Quality Assessment for LCA

Nico W. van den Berg
Gjalt Huppes
Erwin W. Lindeijer
Bernhard L. van der Ven
M. Nicoline Wrisberg

CML
Centre of Environmental Science

TNO Environment, Energy and Process Innovation

ENVIRONMENTAL
IVAM
RESEARCH

Quality Assessment for LCA

Nico W. van den Berg
Gjalt Huppes
Erwin W. Lindeijer
Bernhard L. van der Ven
M. Nicoline Wrisberg

CML Report 152
ISBN:

Content

Foreword

LCA for decision support lacks a systematic quality assessment to guide makers, peer reviewers and users of LCA studies. Trust is the main basis acceptance, which usually is not present in adversary situations. Without systematic quality assessment LCA may easily degrade into a mere PR instrument. The only tool for integrated environmental assessment of technology related choices, would then be lost for really useful applications. The authors, from three main Dutch institutes involved in LCA, came together to see how this situation could be improved. Starting point for our analysis is that LCA is a simple model using data which does not allow for regular statistical analysis. So we had to develop a more general strategy for dealing with this situation. On the other hand, our aim was and is to arrive at an operational method and procedure for quality assessment of LCA outcomes. To avoid double work, we also had to analyse the work which had been done on quality assessment in LCA already. Without financial support, we started the job, two years ago, making the work to a large extent "homework". We now close off this period with the current paper as a result. The paper has three parts, one dealing with the strategy for quality assessment, one part with a survey on what has been done in the field, and a third part indicating what an operational method for quality assessment could look like. We did not integrate the three parts, so in principle each can be read on its own.

Further work we think would require an extended effort, as a substantial project involving specialists from the field of LCA and specialists from the field of quality assessment. We hope to be part of that group. This group would preferably be more international. We started now to look for funding for this essential and still lacking part of LCA, so we hope to come back on the subject.

the Authors

20 September 1999
Leiden
Netherlands

Part 1: A Framework for Quality assessment

Abstract

For the acceptance of LCA results, it is crucial that the level of confidence in outcomes is appropriate for the decision to be made. Currently, there is no general method for establishing the level of confidence in LCA results. A framework for such a general method has been established, distinguishing between reliability and validity of results, and combining these with a number of elements of pedigree. This framework has been used to assess current approaches to data quality assessment in LCA. Most quality assessment methods consider only the quality of input data. Moreover, they generally concentrate on the inventory phase. This paper widens the scope of quality considerations. Besides inventory data characteristics, it is also important to assess the quality of the input data for the impact assessment and the overall model, and, given the restricted nature of quality assessment in LCA, to employ circumstantial evidence from a broader quality perspective: the pedigree. A second paper addresses the operationalisation of this framework.

1. Introduction

LCA today is being used for decision support. With marketing analysis, also used for decision support, a feedback mechanism that weeds out faulty analyses exists. If predictions turn out wrong - products flop, costs are too high, markets smaller than expected, etc - the marketing analyst loses his job. With LCA, there is no such a feedback mechanism. Here, confidence in outcomes can be based only on the quality of the input data and the quality of the models used. For the acceptance of LCA results, it is crucial that the level of confidence is appropriate for the decision to be made. Currently, no general method exists for establishing the level of confidence in LCA results. The aim of this paper is to build a framework for such a general method and assess current approaches to LCA data quality analysis within this framework. It makes use of the survey of existing quality assessment methods (Van der Ven et al., 1999). A second paper will elaborate a still incomplete operationalisation of this framework.

Work on this subject of data quality analysis in LCA has been going on for quite some time. The first highlight was the SETAC Workshop on data quality in Wintergreen in 1992 (Fava 1992). Eleven different approaches, which were subsequently pursued, will be discussed in this paper. However, none of these has been received as being fully appropriate to the need. Two recent developments may prove this point. Quantification of uncertainty has been identified as a top research priority by LCANET (Udo de Haes and Wrisberg, 1997), and SETAC Europe has just installed a new working group on "Data Availability and Data Quality". Why have developments in the field of quality analysis been so disappointing?

In the case of LCA a number of basic difficulties can be distinguished, which render data quality analysis more complicated than in the case of most other decision support systems. First, statistical analysis at the level of basic data gathering is in fact lacking. This implies that usual error propagation techniques are not applicable. Secondly, many different types of input data play a role, not only in the inventory but also in the characterisation, normalisation and, if present, evaluation steps. This makes combining different measures on input data quality difficult in itself and quite impossible if these measures are qualitative. Thirdly, the model used in LCA to transform input data into results cannot be tested. The reasons for this are diverse. The very simplified nature of the model makes a comparison with real life developments cumbersome, as does the arbitrary quantity of the functional unit. In most cases the effects in LCA are not specified in terms of place and time and hence cannot be "seen". This is so in the inventory analysis and even more so in the environmental models for characterisation. The nature of the modelling, usually some type of steady state modelling, does not allow specific predictions. At best, some parts of the model may be tested independently of the LCA context. Fourthly, the number of input data items used in LCA is extremely large. In the inventory alone, medium-sized LCA may already be made up of five hundred processes with two hundred items per process. The task, therefore, is to combine one hundred thousand independent flow items, each with its own level of reliability, into an overall level of confidence in outcomes, together with a number of other confidence-related aspects like completeness and validity of flow types, validity of processes, and overall model validity. Finally, if there is no weighting procedure to transform the characterisation results into a single score, confidence can only be specified for the individual impact categories, specified in the characterisation models. The quality of results may be quite high for global warming, but poor for human toxicity. The overall level of confidence then cannot generally be established. If there is a weighting procedure, one can hardly expect it to be generally accepted, introducing uncertainties of another kind.

Given the fairly poor state of the art in quality assessment, it becomes necessary to take into account indirect evidence for quality. A sensitivity analysis of the data and modelling choices may indicate that results are not dependent on any of the specific choices. A comparison of results with those of allied studies may show similarities, or dissimilarities, the latter leading to a lower level of confidence. Technical reproducibility of results, using the same data and modelling choices but executed by other scientists using other software, increases confidence. This is also the case if external checks, e.g. in the form of a peer review, have supported the results yielded. Anomalies, such as one particular economically inferior process dominating the results, reduce confidence. And finally, the quality judgement itself is more valuable if made by independent experts of high esteem. So even if a more or less formalised method for establishing validity and reliability were established, there would still be a substantial number of qualitative aspects of relevance for the confidence level of the study results.

To further complicate matters data quality analysis may be carried out for a variety of purposes. In the context of decisions-support, one may query how much confidence one has that a certain decision is the right one. This very legitimate query then encompasses the question whether the goal, support for that decision, is properly reflected in the scope of the study, for example in the definition of the functional unit and in the options investigated. Alternatively, one might assess the quality of a study solely in relation to the scope chosen, ignoring the question of the appropriateness of that choice.

None of this means that quality analysis should simply be forgotten, of course. As decisions have to be made anyway, using all information relevant to the quality of output data in the best possible way, is better than ignoring that information. In this paper, we hope that by specifying the full nature of the problem the discussion can be structured. Given the lack of basic data on quality and the complexity of the situation, no perfect solution is possible. However, a comprehensive overview of the field may help develop the most reasonable solution possible today.

The paper is structured as follows: section 2 sets up the framework for a quality assessment model, including the relations with procedural aspects like peer review. Section 3 sketches the framework of a quality assessment model in relation to the main phases of LCA. A discussion closes the paper in section 0.

2. Model and quality: the starting points

2.1. Model

Models are used to reflect certain aspects of the real world. By using information from the real world, application of a model leads to formulation of statements about the real world. When applying a model, input data is generally fed through the model in order to generate the output data, i.e. the results. The model describes which transformations, combinations and calculations are performed on the input data. Results are therefore determined entirely by the combination of input data and model, as shown schematically in figure 1.

This distinction between model and data is by no means self-evident. In LCA, the process flow chart is often seen as the model. Here the term model is used to describe the logical and computational structure of LCA. We treat model parameters here as data, narrowing down the model to the choice of relations. Thus, the inventory system itself is not the model. The model is the way that information on processes is transformed into an inventory system, stating how a functional unit influences the environment, e.g. using technological and possibly economic relations in the model.

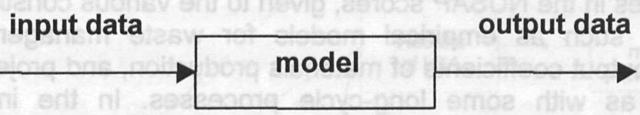


Figure 1 Input data leading to output data by being fed through the model

2.2. The quality of decision support

When using LCA for decision support, LCA information is combined with other information, environmental and non-environmental, to arrive at the decision. Ultimately, one wants to be sure that the right decision is being made, or at least to know how sure one is about a decision being the right one. A case in point is a decision on the most efficient investments to be made for environmental improvements, combining economic and environmental information in one score. This overall level of confidence will not be discussed here; we shall concentrate purely on the level of confidence in the LCA advice, and how this may be determined. Depending on the types of non-LCA information of relevance in a particular case, the final step in assessing overall decision confidence is still to be added.

The quality of decision support depends on the quality of the data and model. Under ideal circumstances, models are corroborated and uncertainty in data is quantified. In LCA, and more generally in decision support in real life circumstances, this is not the case, however. Measures of spread are not available, measures on incompleteness are lacking and the status of models is unclear. Still decisions are being made and, intuitively, most people are able to assess the quality of outcomes. Funtowicz and Ravetz (1990), further referred to as 'F&R', have developed a comprehensive approach to the problem of uncertainty regarding quality. In today's sub-optimal situation, the question is how the most relevant information available on quality, can be processed in such a way as to arrive at the most reasonable overall quality assessment. F&R divide the assessment into five areas, which together form the acronym NUSAP: *Numeral, Unit, Spread, Assessment* and *Pedigree*. The first three relate to the nature of quality assessment in terms of scaling, units and measures of spread. These are also applicable in situations where ratio scale measurements are not possible and are hence relevant for LCA. The next, *Assessment*, covers elements more or less technically related to the validity of the outcomes. Often *Assessment* will function as a correction factor for too much optimism on the first three, e.g., accounting for missing items and indicating the validity of the models used. Finally, *Pedigree* comprises an external view of the research and its outcomes, relating them to comparative validity and overall confidence and credibility of results. It involves placing the model in a range, which is related to the scientific status of the model, the nature of the data inputs and meta-criteria like peer acceptance and colleague consensus.

In flavour, we will follow F&R, but not in terms of their NUSAP acronym. There are two main reasons for this. One is the extremely complicated nature of LCA, which makes the NUSAP scheme a relatively complicated affair for LCA. The second reason is that NUSAP is designed for larger decisions, with a substantial amount of effort dedicated to the quality assessment of results. The function we have in mind for quality assessment methods is a more routine application, requiring only a limited amount of work. The operational quality assessment methods we have in mind should, therefore, not only indicate where the strengths and weaknesses lie, as F&R do, but also aim to aggregate these elements into an overall judgement in a more or less formalised procedure.

The practical difficulties involved in applying the NUSAP scheme go further. There will be fairly substantial differences in the NUSAP scores, given to the various constituent elements of the inventory, such as empirical models for waste management, historical data for input and output coefficients of materials production, and projective sketches of technologies, as with some long-cycle processes. In the impact assessment, toxicity scores have a very different status from climate-forcing scores, and these are very different again from normalisation data and quantitative weighting sets used in evaluation. Hence, the problem of quality assessment in LCA is not just a matter of ensuring that the assessment method is suitably adapted to the level of quality encountered in the LCA sub models, but that very different types of quality aspects are aggregated into an overall pronouncement on quality. For the time being at any rate, using the output of NUSAP as input for this aggregation seems too complicated in the context of LCA.

The general structure of our framework is depicted in Figure 2, with Level of Confidence as the ultimate aim of quality assessment. Going backwards to its constituent factors, to the right in the scheme, the level of confidence in the LCA advice is based on the LCA results themselves and on their overall quality. If a choice is to be made between two alternatives, a low-quality study combined with very large differences between alternatives, gives the same confidence as a high-quality study with much smaller differences. We do not assume that this step is a formalised one, only that it is structured. If it is not formalised, the next step in the decision procedure, considering other, non-LCA types of information and their levels of confidence, is also a qualitative one.

The central question then is what procedure should be adopted to arrive at a judgement on the overall quality of LCA results. We follow traditional lines here in distinguishing between the quality of the model used and the quality of the input data. As 'model' is used in a restricted sense here, the data comprise data on the alternatives to be compared, data on economic processes, data on environmental processes and evaluative data on how to combine different outcomes into one or a limited number of scores, as when using a set of weights.

As a subsequent step, the quality of the model and input data is related to the two main aspects of validity and reliability. *Reliability* usually relates to consistency. It is a measure of the reproducibility of the data and the model computations. *Validity* usually refers to the requirements implied by the decision to be supported. It indicates the extent to which the model and data, and hence the results, properly indicate what one wants to know. The validity and reliability of both input data and model together cover the same ground as Spread and Assessment with F&R. One option for structuring the whole may be to assess the overall quality of the data and the overall quality of the model, combining validity and reliability steps in each, as Morgan and Henrion (1990) and van Asselt et al. (1997) suggest. However, we have chosen to follow F&R in this respect, grouping all the more or less technical aspects of reliability together, similar to their category of Spread, and all the more or less technical aspects of validity together, similar to their category of Assessment. Operationalisation here involves an assessment of the quality of the model and the quality of the input data separately. The more external characteristic of Pedigree covers in a more internal, technical way what cannot be adequately taken into account under the headings reliability and validity. The more complete the analysis of validity and reliability becomes, the smaller will be the role played by Pedigree. Also, the higher the quality of the model and data becomes, the less important Pedigree aspects will become.

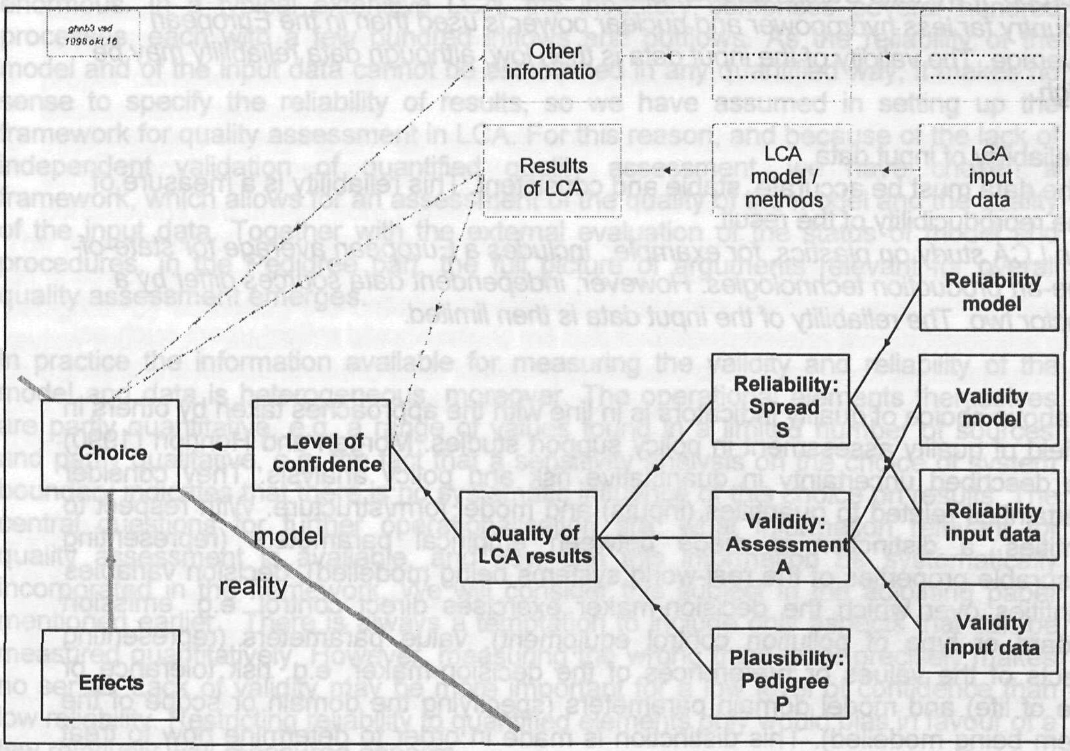


Figure 2 A framework for quality assessment in LCA

Combining validity and reliability on the one hand with input data and model on the other, we can distinguish four basic quality indicators.

