A New Least-Squares Adjustment Procedure to Achieve Internal Consistency of Physico-Chemical Properties

Urs Schenker[1], Matthew MacLeod[1], Martin Scheringer[1], Konrad Hungerbuehler[1]

[1]Swiss Federal Institute of Technology

Introduction

Modeling the fate of chemicals in the environment requires simulation of partitioning among different media such as air, water, and soil. The partitioning behavior is described by chemical partitioning properties (octanol-water partition coefficient, air-water partition coefficient (the dimensionless form of the Henry's law constant), and octanol-air partition coefficient). Given the large number of chemicals in commerce and the labor required to make measurements, complete sets of partition coefficients are not available for all chemicals. Therefore, Cole and Mackay have developed a method to estimate partition coefficients from the vapor pressure (which can be expressed as a solubility in air) and the solubilities in water and octanol: the three-solubility approach[1].

The approach assumes ideal solute-solvent behavior from high dilutions up to the solubility limit (in other words, the activity coefficients are taken to be constant over the whole concentration range)[2]. Under these assumptions, thermodynamic relationships dictate that the partition coefficient (K) between two media must be equal to the ratio of the solubilities (S) in those two media. For air, water and octanol, three relationships can be expressed:

$$S_A / S_W = K_{AW}; \; S_O / S_A = K_{OA}; \; S_O / S_W = K_{OW}$$

The index A stands for air, W for water, and O for octanol. S_A can be calculated from vapor pressure as follows:

$$S_A = P / RT$$

The three-solubility approach allows unknown partition coefficients and solubilities to be estimated from known properties. For instance if no measurement of the K_{OA} is available, it can be estimated from the vapor pressure and the octanol solubility.

If a complete set of solubilities and partition coefficients is available for a given chemical, the three-solubility approach allows one to verify if those properties are internally consistent, and thus to detect coarse errors. This has been explored by Beyer et al[3]. To find the best estimate of the true values of a set of partition coefficient and solubilities, they suggest that measured values be adjusted to be internally consistent. They state that the method they have developed "find(s) the one set of internally consistent partitioning parameters that is minimally divergent from the experimental values." The approach has been applied by Beyer et al. to 50 mostly aromatic chemicals, by Li et al. to a set of PCBs, by Xiao et al. to different congeners of hexachlorocyclohexane, and by Shen and Wania to a set of organochlorine pesticides[4-6].

However, the set of internally consistent values calculated with the method presented by Beyer et al. is not minimally divergent from the measured values[7]. We present an improved method based on the theory of least-squares that provides a set of internally consistent values that diverges less from the measured values (expressed in terms of the sum of squares of the differences between original and adjusted values). Here, we show how the new least-squares method minimizes the deviation from measured values, and what the impact of the new procedure is regarding the final values of the physicochemical properties. We will also present a more rigorous approach for dealing with different uncertainties in the measured quantities.

Materials and Methods

In the field of land surveying, the angles of a triangle are measured with instruments called theodolites. Geometrical constraints dictate that the three angles in a triangle must sum to 180°. However, if such angles are measured in reality, the sum of the three measurements will usually not be exactly 180° because of measurement errors (see figure 1). To find the best estimate of the true angles, the measured quantities are adjusted. The misclosure error (the difference between the sum of the three angles and 180°) is equally distributed to the three measured quantities. If the three measurements are $a = 30°$, $b = 65°$ and $g = 90°$, then the misclosure error is $30° + 65° + 90° - 180° = 5°$, and thus the best estimate of the true angles would be $a = 28.33°$, $b = 63.33°$, and $g = 88.33°$, so that the geometrical constraint is fulfilled. Modern theodolites allow measurements of angles and distances at the same time. To include also the measurements of distances in the calculation of the true angles' best estimate, the same approach can be used, but several constraints have to be fulfilled simultaneously. To this end, Helmert developed a matrix-based algorithm in 1872[8]. His method is still widely used in land surveying and it can also be applied to adjust measurements of partition coefficients to conform to thermodynamic constraints.
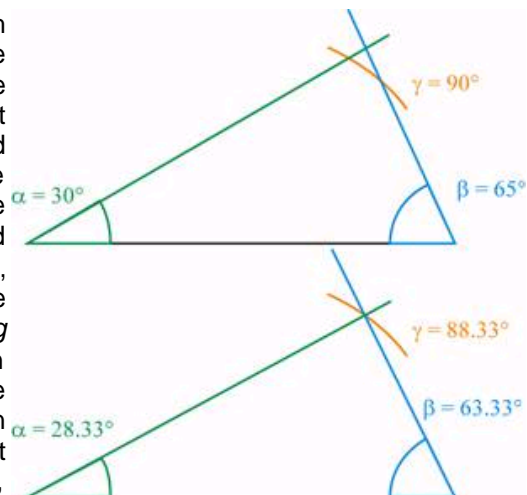

Figure 1: Example of the measurement of three angles in a triangle, before and after the adjustment for internal consistency.

