The Effect of Data Structure and Model Choices on MFA Results: A Comparison of Phosphorus Balances for Denmark and Austria

Klinglmair, Manfred; Zoboli, Ottavia; Laner, David; Rechberger, Helmut; Astrup, Thomas Fruergaard; Scheutz, Charlotte

Published in:
Resources, Conservation and Recycling

Link to article, DOI:
10.1016/j.resconrec.2016.03.009

Publication date:
2016

Document Version
Peer reviewed version

Link back to DTU Orbit

Citation (APA):

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain.
- You may freely distribute the URL identifying the publication in the public portal.

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.
The effect of data structure and model choices on MFA results: a comparison of phosphorus balances for Denmark and Austria

Manfred Klinglmair¹, Ottavia Zoboli², David Laner², Helmut Rechberger², Thomas Fruergaard Astrup¹, Charlotte Scheutz¹

1) Technical University of Denmark, Department of Environmental Engineering
2) Vienna University of Technology, Institute for Water Quality, Resource and Waste Management

KEYWORDS: substance flow analysis; MFA; phosphorus; data quality; uncertainty assessment; data reconciliation
ABSTRACT

Material Flow Analysis (MFA) studies for a particular substance often exist for several different countries or regions, but share a similar goal and scope. In direct comparisons of such regional resource budgets, the importance of the choices made in establishing an MFA system tends to be disregarded. We identify and quantify the effects of choices made in system layout, data material and uncertainty assessment on the outcome of regional MFAs using two recent country-scale MFAs (of Denmark and Austria) of phosphorus as a case study.

We highlight the differences in system boundaries and definition of flows and processes. We quantify types and choice of data sources; analyse the consistency of the data used by looking at the extent of data reconciliation, as a measure of model quality; quantify the effect of different approaches to uncertainty assessment; and show the influence of aggregating/disaggregating flows.

We show that differences in system layout are mostly attributable to varying goals and scope definitions. Direct comparison of uncertainties across studies is problematic: both studies draw on similar types of data sources, yet they show very different uncertainty assessments; the uncertainty assessment in MFA is always subjective to a certain extent. We demonstrate that reconciliation of conflicting data provides a useful measure to assess data consistency and model quality: data are more consistent (5% average change in reconciled data) in the Austrian than in the Danish (9%) case. We suggest an iterative approach to uncertainty assessment. Likewise, we demonstrate the effect of the aggregation of flows on model uncertainty. These findings quantify and emphasise the importance of examining MFA studies’ metadata and suggest an approach to be followed when drawing on such studies as a source of information.
1. Introduction

Material Flow Analysis (MFA; Baccini & Brunner 2012) has become a widespread approach to visualise the anthropogenic turnover of materials across a defined geographical area. For a number of substances, e.g. phosphorus (P), several MFA studies—often called Substance Flow Analysis (SFA) when studying a substance—of regional or national households with similar goal and scope definitions are available. Patterns of resource use are analysed to evaluate resource efficiency and recycling activities (e.g. Chen and Graedel 2012), anthropogenic material stocks are investigated using dynamic analysis to quantify current and future resource potentials (e.g. Müller et al. 2014), and sources and sinks of problematic substances are the focus of MFA studies addressing environmental pollution issues (e.g. Hansen and Lassen 2003). While MFA studies may obviously differ with respect to their geographical scope, the authors’ modelling choices and the database used for MFA are a different set of factors often disregarded in the comparison of regional resource budgets. Country or multi-country level MFA studies are quite consistent with respect to system boundaries, and external trade and national accounts, across substances and countries (Fischer-Kowalski et al. 2011). For P, Seyhan (2009) suggests a standard template regarding processes, flows and system boundaries, for P resource budgets on the national or regional level.

However, there is considerable freedom in data selection and model choices. This implies that the model and data structure of an MFA typically vary from one study to another, an issue tending to get little mention in comparisons of P flow studies, which are often taken at face value (e.g. the review of phosphorus MFAs by Chowdhury et al. 2013). Data uncertainty quantification and consequent uncertainty analysis has become an integral part of MFA, and has recently been analysed in more detail by Laner et al. (2014 and 2015b). Schwab et al. (2015) have moreover presented a detailed framework to systematically characterise data used in an MFA. Still, uncertainties due to model structure and
scenario assumptions (or arbitrary choices of the practitioner), and their influence on MFA results, are rarely addressed and therefore difficult to analyse.

