FUZZY DATA RECONCILIATION IN LIFE CYCLE INVENTORY ANALYSIS

Raymond R. Tan[1] **Alvin B. Culaba**
Center for Engineering and Sustainable Development Research
College of Engineering
De La Salle University-Manila
2401 Taft Avenue, 1004 Manila, Philippines

INTRODUCTION

Data acquisition during life cycle inventory assessment (LCI) is the most time-consuming stage in the conduct of LCA. Practitioners rely on a combination of different data sources, including direct measurements, historical company records, expert estimates, scientific or technical literature and electronic databases [1, 2], and it has been noted that data from diverse sources may prove inconsistent [3, 4]. Data reconciliation methods can be applied to ensure that LCI data do not violate material and energy balance principles [5, 6]. Conventional reconciliation methods rely on stochastic programming and invariably involve highly nonlinear mathematical models. Furthermore, they presuppose that inconsistencies in the data arise from random errors that occur during measurement. Data imprecision in LCI, on the other hand, results from a variety of sources, including random noise and epistemological undertainty [7]. An appropriate data reconciliation method for LCI must be able to operate under the assumption that uncertainty is not purely stochastic in nature.

Probability theory is used by majority of LCA practitioners in dealing with data uncertainties encountered in models. Monte Carlo simulation is often employed to simplify the computational procedure. Alternatively, various forms of sensitivity or perturbation analysis have been proposed [8]. Recently there have also been some attempts to use fuzzy set theory for modelling data uncertainties. For example, Tan et al. [9, 10] demonstrated the use of triangular fuzzy numbers in the LCA of a range of alternative automotive fuels, while Geldermann et al. [11] used trapezoidal fuzzy numbers for the analysis of steel production. (It is notable that although these fuzzy models have proven computationally effective, they have nevertheless failed to gain mainstream acceptance, possibly due to the erroneous impressions of fuzzy set theory. It will be necessary for the mystique surrounding fuzzy sets to be addressed before these methods become more commonplace).

FUZZY NUMBERS

Imprecise values can be expressed by fuzzy numbers when the associated uncertainty is not random in nature. These fuzzy numbers consist of a set of intervals for a range of possibility values ranging from zero to unity; possibility is typically interpreted as the *degree of plausibility* or *degree of truth* of of a particular interval [9, 12]. The intervals are stacked to define a fuzzy membership function (or possibility distribution). Stylized triangular or trapezoidal distributions

[1] Fax: +63-2-524-0563; e-mail: tanr_a@dlsu.edu.ph

are often used for convenience [12, 13]. A triangular fuzzy number can be fully defined by specifying three parameters:

$$\tilde{y} \quad = \quad (a, b, c)_T \tag{1}$$

where a, b and c are the lower limit, mode and upper limit of fuzzy number $\tilde{y}$, and the subscript T denotes the triangular form of the distribution. The interval [a, c] is called the support of the fuzzy number; it gives the range of all values of $\tilde{y}$ that are at least marginally possible or plausible. The value b is the core of the fuzzy number and is gives its single most plausible value. A trapezoidal fuzzy number in contrast can be fully defined by four parameters:

$$\tilde{y} \quad = \quad (a, b, c, d)_{Tr} \tag{2}$$

where a, b, c and d are the lower limit, lower mode, upper mode and upper limit of fuzzy number $\tilde{y}$, and the subscript Tr denotes the trapezoidal shape of the distribution. The interval [a, d] is the support of the fuzzy number and gives the range of all values of $\tilde{y}$ that are at least marginally possible or plausible. The interval [b, c] corresponds to the core of the fuzzy number and gives the range of most plausible values. It will be noted that a triangular fuzzy number is a special case of the trapezoidal distribution with b = c, and that an ordinary non-fuzzy (or crisp) number can also be viewed as a fuzzy number with an infinitesimally narrow distribution. Figure 1 shows the fuzzy numbers $(1, 2.5, 3.5)_T$ and $(0.5, 1, 2, 3)_{Tr}$ to illustrate the preceding discussion.

