Procedural guideline for collection, treatment, and quality documentation of LCA data

Procedural guideline for collection, treatment,
and quality documentation of LCA data

Bo P. Weidema, Francesca Cappellaro, Raul Carlson, Philippa Notten, Ann-Christin Pålsson, Andreas Patyk, Emanuele Regalini, Francesco Sacchetto, Simona Scalbi

2004 ENEA
Italian National Agency for New Technologies, Energy and the Environment
Lungotevere Thaon di Revel, 76
00196 - Rome

ISBN 88-8286-110-4

Photo by Paolo Masoni

Italian National Agency for New Technologies,
Energy and the Environment

Procedural guideline for collection, treatment, and quality documentation of LCA data

Editor:
Bo Weidema

Authors:
Bo P. Weidema[1], Francesca Cappellaro[2], Raul Carlson[3], Philippa Notten[1], Ann-Christin Pålsson[3], Andreas Patyk[4], Emanuele Regalini[5], Francesco Sacchetto[5], Simona Scalbi[2]

The CASCADE Project is supported by funding under
the European Commission's Fifth Framework "Growth" Programme

[1] 2.-0 LCA consultants, Borgergade 6, 1., 1300 Copenhagen K., Denmark
[2] ENEA, Via Martiri di Monte Sole 4, 40129 Bologna, Italy
[3] CPM, Chalmers University of Technology, 412 96 Göteborg, Sweden
[4] IFEU, Wilckensstraße 3, 69120 Heidelberg, Germany
[5] Ecobilan s.r.l., Corso Europa 2, 20122 Milan, Italy

CONTENTS

Foreword

This Volume has been produced in the context of the EU-project CASCADE - Co-operation And Standards for life Cycle Assessment Data in Europe (contract no. G7RTCT-2001-05045). CASCADE is a Thematic Network of the GROWTH program carried out during 2001-2004. The partnership includes practitioners and experts in Life Cycle Assessment (LCA), producers of software for LCA, information modelling experts, industrial companies and two national environment agencies.

Main objectives of the project are:
1. To introduce the environmental data in the design processes
2. To facilitate data exchange and independence from any computer system
Those objectives are both “key points” and “bottle necks” for a wider diffusion of LCA and eco-design tools.
They are achieved adapting the standards developed for the communication of product, process and property data in design and manufacturing to the requirements for LCA. Three types of standards for computerised information representation and communication have been adopted:

• ISO 10303 provides entity-relationship models for product, property and process data;

• ISO 15926 provides classification structures for the description of industrial data

• Web ontology provide web-based classification structures for the description of information.

These standards are used for the representation of examples of LCA data in accordance with ISO 14048 and their communication between different systems.

Distance learning packages has been developed to improve knowledge of the use of LCA and the benefits of information communication standards. A specific training course deals with the collection, treatment, and quality documentation of Life Cycle Assessment data and it has been produced on the basis of the present guideline.

The project has also established international cooperation and liaison with ISO TC184/SC4, responsible for the ISO 10303 and 15926, and with VAMAS (Versailles Accord on Materials and Standards) for cooperation in the evaluation and testing of the outcomes.

Bologna, 22 October 2004

Paolo Masoni (ENEA)

CASCADE Co-ordinator

Chapter 1. Introduction

1.1 Purpose and context of the guideline

The *purpose* of this guideline to support the development, implementation and maintenance of data collection systems within environmental product life cycle assessments (LCAs) and hence improve the quality and comparability of LCA data.

The *target group* for this guideline is managers of LCA databases and data collection systems, as well as the experienced LCA data collector.

The guideline provides technical guidance and recommendations for the procedures to be used for data collection and data treatment (Chapters 2 to 9), and includes minimum requirements for data quality documentation (Section 1.4). This is supplemented by a procedure for certification of data according to the guideline (Section 1.5).

The guideline does *not* cover the more practical aspects of data collection, such as how to design data collection protocols, how to train data collection personnel, or the psychological aspects of how to address data suppliers most efficiently.

The guideline follows the terminology of ISO-standards 14041, 14042, 14043, and 14048, and should be seen as a technical supplement to these standards.

The guideline has been produced in the context of the EU-project CASCADE (contract no. G7RT-CT-2001-05045). The outline for this guideline was originally produced by the CPM group (Carlson & Pålsson 2001).

1.2 The nature of LCA data and the relation to similar types of data

A life cycle assessment is a compilation and evaluation of the inputs and outputs and the potential environmental impacts of a product system throughout its life cycle (ISO 14040). A product system is composed of a number of processes, which together are the human activities necessary and sufficient to deliver a specified product. A product is defined in terms of its functional performance, and can be either a physical good or a service.
In the terminology of the ISO 14040-series of standards, the concept of "inputs and outputs" cover also environmental aspects that are not directly linked to the flows to and from the product system, such as the physical impacts on the land area occupied, or the working hours and working conditions of the personnel.

LCA data thus refer primarily to these environmental "inputs and outputs," but in order to be meaningful, data must also be available that describe the processes that are the destination and origin of these inputs and outputs, as well as data describing the product system that results from a specific combination of these processes. Furthermore, the translation of "inputs and outputs" into environmental impacts requires data that describe the environmental mechanisms (impact pathways) in the ecosystems. And finally, to assess the environmental impacts, data are required on the value systems used in this assessment.

In this way, LCA data are not unique to LCA, but are also available from and applicable in many other contexts. Data on environmental inputs and outputs are also used for environmental reporting, green accounting, and elsewhere in environmental management systems. Technical information on processes is used in process design and optimisation, and data on product systems are used in supply chain management and for national input-output tables. Data that describe environmental mechanisms in ecosystems are used in other environmental impact assessments, and data on value systems are used in opinion research and policy making.

What makes a specific piece of data or a dataset "LCA data" is thus not the nature of the data itself, but the use it is put to. A specific questions asked in an LCA requires a *unique combination* of environmental data from processes into product systems that are meaningful in the context of the question posed. Only in this way are LCA data unique.

1.3 The structure of this guideline

This guideline is structured in the same way as the procedure of data collection, treatment, and quality documentation.

First, a data collection system must be established (Chapter 2) before meaningful data collection can be initiated. The procedure of data collection starts with the collection of individual data values (Chapter 3) for each item, whether by measurement, estimation or calculation. Probability distributions are then formed (Chapter 4) from a number of individual data values for the same item.

The individual probability distributions are then brought together into a description of a process (Chapter 5), an environmental mechanism (Chapter 7), or a value system (Chapter 9). A process may be a unit process (the smallest portion of a product system for which data are collected) or a system of linked unit processes, eventually establishing a complete product system. An environmental mechanism may be a single relationship or a model of linked relationships, eventually establishing a complete characterisation model.

To obtain aggregated data for processes or product systems, the boundaries between unit processes within the systems are removed (Chapter 6). To obtain aggregated data for environmental mechanisms (named "characterisation factors"), the boundaries between the individual relationships within the characterisation models are removed (Chapter 8).

Descriptions of the environmental impacts of processes (named "LCIA profiles" composed of indicator results for each category of environmental impact) are obtained by a procedure named characterisation (ISO 14042), in which the aggregated or non-aggregated "input and output" data are combined with the characterisation factors. The indicator results may be further converted and possibly aggregated across impact categories using the weighting factors resulting from one or more value systems.

Each individual chapter of the guideline is sub-divided into sections covering definitions, a description of the actions to be taken, procedures for documentation of these actions, validation procedures and communication issues.

1.4 Recommended minimum requirements for data quality documentation

Data quality is an agreement between data provider and data user, and there are many different ways in which providers and users may define requirements. The choice depends on the contextual distance between the two, e.g. whether they both are familiar with the processes described, the LCA methodology applied or the impact assessment methodologies implied. If, for example the process describes the production of a specific chemical including electricity production, waste management etc. and the data user resides in a food processing industry, he or she may need more documentation to understand the data set than if he or she were in a similar chemical production company. Also, since there are different LCA methodologies to apply, both on the inventory part and on the environmental impact assessment part, the data provider and the data user may have a need to express and understand each other's choices. Due to this context dependency, it is difficult to recommend general requirements on documentation.

However, recommended minimum requirements for documentation are provided in this document in sections:

- 3.3.1 for individual data values
- 4.3.1 for probability distributions
- 5.3.1 for processes

1.5 Procedures for certifying data according to this guideline

Each individual chapter of this guideline contain details on validation procedures. In addition to these validation procedures, it is possible to establish a certification procedure, based on the concept of self-organised third-party certification.

Certification ascertains and confirms that a particular dataset or group of datasets meets the minimum requirements referred to in the preceding section. It is also possible to add additional certification requirements and thereby creating several levels of compliance, i.e. several levels of data quality.

In self-organised third-party certification, the data supplier chooses an independent third party as certification agent. The independence and qualifications of the certifying agent can be ensured by requiring that the agent is chosen from a list of accredited agents, maintained by an authoritative accreditation body. The certification agent performs the necessary validation against the set requirements, and given a satisfactory result of the validation, the agent issues a certificate, which may have the form of an electronic tag. The certificate, with information on the date of issue and the name of the certifying agent, then follows the dataset, as long as the dataset is kept unmodified.

Chapter 2. Establishing data collection systems

2.1 Definitions

LCA data is data that describe environmentally significant *properties* of either of the following *items*:

- Human activities described as processes, e.g. production, transportation or market processes.
- Environmental mechanisms, e.g. physical, chemical or biological reactions in manmade or natural biotopes.
- Social groups, e.g. communities.

Environmentally significant properties are identified from knowledge of how strongly they contribute to an environmental change and how well they represent or describe the importance of such a change.

LCI data is data that describe a process from its environmentally relevant perspective. An LCI data item consists of a definition of the process, its boundaries etc., and its inputs and outputs. In particular *intermediate flows* indicate mass or energy exchanges between two processes, while *elementary flows* indicate mass or energy exchanges between a process and the natural environment. Processes also make an environmental impact on the nature in other ways than through physical flows, e.g. by occupying land or water areas, or by rearranging such areas so that the productivity of ecosystems is reduced. Such non-flows are also included under the term *inputs and outputs.* According to ISO/TS 14048, a *process* can be both a unit process (the smallest portion of a product system for which data are collected) and a system of linked unit processes. Changes in stocks of raw materials, components and produced goods are usually not part of LCA data, since the LCA methodology is only considering inputs and outputs. However, inputs and outputs may indirectly be estimated from data on changes in stocks.

Data for impact assessment in LCA is

- *Characterisation factors*, i.e. numerical values expressing the aggregated contribution of inputs and outputs of processes to the indicator result of each relevant environmental impact category.
- *Weighting factors*, i.e. numerical values expressing the relative importance (weight) assigned by social groups to indicator results within an LCIA profile.

Data collection is the acquisition of quantitative and/or qualitative values of a property, either by measurement, estimation or calculation.

Primary data are data determined by direct measurement, estimation or calculation from the original source (typically for process emissions).

Secondary data are data collected from literature or other published media (typically for physical constants, carbon content of fuels).

A *data collection system* is a combination of procedures for:

a) identification of the data that needs to be collected,
b) planning when, where, and how data are to be collected and by whom,
c) identification and treatment of data gaps,
d) the actual data collection (measurement or retrieval from book, experience, expert, etc.),
e) documentation of the resulting data, together with possible sources of error, bias or lack of knowledge,
f) validation of the data collection system, the collected data and its documentation,
g) communication of the data and its documentation.

2.2 Action

2.2.1 Scoping a data collection system

In the scope for a data collection system, the following issues should be considered:

- *Objectives* of the data collection system.
- *Data requirements*, in terms of included items (processes, environmental mechanisms, groups of people) and the properties for which data are to be collected, and the quality requirements for these data.
- Procedure for identification and treatment of *data gaps.*
- Required *frequency* of data collection, *location* of data collection points, data collection *methods and units*, and how this matches the nature of the properties for which data are to be collected (type of data to be collected, natural variation versus requirements on precision, accessibility to measurement points, etc.).
- Required *documentation, validation and communication* of the data to fulfil the objective.
- *Personnel* involved, both at management and operational level.

Establishing a data collection system is initiated by identifying the *objectives and data requirements*. For efficiency reasons, only the actually needed data should be collected, unless there are uncertainties about what data is actually needed, or other good reasons for doing a broader than specified data collection. The following issues should be considered:

- What kinds of questions are to be answered by obtaining the data?
- What analyses or assessments need to be done to answer these kinds of question(s)?
- What data is needed to perform these analyses and assessments, and in what quality?
- What data is already available, and is the quality adequate?
- What data is it necessary to collect and with what quality and in what priority?

The answers to the above questions provide an explicit formulation of the expected outcome from the data collection system, in terms of a specification of the data requirements.

For example, the objective of a data collection system for the steel industry may be to provide the data needed to answer questions on the environmental performance of a specified range of steel products to be used in benchmarking between plants and in comparisons with competing materials. This objective can then be specified in terms of the benchmarking procedures and methods for comparisons. The data needs can be specified in terms of product and process specifications (e.g. steel bars from the Basic Oxygen Furnace production route), market trends, elasticities and production costs for these products and processes, their important inputs and outputs, and their characterisation and weighting factors. Data quality requirements can be determined from estimates of importance and uncertainty. Comparing the needs with the available data from the current data collection systems results in a prioritised list of data that needs to be collected.

At this stage it may also be useful to consider the required completeness of the data collection system and how *data gaps* are to be identified and avoided. Data gaps can be identified through:

- Careful examination of the initially available descriptions of the items to be investigated (e.g. reaction formulae, technical specifications, engineering calculations, budgets and/or bookkeeping).
- Comparison with data from similar items.
- Comparison with data from related items at a lower or higher resolution.
- Comparison with relevant checklists.
- Chemical "fingerprints" of inputs and/or outputs.

The required *frequency* of data collection depends on the variation and development in the property for which data are to be collected, compared to precision required. Some properties may be easy to measure on a continuous basis while others can only be determined on an interval basis (e.g. batch-wise). For mature technologies, such as the BOF steel production, data collection at 5-year intervals may be adequate, while annual updates may be required for technologies in more rapid development.

The data collection *locations* should be selected to reflect the desired representativeness of the data in view of the factors that may influence geographical variability (climate, landscape, soil type, density of population, raw material quality, legislation/regulatory differences, availability and costs of raw materials, labour and capital). Accessibility of measurement points and equipment (e.g. own production plants versus a supplier's production plant) and costs of measurement may also influence the choice of location.

The choice of data collection *method* depends on the required precision, the type of data to be obtained (e.g. a continuous flow or discrete batches, large variations or small, physical entities or personal values, simple systems or complex, etc.), the required skills of the data collection personnel, and the acceptance of the method by the intended audience. For measurements, national or preferably international standards should be used whenever possible. The advantage of applying and referring to standard methods is repeatability and comparability of the measurements. Legislation or authoritative requirements may also prescribe specific standards of measurement to be applied. Measured data have the advantage of being up-to-date and specific, while calculated data have the advantage of being based on theoretical models and not being affected by the possible errors of individual measurements (albeit it is affected by other uncertainties). Models are typically built on the basis of a large number of measured data. Models are especially relevant for forecasting, where measurements are not possible. An additional advantage of models is their explanatory power, which allows possible improvement options to be identified and their effects to be modelled. When it is possible to obtain data both from direct measurement and from calculation, it should be considered to use both collection methods in parallel, as they each have their advantages, and as they can be used for mutual validation.

The use of non-SI *units* may be advantageous for internal purposes (e.g. the collection of authentic process data in lbs, gal, BTU etc.), but should be strictly avoided in data sets prepared for publication or (electronic) data exchange. Since these units all have universally valid conversion factors to SI units, it is reasonably easy to convert them. Even when SI units are employed, many properties can still be expressed in ambiguous ways (e.g. nitrate may be reported as "kg nitrate as N" or as "kg nitrate as such", which differ by a factor of 4.4). Insecurity about the underlying measurement base impedes the exchange of data between practitioners, and can introduce serious errors if the base units are accidentally confused (de Beaufort-Langeveld & Bretz 2001).

Personnel for data collection should be assigned on the basis of their technical qualifications (required skills and knowledge of procedures). Links to other parts of the organisation can be advantageous to ensure coordination and resource savings in data collection.

