Title
Characterizing Uncertainties in Life Cycle Assessment

Author
Qin, Yuwei

Publication Date
2019

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Santa Barbara

Characterizing Uncertainties in Life Cycle Assessment

A dissertation submitted in partial satisfaction of the

requirements for the degree Doctor of Philosophy

in Environmental Science and Management

by

Yuwei Qin

Committee in charge:

Professor Sangwon Suh, Chair

Professor Sarah E. Anderson

Professor Arturo A. Keller

March 2019

The dissertation of Yuwei Qin is approved.

Sarah E. Anderson

Arturo A. Keller

Sangwon Suh, Committee Chair

March 2019

Characterizing Uncertainties in Life Cycle Assessment

To my parents and grandparents.

ACKNOWLEDGEMENTS

When I look back on my PhD studies, I feel so blessed to have had a great many people helping and encouraging me along the way. My advisor, Sangwon Suh, is a very nice and open gentleman with all the virtues I would expect from a PhD mentor and allowed me to explore my own research interests. His background in environmental engineering makes him both theoretical and practical, and he has an in-depth grasp of the nature of many problems. Apart from my advisor, I also received a great amount of help from other mentors. Here, I would like to thank Sarah Anderson and Arturo Keller, who guided me greatly during my PhD. Their extensive knowledge and experience in environmental science and management let me realize the importance of understanding multidisciplinary aspects of environmental problems.

This work would not have been possible without collaboration and advice from my fellow colleagues of Suh's group and the CLiCC project. In particular, the mentorship of post-doc researcher Stefano Cucurachi was invaluable when I started my PhD. Many thanks to Mengya Tao, Runsheng Song, Jessica Perkins, Elise Wall, and Dingsheng Li. I would like to specially acknowledge Yuxiong Huang and Kendra Garner for their help and guidance to my PhD study. I want to thank my interns, Jiaheng Tang and Mingyu Lin, who helped a great deal in my work.

Finally, I want to express my greatest gratitude to all my friends and family, especially my husband who always inspires and encourages me in life. I would like to deeply thank my parents and grandparents for giving me unwavering support throughout my life and encouraging me to pursue my dream.

VITA OF YUWEI QIN
March 2019

EDUCATION

Ph.D. in Environmental Science and Management **2018**

Bren School of Environmental Science & Management

University of California, Santa Barbara

Dissertation: *Characterizing Uncertainties in Life Cycle Assessment*

Master of Environmental Science and Management **2014**

Bren School of Environmental Science & Management

University of California, Santa Barbara

Bachelor of Science in Environmental Policy and Management *(cum Laude)* **2012**

Minor in Environmental Economics

School of Environment and Natural Resources

The Ohio State University

PUBLICATIONS

1. Garner, K. L., **Qin, Y.**, Cucurachi, S., Suh, S., & Keller, A. A. (2018). Linking Exposure and Kinetic Bioaccumulation Models for Metallic Engineered Nanomaterials in Freshwater Ecosystems. *ACS Sustainable Chemistry & Engineering,* 6(10), 12684-12694.
2. **Qin, Y.**, Suh, S. (2018). Does the Use of Pre-calculated Uncertainty Values Change the Conclusions of Comparative Life Cycle Assessments? – An Empirical Analysis**.** *PLoS ONE,* 13(12), e0209474.
3. Zhang, H., HUANG, Y., Gu, J., Keller, A. A., **Qin, Y.**, Bian, Y., ... & Zhao, L. (2019). Single Particle ICP-MS and GC-MS Provide New Insight into the Forming Mechanisms for AgNPs Green Synthesis. New Journal of Chemistry.
4. Suh, S., **Qin, Y**. (2017). Pre-Calculated LCIs With Uncertainties Revisited. *International Journal of Life Cycle Assessment*, 22(5), 827-831.
5. Song, R., **Qin, Y.**, Suh, S., & Keller, A. A. (2017). Dynamic Model for the Stocks and Release Flows of Engineered Nanomaterials. *Environmental Science & Technology*, 51(21), 12424-12433.
6. Qin, Y., Suh, S. (2016). What Distribution Function Do Life Cycle Inventories Follow? *International Journal of Life Cycle Assessment*, 22(7), 1138-1145.

7. Cucurachi, S., Yang, Y., Bergesen, J. D., **Qin, Y.**, Suh, S. (2016). Challenges in Assessing the Environmental Consequences of Dietary Changes. *Environment Systems and Decisions*, 36.2 (2016): 217-219.

HONORS and AWARDS

2015 – 2017	Robert L. Boughton, Jr. Educational and Scholarship Fund, UCSB
2016	Best Poster, Bren PhD Symposium, UCSB
2013	Tim Cohen Summer Internship Fellowship, UCSB
2011 – 2012	Buford M. and May Scott Teater Scholarship, OSU
2010 – 2012	Barnebey Family Scholar, OSU
2009 – 2011	Dean's List, OSU
2009 – 2012	International Undergraduate Scholar, OSU
2008	Best Volunteer Award, Jane Goodall's Roots & Shoots

ACADEMIC RESEARCH

Graduate Student Researcher 9/2014-12/2018

Bren School - University of California, Santa Barbara

- Worked with Dr. Sangwon Suh on life cycle environmental impact assessment
- Conducted Monte Carlo simulation and statistical analysis on large-scale environmental data, ecoinvent database, to determine the probability distribution to best describe life cycle inventories
- Worked on a EPA-funded project of rapidly assessing environmental and human health impact of new chemicals
- Developed a web-based survey to collect expert opinions for the use of Pedigree approach to quantify uncertainty in life cycle impact assessment
- Trained and managed 3 long-term graduate interns for creating a web-based tool of assessing chemical life cycle environmental impact

Visiting PhD Student 7/2017-9/2017

Department of Energy Resources Engineering - Stanford University

- Developed optimization model to predict consequential life-cycle impacts of agricultural production under three policy scenarios