A number of methods have been suggested for determining the distributions of fuzzy numbers:

- Elicitation of distribution parameters from expert estimates. An expert can be asked to identify "all possible values" of an imprecise quantity to determine the support, and to estimate the "most plausible value or values" to determine the core [12]. This procedure is particularly compatible with LCI data derived from subjective estimates (e.g., estimating the thermal efficiency of future power plants using the projections of a combustion expert).

- Determining approximate fuzzy distributions from statistical properties or graphical displays [13]. This approach blurs the distinction between probabilistic and fuzzy schools of thought. At best it is an expedient approach that takes advantage of the computational simplicity of fuzzy models compared to probabilistic ones.

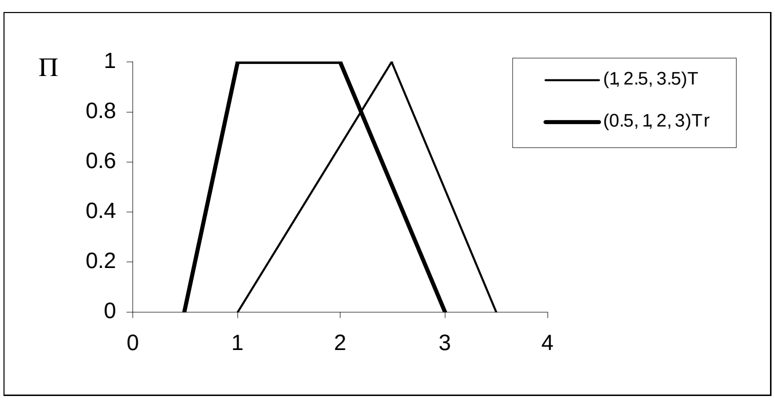

Figure 1. Fuzzy Numbers $(1, 2.5, 3.5)_T$ and $(0.5, 1, 2, 3)_{Tr}$

FUZZY RECONCILIATION MODEL

The task of data reconciliation is to generate a set of consistent stream flowrate values given a set of initially inconsistent flowrate estimates with their associated uncertainty margins. Material and energy balance conditions must also be specified. Data reconciliation approaches assume that inconsistencies arise from uncertainties, and that a consistent set of flowrate values can be found when the error margins are taken into account. The objective is to find the optimal set of reconciled flowrates – those values closest to what are considered as the most plausible values as defined *a priori* by the fuzzy distributions.

Kikuchi [14] first proposed the use of fuzzy linear programming as a means of reconciling inconsistent vehicular traffic flow data. The model is based on the general symmetric fuzzy linear program of Zimmermann [15] and assumes that observed values are imprecise with triangular fuzzy distributions. A modified form of the model for use in LCI data reconciliation problems is used here. The principal change is that the model assumes trapezoidal fuzzy distributions. Since triangular fuzzy numbers are a special case of trapezoidal distributions as discussed in the previous section, the modified formulation is more general than Kikuchi's original model. The mathematical model is:

$$\max \alpha \tag{1}$$

subject to:

$$[x_i - a_i][b_i - a_i]^{-1} \geq \alpha \qquad \forall\, i \tag{2}$$

$$[x_i - d_i][c_i - d_i]^{-1} \geq \alpha \qquad \forall\, i \tag{3}$$

$$\Sigma_i\, e_{ij}\, x_i = 0 \qquad \forall\, i, j \tag{4}$$

$$\alpha \leq 1 \tag{5}$$

$$\alpha, x_i \geq 0 \qquad \forall\, i \tag{6}$$

Variables

α = global satisfaction variable

x_i = adjusted or reconciled flowrate value of stream (i)