As part of the scoping of a data collection system, and certainly before initiating any data collection, it should be described how the data collection system will handle the *documentation, validation and communication* of the resulting data to fulfil the earlier defined objectives. More details on these issues can be found in the following chapters.

For each procedure in a data collection system, the *personnel* responsible for the procedure should be assigned as part of the scoping.

2.2.2 Recommended practice

Identify the objectives and prioritise the data requirements on the basis of importance and uncertainty. Compare to already available data from the any current data collection systems.

Consider the required completeness of the data collection system and how data gaps are to be identified and avoided.

Determine the required frequency of data collection on the basis of the variation and development in the properties for which data are to be collected. Select the data collection locations in view of costs, accessibility, and the factors that influence geographical variability.

Prescribe standardised data collection methods and SI units whenever possible, and avoid ambiguous units and expressions.

Include a description of how the data collection system will handle the documentation, validation and communication of the resulting data.

Assign responsible personnel for all procedures and specify any qualifications required for the operational data collection personnel.

2.2.3 Checklist of things that can go wrong

- Unfocused data collection system without clearly defined objectives
- Lack of accordance between objectives and listed data requirements
- Lacking assessment of the importance and uncertainty of required data
- Lacking analysis of existing data and data collection systems
- Lacking identification of data gaps and procedures for treating these
- No analysis of the properties' variation as basis for selecting frequency and locations of data collection
- Ambiguous data collection methods, units or expressions
- Lack of specification of relevant qualifications for data collection personnel
- Lacking description of how the data collection system will handle documentation, validation and communication
- No responsible assigned for the described procedures.

2.3 Documentation of data collection systems

Documentation of data collection systems should include:

- A description of the objectives of the data collection system, and the underlying technical and scientific understanding.
- Estimates of importance and uncertainty of data in relation to the objectives.
- A description of relations to other data collection systems, and possible redundancy in relation to these.
- A description of the identification and treatment of data gaps.
- A prioritised list of data requirements, in terms of included items (processes, environmental mechanisms, groups of people) and the properties for which data are to be collected, and the quality requirements for these data.
- Locations and frequency of data collection (when necessary for each item and property for which data are to be collected), with indication of arguments for the specifications.
- General specifications of data collection methods and units, and when necessary specific instructions for individual properties.
- References to relevant data collection instructions, manuals, handbooks, standards, etc.
- A description of the procedures for documentation, validation and communication of the data outcome from the system.
- Qualification requirements for the operational data collection personnel.
- A list of personnel responsible, per procedure.
- A description of any subjective choices and compromises to the theoretical data requirements, made to practically install the data collection system.
- A procedure to document any deviations from the stated procedures and specifications.

2.4 Validation

Validation of data collection systems has much in common with validation of management systems. The validation may be made both prior to their implementation (to prevent incomplete and unfocused systems from being implemented) and during their functioning (to correct any deficiencies).

Validation should be done by professionals with experience in establishing and maintaining data collection systems.

Validation may include:

- Check for focus and clarity in objectives and their implementation.
- Check for system completeness.
- Check for clarity of instructions, manuals etc.
- Check on qualifications and competence of personnel and the resources available to them.
- On-site checks that instructions are actually followed.

The names and roles of people involved in the validation procedure(s) should be documented.

2.5 Communication

2.5.1. On the communication of information on data collection systems

The target groups for information on the data collection system are:

- the current and future administrators (managers and validators) of the system itself,
- the personnel managing and performing the procedures of the system,
- users of data provided by the data collection system.

For the system administrators the entire information listed in Section 2.3 would be relevant, with specific emphasis on the rationales for each specification or procedure.

For the operational personnel, it is adequate that the documentation includes a short description of the objectives of the system, while placing emphasis on the responsibilities and procedures for data collection (items and properties for which data are to be collected, the quality requirements for these data, locations and frequency of data collection, data collection methods and units, references to relevant instructions etc.), documentation, validation, communication and personnel management.

In spite of the above distinction between administrators and operational personnel, it may be beneficial for the quality of the data collection that the operational personnel also have access to the rationales behind the specific procedures and specifications. However, this explanatory part of the documentation should be presented in a way that does not hamper the clarity of the operational instructions.

Users of data will seldom request detailed information on the system that supplies the data (provided that the data are themselves adequately documented, see later chapters), but a general description of the originating data collection system, targeted towards users of data, may nevertheless help answer questions such as: "Why are these data not provided?" and "why are the data provided in this form?" also facilitating the interpretation of specific datasets.

2.5.2 Issues of terminology

For chemical substances, SETAC (de Beaufort-Langeveld & Bretz 2001) has proposed a *nomenclature principle*: "The parameter name that gives rise to the least possible misunderstandings is chosen. The name must indicate what is actually measured (especially in sum parameters and indicators)," which is guiding the following *SETAC nomenclature rules*:

"All parameters are first sought in the CAS REGISTRY[6] system. If ever possible, one of the index names registered in CAS has to be used. This needs not be the shortest CAS name; the choice is made according to the following rules:

1. Inorganic anions reported as such (not as salts) carry their trivial names, which also indicate the oxidation level of the central element. Examples: "Chloride", "Chlorate".
2. Ores (mineral resources) are normally expressed as the element that is extracted from them. Example: "Cadmium Cd, in ground". However, mixed ores that yield more than one element or other minerals obtain their mineralogical name, rather than a chemical one. Example: "Chalcopyrite". Also, ores of different composition or "richness" can be distinguished by their mineralogical names (and possibly the content of the parent element).
3. In the cases of (a) very common trivial names (e.g. "Water", "Acetone", "Benzo(a)pyrene, "Methacrylic acid", "Nitrous oxide") and (b) pesticide names that are much shorter or simpler than the systematic names (e.g. "DDT"), these trivial names may be used.

[6] Note added by authors: www.chemfinder.com

4. Chlorofluorocarbons (CFCs and HCFCs) and halons must be identified by their CFC/ halon name *appended* to the chemical name, separated by comma. Example: "Ethane, 1,1,1,2-tetrafluoro-, HFC-134a".
5. Qualifying information follows the parameter name, separated by commas. Depending on the substance, the following possibilities occur:
    - Organics: "parent hydride" i.e. the principal (unsubstituted) chain or the preferred ring (system) comes first, the substituents follow with their positions, each (including the last one) ending with a hyphen. Example: "Ethane, 1,1-dichloro-1-fluoro-". One characteristic group (e.g. sulfonic acid) or functional class name (e.g. nitrile) can follow immediately (without blank) behind the principal chain or ring, before other substituents. Examples: "Benzenesulfonic acid, hydroxy-", "Adiponitrile".
    - Salts of oxygen-containing acids (inorganic and organic): The name of the acid appears first; the cation follows, if necessary including its oxidation level, or a counting prefix ("di, tri"). Examples: "Sulfuric acid, iron (2+) salt", "Carbonic acid, disodium salt".
    - Esters of organic acids: Acid appears first, followed by the alkyl group, possibly with a counting prefix ("bis"). Example: "Sebacic acid, bis(2-ethylhexyl) ester".
    - Isotopes: Name of parent element, followed by a dash and the mass number. Example: "Cobalt-57".
    - Metal ions: Named after the parent metal, followed by the word "ion". The oxidation level must be indicated, if necessary. Example: "Chromium, ion (Cr3+)". Ions are only reported as separate parameters if they need to be distinguished from the parent metal, e.g. for toxicological reasons.
    - Metal oxides and simple salts (sulfides, halogenides, also cyanides7): Named after the parent metal, followed by its oxidation level or a counting prefix if necessary, and the counter-anion. Examples: "Chromium trioxide", "Calcium chloride".
    - Hydrogen compounds: Simple hydrogen compounds follow the same rule as the analogous metal compounds (see above). Examples: "Hydrogen cyanide", "Hydrogen peroxide", "Hydrogen chloride".
6. Chemical formulae should be given whenever possible, but in a separate field and not in the name field. For complicated chemicals, sum formulae can be used. The sequence of elements in such sum formulae must be adhered to: C, H, N, O, P, Br, Cl, F, I, to make them searchable.
7. Salt formulae are written conventionally with the cation first. Formulae of ions indicate the charge as +, ++, 3+, -, --, 3-. To avoid confusion, a blank is left before the charge in complex anions.

The use of deviant synonyms may be appropriate for *internal* purposes (reflecting the language and nomenclature traditions of any particular site), but should be strictly avoided in data sets prepared for publication or (electronic) data exchange. For chemical substances, the conversion from local synonyms to terms that follow the global nomenclature rules should in most cases be very easy, due to the uniqueness of the CAS numbers."

[7] Cyanides, though having a complex anion CN-, are included here because of historic usage.

Chapter 3. Collecting individual data values

This chapter covers the initial collection of any data value that enters the data collection system, whether by measurement, estimation or calculation. Thus, the later chapters deal with ways of combining such individual data values, but no further data collection occurs in these later steps.

3.1 Definitions

Data collection methods fall into three main categories:

- measurement (quantitative)
- estimation (quantitative and/or qualitative)
- calculation (quantitative).

These methods can be used separately or in combination.

Measurement is an act of determination of the magnitude of a quantity by comparison with a standard unit for that quantity. Measurement is also the result of such an act. A measurement directly applied to a given property is considered *direct* with respect to that quantity.
Examples: Direct (continuous) measurement of liquid fuel consumption by a flow meter. Direct (discrete) measurement of coal consumption by weighing incoming vehicle loads. Direct measurement of the number of samples by counting.

Estimation is an act of obtaining the value of a given entity, not involving measured or otherwise known quantities related to the desired quantity. However, more generally, "*estimation*" may also indicate any method of obtaining the quantitative and/or qualitative expression for a given entity not solely based on direct measurements.

Calculation is an act of obtaining the value of a given property through mathematical operations or models involving already known data related to the desired property. Mathematical modelling is used to describe the relation between the desired property and the data already known.
Measurements used for the purpose of calculations are *indirect* measurements with respect to the desired property.
Examples: Thermal energy required by a unit process can be calculated from measurements of both room and operating temperatures as well as the specific heat capacity of the thermal fluid used to transfer heat to the unit process, and the efficiency of the heat exchange. CO_2 emissions can be calculated from measurements of the carbon content of the fuel and the completeness of combustion.

3.2 Action

The actual data collection follows the prescriptions documented in the scoping of the data collection system (see Section 2.5.1) with respect to locations, frequency, methods and units.

3.2.1 Measurements

For measurements, the prescribed national or preferably international standards are used whenever possible. When no measurement standard systems are in place, at least the following rules should be observed for good measurement practice:

- Use relevant, repeatable and documented measurement techniques.
- Use calibrated measurement equipment.
- Adjust for known bias.
- Check and report the conditions of the measurement (meta-data such as geographical location, time, technology, reference flow), especially such conditions that may influence the measurement.

3.2.2. Methods for estimation

Where measurements are not applicable, estimation is necessary. An estimate may be obtained:

- based upon generic expertise on the item and property for which data are to be collected,
- from legislative, standard, or handbook values,
- "rule-of-thumb" estimates.

Generic expertise on the item and property in question: This method relies on the experts' empirical grasp of the process. For processes, a machine operators experience may be adequate to obtain the desired precision for the value of a property. Often, an expert's estimate involves an element of interpolation or extrapolation from her/his knowledge of data for similar or related items (see also on inter- and extrapolation under "3.2.3. Calculation"). The collective experiences of many experts are the basis for empirical formulae or tables in handbooks and technical guidance documents (see also the following paragraph). Example: Emission factors for a chemical substance may be estimated when its use pattern and a few chemical properties of the substance (solubility and vapour pressure) are known. In the technical guidance document of the European Commission (1996), use patterns are specified in terms of "main category" (closed or open systems, inclusion in product matrix, dispersive or non-dispersive use), industrial category, and functional category of the substance.

Legislative, standard, or handbook values: Threshold values from legislation may sometimes be used as a proxy for measured data, when the real values can be expected to be close to the threshold value. Raw material and energy requirements are often specified in standards and technical handbooks. Example: The energy needed to grind a given surface of a metal piece with a desired degree of roughness can be found in engineering handbooks, based on machining parameters and the mechanical properties of the metal.

"Rule-of-thumb": This case applies when no better method is possible. It amounts to a best estimate and relies entirely on the best professional judgement of the practitioners or experts involved. This method should be checked out whenever possible by sensitivity analyses.

3.2.3. Calculations

Calculations may include one or several of the following items:

- Calculations based upon technical knowledge of the features and operating conditions of the machinery or equipment used by a process.
- Calculations based upon knowledge of the material properties or physical-chemical mechanisms involved in a process or mechanism.
- Indirect calculations (residuals calculated by subtraction).
- Interpolation or extrapolation of data from related items.
- Derivation from models.

Some of the most common calculation methods are illustrated below:

Technical knowledge of the features and operating conditions of the machinery or equipment used by a process: This case applies to processes where the performance is mainly determined by the given properties of a machine or piece of equipment, e.g. in the case when the electricity consumption of an electrical hand-drier is determined by the a built-in timer that give a fixed time per hand-drying event to be multiplied with the electrical effect of the device, or the case of water heating in, say, textile manufacturing where the quantities necessary for the calculation are: capacity of the machine water vessel, initial room temperature (or feed temperature), process temperature, and combustion efficiency.

Knowledge of the material properties and physical-chemical mechanisms involved in a process or mechanism: This case applies to processes that can easily be described by the main chemical reactions involved. For example, fuel consumption can be calculated from the heat requirement of the process, the fuel heat value and the combustion efficiency (see the next paragraph). Under certain assumptions (e.g. completeness of reaction), the stoichiometric relations among compounds on both sides of a balanced reaction can be used to calculate the quantity of products generated by a given amount of reactants. Typically, combustion processes burning fossil fuels can be described by the following balanced reaction (under the assumption combustion occurs with excess oxygen):

$C + O_2 => CO_2$

The carbon content in a fossil fuel is usually known from literature, while the molar ratios in the balanced reaction (in the above reaction 1:1:1) are easily converted into mass ratios. This method is also suitable for a wide range of processes designed to produce commercially viable compounds from common minerals or their decomposed forms. For instance, an energy balance could be calculated in order to determine the thermodynamical minimum energy needed to produce sodium hydroxide through caustification of hydrated lime, according to the following reaction and assuming the reaction is complete: $Na_2CO_3 + Ca(OH)_2 \rightarrow NaOH + CaCO_3$

Indirect calculations (residuals calculated by subtraction): This case applies when no direct or indirect measurements or calculations are possible for a given quantity. For instance, the energy consumed by the offices and general services in a plant could be calculated as a residual by subtracting the direct measured consumption of the manufacturing facilities from the measured total consumption of the plant. Likewise, data for an unknown process can be obtained as a residual by subtracting all known processes from a known sector total (e.g. from input/output statistics).

Interpolation or extrapolation of data from related items: When the desired data are not available from measurements, data for related items may form the basis of an interpolation or extrapolation, e.g. data for an industry sector may be used to represent a desired process within that sector, or related processes may be used as basis for interpolation or extrapolation. Likewise data for related areas, or for a larger or smaller area than desired, and for larger, smaller or entirely different time spans than desired may be used as basis for interpolation or extrapolation. The possible errors caused by interpolation and extrapolation, depend on the homogeneity of the populations used and the position of the targeted item in relation to these populations.

Derivation from models: A model is a representation of the relationships between different items, typically used for the analysis of the interactions of several simultaneous cause-effect mechanisms (e.g. explaining why emissions are higher for electricity from a specific coal-fired power plant and how much of the additional emission is due to differences in climate, boiler type, capacity, coal types, heat utilization etc.). Models are typically based on measured data. Models are especially relevant for forecasting, where measurements are not possible. An additional advantage of models is their explanatory power, which allows possible improvement options to be identified and their effects to be modelled.

3.2.4. Recommended practice

Follow the prescriptions documented in the scoping of the data collection system. When no measurement standard systems are in place, observe good measurement practice (see Section 3.2.1).

3.2.5. Checklist of things that can go wrong

The following items can become hindrances in obtaining the desired data:

- Unclear or misunderstood context
- Inadequate standards, procedures, personnel, equipment, calibration, and/or documentation
- Insufficient knowledge of the technical or chemical-physical characteristics of the measured item or property
- Unavailability of models to describe the measured item and property
- Far too great complexity of the phenomena involved
- Unreliable or biased measurements
- Estimation bias
- Calculation errors

3.3 Documentation of individual data values

3.3.1 Recommended practice

The documentation may be available on a paper form and/or electronically in a database. In both cases, an individual data value must be documented by:

- the unit of measurement,
- its amount,
- the necessary information to distinguish the data value from other data in the same series.