It is the aim of this work to examine the effect of the MFA system layout and modellers’ choices on the outcome of regional MFAs, using the country-scale MFAs of P in Denmark and Austria as case studies. The Danish study (Klinglmair et al. 2015) examines the anthropogenic P budgets of three distinct regions in Denmark as subsystems of the P flows on the national scale. The Austrian study (Zoboli et al. 2015a) presents a 22-year timeline of detailed P budgets on the national level with particularly detailed accounts of flows in the waste management sector. Both studies share the objective of examining anthropogenic P flows from a standpoint of resource efficiency for countries of comparable size and population.

The following characteristics of an MFA model are of interest to the above objective: 1) system boundaries, layout (i.e. processes/flows shown, resolution, definition of processes, flow definitions), 2) types and choice of data sources, 3) the closely linked issues of data reconciliation and uncertainty assessment: over-determination, i.e. the difference between the number of independent balance equations and unknown variables (flows) of the model; the results of different approaches to quantify uncertainties in both studies; the extent of data reconciliation as a measure of data consistency and model quality; and the effects of aggregation/disaggregation of flows on model uncertainty. Points 1 and 2 are discussed in qualitative terms; point 3 consists of systematically quantifiable results. It is not the aim of this study to compare the P budgets of the two countries as such; instead, we take the specific examples of these two studies to illustrate, on a more general level, the extent of differences that can occur in the areas above due to the MFA practitioner, the data available, and the model formulation, apart from actual physical differences between the systems. In this way, we demonstrate a series of steps to analyse MFA studies in order to make plain, and quantify, this often-overlooked
distinction. While this work is based on studies of P, the conclusions we draw may be applied to a wider range of country- or regional-scale resource budgets.

2. Materials & methods

2.1 Source material and background

The present study is based on two recent P balances (Klinglmair et al. 2015; Zoboli et al. 2015a) for Denmark and Austria, conducted at the Technical University of Denmark and Vienna University of Technology, respectively. Table 1 shows characteristics pertinent to the two countries’ P budgets. The area used for agriculture is similar in both countries, but relatively smaller in Austria, due to the larger area of the country and the high share of forests. However, more livestock is produced in Denmark, with a particularly high number of pigs compared to Austria.
Table 1  Basic country data pertaining to P households in Austria (AT) and Denmark (DK).

<table>
<thead>
<tr>
<th></th>
<th>AT</th>
<th>DK</th>
</tr>
</thead>
<tbody>
<tr>
<td>Population</td>
<td>8,200,000</td>
<td>5,580,000</td>
</tr>
<tr>
<td>Area</td>
<td>8,387,899</td>
<td>4,309,400</td>
</tr>
<tr>
<td>Agricultural area</td>
<td>2,879,000</td>
<td>2,567,000</td>
</tr>
<tr>
<td>% of total</td>
<td>34%</td>
<td>60%</td>
</tr>
<tr>
<td>Forest area</td>
<td>3,405,752</td>
<td>600,427</td>
</tr>
<tr>
<td>% of total</td>
<td>41%</td>
<td>14%</td>
</tr>
<tr>
<td>Livestock</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cattle</td>
<td>1,976,527</td>
<td>1,567,970</td>
</tr>
<tr>
<td>Pigs</td>
<td>3,004,907</td>
<td>12,931,678</td>
</tr>
<tr>
<td>Sheep</td>
<td>361,183</td>
<td>143,889</td>
</tr>
<tr>
<td>Horses</td>
<td>81,637</td>
<td>61,476</td>
</tr>
<tr>
<td>Goats</td>
<td>72,358</td>
<td>(no data)</td>
</tr>
<tr>
<td>Poultry</td>
<td>14,644,413</td>
<td>18,206,943</td>
</tr>
<tr>
<td>Livestock/ha agricultural land</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cattle</td>
<td>0.7</td>
<td>0.6</td>
</tr>
<tr>
<td>Pigs</td>
<td>1.0</td>
<td>5.0</td>
</tr>
<tr>
<td>Sheep</td>
<td>0.13</td>
<td>0.06</td>
</tr>
<tr>
<td>Horses</td>
<td>0.02</td>
<td>0.02</td>
</tr>
<tr>
<td>Goats</td>
<td>0.03</td>
<td>(no data)</td>
</tr>
<tr>
<td>Poultry</td>
<td>28.8</td>
<td>7.1</td>
</tr>
</tbody>
</table>