Constants

a_i = lower limit of fuzzy flowrate value of stream (i)

b_i = lower mode of fuzzy flowrate value of stream (i)

c_i = upper mode of fuzzy flowrate value of stream (i)

d_i = upper limit of fuzzy flowrate value of stream (i)

e_{ij} = coefficient of stream (i) in process (j)

The model functions by adjusting the global satisfaction variable α as close as possible to unity (Eq. 1 and 5). The first two constraints (Eq. 2 and 3) restrict the value of each variable x_i within the spread of the corresponding trapezoidal fuzzy distribution specified by $(a_i, b_i, c_i, d_i)_{Tr}$ and also force the final values to be as close as possible to the core or mode of the distribution, $[b_i, c_i]$. The third constraint is a non-fuzzy or crisp material balance condition for each process. The values of the coefficients e_{ij} can assume negative values depending on flow direction and can also assume fractional values when there is stream splitting. All variables in the model are nonnegative (Eq. 6).

Significantly, the model shown is a simple linear program that can easily be solved within common spreadsheet software. By comparison, conventional data reconciliation models are nonlinear and thus more difficult to solve.

NUMERICAL EXAMPLES

Two examples are used to illustrate the model. The first case is based on a data reconciliation problem from Madron [6]. The system, illustrated in Figure 2, consists of three processes and twelve streams; the exact nature of each process is immaterial since the principal concern is the reconciliation of mass flows. One material balance constraint is specified to balance the total input and total output of each process. The means and standard deviations of the raw flowrate data are given in the second and third columns of Table 1. These statistics are used to estimate the triangular fuzzy distributions of the flows using the rules described by Mauris et al. [13]. The next three columns in Table 1 show the lower limits, modes and upper limits of these fuzzy numbers. The last two columns show the reconciled values derived using the fuzzy and conventional approaches, respectively.

Table 2 summarizes the total input and output flows of each of the three processes. Note that for the unreconciled data there are flow discrepancies of 52, 26 and 26 kg/h for processes A, B and C, respectively. These discrepancies amount to missing mass that is eliminated by the reconciliation procedure. Figure 3 compares the original and reconciled values for stream 1. The fuzzy flowrate in kg/h is $(3815.3, 3849.0, 3882.7)_T$ while the reconciled value of 3872.5 kg/h falls between the mode and the upper bound.

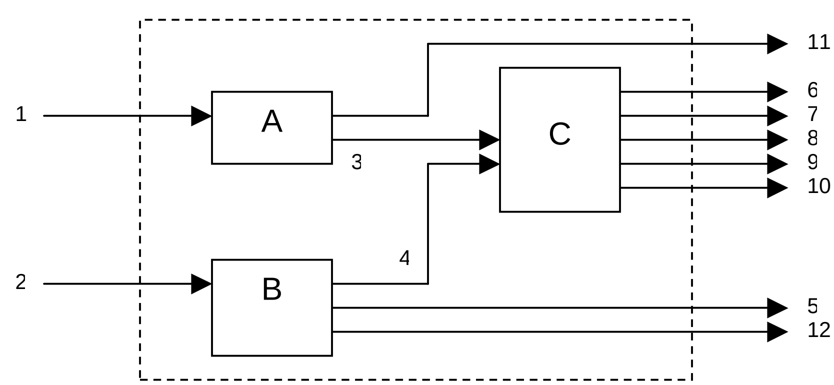

Figure 2. Process Cluster for Example 1 (Madron, 1992)

