. First, data sets for the solubility of hexanedioc acid in ethanoic acid are shown to be fitted by the van't Hoff equation. The second example, hexanedioc acid in propanone shows clear curvature and is fitted to the Apelblat equation for five out of six sets of data. Calculation of the significance of the coefficient of ln *T* (*C*) should guide the choice of model, however, the difference is well shown in plots of the residuals from fitting each equation. We take as an example hexanedioic acid in ethanoic acid of Luo [7] (Fig. 1a), and hexanedioic acid in propanone of Li [8] (Fig. 1b). Fig. 1a shows a random scatter of residuals about zero, with no improvement using the Apelblat equation. In Fig. 1b the residuals for the Apelblat model are random about zero, but van't Hoff residuals show the typical parabolic shape indicating curvature.

For Luo the probability of H_0 (the ln *T* coefficient in the Apelblat equation is not significant) is 0.45, while for Li the probability of H_0 is 0.0001.

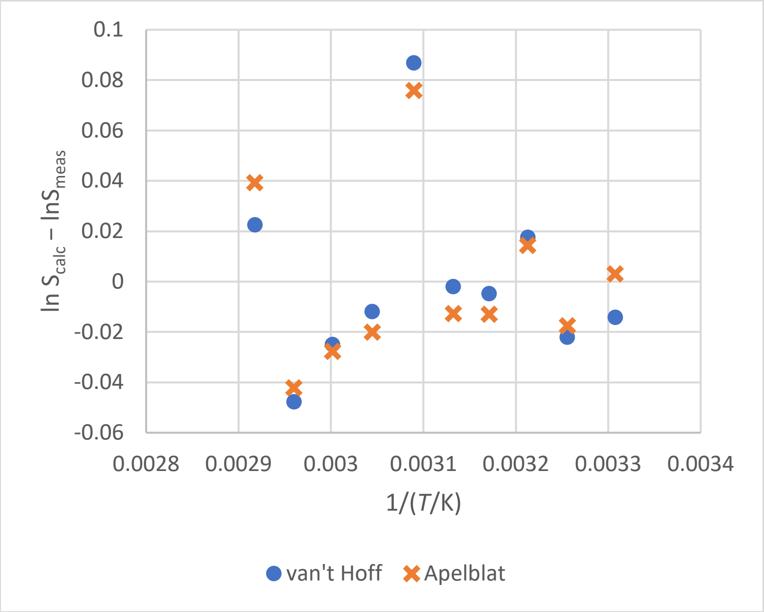

Fig. 1 (a) Residuals of data for hexanedioic acid in ethanoic acid of Luo [7] fitted to the van’t Hoff model Eq. 1 (filled circles) and Apelblat model Eq. 2 (crosses).

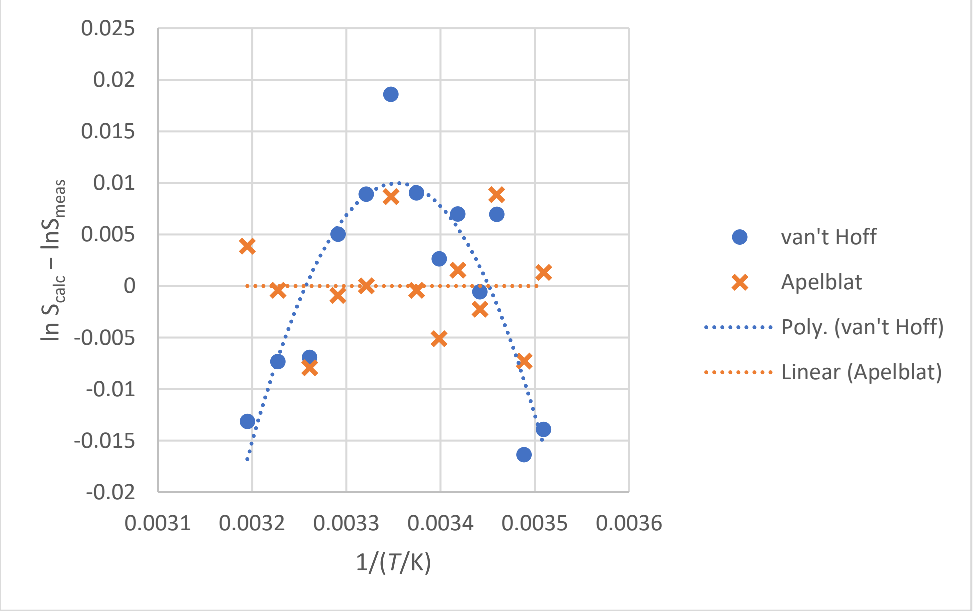

Fig. 1 (b) Residuals of data for hexanedioic acid in propanone of Li [8] fitted to van’t Hoff model Eq.1 (filled circles) and Apelblat model Eq. 2 (crosses). Dotted line is a parabolic fit to van’t Hoff residuals to show curvature of regression.