While the geographical scope and the areas under scrutiny are comparable, the objectives of the studies differ in their respective focus. In the Austrian case, Zoboli et al. (2015a) draw upon an existing P budget (Egle et al. 2014) to examine the utility of multi-year timelines in MFA studies, highlighting considerable variation in the period from 1990–2011. The model layout was only slightly modified from the earlier study, and comprises 56 processes and 122 flows, including 10 ‘virtual’ processes to aggregate or split flows for visualisation purposes. 11 processes contain subsystems, where the process is further disaggregated into flows and processes on a sub-system level. The Danish study (Klinglmair et al. 2015), conducted for the year 2011, aims for a higher spatial resolution by incorporating three regional subdivisions, for the agriculture and waste management processes, into the national P balance. The purpose of this study was to investigate imbalances in the P household and potential for P recovery.
resulting from differences in agricultural practice, and population and industrial density between the
Danish regions. The model layout comprises a total of 50 processes and 166 flows, including 12
‘virtual’ processes and 9 processes containing subsystems (see Tables S.1 and S.2 in the supplementary
data for flows and data sources). For the purpose of this comparison, we look at the results for the year
2011 in both studies. Both systems are overdetermined, i.e. more variables (flows) are known than are
necessary to solve the balance equations. In the Danish model, all 166 flows are determined, while all 7
stock changes are unknown and therefore calculated using mass balance equations. This means that
there exist 7 unknowns for 41 independent balance equations (processes not containing a sub-system).
In the Austrian case, 16 of 122 flows and 7 out of 8 stock changes are calculated, so that we find 44
independent balance equations and 23 unknowns. The degree of overdetermination is the number of
independent balance equations (constraints) minus the number of unknown variables; the Danish model
can therefore be said to be overdetermined to a higher degree. Figure 1 shows simplified qualitative
representations of both MFA systems. It has to be noted that several processes shown in the main
system in the Austrian study (animal husbandry, crop farming, forestry & miscellaneous soils, and
bioenergy; Figure 1b) are found in sub-systems in the agriculture process in the Danish study (see
Figure 1a).
Figure 1a
Figure 1. Simplified qualitative representation of the MFA systems for P flows and stocks in Denmark (a) and Austria (b). Main systems are shown, i.e. without subsystems contained in the processes. The number of processes, flows, and stocks in the subsystems are listed below the process name. The
numbers of flows and processes in each of the 3 regional subsystems are shown separately in Figure 1a (Denmark), where applicable. Arrows are marked (e.g. 1F, 4F ...) with the number of flows they represent, with arrow width proportional to the number of flows. The dashed lines denote the system boundary. Processes pertaining to crop farming, forestry, animal husbandry, and biogas are found in subsystems in the Danish study.

2.2 Material Flow Analysis (MFA)

Both studies use the STAN 2.5 software (Cencic & Rechberger 2008) for balancing material flow systems with options of data reconciliation and uncertainty propagation, assuming all uncertain flows to be random, normally distributed variables given by mean value and standard deviation. Figure 1 shows a simplified graphical representation of the two MFA models.

Since flow values are based on data of varying quality (and hence uncertainty) from multiple sources, these data will generally not be fully consistent with the model, but contradict each other to a certain degree, i.e., the input and output flows of processes will not balance completely. This problem can be solved by data reconciliation as implemented in STAN. Data reconciliation in STAN is a weighted least-squares optimisation of the measurement adjustments subject to model constraints (system balances, where the weights are chosen to be the inverse of the measurement variance; see Laner et al. 2014). A further result of data reconciliation is the reduction of the uncertainty of the reconciled flows.