1. Validity of model
   The model must indicate what we indeed want to know in the context of decision support. This validity is a measure of the extent to which the model matches external reality.
   In the inventory step most LCA models do not take secondary reactions into

account: e.g., extra demand because of a certain choice is taken at its full extent, disregarding market responses to the higher prices induced by that extra demand. Ignoring such real world mechanisms reduces model validity.
In the impact assessment step, to give another example, the problem of 'climate change' is modelled using the Global Warming Potential (GWP) of substances as established by the IPCC on the basis of a number of models. GWP indicates the absorption of infrared radiation, integrated over a certain period of time, taking into account its removal from the atmosphere. The validity of this part of the LCA is based on the extent to which this global warming potential is the right predictor for absorption and whether absorption is the right predictor for climate change.

2. Reliability of model
   The model must indicate the same results each time it is applied to the same set of input data. This reliability is a measure of the model's internal consistency.
   Computational procedures involving multiple loops may, for example, depend on where one starts computing, back down strings of processes. In these cases reliability is lower than in computational procedures involving matrix inversion.

3. Validity of input data
   Relevant input data must be used. This validity is a measure of the extent to which the types of input data chosen, are appropriate for the external requirements.
   In a specific LCA we may need a figure for power generation emissions valid for Europe. If we take Dutch national data as a sample, we ignore the fact that in this country far less hydropower and nuclear power is used than in the European average. The validity of the input data is then low, although data reliability may be high.

4. Reliability of input data
   The data must be accurate, stable and consistent. This reliability is a measure of the reproducibility of the result.
   An LCA study on plastics, for example, includes a European average for state-of-the-art production technologies. However, independent data sources differ by a factor two. The reliability of the input data is then limited.

The above choice of quality indicators is in line with the approaches taken by others in the field of quality assessment in policy support studies. Morgan and Henrion (1990) have described uncertainty in quantitative risk and policy analysis. They consider uncertainties related to quantities (inputs) and model form/structure. With respect to quantities, a distinction is made between empirical parameters (representing measurable properties of the real-world systems being modelled), decision variables (quantities over which the decision-maker exercises direct control, e.g. emission standard or type of pollution control equipment), value parameters (representing aspects of the values or preferences of the decision-maker, e.g. risk tolerance or value of life) and model domain parameters (specifying the domain or scope of the system being modelled). This distinction is made in order to determine how to treat uncertainty. The only quantities whose uncertainty may be appropriately represented in statistical terms are empirical quantities measured on a ratio scale. Uncertainties related to the other types of variables may be expressed by parametric sensitivity analysis. Overall, quality assessment of these quantities is not very structured.
Van Asselt et al. (1997) describes uncertainty in results, likewise distinguishing between quantities and structure, termed *loci of uncertainty*. Quantity and structure are somewhat comparable to our data and model. Three *types* of uncertainty are involved: technical, methodological and epistemological. They each have different *sources*, such as statistical variation, linguistic imprecision, subjectivity, approximation or disagreement. To some extent standardisation, as currently taking

place for LCA in ISO, can thus reduce uncertainty. The main message of Van Asselt et al. with respect to the quality of outcomes is, that models never provide the full truth. They can be used to gain a partial understanding of the world around us. Methodological and epistemological uncertainty will be reflected mainly in the Pedigree measure on quality, discussed below.

After establishing the main framework for quality assessment of LCA results, the real work starts: how to practically measure the four main indicators, how to arrive at a Pedigree score, and how to check whether the resultant overall quality measure is "right"? Very good predictions combine high validity with high reliability. If predictions are less solid, this may be the result of low validity, low reliability or a combination of both. The latter then have to be established independently. In practice, there is no way to measure the factors determining quality on a ratio scale. This is not typical for LCA, but is the case generally in interdisciplinary decision support. The validity of models and data can only be established qualitatively; there are no ratio-scale measures for this purpose. With reliability, quantification is possible in principle, using the relevant branches of statistics. However, in LCA the options to do so are limited. First, there is no good measure of model reliability. Even a remark such as that made above on the superiority of matrix inversion over computing consecutive rounds of process strings, is not universally accepted. To assess the reliability of input data, in principle the option of quantification is available. In practice, however, the spread in data is not measured in a way amenable to statistical analysis. The work involved would be enormous. In a typical extensive LCA, the inventory comprises several hundred processes, each with a few hundred inflows and outflows. As the reliability of the model and of the input data cannot be established in any quantified way, it makes no sense to specify the reliability of results, so we have assumed in setting up the framework for quality assessment in LCA. For this reason, and because of the lack of independent validation of quantified quality assessment, we have chosen a framework, which allows for an assessment of the quality of the model and the quality of the input data. Together with the external evaluation of the status of model and procedures, in the Pedigree part, the full picture of arguments relevant for overall quality assessment emerges.

In practice the information available for measuring the validity and reliability of the model and data is heterogeneous, moreover. The operational elements themselves are partly quantitative, e.g. a range of values found in a limited number of sources, and partly qualitative, e.g. the fact that a sensitivity analysis on the choice of system boundary indicates that there is no systematic influence of this choice on results. The central questions for further operationalisation are: what information relevant for quality assessment is available, and how can this information be systematically incorporated in the framework. We will consider this subject in the adjoining paper mentioned earlier. There is always a temptation to include only aspects that can be measured quantitatively. However, measuring the wrong thing quite precisely makes no sense. Lack of validity may be more important for a low level of confidence than low reliability. Restricting reliability to quantified elements only would bias in favour of a few relatively well-measured aspects.

Given the mainly qualitative nature of quality assessment, why bother so much at all? Can a general view by an experienced practitioner not suffice? There are two main reasons for insisting on a stepwise, repeatable procedure. One is that, without such a procedure, there is no feedback for improving practitioners' judgement. Also, one often wants to identify the factors contributing to an overall judgement on the quality of results, to see how this quality can be improved effectively and efficiently. This last point also is important indirectly, for example in setting criteria for the minimum quality of the databases to be used in LCA.

Does our framework now cover all aspects potentially relevant for quality assessment? It seems so. Some types of information, relevant for quality assessment, do not fit into the framework developed for assessing model and input data quality. Procedural aspects are one example. The level of confidence in outcomes is higher if the party performing the LCA is unrelated to the firm making the superior product. Also, confidence increases when independent peer reviews have taken place. External comparisons may also contribute. Confidence increases further if the outcomes of similar studies indicate the same directions or even magnitudes. Such aspects, if not given due place in the assessment of quality and input data, should be involved in the *Pedigree* measure on quality. Some procedural aspects would use the quality assessment as an input and hence could not serve as an element of the equality assessment itself. Several factors, potentially relevant for Pedigree analysis in LCA, will be investigated in the next section. The operationalisations suggested by F&R (especially as formulated in Chapter 10) will be taken into account in our second paper, on operationalising the framework for quality assessment (Wrisberg et al. 1999).

2.3. Pedigree and procedure

The analyses for quality assessment in LCA cannot be specified in a 'hard' fashion. Information on the models and data, relevant for quality assessment, is heterogeneous and cannot always be combined in an unequivocal way. In practice, there is not even any operational method for quality assessment. Hence, several procedural safeguards have been developed for quality control, including various forms of peer review. In establishing the level of confidence of the outcome of an LCA study, such procedural aspects may play an independent role.

Major procedural aspects are addressed here, showing on the one hand their relation to the quality assessment of LCA results and on the other how an operationalisation of the framework can contribute to such procedures.

Role of the commissioner

It is obvious that the commissioner should have a clear view of the goal of the study. There are several goal-related questions that he or she should reflect upon, such as:

- Is it sufficient to limit the goal to predicting potential rather than of actual impacts?
- Can the goal really be met adequately with the available data and model?
- Does the stated goal correspond to the actual question or impulse for performing the study; which decisions are to be supported?

Ideally, such questions should be discussed with experienced LCA practitioners, other LCA commissioners and interested parties, forming a bridge between the LCA and the 'outside world'. Such a procedure may lead to restatement of the goal and possibly to other types of environmental analysis than LCA, such as Risk Assessment, Substance Flow Analysis or some form of Cost-Benefit Analysis.

Subsequently, as with any study, the commissioner should check whether the study is indeed being performed in accordance with the goal. A guiding committee, consisting of relevant experts and/or stakeholders, may be of help here, providing data, expert knowledge and/or feedback based on public opinion. Depending on the goal, a Peer Review procedure (see below) can be initiated at the start of the study to ensure direct feedback to the commissioner and LCA practitioner(s) on the choices made during the study (from scope to interpretation).

In any case, the commissioner should be clear about the required quality of the LCA, preferably in terms of a systematic quality assessment model, e.g., based on the framework developed here.

Peer Review

In the SETAC Code of Practice (Fava et al. 1992) the LCA Peer Review is laid down as a process accompanying the study rather than merely being the more traditional review carried out afterwards. In ISO 14040 (ISO, 1997) three different types of review are distinguished: an internal expert review (for internal use of the study only), an external expert review (e.g. for comparative assertions disclosed to the public) and a review by interested parties (the need for which depends on the goal and scope of the study). The Peer Review should check (Klöpfer, 1997):

- the completeness and consistency of the study
- the transparency of the report
- whether the data quality is adequate for the goal and scope stated
- whether SETAC guidelines, ISO standards and other relevant external requirements are adequately met
- whether the impact assessment is appropriate for the systems studied.

ISO states that the Peer Review will enhance the study's credibility by helping to focus the scope, data collection and model choices (improving the first 3 phases of an LCA) and by supplying critical feedback on the conclusions and how they have been reached, thus improving interpretation.

In the Peer Review a qualitative assessment is made of the LCA quality, based on expert experience. Checks will concern such aspects as:

- inventory-related items, including detailed verification of selected data(sets) and plausibility checks on other data and inventory results (including ad hoc checks on possible mistakes)
- proper consideration being given to with the possibilities and limitations of impact assessment methodologies
- interpretation items such as the dominance analysis, uncertainty analysis, sensitivity analysis, consistency checks, completeness checks and how conclusions are drawn from these.

The importance of such items for the quality of LCA results is generally estimated in an unstructured manner. The result of a peer review can only be trusted as matter of pedigree: by accepting the credibility of the peer reviewers and by judging the peer review report. An additional aid to ensure the quality of the peer review and increasing its transparency is to have the peer reviewers use an operational quality assessment procedure during assessment and reporting. Again, the framework developed here may guide that procedure.

Data verification

Verification of the process data used in LCA may be performed as part of the (peer-reviewed) LCA. It may also be performed separately. In the latter case, data verification serves to specify data quality prior to inclusion in larger databases like ETH, BUWAL and EcoQuantum, or prior to external communication of environmentally relevant information, as to suppliers, customers or consumers. Dutch examples of the second category are the DALCA project (chemical industry) and the MRPI project (building sector).

Data verification can be seen as a kind of peer review on individual processes in a database. One way or another the quality of the process data has to be assessed in a credible manner. Again, by using an explicit quality assessment framework this data verification process can be structured and made more transparent.

LCA interpretation

In the interpretation phase of LCA as proposed by ISO, many checks are to be made on quality-related issues. Examples are dominance analysis, which investigates if one alternative is better than another with respect to all relevant respects, and sensitivity analysis, which shows how assumptions regarding models and variations in input data affect results. Consistency checks are relevant for assessing the reliability of the model. Internal plausibility checks indicate what might be wrong, e.g. one small process dictating the overall score for an environmental impact category. Such checks may be performed by the LCA practitioners themselves and contribute to the quality of the conclusions drawn. Some of these checks can be performed within the same quality assessment framework discussed in this paper.

Once a more or less well-structured method for data quality assessment has been developed, procedural and analytic aspects can be better separated. At the moment some procedural elements are analytical but not very operationally formulated, like "consistency checks", while others are procedural, as when ISO norms have been followed and stakeholders involved. Once a data quality assessment method becomes available, it can be used directly by practitioners and may become part of quality assessment procedures in the same way as peer reviews today. Given the state of the art, some analytic elements cannot be adequately incorporated in the method, such as dominance analysis and sensitivity analysis. These can play an independent role in establishing the level of confidence in results, as elements in the Pedigree part of quality assessment.

2.4 A survey of relevant factors for quality assessment

As a summary of this section, Table 1 reviews how the Spread, Assessment and Pedigree aspects of F&R are related to the data quality indicators described in this paper.

Table 1 Survey of factors relevant for quality assessment in LCA

	Main quality aspects	DQ indicator	Factors
overall quality	S Reliability ≈ Spread	Model reliability	• Reproducibility of transformation • Reproducibility of computation
		Input data reliability	• Uncertainty • Completeness • Variability
	A Validity ≈ Assessment	Model validity	• Steady state versus real dynamics • Linearity • Goal and scope match • Scope properly elaborated in functional unit, allocation methods and characterisation models • Potential vs. actual effects • Disregarding local circumstances • All relevant empirical mechanisms included? • Models behind equivalency factors
		Input data validity	for 4 types of input data: • System boundaries • Representativeness
	P Pedigree	Procedural aspects	• Data verification • Sensitivity analysis • Gravity analysis • Dominance analysis • External plausibility • Parts of model tested • Comparison of outcome with similar models • Status of software provider

3. Framework for quality assessment in LCA

This section outlines the framework of a quality assessment model in relation to the main phases of LCA. It concentrates on the Spread and Assessment look-alike parts of F&R as described in the previous sections (see Table 1). It does not deal with any Pedigree aspects. It first describes LCA, its phases, and its shortcomings. Next LCA is related to the aforementioned individual quality indicators for input data, model data and output data.

3.1. What LCA is about

LCA indicates how the fulfilment of a certain function by a product or service can influence the environment. One or more alternatives that might be employed to fulfil that function are specified first in the goal definition and then, in greater detail, in the scope of the study. The function is fulfilled by the use of a product or service, itself produced by means of other goods and services. The use of all these goods and services causes environmental interventions, during resource extraction, in manufacturing, while using the product, and in processing the wastes from all these stages of the life cycle. The aim of LCA is to specify all the interventions caused and assess their environmental impacts.

3.2. LCA phases

3.2.1. Goal and scope

In the ISO definitions the goal and scope define the subject of the study, for whom it is intended, who does the work, etceteras. In this phase, decisions are made on such aspects as the allocation methods and impact assessment models to be used. All kinds of crucial assumptions and statements are made here. The goal and scope phase deals with four different kinds of choices:

1. Goal choice, determining the study topic and the reasons for performing the study.
2. Goal-related choices, determining the central object of analysis, the functional unit. These choices act as input parameters for the scope, and do not influence the quality of the results for a given scope.
3. Reference flow choices, fulfilling the functional unit.
4. Methodological choices that influence the quality of the results, e.g. choice of allocation method, characterisation methods and data to be used for the analysis.