Furthermore, the individual data or the data series must be provided with the necessary identifiers (meta-data) to place the data in its appropriate context (its data collection system and/or applications), cf. Section 1.4, For example, reference must be given, either directly or indirectly, to the data collection method, geographical location, time, and other conditions pertaining to the validity of the collected data.

Unless covered in the description of the data collection method, the documentation must also include any specific assumptions, limitations and/or bias that affects the representativeness and/or applicability of the data.

Unnecessary calculations (e.g. conversion of units) that can introduce errors and obscure the original data and collection procedures should be avoided at this stage.

Individual data fields should be named according to ISO 14048.

3.3.2 Checklist of things that can go wrong

- Leaving out information necessary to distinguish the data value from other data in the same series
- Leaving out information necessary to place the data in its appropriate context
- Leaving out information on conditions that affects the data representativeness and/or applicability
- Unnecessary calculations (e.g. conversion of units) that can introduce errors and obscure the original data and collection procedures
- Use of local file names without reference to ISO 14048

3.4 Validation

Individual data values should be validated after collection and prior to further aggregation and calculation.

Validation should be done by professionals with data collection experience pertaining to the item and property in question.

Validation may include:

- Check for outliers by comparing to other data in the same series
- Recalculation for calculated data
- Identification of data gaps through logical checks, such as mass balances (total or broken down at elementary level) of inputs and outputs
- Cross-checking with data obtained through different sources and/or collection methods, e.g. comparing measured data with expected values from estimates or theoretical calculations and vice versa
- Benchmarking data collected for the given process versus data relative to the same industrial sector or technology
- Proof-reading

Validation should be performed both before and after any unit conversions or other calculations made for modelling purposes.

The names and roles of people involved in the validation procedure(s) should be documented.

3.5 Communication

The objective of communication is to transfer the information from data supplier to target group without any corruption, incompleteness or misunderstanding. To reach this objective, it is important to consider the context of the communication, especially the situation and needs of the target group for the communication, and the intended further transformation and use of the data.

The principle to be followed is that the person closest to the data origin, i.e. the data supplier, should provide as much information for the correct interpretation of the data as possible, so that the target group does not need to make any unnecessary guesses or unfounded interpretations.

With this in mind, the communication should be unambiguous, reliable, complete and referenced, cf. Section 3.3. Special care should be taken in defining the scope of applicability and the representativeness of data, so as to ensure that the data can be used appropriately. All terminology should refer to international standards, for life cycle data especially ISO 14048.

If the communication takes place in a public context, data suppliers should always tell whether data has been independently reviewed or peer-reviewed prior to disclosure.

3.5.1 Checklist of things that can go wrong

Communication can be unsatisfactory when one or more of the following situations occur:

- When the data supplier misunderstands the communication context, i.e. the needs of the target group and what the data are to be used for
- When the data supplier provides incomplete or ambiguous information on the data and the conditions of the data collection
- When the data recipient misunderstands the original conditions of the data collection and/or adds unfounded interpretations

The above three points may to some extent be seen as a cause-effect chain, one problem leading to the next.

Chapter 4. Forming a probability distribution

4.1 Definitions

Uncertainty: Data-related uncertainty may result from inherent variability in the item and property studied, from incomplete or biased data collection, and from disagreement between information sources. Uncertainty can also be about the structure of a model, which arises from the simplifications and approximations made during constructing a model. In addition to model and data uncertainty, subjective elements, such as values and preferences, may also be uncertain. It is important to recognise that only data related uncertainties are captured in a probability distribution. Thus, where used in this chapter, the term uncertainty is to be understood as data-related uncertainty.

Empirical quantity: Empirical or chance quantities represent properties or states of the world, and are measurable, at least in principle, now or at some time in the past or future. All continuous empirical quantities are uncertain, because no matter how great its precision, no experiment can measure a real-valued quantity with zero error (although in many instances the uncertainty is negligible for practical purposes). Empirical variables, as distinct from decision or choice variables, are the only type of variables whose uncertainty may appropriately be represented in probabilistic terms, as they are the only type of quantity that is both uncertain and can be said to have a true, as opposed to an appropriate or good, value. This chapter therefore only deals with empirical quantities. As used in this chapter, the term quantity is used to encompass all theoretically measurable data inputs into LCA models, i.e. the quantity may equally be an input into an inventory model or an impact model.

Probability: Probability is the most common measure of uncertainty. The classical or frequentist view defines the probability of an outcome in a particular measurement as the frequency with which it occurs in a long sequence of similar measurements. Classical probability estimates require relevant populations of similar measurements to be known, and so are not able to consider a lack of knowledge. Thus, when dealing with uncertainty arising from incompleteness, it is only meaningful to define probability according to the Bayesian view. However, according to the Bayesian or subjectivist view, the probability of an outcome is the degree of belief that a person holds that it will occur, given all the relevant information currently known to that person.

Probability distribution: A probability distribution is a commonly used way to represent the uncertainty about an empirical quantity, and gives the relative likelihood of the quantity having different possible values. Probability distributions are either represented as cumulative distribution functions (CDF), probability density functions (PDF), or they can be described in terms of their defining parameters. The CDF gives the probability that the random variable described by the distribution will be less than or equal to a particular value. The PDF is the derivative of the CDF, and gives the probability that the random variable will fall within some particular small interval. Parameters used to define probability distributions include the fractiles and moments of the distribution. The *fractile* of a distribution is a value corresponding to a particular probability that the value of the random variable will be less than the fractile value. If the probability is expressed as percentages, then the fractile is referred to as a *percentile*. For example, if the 50^{th} percentile has a value of 5, then there is a 50% probability that the value of the variable will be less than 5 (see Figure 4.1). The 0.25 and 0.75 fractiles are often termed quartiles, and the range between them referred to as the interquartile range (see Figure 4.1). The first *moment* of a distribution characterise its central tendency (e.g. the *mean* or *expected value,*

median and *mode*), the second moment characterise the spread or dispersion of the distribution (e.g. the *variance* or *standard deviation,* which is the square root of the variance), while the higher moments characterise its skewness (e.g. the *coefficient of skewness* which gives a measure of the degree of positive or negative skew relative to a normal distribution which has a skewness coefficient of zero) and its kurtosis (e.g. the *coefficient of kurtosis* which gives a measure of the "peakedness" or flatness of the distribution relative to a normal distribution which has a kurtosis coefficient of zero). Figure 4.1 gives an example of a CDF and its corresponding PDF, and some examples of the more common parameters used to describe them.

Frequency distribution: A probability distribution is referred to as a frequency distribution when its parameters are determined using classical statistical estimators alone, to distinguish from probability distributions that include also subjective probability estimates.

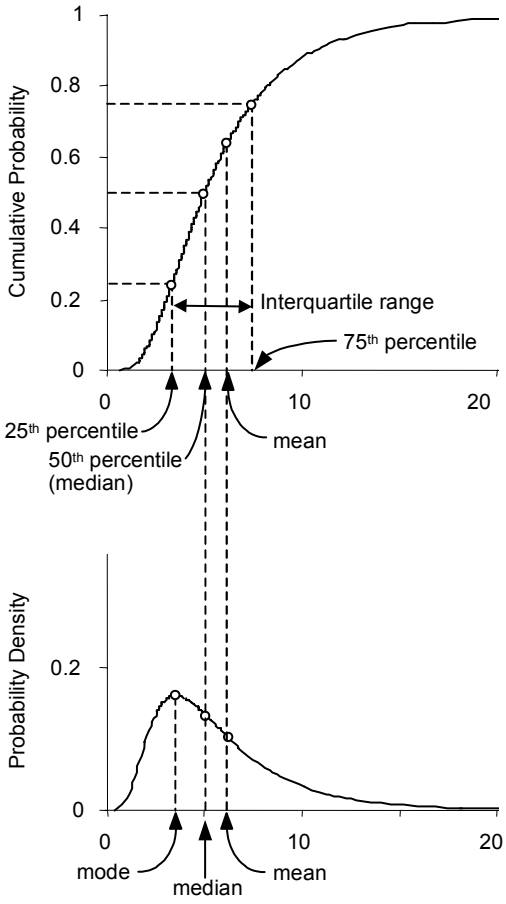


Figure 4.1 *Example of a probability density function and a cumulative density function, and some commonly used terms to describe them*

4.2 Action

4.2.1 Frequency and probability distributions

Any single value assigned to an empirical quantity, be it measured, estimated or calculated, is only one out of many possible values that the quantity could have. All possible values that a variable can have is termed its population, and a frequency distribution is a representation of this parent population. Typically a small number of individual data are taken to characterise a particular quantity, which together comprise the data sample. The data sample is then used to infer the size and shape of the parent population, and thus to estimate the parameters governing its frequency distribution.

Determining a frequency distribution is a relatively straightforward procedure. The many well established statistical methods are simple to apply, and give robust results provided a sufficiently large data sample is available (a minimum of 30 data points is recommended for homogeneous populations).

However, these methods are only able to represent the variability encountered in the data sample, and the frequency distribution formed does not account for such aspects as the reliability of the data collection. For example, if data are missing for a particular geographical area within a larger region for which data is required, the frequency distribution will not account for the uncertainty of using the data as representative of the larger region. In this case, an estimate of the uncertainty that will be introduced by applying the data to a different region must be made. Incorporating such essentially "un-measurable" elements of uncertainty involves an unavoidable degree of subjectivity. Because of this, extending the analysis to include all possible sources of uncertainty may not be desirable in all situations, i.e. it may be preferred to present the result as a frequency distribution.

However, where only a single estimate or small data sample is available, basing the analysis on subjective probability estimates may be unavoidable. Large representative data samples are not typical for LCA data, where often only a few sample values are available, frequently from a different time frame from the one under consideration (e.g. historical data used to predict future conditions), or data are taken from a system similar to the one directly of interest (e.g. data from a different geographical location or technology). In addition, often only a proxy quantity is measurable (e.g. laboratory or pilot-scale data used to model a future process not yet operated at full scale). The data sample is therefore often inadequate, in which case it is necessary to use additional information to infer the parameters of the probability distribution if the full uncertainty associated with the quantity is to be represented.

4.2.2 Describing frequency distributions by statistical methods

Statistical methods are used to help select a distribution type and to estimate its parameters. The below brief summary is based on that of Morgan and Henrion (1990), and readers are referred to this text for a more detailed coverage of these topics.

Classical statistical *estimators* are concerned with predicting the parameters of the frequency distribution, so that a random sample from the predicted population will match the data sample as closely as possible. A number of well-established methods are available to derive estimators, with the method of *matching moments*, the method of *least squares*, and the method of *maximum likelihood* used the most widely. For the relatively imprecise and data deficient situations encountered in LCA modelling, the method of matching moments is judged sufficiently accurate, and because of its simplicity, is the recommended method.

The estimators for the first and second moments are given by simple formulas, which calculate estimates of the mean and variance from the data sample values. The formulas can be found in essentially all statistical textbooks (e.g. Ott, 1995), and are the "basic statistics" calculated by practically all statistical software and most spreadsheet programmes. Once the mean and variance (or higher moments, if required) have been calculated from the data sample, these are used to infer the parameters of the relevant distribution type. The parameters of some common distribution types are shown in Annex 1.

4.2.3 Selecting a distribution type

For samples above 30, the method described above provides an estimation of the parameters of a distribution to best fit a set of data observations. This assumes that the type of distribution from which the observations are generated is already known, and that the only question is the value of the mean and variance. It is therefore first necessary to select a particular distribution type, for which the parameters can then be calculated. This can most simply be achieved by plotting a histogram from the data sample and judging by eye which distribution type best fits the data (see Annex 1). A histogram presents a graphical representation of the frequency distribution of the variable, and is drawn from a frequency table of the data. The columns are drawn over a predetermined number of class intervals, and the heights of the columns are proportional to the class frequencies. An example of such a plot is given in Figure 4.2.

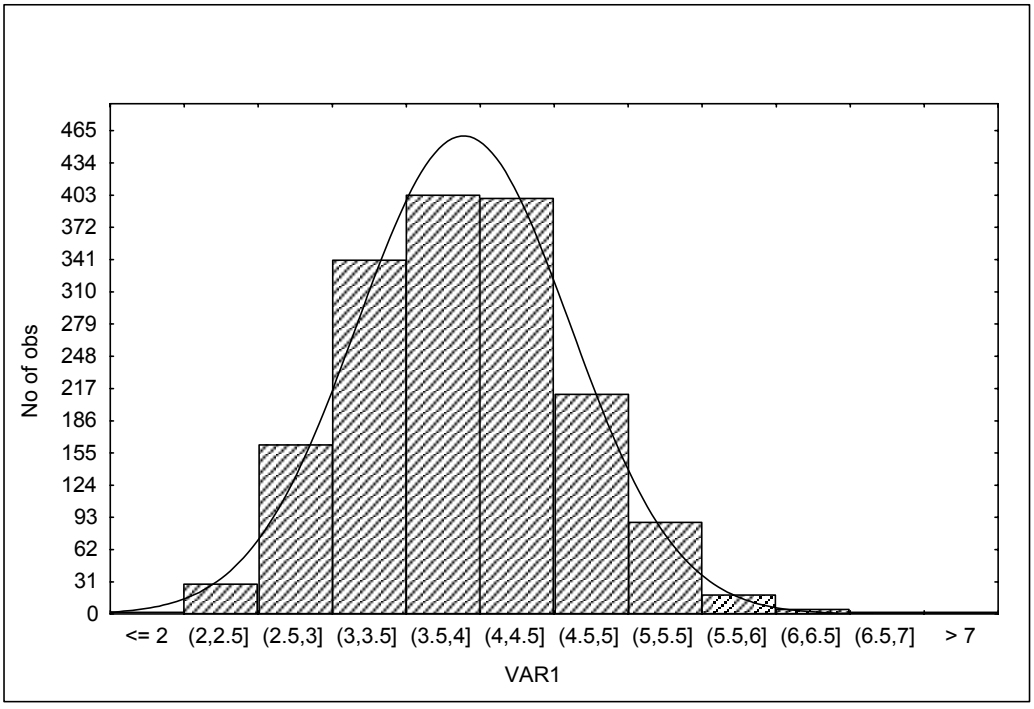


Figure 4.2 *Example of a frequency plot (histogram). The y-axis gives the number of observations (their frequency), and the x-axis the intervals into which the variable range has been divided. The chosen probability distribution type (in this case, a normal distribution) with its parameters estimated from the data sample is superimposed over the histogram to allow the goodness of fit to be judged by eye (plot from STATISTICA software, StatSoft, Inc.)*

Alternatively, statistical "goodness of fit" tests can be used to determine the probability that the data sample came from a particular type of probability distribution (e.g. the Kolmogorov-Smirnov test, or the chi-squared test). Other tests include the use of probability plots, where the degree of "straightness" gives a measure of the fit, or correlation tests. Most statistical software programs include these tests (and more besides). Their results are typically easy to calculate, but they can be difficult to interpret, and can give misleading results. Thus none of these tests can entirely substitute for a visual examination of the data using a histogram. These plots are simple to construct and easy to analyse, and are therefore the recommended method for choosing a probability distribution type.

It is only possible to use statistical methods to determine the distribution type if a large data sample is available. For smaller samples, or when only a single estimate is available, it is necessary to infer the distribution type from the knowledge of the variable properties. For example, many natural processes arise from the product of many independent random variables multiplied together, and are therefore most suited to a lognormal distribution. A short description of the features of some commonly applied distribution types is given in Annex 1, which give some guidelines as to when a particular distribution may be preferred over another.

The methods discussed above are only applicable to univariate distributions, that is, to single-dimensional quantities. If the value of one variable is probabilistically dependent on another, it may be necessary to represent their uncertainty with a multivariate distribution (e.g. if the correlated variables are to be used in a stochastic model). However multivariate distributions are complex to determine, so discrete conditional probability distributions are recommended for use in LCA models. These are simply tables that specify the value of the dependent variable given the probability of the other variable occurring. Although these result in a loss in accuracy due to discretising the distributions, they are far simpler to specify than continuous conditional probability distributions.