2.3 Approach
As a first step, attention needs to be paid to the differences in the definitions of system layout and system boundaries, even as conventions about boundaries at various geographical scales have been developed (see Chowdhury et al. 2013). Nevertheless, the overall system layout warrants attention in any comparison of MFA studies. Furthermore, a (mostly qualitative) evaluation of the essential differences in the system layouts in representing the respective P flows and stocks, and in the definitions of processes and flows, is necessary.

Second, we examined the data material on which the models were built. We categorised the data sources used in the quantification of P flows and stocks (material/bulk flows as well as P concentrations) for both systems and ranked them according to their relative importance. Since flow values entered in the model are generally based on several sources, we looked at how frequently data sources from each respective category were cited as references for flow values.

Third, we compared the consistency of the models, their layouts and the approaches to uncertainty assessment. ‘Consistency’ here means the fit of the data given the mass balance constraints of the model in consideration of the defined uncertainty ranges for the input data. The more the entered values contradict each other, the higher is the conflict in the database in view of the mass balances of the model. The effect of data reconciliation can be quantified as the change of individual flow values after data reconciliation in STAN (in % of the entered values); the mean of all deviations gives a measure of the extent of data reconciliation or ‘fit’ of the model and data (see Eq. 1; see also Zoboli et al. 2015a), with $\bar{D}$ being the mean relative deviation of the reconciled data from the entered values, $m$ the total number of flows, $x_n$ the entered value for flow n, and $\hat{x}_n$ the reconciled value for flow n.

$$\bar{D} = \frac{\sum_{n=1}^{m} \left| \frac{x_n - \hat{x}_n}{x_n} \right| \ast 100}{m}$$ (1)
Fourth, the quantification of uncertainty in MFA studies, even if reproducible and systematic, is subjective to a certain degree (see Laner et al. 2015b), which puts limitations on a direct comparison of uncertainty ranges alone. We therefore compare and evaluate the approaches to uncertainty assessment as well as the quantitative effects thereof, in conjunction with data reconciliation in STAN.

Finally, we evaluate the effects of aggregation/disaggregation of flows: often, a flow is an aggregate of several smaller flows (as is often the case in the Danish study, for example, where certain flows on the national scale are disaggregated into regional flows). This aggregate flow is either known and it may be subject to data reconciliation, such as other determined flows; or, the flow is unknown and is calculated as the sum of the flows it is composed of. In the former case, data reconciliation will result in a reduction of uncertainties, because more information is given (i.e. the constituent flows are known as well as the total (=aggregated) flow, hence, resulting in over-determination). In the latter case, the uncertainties of the constituent flows are propagated to derive the uncertainty of the total (aggregated) flow. Depending on how many aggregate flows are present in a model and whether they contribute to overdetermination or not, this may have a considerable effect on the degree of data reconciliation and the consequent uncertainty of the model results.

3. Results & discussion

3.1 Qualitative aspects

3.1.1 System layout

The system boundaries for both studies could be expected to be defined in similar ways. Indeed, they differ only with regard to water bodies. The Danish study places the hydrosphere outside the
system boundary, considering P flows to water bodies simply as losses. The Austrian study, on the other hand, includes water bodies in the system; this can be attributed to the slightly different goal & scope definitions of the two studies. Whereas the Austrian study put a focus on P as a major contributor to eutrophication in addition to the P resource management issues (this is also highlighted by a follow-up study on the effect of anthropogenic P emissions on P load dynamics in the river Danube; Zoboli et al. 2015b), the Danish study put emphasis mainly on the resource aspects of P management in Denmark. Figure 2 shows the summed in- and outflows, and stock changes, of the main processes in the Danish and Austrian models. Note that ‘Agriculture & forestry’ has been combined in the figure, although in the Austrian case it is divided into three processes in the main system. The high turnover in the industry & trade process results from the process handling most imports and exports, as well as most exchanges with the consumption and agriculture and forestry processes.
Figure 2. Overview of summed in- and outflows and stock changes per main process, for both studies examined, shown in kt/yr (AT: Austria; DK: Denmark).

When compared to the template for country-scale P balances put forward by Seyhan (2009), the main systems of both studies conform to the suggested types of processes, flows and stocks. Exceptions are mineral P reserves (non-existent in both countries) as well as a Bioenergy process in the main system for Austria (not part of the template; of similar magnitude in Denmark, but located in the Agriculture subsystem), and minor differences in the classification of various wastewater flows.