3.2.2. LCA inventory phase

The entire system studied is considered as consisting of unit processes. These unit processes define both the environmental interventions and the mutual linkages between these unit processes in the economy. In order to construct these unit processes, a wealth of potential input data is assessed as to their potential use. They are filtered during data selection, using choices and criteria set in the goal and scope phase. The intermediate result is determined by the choice of input data.

Next, the first steps of the model are performed; involving both transformation and calculation. In the transformation step the process flow chart is constructed, using the data selected and applying the choices and criteria for allocation. Having compiled the relevant process data, the inventory table is calculated, using appropriate algorithms. This step combines all the environmental interventions due to each of the unit processes into one aggregated set of environmental interventions for the system analysed.

3.2.3. LCA classification/characterisation phase

The next LCA phase yields an assessment of the environmental interventions, relating these to environmental problems. It makes use of basic data on the various environmental problems, and a selection of basic data and model must therefore be made. The intermediate result is formed by the chosen input data.

Next, the model is applied by performing the transformation and calculation step, comprising transformation of the basic data according to the models chosen in the scope and derivation of equivalency factors indicating the contributions of substances to the respective environmental problems. Equivalency factors are derived in a variety of ways, for instance using the LC50 as an input parameter for the toxicity measure. In principle the specialised models underpinning characterisation form part of the LCA model. There may be a lack of confidence in the IPCC climate models used to compute GWPs, for example. Finally, in the calculation-step, the inventory figures are multiplied by the equivalency factors found, resulting in the environmental profile.

3.2.4. LCA weighting phase

An optional LCA phase is weighting of the environmental problems into a one-figure score for the environmental load: the environmental index. For this final transformation, specific information is used to find the relevant set of weights for each of the problems/impact categories. For instance, policy reduction targets may be used to derive the weighting factors for various environmental problems. The various elements of the environmental profile are then multiplied by the respective weighting factors, to yield the environmental index.

3.3. Shortcomings and approximations of the LCA model

- Being a simplified model, LCA yields result that differ in several respects from "what will really happen", but how they differ is too difficult to predict and evaluate. The inventory model is a generally comparatively static model, for example, built from a number of processes each described in terms of linear input-output relations, describing a sort of steady state. In reality, however, we know there are dynamic non-linearities, market mechanisms, continuous technological developments, etc.
- actual versus potential effects: the environmental models used in LCA describe potential environmental effects of emissions.
- linearity: LCA presumes linearity of production scale and of environmental effects related to the functional unit.
- local versus global: LCA generally treats local and global information and effects in the same way, abstracting mainly from local aspects.

3.4. LCA input data and its quality

The input data for the LCA model consists of the filtered information used as input for the model transformations and calculations.

- For the scope phase, the input data consists of the information necessary to properly specify a functional unit.
- For the inventory phase, the input data consists of the information necessary to compile the unit process descriptions, including both the technical production data and the environmental interventions.
- For the classification/characterisation phase, the input data consists of all the information necessary to compile operational equivalency factors for the characterisation.
- For the weighting phase, the input data consists of all the information necessary to construct weighting factors.

The validity of input data should indicate whether the proper input data has been chosen. It can be seen as a measure of the extent to which the raw data has been made correctly selected. The criteria relate to the scoping choices made: to what extent do these choices match the scope of the study? It should be noted that in practice the validity of input data is never perfect. For instance, cut-offs are made in every LCA study; their influence on the results can only be estimated. The validity of the input data should therefore indicate how imperfectly the data and system boundaries have been chosen, and the extent to which the results are influenced. This indicator should therefore cover:

- Validity of system boundaries:

Has data relevant to the scope been excluded? (This is also called completeness at the process level.)
Have system cut-offs been made in accordance with the scope?

- Representativeness of chosen data:

Has data been chosen in accordance with the scope?

The reliability of input data should indicate whether this data is stable and consistent. Relevant factors are:

- uncertainty: What is the measurement error or estimation range?
- completeness: Is there any data lacking?
- variability: What are the ranges for a given representativeness?

Remark

The subdivision into validity versus reliability presumes their independence. With regard to input data, however, there is a trade-off between representativeness (validity) and variability (reliability), since a measure of variability can only be given for a certain representativeness, as is illustrated by following example. Suppose European power generation data gives an emission of 0.5 ± 0.2 kg CO2 per kWh_e. While national data gives 0.4 ± 0.1 kg CO2 per kWh_e. If we accidentally take the national data instead of the European, we should adjust either the variability or the representativeness. This problem is not too serious for the conceptual part of quality assessment, but should be kept in mind during operationalisation.

3.5. LCA as a model and its quality

The LCA model makes an assessment of input data. The goal and scope phase sets all the choices and criteria, and data collection defines the set of input data to be used. Here we assume that data collection is not included in the LCA model itself. The LCA model aims to describe how fulfilment of a given function influences the environment, or at least assess the relative influence of different options. Quality assessment, which generally pertains to the generic question "how valuable is the result?", can therefore be specified for LCA more precisely: "to what extent do LCA outcomes correspond with real environmental risks?"

The model validity should indicate whether use of LCA really can provide solid indication of environmental harm, and whether the model is applied consistently. This indicator should therefore cover:

- Validity of scope:

To what extent does the scope match the goal?

- Validity of modelling choices:
  Have the relevant empirical mechanisms been incorporated in the model?
  Does the functional unit choice match the scope?
  Do the chosen allocation methods match the scope?
  Do the chosen characterisation methods match the scope?

The model reliability should indicate whether the LCA model correctly yields the true environmental load and if it gives reproducible answers. This indicator should therefore cover:

- Reproducibility of the model results

To what extent does the transformation model employed give reproducible answers?
To what extent does the calculation model employed give reproducible answers?

Data transformation

- general: The transformation step comprises the transformation of the chosen input data into a manageable form, on the basis of the criteria defined in the scope.
- inventory: All the chosen inventory data is fed into unit processes, which are later linked during the calculation step.
- characterisation: The data chosen for the impact assessment stage is converted to equivalency factors. This transformation uses the characterisation models chosen during the scope.
- weighting: The data chosen for weighting is converted to weighting factors. This transformation uses the models chosen during the scope.

Calculation on transformed data

- general: the calculation sums the respective transformed input data (for the inventory) and combines it with the intermediate results (for the characterisation and weighting). It also makes use of some of the choices and criteria defined in the scope.
- inventory: All the unit process data, i.e. the transformed inventory data, is combined. Thus an inventory table for the entire system is constructed. This calculation uses the allocation procedures and calculation algorithms chosen during the scope.
- characterisation: All the equivalency factors, i.e. the transformed characterisation data, are combined with the inventory table. Thus the environmental profile is constructed.
- weighting: The weighting factors, i.e. the transformed weighting data, are combined with the environmental profile. Thus the environmental index is constructed.

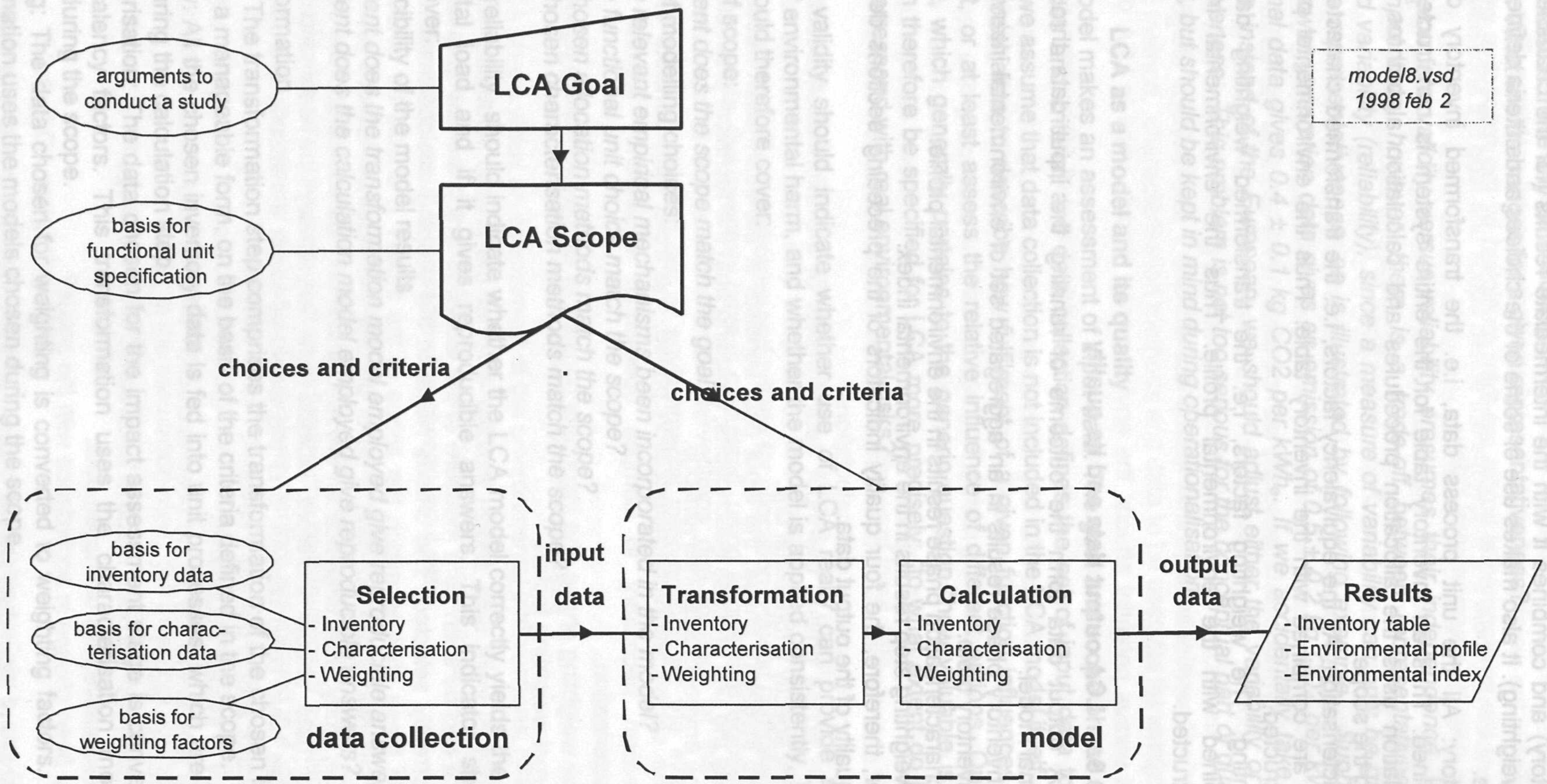
3.6. LCA output data and its quality

The LCA output data form the outcome of running the input data through the operational model.

- The inventory phase results in an aggregated set of environmental interventions: the inventory table.
- The characterisation phase results in the environmental profile.
- The weighting step results in the environmental index.

Together, therefore, the four quality indicators of the preceding sections define the overall quality of the output data.

Figure 3. Schematic representation of how LCA input data leads to LCA output data by being fed through the LCA model. The figure does NOT illustrate the difference between reliability and validity.



4. Discussion

Use of LCA in decision support requires yardstick to measure the confidence (= uncertainty mirrored) one may have in the advise based on its outcomes. This is a complex matter because of the many inhomogeneous sources of uncertainty stemming from different types and often large numbers of input data and the various models used in the different phases of an LCA. This explains why LCA studies accompanied by a structured quality assessment are still an exception.

The discussion about quality assessment to date has focussed on inventory input data. There has been little discussion of the validity of models, and it has been, in other contexts, e.g. whether weighting makes sense. Several quality assessment approaches have been proposed. Explicitly or implicitly, they all employ quality indicators, and some authors are also considering use of statistical analysis. We think that statistical measures of spread may be of some use. However, current data does not have an indication of spread and it is not even clear the spread of what exactly is to be specified. Key concepts such as representativeness, completeness, etc. are interpreted in very different ways.

The framework developed in this paper covers not only the quality aspects related to the input data in the inventory, as is generally the case. It also addresses quality aspects related to the input data used for the other LCA phases and to the model used. For analysing the quality of LCA results, four basic factors are proposed: validity and reliability of models, and validity and reliability of input data. The LCA phases form the third dimension of the structure. Ideally, these four aspects suffice for a quality analysis of LCA results. In practice, however, they do not, and many relevant aspects can find a place in the Pedigree category. One can go a step further, specifying the confidence in the advice based on the LCA results in relation to the goal and decision to be supported as the ultimate function of conducting LCAs. In this approach, the outcomes themselves then also play an independent role: for a given overall quality, large differences between alternatives give a higher confidence.

In the course of this paper we have developed a framework for quality assessment and have elaborated it, although not yet operationally, with reference to a number of factors that seems intrinsically relevant. The framework does not cover all the quality aspects relevant for confidence in LCA outcomes.

The question now is what the overall structure looks like and how the different elements can be combined to achieve the ultimate aim: a statement on the level of confidence of advice based on the outcome of an LCA study. Table 1 (Section 2.4) reviews the elements. Several of the factors specified are not independent, and this should be made clear to the quality assessor; otherwise, operationalisation must be further refined.

To our mind a more systematic approach is possible within the framework developed here. Operationalisation of such a method forms the subject of our forthcoming paper.

5. References

van Asselt 1996
van Asselt, M. B. A., A. H. W. Beusen and H. B. M. Hilderink, *Uncertainty in integrated assessment: A social scientific perspective.* In Environmental Modelling and Assessment 1 (1996) pp.71-90.

Beck 1992
Beck, C., *Risk society: towards a new modernity*. SAGE publications Ltd., London, 1992. pp.260.

Bevington and Robinson 1992
Bevington, P.R. and D.K. Robinson, *Data reduction and error analysis for the physical sciences*, McGraw-Hill, Singapore, 1992. ISBN 0-07-911243-9.

Clift et al., 1997
Clift, R. et al., *Towards a coherent approach to life cycle inventory analysis*. Final document of SETAC-Europe Working Group on Data Inventory, April 1997.

Consoli 1993
Consoli, F., *Guidelines for Life Cycle Assessment, a Code of Practice,* Society of Environmental Chemistry and Toxicology (SETAC). Brussels, 1993.

Coulon 1997
Coulon, R., V. Cambobreco., H. Teulon and J. Besnainou. *Data quality and uncertainty in LCI*. Int.J.of LCA, Vol.2, no. 3, p. 178.

Fava 1992
Fava, James, et al., *Life cycle assessment data quality: a conceptual framework.* SETAC workshop 1992, Wintergreen, USA.

Funtowicz and Ravetz 1990
Funtowicz, S.O. and J.R. Ravetz, *Uncertainty and quality in science for policy*, ISBN 0-7923-0799-2. Dordrecht, 1990.

Heijungs 1996
Heijungs R., *Identification of key issues for further investigation in improving the reliability of life-cycle assessments.* Journal of Cleaner Production, 1996 Volume 4, Number 3-4, p.159.

ISO 1997
International Organisation for Standardization, *Standard on Environmental Management - Life Cycle Assessment*, DIS 14040, 14041, 14043.

Kennedy et al. 1996
Data Quality, stochastic environmental life cycle assessment modelling, Kennedy et al. International Journal of Life Cycle Assessment 1 (4), 1996, p. 199.

Kennedy et al. 1997
Kennedy, D.J., D.C. Montgomery, D.A. Rollier and J.B. Keats, *Data quality, Assessing input data uncertainty in Life Cycle Assessment Inventory Models*. In International Journal of Life Cycle Assessment 2 (4) pp. 229-239, 1997

Kidder and Judd 1986
Kidder, L.H. and C.M. Judd, *Research Methods in Social Relations* ISBN 0-03-910714-0. CBS College publishing. New York, 1986.

Klöppfer 1997
Klöppfer, W., *Peer (expert) review in LCA according to SETAC and ISO 14040, theory and practice.* In International Journal of Life Cycle Assessment 2 (4) pp. 183-184, 1997.