4.2.4 Estimation of probability

Classical statistical methods are concerned with estimating the frequency rather than the probability of a variable's occurrence (where probability is defined according to the subjectivist or Bayesian view). To obtain an estimate of a variable's full empirical uncertainty, it is necessary to augment the above methods with subjective estimates of uncertainty. Although inherently a subjective procedure, this subjectivity can be minimised by following a structured framework. An example of such a structured approach is given below. The method outlined is based on that of Weidema and Wesnæs (1996), and builds on the data quality indicator (DQI) approach to evaluating data uncertainty, as this is a well established and accepted procedure for analysing LCA data quality.

To select a necessary and sufficient set of DQIs to take into account all aspects of a variable's uncertainty requires an understanding of the various sources of empirical uncertainty. Whilst some aspects of data uncertainty are independent of the particular study in hand (i.e. those associated with its acquisition), others depend strongly on the goals of the study (i.e. the applicability of the data). Weidema (1998) defines a minimum set of DQIs to ensure all relevant aspects of data uncertainty are taken into account in a "pedigree matrix", which is shown in Table 4.1. The guidelines for "scoring" the data sample (or single estimate) minimise the subjectivity of assigning DQI scores. Another even more important feature of the pedigree matrix is that the DQIs are independent of each other. This means that the influence of each indicator is additive, which allows for an overall estimate of the uncertainty to be determined. However, the DQI scores do not represent a "quantity" of uncertainty, so it is only their influence and not the scores themselves that are additive. It is thus first necessary to extend each DQI score to a quantitative estimate of the uncertainty it introduces, before they can be combined to obtain an estimate of the overall variance. This is done by assigning

each score a *coefficient of variation*, which is a dimensionless measure used to characterise the spread of a distribution (given by the ratio of the standard deviation to the mean). Essentially this is an estimate of the variance each particular source of uncertainty adds to the overall uncertainty of the variable. The overall CV of the variable is calculated from the individual CV estimates by taking the square root of the sum of the squares of the individual coefficients.

It is generally not meaningful to pre-assign CVs to each pedigree score (i.e. to define a default matrix of CV estimates), as the score assigned will depend on the nature of the specific data sample under consideration. For example, two quantities may each score badly on technological variability, because data on each are obtained from a different process to the one under study. However the one quantity is known to exhibit high variability across technologies and is assigned a CV of 0.6, whilst for the other, variability across technologies is known to be small, and it is only assigned a variability of 0.1. Thus the CV estimates differ widely even though both had a DQI score of 4. The CV can therefore not be assigned in isolation, and a good understanding of the context is required to determine the degree of uncertainty that should be assigned to the data (e.g. knowledge of its geographical features, so that the degree of uncertainty introduced by applying data from other geographical locations can be estimated). Wherever possible, the statistical methods outlined above are used to inform the CV estimates, and the variance of the data sample used to inform the estimate of variance due to each indicator. The overall variance calculated from the combined CV estimates cannot feasibly be less than that calculated for the data sample, and in most cases it is much larger (i.e. the data sample very rarely incorporates all sources of uncertainty). The CV estimates must typically be based on "expert judgement", informed by the variance of the data sample in hand, estimates of variance across relevant literature data samples, and estimates of variance in data samples of similar or related quantities.

The revised estimate of the variance is then used, together with the estimate of the mean, to calculate the parameters of the distribution. As before, a relevant distribution type first has to be selected. In some cases, in addition to adjusting the estimate of the variance, the analysis may require that the mean or the distribution type also be adjusted. For example, in adjusting the parameters to account for the uncertainty of applying current data to future conditions, it may be decided to adjust the data sample of a particular quantity to account for significant increases in efficiency that are predicted to occur in that quantity over the next few years. In this case, in addition to a high temporal correlation DQI score, and thus a high CV estimate, a 10% increase in the mean is also estimated.

Table 4.1 *Pedigree matrix with guidelines for scoring the DQIs (Weidema, 1998)*

Indicator Score	**1**	**2**	**3**	**4**	**5**
Indicators which are independent of the study in which the data are applied:					
Reliability of source	Verified data based on measurements	Verified data partly based on assumptions or non-verified data based on measurements	Non-verified data partly based on assumptions	Qualified estimate (e.g. by industrial expert)	Non-qualified estimate or unknown origin
Completeness	Representative data from a sufficient sample of sites over an adequate period to even out normal fluctuations	Representative data from a smaller number of sites but for adequate periods	Representative data from an adequate number of sites but from shorter periods	Representative data but from a smaller number of sites and shorter periods or incomplete data from an adequate number of sites and periods	Representativeness unknown or incomplete data from a smaller number of sites and/or from shorter periods
Indicators relating to the technological and natural production conditions under which the data are valid, and therefore dependent on the data quality goals for the study in which the data are applied:					
Temporal correlation	Less than 3 years of difference to year of study	Less than 6 years of difference	Less than 10 years of difference	Less than 15 years of difference	Age of data unknown or more than 15 years of difference
Geographical correlation	Data from area under study	Average data from larger area in which the area under study is included	Data from area with similar production conditions	Data from area with slightly similar production conditions	Data from unknown area or area with very different production conditions
Further technological correlation	Data from enterprises, processes and materials under study	Data from processes and material under study but from different enterprises	Data from processes and materials under study but from different technology	Data on related processes or materials but from same technology	Unknown technology or data on related processes or materials, but from different technology

4.2.5 Fluctuations in time

The variability captured in data samples is strongly dependent on the time intervals over which the data are collected. Typically, the longer the time interval, the more the variability will have been "dampened" out (e.g. annual variability is typically less than monthly or daily variability). Thus it is important to decide over which time interval data variability should be represented in a probability distribution. For life cycle inventory modelling, where studies typically have medium to long time frames (e.g. a product or technology assessed over its entire life time), annual variability is generally sufficient. Fluctuations over shorter time intervals, such as differences in industrial emissions on weekdays versus weekends or even short calamitous emissions, are not usually taken into account. Such shorter time intervals (daily or hourly) are probably more applicable to impact models. For the probability distributions of the various quantities to be comparable, the choice of time interval must be consistent across these quantities. Not incorporating temporal variation in the inventory analysis also has consequences for the possibilities to incorporate temporal variability in the impact assessment (Huijbregts, 1998).

4.2.6 Bias

Bias gives rise to systematic errors in the data set, where a systematic error is the difference between the true value of the variable and the value to which the mean of the measurements converges as increasing numbers of measurements are taken. Systematic errors may be due to e.g. imprecise calibration, faulty reading of scales or inaccurate assumptions used to infer the actual quantity from the observable measurements. Bias cannot be reduced by collecting more data. Whilst large data samples may be able to accurately represent data frequency or variability, they tell us nothing of the systematic errors that may be present in the data set. Other than a careful design of data collection protocols, it is usually not possible to eliminate all sources of bias. Whilst it is possible to include (under the "reliability of source" DQI) a subjective estimate of the uncertainty introduced by systematic errors, it is extremely difficult to estimate their possible magnitude, as the sources of bias are typically unknown or merely suspected (assuming all known sources of bias have been adjusted for in the data collection). It has been shown that there is a consistent tendency to underestimate the systematic error in almost all measurements of physical quantities (Morgan and Henrion, 1990). Thus, at best, all that can be done is to bear this in mind when making subjective estimates of the uncertainty that bias introduces.

An additional source of bias is that introduced by the statistical estimators, i.e. there is some degree of systematic error between the actual distribution parameter values and those estimated from the data sample. The formulas of the estimators have been corrected for known biases wherever these are known to occur, although these may still be significant for small data samples, or where a large number of outliers (extreme values from the edges of the distribution) are present in the sample.

4.2.7 Inhomogeneous data

When faced with a particular data sample, it may be possible that its elements come from two different parent populations (i.e. that it is not homogenous). It is therefore always recommended to first draw up a frequency table of the data, which can then be plotted in a histogram (an example of which is shown in Figure 4.2). This allows a good "first look" at the data, and allows a qualitative examination of its various features. For example, a bimodal distribution (i.e. two distinct peaks in the distribution) may suggest that the sample is not homogeneous. This should alert the analyst to a closer examination of the causes of the variability underlying the particular quantity, and should provide the solution as to whether the variable is genuinely characterised by a bimodal distribution, or whether two distinct sub-populations can be separated out.

In LCA data, inhomogeneous data samples may arise when different quantities with similar properties have been aggregated into a single category (e.g. Hg and Pb combined into a "heavy metals" category). To obtain meaningful representations of probability it is recommended that probability distributions always be based on the lowest level of LCA data (i.e. on the measured data before any aggregation has taken place). Probability distributions on higher levels of data aggregation should rather be obtained by combining the distributions of the lower level data (e.g. through the use of stochastic models), than by attempting to define probability distributions for inhomogeneous quantities. This is also recommended because it is much less meaningful to make subjective estimates of uncertainty on aggregated quantities, than on measured data samples.

4.2.8 Recommended practice

The following is recommended practice when forming a probability distribution. Arguably more accurate methods are available, however much more complicated and therefore easier to abuse (even if easy to calculate with statistical software). A straightforward procedure is deemed most appropriate in an LCA context, and the following steps are therefore recommended.

Quantities with data samples in excess of 30 elements:

1. Draw up a frequency table and histogram plot of the sample to obtain an overview of the data and to check for possible problems in the data sample (e.g. inhomogeneities)
2. Select the most applicable distribution type using the histogram plot.
3. Calculate an estimate of the mean and variance from the data sample.
4. Calculate the parameters of the chosen distribution from the estimates of the mean and variance.
5. Plot the frequency distribution using the parameters calculated. If the data sample is deemed sufficiently representative, the analysis may end at this step.
6. Estimate the additional uncertainty present but not represented in the data sample by making subjective estimates of the additional sources of variance. This should be done using a structured, reproducible approach, e.g. the pedigree matrix of Weidema (1998).
7. Re-calculate the parameters of the distribution with the adjusted estimate of the variance.
8. Plot the probability distribution using the parameters calculated.

Quantities with few data values, or a single estimate:

1. Estimate the uncertainty present by making subjective estimates of the magnitude of the possible sources of variance. This should be done using a structured, reproducible approach, e.g. the pedigree matrix of Weidema (1998), and informed by whatever literature data or data on related quantities that are available.
2. Select an appropriate distribution type, according to the level of information known about the variable (see guidelines in Annex 1).
3. Calculate the parameters of the chosen distribution from the estimates of the variance and the mean (which is the single estimate value, or the most likely value for small data samples).
4. Plot the probability distribution using the parameters calculated.

4.2.9 Checklist of things that can go wrong

- Analysis based on an inadequate data sample (not representative, too small or inhomogeneous).
- Inappropriate distribution type chosen.
- Mean and variance estimated from an insufficient sample (too few values to give accurate estimates).
- Conditional probabilities between variables ignored.
- Additional sources of uncertainty (e.g. biases) not accounted for.
- Estimates of the additional variance not based on a structured procedure.
- Inadequate estimates of the additional variance (typically underestimates).

4.3 Documentation of probability distributions

4.3.1 Recommended practice

When documenting the formation of probability distributions the following should be included:

- The adequacy of the sample (i.e. its size and whether it can be considered to be representative of the population).
- The distribution type and its main defining parameters.
- The reasons for selecting the particular distribution type.
- The method by which the parameters of the distribution are estimated.
- Whether the distribution is an objective frequency distribution or whether it includes subjective probability estimates, and in the latter case the method by which the frequency distribution parameters are updated with the subjective probability estimates.
- The method by which subjective probability estimates are made (when applicable).

4.3.2 Checklist of things that can go wrong

- Documentation incomplete (relevant information not included for each probability distribution)
- Description of methods unclear (too brief or lost in too much statistical detail)

4.4 Validation

A frequency distribution is fairly easy to validate, in that all that needs to be checked is that the statistical estimates are based on a sufficiently large, representative data sample and that the sample generated from the distribution is consistent with the data sample from which it is inferred, e.g. the mean and selected fractiles of the original data sample and a sample generated from the distribution should closely resemble each other. Any software able to do stochastic modelling (e.g. Analytica, Crystal Ball etc.) is able to generate a random sample from a given distribution.

Validating a probability distribution based on subjective estimation is more difficult, as in this case the range covered by the variable will be greater than that covered by the sample on which it is based. Again, these are best validated by looking at a random sample generated from the probability distribution, but in this case it should be checked that the sample values fall within the feasibility range of the quantity. For example, many quantities encountered in LCA are constrained to be positive, thus negative sample values warn of an invalid distribution. Similarly, many variables have upper and/or lower bounds that the distribution should not violate. In addition to checking the feasibility of the sample generated, it should be determined that a structured, reproducible approach was followed to elicit subjective probability estimates, and that adequate supporting information was used to inform the estimates.

4.4.1 Recommended practice

The validation of a frequency distribution should include:

- The size of the sample on which the distribution is based.
- The statistical methods used to infer the distribution.
- That the distribution produces a sample consistent with that on which its parameters are based.

In addition, the validation of a probability distribution with subjective probability estimates should include:

- The procedure followed to elicit the probability estimates.
- The information used to support the estimates.
- That the distribution produces a sample within the feasibility range of the quantity.

The names and roles of people involved in the validation procedure(s) should be documented.

4.4.2 Checklist of things that can go wrong

- Insufficient documentation to allow a meaningful validation of the methods and practices followed
- Insufficient knowledge on the properties of the variable to determine whether the distribution represents a feasible operating range

4.5 Communication

Essential to the good communication of probability distribution is to impart a clear understanding of what has been included in their determination. A common fault with probability distributions is that they give a false sense of accuracy and completeness, although they are only as good as the estimates that underpin them. Typically not well communicated is that a well characterised frequency distribution (i.e. one based on many measurements) does not necessarily mean that it captures the full range of uncertainty in the variable, since well characterised data can introduce significant uncertainty when applied in a different context to that for which the data sample is representative. For this reason it is necessary to clearly communicate whether frequency or probability distributions have been formed. In the case of frequency distributions, it needs to be communicated clearly what situation the data sample are regarded as representative of. In the case of probability distributions where estimates have been made of the potential effect of non-represented sources of variability, the inherently subjective nature of this procedure should be communicated.

4.5.1 Issues of terminology

It is not possible to communicate probability distributions without some use of statistical terminology. However, this can be kept to a minimum by using graphical methods and more intuitively understandable concepts, such as confidence intervals.

4.5.2 Recommended practice

Communication of probability distributions should as a minimum include:

- Information on the distribution type and its parameters (e.g. the mean and variance), and/or confidence intervals between selected fractiles, e.g. “50% confidence that the variable lies between 3 and 6.5” (see Figure 4.1).
- A statement of the degree of confidence that can be held in the distribution, i.e. there should be a clear understanding of whether the distribution is based purely on objective statistical methods, or whether subjective judgement played a role.

Whenever possible a graphical communication (plot of the PDF and/or CDF) and confidence intervals are preferred for communication, since these do not require much statistical knowledge.

4.5.3 Checklist of things that can go wrong

- Inadequate parameters
- Communication couched in confusing statistical terminology
- Probability distributions conveying a false sense of accuracy and completeness (no information whether the distributions are based on a representative data sample, and/or whether subjective estimates are included)

Chapter 5. Describing a process

A process description brings together all the individual data related to the functioning of a process, expressed by a functional unit or a reference flow. This chapter deals with the linking of such individual data (the collection of which was the topic of the previous chapters) and describes how this is done.

5.1 Definitions

A *process* is a set of interrelated or interacting activities which transforms inputs into outputs (ISO 9000:2000). Processes may be individual process steps or production lines within a site, entire plants, markets, transports and transportation routes, and complex composite systems such as production systems for specific products from cradle to gate.

Inputs and outputs are materials or energy which enters and leaves a unit process, respectively. Examples of inputs are natural resources, raw materials, energy, ancillary material etc. Examples of outputs are products and by-products, emissions to air, water and soil, waste etc. However, an inventory of inputs and outputs may also include environmental aspects not directly related to the inputs and outputs. (ISO14040).

A *unit process* is the smallest portion of a product system for which data are collected when performing a life cycle assessment (ISO14040). This implies that it is the detail of data collection that determines whether a process is a unit process.