- Predicted future production situation by analyzing historical trends of land use and water use for agricultural production

Visiting PhD Student 6/2016-9/2016

School of Environment – Tsinghua University

- Helped streamline the data collection process for 1000+ chemicals
- Analyzed the environmental data to determine the uncertainty in electricity-generation processes

---

TEACHING and MENTORING EXPERIENCE

Graduate Teaching Assistant 3/2015-6/2015

ESM 273 – Life Cycle Assessment

Bren School - University of California, Santa Barbara

- Assisted professors and students in modeling exercises using GaBi
- Led linear algebra workshops and provided students with one-on-one tutoring
- Evaluated and graded examinations, assignments, and term projects

Group Project Mentor 9/2015-6/2016

CarbNewt

Bren School - University of California, Santa Barbara

- Mentored a Master's Group Project on achieving carbon neutrality at UCSB by 2025 – a critical analysis of technological and financial strategies
- Provided technical consultant on conducting life cycle assessment and cost-benefit analysis for emission reduction project

---

WORK EXPERIENCE

Life Cycle Assessment Data Analyst Intern 4/2013 – 6/2014

Industrial Ecology Research Services - IERS LLC., Goleta, CA

- Assisted with the project of reducing environmental impacts of U.S. government spends
- Participated in conducting sustainability analysis for the project associated with green buildings using hybrid input-output LCA database, CEDA
- Contributed to data collection, analysis, and visualization
- Identified major contributors to environmental degradation

Data Manager 4/2013 – 6/2014

DNV KEMA Energy & Sustainability

Bren School, University of California, Santa Barbara

- In charge of data collection and analysis and helped the management of project budget, timeline and deadline
- Collected local governmental data of current Climate Action Plans through survey and interviews as part of the Master's Thesis Group Project process
- Performed statistical tests and analysis to determine the key contributors to the success and failure of their Climate Action Plan
- Contacted local climate plan managers to identify common barriers to implementation
- Analyzed the best practices for implementing Climate Action Plans in California

---

TALKS

1. **Qin, Y**., Suh, S. Does the Use of Pre-calculated Uncertainty Values Change the Conclusions of Comparative LCAs? *Gordon Research Conference: Proceedings, Les Diablerets, Switzerland, 2018*
2. **Qin, Y**., Suh, S. The Debate on the Use of Pre-calculated Uncertainty Values. *American Center for Life Cycle Assessment XVII: Proceedings, Portsmouth, New Hampshire, 2017*
3. **Qin, Y**., Garner, K., Song, R., Tao, M., Suh, S. Chemical Life Cycle Collaborative: A New Web-Based Tool for Assessing Chemicals' Life Cycle Impacts. *American Center for Life Cycle Assessment XVII Pre-conference Workshop: Proceedings, Portsmouth, New Hampshire, 2017*
4. **Qin, Y**., Garner, K., Song, R., Suh, S. Rapid Assessment of Chemical Risks and Life-cycle Impacts. *Life Cycle Management Conference: Proceedings, Luxembourg, Luxembourg, 2017*
5. **Qin, Y**., Yang, Y., Cucurachi, S., Bergesen, J., Suh, S. How to Model Marginal Production in Life Cycle Assessment?: A Case Study of Potato Production in the U.S. *International Society for Industrial Ecology Conference: Proceedings, Chicago, Illinois, 2017*
6. **Qin, Y**., Cucurachi, S., Garner, K., Li, D., Song, R., Tao, M., Keller, A., Suh, S. Monte Carlo-based Uncertainty and Global Sensitivity in Environmental Modelling. *CLiCC Chemical LCA Workshop: Proceeding, Santa Barbara, California, 2016*
7. **Qin, Y**., Yang, Y., Cucurachi, S., Suh, S. Where Would the Potato be Grown?: Understanding the Potential Consequences of Large-scale Dietary Shift. *Life Cycle Assessment XV: Proceedings, Charleston, South Carolina, 2016*

8. Cucurachi, S., **Qin, Y**., Garner, K., Keller, A., Suh, S. Joint Monte Carlo Uncertainty And Global Sensitivity Analyses For a Novel Multi-Media Fate And Transport Model. *Life Cycle Assessment XV: Proceedings, Charleston, South Carolina, 2016*
9. Song,. R., **Qin, Y**., Bachmann, M., Cucurachi, S., Suh, S. Strategy for Developing Gate-to-Gate Chemical Life Cycle Inventory. *Life Cycle Assessment XV: Proceedings, Charleston, South Carolina, 2016*
10. **Qin, Y**., Suh, S. What Distribution Function Do LCIs Follow? *Gordon Research Conference: Proceedings, Stowe, Vermont, 2016*
11. Cucurachi, S., **Qin, Y**. Characterizing and Communicating Uncertainty in Quantitative Sustainability Assessment: the Case of LCA Results. *Gordon Research Conference: Proceedings, Stowe, Vermont, 2016*
12. **Qin, Y**., Suh, S. Using Monte Carlo Simulation to Find the Best Distribution Function in Life Cycle Inventories. *Life Cycle Assessment XV: Proceedings, Vancouver, Canada, 2015*
13. **Qin, Y**., Cucurachi, S., Garner, K., Li, D., Song, R., Tao, M., Keller, A., Suh, S. Uncertainty Assessment in CLiCC Project. *Chemical Life Cycle Collaborative (CLiCC) Annual Program Review: Proceedings, Santa Barbara, California, 2015*

PROFESSIONAL ACTIVITIES

ISIE Student Chapter – International Society for Industrial Ecology, 2016 - present
Student Board – Chinese Society for Industrial Ecology, 2016 – 2018

ABSTRACT

Characterizing Uncertainties in Life Cycle Assessment

by

Yuwei Qin

Life cycle assessment (LCA) aims to support corporate and public policy decisions by quantifying the environmental performance of a product. Understanding uncertainties in LCA results is therefore important for making informed decisions. Monte Carlo simulation (MCS), which uses random samples of the parameters from a pre-determined probability distribution, has been widely utilized to characterize uncertainties in LCA. However, as the size of an LCA database grows, running a full MCS is becoming increasingly challenging. Furthermore, the uncertainty literature in LCA has focused on life cycle inventory (LCI), while the uncertainties from the remaining steps—including characterization, normalization, and weighting—have not been addressed, despite their perceived relevance in overall uncertainty characterization in LCA.