5.1 Hexanedioic acid in ethanoic acid

For the nine sets of data, eight were sufficiently modelled by the van't Hoff equation. The data of Luo [9] required the additional coefficient of the Apelblat model (Eq. 2). Fitted values at 320 K, which is in the temperature interval of each set, are given in Table 1.

Table 1 log solubility values with standard uncertainties for measurements of hexanedioic acid in ethanoic acid fitted to the van't Hoff equation, calculated at 320 K

		Calculated from regression	
Study	Temperature range /K	ln S (320 K)	$u_{\ln S}$
1 Fan, *et al.* (2007) [10]	297.93 – 347.65	−3.084	0.004
2 Gaivoronskii, *et al.* (2005) [11]	293.15 – 333.15	−3.063	0.012
3 Suren, *et al.* (2013) [12]	303 – 383	−2.975	0.022
4 Zhang, *et al.* (2014) [3]	288.15 – 323.15	−3.036	0.015
5 Luo. , *et al.* (2019a) [9]	300.75 – 342.25	−3.028	0.003
6 Luo, *et al.* (2019b) [7]	302.35 – 342.75	−3.020	0.012
7 Luo, *et al.* (2016) [4][a]	301.85 – 338.25	−3.016	0.003
8 Luo, *et al.* (2017) [13]	301.85 – 334.05	−3.079	0.003
9 Shen, *et al.* (2013) [14]	303.2 – 343.2	−3.041	0.009
Consensus		−3.040	0.011
τ		0.032	

[a] Fitted to the Apelblat equation (Eq. 2)

Applying the DerSimonian Laird procedure gave a value of τ of 0.032 which is seen to be greater than the standard uncertainty of any individual ln$S$ value (i.e. $\tau > u_{\ln S}$). The results are shown in Fig. 2. Note that considering the uncertainties including the between-laboratory uncertainty, the consensus value has about a three-times smaller uncertainty than any individual value, $u_{\ln S^*} = \sqrt{\left(u_j^2 + \tau^2\right)}$, as expected from the combination of nine values.

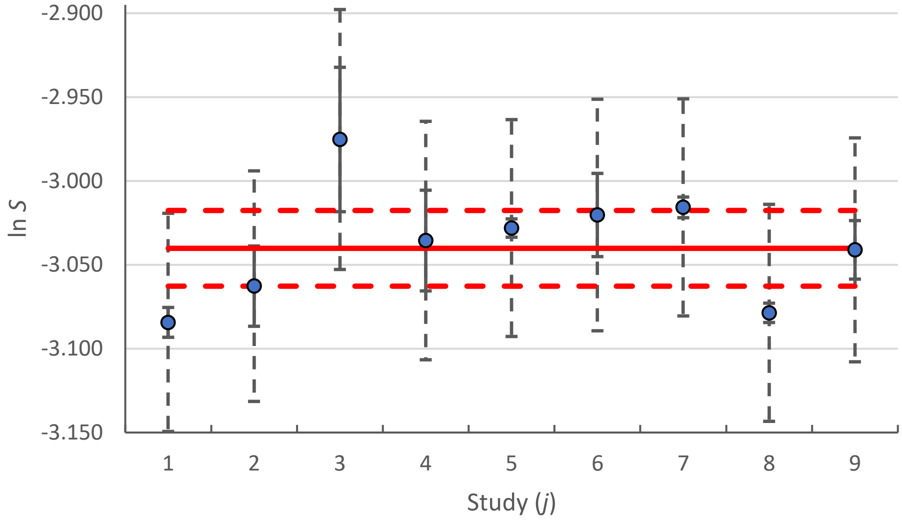

Fig. 2 ln *S* values for studies of the solubility of hexanedioic acid in ethanoic acid at 320 K. Study number (*j*) is in Table 1. Value (point) is calculated from the fitted equation. Solid error bar is 95 % coverage interval of the calculated value ($\pm 2u_j$). Dashed error bar includes the between-laboratory uncertainty (τ), $\pm 2\sqrt{\left(u_j^2 + \tau^2\right)}$. Solid red line is at the consensus value (ln S^*) with the 95 % coverage interval between the horizontal dashed red lines ($\pm 2u_{\ln S^*}$).

5.2 Hexanedioic acid in propanone

Out of six data sets, five were sufficiently curved to require the third coefficient in ln *T* of the Apelblat equation.

Fitted values at 300 K, which is in the temperature interval of each set, are given in Table 2.