Wastewater treatment is the only main process without a stock in both systems. Minor differences between the two studies can be more frequently found in the detailed subsystems, and process definitions. Wood & paper industry is found in the Forestry process in the Austrian case; in the Danish case, any industry is located in the Industry and trade process, with wood from forestry (which plays a considerably smaller role in Denmark) subsumed under the Agriculture process; see Figure 1. A second
such example is animal processing: slaughtering is not identified as a separate process in either study, thus the activity is not explicitly located in the model.

The availability or use of specific datasets is linked to the layout of the MFA model. This is highlighted by the use of datasets provided by national statistical offices regarding agricultural production (with one of the main turnovers of P). In the Austrian study, the datasets used were national supply balances; the Danish study draws on production data for domestic agricultural production. The statistics referenced by the Austrian study list national data only; the corresponding Danish statistics give regionalised data. Other, regionalised datasets also exist for Austria (e.g. Statistics Austria 2015). These differences in model layout were therefore not caused by data availability. However, the datasets used would not readily allow the reproduction of a regionally differentiated model layout, such as the Danish one. Table 2 compares the datasets by national statistical offices from which the respective values were drawn. In cases where regionalised statistical data for agricultural production exist in the Danish case, they are reflected in the model. Moreover, these datasets do not list private production of vegetables in the kitchen gardens for Denmark. The flow ‘garden vegetables’ exemplifies the influence of data availability on model layouts: the flow is not shown in the Danish model (i.e. it is located within the Consumption process), whereas in the Austrian case, a subsystem of consumption shows production from private gardens (see Figures 1 a & b; see also Table S.1). A direct comparison of the MFA models’ layouts can yield insight into such data gaps, or otherwise draw attention to erroneously omitted (‘forgotten’) flows.
Table 2. Comparable flows based on datasets from national statistical offices (see Tables A.1 and A.2 for references). The flow ‘Garden vegetables’ has no counterpart in the Danish study.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Agricultural products</td>
<td>no</td>
<td>no</td>
<td>Plant products (food)</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>Seeds</td>
<td>no</td>
<td>no</td>
<td>Seeds &amp; planting material</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>Food</td>
<td>no</td>
<td>no</td>
<td>Food</td>
<td>no</td>
<td>no</td>
</tr>
<tr>
<td>Import food</td>
<td>no</td>
<td>no</td>
<td>Imported food</td>
<td>no</td>
<td>no</td>
</tr>
<tr>
<td>Export food</td>
<td>no</td>
<td>no</td>
<td>Exported food</td>
<td>no</td>
<td>no</td>
</tr>
<tr>
<td>Eggs and milk production</td>
<td>no</td>
<td>no</td>
<td>Milk &amp; eggs</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>Meat production</td>
<td>no</td>
<td>no</td>
<td>Meat &amp; slaughterings</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>Garden vegetables</td>
<td>no</td>
<td>no</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Import live animals</td>
<td>no</td>
<td>no</td>
<td>Other imports</td>
<td>no</td>
<td>no</td>
</tr>
<tr>
<td>Export live animals</td>
<td>no</td>
<td>no</td>
<td>Other exports</td>
<td>no</td>
<td>no</td>
</tr>
</tbody>
</table>

Without a clear justification in the respective studies, and if data availability is similar, such differences in subsystems can be reasonably attributed to the authors’ judgement. When compared to each other, and held against earlier templates such as those shown by Seyhan (2009) and Chowdhury et al. (2013), there is evident consensus on the basic system boundaries and layouts in P MFAs.

A notable aspect, however, is the higher spatial resolution of the model layout through the multi-regional aspect in the Danish case, leading to a multiplication (by the number of regions considered) of all flows quantified on the regional level. This engenders higher requirements on data consistency for these flows. In the example at hand, the resolution in terms of different types of flows (i.e. of different compositions/functions) is lower in the Danish than in the Austrian case, but regional sub-division results in a higher overall number of flows (and the resulting higher risk of data conflicts between connected flows).