Table 1. Raw Data and Reconciled Results for Mass Flows in kg/h for Example 1

Stream	Observed Values		Triangular Fuzzy Distribution Parameters			Reconciled Value	
	Mean	Standard Deviation	Lower Limit	Mode	Upper Limit	Fuzzy Method	Conventional Method
1	3849.0	19.2	3815.3	3849.0	3882.7	3872.5	3864.0
2	1848.0	9.2	1831.8	1848.0	1864.2	1837.3	1833.0
3	3852.0	19.1	3818.5	3852.0	3885.5	3828.6	3816.0
4	1362.0	6.8	1350.1	1362.0	1373.9	1370.3	1371.0
5	359.0	3.6	352.7	359.0	365.3	363.4	362.0
6	1022.0	5.1	1013.1	1022.0	1030.9	1028.2	1022.0
7	1021.0	5.1	1012.1	1021.0	1029.9	1027.2	1021.0
8	1033.0	5.2	1024.0	1033.0	1042.0	1039.3	1033.0
9	1048.0	5.2	1038.8	1048.0	1057.2	1046.7	1048.0
10	1064.0	5.3	1054.7	1064.0	1073.3	1057.5	1064.0
11	49.0	4.2	41.7	49.0	56.4	43.9	48.0
12	101.0	2.1	97.3	101.0	104.7	103.6	102.0

Table 2. Total Mass Flows of Unit Processes in Example 1

	Original		Reconciled	
Process	Total Input	Total Output	Total Input	Total Output
A	3849.0	3901.0	3872.5	3872.5
B	1848.0	1822.0	1837.3	1837.3
C	5214.0	5188.0	5199.0	5199.0

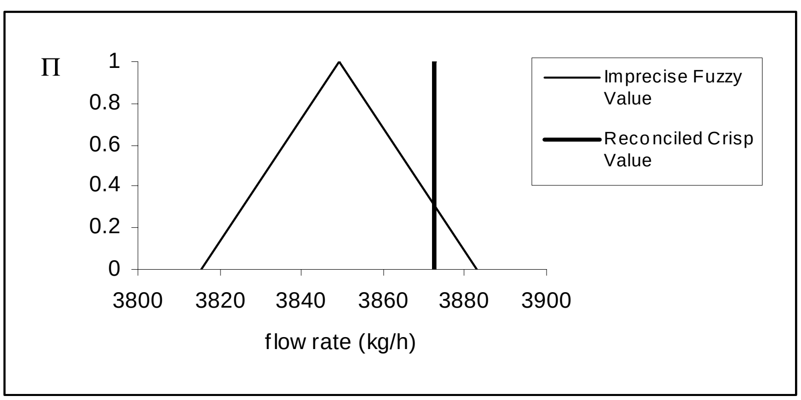

Figure 3. Comparison of Imprecise and Reconciled Stream 1 Flowrates in Example 1

The second example is based on the Bayer process for the production of hydrated alumina from bauxite. The process has five input and five output streams and is shown in Figure 4. Flow data and statistics are derived from the study of Wilson and Jones [4]. Trapezoidal fuzzy distributions are determined from these statistics using the medians, standard deviations as well as the maximum and minimum values.

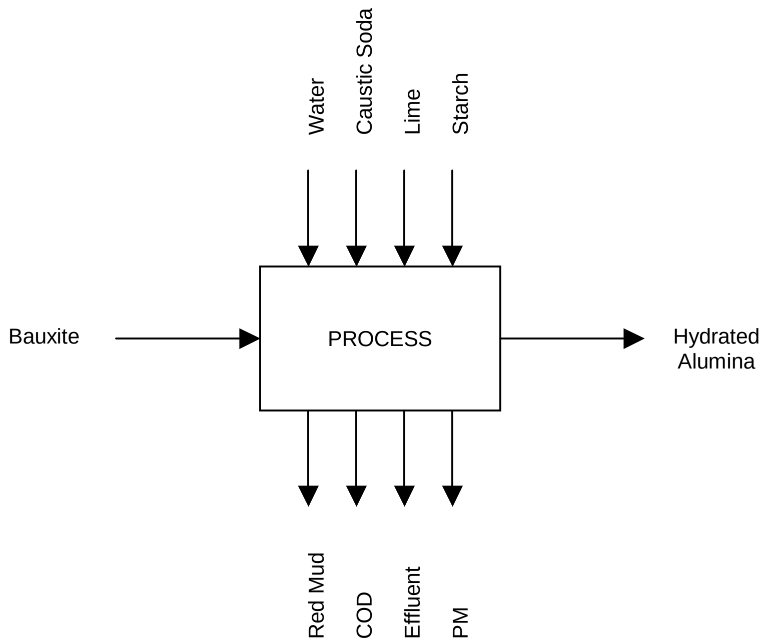

Figure 4. Input and Output Flows of Bayer Process (Wilson and Jones, 1994)