The objectives of my dissertation are: (1) to develop a new method to improve the computational efficiency of large-scale MCS in LCA, (2) to empirically test the reproducibility of comparative decisions obtained using the method, and (3) to develop and test an analytical method to decompose the overall uncertainty in LCA into its constituents. The new method for uncertainty characterization in LCA involves pre-calculating and

storing the distribution profiles of the most widely used LCA database, ecoinvent. Using parallel computing, I have generated the distribution functions for 22 million life cycle inventory (LCI) items of the database. I then tested 20,000 randomly selected comparative LCI cases, and showed that pre-calculated uncertainty values can be used as a proxy for understanding the uncertainty and variability in a comparative LCA study without compromising the ability to reproduce the comparative results.

One key barrier to conducting uncertainty analysis in LCA occurs in life cycle impact assessment (LCIA), an important step of LCA calculation followed LCI phase, because characterization models for LCIA do not typically provide uncertainty information for the input parameters and lack detailed information about the relationships between those inputs. A Pedigree matrix for characterization factors in LCIA was developed to fill in the gap in the uncertainty characterization in LCA. Expert opinions on the use of the Pedigree method in estimating uncertainty in LCIA and the Pedigree scores for both LCI and LCIA were collected through an online survey.

Finally, I demonstrated a new method to decompose the overall uncertainties of an LCA result over the contributing factors including those from LCI, characterization, normalization, and weighting, which are the steps involved in LCA calculation. To do so, I incorporated the logarithmic mean Divisia index (LMDI) decomposition method into MCS to parse out the overall uncertainty into its constituents.

This research helps improve the efficiency and analytical power of uncertainty analysis in LCA. The findings can be applied to other problems outside of LCA that utilize MCS.

TABLE OF CONTENTS

Chapter 1. Introduction

Life cycle assessment (LCA) is a method to analyze the environmental impacts of products throughout products' life cycles (ISO 2006). Traditionally, LCA studies only include deterministic values in results. However, sound decision-making can benefit from an understanding of the magnitude of the uncertainty of LCA results. For example, when making comparisons among products, ignoring uncertainty may lead to a misleading decision if the distributions of the two LCA results significantly overlap, although their deterministic values favor one versus another (Heijungs and Kleijn 2001; Geisler et al. 2005; Sugiyama et al. 2005; Finnveden et al. 2009). Therefore, many LCA studies have implemented uncertainty analysis for sound decision-support (Hertwich and Hammitt 2001; Huijbregts et al. 2003; Basson and Petrie 2007; Cellura et al. 2011; Clavreul et al. 2012; Noshadravan et al. 2013).

Though the importance of analyzing uncertainty in LCA is broadly accepted, uncertainty assessments are not yet standard practice in LCA. This dissertation aims to discuss the current challenges of performing uncertainty analysis and contribute to the literature of improving the feasibility of conducting uncertainty analysis in LCA. A fast, feasible, efficient uncertainty assessment approach for LCA practitioners was developed to save computation time and cost, and the comparison between the proposed approach and the traditional approach was provided in the thesis. Uncertainty estimation for characterization factors was also generated to incorporate the uncertainty from the impact assessment phase.

1.1. Recent development of uncertainty analysis in life cycle assessment

The concept of uncertainty in LCA was first discussed in a workshop of Society of Environmental Toxicology and Chemistry (SETAC) in 1992 in the context of data quality (Fava 1994). Recognizing the significance of incorporating uncertainty, the LCA community formed the SETAC LCA working group on data availability and data quality in the early 90s. Heijungs (1996) illustrated how uncertainty is propagated from input parameters of an LCA model to its outputs. Weidema and Wesnæs (1996) addressed the problem of data quality concerns by introducing the pedigree method, which has been widely incorporated into various Life Cycle Inventory (LCI) databases to-date. European Network for Strategic Life-Cycle Assessment Research and Development (LCANET) has suggested making uncertainty quantification a top research priority. During those early years, many efforts were devoted to the setting-up the scheme for data quality indicators. Based on such efforts, Huijbregts (1998) established a framework for parameter uncertainty analysis. Subsequently, a framework for quantifying data quality in LCI was also developed.

More recently, the literature focused more on typologies of uncertainty and approaches to handling uncertainty (Björklund 2002; Huijbregts 2002; Baker and Lepech 2009). In general, two types of uncertainties have been distinguished: stochastic uncertainty (due to inherent randomness) and epistemic uncertainty (due to lack of knowledge) (Clavreul and Guyonnet 2013; Heijungs and Lenzen 2014). Among them, stochastic uncertainty has been the focus of many LCA studies, while the literature on epistemic uncertainty in LCA is scarce (Laner et al. 2014; Gavankar and Suh 2014). According to the survey of 24 LCA studies that incorporated uncertainty analysis, parameter uncertainty is the most addressed one compared to model and scenario uncertainties, and sampling method is the most frequently used technique to quantify uncertainty (Lloyd and Ries 2008).

Furthermore, measurement errors in input data, choices of system boundaries, underlying assumptions, model incompleteness, all contribute to the reliability and accuracy of LCA results (Clavreul et al. 2012, 2013). In particular, the LCI and life cycle impact assessment (LCIA) phases of LCA are the most data- and calculation-intensive phases, involving many model and data assumptions that could cause errors in LCA results (Huijbregts 1998; Heijungs and Huijbregts 2004; Mendoza Beltran et al. 2018).