A *product system* is a collection of materially and energetically connected unit processes which performs one or more defined functions (ISO14040). The connections between the unit processes are not necessarily directly physical, but may also be indirect through changes in supply and demand on the market that connects the processes. A market may in itself be described as a process.

A *functional unit* is the quantified performance of a product system for use as a reference unit in an LCA study (ISO 14040). The function of a process may be the production, supply, use or waste handling of a certain product, and the functional unit may have been chosen as for example 1 kg of that product.

A *Flow diagram* is a graphic representation of the interlinked unit processes comprising the product system (Guinèe et al. 2001).

A system boundary is the interface between a product system and the environment or other product system (ISO14040).

Allocation is the partitioning of the inputs or outputs of a unit process to the product system under study (ISO14040).

System expansion is the adding to a multi-functional product system the processes necessary to balance out the undesired output of co-products, in order for the product system to produce only the product of interest while keeping the output of the other co-products constant.

5.2 Action

5.2.1 Selecting the level of process modelling

The level of detail in process modelling (e.g. whether a production process is described at the level of individual machines or production lines, at plant level or at sector level) depends on one hand on the needs of the user and on the other hand on the data availability.

In the context of LCA, process data are required with the aim of suggesting improvements to the investigated processes and systems. Thus, the degree of detail required in process data depends on the level at which the investigated processes can be influenced. If controls can only be made at plant level, detailed information on machine level is unnecessary. However, when the end user or final application of the data is unknown, the largest possible degree of detail should be strived for. This applies both to the choice of detail in process modelling and to the issue of transparency in aggregated data (see Chapter 6).

The individual data to describe a process (e.g. data on specific inputs and outputs) may be available at different levels (e.g. data on energy use may be available at production line detail, VOC emissions only available at plant level, while other emissions are only calculated at sector level).

Following the principle of "the lowest denominator," the least detailed data determine the level of detail possible for the entire process. However, data with a low degree of detail may be regarded as representative for the more specific situation. This assumption will be least problematic in a homogeneous population.

A more serious problem arises when data for different properties are representing incongruent system boundaries, e.g. when energy use is measured for machine A+B together and VOC emissions for machine B+C. Assuming that separate energy data for machine C and emission data for machine A are available, two strategies can be followed to obtain a consistent process description:

1. To describe machine A, B and C separately, the originally measured data are allocated, separating from the original measurements that part of the energy and emissions that belong to machine B.
2. To form a process description for machines A+B+C together, the energy data for machine C and emission data for machine A are added to the original measurements.

The choice between strategy 1 and 2 depends on the need for separate process descriptions for each machine and the additional work and uncertainty involved in performing the allocation.

5.2.2 Defining system boundaries in time and space

System boundaries can be divided into:

- system boundaries towards the environment system (nature),
- system boundaries in time,
- geographical system boundaries, and
- system boundaries towards other processes, distinguishing between processes that are relevant and irrelevant to the system, i.e. boundaries between the life-cycle system of the studied product and the life-cycle systems of other products.

System boundaries towards the environmental system: Only human activities are described as processes. The environmental system and its mechanisms is what would also exist if the human activities were non-existing, i.e. the natural background mechanisms. It is important to document whether a specific process includes all human activities that deviate from the natural background, e.g. emission control activities, the processes that occur in a landfill, and the changes that result from a specific land use.

System boundaries in time: System boundaries in time describe different time related aspects of the studied process, such as the time period during which the description of the process can be considered valid. Examples of such descriptions are an estimation of the "best before" date for the process, which can be determined and described by the technology level of the process combined with knowledge on investment plans or technology developments that will have a significant effect on the process and consequently the data. With such information it is possible to form an opinion on the period for which the data will be representative. Also, a process may only be representative of a specific season or time of day, as e.g. peak hours of electricity use, where a different electricity generation pattern applies compared to the non-peak hours.

Geographical system boundaries: The geographical extension and/or geographical limitations of the studied process or system. The motives for choosing the geographical system boundaries should be described. Also, if sub-processes that are included in the process operate in different geographical places, a description should be given to where geographically each included subsystem operates.

System boundaries towards other processes: This point includes rationales for exclusions of sub-processes and the criteria for identifying flows to and from other processes.

Defining system boundaries based on data availability is an unacceptable method because it is not repeatable, has no scientific justification, and is not rigorous.

Reflecting the iterative nature of LCA, decision regarding the data to be included shall be based on a sensitivity analysis to determine their significance, thereby verifying the initial analysis.
The initial product system boundaries shall be revised as appropriate in accordance with the cut-off established. (ISO 14041).

To be efficient and provide a repeatable and rigorous basis for comparison between systems, a system boundaries selection method must (Raynolds et al. 2000):

1. be quantitative (i.e. qualitative judgment rules are not adequate),
2. not require the quantification of environmental outputs from every unit process in the life-cycle system before system boundary selection,
3. be simple and allow for streamlining (i.e. allow for different degrees of rigor due to varying availability of time and resources to make decisions),
4. consider the significance of inputs and outputs relative to the system as a whole, not only to an individual unit process,
5. provide the ability to define measurable levels of system completeness.

5.2.3 Allocation

When data relates to an input or output that belongs more than one process (and thus to more than one product), or the process that we have input data for is larger than the process we wish to describe (and thus include more product than required), a choice must be made, either to

- leave the data as they are, but report the lacking fit in the available data, or
- perform an allocation to estimate and/or calculate the part of the available data that should be included in the process description.

Many different allocation procedures have been suggested. The choice of procedure can have a decisive impact on the results of an LCI. The ISO 14041 requires the following procedure be used for allocation in multifunction processes:

- Allocation should be avoided, wherever possible, either through division of the process into relevant sub-processes, and collection of separate data for each sub-process, or through expansion of the process, balancing out the undesired functions so that the expanded process deliver only the desired functions.
- Where allocation cannot be avoided, the allocation should reflect the physical relationships between the environmental burdens and the functions, i.e., how the burdens are changed by quantitative changes in the functions delivered by the system.

Both of these procedures are based on the principle that what is to be reported for a process is the effect of changing the output of a specified product while keeping the other co-product outputs constant. When these effects can be isolated within a process, subdivision of the process is straightforward. When the different product outputs can be varied independently, it is often possible to identify a physical relationship that determines the changes in the process, and which can therefore be used as allocation key. When the different product outputs are joint, and thus not independently variable, the additional output of the other co-products can only be kept constant by expanding the process to include processes that will balance out this output (the procedure know as system expansion).

The estimates and/or calculations made as part of the allocation procedures (or to avoid allocation) should be reported in the same way as any other estimate or calculation (see Chapter 3).

5.2.4 Estimating system boundaries

When information related to the system boundaries is missing, it may be necessary to estimate the boundaries.

For example, when the timing of measurements are missing, the timing must be estimated with a best estimate of equal intervals between measurements and an error estimate based on a less ideal temporal distribution of measurements.

A similar example of missing geographical information is a measurement of BOD and Total-N in two wastewater streams from a factory, with 10 measurements for each waste stream. To obtain a total value, the measured values for each waste stream would need to be multiplied by the wastewater volume in each stream. However, only the total wastewater volume was known; no information was available on its distribution between the two waste streams. In this case, a best estimate was made by assuming equal wastewater volume in the two flows and error estimates made by assuming 90% of the volume in each of the two streams.

5.2.5 The environmental properties to include in the process description

Some properties may be excluded from the process description when not considered environmentally relevant or significant. This choice may pragmatically be based on legislature for the studied process. In practice, data are typically only available for those properties that are of economic interest or where there is a requirement to report to the authorities. Some properties that may have a potential influence on the environment may therefore be excluded when describing a process. Care should be taken that this choice is made consciously and reported transparently.

As a general rule, sum parameters should be only used when the group as such is measured, e.g. “total VOC” with an FID detector, or “total C_xH_y chloro” via adsorption and subsequent halogen determination. Individually measured emission parameters (e.g. tetrachloroethene) should *not* be hidden in group or sum parameters such as “C_xH_y chloro", "C_xH_y halogenated", "NMVOC", or "VOC” (de Beaufort-Langeveld & Bretz 2001).

5.2.6 Overlapping property data

If the same substance is both determined as individual molecules and a sum parameter for the same process, overlap occurs inevitably. This overlap can only be overcome, when a *corrected* sum parameter are reported (e.g. total VOC *minus* individually measured values). Preferably, this should be made explicit, with the annotation: "sum parameter corrected for individually reported species" (de Beaufort-Langeveld & Bretz 2001).

5.2.7 Correlations between property data

Some emissions may be correlated, so that a change in one emission must necessarily lead to a proportional or reverse proportional change in the other(s). For example, the different emissions of nitrogen from an agricultural process (N_2, N_2O, NH_4, NO_3 to surface-water and NO_3 to ground-water) are correlated, since these emissions must add up to 100% of the nitrogen surplus as calculated from a mass balance. In cases like this, it is not meaningful to report emissions separately, since this could lead to over- or underestimates.

5.2.8 Linking processes by their intermediate flows

The processes to include in a *product system* are identified by starting from a specific product or service. From this starting point, the matter and energy paths are followed, upstream to the cradle of the raw material and downstream to waste treatment. The paths between the unit processes are not necessarily directly physical, but may also be indirect through changes in supply and demand on the market that connects the processes. The result is a *system model* that is not merely a *collection* of processes and subsystems, but rather a *network* of processes. Each process is incorporated into the *system model* on the basis of its causal relationship with the previously incorporated processes, in accordance with the principles described in Sections 5.2.1 to 5.2.3. A network is built by *connecting* intermediate flows between processes. Since resource use, emissions, wastes, and by-products are also flows, there will be a number of flows not connected to any other flows of other internal processes. These *free flows* represent resource use, use of raw materials, emissions, waste and by-products out of or into the overall process.

This network of processes and flows may be represented by a *flow diagram*. A flow diagram provides an outline of all the major unit processes to be modelled, including their interrelationships. It may also illustrate which processes are included and which are excluded, i.e. the system boundaries. Drafting flow diagrams can be done at different levels of complexity, also in relation to level of sophistication (detailed, simplified), the most complete being a diagram in terms of unaggregated unit processes. Environmental interventions are often omitted from the flow diagram, because the main function of the flow diagram is to illustrate the structure of the product system and the relationships between the unit processes.

For building an appropriate flow diagram, these guidelines apply (Guinèe et al. 2001):

- Report process flow diagrams describing the complete process under study.
- Large numbers of unit processes and complex interrelationships can be managed by zooming in from a general flow diagram to the underlying unit processes in partial flow diagrams.

- Use boxes to represent processes, and arrows to represent product flows[8]. It is especially important to indicate clearly which of the boxes refers to a unit process and which to a partial flow diagram with underlying unit processes.
- Text labels in boxes should clearly indicate the names of the unit processes, and text labels attached to arrows show the names of the product flows.
- Include system boundaries, major input, products and co-products in the process flow diagrams. The process flow diagrams should include the main production sequence, ancillary materials and energy/fuel production.
- Check the flow diagram for completeness and consistency, internally and compared to the data sheets.

5.2.9 Recommended practice

The largest possible degree of detail should be strived for, unless the data are to be used for a specific application in which a lower degree of resolution is adequate.

Allocations should only be made when necessary and should follow the ISO 14041 procedure. When the production volume of the different functions cannot be independently changed, system expansion, or an approximation thereof, is the only approach that gives comprehensive information of the environmental consequences of an action. When the effect of the system expansion is significant but the uncertainties very large, different scenarios could be used to describe the uncertainty. As a minimum, it should be clearly stated that the system expansion constitutes a source of uncertainty that should be taken into account.

Estimates and calculations in relation to system boundaries and allocation procedures should be reported in the same way as any other estimate or calculation (see Chapter 3).

If the same substance is determined as both individual molecules and a sum parameter for the same process, the sum parameter should be corrected for the individually reported values.

[8] Often, arrows have been used to indicate the direction of physical rather than economic flows. For instance, an arrow from consumption to waste treatment has been used to indicate the flow of disposed products, and not the other way around to indicate the flow of the waste treatment service. However, it is recommended to indicate service-providing processes as delivering a product input (the service) to the serviced process, so that e.g. a cleaning process does not absorb dirt (arrow from the dirt-generating process to the cleaning process), but provides a cleaning service (arrow from the cleaning process to the dirt-generating process).

5.2.10 Checklist of things that can go wrong

- Cut-offs based on mass or energy inputs, ignoring service inputs and environmental significance
- Cut-offs based on inputs to each unit process, rather than on the cumulated contribution
- Cut-offs based on availability of data rather than significance
- Unnecessary or inappropriate allocation procedures
- Undocumented introduction of additional estimates and calculations in relation to system boundaries and allocation procedures.
- Overlapping or inconsistent property data

5.3 Documentation of a process description

ISO 14048 specifies how to document processes for use in LCA. The information is structured under the following main headings:

- A *process description*
- Data on *inputs and outputs*
- The documentation of *modelling and validation*
- *Administrative information*

5.3.1 Recommended practice

As a minimum, the process description should include:

- A descriptive *name* (e.g. 'Power plant, combined heat and power' or 'Transport, long distance, heavy truck').
- The *quantitative reference* (functional unit or reference flow) that the input and output data refer to.
- The *valid time span* to which the model of the process applies. Unless projections or other forecasts have been applied, the valid time span is identical to the time of the primary data collection. Limitations for the validity in time may be set by e.g. future technology shifts, planned measurement improvements, or specific seasons.
- The *valid geography* describing the geographical area or location for which the process and data is valid. This is identical to the area or location of the data collection, unless extrapolations from other areas have been performed.
- The *technical scope* of the process (in terms of specified operations or transformations included in the data, or simply as "from ... to ..." stating the first and the last operation of a chain, e.g. in a transport route, on a site, or in a production line).
- The detailed *technical content and functionality* of the process.
- Documentation of the *data acquisition* (data collection and treatment) procedures at the process level.

Also recommended is a *technology picture,* as a very efficient way to show what is included and what is not in the system boundaries of the process, see Figure 5.1.

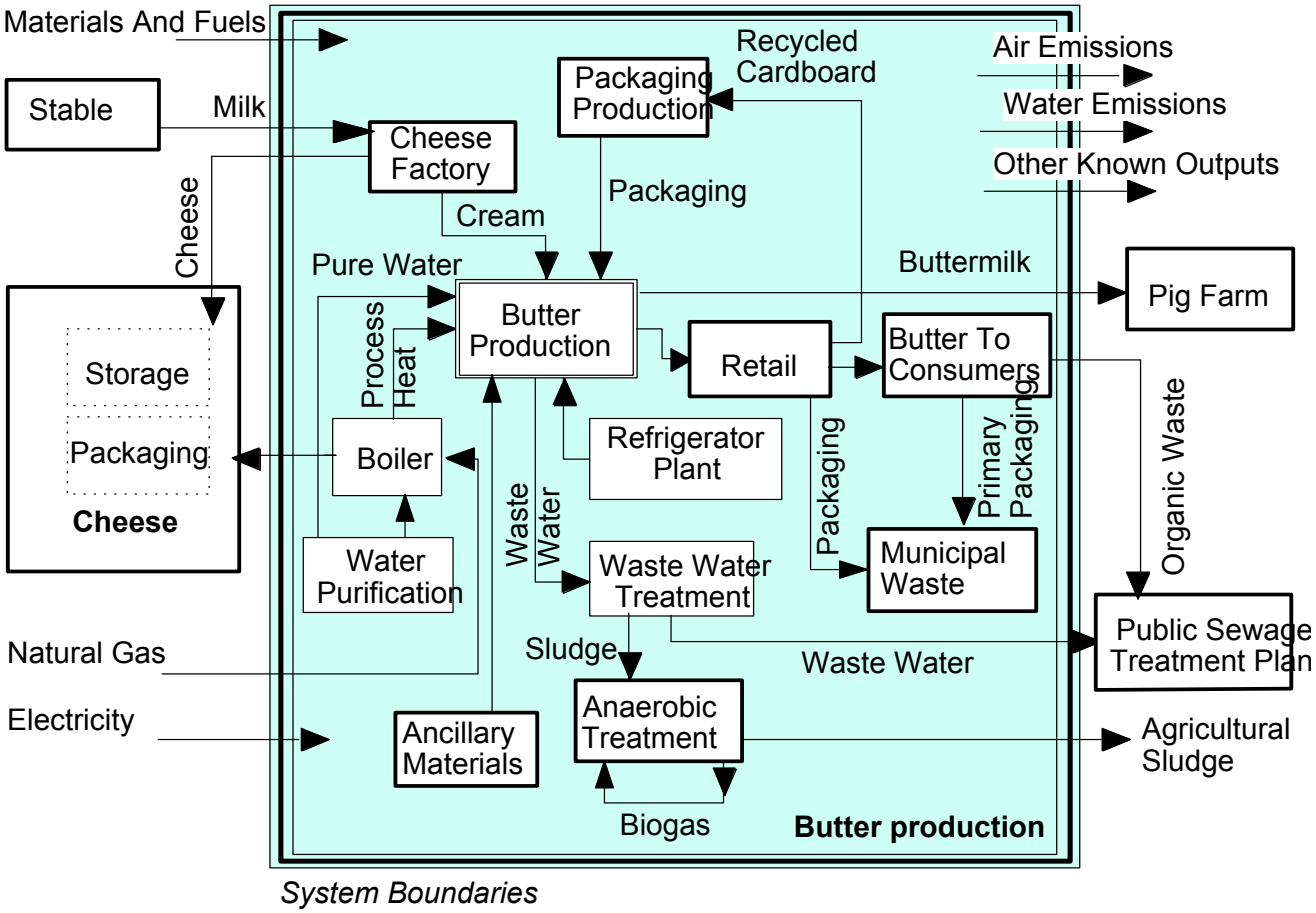


Figure 5.1 *Example of technology picture for the butter production*

The documentation of the inputs and outputs data is based on the documentation described in Sections 3.3 and 4.3. The relevant parts of the original documentation should be repeated or referenced as part of the process documentation, together with any additional estimates and calculations made to establish consistent system boundaries.