Table 3 shows the original and reconciled mass flow data per unit of the final product. Parameters of the trapezoidal distributions are given in the second through fifth columns, while the original (average) values are in the sixth column. Note that there is a discrepancy or missing mass of 0.403 kg when the original are used. Recognizing this imbalance, Wilson and Jones adjusted the flow values using as hoc and subjective considerations and used the values in the seventh column in their final model (a slight discrepancy remains in the data shown due to round off errors). Use of the fuzzy data reconciliation results in the flows given in the last column of Table 3. Figure 5 compares the original and reconciled values for bauxite usage. The fuzzy value in kg is $(1.255, 1.307, 1.641, 2.209)_{Tr}$ while the reconciled value is 1.662. The latter falls just beyond the upper limit of the core of the fuzzy value.

Table 3. Mass Flow Data and Reconciled Results for Example 1

Streams	Trapezoidal Fuzzy Number Parameters				Original Values	Values Used in Study	Reconciled Values
	Lower Limit	Lower Mode	Upper Mode	Upper Limit			
Inputs							
Bauxite	1.255	1.307	1.641	2.209	1.474	1.55	1.662
Caustic Soda	0.029	0.046	0.147	0.163	0.096	0.07	0.148
Lime	0.022	0.033	0.033	0.163	0.033	0.03	0.038
Water	2.353	4.281	4.281	6.209	4.281	4.31	4.353
Starch	0.007	0.007	0.007	0.007	0.007	0.01	0.007
TOTAL					5.891	5.98	6.207
Outputs							
Red Mud	0.523	0.869	0.869	1.144	0.869	0.85	0.856
COD	0.098	0.111	0.111	0.124	0.111	0.11	0.111
PM	0.001	0.033	0.033	0.064	0.033	0.03	0.031
Effluent	2.353	4.281	4.281	6.209	4.281	4.01	4.209
Hydrated Alumina	1.000	1.000	1.000	1.000	1.000	1.00	1.000
TOTAL					6.294	5.98	6.207

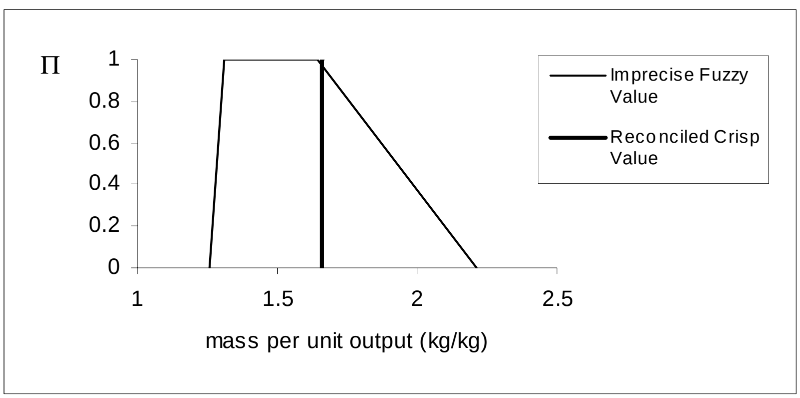

Figure 5. Comparison of Imprecise and Reconciled Bauxite Usage in Example 2

CONCLUSION

A simple and computationally efficient method for data reconciliation using fuzzy linear programming has been demonstrated. The model is suitable for life cycle inventory analysis as well as related procedures such as substance flow analysis. Unlike conventional data reconciliation methods which typically result in highly nonlinear stochastic models, the method shown retains its linearity for basic triangular or trapezoidal fuzzy distributions and can thus be implemented with ease in commercial spreadsheet software.