3.1.2 Data sources
We divided the data sources used in the MFA models into 6 categories (Figure 3). Typically, flow values entered into STAN are based on data from more than one source. Figure 3 gives an overview of these sources, showing the relative frequency of references from each category being cited as sources for flow values; this includes both references cited for mass (bulk) flows of materials and P concentrations. In total, 399 citations are given in the Danish and 500 in the Austrian study. Generally, flow values are based on several data sources; for Figure 3, each single reference for each flow was counted. Both studies rely strongly on (usually non-peer-reviewed) technical reports from either academic or official sources as well as data from statistical offices, together comprising approximately 80% of the data sources in both cases. The Danish study draws more strongly upon official statistics.

Figure 3. Relative use of data sources in both systems (by the total number of citations). For a complete list of flows with corresponding references, see supplementary data (Tables S.1 and S.2).
3.2 Quantitative discussion

3.2.1 Data consistency

Figure 4 shows an evaluation of data consistency per process: the smaller the deviation of a flow value after data reconciliation from the entered value, the better the consistency of the input data given the balance constraints of the model. Thus, the consistency can provide an indication of the ‘quality’ of the mass balancing. A grouping around processes makes sense in order to identify possible ‘weak spots’ in the system. In the Danish model, the average deviation of flows, as a result of data reconciliation over all flows, amounts to 8.6%, ranging from a 4.4% average of flows connected to, or contained in subsystems within, the Consumption process to 8.9% for flows connected to the Industry and trade process. For 25% of the flows, the extent of data reconciliation is stronger than average. In the Austrian model, the average deviation across all flows is 4.9%, with flows pertaining to Wastewater treatment being the lowest at 2.2%, and flows connected to and contained in Industry and trade being the highest at 7.7%, with 30% of flows showing stronger than average data reconciliation. In light of this, the Austrian study shows a slightly higher overall consistency in mass balancing. Additionally, a higher degree of overdetermination may result in a higher extent of data reconciliation, since it creates more possibilities for data conflicts; the Danish model is overdetermined to a higher degree than the Austrian model, particularly because of the aggregation of regional flows into known total flows on the national level (see Section 2.1).
Figure 4. Deviation of reconciled from entered flow values per process in the main system and associated subsystems. Each column shows the average deviation $\bar{D}$ (see Section 2.3, Eq. 1), in % of the entered values, of all flows connected to or belonging to a subsystem in a process. In the Austrian case, ‘Agriculture & forestry’ consists of 3 discrete processes (Animal husbandry, Crop farming, and Forestry & miscellaneous soils) in the main system that are grouped together here; for Denmark, such differentiation takes place in subsystems.

The extent of measurement adjustments in data reconciliation is influenced (weighted) by the uncertainties of entered data (see Section 2.2). Hence, the more heterogeneous the uncertainties of
individual flows throughout the model, the more reconciliation will concentrate on the most uncertain
flows. It has to be noted that this limits, to some extent, the explanatory power of an average value of
data reconciliation to measure consistency. The more heterogeneous the uncertainties of flows (the
larger the difference between the smallest and highest), the higher will be the changes in the few very
uncertain flows, while less uncertain flows change very little. On average, these changes may be
similar to a system where uncertainties are more homogeneous, and may obscure these underlying
differences.

3.2.2 Uncertainty assessment

The Danish study draws upon the uncertainty concept developed for MFA by Hedbrant &
Sörme (2001). Depending on the data source, all the data are assigned a specific uncertainty level.
Klinglmair et al. (2015) use 5 such levels; each uncertainty level corresponds to a category of data
sources (e.g. data from national statistical offices, information from plant operators, etc.). Each
uncertainty level is associated with an uncertainty factor, which is used to calculate asymmetric
uncertainty ranges for the input data. Since STAN relies on the input data given as the mean value and
standard deviation (i.e. symmetric ranges), the Danish study draws upon the adaptation of this approach
by Laner et al. (2015b), converting these factors into coefficients of variation (CV). In this adaptation,
the product of the mean value and the uncertainty factor is defined as the mean value plus two standard
deviations. Thus, the symmetric interval around the mean corresponds to a 95% confidence interval.
Based on this approach, an uncertainty level (and corresponding CV) is assigned to both material (bulk)
flows and P concentrations therein, resulting in one CV for the P flow. Calculating the uncertainty of a
P flow shown in the model from two uncertain values in this way resulted in 11 different CVs
(uncertainty ranges) for characterising the uncertainty of flow values in the model.
The Austrian study makes use of a more refined approach for data quality assessment and uncertainty characterisation (Laner et al. 2015a) based on the pedigree matrix by Weidema & Wesnæs (1996), developed to express uncertainty due to data imperfection in Life Cycle Assessment (LCA).