Lindeijer et al. 1997
Lindeijer, E. , N.W. van den Berg and G. Huppes, *Procedure for Data Quality Assessment*, Report for Rioned (in Dutch) September 1997.

Meier 1997
Meier, M. A., *Eco-efficiency evaluation of waste gas purification systems in the chemical industry.* LCA documents vol 2. Ecomed, Landsberg, Germany 1997. ISBN 3-928379-54-2.

Morgan and Henrion 1990
Morgan, M. G. and M. Henrion, *Uncertainty. A guide to dealing with uncertainty in quantitative risk and policy analysis*. Cambridge University Press, 1990. pp. 332.

Weidema 1996
Weidema, B. and M. Suhr Wesnus, *Data quality management for life cycle inventories, an example of using data quality indicators*. Journal of .Cleaner Production, 1996 , Vol. 4, no. 3-4, p. 167-174.

Spold 1997 : *Spold data format* (1997-10-08) published on the website of IPU-Denmark .

Udo de Haes and Wrisberg, 1997
Udo de Haes, H.A. and M.N. Wrisberg (eds.), *Life cycle Assessment: State-of-the-Art and Research Priorities; Results of LCANET, a Concerted Action in the Environment and Climate Programme (DGXII)*. LCA documents, vol 1. Landsberg, 1997.

van der Ven and van Dam, 1996
DALCA, DAta for LCA, TNO report BU3.96/002461-1/AD, 31 July 1996. *Data and LCA, generation and presentation of data in view of exchange of information along the production chain.* Van der Ven , Van Dam, Proceedings of SETAC Case study symposium, December 2, 1997, Brussels.

van der Ven et al. 1999
van der Ven, B.L., Wrisberg, M.N., E. Lindeijer, G. Huppes and N.W. van den Berg, *Survey of quality assessment method in LCA.* In preparation, to be submitted to the International Journal of LCA.

Wrisberg 1997
Wrisberg, M.N., *A semi-quantitative approach for assessing data quality in LCA* Proceedings 7th Annual Meeting of SETAC-Europe, Amsterdam , April 6-10, 1997.

Wrisberg et al. 1999
Wrisberg, M.N., B.L. van der Ven, E. Lindeijer, G. Huppes and N.W. van den Berg, *Operational quality assessment in LCA.* In preparation, to be submitted to the International Journal of LCA.

Part 2: Survey of literature on quality assessment

1. Introduction

In the article "Quality assessment in LCA" (van den Berg et al. 1999), a framework for quality assessment has been given. The subject is known as a major area for improvement. Consequently other contributions to this subject exist. Therefore it is worthwhile to have an overview of the mainstream of concepts and approaches.
This article gives an overview of a selection of publications in the area of quality assessment in LCA. The overview is not exhaustive, it is primarily meant to determine the position of the framework of van den Berg et al.
First a general overview is given, then a more detailed description per literature source.

2. Overview

The subject of data quality has a clear starting point in the SETAC workshop held at Wintergreen in October 1992 (Fava 1992). One of the main results of this workshop was the conceptual distinction between quality indicators and quality goals: the goal and scope of a study defines the required quality of the data, whilst the individual data quality indicators determine the "fitness for use".
This distinction is important and is referred to in various publications and studies. However, it is not a sufficient basis for developing a quality assessment system. The main aim of such a system is to obtain insight into the confidence with which the results of a study can be used to help make the right decisions. The quality of the end result is a combination of the quality of input data and system parameters, i.e. the quality of the models used . It appears that this multi-layer quality aspect hampers the development of a straightforward quantitative approach. Data quality assessment methods based on statistical methods or, better, probability distributions, cover only part of the problem, viz. the quality of process input data as specified in the inventory.

This explains why there are basically two different approaches to the operationalisation of data quality assessment, viz. a ***qualitative indicator*** method and a ***probability distribution function*** method. Although both methods are based on indicators, the qualitative indicator method seems to be able to deal with indicators at different system levels, whilst the probability distribution method employs indicators with an explicit functional relationship.

A qualitative indicator method consists of defining the attributes of the data in question (e.g., at a product or substance flow level), with these attributes are subsequently being assigned a score, qualitatively or quantitatively. These scores can be used to assess the data at the substance flow, process or product system level, using suitable algorithms. Typical examples of this indicator approach are Weidema et al., 1996 and Clift et al., 1997. The indicators used by these authors are :

- reliability, a measure of the reliability of the data sources, acquisition methods and verification,
- completeness, a measure of the "representativeness" of the sample,
- temporal, geographical and technological correlation, a measure of the degree of correspondence between the data and the goal and scope of the study, as other aspects of representativeness.

Although the scores suggest a quantitative ranking, no aggregation of the scores is allowed. The pedigree matrix serves as an identification set. In order to close the gap between the qualitative indicator score and a quantitative final score, the authors propose the use of estimated distribution functions of the various datasets. The same indicator method has been used in a simplified form by van der Ven and van Dam (1997).

A step forward has been given by Wrisberg et al., (1997). This paper examined the presence of different data levels in the inventory: data on flow, process and system. This means that different indicators may be introduced for the various levels. The question still remains what the aggregated scores mean. The most detailed "indicator method" is presented in Lindeijer et al. (1997). His assessment consists of 5 steps, covering the quality of the inventory input data:

- Data quality parameters are established describing relevant data attributes (source, time, etc.).
- Data quality indicators are established for each individual process (reliability, representativeness and completeness).
- Indicator scores are aggregated for the (sub)system.
- The result is compared with quality goals.
- If necessary, the cycle is repeated to improve the data.

Each indicator is scored from 1 to 5. The aggregated data quality for a given system can be calculated by summing the individual indicator scores with the aid of a weighting factor. This factor is determined from the ratio of the normalised impact score of the individual process to the total score for that impact category, resulting in a quality score per impact category. Indirect contributions to emissions through other processes, e.g. incinerator emissions from industrial waste, are not considered in this ratio. Assuming some set of weights between impact categories, a single quality score on inventory input data results.

The probabilistic approach has been presented by several authors. Kennedy et al. (1997) describe the use of "beta probability distributions" to characterise the uncertainty of data. Beta probability distributions are described by 4 parameters, the maximum and the minimum variable (a and b) and 2 shape parameters α and β. The calculation procedure consists of three steps. First, for each data element a quality indicator is defined, second these indicator values are converted into a particular probability distribution, and finally the resulting probability function is determined. Heijungs (1996) considers the question of which elements determine the confidence in the result of an LCA within the context of the screening process. Heijungs presents an algorithm to describe the propagation of uncertainties in a LCI. The aggregated intervention, y, is a function of the processes used and the interventions of the processes. These can be presented as matrices. The individual uncertainties (or errors) of the elements of these matrices yield a final uncertainty value for the aggregated function, y, as well as for the individual contributions to the final result. This permits prioritisation of those processes for which data quality improvement may contribute most to overall data quality improvement. This approach is definitely on the level of flows and processes. Meier (1997) presents an extensive theory, as well as an operationalisation of the probabilistic approach. He uses normal and lognormal distributions and scenarios for discrete options in order to characterise uncertainty. The author argues that this characterisation method is better in line with the information available in other fields than the alternative approach, viz. the beta probability distribution method described by Kennedy et al. (1997). The data quality indicators used are similar to those of Weidema. For all types of data elements (and corresponding uncertainties) Meier has estimated the uncertainty range (equal to 2*coefficient of variation). Variables contributing to confidence/uncertainty are related both to the model and to the input data.

Qualitative uncertainties (e.g. model assumptions) cannot be characterised by probability distribution functions. Their importance can only be discussed, as their degree of imprecision is not predictable. The uncertainty analysis, as given by the mentioned authors, deals only with quantifiable elements. The estimates for the various coefficients of variance are based on literature. Using these estimates, the total effect of the uncertainties can be calculated by a Monte Carlo simulation. The final uncertainty can be expressed as a range for the eco-indicator (or for the individual theme scores).

The approach described in Meier (1997) is the most quantitative and detailed analysis of data quality assessment and the ensuing effect on the final result published so far. The method as such appears worthwhile, but the definition of the variables and the estimation of the uncertainty range are debatable. For example, literature data do not necessarily have a larger uncertainty range than measured data, as these have ultimately also been measured or estimated.

The two approaches, "qualitative indicator" and "probabilistic", are not mutually exclusive but may be considered complementary. Both approaches use "indicators" in order to operationalise quality characteristics. The probabilistic calculations can be considered as a method to evaluate a specific group of parameters (or indicators), which allows for a quantitative calculation. In the case of missing data or data from unknown (or poorly described) sources a qualitative indicator method is the best way to proceed. In the case of well-defined systems with various datasets, the probabilistic approach can be taken. Coulon (1997) emphasises the fact that these indicator methods do not end by estimating the quality of the final result of an LCA. In a certain way the indicator method could serve as a first step in a probabilistic approach.

These considerations vis-à-vis data quality focus mainly on the "uncertainty" of the results. It may be necessary to make a distinction between the "quality" of databases and the uncertainty of a particular LCA result.

The table below gives an overview of the various quality indicators, described in the literature. As can be seen relatively few authors have given attention to model validity and model reliability, while most do not investigate the role of selection, classification, characterisation, normalisation and weighting in the impact assessment.

Table 2 Comparison of the various indicators used in literature

	model validity	Model reliability	data validity	Data reliability
van den Berg current proposal	qualitative LCA model, choices	Reproducibility	Representativeness	Uncertainty completeness variability
Lindeijer/ Wrisberg	representativeness on system level	Completeness on system level	Completeness on process level	Reliability on process level
Weidema	-	-	Representativeness (time,geography,techn.)	Reliability Completeness
Meier	-	-	?	Coefficients of variance
ISO	Qualitative	?	Representativeness	Precision Completeness Consistency(reproducibility)

3. Literature Survey

This chapter gives a brief overview of data quality assessment models, published to date. The description is not exhaustive, but serves to illustrate the mainstream of the ongoing process and link it to this article.

3.1. SETAC 1, Wintergreen, (Fava , 1992)

Introduction

This workshop report can be considered one of the first efforts to describe the quality issue within LCA and to indicate options for solutions.

Content

The concepts of data quality goals (DQG) and data quality indicators (DQI) are distinguished. First define DQG , including :

- identification of decision types
- identification of data uses and needs
- design of data collection programmes

DQI is considered as a fixed set of labels as part of the data format. Evaluation of the DCG and the available DQIs leads to a quality assessment. Table 3 gives a set of DQIs and this description. This framework has been applied to energy, raw materials, emissions and ecological and human health.

Table 3 Quality Indicators

Quantitative	
Accuracy	Conformity of an indicated value to an accepted standard value. For many LCI data an accepted standard is not available, in that case the applicability of this indicator is limited.
Bias	Systematic or non-random deviation that makes data values different from the real value.
Completeness	The percentage of data made available for analysis compared to the potential amount of data in existence.
Data distribution	The theoretical pattern which provides the best estimate of a real variation of the data set.
Homogeneity*	Statistical outliers or large variance may be an indication that more than one pattern is represented by the data.
Precision	Measure of spread or variability of the data values around the mean of the data set.
Uncertainty*	Levels of uncertainty can be calculated from statistical tests on the data set.
Qualitative	
Accessibility **	The actual manner in which the data are stored or recorded.
Applicability/ Suitability/ Compatibility	Relevance of the data set within a study to the purpose of that study.
Comparability**	The degree to which the boundary conditions, data categories, assumptions and data sampling are documented to allow comparison of the results.
Consistency**	The degree of uniformity of the application of methodology in the various components of the study.
Derived models**	The differences between models generating potentially similar data.
Anomalies	Extreme data values within a data set.
Peer review**	
Representativeness	The degree to which the data set reflects the true population of interest.
Reproducibility**	The extent to which the available information about methodology and data values allows a researcher to independently carry out the study and reproduce the results.
Stability	Measure for consistency and reproducibility of data over time.
Transparency	The degree to which aggregated data can be traced back to the original values.

* The description given is not a definition.
** These are system indicators rather than data indicators.

Remarks

The concept of a separate set of DQGs and DQIs is good and certainly serves as a helpful tool for any quality model. Unfortunately, there are unresolved problems and disadvantages:

- The number of indicators is too large for a workable approach. Besides, the indicators do overlap.
- No coherent framework is presented for assessing the quality of the different levels of a LCA. How does assessment at the substance level relate to assessment at the process and system level ?

3.2. SETAC 2 (Clift et al. 1997 ; Weidema et al. 1996)

Introduction

Over the past few years a SETAC working group on Data Inventory has been active in the field of data quality. Bo Weidema participated in this working group. It is therefore understandable that the final document of the working group and the publication of Weidema and Wesnaes bear a close resemblance.

Content

First data quality goals (DQG) must be defined. The collected data must then be related to the DQG's. This can be done using a set of data quality indicators (DQI) to characterise the individual data. In addition, these DQIs can be used to calculate the uncertainty of the overall result.
A "pedigree matrix" is proposed to describe the DQIs. See Table 4.

These indicators are:

- reliability, a measure for the data sources, acquisition methods and verification
- completeness, a measure for the "representativeness" of the sample.
- temporal, geographical and technological correlations, measures for the degree of correspondence between the data and the goal and scope of the study.

Although the scores suggest a quantitative ranking, no aggregation of the scores is allowed. The pedigree matrix serves as an identification set.
The next step is to estimate the overall uncertainty. Uncertainty consists of two elements :

- basic uncertainty (measurements errors, fluctuations of the data, etc).
- additional uncertainty (sub-optimal quality of data, reflected in a pedigree different from 1.1.1.1.1).

For both types of uncertainty a C.V. (coefficient of variance) can be estimated, based on expert judgement (N.B. this approach has also been used by Meier) . Default values of relevant C.V.s could be made available for various data sets in the future. With these estimates, the aggregated C.V. can be calculated.

Table 4 **Pedigree matrix of DQI**

DQI	score				
	1	2	3	4	5
Reliability	verified data based on measurements	verified data based partly on assumptions or non-verified data based on measurements	unverified data based partly on assumptions	qualified estimate	unqualified estimate
Completeness	representative data from an adequate sample of sites over an adequate period	representative data from a smaller number of sites over an adequate period	representative data from an adequate number of sites but over a shorter period	representative data from a small number of sites over a shorter period or inadequate data from adequate number of sites	unknown or incomplete data from a small number of sites
Temporal	< 3 years difference	<6 years difference	<10 years difference	<15 years difference	unknown or > 15 years
Geographical	data from an adequate area	average data from a larger area	data from an area with a similar production structure	data from an area with a slightly similar production structure	unknown or different area
Technological	data from processes under study and company-specific	data from processes under study for different companies	data from processes under study with different technologies	data from related processes and materials, same technology	data from related processes and materials, different technology

Remarks

The given approach emphasises the need to distinguish between data indicators and data quality goals.

Data are not intrinsically good or bad, but more or less suited to the goal and scope of the study. However, using numbers for the identification of an indicator in combination with the description of the various indicator values, it is suggested very strongly that there is quality difference between the indicator values. Measured data are likely to be better (i.e. more precise) than calculated values or values from literature. This is not true, at least not true in a general way.

The concept of the estimation of an overall "uncertainty" factor, based on estimated C.V's of the individual data looks meaningful. It certainly produces a "number". It is clear that the knowledge of the C.V's of the individual data is absolutely insufficient. Besides, it should be remembered that the statistical approach of the data is only one element of the total "reliability" of the result of a LCA.

3.3. SPOLD (SPOLD, 1997)

Introduction

The present format is an electronic version of the previous paper format. The format (including data quality indicators) is intended for LCI data exchange.