Table 2 log solubility values with standard uncertainties for measurements of hexanedioic acid in propanone fitted to the Apelblat equation, Eq. 2, (Studies 1 – 4 and 6) and van't Hoff equation, Eq. 1, (Study 5), calculated at 300 K

		Calculated from regression	
Study	Temperature range /K	ln *S* (300 K)	$u_{\ln S}$
1 Wei, *et al.* (2009) [15]	298.15 – 323.15	−4.020	0.003
2 Mao, *et al.* (2009) [16]	288.05 – 325.85	−3.914	0.007
3 Gaivoronskii, *et al.* (2005) [11]	273.15 – 313.15	−4.059	0.000
4 da Silva, *et al.*, (2005) [17]	297.9 – 322.8	−4.024	0.011
5 Zhang, *et al.* (2014) [3]	288.15 – 323.15	−3.734	0.019
6 Li, *et al.* (2014) [8]	284.95 – 313	−4.054	0.002
Consensus		−3.975	0.016
τ		0.038	

As with the first example the value of τ is greater than the uncertainty of any study.

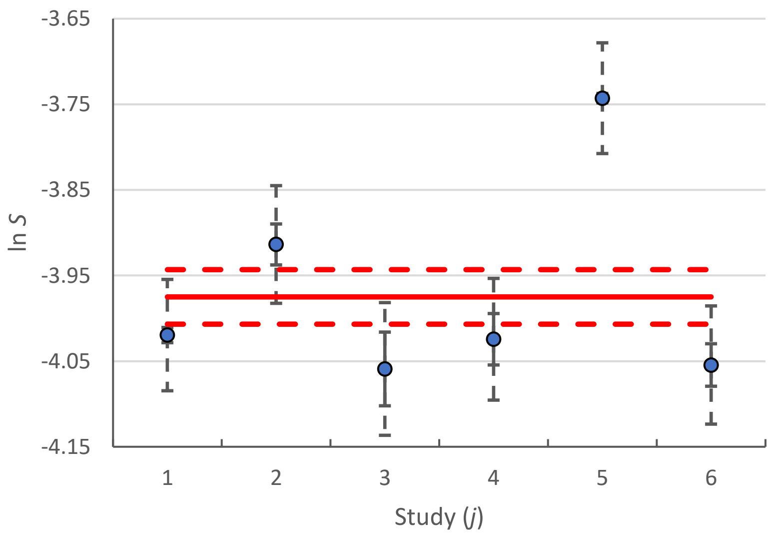

Fig. 3 ln *S* values for studies of the solubility of hexanedioic acid in propanone at 300 K. Study number (*j*) is in Table 2. Value (point) is calculated from the fitted equation. Solid error bar is 95 % coverage interval of the calculated value ($\pm 2u_j$). Dashed error bar includes the between-laboratory uncertainty ($\hat{\tau}$), $\pm 2\sqrt{\left(u_j^2 + \hat{\tau}^2\right)}$. Solid red line is at the consensus value (ln S^*) with the 95 % coverage interval between the horizontal dashed red lines ($\pm 2u_{\ln S^*}$).

The value from study 5 (Zhang et al. [3]) raises an interesting question. There are no obvious flaws in the reported experimental method, which would provide objective grounds for rejecting the results of study *j*, but the reported value(s) lie well away from those of the other studies. Thus, the evaluator faces the dilemma of whether to exclude the value from the evaluations as an outlier which could compromise the consensus value or to include it since the reported scientific method is sound.

We recommend testing the value having the greatest deviation from the mean of the results by a Grubbs' (or maximum normed residual) test [18]. The assumptions of the test are that the data are normally distributed, which should be checked by, for example, a normal probability plot, and there is a single outlier (maximum or minimum). (See Appendix III). The Grubbs statistic is $g = |x_? - \bar{x}|/s$, where $x_?$ Is the suspect point (perforce the maximum or minimum of the data), $\bar{x}$, the mean of the data, and *s*, the sample standard deviation of the data. It is tested against a critical value calculated from the number of data (*n*) and the required confidence (α). Note for 95 % confidence α = 0.05, and the Student-t value is two-tailed.

$$G_{\text{crit}} = \frac{n-1}{\sqrt{n}} \sqrt{\left(\frac{t^2_{(\alpha/n)",n-2}}{n-2+t^2_{(\alpha/n)",n-2}} \right)} \tag{9}$$

For multiple suspected outliers a generalized extreme Studentized deviate ESD test [19] applies the Grubbs formula recursively to *r* data, where *r* is the maximum number of outliers suspected. Of course, there comes a point where the number of outliers suggests the whole data set is suspect.

In this case, the six data cannot be considered normally distributed (see Fig A2 in Appendix III) and therefore a Grubbs' test is not appropriate. We therefore fall back on expert judgement and keep the result from study 5, noting that the result does not make a great difference to the consensus value. (If Study 5 is removed the consensus values are ln S^* = −4.015 and *u*(ln S^*) = 0.014, with τ = 0.031).

6 Conclusion

The procedure outlined above provides a rigorous way to evaluate solubilities from different data sets, reported over a range of temperatures but with values reported at different specific temperatures. Further the estimation of the statistical uncertainties in the individual studies allows estimation of that in the consensus values and any between-laboratory uncertainty.

In principle the procedure can be used with any experimental property where data are not directly comparable but can be represented as a function of experimental variables to allow interpolation to common experimental conditions (temperatures in this case).