The documentation of modelling and validation should as a minimum include:

- The *criteria for excluding elementary flows*, i.e. the criteria used in the choice of properties that are considered environmentally relevant and vice versa.
- The *criteria for excluding intermediate product flows*.

When applicable, an *allocation explanation* or a *process expansion* explanation should be provided. Also, any applied *validation procedure* should be documented in order to avoid duplication of work.

As a minimum, the administrative information should include:

- Identification of the *data commissioner* (the person or organisation requesting and funding the data collection)*, data generator* (the person or organisation responsible for the modelling of the process and the compilation or the updating of the data), and *data documentor* (the person responsible for entering the data into the current data documentation format).
- The *date completed*, i.e. the last date at which the data was entered, edited or updated.
- Identification of the *copyright* holder.

5.3.2 Checklist of things that can go wrong

- Use of local file names without reference to ISO 14048
- Leaving out information on the temporal, geographical or technological validity of the process data
- Leaving out information on the data acquisition procedures
- Lacking documentation of excluded properties or missing data
- Insufficient administrative information to trace back the data source and judge the data applicability

5.4 Validation

Process descriptions should be validated prior to further aggregation and impact assessment.

Validation should be done by professionals with experience of the process in question.

Validation may include:

- Independent remodelling or recalculation of the process model.
- Identification of data gaps through logical checks, such as mass balances (total or broken down at elementary level) of inputs and outputs.
- Benchmarking the process description against descriptions for similar processes.
- Cross-checking with data obtained through different sources and/or collection methods, e.g. comparing site specific plant data with data from statistical input-output tables or theoretical models and vice versa.
- Proof-reading.

The names and roles of people involved in the validation procedure(s) should be documented.

5.5 Communication

The objectives and principles for communication of processes are the same as for individual data (see Section 3.5). Data about processes will be used either for further analytical treatment, e.g. aggregation into averages or aggregates of linked processes, or for final reporting of environmental performance. Thus, to prepare for further analytical treatment, the communication should be unambiguous, reliable, complete and referenced, cf. Section 5.3.

Special care should be taken in defining the scope of applicability and the representativeness of data, so as to ensure that the data can be used appropriately. A false sense of completeness or detail can be conveyed when unit processes known to be relatively insignificant are included. The audience of the LCA study might interpret this as the entire LCA having been investigated to a similar level of detail.

All terminology should refer to international standards, especially ISO 14048.

If the communication takes place in a public context, data suppliers should always tell whether process descriptions have been independently reviewed or peer-reviewed prior to disclosure.

5.5.1 Checklist of things that can go wrong

Communication can be unsatisfactory when one or several of the following situations occur:

- When the data supplier misunderstands the communication context, i.e. the needs of the target group and what the process description is to be used for.
- When the data supplier provides incomplete or ambiguous information on the process and the conditions of the process modelling.
- When the data recipient misunderstands the original conditions of the process description and/or adds unfounded interpretations.

The above three points may to some extent be seen as a cause-effect chain, one problem leading to the next.

Chapter 6. Aggregating models of processes

6.1 Definitions

Aggregation

Aggregation is the removal of internal boundaries.

A process that is composed of other (sub-)processes can be presented as both:

- unaggregated (i.e. with all its sub-processes intact), by providing information on each sub-process as well as information on how the sub-processes are linked, and
- aggregated (i.e. with loss of information on the sub-processes), by adding the information for each property from each sub-process into an aggregated value for the aggregated process, in such a way that the information for each sub-processes cannot be reconstructed (i.e. the information is lost on how the values of the aggregated properties shall be subdivided over the sub-processes).

6.2 Action

6.2.1 How to aggregate models of processes

Description of unaggregated processes was treated in Chapter 5. As mentioned there, an unaggregated process can be both a unit process and a complex composite system, such as a production system for a specific product from cradle to gate. Unaggregated processes are simply described by providing information on each unit process as well as information on how these unit processes are linked.

Aggregation of processes can be done for different purposes, for example, when compiling the product system for a life cycle inventory (LCI) or when compiling different types of averages of processes. These compiled aggregates are new models of processes, with specific purpose, properties, functions and scopes.

As examples, two important types of aggregates are discussed; averages and aggregation of linked processes.

Averaging: Industrial averages are compiled by averaging different but physically similar models of processes into a new system model. Industrial averages may describe a specific sector, market or product. Averaging is the mathematical aggregation of similar models of processes into a new virtual model, representing an average process. The modelling includes specification of the scope and requirements on the included processes and the mathematical summation of frequency functions of flows. Mathematical treatment for formation of averages requires each of the summed models to be equal with regard to a set of specified physical and mathematical aspects (Carlson & Pålsson 2001). Examples of averages are:

- *Interval averages - the same process:* Interval averages are compiled from process models describing the same physical process during different intervals, for example, time intervals or product volume intervals. The resulting process model describes a specific time or product volume interval. An example of such an average is a corporate annual environmental report describing the environmental performance of a plant.
- *Industrial averages - different processes:* Industrial averages are compiled by averaging different but physically similar models of processes into a new process model. Industrial averages may describe a specific sector, market or product.

Aggregation of linked processes: The flows of an aggregated process are the sum of the free flows of its internal processes (Carlson et al. 1998). An aggregated result of environmental inputs and outputs for linked processes is produced by quantitatively scaling each process to the references flow following from the functional unit. In this procedure internal loops may be encountered, for instance when the extraction of coal requires electricity and the production of coal requires coal. In a matrix method, the calculation takes this accurately into account, while in a sequential method, the solution is only an approximation, and the accuracy depends on the number of times the loop is passed. Depending on the software used, the number of passes may be one, a fixed number (e.g. 5), or it may depend on the size that each additional pass adds to the results.

6.2.2 The limits of meaningful aggregation

To be meaningful, *averaging* must take place over a reasonably homogeneous population, e.g. an average of similar production lines within the same plant, or an average of plants with approximately the same scale, technology, capacity utilisation, and effectiveness of (emission) control, within the same administrative region (ensuring homogeneous cost structure and legislative regime) and under approximately the same natural production conditions (e.g. climate and raw material quality and availability).

When aggregating the flows from different interlinked processes, information is lost on the specific conditions relating to each individual flows. Whether this loss of information is acceptable, depends on the application of the aggregated values. For example, the loss of geographical or technological detail may be quite acceptable for substances such as CO_2 that have mainly global impacts, while it may be quite unacceptable for substances with effects that depend on local conditions, such as is the case for many toxic substances. Sum parameters such as VOC and AOX should never be aggregated over different processes, e.g. solvent discharges and bleaching processes, since their composition depend on the nature of the process (de Beaufort-Langeveld & Bretz 2001).

6.2.3 Completeness and consistency

Completeness of a system of processes is ensured by applying the ISO criteria of cumulated mass, energy and environmental relevance, justifying each cut-off in comparison to the cumulated inventory for each property (or environmental effect, when an impact assessment method has been chosen). The same principle can be applied to economic value, so that each input is compared by economic value to the total economic value of the functional unit. The advantage of this procedure is that it is quantifiable and repeatable and provides results that are comparable across different systems.

When aggregating process descriptions, the underlying assumptions for each of the original processes must be consistent, in order to make a meaningful and adequate description of the aggregated process. Thus, it is not meaningful to aggregate processes with e.g. different time horizons, unless precautions have been made to compensate for these differences prior to the aggregation.

The links in a system of processes are consistent when:

- all processes are connected to one another, i.e. there are no unconnected intermediate flows,
- intermediate flows that are shared among several processes sum to 100%,
- the system is not overdetermined or underdetermined (that each share of an intermediate output flow is connected to one and only one intermediate input flow and vice versa).

6.2.4 Recommended practice

Before averaging, consider whether the population in question is adequately homogenous to provide a meaningful aggregate.

Before aggregating flows across different interlinked processes:

- Check for completeness and consistency of the processes in question.
- Consider whether the implied loss of process specific information (e.g. geographical or technological detail) is acceptable for the intended application. Sum parameters such as VOC and AOX should never be aggregated over dissimilar processes.

When producing an aggregated result of environmental inputs and outputs for interlinked processes, it is recommended:

- to use matrix-based calculations, because they easily allow for recursive relationships,
- to maintain the underlying unaggregated information, that allows the data to be recalculated, partially revised and/or used in other contexts.

If it is not possible to use matrix-based calculations, check whether there are any loops within the system described, also in the data from literature and reference databases. If deficits are identified, these can be adjusted directly by using a different calculation method. If this is not possible for practical reasons, assess the influence that ignoring loops might have on each of the system studied, by means of a sensitivity analysis (Guinèe et al. 2001).

6.2.5 Checklist of things that can go wrong

- Averages of inhomogeneous populations, e.g. from plants of widely different scale, technology, capacity utilisation, or effectiveness of (emission) control, from different administrative regions or under different natural production conditions.
- Loss of relevant information on the specific conditions relating to each individual process or flow.
- Aggregation over data that is incomplete or inconsistent with respect to the aggregated process.

6.3 Documentation of process aggregation

Documentation of an aggregated process follows the same structure as for unaggregated processes, see Section 5.3, but should additionally contain information on the rationales and procedures used during the aggregation.

Estimates and calculations used in the aggregation procedure should be reported in the same way as any other estimate or calculation (see Chapter 3).

It is recommended to maintain the underlying unaggregated information that allows the data to be recalculated, partially revised and/or used in other contexts.

If it is not possible to use matrix-based calculations for aggregation of interlinked processes, use other documented software, or specify the algorithm or algebra used for the calculations, and specify whether the outcomes are compatible with the outcomes of matrix inversion. Document explicitly any software used for the calculation, along with the calculation procedures involved.

6.4 Validation

Validation of aggregated processes follows the same structure as for unaggregated processes, see Section 5.4.

6.5 Communication

The objectives and principles for communication of aggregated processes are the same as for individual data (see Section 3.5). Data on aggregated processes will be used either for further analytical treatment, e.g. impact assessment, or for final reporting of environmental performance. Thus, to prepare for further analytical treatment, the communication should be unambiguous, reliable, complete and referenced, cf. Section 6.3. Special care should be taken in defining the scope of applicability and the representativeness of data, so as to ensure that the data can be used appropriately.

All terminology should refer to international standards, especially ISO 14048.

If the communication takes place in a public context, data suppliers should always tell whether process descriptions have been independently reviewed or peer-reviewed prior to disclosure.

6.5.1 Checklist of things that can go wrong

Communication can be unsatisfactory when one or several of the following situations occur:

- When the data supplier misunderstands the communication context, i.e. the needs of the target group and what the process description is to be used for
- When the data supplier provides incomplete or ambiguous information on the process and the conditions of the process modelling
- When the data recipient misunderstands the original conditions of the process description and/or adds unfounded interpretations

The above three points may to some extent be seen as a cause-effect chain, one problem leading to the next.

Chapter 7. Describing environmental mechanisms

In contrast to the description of a technical system, which has to be provided in each LCA study, the description of environmental mechanisms is usually not part of an individual LCA study. Environmental mechanisms and models form a research field for itself. The LCA practitioner usually takes what is recommend as best practise, without developing new models. Therefore, the intention of this chapter of the guideline is not to describe how data are treated in environmental science to develop a model of a mechanism, but only to give guidance on how these data are translated into the form required by life cycle impact assessment (LCIA).

7.1 Definitions

Category endpoint: attribute or aspect of natural environment, human health, or resources, identifying environmental issues of concern (ISO 14042)

(Life cycle impact) Category indicator: quantifiable representation of an impact category (ISO 14042)

Characterisation factor: factor derived from an characterisation model which is applied to convert assigned LCI results to the common unit of the category indicator (ISO 14042)

Characterisation model: procedure to convert assigned LCI results to the common unit of the category indicator

Environmental mechanism: system of physical, chemical, and biological processes for a given impact category, linking the LCI results to category indicators and to category endpoints (ISO 14042)

Impact category: class representing environmental issues of concern into which LCA results may be assigned (ISO 14042)

System boundary: interface between a product system and the environment or other product systems (ISO 14042)

7.2 Action

7.2.1 How to describe an environmental mechanism

Based on the ISO definition an environmental mechanism of a certain impact category is described by :

- the category indicator representing the impact,
- the LCI results causing the impact, and
- the description in formulas of all physical, chemical and/or biological processes linking the LCI results to the category indicator,

which form the characterisation model.

The indicator can be selected to be more or less close to the category endpoint - the real damage of human beings, animals etc. Indicators close to the endpoint should be preferred. However, up to now for many categories there are methodological problems to achieve this goal. The fate of emissions is sometimes, the exposure often not included. Without exposure information for most of impacts no real damages can be assessed; exceptions are global effects as greenhouse warming for

which the term exposure is not meaningful. Therefore, the derivation of characterisation factors for the aggregation of substances to equivalents with respect to a certain physical property is the common procedure (e.g. GWP: IR absorption). Such indicators are called mid-point indicators.

From the broad variety of environmental mechanisms of different complexity it is clear that characterisation factors for end- as well as for mid-point indicators can cover variable number of specific mechanisms and factors for individual substances and individual environmental processes. Derivation of characterisation factors can thus be seen as an aggregation of environmental processes in parallel to the aggregation described in Chapter 6 for technical processes.

The different steps of environmental mechanisms are usually researched by different sciences (for example, the environmental mechanisms involved in human toxicology are studied in meteorology, soil science, plant physiology, and human toxicology). Further research, specifically for LCA purposes and with participation of LCA experts, is not common. Therefore, in some cases characterisation factors are taken directly from the environmental scientists studying the mechanism in question (e.g. IPCC for global warming), in other cases they are proposed by LCA experts (e.g. for acidification). Consequently, the compatibility of information related to different steps must be considered when calculating and using the factors.

Table 7.1 Important impact categories and characteristics of common indicators

	Focal point	**Time span**	**Fate & exposure**	**Time & place independent**
Climate change	mid	100 a	fate incl., exp. not relevant	Y
Depl. strat. ozone	mid	Eternity	fate incl., exp. not relevant	Y
Human toxicity	near end	Eternity	fate & exp. incl.	y (in fact Europe)
Eco-toxicity	near end	Eternity	fate & exp. incl.	y (in fact Europe)
Photo Smog	mid	5 d	fate incl.	y (in fact Europe, high NOX)
Acidification	near end	Eternity	fate & exp. incl.	y (in fact Europe)
Nutrification	mid	Eternity	fate & exp. not incl.	Y
(Guinée et al. 2001)				

7.2.2 Defining system boundaries in time and space for each input and output

Generally it should be possible to carry out an LCIA without any information about the time and site/place/location of the effects because an LCIA should cover all potential impacts independent of time and place. Therefore, the preferred time span is eternity and the preferred geographical reference is no specific reference (or: the whole world).