Content

The basic ideas of data quality assessment are similar to the concept of Weidema. The Spold format contains a large number of data fields to be completed. Data fields in various sections relate to data quality. Text fields for *Time period*, *Geography* and *Technology* are connected to the main process. A data field for *Representativeness* also relates to the entire data set. The *R*-percentage reflects the relation between the

actual data set and the potential data population. The text file reports the number of sites for which data were collected, the sampling method and possible bias.
The main structure of the description of "flows" is :

- reference code
- name of flow
- unit
- mean
- uncertainty type (distribution type of the data)
- coefficient of variance (σ/μ)
- geographical location (the location of the delivering and the receiving processes allows for more precise identification of related up- and downstream processes).

For each intervention, so-called "adjustments " can be made: fields reporting any deviations of the intervention from the general entries in terms of representativeness, time, geography and technology.

Remarks
The SPOLD format does not contain any assessment procedure for data quality. The format should be considered as a - nearly complete - description of a set of data, included data quality indicators.
The indicators are the same as those described in SETAC 2.

3.4. ISO (ISO, 1997)

Introduction
ISO is in the process of defining a standard for an LCA methodology. Elements of the subject of data quality assessment are described in various parts of the documents.

Content
Document 14041 sets out guidelines on how to formulate goal and scope, define and model the systems, collect and verify the data, evaluate the reliability of the inventory, and interpret and report the results. In the goal and scope phase data quality requirements should be formulated. The following parameters are given :

- time
- geography
- technology

Data quality indicators to be covered in each study are:

- precision: measure of variability of data values for each data category expressed as e.g. variance.
- completeness: percentage of locations reporting primary data from the potential number in existence for each data category in a unit process.
- representativeness: qualitative assessment of degree to which the data set reflects the true population of interest (time, geography and technology coverage).
- consistency: qualitative assessment of how uniformly the study methodology is applied to the various components of the analysis.
- reproducibility: qualitative assessment of the extent to which information about the methodology and data values allows an independent practitioner to reproduce the results.

During the inventory data should be validated. Instruments for validation are balance checks (mass and energy) and comparative analysis of emissions

Document 14043 gives guidelines for interpreting results. Elements in the evaluation process are:

- completeness: check process of verifying that information from the different phases is sufficient for interpretation.
- consistency: check process of verifying that interpretation is in accordance with the goal and scope definition.
- sensitivity analysis: systematic procedure for estimating the effects on the outcome of a study of the chosen methods and data.
- uncertainty analysis: systematic procedure to ascertain and quantify the uncertainty introduced in the results of an LCI due to the cumulative effects of input uncertainty and data variability. It uses either ranges or probability distributions to determine the uncertainty of the results.

Remarks
Although the guidelines of the present draft documents mention the need to take these steps, they do not prescribe detailed procedures. Neither the concept and theory of quality assessment nor the various levels of uncertainty are discussed in these documents.

3.5. DALCA (van der Ven and van Dam 1996)

Introduction
The DALCA project was commissioned by VNO-NCW (the Dutch employers association). The research question was :

> *Is it feasible for companies to generate generally accepted, reliable environmental data from processes which can then serve as input for environmental analyses, such as life cycle analyses (LCAs), and also as a basis for the exchange of environmentally specific product information between companies?*

Inevitably, data quality was one of the topics considered.

Content
This project developed a pragmatic approach. An essential distinction is made between data quality goals (DQG) and data quality indicators (DQI). DQGs are project dependent, whilst DQIs are an intrinsic function of the data.
A minimal set of DQIs should cover the **representativeness** of the data , the **source** (how are the data generated ?) and the **completeness** (which part of the total is being covered).
N.B. The question how to deal with the various levels of aggregation within LCA (e.g. substance flows, process blocks, systems) is not answered in this approach. The described method refers to the level of substance flows and processes.
Representativeness is broken down into a temporal, geographical and technological indicator. These indicators can be described in the heading of the process description.
Figure 4 shows the structure.

Figure 4 Structure of DQA method

The indicators source and completeness are expressed qualitatively in 5 categories. This typology does **not** refer to a difference in the quality of the score. Table 5 gives the score system.

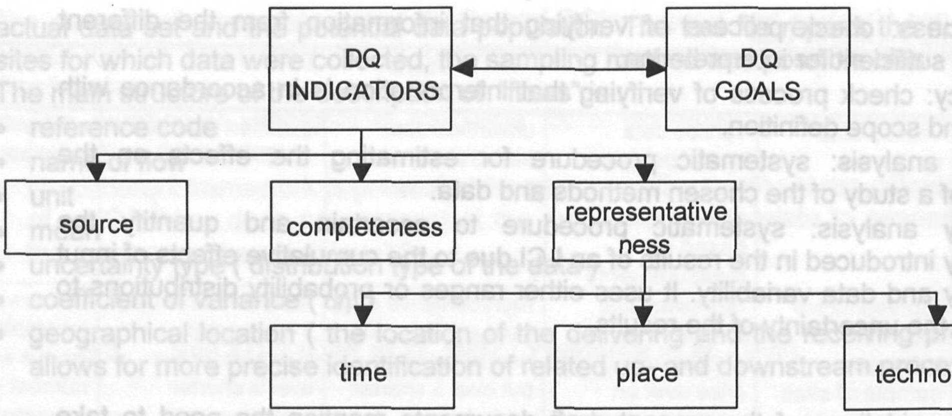


Table 5 **Score system for indicators *Source* and *Completeness***

	Class				
	A	B	C	D	E
Source	measurements, standardised	measurements, unspecified method	calculated data based on measurements	calculated data based on literature	estimated data , experts' guess.
Completeness	sufficient number of measurements, average with confidence limits	limited amount	single measurement		

Remarks

This approach is based on SETAC 2. The difference is that the indicator scores in DALCA are less detailed. Besides, a qualitative approach is described to estimate the "reliability" of a process, based on the indicator scores of the substance flows and the verification results.

3.6. Meier (Meier, 1997)

Introduction

In a special publication of the Journal of LCA , so-called LCA documents, Markus Meier published a very comprehensive study on the eco-efficiency of various waste gas purification technologies. This technology comparison was based on LCA methodology. In addition, a quantitative uncertainty analysis was carried out and described.

Content

The uncertainty analysis has been demonstrated with the Eco-indicator 95 method. Qualitative uncertainties (e.g. regarding model assumptions) may not be characterised by probability distribution functions. Their importance can only be discussed and the degree of imprecision is not predictable. This uncertainty analysis deals only with quantifiable elements. Normal, log normal distributions and scenarios for discrete options are used to characterise uncertainty.

These elements from the classic statistical approach can be used, but the aim is different. In the classic approach a probability distribution defines the probability of an event. The *subjective* method of this study considers the fact that there is no true model for the system. Distributions are estimated by subjective probability methods rather than by single-point estimates.

The data quality indicators used are similar to those of Weidema. For all types of data elements (and corresponding uncertainties) Meier has made an estimation of the uncertainty range (2 * coefficient of variation). Variables contributing to uncertainty relate to both the model and data:

model
- general assumptions (e.g. on functionality of system)
- omittance of processes in the inventory
- definition of functional unit
- allocation
- valuation model and assumptions
- consideration of impact categories

data
- site-specific process data
- temporal variation
- data modules of background data
- measured emissions of background processes
- average background data; space and time aspects
- omittance of emissions

Table 6 **Example data variables and the uncertainty range**

	variable	uncertainty range 2* c.v.
d1	site-specific process data	
	verified and measured	± 0.05
	verified assumptions / unverified measurements	0.10
	unverified assumptions	0.20
	qualified estimates	0.30
	estimates	0.50

The estimates for the various c.v.'s are based on literature. With these estimates the overall effect of the uncertainties can be calculated by a Monte Carlo simulation. The final uncertainty can be expressed as a range for the eco-indicator (or for the individual theme scores).

Remarks
The approach described in this study is the most quantitative and detailed analysis of data quality assessment and the impact of data quality on the final result published to date. As such, the method looks promising. The definitions of variables and the estimated uncertainty range are debatable. For example literature data do not necessarily have a larger uncertainty range than measured data.

3.7. Heijungs (Heijungs 1996)

Introduction
Heijungs' article is based on a discussion paper for the SETAC working group on Screening and Streamlining. Comments from the working group and others have been included in this final article.

Content
The ultimate question concerns the reliability of the result of an LCA. Many elements at different levels contribute to the final result. What are the key issues ? This question is discussed in this paper within the context of the screening process. Key issues are formed by (the issues which represent highly) sensitive parameters with respect to the end result. Heijungs presents an algorithm to describe the propagation of uncertainties in an LCI. The aggregated intervention y is a function of the processes used and the interventions of the processes These can be presented as matrices. The individual uncertainties (or errors) in each of the elements of the matrices gives a final uncertainty value for the aggregated function y as well as the individual contributions to the final result. This allows for prioritisation processes requiring further research to decrease uncertainty.

Remarks
This paper gives a clear and formal method for dealing with uncertainties in LCI. Although the aim of the paper is to develop a screening instrument, the proposed method can be used for an overall quantitative uncertainty analysis of the LCI. The paper is restricted to the domain of process and interventions. Uncertainty due to system definitions is not taken into account.

3.8. Wrisberg (Wrisberg, 1997)

Introduction
A practical semi-quantitative approach has been developed for Philips Electronics with the aim of providing an overall indication of the quality of data used in an LCA and permitting identification of data contributing significantly to poor data quality.

Content
In this framework a distinction is made between 4 elements :

- DQG, data quality goal, dependent on goal and scope.
- DQP, data quality parameters, comprising information on the data and included in the data format.
- DQI, data quality indicators, scores based on the parameters.
- DQA, the assessment procedure.

A data quality parameter is given a score (1-5) according to a subjective but transparent break-down into information quality categories. The data quality parameters are aggregated stepwise to the system level, resulting in data quality indicators relating to **reliability**, **completeness** and **representativeness**. Aggregation is performed by summing up the scores of the data quality parameters relating to a specific indicator and dividing this figure by the number of data quality parameters (giving equal weights to each parameter). By aggregating the reliability parameters, a distinction is made between environmental flows and economic flows.

Table 7 Indicators and parameters

indicator	parameters
reliability (per flow)	uncertainty statistical representativeness age collection method
completeness of flows (process level)	included /excluded flows aggregated data mass balance information
completeness of processes (system level)	included/excluded processes allocation rules verification of cut-off/allocation rules
representativeness	geographical coverage temporal coverage technological coverage

The end result is a set of indicators, per flow, per process and per system.

Remarks
The proposed method can be considered a further expansion of the indicator concept. Based on subjective parameters, a quantitative assessment procedure has been developed. The question remains: what is the meaning of the aggregated scores. Or, may average scores be used?

3.9. Kennedy (Kennedy et al 1996,1997)

Content
With a view to transforming the deterministic data models into stochastic models, this paper describes the use of "beta probability distributions" to characterise data uncertainty. Beta probability distributions are described by 4 parameters, the maximum and the minimum variable (a and b) and 2 shape parameters alfa and beta. The calculation procedure consists of three steps. First, for each data element a quality indicator is defined; second these indicator values are translated into a particular probability distribution; and finally the resulting probability function is determined.

Remarks
The proposed method is practicable and correct. Nevertheless, the question arises whether the method is adequate. Determination of the 4 parameters introduces many subjective assumptions. In addition, the distribution itself introduces new uncertainties because of its prescribed shape.

3.10. Rioned (Lindeijer et al. 1997)

Introduction
The Dutch sewerage branch organisation intends to develop a database on its operating area. Part of the project has been the development of guidelines for data generation, verification and assessment. This project has been carried out by IVAM and CML.

Content
The main assessment procedure consists of 5 steps :

- Data quality parameters are established, describing relevant data properties (source, time, etc.).
- Data quality indicators are established per process (reliability, representativeness and completeness).
- Indicator scores are aggregated for the (sub)system.
- The result is compared with quality goals.
- If necessary, the cycle is repeated to improve the data and thus the results.

Indicators are :

reliability (process level)
1 uncertainty of data
2 representativeness of data : have variations in substance flows been considered?
3 verification of data.
completeness (process level)
4 substance flows : the extent to which all flows are included.
5 aggregated substances: extent to which substances are aggregated into groups instead of being taken as individual substances (e.g. N_2O - - NO_x)
6 mass balance
7 allocation
completeness (system level)
8 processes : extent to which all processes are included
9 allocation
representativeness (system level)
10 time
11 technology

Each indicator is scored from 1-5. The aggregated data quality of a system can be calculated by the summing the individual indicator scores by means of weighing factors. These weighing factors are calculated as the ratio of the impact score of the individual process to the total score for that impact category.

Remarks
The proposed assessment method is clearly the ultimate "indicator method", with indicator scores being aggregated in a quantitative manner.

4. References

Van den Berg et al., 1999
Van den Berg. N.W. et al., *Quality Assessment in LCA*, Int. Journal of LCA, to be published.

Clift et al., 1997
Clift, R. et al., *Towards a coherent approach to life cycle inventory analysis*. Final document of SETAC-Europe Working Group on Data Inventory, April 1997.

Coulon 1997
Coulon, R., V. Cambobreco., H. Teulon and J. Besnainou. *Data quality and uncertainty in LCI.* Int.J.of LCA, Vol.2, no. 3, p. 178.

Fava 1992
Fava, James, et al., *Life cycle assessment data quality: a conceptual framework.* SETAC workshop 1992, Wintergreen, USA.

Heijungs 1996
Heijungs R., *Identification of key issues for further investigation in improving the reliability of life-cycle assessments.* Journal of Cleaner Production, 1996 Volume 4, Number 3-4, p.159.

ISO 1997
International Organisation for Standardization, *Standard on Environmental Management - Life Cycle Assessment*, DIS 14040, 14041, 14043.

Kennedy et al. 1996
Data Quality, stochastic environmental life cycle assessment modelling, Kennedy et al. International Journal of Life Cycle Assessment 1 (4), 1996, p. 199.

Kennedy et al. 1997
Kennedy, D.J., D.C. Montgomery, D.A. Rollier and J.B. Keats, *Data quality, Assessing input data uncertainty in Life Cycle Assessment Inventory Models*. In International Journal of Life Cycle Assessment 2 (4) pp. 229-239, 1997

Lindeijer et al. 1997
Lindeijer, E. , N.W. van den Berg and G. Huppes, *Procedure for Data Quality Assessment*, Report for Rioned (in Dutch) September 1997.

Meier 1997
Meier, M. A., *Eco-efficiency evaluation of waste gas purification systems in the chemical industry*. LCA documents vol 2. Ecomed, Landsberg, Germany 1997. ISBN 3-928379-54-2.

Spold 1997 : *Spold data format* (1997-10-08) published on the website of IPU-Denmark .

van der Ven and van Dam, 1996
DALCA, DAta for LCA, TNO report BU3.96/002461-1/AD, 31 July 1996.
Data and LCA, generation and presentation of data in view of exchange of information along the production chain. Van der Ven , Van Dam, Proceedings of SETAC Case study symposium, December 2, 1997, Brussels.

Weidema 1996
Weidema, B. and M. Suhr Wesnus, *Data quality management for life cycle inventories, an example of using data quality indicators*. Journal of .Cleaner Production, 1996 , Vol. 4, no. 3-4, p. 167-174.

Wrisberg 1997
Wrisberg, M.N., *A semi-quantitative approach for assessing data quality in LCA* . Proceedings 7th Annual Meeting of SETAC-Europe, Amsterdam , April 6-10, 1997.