It is always important to document any data-analysis procedures, including the procedure for estimating the uncertainty budget, to allow researchers to reevaluate data as new methods, or requirements present themselves.

Supplementary Material

All data and calculations given in section 5 are available in “SupplementaryMaterial-WaghorneConsensusSolubility.xlxs”.

References

1. Shaw, D. G., Bruno, I., Chalk, S., Hefter, G., Hibbert, D. B., Hutchinson, R. A., Magalhães, M. C. F., Magee, J., McEwen, L. R., Rumble, J., Russell, G. T., Waghorne, E., Walczyk, T. Wallington, T. J.: Chemical data evaluation: general considerations and approaches for IUPAC projects and the chemistry community (IUPAC Technical Report). Pure Appl. Chem. 95,1107-1120 (2023) https://doi.org/10.1515/pac-2022-0802
2. BIPM, IEC, IFCC, ILAC, ISO, IUPAC, IUPAP OIML: Evaluation of measurement data - Guide to the expression of uncertainty in measurement (GUM). BIPM, Sèvres, France (2008) https://doi.org/10.59161/JCGM100-2008E
3. Zhang, H., Yin, Q., Liu, Z., Gong, J., Bao, Y., Zhang, M., Hao, H., Hou, B. Xie, C.: An odd–even effect on solubility of dicarboxylic acids in organic solvents. The Journal of Chemical Thermodynamics 77,91-97 (2014) https://doi.org/10.1016/j.jct.2014.05.009
4. Luo, W., Ruan, D., Liu, D., Xie, K., Li, X., Deng, W. Guo, C.: Measurement and Correlation for Solubilities of Adipic Acid in Acetic Acid + ε-Caprolactone Mixtures and Cyclohexanone + ε-Caprolactone Mixtures. Journal of Chemical & Engineering Data 61,2474-2480 (2016) https://doi.org/10.1021/acs.jced.6b00177
5. Thompson, M. Ellison, S.: Dark uncertainty. Accreditation and Quality Assurance: Journal for Quality, Comparability and Reliability in Chemical Measurement 16,483-487 (2011) https://doi.org/10.1007/s00769-011-0803-0
6. DerSimonian, R. Laird, N.: Meta-analysis in clinical trials. Controlled Clinical Trials 7,177-188 (1986) https://doi.org/10.1016/0197-2456(86)90046-2
7. Luo, W., Yan, M., Tao, B., Shi, S. Chen, B.: Solubility of succinic acid, glutaric acid and adipic acid in propionic acid + ε-caprolactone + water mixtures: Experimental measurement and thermodynamic modeling. The Journal of Chemical Thermodynamics 138,332-344 (2019) https://doi.org/10.1016/j.jct.2019.06.033
8. Li, Y., Wang, Y., Ning, Z., Cui, J., Wu, Q. Wang, X.: Solubilities of Adipic Acid and Succinic Acid in a Glutaric Acid + Acetone or n-Butanol Mixture. Journal of Chemical & Engineering Data 59,4062-4069 (2014) https://doi.org/10.1021/je500682v
9. Luo, W., Yan, M., Sheng, X., Tao, B., Shi, S., Deng, W., Yang, W. Luo, S.: A Unified Thermodynamics Model for Solid–Liquid Equilibrium, Liquid–Liquid Equilibrium, and Vapor–Liquid Equilibrium of Cyclohexane Oxidation Systems: NRTL Model. Industrial & Engineering Chemistry Research 58,10018-10030 (2019) https://doi.org/10.1021/acs.iecr.9b00921
10. Fan, L., Ma, P. Xiang, Z.: Measurement and Correlation for Solubility of Adipic Acid in Several Solvents. Chinese Journal of Chemical Engineering 15,110-114 (2007) https://doi.org/10.1016/S1004-9541(07)60042-1
11. Gaivoronskii, A. N. Granzhan, V. A.: Solubility of Adipic Acid in Organic Solvents and Water. Russian Journal of Applied Chemistry 78,404-408 (2005) https://doi.org/10.1007/s11167-005-0305-0