The approach quantifies data uncertainty via a matrix assigning discrete scores, expressed as coefficients of variance, to each of the data quality indicators: reliability, completeness, composition, temporal correlation, geographical correlation, and further correlation. CVs are determined based on the evaluation scores, and the overall CV of each datum is calculated as the square root of the summed squares of the individual CVs. Due to the combination of various indicator scores for each datum, a greater variety of uncertainty ranges is obtained for the P flows of the Austrian model in comparison to the Danish model.

While both methods to quantify uncertainties follow a well-defined procedure, in their choice of uncertainty levels associated with the respective quality of data sources, they are to some degree based on the modellers’ best judgement and assessment of data quality. This makes it difficult to compare the uncertainty levels of corresponding flows directly between the two systems: a type of flow may occur in both systems and be based on the same data source, yet have a different uncertainty range assigned in each model. Both studies show a similar reliance regarding which types of data sources they draw upon (see Section 3.1.2). However, the distribution of the uncertainty ranges assigned to flows throughout the model differs considerably. The relative majority of flows are assigned relative standard deviations of 30–50% in the Danish study; for Austria, the relative majority of flows are assigned relative standard deviations of 10–20% (Figure 5). The different distributions of uncertainty ranges in Figure 5 clearly indicate that the uncertainty assessment is not directly comparable between the studies, regardless of a largely similar composition of the data material (Section 3.1.2).
Figure 5. Distribution of uncertainty ranges (coefficient of variation, or CV) as entered in the model: percentage of flows with CV of 5–10%, 10–20%, 20–30%, 30–50%, 50–70%, >70%, respectively.

A high confidence in the quality of input data, and assigning low uncertainty ranges (see Figure 5) to a majority of flows as a consequence, will have a detrimental effect on the consistency of the mass balancing if there is substantial conflict between the inserted mean values for the flows. In such cases, overly strong reconciliation (i.e. the reconciled value is associated with a very low probability, given the probability density function of the input flow) may occur and the data quality assessment, the quantitative uncertainty estimates, and/or the model equations would need to be critically evaluated.
Moreover, data reconciliation in STAN leads to a reduction of uncertainty throughout the system (see Section 2.2); Figure 6a shows the effects on the two systems we examined. In both cases, relative uncertainties after data reconciliation are generally lower, with the exception of 38 flows in the Danish and 34 flows in the Austrian model. The most notable exceptions occur for highly uncertain flows, where reconciliation resulted in much higher relative uncertainties for individual flows. (Data reconciliation always results in a reduction of absolute uncertainty; an increase in relative terms can only be due to a decrease in the flow value).

The further to either side of the normal distribution—as defined by the uncertainty range entered by the user—a reconciled value lies, the less probable it becomes given the original information. Figure 6b compares the entered flow values with the reconciled values. If data reconciliation results in changes of flow values larger than the defined uncertainty range, it is a very strong indication that the original uncertainty characterisation was overly optimistic, and that the reconciliation is too strong. Conversely, if the uncertainty ranges defined by the modeller are much larger than the changes of flow values resulting from data reconciliation, such a pessimistic determination of data quality results in a loss of model precision. Seen in conjunction with Figure 6b, this shows that the uncertainty assessment in both studies resulted in similar uncertainty ranges; the Austrian study, however, is somewhat more optimistic regarding the flows with the smallest uncertainties. In order to ensure a quantification of uncertainties corresponding to the extent of data reconciliation, and to avoid overly optimistic (and pessimistic) uncertainty estimates, an iterative approach to uncertainty (cf. Laner et al. 2014) is advisable. The main goal of such a process is to ensure that all necessary changes of flow values remain within the specified uncertainty ranges.
Figure 6a
Figure 6b
Figure 6. a) Relative uncertainty (coefficient of variation, or CV) for Denmark (DK) and Austria (AT), as entered and after data reconciliation; b) CV as entered, and relative deviation of reconciled flow values from entered values (in % of the entered value). Each point on the x-axis represents a flow, with flows numbered and sorted from the lowest to the highest uncertainty as entered. Flows calculated in STAN (i.e. no flow values entered) are not shown here.