Despite its simplicity, the procedure in its present form has a number of limitations. One fundamental difficulty is that typical LCA practitioners are more comfortable with the use of probability measures rather than fuzzy numbers in modeling data uncertainty. Current research efforts related to the method itself include:

- Utilization of the reconciliation method in the analysis of inventory data in the gate-to-gate assessment of an adhesive tape manufacturing plant.
- Improvement of the computational procedure. Unlike conventional probabilistic data reconciliation methods, it does not generate corrected uncertainty margins after optimization. Such adjusted uncertainty margins may be needed in subsequent interpretation of the LCA data.
- Integration of the data reconciliation method with the generalized matrix-based computational framework described by Heijungs and Suh [8].

REFERENCES

[1] Boguski, T. K., Hunt, R. G., Cholakis, J. M. and Franklin, W. E. (1996). LCA Methodology. In: Curran, M. A., ed. Environmental Life Cycle Assessment, McGraw-Hill, New York.

[2] Lindfors L. G., Christiansen, K., Hoffmann, L., Virtanen, Y., Junitlla, V., Hanssen, O. J., Ronning A., Ekvall, T. And Finnveden, G. (1995). Nordic Guidelines On Life Cycle Assessment. Nordic Council Of Ministers, Copenhagen.

[3] Culaba, A. B. And Purvis, M. R. I. (1999). A Methodology For The Life Cycle And Sustainability Analysis Of Manufacturing Processes 7: 435 – 445.

[4] Wilson, B. and Jones, B. (1994). The Phosphate Report. Landbank Environmental Research & Consulting, London.

[5] Bagajewicz, M. J. (1996). On the Probability Distribution and Reconciliation of Process Plant Data. Computers and Chemical Engineering 20: 813 – 819.

[6] Madron, F. (1992). Process Plant Performance – Measurement Data Processing for Optimization and Retrofits. Ellis Horwood, West Sussex.

[7] Hoffmann, L., Weidema, B. P., Kristiansen, K., Kruger, A. S. and Ersboll, A. K. (1995). Statistical Analysis and Uncertainties in Relation to LCA. LCA-Nordic Special Report No. 1, Nordic Council of Ministers, Copenhagen.

[8] Heijungs, R. and Suh, S. (2002). The Computational Structure of Life Cycle Assessment. Kluwer Academic, Dordrecht.

[9] Tan, R. R., Culaba, A. B. And Purvis, M. R. I. (2002). Application Of Possibility Theory In The Life Cycle Inventory Assessment Of Biofuels. International Journal Of Energy Research 26: 737 – 745.

[10] Tan, R. R., Culaba, A. B. And Purvis, M. R. I. (2004). POLCAGE 1.0: A Possibilistic Life-Cycle Model For Evaluating Alternative Transportation Fuels. Environmental Modelling And Software (in press).

[11] Geldermann, J., Spengler, T. And Rentz, O. (2000). Fuzzy Outranking For Environmental Assessment. Case Study: Iron And Steel Making Industry. Fuzzy Sets And Systems 115: 45 – 65.

[12] Bardossy, A. and Duckstein, L. (1995). Fuzzy Rule-Based Modeling with Applications to Geophysical, Biological and Engineering Systems. CRC Press, Boca Raton.

[13] Mauris, G., Lasserre, V. and Foulloy, L. (2001). A Fuzzy Approach for the Expression of Uncertainty in Measurement. Measurement 29: 165 – 177.

[14] Kikuchi, S. (2000). A Method To Defuzzify The Fuzzy Number: Transportation Problem Application. Fuzzy Sets And Systems 116: 3 – 9.

[15] Zimmermann, H. J. (1996). Fuzzy Set Theory And Its Applications, 3rd Ed. Kluwer Academic Publishers, Norwell.