12. Suren, S., Sunsandee, N., Stolcova, M., Hronec, M., Leepipatpiboon, N., Pancharoen, U. Kheawhom, S.: Measurement on the solubility of adipic acid in various solvents at high temperature and its thermodynamics parameters. Fluid Phase Equilibria 360,332-337 (2013) https://doi.org/10.1016/j.fluid.2013.10.003
13. Luo, W., Li, X., Ruan, D., Liu, D., Xie, K., Wang, J., Deng, W., Liu, Q. Chen, Z.: Measurement and Correlation for Solubilities of Adipic Acid, Glutaric Acid and Succinic Acid in Acetic Acid + Cyclohexanone Mixtures. Journal of Chemical & Engineering Data 62,1269-1277 (2017) https://doi.org/10.1021/acs.jced.6b00800
14. Shen, B., Wang, Q., Wang, Y., Ye, X., Lei, F. Gong, X.: Solubilities of Adipic Acid in Acetic Acid + Water Mixtures and Acetic Acid + Cyclohexane Mixtures. Journal of Chemical & Engineering Data 58,938-942 (2013) https://doi.org/10.1021/je301202v
15. Wei, D. Cao, W.: Solubility of Adipic Acid in Acetone, Chloroform, and Toluene. Journal of Chemical & Engineering Data 54,152-153 (2009) https://doi.org/10.1021/je8006887
16. Mao, Z., Sun, X., Luan, X., Wang, Y. Liu, G.: Measurement and Correlation of Solubilities of Adipic Acid in Different Solvents. Chinese Journal of Chemical Engineering 17,473-477 (2009) https://doi.org/10.1016/S1004-9541(08)60233-5
17. da Silva, A. d. P. M. Cajaiba da Silva, J. F.: Determination of the Adipic Acid Solubility Curve in Acetone by Using ATR-FTIR and Heat Flow Calorimetry. Organic Process Research & Development 15,893-897 (2011) https://doi.org/10.1021/op2001029
18. Grubbs, F. E.: Sample Criteria for Testing Outlying Observations. The Annals of Mathematical Statistics 21,27-58 (1950) https://www.jstor.org/stable/2236553
19. Croarkin, C. Tobias, P.: Generalized ESD Test for Outliers. https://www.itl.nist.gov/div898/handbook/eda/section3/eda35h3.htm (2012). Accessed
20. Thode, H. C.: Testing For Normality. CRC Press, Boca Raton (2002) https://doi.org/10.1201/9780203910894

Appendix I
Derivation of the uncertainty of a calculated value from a regression.

We use an equation attributed to Gauss to combine several uncorrelated uncertainties (See section 5 of [2]). For $y = f(x_1, x_2 \dots x_N)$

$$u_Y^2 = \sum_{i=1}^{N} \left(\frac{\partial f}{\partial x_i}\right)^2 u_{x_i}^2 \qquad \text{(A1)}$$

If there are correlations between the quantities *x*

$$u_Y^2 = \sum_{i=1}^{N} \left(\frac{\partial f}{\partial x_i}\right)^2 u_{x_i}^2 + 2\sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \left(\frac{\partial f}{\partial x_i}\right)\left(\frac{\partial f}{\partial x_j}\right) u_{x_i,x_j} \qquad \text{(A2)}$$

where $(\partial f/\partial x_i)$ is the partial derivative of the function *y* with respect to x_i. u_i is the standard uncertainty of quantity x_i and u_{x_i,x_j} is the covariance associated with x_i and x_j. We apply these formulae to the problem at hand.

I.1 Van't Hoff equation

For the regression equation

$$Y = a + bX \qquad \text{(A3)}$$

classical linear regression assumes no uncertainty in *X*, and so when a value of *Y* is calculated at a given value of *X* the uncertainty in the value of *Y* arises only from uncertainties in the estimates of the coefficients *a* and *b*. Applying Eq. A2

$$u_Y^2 = \left(\frac{\partial Y}{\partial a}\right)^2 u_a^2 + \left(\frac{\partial Y}{\partial b}\right)^2 u_b^2 + 2\left(\frac{\partial Y}{\partial a}\right)\left(\frac{\partial Y}{\partial b}\right) u_{a,b} \qquad \text{(A4)}$$

where $u_a$and u_b are the standard uncertainties of the coefficients *a* and *b* respectively, $(\partial Y/\partial a)$ and $(\partial Y/\partial b)$ are the partial derivatives of *Y* with respect to *a* and *b*, and $u_{a,b}$ is the covariance between *a* and *b*.

If there is no correlation between *a* and *b* then $u_{a,b} = 0$ and Eq. A4 becomes

$$u_Y^2 = \left(\frac{\partial Y}{\partial a}\right)^2 u_a^2 + \left(\frac{\partial Y}{\partial b}\right)^2 u_b^2 \tag{A5}$$

For Eq. A3, $\partial Y/\partial a = 1$ and $(\partial Y/\partial b) = X$, and from the regression the standard uncertainties in *a* and *b* are estimated by the standard errors of the coefficients s_a and s_b.

For the van't Hoff equation (Eq. 1), $Y = \ln S$, $(\partial \ln S/\partial A) = 1$ and $(\partial \ln S/\partial B) = 1/T$ so that

$$u_{\ln S} = \sqrt{s_A^2 + s_B^2/T^2 + 2\, s_{A,B}/T} \tag{A6}$$

where $s_{A,B}$ is the covariance between *A* and *B* which may be calculated by Eq. A17 below.

Avoiding correlation

Correlation between the coefficients is removed by subtracting the mean of *X* ($\bar{X}$) from each value of *X*. The regression equation is now

$$Y = a' + b'(X - \bar{X}) \tag{A7}$$

and Eq. A5 becomes

$$u_Y = \sqrt{s_{a'}^2 + s_{b'}^2 (X - \bar{X})^2} \tag{A8}$$

In the case of the van't Hoff equation written as

$$\ln S = A + B\left((1/T) - \overline{(1/T)}\right) \tag{A9}$$

$$u_{\ln S} = \sqrt{s_A^2 + s_B^2\left((1/T) - \overline{(1/T)}\right)^2} \tag{A10}$$

where $\overline{(1/T)}$ is the mean of the reciprocal temperatures.