Characterisation models without the parameter time cover the eternity. If time is a parameter the functions can be integrated. The integration boundaries have to be selected according to the mechanism and to be tested in sensitivity analyses. In some cases different spans are common (GWP: 20, 100, 500 years); uncertainties increase with the time span. In the case of very short lifetimes the span can be set very short; example: photo smog: 5 days). In a similar way characterisation models without parameters related to certain places as temperature, rainfall, background concentrations etc. need no (and cannot use) spatial definitions.

Although the LCA is a time and site independent tool, temporally and spatially differentiated information is requested by decision makers (in reality different phases of a life cycle happen at different sites and at different times; real impacts may be eternal and global, but are more often specific in time and space). Further, as Table 7.1 shows, endpoint indicators most often do require spatially and temporally resolved information. In such cases the system boundaries have to be defined according to the goals of the study. Generally temporal and spatial boundaries are linked, because of the simple fact that all environmental processes, including transport, are time-related. The longer the lifetime of an emission (and its decomposition products) and the bigger its specific mobility, the larger is the area it will impact. Sensitivity analyses have to be carried out to avoid underestimating by using to narrow boundaries.

To obtain consistent models from environmental process descriptions from different sources, the conditions of the underlying measurements and calculations must be checked. Therefore, the sources must provide the necessary information. Relevant conditions should be documented. The quantitative effects of variation should be assessed. If meaningful, characterisation factors may be derived for different conditions (e.g. for POCP: low and high NOX background concentrations).

It is important to note that characterisation models with high temporal and spatial resolution need corresponding highly resolved LCI data. A simplified procedure for time and space differentiated LCIAs is:

- Distinguish between sensitive and non-sensitive areas: exclude emissions in non-sensitive areas from further analysis,
- Use effect-oriented site-factors, weighting the emissions by a factor between 0 and 1.

7.2.3 Allocation

When data belongs to more than one environmental process, or the process that we have data for is larger than the process we wish to describe, it may be necessary to estimate and/or calculate the part of the available data that should be included in the description. For this, the same procedures as those used in life cycle inventory analysis (section 5.2.3) should be used.

The characterisation factors should not contain hidden assumptions on the allocation of flows among different impact paths. Such allocations can be expected to be more time and site dependent than other data describing the mechanism. Therefore, it is important to check the temporal and spatial compatibility of the base model parameters and data related to allocations. Full documentation is required for the conditions of each environmental process and the effective allocations.

7.2.4 Estimating system boundaries

Generally, the effort to establish a model for an environmental impact is much larger than for the model of a technical process. Further, an impact model is developed for the assessment of more than one technical process (ideally for use in all LCIAs subsequently carried out). Therefore, the use of estimates of system boundaries (when actual system boundaries are unknown) is less acceptable than for technical processes. However, if there is an urgent need to compile an impact model from data without sufficient meta-data, temporal and spatial system boundaries can be estimated e.g. from the year of publication, the location of the authors institute etc. The model resulting from such estimates should be used only in very restricted application fields; sensitivity analyses are a must.

More important is the estimation of system boundaries in the sense of selection of characterisation factors with the most suitable boundaries for an LCI with given time and spatial boundaries. In the ideal situation, where sets for different boundaries are available, having all the same quality regarding experimental and assessed base data, the one with the most similar boundaries (referring to that of the LCI) is obviously to be selected; sensitivity analyses are recommended. Also, if no fully consistent data sets are available, the effect of deviating temporal and spatial boundaries must be checked in relation to the underlying model. Rules of thumb for the relation between system boundaries and characterisation factors may be derived. Advanced users of the characterisation model should apply these rules but also carry out sensitivity analyses. For basic user, sets of characterisation factors suitable for typical situations should be provided.

7.2.5 The relationship between concentrations and flows

The impact of a substance depends on its concentration in the medium where it has its effects. Usually, the flow is known (or data from which the flow can be calculated). From the flow the concentration may be calculated by dispersion models for specific sites and/or media. The selection of representative situations for these calculations is important. Rules of thumb for the transfer to different situations should be derived.

7.2.6 Recommended practice

It is recommend to describe environmental mechanisms in LCAs by models and model-derived factors, which fulfil the following requirements (Guinée et al. 2001):

- Scientific and technically valid
- Environmentally relevant
- Internationally accepted
- Value-choices and assumptions minimised
- Free choice of the focal point
- Linearity
- Time span preferred eternity
- Fate & exposure should be included
- No threshold (less is better)
- Time & place independent
- Operational
- Uncertainty as small as possible

Although temporal and spatial independence is a main point of the default practice, future environmental research should provide supplementary time and site resolved data, referring to representative system boundaries.

7.2.7 Checklist of things that can go wrong

- Fundamental but unlikely: The environmental mechanism is misunderstood completely. Based on wrong interpretations of preliminary results wrong parameters are measured and wrong correlations are derived.
- Relevant situations are neglected and therefore rare or non-typical situation overestimated with respect to the boundaries for which the final characterisation factors should be valid.
- The statistical treatment of the measured data is insufficient.
- Wrong characterisation factors or misleading information on their reliability result.

7.3 Documentation

Documentation of environmental mechanisms should be made in accordance with the structure described in the international standard ISO 14042 Environmental management – Life cycle assessment – Impact assessment. This text is adapted from Carlson & Pålsson (2002). The structuring has been developed as an addition to the SPINE data documentation format, to include the capability to generally handle information about environmental impact assessment (Carlson & Steen, 1998).

An impact category should be documented with its *name*, a *description*, and its *category indicator*(s).

A category indicator should be documented by its name, a description, the default unit in which the category indicator is measured, and a reference to the impact category to which the category indicator belongs.

An environmental mechanism should be described in terms of the *characterisation model* and the derived *characterisation factors* which convert LCI results to the common unit of the category indicator.
A characterisation model should be documented by its *name, version* (successive version numbers if the model is updated), *date completed* (date when the method was finalised or published), *method description, literature reference* (to a detailed description of the method), methodological range (geographical, technical, environmental etc. conditions under which the characterisation method may be applicable), *geographical area*, and a *mathematical expression*.

A characterisation factor should be documented by a *reference to input or output* for which the characterisation factor is valid, *reference to characterisation model* to which the characterisation factor belongs, *reference to category indicator* for which the characterisation factor is valid, the *quantity* of the characterisation factor, together with an explicit specification of the *category indicator unit* and the *unit of input and output* for which the factor is calculated.

7.4 Validation

Descriptions of environmental processes and mechanisms (impact chains) should be validated both before and after any aggregation steps.

Validation should be done by professionals with experience of the environmental process or mechanism in question.

Validation may include:

- Recalculation of the models involved.
- Identification of data gaps.
- Benchmarking the description against descriptions for similar processes or mechanisms.
- Cross-checking with data obtained through different sources and/or collection methods.
- Proof-reading.

The names and roles of people involved in the validation procedure(s) should be documented.

7.5 Communication

The objectives and principles for communication of descriptions of environmental processes and mechanisms are are the same as for individual data (see Section 3.5). Information about environmental mechanisms is mainly used by experts in environmental impact assessment and should be precise in terms of geographical, ecological scope of applicability and the representativeness of data, so as to ensure that the data can be used appropriately.

All terminology should refer to international standards, especially ISO 14042.

If the communication takes place in a public context, data suppliers should always tell whether the descriptions have been independently reviewed or peer-reviewed prior to disclosure.

Chapter 8. Describing a value system

The subject of this chapter is restricted to the use of value systems in LCA, not value systems and social theories on how values are formed, communicated, fixed, and measured in general. Values or value-choices in LCA are very closely related to ranking and weighting. Therefore in the following the focus will be on these items.

8.1 Definitions

Ranking: special case of grouping; to rank the impact category in a given order or hierarchy, such as high, medium or low priority (based on value-choices) (ISO 14042)

Weighting: converting indicator results of different impact categories by using numerical factors based on value choices (ISO 14042)

8.2 Action

8.2.1 How to describe a value system

Values are not objective. For each person they are the subjective result of social and cultural imprint, personal experiences, information on the world out of the individual life, and reflections on all of this.

Definition and description of value systems are especially relevant for ranking of impact categories and weighting of indicator data, although other phases of LCA may also contain choices that based on opinions, interests and values, e.g. the goal and scope definition. Therefore, value-choices play an important role deriving recommendations and decisions from LCI and LCIA results.

For ranking, the following issues may be considered:

- Ecological threat potentials,
- Reversibility of the effect,
- Scale of the effect (global, regional, local),
- Environmental preference of the population,
- Relationship of present and/or previous pollution to quality goals.

For weighting of LCIA results the value-choices can be based on or expressed in e.g.:

- Monetary values
- Authorised standards or targets
- Weights established in an authoritative panel procedure.

Because of the subjectivity of values it is a must to mark all steps containing value-choices. But subjective does not mean arbitrary. A reasonable value-choice can be reasoned and communicated to other people of a similar information level. It can be made plausible and also differences can be made plausible. Therefore, full information of all stakeholders is the precondition for setting up values for sound decisions.

Example: The German Federal Environmental Agency developed in a panel procedure a valuation procedure, which defines three criteria for each impact category describing its relevance (UBA 1999). For the two non-quantitative criteria sub-criteria were derived similar to those listed above.

8.2.2 Defining system boundaries in time and space

The system boundaries of the values must be compatible (but not necessarily identical) to the boundaries of the project in which they are applied.

The value system need not be of recent date. Usually it is sufficient to check whether there were principal changes since the definition of the system with respect to the real environmental situation or the knowledge about environmental mechanisms (or on the level of pure subjectivity). Quantitative criteria, such as normalisation data, should be up to date.

In a similar way for the geographical area: a given value system can be used in other areas than the original, if the environmental, social etc. situations are sufficient similar. Spatial resolution is possible but not common; example: 1 km passenger car driving in the EU with gasoline from Arabian oil; use of European and Arabian specific contributions and environment-related value-choices; alternatively in one panel all relevant regions could be represented.

8.2.3 Recommended practice

No specific procedure is recommended for ranking. Generally, sub-criteria behind the criteria and plausible reason for the selection of the (sub-)criteria and the assessment as well as complete and transparent documentation is required.

In general, weighting should be avoided. This is in compliance with Guinée et al. (2001), UBA (1999) and others. If it cannot be avoided, national or international authorised factor sets should be used. Because they are not available currently, it may be necessary to create such a set in the frame of a project by a panel-like procedure or adopt a "less authorised" set. In such a case the full argumentation leading to the results have to be documented completely and transparent.

For ranking and weighting systems that are intended to form a standard (e.g. for all LCAs in a certain country), agreements among the main stakeholders on the following points should be achieved and documented:

- The involved stakeholders are really the main players (important for general acceptance)
- The overall method: selection, definition (example "weighting": monetary values or authorised standards or targets or established in an authoritative panel-procedure)
- Main criteria: selection, definitions, relative weights (example "ranking" (UBA 1999): ecological endangering, distance to target, specific contribution)
- Sub-criteria for each main criterion: selection, definitions, relative weights (example "ecological endangering" (UBA 1999): ecological threat potentials, reversibility of the effect, scale of the effect etc.)
- Panels: procedure (questionnaire, discussion etc.), selection of participants (overlap with the stakeholders mentioned above), number of rounds etc.
- Classification schemes: number and definition of classes
- Derivation of weighting factors / allocation to classes: rules to convert questionnaire results into classifications or factors etc.

Evidently at least some of these points are trivial. However where LCA borders social science they are still to be implemented as standard procedures.

8.2.4 Checklist of things that can go wrong

- Criteria based on unclear goal and scope definition.
- Insufficient information and/or freedom in a panel procedure, resulting in non-plausible decisions.
- Use of values derived for other, non-comparable system boundaries.

8.3 Documentation

Documentation of value systems should be made in accordance with the structure described in the international standard ISO 14042 Environmental management – Life cycle assessment – Impact assessment. This text is adapted from Carlson & Pålsson (2002). The structuring has been developed as an addition to the SPINE data documentation format, to include the capability to generally handle information about environmental impact assessment (Carlson & Steen, 1998).

A ranking or weighting method should be documented by its name, *version* (successive version numbers if the method is updated), *method description, literature reference* (to a detailed description of the method), methodological range (geographical, temporal, cultural, political etc. scope considered for the principle or policy), and *geographical area*.

A weighting factor should be documented by a *reference to the category indicator* for which the weighting factor is valid, *reference to weighting method* to which the weighting factor belongs, the *quantity* of the weighting factor, together with an explicit specification of the *category indicator unit* and the *unit of weighting measure* for which the factor is calculated.

8.4 Validation

Descriptions of value systems should be validated prior to communication.

Validation should be done by professionals with experience of the methods in question.

Validation may include:

- Quality assurance of the procedures applied in data collection and treatment.
- Completeness of the stakeholders and issues covered. Identification of data gaps in relation to the system boundaries (geographical, temporal, cultural, political etc.).
- Benchmarking the description against descriptions for similar value systems and/or stakeholder groups.
- Cross-checking with data obtained through different sources and/or collection methods.
- Proof-reading.

The names and roles of people involved in the validation procedure(s) should be documented.

8.5 Communication

The objectives and principles for communication of descriptions of environmental processes and mechanisms are are the same as for individual data (see Section 3.5). Information about value systems is mainly used by experts in environmental impact assessment and should be precise in terms of geographical and cultural scope of applicability and the representativeness of data, so as to ensure that the data can be used appropriately.

All terminology should refer to international standards, especially ISO 14042.

If the communication takes place in a public context, data suppliers should always tell whether the descriptions have been independently reviewed or peer-reviewed prior to disclosure.

References

de Beaufort-Langeveld A S H, Bretz R. (ed.) (2001). Code of Life Cycle Inventory Practice. Pensacola: SETAC.

van den Berg N W, Huppes G, Lindeijer E W, van der Ven B L, Wrisberg N. (1999). Quality Assessment for LCA. Leiden: CML, Leiden University. (CML Report 152). (http://www.leidenuniv.nl/interfac/cml/ssp/publications/quality.pdf)

Carlson R, Pålsson A-C. (2001). Industrial environmental information management for technical systems. Journal of Cleaner Production 9(5):429-435.

Carlson R, Pålsson A-C. (2002). Documentation of environmental impact assessment, compatible with SPINE and ISO/TS 14048. Göteborg: Chalmers University of Technology. (IMI-report 2002:1).

Carlson R, Steen B. (1998). A Data Model for LCA Impact Assessment. Presentation at the 8th Annual Meeting of SETAC-Europe, Bordeaux, 1998.04.14-18.

Carlson R, Löfgren G, Steen B, Tillman A-M. (1998). LCI Data Modelling and a Database Design. International Journal of Life Cycle Assessment 3(2):106-113.

European Commission. (1996). Technical guidance document in support of commission directive 93/67/EEC on risk assessment of new notified substances and commission regulation (EC) No. 1488/94 on risk assessment for existing substances. Part II. Luxembourg: Office for Official Publications of the European Communities.

Ecobilan (1999). TEAM User's Manual – Version 3.0.

Ekvall T, Finnveden G. (2001). Allocation in ISO 14041 – a critical review. Journal of Cleaner Production 9:197-208.

Erixon M. (1999). *Practical Strategies for Acquiring Life Cycle Inventory Data in the Electronics Industry.* Göteborg: Centre for Environmental Assessment of Product and Material Systems (CPM), Chalmers University of Technology. (CPM Report 1999:3).

Guinée J B, Gorrée M, Heijungs R, Huppes G, Kleijn R, de Koning A, van Oers L, Sleeswijk A W, Suh S, Udo de Haes H A, de Bruijn H, van Duin R, Huijbregts M A J, Lindeijer E, Roorda A A H, van der Ven B L, Weidema B P. (2001). Life cycle assessment- an operational guide to the ISO standards. Leiden: CML, Leiden University.