1.2. Methods of addressing uncertainties in life cycle assessment

Two types of techniques for addressing uncertainties in life cycle assessment have emerged: the sampling method and the analytical approach (Ross et al. 2002; Heijungs and Frischknecht 2004; Clavreul and Guyonnet 2013; Jung et al. 2013). Heijungs and Huijbregts (2004) presented a review of four general uncertainty treatments for stochastic uncertainty and Ciroth and colleagues (2004) proposed a method for uncertainty calculation.

Among the various statistical methods, Monte Carlo simulation (MCS) is the most commonly used approach, which relies on pre-defined probability distributions and runs the model repeatedly for a sufficiently large number of times to allow statistical analysis of the results (Huijbregts 1998; Sonnemann et al. 2003; Peters 2007; Hung and Ma 2009; Imbeault-Tétreault et al. 2013; Heijungs and Lenzen 2014; Prado-Lopez et al. 2014; Vinodh and Rathod 2014; von Pfingsten et al. 2017). Many studies included probability distribution in uncertainty analysis through MCS (Maurice et al. 2000; McCleese and LaPuma 2002; Sonnemann et al. 2003; Hung and Ma 2009; Cucurachi and Heijungs 2014). For example, Noshadravan et al and Gregory et al performed MCS to compare two pavement designs using the distributions of expected LCA results (Noshadravan et al. 2013; Gregory et al. 2016).

MCS usually takes three steps: (1) extract distribution functions of the raw data, which are the data on unit process-level intermediate flows and elementary flows, (2) create random samples based on the probability distributions of the raw data, and (3) iterate the process and collect the sample results. With the help of advancement in computer hardware and software, MCS of large datasets became viable (Gentle 2013). Many professional LCA software like SimaPro and OpenLCA can now perform uncertainty analysis using Monte Carlo simulations for sampling foreground and background LCI data (SimaPro 2016; OpenLCA 2018). For example, the most widely used data source, ecoinvent database, includes distribution information for 90% of the unit process data (Weidema et al. 2013; Qin and Suh 2017).

1.3. The challenge in performing uncertainty analysis in LCA

Though a growing number of LCA scholars and practitioners now address the uncertainty issue in LCA (Cooper et al. 2012; Sills et al. 2012; Groen et al. 2014), quantitative uncertainty assessments are only rarely performed in practice. Two main challenges in performing uncertainty analysis in LCA that will be addressed in this dissertation are (1) the long computational time of Monte Carlo simulation for both background and foreground data and (2) the near complete lack of uncertainty information for characterization factors. The first challenge of computation time will be addressed in Chapter 2 and 3, and the second challenge associated with uncertainty for characterization factors will be addressed in Chapter 4.

1.3.1. Computation time of running Monte Carlo simulation

Although Monte Carlo simulation can generate the range of possible results to support corporate and public policy decision, performing a full MCS is becoming a computational burden to lay LCA practitioners. In LCA, performing an MCS using fully dependent sampling typically involves repeated inversion of a technology matrix for each run. As the dimension of the technology matrices used in LCA databases grows, however, time spent on MCS would greatly increase. The time required for each matrix inversion in a modern computer is known to have an order of $n^{2.73}$ time complexity, where n is the dimension of a irreducible, invertible square matrix (Stothers 2010; Williams 2012; Wu et al. 2014), which is generally used in LCAs (Suh and Heijungs 2007). This means that doubling the dimension of a technology matrix increases the computational time at least 4.8 times. Given that the number of processes in LCI databases continues to grow, running full MCSs will increasingly become a challenge.

Several studies considered parameter uncertainty and applied MCS using a fully dependent sampling approach, and the running time for MCS varied based on the number of processes included in their studies and computing power of their computers. Imbeault-Tetreault et al. (2013) performed MCS for an LCA case study with nearly 900 unit processes using fully dependent sampling and compared two scenarios around the use of Global Position System (GPS). The fully dependent sampling used by Imbeault-Tetreault et al. required several hours to complete the MCS. Henriksson et al. (2015b) conducted 1,000 Monte Carlo simulations with fully dependent sampling for a comparative LCA of Asian aquaculture products. Ren et al. also performed fully dependent sampling using OpenLCA, which took the team 16 hours for 1,000 Monte Carlo simulation runs on a personal computer.

1.3.2. Uncertainty in characterization factors

The other challenge in performing uncertainty analysis in LCA is due to the lack of uncertainty information for characterization factors. Uncertainty assessments have largely failed to consider the contribution to the output uncertainty of other phases of LCA, and of the LCIA phase in particular (Lloyd and Ries 2007; Reap et al. 2008). Since only LCI data regularly contains uncertainty information, LCA studies that do include uncertainty analysis, typically only focus on the LCI phase (Heijungs 1996; Maurice et al. 2000; Sonnemann et al. 2003; Gavankar et al. 2014; Scherer and Pfister 2016; von Pfingsten et al. 2017). The most widely used LCI database, ecoinvent (Frischknecht and Rebitzer 2005), includes uncertainty values, for example, geometric standard deviation for lognormal distribution, for 62.7% of its unit process data in ver. 3.4 (Wernet et al. 2016).

However, characterization models for LCIA do not typically provide uncertainty information for the input parameters and lack detailed information about the relationships between those inputs. Such lack of information limits the possibility of researchers and practitioners to regularly assess characterization models when conducting an uncertainty analysis over the full scale of an LCA study (Hung and Ma 2009; Noshadravan et al. 2013; Henriksson et al. 2015; Gregory et al. 2016). As a result, the influence of characterization models on LCA results is largely unknown, and LCA results that only contain an analysis of parameter uncertainty may be misleading for decision-makers (Huijbregts 1998).