It is noted that the same result is obtained from the van't Hoff equation ($\ln S = A + B/T$) using the full form of the uncertainty calculation (Eq. A6), including correlation between *A* and *B* ($u_{A,B}$).

Subtracting the mean of $1/T$ simplifies the calculation of $u_{\ln S}$, allowing the use of the regression statistics s_A and s_B (given, for example in Excel by LINEST) without need for matrix multiplication.

I.2 Apelblat Equation

For a linear regression with two 'X' variables, X_1 and X_2

$$Y = a + bX_1 + cX_2 \qquad \text{(A11)}$$

and

$$\begin{aligned} u_Y^2 = {} & \left(\frac{\partial Y}{\partial a}\right)^2 u_a^2 + \left(\frac{\partial Y}{\partial b}\right)^2 u_b^2 + \left(\frac{\partial Y}{\partial c}\right)^2 u_c^2 + 2\left(\frac{\partial Y}{\partial a}\right)\left(\frac{\partial Y}{\partial b}\right) u_{a,b} \\ & + 2\left(\frac{\partial Y}{\partial a}\right)\left(\frac{\partial Y}{\partial c}\right) u_{a,c} + 2\left(\frac{\partial Y}{\partial b}\right)\left(\frac{\partial Y}{\partial c}\right) u_{b,c} \end{aligned} \qquad \text{(A12)}$$

Therefore as $(\partial Y/\partial a) = 1$, $(\partial Y/\partial b) = X_1$, and $(\partial Y/\partial c) = X_2$

$$u_Y^2 = u_a^2 + X_1^2 u_b^2 + X_2^3 u_c^2 + 2X_1 u_{a,b} + 2X_2 u_{a,c} + 2X_1 X_2 u_{b,c} \qquad \text{(A13)}$$

It is not possible to manipulate the equation to remove correlations among *a*, *b* and *c*. For the Apelblat equation

$$\ln S = A + B/T + C \ln T \qquad \text{(A14)}$$

$\partial \ln S/\partial A = 1$, $(\partial \ln S/\partial B) = 1/T$, and $(\partial \ln S/\partial C) = \ln T$ and the uncertainty in ln *S* is, by Eq. A13

$$u_{\ln S}^2 = u_A^2 + u_B^2/T^2 + u_A^2(\ln T)^2 + 2\,u_{A,B}/T + 2u_{A,C}\,(\ln T) + 2u_{B,C}\,(\ln T)/T \qquad \text{(A15)}$$

Equating the standard errors of the coefficients with u_A, u_B, and u_C , and elements of the covariance matrix with $u_{A,B}$, $u_{A,C}$ and $u_{B,C}$

$$u_{\ln S} = \sqrt{s_A^2 + s_B^2/T^2 + s_C^2(\ln T)^2 + 2\,s_{A,B}/T + 2s_{A,C}\,(\ln T) + 2s_{B,C}\,(\ln T)/T} \qquad \text{(A16)}$$

The covariance matrix for the coefficients of a linear regression may be estimated as

$$s_{\mathrm{y/x}}^2(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1} \qquad \text{(A17)}$$

where $s_{\mathrm{y/x}}$ is the standard error of the regression and **X** is the matrix of x-values with an additional column of ‘1’s to allow for the intercept. See Supplementary Material for Excel Code.

Appendix II
DerSimonian–Laird (DL) Method for Determination of Consensus Values and their Uncertainties

A method of introducing a component of uncertainty that may be attributed to systematic errors in laboratories, that are not manifest in repeatability studies, takes the simplified view that a single between-laboratory uncertainty may be added, and then used to determine the consensus value and its uncertainty.

A particular laboratory has a standard uncertainty $u_{\ln S}$ associated with a measured solubility ln *S*. In this paper, we have obtained $u_{\ln S}$ from a regression analysis of data at different temperatures (See Appendix I).

In the DL method [6], the between-study standard uncertainty among the values from different studies (sometimes called the dark uncertainty), symbol τ, is estimated $(\hat{\tau})$ and combined with the standard uncertainty $u_{\ln S}$ from the regression (Eqs. A10, A15 or A16). For data set *i* the standard uncertainty of ln *S*, $u_{\ln S_i,\mathrm{DL}}$, is now

$$u_{\ln S_i,\mathrm{DL}} = \sqrt{u_{\ln S_i}^2 + \hat{\tau}^2} \tag{A18}$$

w_i^* is a weighting factor which is used to calculate the consensus value $\ln S^*$

$$w_i^* = 1/u_{\ln S_i,\mathrm{DL}}^2 \tag{A19}$$

$$\ln S^* = \frac{\sum_{i=1}^{N} w_i^* \ln S_i}{\sum_{i=1}^{N} w_i^*} \tag{A20}$$