3.2.3 Aggregation effects

In the case of multi-level studies, such as the Danish multi-regional balance, the aggregation or disaggregation of flows may be of particular importance for the degree of data reconciliation and the resulting flow uncertainties. In the Danish study, in- and outflows to and from regional subsystems are aggregated to flows on the national level, which are also known. However, because the flows on the national level are generally based on the same data sources as regional in- and outflows, treating them as independent estimates may not be fully justified. Therefore, the overdetermination related to the flows on the national level, which are aggregates of regional flows (see Table S.2), causes the resulting uncertainties (after reconciliation) to be lower than in the case of treating the national flows as unknowns (see Section 2.3). In the Danish study, the available information was used to enter values for regional flows as well as aggregate flows in the model, though in many cases the original data stemmed from the same sources. The assumption of independent variables in the reconciliation step may have resulted in lower uncertainties than justified by the original level of information. An alternative approach would be to specify the national flows (with identical data sources as the regional flows) as
unknowns. This would reduce the degree of overdetermination and also the potential for conflicting
data. Furthermore, it would result in slightly higher uncertainties of the respective flows.

4. Conclusions

In this study, we showed how choices made in establishing country-scale material balances, i.e.
model properties not resulting directly from the physical properties of the systems studied, can
influence results. While this may not impact the overall conclusions drawn by such studies, such
variance is of particular importance when the values published in an MFA are used in other works. In
other words, metadata matter. Regarding the qualitative analysis of system layouts, we showed that the
choice of data material is closely linked to the layout of the model and vice versa. We identified the
following, more widely applicable, quantitative implications: first, conflicting flow data are practically
inevitable in country-scale MFA or similar material balances of a complexity comparable to the studies
examined in this article. The way in which these conflicts are handled—for example, through the
choice of approach for uncertainty characterisation, propagation, and reconciliation—requires
transparency. This process reflects the authors’ belief in the data underlying the MFA. The analysis of
data reconciliation is an important step to judge the consistency and quality of an MFA model, and to
assess both the data and the system. The average extent to which reconciled data deviate from input
data in an MFA model provides a useful measure of the data consistency and the quality of the mass
balances. The higher the inconsistency between the data given the mass balance constraints, the more
data reconciliation becomes analogous to fitting the proverbial square peg in a round hole. A second
implication of uncertainty assessment being subjective is that uncertainties can be evaluated relative to
the uncertainties of other flows in the same model; however, they may not be comparable to the
uncertainties specified for a similar flow in another model due to different assumptions and choices
made in each study. Third, an iterative approach to quantifying the uncertainty of flow values, i.e. adjusting uncertainty ranges to more closely correlate to the extent of data reconciliation, can ensure the representation of data uncertainty in compliance with the level of information available, while also avoiding overly optimistic or pessimistic estimates of data quality. Finally, and linked to uncertainty assessment, the treatment of multi-level flows as independent or not, has an effect on the resulting uncertainty of the flows. This is of particular interest if regional flows are aggregated to national flows, with data for flows on both levels originating from the same sources. As a concluding remark, we stress the need for a transparent system definition and data characterisation in MFA. To better comprehend the results of MFA studies, the authors have the responsibility to document their choices, especially those that are not immediately evident from the model results. This would enable consistent comparisons between key indices (such as recovery rates, etc.). In this article, we have put forward a quasi-standard approach to transparently and systematically document and analyse the choices underlying an MFA model.

**Acknowledgements**

This research was funded by the Danish Council for Strategic Research (Innovation Fund) as part of the Integrated Resource Management & Recovery (IRMAR) project at the Technical University of Denmark. We would like to thank Oliver Cencic (Vienna University of Technology) for his valuable inputs.
References