1.4. The gap in uncertainty analysis in life cycle assessment

1.4.1. Distribution type for LCI results

When using Monte Carlo simulation, the shape of the distribution in the aggregate LCI results becomes an important issue for efficient storage of such data. In the study of waste incinerators by Sonnemann et al. (2003) the distribution of aggregate LCI results from Monte Carlo simulations looks like a lognormal distribution. Several reports suggest that the lognormal distribution could be an appropriate distribution type in inventory data, risk assessment, and impact pathway analysis because lognormal distributions can avoid negative values for emissions and impacts (Hofstetter 1998; Frischknecht et al. 2004).

Many LCA studies used lognormal distribution for LCI results (Rosenbaum et al. 2004; Hong et al. 2010; Ciroth et al. 2013; Imbeault- Tétreault et al. 2013; Heijungs and Lenzen 2014); however, such an assumption has not been empirically tested in the LCA literature. In the literature, it was shown that the product of lognormally distributed data result in a lognormal distribution (Limpert et al. 2001). However, there is no theoretical underpinnings on the types of distribution for the product of two matrices of which the data are lognormally distributed, which is basically a set of linear combinations of products between lognormally distributed data (Hong et al. 2010). Furthermore, LCA data exhibit not only lognormal distributions but also other types of distributions such as normal and triangular distributions, for which the distribution of the products cannot be determined analytically. The objective of the Chapter 2 is to determine the distribution type of LCI result using empirical data and the overlapping coefficient technique.

1.4.2. The use of pre-calculated uncertainty values in comparative LCA studies

As Chapter 2 proposed, using pre-stored distributions for LCI could significantly reduce computation time for LCA uncertainty analysis. However, it remains a question whether the additional errors due to the use of pre-calculated uncertainty values are small enough to maintain the conclusions of a comparative study, and, if not, what the odds of misinterpreting a comparative LCA results due to the use of pre-calculated uncertainty values are. In particular, the use of pre-calculated uncertainty values does ignore the presence of internal dependency within a technology matrix (Heijungs and Lenzen 2014). Henriksson and colleagues (2015) highlighted the importance of dependent sampling in understanding the distribution of comparative LCA results. There are two main issues to consider. First, when performing an MCS, a data point from the same process commonly used by the two products under comparison can be perturbed independently. In principle, however, they should be perturbed in the same direction and magnitude, which is referred to as 'dependent sampling.' Second, in a comparative LCA setting, the distribution of the difference between the results by the two product systems being compared helps distinguish the real difference of the two results.

Therefore, the objective of Chapter 3 is to empirically test the hypothesis that the use of partially independent sampling using pre-calculated uncertainty values in a life cycle inventory alters the conclusion that would have been drawn if the uncertainty values are sampled dependently.

1.4.3. The uncertainties of characterized LCA results

Even though empirical evidence is limited, it has been claimed that the impact assessment phase of LCA is the phase that contributes the most to the uncertainty of a LCA result (Owens 1996, 1997; Clavreul et al. 2012). Characterization factors are calculated from simplified models of complex interacting physical and chemical systems and do often require modelers to resort to a process of linearization of non-linear relationship to fit a characterization model to the linear computational structure of LCA (Cucurachi et al. 2016). Characterization models, thus, are likely to carry large model uncertainties (Huijbregts 1998).

For most LCA studies that have included quantitative uncertainty assessment, only uncertainties during the inventory phase were considered. Only a few studies have considered quantitative uncertainty from the characterization phase (Cellura et al. 2011; Hauschild et al. 2013). For example, the uncertainties from the characterization factors for the global warming and acidification impact categories were assessed in the study of two types of roof gutters (Huijbregts 1998). Later, the same author and his colleagues further performed an uncertainty analysis including parameter, scenario, and model uncertainties involved in characterization factors for two insulation options for a Dutch dwelling (Huijbregts et al. 2003). The most other LCA studies lacked the consideration of the uncertainties from the characterization factors in their uncertainty analysis due to the unavailability of quantitative uncertainties in characterization factors. Therefore, there is a need to develop a method to estimate the uncertainty in the LCIA phase, which is mainly from the characterization model, in order to consider the uncertainty during both inventory and impact assessment in the absence of measurement data.

To fill in the gap in the uncertainty characterization in LCA, a Pedigree matrix for the characterization phase of LCIA was developed in Chapter 4. New indicator scores were defined based on the Numeral Unit Spread Assessment Pedigree (NUSAP) literature and environmental risk assessment common practice (Funtowicz and Ravetz 1990; Jaworska and Bridges 2001; Van Der Sluijs et al. 2005; Ragas et al. 2009). We used expert elicitation method to collect expert judgements of the use of Pedigree method in estimating uncertainty in characterization factors and the Pedigree scores for both LCI and characterization factors based on the experience and knowledge of respondents.

1.4.4. Decomposing LCA uncertainty using logarithmic mean Divisia index (LMDI) method

Some LCA scholars claimed that the impact assessment phase of LCA has a larger influence on the uncertainty of a LCA result than the inventory phase but they didn't provide any empirical analysis or evidence (Owens 1996, 1997; Clavreul et al. 2012). Because most of the uncertainty analysis in LCA failed to consider the uncertainties from the characterization factors, it is difficult to understanding the contribution of LCI and characterization factors to the uncertainty in LCA results (Maurice et al. 2000; Björklund 2002; Lloyd and Ries 2008). Even if we have the uncertainty for characterization factors, the current sensitivity analysis that is used to understand the importance of the parameters to the uncertainty cannot tell us the exact contribution of each parameter to the uncertainty. Chapter 5 provided a methodology for quantifying the contribution of each LCA phase to the overall LCA uncertainty using logarithmic mean Divisia index (LMDI) method.

A case study on natural gas production was used to demonstrate LMDI method by focusing on two impact categories: global warming and USETox human health impacts. The proposed method uses the simulations from MCS to calculate the difference between the simulated and deterministic category indicator results as the changes in the characterized results for each run. Applying the LMDI decomposition method, we can calculate the decomposition of the differences, which returns the contribution of each factor to the difference of the run. Then, the statistical properties of the contributions by each factor can be analyzed after a large number of MCS runs.