The between-laboratory uncertainty is calculated from the uncertainties of each calculated value, via the weights w_i

$$w_i = 1/u_{\ln S_i}^2 \tag{A21}$$

$$\hat{\tau} = \sqrt{(Q - df)/c} \tag{A22}$$

where, *df* are degrees of freedom = *N* – 1

$$Q = \sum_{i=1}^{N} w_i (\ln S_i)^2 - \frac{\left(\sum_{i=1}^{N} w_i \ln S_i\right)^2}{\sum_{i=1}^{N} w_i} \tag{A23}$$

$$c = \sum_{i=1}^{N} w_i - \frac{\sum_{i=1}^{N} w_i^2}{\sum_{i=1}^{N} w_i} \tag{A24}$$

Appendix III
Testing for an outlier by the Grubbs' test

Before outlining the procedure that is undertaken to test an outlying result, we recall the advice given by Grubbs himself [18]:

Actually, the rejection of "outlying" observations may be just as much a practical (or common sense) problem as a statistical one and sometimes the practical or experimental viewpoint may naturally outweigh any statistical contributions. In this connection, the concluding remarks of Rider's survey are pertinent: "In the final analysis it would seem that the question of the rejection or the retention of a discordant observation reduces to a question of common sense. Certainly, the judgment of an experienced observer should be allowed considerable influence in reaching a decision. This judgment can undoubtedly be aided by the application of one or more tests based on the theory of probability, but any test which requires an inordinate amount of calculation seems hardly to be worthwhile, and the testimony of any criterion which is based upon a complicated hypothesis should be accepted with extreme caution."

The Grubbs' test outlined below assumes a single potential outlier in a data set that is normally distributed. Therefore, it is good practice to perform a normality test. There are many possible tests [20], and we recommend the visual assessment of a normal probability plot. A normal probability plot plots expected values of *n* data if they were perfectly normally distributed. These are plotted against the experimental data, a straight line being an indicator that the experimental data is indeed normal. Outliers appear off the line, and curvature implies various deviations from normality. The function calculated for each of *n* values is

$$z_i = \boldsymbol{\Phi}^{-\mathbf{1}}\left(\frac{i-a}{n+1-2a}\right) \tag{A25}$$

where $\boldsymbol{\Phi}^{-\mathbf{1}}$ is the standard normal quartile function (NORM.S.INV() in Excel), *i* is the rank of the value *i* , and *a* is 3/8 (0,375) for $n \leq 10$ and 0.5 for $n > 10$. The function may be created in Excel by

=NORM.S.INV((RANK(ln S_i, *range of values*, 1)-0.375)/(COUNT(*range of values*)+1-2*0.375)).

We use two examples from the paper.

The z and ln *S* values for hexanedioic acid in ethanoic acid are given in Table A1

Table A1 Data (not sorted) for normal probability plot for the solubility of hexanedioic acid in ethanoic acid. See Eq. A25

Study	z	ln *S*
1 Fan, *et al.* (2007) [10]	-1.49415	-3.084
2 Gaivoronskii, *et al.* (2005) [11]	-0.57164	-3.063
3 Suren, *et al.* (2013) [12]	1.494155	-2.975
4 Zhang, *et al.* (2014) [3]	0	-3.036
5 Luo. , *et al.* (2019a) [9]	0.274391	-3.028
6 Luo, *et al.* (2019b) [7]	0.571638	-3.020
7 Luo, *et al.* (2016) [4][a]	0.931971	-3.016
8 Luo, *et al.* (2017) [13]	-0.93197	-3.079
9 Shen, *et al.* (2013) [14]	-0.27439	-3.041

Fig. A1 shows the normal probability plot, which has no obvious curvature or outliers.