As an example, consider independent measurements of S_A, S_W and K_{AW} (the log form is chosen because measurement errors are expected to be log-normally distributed): $\log(S_A) - \log(S_W) - \log(K_{AW}) = w$ where w is the misclosure error. Assuming that the best estimate of the true values will form an internally consistent set, the misclosure errors w will be zero. For example, $\overline{\log(S_A)} - \overline{\log(S_W)} - \overline{\log(K_{AW})} = 0$. To link the measured values to the adjusted values, the adjustment v_i is introduced:

$$\overline{\log(S_A)} = \log(S_A) - v_1 \; ; \; \overline{\log(S_W)} = \log(S_W) - v_2 \; ; \; \overline{\log(K_{AW})} = \log(K_{AW}) - v_3 \, .$$

To find the best estimate of the true values, the sum of squares of the v_i has to be minimized, under the constraint that $v_1 - v_2 - v_3 = w$. For that purpose, the Lagrange theorem is used and a function W is introduced: $\Omega = v_1^2 + v_2^2 + v_3^2 - 2 \cdot k(v_1 - v_2 - v_3 - w)$. If the partial derivatives of W (the Lagrangian) are set to zero, the resulting v_i will allow one to calculate an internally consistent set of partition properties that differ minimally from the measured quantities.

To minimize several equations simultaneously, the above equations are written in matrix form: $\mathbf{B} \cdot \mathbf{l} = \mathbf{w}$, where

$$\mathbf{B} = \begin{pmatrix} 1 & -1 & 0 & -1 & 0 & 0 \\ 1 & 0 & -1 & 0 & 0 & 1 \\ 0 & 1 & -1 & 0 & 1 & 0 \end{pmatrix}, \quad \mathbf{l} = \begin{pmatrix} \log(S_A) \\ \log(S_W) \\ \log(S_O) \\ \log(K_{AW}) \\ \log(K_{OW}) \\ \log(K_{OA}) \end{pmatrix} \quad \text{and} \quad \mathbf{w} = \begin{pmatrix} w_1 \\ w_2 \\ w_3 \end{pmatrix}.$$

It can be shown that the adjustments v_i that will result in internally consistent values that are minimally different from the measured quantities can be found with the following equation: $\mathbf{v} = \mathbf{Q} \cdot \mathbf{B}^T \cdot (\mathbf{B} \cdot \mathbf{Q} \cdot \mathbf{B}^T)^{-1} \cdot \mathbf{w}$, where **Q** is a square matrix with the variances of the measured quantities on its diagonal [8].

Results and Discussion

We have applied the new least-squares adjustment procedure to the partitioning properties of 60 compounds

1872

presented in Beyer et al., Li et al., Xiao et al., and Shen and Wania[3-6]. The results obtained are compared to the values that were obtained using the iterative adjustment procedure presented by Beyer et al[3]. We have first assessed whether the adjusted values obtained with the two methods were significantly different, and by how much, and second we have analyzed which of the two datasets were closer to the initially measured values, which would be a quality criteria for one or the other method.
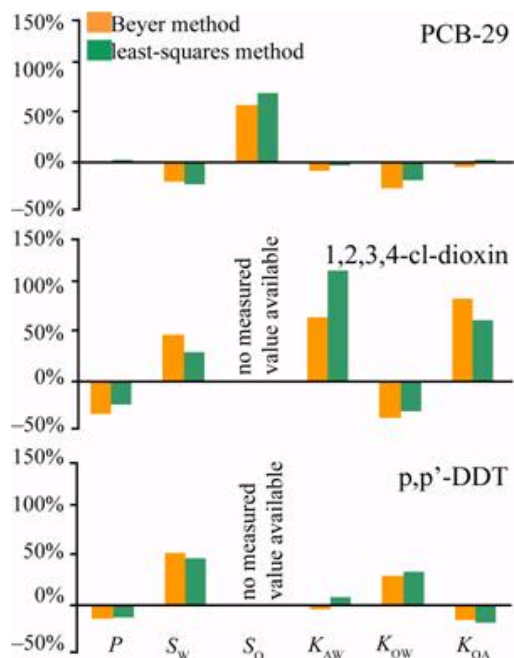

Figure 2: Relative adjustment of the partitioning properties with the Beyer method and the new least-squares method. The compounds are (top to bottom) PCB-29[4], 1,2,3,4-cl-dioxin[3], and p,p'-DDT[5].