This work is the first attempt to apply the technique of decomposition analysis using the MCS samples to decompose the uncertainty of LCA from LCI, characterization, normalization, and weighting phases. The procedure proposed in this study can also serve as a practical guide for future LCA practitioners to use LMDI approach to decompose the effect of each intermediate LCA step to the uncertainty of final LCA output.

1.5. Organization of the dissertation

The dissertation mainly focuses on the characterization of uncertainty in life cycle assessment. The objective is to develop a fast, feasible, efficient uncertainty assessment approach for LCA practitioners to save computation time and cost when running Monte Carlo simulations and provide uncertainty estimation for characterized life cycle assessment results.

Chapter 1 provides an overview of the current development in uncertainty analysis in LCA including the methods of treating uncertainties in LCA and the gaps in the literatures, especially in the area of life cycle impact assessment.

In Chapter 2, a method of storing distributions as uncertainty information for life cycle inventory was proposed for the purpose of saving computation time and cost for running uncertainty analysis. The study investigates the probability distribution that best describes LCI results, which is the first attempt to generate the distribution profiles for the entire aggregate LCIs of Ecoinvent database. This chapter has been published in International *Journal of Life Cycle Assessment* (Qin and Suh 2016).

Currently, three methods were proposed to sample inventory data; (1) the fully dependent sampling for both background and foreground data (Henriksson et al. 2015), (2) the use of pre-calculated uncertainty values proposed in Chapter 2, and (3) the use of pre-stored dependent LCI samples (Lesage et al. 2018). The three methods can generate similar results for uncertainty analysis of non-comparative LCAs (Suh and Qin 2017). However, for comparative LCA studies, the second approach, the use of pre-calculated uncertainty values due to independent sampling may overestimate the uncertainty than the fully dependent sampling which raised in Heijungs et al's paper (2017). In the study of 20,000 randomly selected comparative LCI cases, the results showed that pre-calculated uncertainty values can be used as a proxy for understanding the uncertainty and variability in a comparative LCA study especially when adequate computational resources are lacking. Then, the hypotheses raised in Heigjungs's paper were tested in the Chapter 3. The study examines whether the use of partially dependent sampling using pre-calculated uncertainty values in a life cycle inventory alters the conclusion that would have been drawn if the uncertainty values are sampled dependently.

Chapter 2 and 3 addressed the issue of uncertainty analysis occurred in the LCI phase, while Chapter 4 discussed the use of Pedigree approach to estimate uncertainty for characterization factors in LCIA phase which is an important step for LCA calculation after the LCI phase. A Pedigree matrix for characterization factors in LCIA was developed to fill in the gap in the uncertainty characterization in LCA. Expert opinions regarding the use of Pedigree method in estimating uncertainty in characterization factors and the Pedigree scores for both LCI and characterization factors were collected through an online survey.

After the uncertainty values are determined for LCI and LCIA phases, what is the influence of the uncertainty from each LCA step on the final LCA result? Chapter 5 demonstrates the application of logarithmic mean Divisia index (LMDI) method to quantify the contribution of each LCA phase to the overall LCA uncertainty.

Chapter 6 concludes the dissertation.

Chapter 2. What Distribution Do Life Cycle Inventory Follow?[1]

2.1. Introduction

Assessing uncertainty in Life Cycle Assessment (LCA) is important for understanding reliability and robustness of the results in the context of decision making (Finnveden et al. 2009). When making comparisons among products, ignoring uncertainty may lead to a misleading decision if the distributions of the two LCA results significantly overlap, though their deterministic values favor one versus another (Heijungs and Kleijn 2001). Therefore, many LCA studies have included uncertainty analysis for making sound decisions (Hertwich and Hammitt 2001; Cellura et al. 2011).

In addition to the development of frameworks and methodologies of uncertainty assessment, a number of empirical studies have implemented uncertainty analysis in LCA. Geisler et al. (Geisler et al. 2004) applied uncertainty assessment to a case study of plant-protection products using generic uncertainty factors for inventories. Huijbregts and his colleagues (2003) performed uncertainty quantification considering parameter, scenario, and model uncertainties in a comparative study of building's insulation options. Many studies included probability distributions in uncertainty analysis through Monte Carlo Simulation (Maurice et al. 2000; McCleese and LaPuma 2002; Sonnemann et al. 2003; Hung and Ma 2009; Cucurachi and Heijungs 2014).

When using Monte Carlo Simulation (MCS), the shape of distribution in the aggregate LCI results becomes an important issue for efficient storage of such data. In a study of waste

[1] This chapter was published in the International Journal of Life Cycle Assessment. Qin, Y., & Suh, S. (2017). What distribution function do life cycle inventories follow?. *The International Journal of Life Cycle Assessment*, *22*(7), 1138-1145.

incinerators by Sonnemann et al. (2003) the distribution of aggregate LCI results from Monte Carlo simulations looks like a lognormal distribution. Several reports suggest the lognormal distribution could be an appropriate distribution type in inventory data, risk assessment, and impact pathway analysis because lognormal distribution can avoid negative values for emissions and impacts (Hofstetter 1998; Frischknecht et al. 2004). Many LCA studies following Sonnemann et al. (2003) assumed that LCI results are lognormally distributed (Rosenbaum et al. 2004; Hong et al. 2010; Ciroth et al. 2013; Imbeault-Tétreault et al. 2013; Heijungs and Lenzen 2014). However, such an assumption has not been empirically tested in the LCA literature. In the literature, it was shown that the product of lognormally distributed data result in a lognormal distribution (Limpert et al. 2001). However, there is no theoretical underpinnings on the types of distribution for the product of two matrices of which the data are lognormally distributed, which is basically a set of linear combinations of products between lognormally distributed data (Hong et al. 2010). Furthermore, LCA data exhibit not only lognormal distribution but also other types of distribution such as normal and triangular distributions, of which distribution types of the products cannot be determined analytically.