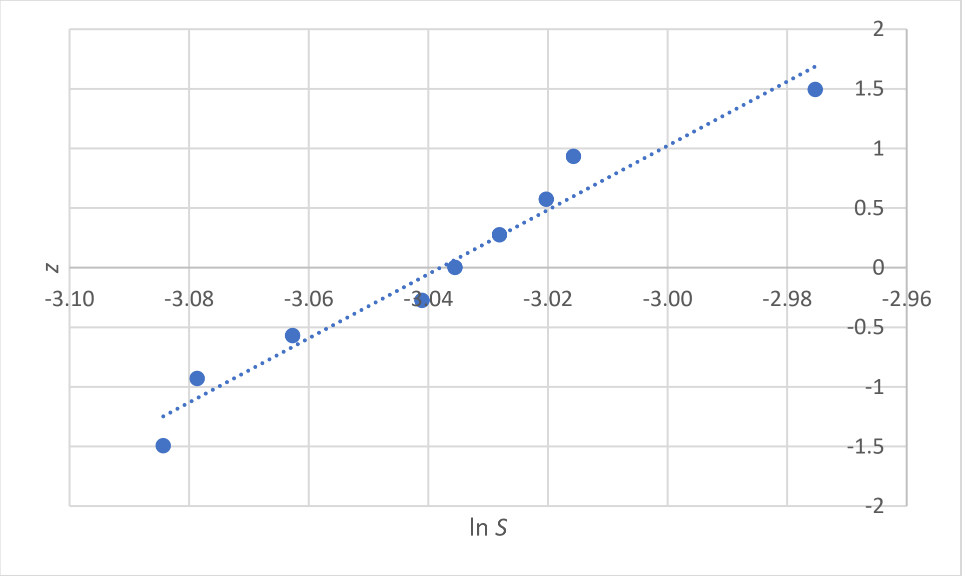

Fig. A1. Normal probability plot for the solubility of hexane in ethanoic acid measured in nine laboratories.

The data in Fig. A1 appears normally distributed with no outliers, but we will test the most extreme value -2.975 (Study 3). As with many statistical tests a parameter is calculated and then compared with a critical value at a particular probability. The Grubbs parameter is simply a *z* score calculated using the mean and sample standard deviation of the entire data set (i.e. including any potential outliers). The critical value for a probability of $1 - \alpha$ (i.e. for 95 % confidence, α = 0.05) is obtained from tables or calculated using the formula

$$g_i = \left(\ln S - \overline{\ln S}\right)/s_{\ln \mathrm{S}} \tag{A26}$$

$$G_{\mathrm{crit}} = \frac{n-1}{\sqrt{n}} \sqrt{\left(\frac{t^2_{(\alpha/n)'',n-2}}{n-2+t^2_{(\alpha/n)'',n-2}}\right)} \tag{A27}$$

where t is the two-tailed Student t value at a probability of (α/n) and $n - 2$ degrees of freedom. In Excel for 95 % confidence this becomes

=(*n* -1)/SQRT(*n*)*SQRT((T.INV.2T(0.05/*n*,*n* -2))^2/(*n*-2+ T.INV.2T(0.05/*n*,*n* -2)^2)).

For the value ln S = -2.975 (Study 3) and n = 9, g = 1.84 (Eq. A26), G_{crit} for 95 % confidence = 2.22. As $g < G_{\mathrm{crit}}$ we conclude the value is not an outlier.

The problem with the solubility of hexanedioc acid in propanone is seen in the normal probability plot (Fig. A2), taking data from Table A2.

Table A2 Data for normal probability plot for the solubility of hexanedioic acid in propanone. See Eq. A25

Study	z	ln *S*
1 Wei, *et al.* (2009) [15]	0.202	-4.020
2 Mao, *et al.* (2009) [16]	0.643	-3.914
3 Gaivoronskii, *et al.* (2005) [11]	-1.282	-4.059
4 da Silva, *et al.*, (2005) [17]	-0.202	-4.024
5 Zhang, *et al.* (2014) [3]	1.282	-3.743
6 Li, *et al.* (2014) [8]	-0.643	-4.054

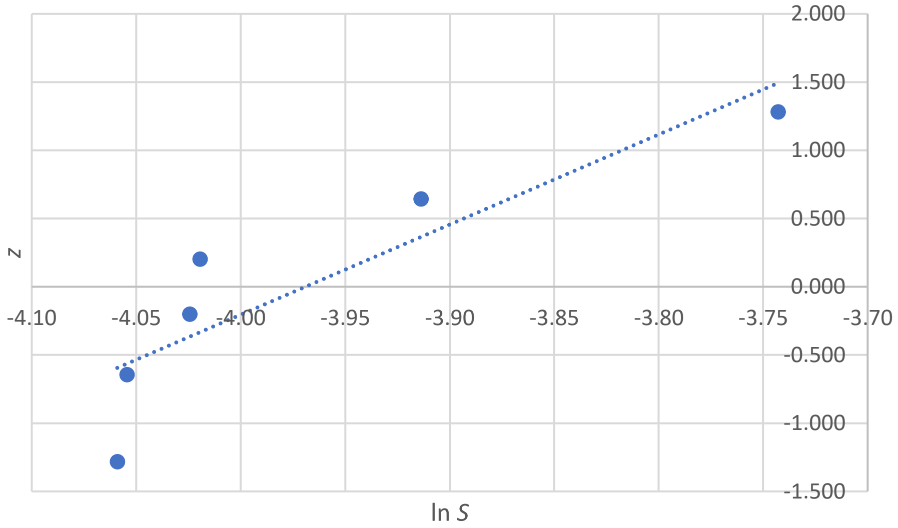

Fig. A2 Normal probability plot for the solubility of hexane in propanone measured in seven laboratories.

The plot is clearly curved which suggests a non-normal distribution in which the extreme point may not be an outlier. It is beyond the scope of this paper to speculate on why there is a long tail on the distribution, but it is now not possible to use Grubbs, or any test based on the assumption of normality, to test the potential outlier.