Working Paper No. 6
Data Quality, Variability and Uncertainty in LCI

LCI results are limited by the quality of the data. Because LCI results are generally used for comparative purposes, it is essential to gain some knowledge of data quality in order to determine if comparative results are potentially valid.

Unfortunately, modeling data uncertainty is not common practice in life cycle inventories (LCI), although different techniques are available for estimating and expressing uncertainties, and for propagating the uncertainties to the final model results. To clarify and stimulate the use of data uncertainty assessments in common LCI practice, the SETAC working group 'Data Availability and Quality' (Huijbregts 2001) has presented a framework for data uncertainty assessment in LCI which is typical of the current state of the development. In the SETAC analysis, data uncertainty is divided into two categories: (1) lack of data, further specified as complete lack of data (data gaps) and a lack of representative data, and (2) data inaccuracy.

Lack of Data

Filling data gaps can be done by a variety of methods, including using data from similar operations, using surrogate data from related or similar processes or engineering analysis. Other options include input-output modeling, using statistical information for similar products or the main ingredients of a product, and applying the law of mass conservation. If possible, the use of such gap-filling data should be accompanied by data quality indicators, such as a range of values or statistical measures, that conveys information about the possible error incurred by using the chosen method.

A major point is that missing data should not be ignored, or replaced with zeroes unless it is subsequently found by sensitivity analysis to be below the cut-off criteria for significance warranting inclusion (see section 6.4.5 of ISO 14041).

The problem of lack of representative data is more likely to arise during *use* of the US LCI database, rather than in its creation. Such lack may be caused by inappropriate temporal, geographical or technological correlation between the data used and data needed. The table below illustrates the three causes of misrepresentation of data discussed in the SETAC document, and places the relative degree of departure from ideal into 5 semi-quantitative categories called an "indicator score." Note that a particular piece of data may score differently on each of the three dimensions of correlation.

For our purposes in creating the US LCI database, the main consideration on this subject will be to provide sufficient documentation about the temporal, geographic, and technological basis of the data so that future users may employ methods such as suggested by Weidema, Kennedy, Kusko and others.

Table 1: Pedigree matrix with three data quality indicators (taken from Weidema, 1998)

Indicator score	1	2	3	4	5
Temporal correlation	Less than 3 years of difference to year of study	Less than 6 years difference	Less than 10 years difference	Less than 15 years difference	Age of data unknown or more than 15 years of difference
Geographical correlation	Data from area under study	Average data from larger area in which the area under study is included	Data from area with similar production conditions	Data from area with slightly similar production conditions	Data from unknown area or area with very different production conditions
Further technological correlation	Data from enterprises processes and materials under study	Data from processes and materials under study but from different enterprises	Data from processes and materials under study but from different technology	Data on related processes or materials but from same technology	Data on related processes or materials but from different technology

Data Inaccuracy

Data inaccuracy may be caused by imprecise measurement methods, (expert) estimations and assumptions, measurements from a small number of sites, and inadequate time periods of measurements pertinent to the processes involved. Various methods have been proposed to make data inaccuracy operational in LCA outcomes, such as analytical uncertainty propagation methods (Hoffman *et al.*, 1995; Heijungs, 1996), calculation with intervals and fuzzy logic (Chevalier & Le Teno, 1996; Becalli *et al.*, 1997), and stochastic modeling (Kennedy *et al.*, 1996; Kusko 1997, Huijbregts, 1998b; Maurice *et al.*, 2000). In particular, stochastic modeling, which can be performed by Monte Carlo simulation, seems to be a promising technique for making data inaccuracy in LCIs operational, as Monte Carlo simulation is widely recognized as a valid technique and the level of mathematics required to perform a Monte Carlo simulation is quite basic (Vose, 1996).

In development of the US LCI database, it is worth considering calling generally for recording of the sample size (number of processes on which an average is based), the minimum and maximum values reported, whether the sample of processes was random or not, and the estimated size of the universe of processes from which the sample was drawn. This information could provide a minimum basis for subsequent efforts by other users to quantitatively model and address implications of variability and uncertainty in their LCI applications.

Critical Review

The role of expert review is also essential in reducing errors and uncertainty in data. In this project, data will be reviewed by multiple experts knowledgeable in the processes under review. This includes not only internal checking, comparison and review, but also review by external experts and most importantly, experts from the sectors which have provided the data in the first place

Acknowledgment

Sections of this paper were copied from Huijbregts 2001

References

Beccali, G, Beccali M, Cellura M (1997): Fuzzy set application in life cycle inventory of building materials. Paper. Building and the Environment: Proceedings of the Second International Conference. Volume 1: Assessment methods and natural resources. Centre Scientifique et Technique du Bâtiment, Paris

Chevalier JL, Le Téno JF (1996). Life cycle analysis with ill-defined data and its application to building products. Int J LCA 1 (2): 90-96.

Huijbregts, M., G. Norris, et al., "Framework for Modeling Data
Uncertainty in Life Cycle Inventories", Int J LCA 6(3):127-132, 2001.

Heijungs R (1996): Identification of key issues for further investigation in improving the reliability of life cycle assessments. J Cleaner Prod 4 (3-4): 159-166

Kennedy DJ, Montgomery DC, Quay BH (1996): Data quality. Stochastic environmental life cycle assessment modeling. A probabilistic approach to incorporating variable input data quality. Int J LCA 1 (4): 199-207

Kusko, B, and R. Hunt (1997) Managing Uncertainty in Life Cycle Inventories, Society of Automotive Engineers, paper 970693

Maurice B, Frischknecht R, Coehlo-Schwirtz V, Hungerbühler K (2000): Uncertainty analysis in life cycle inventory. Application to the production of electricity with French coal power plants. J Cleaner Prod 8 (2): 95-108

Vose D (1996): Quantitative risk analysis. a guide to Monte Carlo simulation modeling. John Wiley & Sons, West Sussex

Weidema BP (1998): Multi-user test of the data quality matrix for product life cycle inventory data. Int J LCA 3 (5): 259-265

Weidema BP, Wesnæs MS (1996): Data quality management for life cycle inventories: an example of using data quality indicators. J Cleaner Prod 4 (3-4): 167-174