This study aims to determine the probability distribution that best describes LCI results. The paper is the first attempt to generate the distribution profiles for the entire aggregate LCIs of ecoinvent version 3.1. In this study, we performed MCS to simulate random samples of unit process data and to estimate the distribution profiles of LCI results. We tested the hypothesized distributions of LCIs using the overlapping coefficient method, and identified the most suitable distribution type to present LCIs.

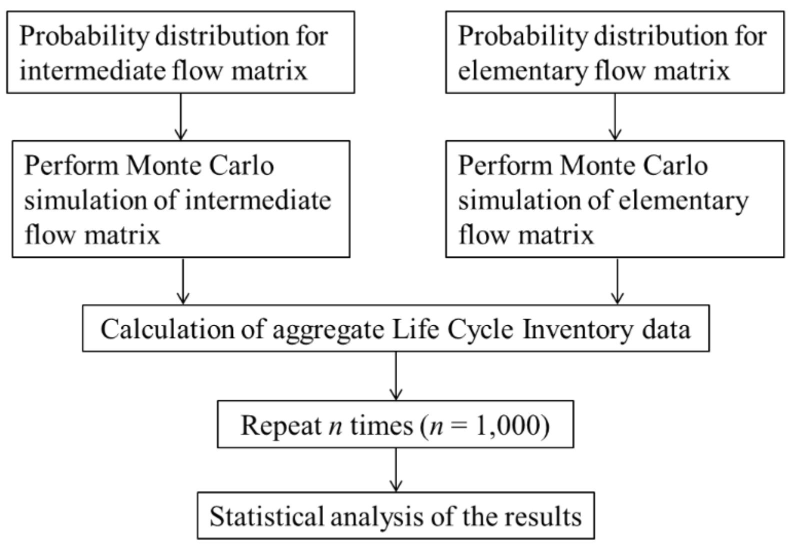
In the next section, the ‘method and data’ used in this study is presented, followed by a ‘results and discussion’ section. In the ‘conclusions’ section, the main findings are presented and a set of recommendations are discussed.

2.2. Method and data

2.2.1. Monte Carlo simulation

In this study, MCS is used to create the distribution of each aggregate LCI result from the entire ecoinvent data v3.1. Figure 1 demonstrates the procedure for the statistical analysis used in this study.

Figure 1. Monte Carlo procedure for uncertainty assessment of aggregate LCI



Each and every input parameter for calculating LCI results is considered as a stochastic parameter. For one iteration, every unit process data in intermediate flow matrix A and elementary flow matrix B are reconstructed based on their distribution functions. Aggregate LCI results are calculated through the equation, $M = BA^{-1}$ (Heijungs and Suh 2002).

This process can be summarized as in Eq.1:

$$M_i^* = (B + \delta B_i)\,(A + \delta A_i)^{-1} \quad (1)$$

δB_i: randomly sampled deviation matrix for the elementary flows

B: deterministic elementary flow matrix

δA_i: randomly sampled deviation matrix for the intermediate flows

A: deterministic intermediate flow matrix

i: number of simulation, $i = 1, \ldots n$ ($n = 1{,}000$)

The resulting M matrix has the dimension of 1,869 (elementary flows) × 11,332 (processes), and we have generated 1,000 of them, $\{M_1^*, M_2^*, \ldots M_{1000}^*\}$. To ensure efficiency, we further sampled 1,000 data points from each M_i^*. To do so, we have extracted 1,000 randomly chosen elementary flow-process pairs, and used them to extract 1,000 data points for each run. The sampled 1,000 elementary flow-process pairs can be found in the SI Excel file. The number of data points that underwent the following statistical analyses were therefore 1,000 (elementary flow-process pairs) by 1,000 (runs) = 1,000,000. One whole iteration including simulation, calculation of entire LCI results, and storage of randomly chosen 1,000 points takes about 1 minute in Python 2.8 in Windows PC with 16 cores. The total time for completing 1,000 times of simulations is 1,000 times of it, which is about 1,000 minutes ≈ 17 hours.

2.2.2 Distribution functions

A probability distribution function $f(x)$ is a function describing the probability distribution of a random variable X. The most frequently used statistical distribution for the unit process level inventory in ecoinvent is the lognormal distribution (Table 1). Normal and triangular distributions are also considered as the input parameter distributions, though they are less common than lognormal distribution. The other two distributions similar to lognormal distribution are gamma and Weibull distributions, which will be used to test the distribution of aggregate LCI results in this study. Details about the five distributions are presented in the Appendix.

Table 1. Summary of probability distribution in ecoinvent v3.1 unit process data

	A matrix	B matrix
Number of columns	11,332	11,332
Number of rows	11,332	1,869
Lognormal distribution	94.7%	60.5%
Normal distribution	0.5%	0.07%
Triangular distribution	0.05%	0.002%
Undefined	4.8%	39.4%

2.2.3 Statistical analysis of fitting the distribution

After the 1,000,000 samples as described in the previous section are obtained, statistical analysis is performed to discover the probability distribution of the aggregate LCIs of ecoinvent v3.1. A general method of finding the best fitting distribution involves the following three steps: (1) Plot the data in frequency histogram or density plot to narrow down the list of possible distribution types (Singh et al. 1997); (2) To ensure that the sample is not biased, run a normality test using Shapiro-Wilk normality test following Razali and Wah (2011); (3) Generate LCIs based on the hypothesized distributions and test the fitness of each distribution with the original data using overlapping coefficient method.