Part 3: First Steps towards Operationalisation

Abstract

A framework for quality assessment in LCA based on the Spread-Assessment-Pedigree approach of Funtowicz and Ravetz (1990) has been described in an earlier paper (van den Berg *et al.*, 1999). This framework, consisting of four quality elements related to reliability and validity of the LCA model and data, is further operationalised in the present paper. The operationalisation involves a number choices. One choice concerns the intended application, which has implications for the choice of quality factors, and another relates to the level of application, e.g. whether the operationalisation is directed at the flow, process or system level. A further important choice concerns the extent to which the assessment model should aggregate the quality factors, and how. The processing of quality-related information can in principle take place via two different strategies, which we refer to as the qualitative indicator method and the probability distribution method.

This paper provides an example of the operationalisation of the framework. It is not comprehensive, but is clearly more complete than previously described quality assessment approaches. In this example, the goal of the quality assessment is to assess the overall quality of an LCA result. The proposed model uses 15 different quality factors related to unit processes or whole systems. The quality factors are aggregated over the scales according to the qualitative indicator method. This choice is made since most quality factors are inherently qualitative and the conditions for statistical operations can therefore not be met.

1. Introduction

Quality assessment in LCA is rarely applied, although it is recognised as important. The importance of quality assessment and the specific requirements differ according to the type of application, e.g. LCIs and (screening or complete) LCAs for product/service improvements, for comparisons, or for information exchange. For instance, building and maintaining a database involve specific quality aspects that will need to be assessed (partially) with criteria other than those used for improvements and comparisons. Furthermore, it must be kept in mind that quality assessments may be applied to different phases in an iterative LCA. At one stage the assessment may serve to state the initial quality and the priorities for quality improvement in an iterative process, whereas in the end it may serve to check whether the final quality is sufficient for the goal of the LCA.

This paper aims to show the options for and the consequences of applying LCA quality assessment methods, following up on a theoretical paper on LCA quality assessment (van den Berg *et al.*, 1999). In this paper, validity and reliability are identified as the main elements of quality to be assessed. These elements can be seen to consist of different quality indicators. Each indicator can be monitored with one or more quality factors, expressing the extent to which that quality aspect is optimised. These may relate to the input data (expressing data quality) and to the LCA model (expressing model quality). Finally, this information contributes to the confidence in the assessment of whether the overall quality is sufficient for the goal of the study/project. Project-specific aggregation of quality aspects may be performed to achieve such an overview. The considered model and data aspects, their operationalisation and importance for the overall quality, and the need for aggregation determine what specific approach is used, which may differ from application to application. For effective communication, it is however important to achieve a common model to start with. Building such a model is the aim of chapter 2, which provides a framework for the operationalisation of quality assessment methods, the choice of which will depend on the application/goal of the study. Chapter 3 reports a tentative operationalisation of a semi-quantitative quality assessment method. Chapter 4 discusses the application of the Spread-Assessment-Pedigree (SAP) approach introduced by Funtowicz and Ravetz (1990) in LCA, and touches upon procedural aspects.

2. Strategies for the operationalisation of quality assessment

2.1. A framework for assessing reliability and validity in LCA

Van den Berg *et al.* (1999) provide a general framework for LCA quality assessment aiming at determining the level of confidence in the outcome. The framework consists of three quality aspects: Spread (S), which describes the reliability of the data and model; Assessment (A), which refers to the validity of the data and model choices; and Pedigree (P), which refers to more procedural quality elements such as data verification, sensitivity analysis, dominance analysis and peer review. All three quality aspects are required to determine the level of confidence in an LCA outcome.

This paper addresses the operationalisation of the S and A aspects, and less attention is paid to the P aspect. The four quality indicators – the reliability of the model, the reliability of the input data, the validity of the model and the validity of input data – may be operationalised according to the quality factors shown in table 1.

Table 1: The general LCA quality framework considers four quality indicators in each of the four phases of the LCA: examples of quality factors (see van den Berg *et al.*, 1999).

Phases of LCA		Validity	Reliability
Scope	Model	Validity of the functional unit (FU)	
	Input data	Representativeness in relation to FU	Uncertainty, completeness, variability
Inventory	Model	Validity of allocation method and system boundary	Reproducibility of model results
	Input data	Representativeness of process data	Uncertainty, completeness, variability
Characterisation	Model	Validity of characterisation method	Reproducibility of model results
	Input data	Representativeness in relation to scope	Uncertainty, completeness, variability
Evaluation	Model	Validity of evaluation method	Reproducibility of model results
	Input data	Representativeness in relation to scope	Uncertainty, completeness, variability

Figure 1 illustrates a general framework for assessing reliability and validity (the Spread and Assessment aspects), which may be operationalised for each of the LCA phases. The elements of the framework are:

- *quality factors*,
- *quality indicators*,
- the *main quality aspects* of the LCA model or the LCA input data,
- providing the *overall quality* of the LCA result.

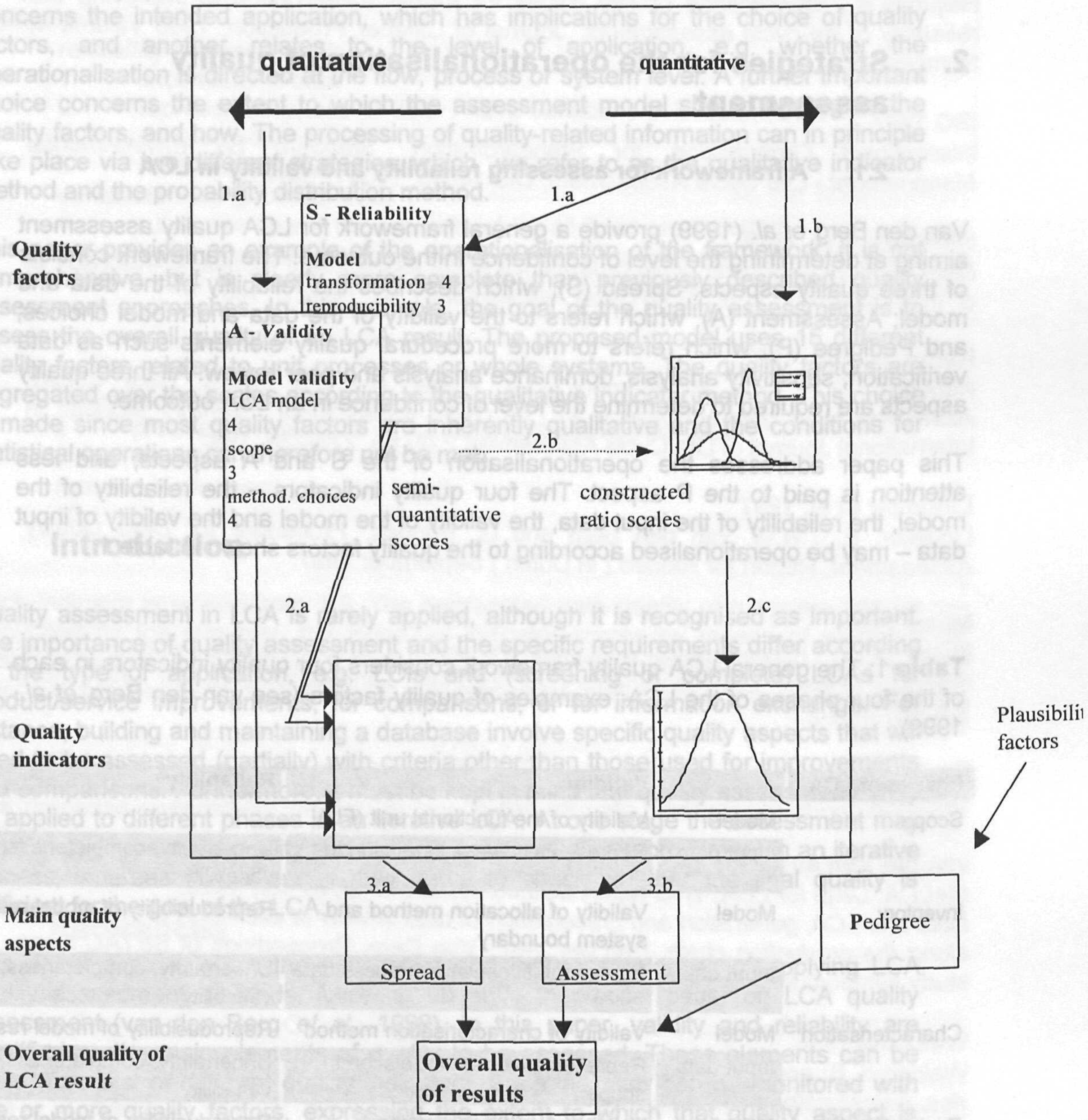


Figure 1: Framework for assessing the quality of the LCA model and LCA input data. The arrows on the left refer to the qualitative indicator method, and those on the right to the probability distribution function method.

It should be noted that this terminology is used to indicate specific quality elements and may in this sense differ from others (e.g. Weidema and Wesnæs, 1996). *Quality factors* are used as headings for specific quality-related information on the model and the data, e.g. the spread of an input flow, the variability of LC_{50} data, or the representativeness of an eco-toxicity model. Most quality factors are inherently qualitative in character, and only a few are quantitative, while relevant items may be specified at very different levels of sophistication. Quality factors related to data reliability are to some extent quantitative in nature, although they may not be available in such a form. Quantitative measurements of the flows of a unit process may be accompanied by, for example, a specified probability distribution, a range, or a qualitative indication of robustness. Quality factors related to model reliability and validity are inherently qualitative in nature, e.g. representativeness may be specified in terms of the characteristics of a specific time and place, or they may be more ad hoc, like "usual in a specific industry".

The model and data aspects are processed in two steps to obtain a measure of the *quality indicator*. The first step involves measurements of the contributions of individual factors to the quality (arrows 1). The second step concerns the aggregation of the different factors that contribute to a quality indicator (arrows 2). For instance, the validity of the boundaries of the system, which are determined by assessing whether the inclusion/exclusion and cut-off of data are in accordance with the scope of the study, together with the representativeness of the data, which is determined by assessing the extent to which the level of aggregation of the data or temporal and technological coverage, are in accordance with the scope, contribute to the quality indicator "validity of input data". The quality indicators are further processed to provide the overall quality of the LCA output (arrows 3).

The level of sophistication for quality assessment in LCA is linked to the level of ambition in the scope of the study. The assessment may be more or less complete in terms of the number and type of indicators, and may be more qualitative or more quantitative in their operationalisation and/or presentation. The basic reasons for applying an LCA quality assessment model are as follows:

- to determine the level of confidence in an LCA outcome, when making a comparative assertion. Very often the differences between the considered alternatives is less than their uncertainty;
- to compare the quality of an LCA outcome with quality standards. There may be reasons for requiring a certain minimum quality of data and results; and
- to identify areas for improving quality in order to increase the overall quality of a specific study, including the selection of data sets for an LCA study with a sufficiently high quality level.

Furthermore, part of the model may also be used to identify priorities for database maintenance and improvement. We assume that the latter (partial) applications of quality assessment can be covered by the framework presented here.

Ultimately, the overall quality of the result of the analysis is to be established, ideally in an unequivocal, derived way. The ways in which independent quality factors may be combined into such an overall assessment are not considered here. There is no formalised way of establishing the overall quality of a result. As there are several qualitative aspects, such a formalisation will be a difficult affair. One can establish some rules of reasonableness – for example, if the quality of one independent quality item increases, the overall quality score should not decrease – but even such a seemingly self-evident rule may be problematic. If in a comparative analysis the completeness of one alternative is improved over that of the other, one might argue that the comparison will become more haphazard exactly for that impact quality reason.

This brings us to the next step in quality assessment for real decisions. At the LCA results level, first there is the quality of outcomes on single alternatives as discussed above. For decision making, however, it is not the quality of the result for a single alternative which is relevant, but the quality of the set of alternatives, where the quality will usually differ between alternatives. This question of how to combine quality aspects in an overall comparative assessment is not addressed here. One fundamental problem is that a lower quality, especially if this is due to incompleteness, may lead to a favourable score in comparison with alternatives with better quality in this respect.

2.2. Different levels of application

An assessment of the quality of LCA results ultimately depends on the quality of the models and the data used. For an operational data quality assessment method, one has to fix the items which are to be measured[1] operationally. Here we first state an ideal situation, in which all basic quality aspects are measured independently. In chapter 3 we show what is now possible in LCA in practice.

There is a large number of quality aspects, related to all elements of an LCA, from inventory to, ultimately, evaluation. Within each of these elements, several independent quality aspects can be distinguished, not only in terms of the spread and validity of the model and data, but also in relation to the level of application of these concepts. In the inventory, for example, there is the validity of the data on environmental and on economic flows; there is the validity of the process model; and there is the validity of the inventory system as built from these process models. Some quality aspects are not independent, but are derived, just as the spread in system data is based on the spread in all environmental and economic flows. The number of these flows may easily exceed 10^4 so measuring the quality of individual flows may never be feasible for practical reasons. We make a distinction here between economic and environmental flows, as their quality works out differently on the results. For environmental flows, it is their direct contribution to the total results which indicates their importance to the overall quality. For economic flows there is no such easy relation, as such flows influence the results through environmental interventions in *other* processes. Table 2 provides an overview of the independent and derived quality aspects of LCA data and model.

1 Measurement does not imply a ratio scale; ordinal scales or classes may also be used. Of course there is then no formalised procedure to arrive at an aggregate quality score.

Table 2: Independent and derived quality aspects of the data and model in the various steps of the LCA model, ideal set-up.

Quantified elements of LCA		Ideal			
		Reliability		Validity	
		Data	Model	Data	Model
Scope definition	Reference flow	⊕	⊕	⊕	⊕
Inventory	Economic flow	⊕	⊕	⊕	-
	Environmental flow	⊕	⊕	⊕	-
	Process	Ʊ	⊕	⊕	⊕
	System	Ʊ	Ʊ	Ʊ	⊕
Classification	Env. flow selection	-	-	-	⊕
	Env. issue selection	-	-	-	⊕
Characterisation	Environmental flow	⊕	-	⊕	⊕
	Environmental problem	Ʊ	Ʊ	-	⊕
	Set of env. problems	Ʊ	Ʊ	-	⊕
Normalisation	Env. flow	⊕	-	⊕	-
	Total of env. flows	Ʊ	-	-	-
Evaluation	Env. problem	⊕	⊕	⊕	⊕
	Set of env. problems	Ʊ	⊕	Ʊ	Ʊ
Key: ⊕ = independently measured Ʊ = measurement derived from less aggregated quality measurements - = not applicable					

In principle, as an ideal, a measurement on all independent quality factors is to be established. In practice this is hardly possible. For example, the spread on the individual flows to and from a process is not available, and it is often difficult and always time consuming to establish. One may then make an independent quality assessment at the level of each individual process. The nature of such an assessment is more subjective but can be objectified by defining specific quality characteristics and indicating the scale at which these may be measured. For large numbers of processes, even this procedure may be too cumbersome. An objectified assessment is then possible at the level of the inventory system as whole. Some quality aspects, such as those related to the validity of the inventory model, can only be established at the level of the system as whole.

For classification, characterisation, normalisation and evaluation, similar types of reasoning may be used to establish which information is relevant for quality assessment independently, and which only in a derived way. Some of the steps have limited relevance for quality assessment. Classification involves a decision not to take some substances into account, or to take them into account only to a certain degree for some environmental issues, or to leave out some environmental issues altogether. Classification is either an empty step, or it reduces the validity of the characterisation model. Thus, only the model validity is relevant for this step, either at a substance level or at the level of environmental issues.