Huijbregts M A J. (1998) Part I: A general framework for the analysis of uncertainty and variability in life cycle assessment. International Journal of Life Cycle Assessment 3(5):273-2800

Huijbregts M A J, Norris G, Bretz R, Ciroth A, Maurice B, von Bahr B, Weidema B, de Beaufort A S H. (2001). Framework for Modelling Data Uncertainty in Life Cycle Inventories. International Journal of Life Cycle Assessment 6(3):127–132

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

Ott W R. (1995). Environmental Statistics and Data Analysis. Lewis Publishers (CRC Press LLC), New York.

Pålsson A-C. (1999). *I*ntroduction and guide to LCA data documentation using the CPM data documentation criteria and the SPINE format. Göteborg: Centre for Environmental Assessment of Product and Material Systems (CPM), Chalmers University of Technology. (CPM Report, 1999:1).

Raynolds M, Checkel M D, Fraser R A. (2000). The Relative Mass-Energy-Economic (RMEE) Method for System Boundary Selection. Part 1: A Means to Systematically and Quantitatively Select LCA Boundaries. International Journal of Life Cycle Assessment 5(1):37–46.

StatSoft Inc. (1999). Electronic Statistics Textbook. StatSoft, Tulsa, OK. (http://www.Statsoft.com/textbook/stathome.html)

Suh S. (2000). Missing Inventory Estimation Tool (MIET 1.2). http://www.leidenuniv.nl/cml/ssp/software

Tillman A-M, Ekvall T, Baumann H, Rydberg T. (1994). Choice of system boundaries in life cycle assessment. Journal of Cleaner Production 2(1):21–29.

UBA. (1999). Bewertung in Ökobilanzen. Berlin: Umweltbundesamt. (UBA-Texte 92/99)

Weidema B.P. (1998). Multi-User Test of the Data Quality Matrix for Product Life Cycle Inventory Data. International Journal of Life Cycle Assessment 3(5):259-265.

Weidema B P, Lindeijer E. (2001). Physical impact of land use in product life cycle assessment. Final report of the EURENVIRON-LCAGAPS sub-project on land use. Lyngby: Department of Manufacturing Engineering and Management, Technical University of Denmark. (IPL-033-01).

Weidema B P, Wesnæs M S. (1996). Data Quality Management for Life Cycle Inventories - An Example of Using Data Quality Indicators. Journal of Cleaner Production 4(3-4):167-174.

Annex 1. Common distribution types

Uniform (min, max)

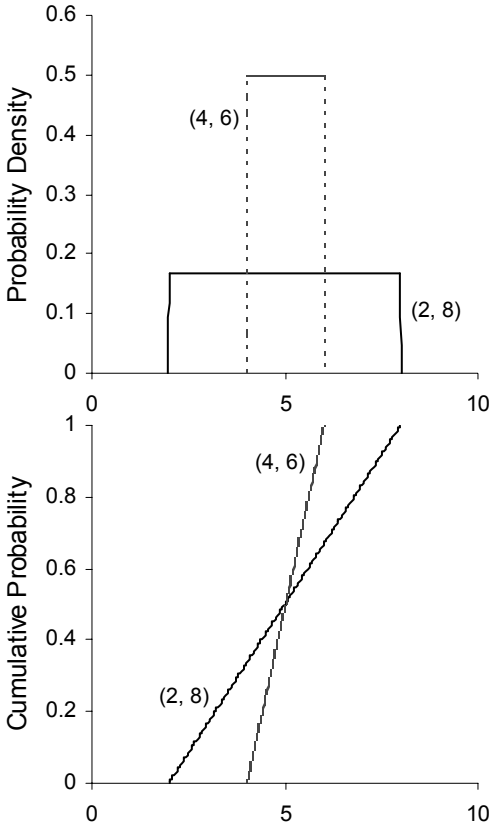
The uniform distribution provides the simplest representation of uncertainty of all, and is appropriate when a range of possible values can be identified, but it is not known whether any values within this range are more likely to occur than others. It is defined by specifying minimum and maximum possible values.

Probability density function:

$$f(x) = \frac{1}{b-a}; \quad a \le x \le b$$

Expected value: $E[X] = \frac{a+b}{2}$

Variance: $Var(X) = \frac{(b-a)^2}{12}$



Triangular (min, mode, max)

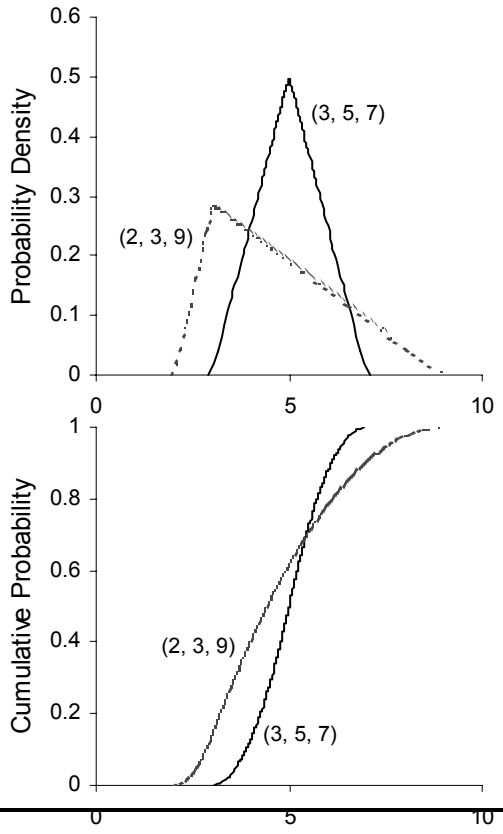
A triangular distribution is a simple and convenient distribution to use when little is known about a parameter, other than a central tendency in its range, or that a certain value is more likely to occur than values near either extreme. It is characterised by specifying minimum and maximum values, and the mode (most likely value). The triangular distribution tends to overemphasise the tails and under-emphasise the shoulders of the parameter range, however, the arbitrary shape and sharp corners can be a convenient way to convey that the details of the distribution are not known, and thus help to prevent a false sense of accuracy.

Probability density function (symmetric distribution):

$$f(x) = \frac{b - |x-a|}{b^2}; \quad a-b \le x \le a+b$$

Expected value (symmetric distribution): $E[X] = a$

Variance (symmetric distribution): $Var[X] = \frac{b^2}{6}$



Normal (μ, σ)

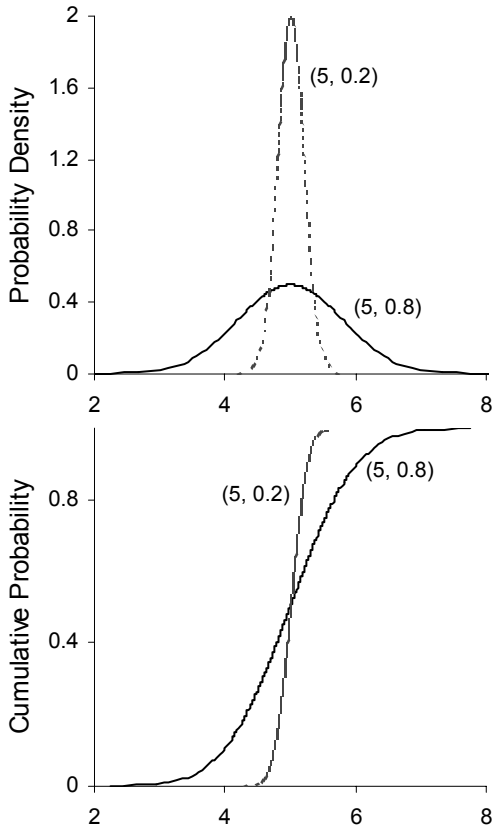
The normal or Gaussian distribution is commonly employed to represent uncertainty resulting from unbiased measurement errors. It results when a number of unrelated, continuous random variables are added together. However, it is inappropriate for many quantities because negative values can occur. The parameters are estimated from the sample mean and standard deviation.

Probability density function:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}; \quad -\infty < x < \infty$$

Expected value: $E[X] = \mu$

Variance: $Var(X) = \sigma^2$



Lognormal (median, σg)

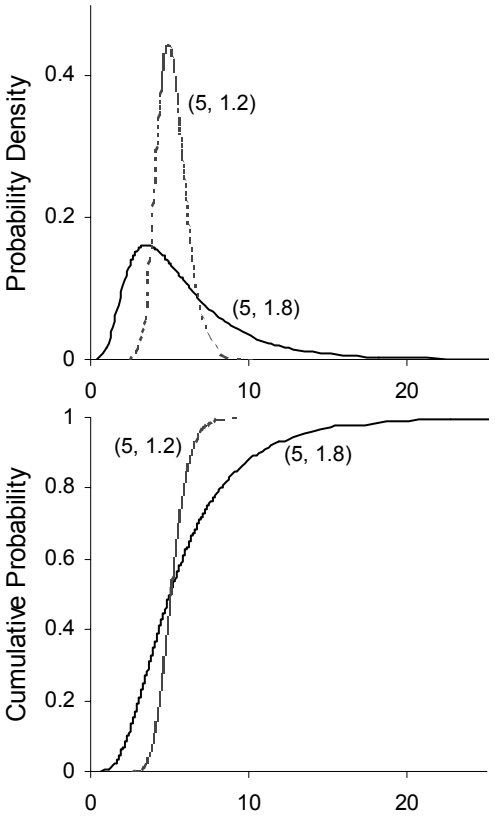
The lognormal distribution provides a good representation for quantities that are non-negative and positively skewed. It results from the product of many independent random variables multiplied together, and is typical of many natural processes, e.g. pollutant concentrations. It is particularly appropriate for representing large uncertainties the are expressed on a multiplicative or order-of-magnitude basis (e.g. "within a factor of two"). The parameters of the distribution are the median and the geometric standard deviation. These are equivalent to the mean and standard deviation of $Y = \ln X$, since the distribution arises when the logarithm of the random variable is described by a normal distribution.

Probability density function:

$$f(x) = \frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\ln x-\mu}{\sigma}\right)^2}; \quad -\infty < \mu < \infty$$

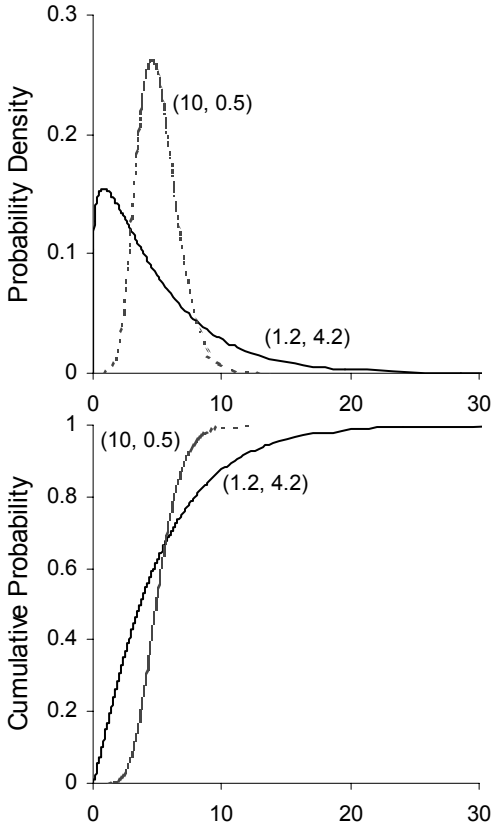
Expected value: $E[X] = e^{\mu+\frac{\sigma^2}{2}}$

Variance: $Var(X) = e^{2\mu+\sigma^2}(e^{\sigma^2} - 1)$



Gamma (η, λ)

The Gamma distribution, like the lognormal distribution, is applicable to many physical quantities, since it returns only non-negative values. However, it is less positively skewed and less "tail-heavy" than the lognormal distribution, and thus generally prescribes a lower probability to the extremes of the distribution than the lognormal distribution. The gamma distribution has a flexible shape, and is described by a shape parameter, η, and a scale parameter, λ. The shape of the distribution runs from highly right skewed to symmetrical and bell-shaped as η increases (it is equivalent to an exponential distribution when η=1, but approximates a normal distribution when η>10).



Probability density function:

$$f(x) = \frac{\lambda^{\eta}}{\Gamma(\eta)} x^{\eta-1} e^{-\lambda x}; \quad x \geq 0, \eta > 0, \lambda > 0$$

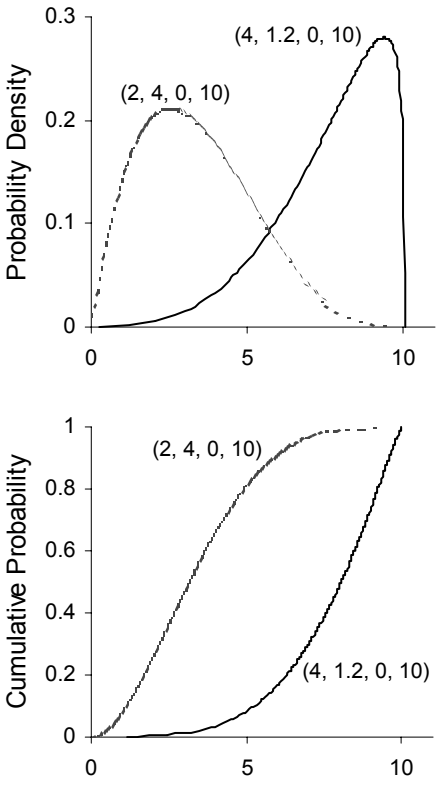
where $\Gamma(\eta) = \int_0^{\infty} u^{\eta-1} e^{-u} du$ or, for integers, $\Gamma(\eta) = (\eta - 1)!$

Expected value: $E(X) = \frac{\eta}{\lambda}$

Variance: $Var(X) = \frac{\eta}{\lambda^2}$

Beta (c, d, min, max)

The Beta distribution is a useful distribution where a bounded distribution is required. It can be made to reflect both positively skewed or negatively skewed quantities, and thus provides a flexible means of representing variability over a fixed range. The first two parameters specify the shape, whilst the second two parameters specify the range endpoints. If not specified, the range defaults to zero to one.



Probability density function:

$$f(x) = \frac{1}{\mathrm{B}(c,d)} x^{c-1}(1-x)^{d-1}; \quad 0 \leq x \leq 1$$

where $\mathrm{B}(c,d) = \frac{\Gamma(c)\Gamma(d)}{\Gamma(c+d)}$

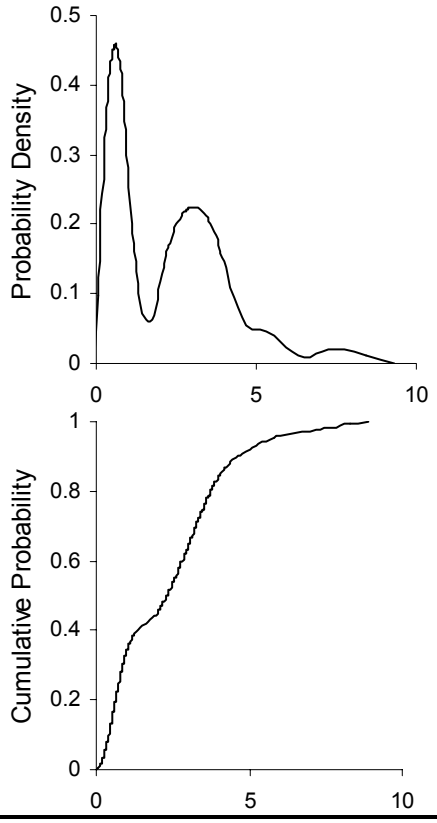
Expected value: $E(X) = \frac{c}{c+d}$

Variance: $Var(X) = \frac{cd}{(c+d)^2(c+d+1)}$

Custom fit Distribution

If a variable does not fit any of the above distribution shapes (e.g. a bimodal distribution), the distribution shape can be custom fit to the data. This is most simply done by defining points on the cumulative probability density curve. For example, the table below specifies the PDF and CDF shown, where the first column gives the probability that the outcome will be less than or equal to the corresponding element in the second column.

Probability	*Outcome*
0%	0
34%	1
48%	2
64%	3
84%	4
92%	5
96%	6
97%	7
99%	8
100%	9



Published by ENEA – Communication Unit

Lungotevere Thaon di Revel, 76 – 00196 Rome
www.enea.it

Editing: Giuliano Ghisu

Printing: Frascati ENEA Research Centre

Printed in October 2004