First, it can be seen that in some cases both methods considerably adjust the measured quantities (figure 2). For 1,2,3,4-cl-dioxin several properties are increased by 50 % or more. For PCB-29, on the other hand, the adjustments are lower than 10% except for the octanol solubility, which is changed by about 60%.

When the adjusted values of the two methods are compared, it can be seen that considerable differences exist, for instance for 1,2,3,4-cl-dioxin: the K_{AW} is increased by about 50% in the Beyer method, whereas it is increased by more than 100% in the new least-squares method (figure 2).

To measure how much the adjustment procedures change the measured quantities, the sum of squares of the v_i (the differences between original and adjusted values) have been calculated with the two methods. For all compounds, this value is smaller or equal if the least-squares method is used (see figure 3). Thus, the adjusted values from the Beyer method deviate unnecessarily from the measured quantities. We therefore suggest that, in the future, the new least-squares method be used when partition coefficients are adjusted.
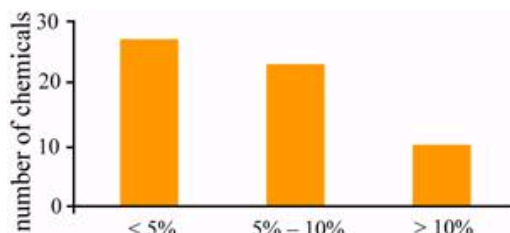

Figure 3: 60 chemicals have been adjusted with the Beyer method and the new least-squares method. For all compounds, the least-squares method resulted in smaller deviations from the measured quantities than the Beyer method. However, for almost half of those compounds, the difference in the deviations between the two methods was less than 5%.

The partition coefficients of a given compound usually are measured with different measurement errors. This must be taken into account in the adjustment procedure. Beyer et al. therefore have introduced uncertainty factors with ordinal values from one to five. Small values represent small uncertainties, high values high uncertainties. Although Li et al. give indications how such uncertainty factors can be calculated, uncertainty factors remain somewhat subjective[4]. Consistent estimation of the true uncertainty and variability associated with different analytical methods measuring different chemical properties is a challenging task, and removing the subjectivity entirely is likely impractical.

However, the least-squares procedure provides the mathematical framework for a more quantitative treatment of measurement uncertainties during the adjustment: Instead of uncertainty factors, the least-squares adjustment procedure can incorporate variances for all the measured quantities. The method is capable of propagating the variances of the measured quantities through the calculations, and provides variances of the adjusted values. It is therefore possible to calculate confidence intervals for the adjusted partition properties, a feature that would be highly

desirable for uncertainty analyses in the field of environmental fate modeling.

The new least-squares adjustment procedure allows estimation of internally consistent sets of physicochemical properties that are minimally different from the measured quantities. It is an improvement compared to the previous method because the adjusted values are closer to measured values. In addition, the possibility of quantitatively propagating uncertainty in measured properties to the adjusted values will help to determine the impact of uncertain partition properties in environmental fate and exposure models.

References

1. J. G. Cole and D. Mackay (2000) *Environmental Toxicology and Chemistry* 19: 265-270.

2. R. P. Schwarzenbach, P. M. Gschwend and D. M. Imboden (2003) in: Environmental organic chemistry, Hoboken, New Jersey

3. A. Beyer, F. Wania, T. Gouin, D. Mackay and M. Matthies (2002) *Environmental Toxicology and Chemistry* 21: 941-953.

4. N. Q. Li, F. Wania, Y. D. Lei and G. L. Daly (2003) *Journal of Physical and Chemical Reference Data* 32: 1545-1590.

5. L. Shen and F. Wania (2005) *Journal Of Chemical And Engineering Data* 50: 742-768.

6. H. Xiao, N. Q. Li and F. Wania (2004) *Journal of Chemical and Engineering Data* 49: 173-185.

7. U. Schenker, M. MacLeod, M. Scheringer and K. Hungerbuehler (2005) *Environmental Science & Technology* submitted

8. F. R. Helmert (1872) in: Die Ausgleichungsrechnung nach der Methode der kleinsten Quadrate, Leipzig, Germany