For the characterisation, three levels of application may be distinguished. At the flow level, there is the reliability and validity of both the model and the data. For example, a fate model as part of a characterisation model may not take into account the relevant breakdown characteristics of a substance, thus limiting the validity of that model, while the validity of the data may be hampered by the fact that the flows considered are not the most relevant for the environmental problem being modelled, e.g. sulphur dioxide emissions at sea are not relevant to the problem of acidification. The reliability of the data at the flow level is derived from the reliability of the inventory data. It is difficult to imagine a reliability aspect of the characterisation model at the flow level.

In the normalisation, the validity and reliability of the data are independent quality factors. The reliability may be limited by faulty sampling procedures, while the validity may be limited, as in generalising emissions data from one country to those in another. In the evaluation, the reliability of environmental issue scores may be limited by a spread in panel responses, if these are used, while the validity of the data may be limited by a too simple explanation of what was to be evaluated by the panel members, as is usually the case. If an economic evaluation is used, the valuation model may have limited validity because, for example, hedonic pricing leaves out a number of relevant aspects. Or an economic valuation model for health effects may have limited validity because relevant health items have not been incorporated, e.g. the effects of heavy metals on mental development have been omitted. Even if each item is included in a reasonable way, the set of environmental issues may not include all relevant value areas, or they may overlap; e.g. ozone layer depletion may be used as a valued effect parameter, as well as measured in terms of the health effects of increased UV radiation.

One quality aspect that is easily overlooked is the quantification of the function, in terms of the functional unit and its reference flow. When analysing the appeal of a consumer good such as ice cream, for example, the function used is psychological satisfaction. It looks good, tastes good and, especially, it makes you feel good. What is the functional unit, and in what terms is it measured? If one takes "(*X* times) one ice cream" as the reference flow, the relation with the function is very poor. If one takes "*X* kg of ice cream" the relation is slightly better, but neither of these reference flows catches the essential appeal of ice cream.

2.3. Different strategies for filling in the framework

Measured quality aspects at the flow or the process level need to be processed to the system level – similar to the processing steps shown in the framework in figure 1. There are various more or less sophisticated strategies that can be chosen for each of these different processing steps. The first processing step in figure 1 concerns the measurement of the quality factor. The quantitative factors may be represented by probability distributions, possibly a subjective probability distribution if there are no relevant observations of the quantity; this step is represented by arrow 1.b in figure 1.

However, probability distributions are inappropriate for qualitative factors (Morgan and Henrion, 1990). Their contribution to quality may be represented in a semi-quantitative way, such as by using a "pedigree table", as introduced by Weidema and Wesnæs (1996);[2] arrow 1.a in figure 1. Another way to address qualitative factors involves the use of parametric sensitivity analysis (Morgan and Henrion, 1990). The latter is in fact a more procedural aspect, which we refer to as the P aspect. It should be noted that the representation of both quantitative and qualitative factors may involve some degree of subjectivity.

The last two processing steps (arrows 2 and 3 in figure 1) concern the aggregation of all the factors that contribute to a specific quality indicator and the assessment of their contributions to the overall quality of the LCA outcome. Different techniques can be used for these means of processing, according to the relative importance of the different quality factors, the input data and the model choices for the overall outcome.

The simplest technique involves the simple aggregation of the semi-quantitative numbers (arrows 2.a and 3.a). In this case the relative importance of the different quality factors, their reliability versus the validity, and the data versus the model aspect, are determined in a subjective manner. The processing of the last two steps by simple aggregation is referred to as the *qualitative indicator* method (van der Ven *et al.*, 1999). With this method, all quality factors may be aggregated into one figure representing the Spread aspects, and into another representing the Assessment aspects.

The use of Monte Carlo simulations is one of the most sophisticated means of carrying out the two last processing steps (arrows 2.c/3.b). This approach allows a large number of quality factors and can treat various parameter distributions, but requires that all the quality factors are specified as probability distributions. This implies that the qualitative quality factors will have to be transformed from qualitative or semi-quantitative measures into probability distributions (dotted arrow 2.b in figure 1). Such a transformation will also require estimations of the relative importance of the different quality factors/indicators, input data and model choice. This approach, referred to as the *probability distribution function* method by van der Ven *et al.* (1999), may be used to aggregate all the transformed contributing quality factors into one figure in one processing step (2.c/3.b), and may not make a distinction between the Spread and Assessment aspects. Fuzzy set theory (Pohl and Ros, 1997; Bilitewski and Hauptmann, 1999) is an alternative way to transform semi-quantitative scores into constructed ratio scales (2.b), which may be added to obtain one single result (2.c/3.b). We also refer to this approach as the probability distribution function.

It should be noted that this framework allows for a different mix between simple aggregation and statistical calculations (see figure 1). For instance, the only inherently quantitative indicator, data reliability, may be presented using statistical distributions processed according to the probability distribution function method (1.b, 2.c, 3.b), while the qualitative factors may be processed according to the qualitative indicator method (1.a, 2.a, 3.a).

Some proposed LCA quality assessment approaches stop after the first processing step (e.g. Weidema and Wesnæs, 1996; van der Ven and van Dam, 1996), while others include all processing steps (e.g. Wrisberg *et al.*, 1997; Lindeijer *et al.*, 1997; Meier, 1997), although data verification is the only P aspect that is addressed in these approaches. The approaches of Wrisberg *et al.* (1997) and Lindeijer *et al.* (1997), which aim to provide information on data quality and not on the overall quality of the

2 In this paper we use the definition of Pedigree given by Funtowicz and Ravetz (1990), which is different from the one used by Weidema and Wesnæs (1996).

LCA, are the only ones that follow the *qualitative indicator* method. The use of the probability distribution function method to obtain information on the overall quality of an LCA outcome has been illustrated by Kennedy *et al.* (1996, 1997), Meijer (1997) and Huijbregts (1998a,b), although they did not use the full scale of quality indicators as proposed here. The quality indicators used by Meier (1997) are similar to those of Weidema and Wesnæs (1996), while Huijbregts (1998a,b) also addresses model choices such as the allocation method and the time horizon for the global warming potential.

2.4. Choices for operationalisation

The framework presented in figure 1 may be further operationalised according to the requirements proposed in table 3. The requirements are general in character or are related to the intended application of the model. A distinction can be made between theoretical and practical requirements. The theoretical requirements indicate the formal correctness of the quality assessment results, while the practical requirements concern the usefulness and practicable aspects of the model.

Table 3: Requirements for the operationalisation of the quality assessment framework.

Theoretical requirements	Practical requirements
1. The quality assessment model should be reliable and valid. This means that the results of the model should be reproducible and should unambiguously predict the quality of the results.	1. The description of quality indicators should be detailed and transparent, so that the assessment result is independent of the assessor.
2. Subjective elements should be stated, their role(s) analysed and, if possible, should be based on scientific consensus.	2. Preferably, the overall assessment should be *expressed* as quantitatively as possible.
3. An improvement in the quality of one particular input data element or model element should retain or result in an improvement in the overall quality.	3. The model should allow the user to trace back the causes of low quality.
4. Depending on the intended aim of the model, it should be able to give results for the entire system under study, by combining flow data and process data into system data.	4. The model should allow an upgrading of quite coarse basic information categories on data quality, while keeping intact the operational procedures for assessing the quality of LCA outcomes. Uniformity in the sense of using only sophisticated ideal indicators is unrealistic, also in the long term.
5. The model should be complete with respect to the relevant quality aspects according to the intended application.	5. The model should be easily applicable and easy to use.
The quality assessment models operating for one or more quality elements in each of the four LCA phases mentioned should be compatible.	

The requirements listed in table 3 give guidance to a number of different choices which are to be made concerning the operationalisation of the framework shown in figure 1. The first choice is related to the goal of the quality assessment, which may first of all have implications with respect to the choice of quality factors shown in table 1 for operationalisation. For instance, a quality assessment model for the purpose of data exchange may only concern the validity and reliability of the inventory input data.

The second choice concerns the level of application, as discussed in section 2.2. This choice is also influenced by the goal of the quality assessment, although practical considerations may dominate. For instance, in an LCA with a few hundred processes and about one hundred independent data per process, starting at the level of individual flows makes quality assessment quite cumbersome, if not impossible. Against this background, we believe that the assessment may start at the level of an individual (unit) process and environmental issue.

The next important choice concerns the extent to which the assessment model should aggregate the quality factors. Any aggregation of qualitative quality factors implies subjective choices. Therefore, a strategy may be chosen in which subjective aggregation is avoided as much as possible in exchange for a more complex quality assessment outcome. Only data uncertainty and variability specified in terms of probability distributions can be aggregated in an objective manner.

If it is decided to aggregate the quality factors into quality indicators and into main quality aspects, the next choice to be made is how to do this – whether to use the qualitative indicator method or the probability distribution function method. We recommend that the qualitative indicator method is used for the qualitative quality factors, and that the probability distribution function method is applied only when probability distributions are available. The reason for this is that the subjective transformation of qualitative quality factors into probability distributions is not transparent, and may give the false impression that it is scientifically well founded.

With the qualitative indicator method, the aggregations can be performed in various ways, ranging from simple averaging to weighted averaging according to a certain principle. For instance, the relative importance of the contribution of specific input data to the overall quality can be assessed using information on its contribution to the LCA outcome, e.g. using normalisation factors or an eco-indicator for making a weighted average. This approach will at the same time provide information on areas where quality improvements will lead to an improvement in the overall quality of the LCA result.

Assessing the relative importance of model choices is more difficult. Models simulate reality, although it is impossible to determine what is reality in absolute terms. Instead, reference is made to scientific consensus or to conventions such as those laid down in ISO standards, against which the quality related to a model choice may be assessed. It should be noted that scientific consensus on a number of methodological issues, such as methods of impact assessment is often lacking. The lack of experience on how to assess the quality of model choice is a general problem, which indicates the importance of procedural quality elements (the P aspect), such as peer review, sensitivity analysis and scenarios.

3. Example of the proposed semi-quantitative quality assessment system

3.1. Introduction

Based on the framework and requirements described in chapter 2, we have developed an operationalised example to show how such a quality assessment method could look. In this operationalisation some choices are pragmatic, and others could be possible. In fact, this will always remain the case. Therefore, an international harmonisation process is required to make a generally acceptable proposal. For now, this example is intended only to illustrate the consequences of the framework described in chapter 2.

The proposed semi-quantitative quality assessment system is based on various inputs from the literature (Weidema and Wesnæs, 1996; van der Ven and van Dam, 1996; Wrisberg *et al.*, 1997; Ruyter, 1997), and on our own experiences during a project to combine these inputs with the practical restrictions of a case study for RIONED, a foundation for co-operation among sewerage system stakeholders' in the Netherlands (Lindeijer *et al.*, 1997). For this paper, we have elaborated the approach to bring it into line with the framework developed here.

The intended applications of the quality assessment model presented here are as follows:

- to assess the overall quality of an LCA result;
- to identify data which significantly detract from the quality of the LCA outcome; and
- to assess the quality of data in new databases for unit processes, allowing the setting of a minimum quality standard; this may also be applicable to existing databases.

In this example, we focus on the first application, since this is the most general application of a quality assessment model. This is a first practical attempt to operationalise the whole quality issue related to LCA.

In its present form, the quality assessment model uses 15 different data quality factors related to unit processes or to whole systems. The quality factors could be applied at the flow level, but in our experience this is too time-consuming for most purposes. Each factor is assigned a score from 1 to 5, according to a pedigree table. Poor data quality are assigned a score of 1, and the best quality data a score of 5.

The input data reliability and validity have been operationalised for the inventory data only, but the data used in the impact assessment can and should also be scored. The uncertainty in these data may be fully quantified, as with the inventory data, in which case the uncertainty ranges should be included in the reporting of the results rather than being scored in a separate table. If they are not included, separate quality factors for these data should be used, adding to the number of factors shown here.

Table 5 shows the operationalisation of the quality aspects for Spread/reliability and Assessment. However, only a limited number of the total set of quality aspects have been included, often at a higher level of aggregation than ideally required. For instance, once the data have been gathered and aggregations have been performed, the uncertainty in all individual flows can not be determined practically. Establishing quality at a more aggregate level than the flow level reduces the amount of work involved in establishing the scores. Such aggregated scores, for instance at the process or system level, can however only be made with a certain fuzziness. The limited number of quality aspects covered is illustrated in table 4, where the ideal scheme of table 2 is filled in for this practical operationalisation.

In the scheme in table 5, a comparison is also made with another important quality assessment approach suggested by Weidema and Wesnæs (1996), where an asterisk (*) is used to indicate if the parameters are comparable, and *italics* if they are additional. It can be seen that the latter approach and the present one overlap for two parameters. Weidema's geographical representativeness is not considered to be a separate parameter, as it can be described by technological and temporal representativeness, except for climate-dependent processes such as in agriculture; we consider climate-related issues to be incorporated in the technical description. In its operationalisation, Weidema's completeness parameter (see also van der Ven *et al.*, 1999) is considered to refer to data variability (S2). Data reliability is operationalised as a P aspect (including verification) in Weidema and Wesnæs

(1996), and does not fit in the S and A table below. Here reliability is considered to include uncertainty (when not fully quantified), variability and completeness, and model reliability is added to the data reliability. Many other quality aspects are also included here, and many others are missing from the present operationalisation, as can be seen from table 4.

Table 4: Operational elements of a method for the quality assessment of LCA results as described in the following.

Quantified elements of the LCA	Level of application	Operationally indicated below: *Reliability* *Data*	*Model*	*Validity* *Data*	*Model*
Scope	Reference flow	☒	☒	☒	☒
Inventory	Economic flow	☒	☒	A5 level of aggregation	-
	Environmental flow	☒	☒	☒	-
	Process	S2 data variability *(*) completeness*	☒	A1 data exclusion A3 (*) temporal representativeness A4 (*) technological representativeness *(*) geographical representativeness*	☒
	System	S1 uncertainty S3 completeness	S4a reproducibility	A2 cut-offs ⦶	A6a LCA invent. A9 allocation
Classification	Env. flow selection	-		-	☒
	Env. issue selection	-			☒
Characterisation	Environmental flow	☒		☒	☒
	Environmental issue	⦶	⦶	-	☒
	Set of envissues	⦶	S4b reproducibility	⦶	A10 characterisation
Normalisation	Env. flow	☒	-	☒	-
	Total of env. flows	⦶		⦶	-
Evaluation	Env. issue	☒	-	☒	☒
	Set of env. Issues	⦶	☒	⦶	⦶

Key: ☒ = to be measured independently, but operational methods for measurement lacking.
⦶ = measurement derived from less aggregated quality measurements.
\- = not applicable.

The more general aspects such as the validity of the scope and the functional unit can also be scored, but these can hardly be regarded as separate from genuine Pedigree aspects such as the verification of data, critical review and comparison with related studies. Also the partial, qualitative, and fuzzy way of establishing the relevant quality parameters makes it relevant to add aspects of Pedigree. These P aspects have not yet been worked out, and are omitted from the scheme in table 5. In fact, for the Pedigree quality aspects (incorporating, for instance, interpretation issues such as sensitivity analysis, gravity and dominance analyses, and verification issues such as the status of the practitioner, external plausibility, parts of the model tested, comparison with similar models), a scoring system does not seem adequate. A checklist with a reporting per item is more suitable. Thus, although the Pedigree aspect is part of our approach, we have not elaborated it further here, since the procedural aspects require a separate discussion.

Table 5: Example of the proposed LCA quality assessment method, including quality aspects, indicators, factors and scores

Quality aspect; Indicator		Score 5	Score 4	Score 3	Score 2	Score 1
Spread/reliability Input data reliability	**S1 Inventory data uncertainty**	Fully quantified or 5% or less	Fully quantified or 5-10%	Fully quantified or 10-25%	Fully quantified or 25-50%	Fully quantified or unknown or larger than 50%
Spread/reliability Input data reliability	**S2 Data variability** (statistical representativeness)	Variability in flow absent or known; adequate measurements/ calculations performed	Variability known to a certain extent; adequate measurements/ calculations performed	Variability only qualitatively known; a limited number of measurements performed or inadequate calculations performed	Data are gathered without taking variability into account	Unknown
Spread/reliability Input data reliability	**S3 Data completeness**	All known flows are included	Comparison with other data; missing data not from priority lists*	Comparison with other data; missing data from priority lists*	Comparison with other data sources limited	Unknown
Spread/reliability Model reliability	**S4 Reproducibility of the transformation model**	The process tree and the allocation methods can be reproduced easily and exactly	The process tree can be reproduced completely only with difficulty, and the allocations performed are clear but can not be reproduced	The process tree can be reproduced almost completely but the allocations performed on it are not transparent	The process tree can be only partly reproduced and the allocations stated to have been performed can not	The transformation model is not transparent and can not be reproduced
Spread/reliability Model reliability	**S5 Reproducibility of the calculation model**	The software is transparent and all methodological choices can be reproduced easily	Data, allocation and cut-off rules can be reproduced but the IA calculations can not	Data are available for reproduction but the allocation and cut-off rules can not be reproduced	The data calculations can not be reproduced due to confidentiality, but the IA calculations are reproducible	The calculations were performed by hand or the software is inadequate to reproduce the results
Assessment Input data validity	**A1 Process data in/excluded** consistently and according to the scope	Data explicitly agreed upon to be ex/included; applied consistently	Adequate data in/excluded, but inconsistently applied	Consistent in/exclusion of data but without clear links to the goal and scope	Inconsistent in/exclusion of data, without links to the goal and scope	Unknown or no guideline stated in scope
Assessment Input data validity	**A2 Cut-offs** consistent and according to scope (In/excluded processes)	Adequate, non-arbitrary cut-off criteria have been applied consistently	Adequate cut-off criteria are inconsistently applied	Cut-off criteria consistently applied but without clear links to the goal and scope	Non-relevant cut-off criteria, also inconsistently applied	Unknown or no guideline stated in scope

Assessment Input data validity	**A3 Temporal representativeness** of the data	Process data on average ≤3 years older than reference date in scope of the study	Process data on average 3 to 6 years older than reference date in scope of the study	Process data on average 6 to 10 years older than reference date in scope of the study	Process data on average 10 to 15 years older than reference date in scope of the study	Process data on average more than 15 years older than reference date in scope of study
Assessment Input data validity	**A4 Technological representativeness** of the data	Adequate company-specific data	Data on the same process and outputs, but from a different company	Data from processes with the same outputs but using different technologies	Data on comparable outputs using the same technology	Unknown or data on comparable products but using different technologies
Assessment Input data validity	**A5 Level of aggregation**	No substance flows are aggregated	Flows with comparable environmental properties are aggregated	Flows with incomparable properties are aggregated	Flows from priority lists* are aggregated	Unknown
Assessment Model validity	**A6 Validity of LCA model**	The model used adequately matches the validity statement made in the goal	Detailed validity statements have been made but the model poorly matches them	Vague statements on the validity have been made, but the model seems to fit them	Vague statements on the validity have been made, and the model does not even fit these	No requirements for the validity of the LCA model have been stated
Assessment Input data validity	**A7 Validity of reference flow**	The reference flow is quantified according to representative data for all alternatives and ranges are included	The reference flow is quantified according to representative data for all alternatives, but ranges are not included	The reference flow is quantified according to representative data for most alternatives and ranges are included	The data to quantify the reference flow have only limited representativeness, but ranges are included	The data to quantify the reference flow have only limited representativeness, and ranges are excluded
Assessment Model validity	**A8 Validity of functional unit**	F.U. is in line with the detailed scope of the study and the process trees match well with this	F.U. is in line with the detailed scope of the study but the process tree poorly matches it	F.U. matches the scope but the process tree details do not match the functional unit	The function is poorly defined but the F.U. matches reasonably with the function as stated	The function is not stated explicitly; the functional unit is inappropriate
Assessment Model validity	**A9 Multi-output and open loop allocation rules**	At least two relevant allocation approaches can be or are applied or no allocation needed	One allocation approach has been applied consistently	Allocation is performed inconsistently or is inadequate	Allocation is performed inconsistently and is inadequate	Unknown or allocation is not performed
Assessment Model validity	**A10 Adequate characterisation (IA) methods**	One or more adequate IA methods applied	Two or more IA methods applied but adequacy not properly assessed	One poorly adequate IA applied due to lack of alternative methods	One inadequate IA method applied when more adequate ones were available	IA method not performed when necessary for scope, or inadequately performed when not required

* For instance, Dir. Econ. Comm. of Europe, Commission for Environmental Policy (ECE/CEP2): Recommendations on implementation of the Convention on the Protection of Transboundary Watercourses and International Lakes, 1994. Part 1: List of hazardous (priority) substances.

Aggregation of quality aspects

In order to arrive at an assessment for the overall system, an aggregation over the scales for each factor needs to be made. Moreover, when required, an aggregation could be made over the various factors per quality aspect, in order to arrive at single scores per quality aspect.

For most factors, the aggregation over the five scales is already a subjective step, as the ranking in scores in our proposal is not based on a linear ratio scale. We strongly doubt whether a fully quantitative system (justifying an objective aggregation over the scales) will ever be possible. However, this aggregation is necessary in order to add the scores per factor over the life cycle, for those factors that are to be scored at the process level.[3] As long as one can agree on the scaling, and it can be applied consistently, some subjectivity is acceptable here. The main problem with this aggregation is that the uncertainties in the economic inputs and outputs of a process are generally not given. This means that quantitative propagation of the quality scores can not be done properly.

Accepting this, there are two options for performing this "second best" aggregation: the first is based on the direct environmental burden of all processes, and the second is based on the "cradle to gate" environmental burden of all processes. Each option has its specific drawbacks, as are explained in the following.

(a) The direct environmental burden method

$$QF_{x,\,tot} = \Sigma\{(QF_{x,\,i}) * (EB_i \,/\, EB_{tot})\},$$

where x is a certain quality factor (QF), i is an index over all processes, and EB_i is the total impact assessment (IA) score for all processes i in the process tree (the total contribution of i to the environmental problem).

Thus the quality factors are weighted with the total environmental relevance of the process, taking the amount of process i required into account. The drawback is that the consequences for the process tree of the uncertainty in the economic flows cannot be incorporated, even if these uncertainties could be known or estimated.

(b) The "cradle to gate" environmental burden method is a procedure for performing an aggregation of each QF over the whole system:

1) Determine the total environmental burden of the system (EB_{tot}).
2) For every process j, determine its upstream environmental burden, i.e. up to the cradle per main unit of output ($\Sigma\, EB_{j,u}$).
3) Score each process level quality factor x (QF_x) and multiply by ($\Sigma\, EB_{j,u}/EB_{tot}$) to weight with the environmental relevance of the process tree up to j, and add this information to the process file.
4) The total score per QF_x up to a certain process can now be determined by calculating through the scores over the process tree up to that process.

The drawback of this method is that processes are included several times, depending on the length of the chain, which implies double counting. Dividing by j is a rough way of compensating for this. The advantage of the method is that any known quantitative uncertainty in the economic inputs and outputs can be inherently incorporated in the assessment.

---

3 Scoring at the flow level has proven not to be feasible as long as quality information is not given with the data beforehand (Lindeijer *et al.*, 1997).

If this aggregation procedure were to be performed rigorously, a sensitivity analysis of different IA methods would be necessary to determine the weight per process. As the purpose of this aggregation (to get a general idea of the quality) is, in our view, limited, it is generally not relevant which IA method is used, and it may be sufficient to use the IA method envisaged for the LCI to be performed.

The major problem with this procedure is that completeness is included, whereas the weighting with environmental relevance excludes the missing data. This may be a valid argument for including completeness only in the P aspects.

Adding scores from different factors requires an even greater degree of subjectivity. How can scores on the reproducibility of the calculation model logically be added to scores on the reproducibility of the transformation model? In our view this is possible only when the overall target scores per factor have been set. This requires an explicit statement on the relative importance of each factor, which can be stated in the scope of the LCA or quality assessment study. Such an explicit quality statement in terms of quality factor scores is in line with the requirement to state such quality targets for the data in the goal and scoping phase of an LCA (see ISO 14040 (1997), section 5.1.2). When required, the relative target scores can then be used directly as weights.

4. Applying the Spread-Assessment-Pedigree approach

When surveying the battlefield, one may have the uneasy feeling that more complexity has been added than light has been shed. However, the issue cannot be avoided as differences in quality are there and have to be reckoned with in the real-life decision support that LCA is about.

Let us now evaluate what has been achieved, and what this means for practical quality assessment in LCA. First, the method as worked out here makes us sensitive to the different aspects related to quality, making easy ways out less acceptable. Second, by ordering the field, the first steps towards filling in the operational aspects may be taken, by a more orderly filling in of the field. Third, the limitations on establishing a firmly based quality analysis have been indicated. There is no easy solution and, because of the work involved, this remains a distant ideal. Fourth, the subjective and fuzzy elements involved in establishing the reliability and the validity of the results – the S and A of SAP – indicate that the internal and external plausibility of the LCA results are additional necessary elements in quality assessment. The internal plausibility relates to the cumbersome work of sensitivity analysis and other interpretation issues, which may indicate that some weaknesses in the analysis, such as unsatisfactory choices of methods, may not influence the results very much. The external plausibility relates to the similarity of the outcomes of other studies of related product systems and on "gut feelings". For such an assessment, the opinions of peers (e.g. critical reviewers) are very relevant.

In procedural terms, there are two very different applications of the above SAP approach. First, for those conducting the LCA, it provides guidelines for performing a better LCA, and it can help to structure the self-assessment of those executing the study, in the interpretation of results. The second application is external peer review. Here also, the method provides guidance on how to establish an overall judgement on quality, and, in the Pedigree aspect, invites critical reviewers to add their subjective but informed judgements on the results.

5. Conclusions

The framework for LCA quality assessment presented in this paper encompasses all quality-related information in an LCA, although the procedural/Pedigree (P) quality aspects, such as data verification and sensitivity analysis, have not yet been worked out. It allows the operationalisation of quality assessment models for different purposes, different levels of application and different strategies for processing quality-related information. The main aim of this framework is to enable the level of confidence in an LCA outcome to be determined when making a comparative assertion. The operationalisation of the framework has to fulfil a number of technical and practical requirements, such as subjectivity, which should be stated and if possible based on scientific consensus, or the transparency of a quality indicator, as indicated in table 3. This implies that theoretical considerations and practical options are to be combined in the "best" possible way.

In principle, the validity and reliability of all independent quality factors should be measured, which implies that all levels of application should be considered. In practice, however, this is hardly possible to achieve due to the lack of quality information and the large number of flows involved in an LCA. Therefore, the LCA quality assessment model should start at the level of individual processes and environmental issues. Quality-related information can in principle be processed via two different strategies, which we call the *qualitative indicator* method and the *probability distribution function* method. The validity-related quality factors and the reliability of model choice are inherently qualitative. Furthermore, for some of the factors related to the reliability of data, we assume that stochastic data will not be available for a long time to come, so that the conditions for statistical operations cannot be met. The assessment of the quality of the results can therefore only be based on softer procedures, with a prime role for qualitative elements, as outlined in the example of a semi-quantitative quality assessment in chapter 3 above. Combining several qualitative aspects in several steps into an overall judgement is inherently a subjective affair. Constructing a method as we have done cannot create objectivity; the subjectivity remains. However, the transparency of the quality assessment is improved, allowing for a more detailed discussion of the reasons given for the ultimately subjective assessment.

There are very many aspects related to quality assessment in LCA. Although the illustrative operationalisation of these aspects given in chapter 3 is not comprehensive, it is clearly more complete than other quality assessment approaches. A clearer distinction between quality aspects related to reliability, here called the spread (S), and validity, referred to as Assessment (A) on the one hand, and procedural (P) aspects on the other is required, especially with respect to data completeness, for inventory and impact assessment. An assessment at the flow level does not seem impossible as long as the providers of the data do not determine the scores while collecting and managing the data. The use of this comprehensive approach is likely to be very limited, however, unless there is general acceptance and harmonisation of the scoring system.

6. References

Bilitewski, B. and T. Hauptmann, 1999. *Realistic results in LCA by facing the uncertainty of our knowledge*. Research paper, TU Dresden.

Funtowicz, S.O. and J.R. Ravetz, 1990. *Uncertainty and Quality in Science for Policy*, Kluwer, Dordrecht. ISBN 0-7923-0799-2.

Huijbregts, M.A.J., 1998a. *Application of uncertainty and variability in LCA. Part I: A general framework for the analysis of uncertainty and variability in life-cycle assessment*. Int. J. LCA, 3(5), 273-280.

Huijbregts, M.A.J., 1998b. *Application of uncertainty and variability in LCA. Part II: Dealing with parameter uncertainty and uncertainty due to choices in life-cycle assessment*. Int. J. LCA,3(6), 343-351.

ISO 14040, 1997. *Standard on Environmental Management: Life Cycle Assessment*, DIS 14040. International Organisation for Standardisation.

Kennedy *et al.*, 1996. *Data quality: Stochastic environmental life cycle assessment modelling*, Int. J. LCA, 1(4), 199.

Kennedy, D.J., D.C. Montgomery, D.A. Rollier and J.B. Keats, 1997. *Data quality: Assessing input data uncertainty in life cycle assessment inventory models*. Int. J. LCA, 2(4), 229-239, 1997

Lindeijer, E., N.W. van den Berg and G. Huppes, 1997. *Procedure for Data Quality Assessment*, Report for RIONED, September 1997 (in Dutch).

Meier, M.A., 1997. *Eco-efficiency Evaluation of Waste Gas Purification Systems in the Chemical Industry*. LCA documents, vol 2. Ecomed, Landsberg, Germany. ISBN 3-928379-54-2.

Morgan, M.G. and M. Henrion, 1990. *Uncertainty: A Guide to Dealing with Uncertainty in Quantitative Risk and Policy Analysis*. Cambridge University Press, p.332.

Pohl, C. and M. Ros, 1997. *Are LCAs too precise?* Presentation at the 7th Ann. Mtg of SETAC Europe, Amsterdam , April 6-10, 1997.

Ruyter C.P.T., 1997. *The development of a Data Quality Indicator System in Life Cycle Assessment*. M.Sc. Dissertation, Hogeschool Zeeland, University of Lincolnshire & Humberside, June 1997.

Van den Berg, N.W., G. Huppes, E.W. Lindeijer, B.L. van der Ven, M.N. Wrisberg,1999. *Quality assessment in LCA. Framework and survey*, Int. J. LCA (submitted).

Van der Ven, B.L. and A. van Dam, 1997. *Data and LCA: Generation and Presentation of Data in view of Exchange of Information along the Production Chain*. Proc. SETAC Case Study Symposium, December 1997, Brussels.

Van der Ven, B.L., Wrisberg, M.N., E. Lindeijer, G. Huppes and N.W. van den Berg, 1999. *Survey of quality assessment methods in LCA*, Int. J. LCA (submitted).

Weidema, B. and M. Suhr Wesnæs, 1996. *Data quality management for life cycle inventories, an example of using data quality indicators*. J. Cleaner Production, 4(3-4), 167-174.

Wrisberg, M.N., E. Lindeijer, P. Mulders, A. Ram, B. van der Ven and H. van der Wel, 1997. *A semi-quantitative approach for assessing data quality in LCA* . Paper presented at the 7th Ann. Mtg of SETAC Europe, Amsterdam , April 6-10, 1997.