LCI results that follow a perfect lognormal distribution can be generated by applying the log-mean and log-standard-deviation of the LCI results. To estimate Weibull and gamma distributions, shape and scale, and shape and rate of the LCI distribution are calculated, respectively. The coefficient of overlapping (OVL) is a measure to evaluate the similarity of two probability distributions, which can be used to calculate the percentage of overlapped area between the distribution of LCI sample results and the expected distribution. The greater the value of OVL, the more similar of the two distributions. In equation (2), Δ is the OVL that represents the common area under both density curves. If the two density functions are *f(x)* and *g(x)*, then

$$\Delta(f, g) = \int min\{f(x), g(x)\}\, dx \tag{2}$$

The OVL of the distribution estimate and the sample aggregate LCI results are calculated in R program. Detailed explanation of overlapping coefficient method can be found in Ridout and Linkie (2009).

2.2.4. Data sources

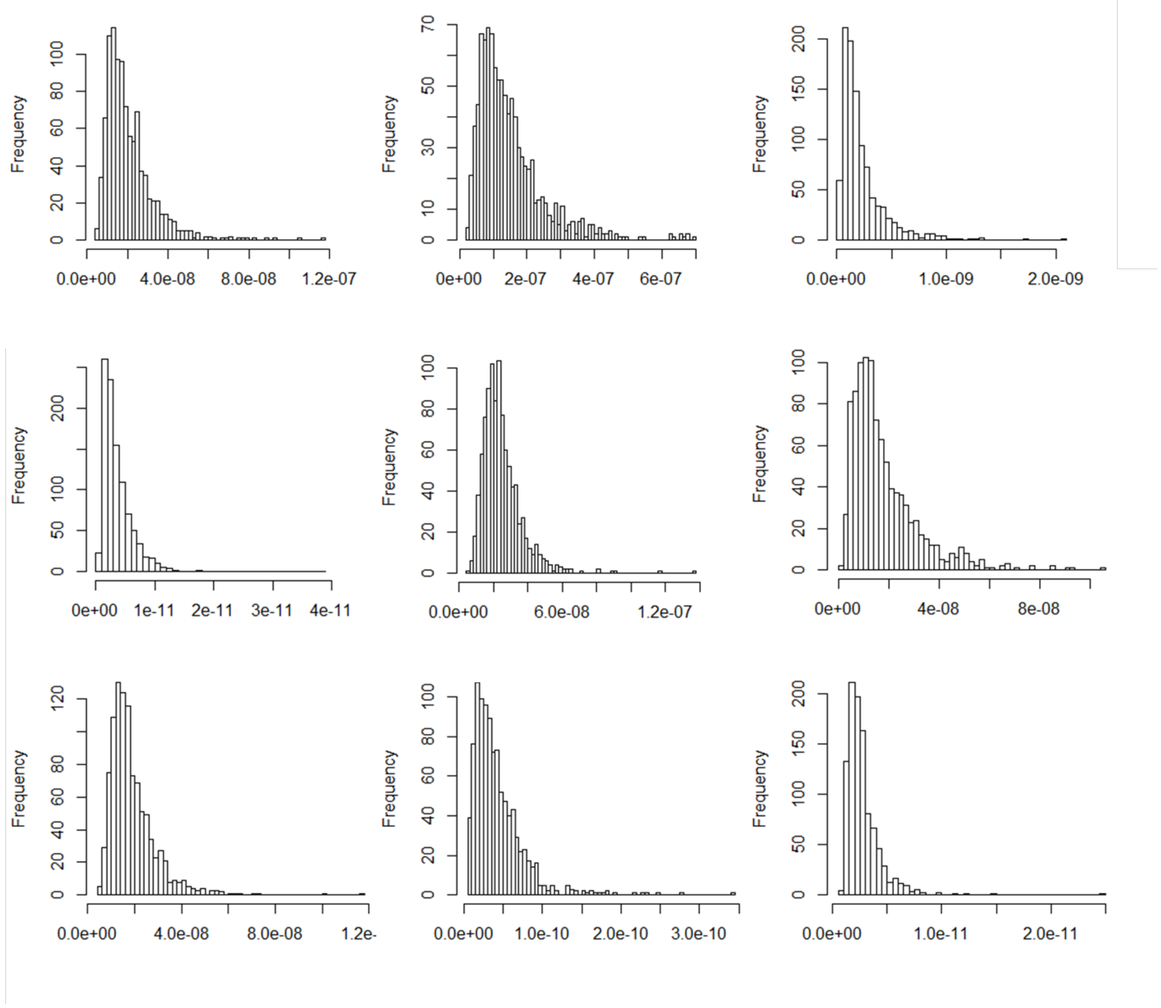
We use the unit process inventory data obtained from the ecoinvent database v3.1 (default allocation method) as our input data. The version 3.1 contains more than 11,000 unit processes and nearly 2,000 types of environmental exchanges (Weidema et al. 2013). Uncertainty information including uncertainty type and corresponding distribution parameters are given for each unit process data. The unit process data includes both intermediate flow matrix (A) and elementary flow matrix (B) and their distributions. For unit process data in lognormal distribution, all the geometric standard deviations of them are calculated based on their variance in pedigree uncertainty.

We also corrected a few extremely high uncertainty values in the database, which are likely to be erroneous, into reasonable values in order to make the A matrix invertible. For example, one of the intermediate flow in the database follows a lognormal distribution with GSD = 4.1E+22, which is highly unlikely to be reflective of the reality. Furthermore, such high GSDs will lead to extreme values in the $(A + \delta A_i)$ matrix that will make it noninvertible. Therefore, we adjusted the GSDs of those intermediate flows into reasonably high value (GSD = 5), which is still about 4 times higher than average GSD, 1.3. For consistency, we also corrected uncertainty values in the B matrix. Because elementary flows have relatively higher GSD values than intermediate flows in the database, we assign GSD = 10 to those GSDs greater than 10 in the B matrix (average GSD of the elements in B matrix is 1.8).

2.3. Results and discussion

As the first step of our analysis, we constructed frequency and probability density plots of simulation results of LCIs to see their distribution shapes. Figure 2 presents the histograms of LCI results of 9 random elementary flow-process pairs. The distribution results are similar to the previous LCI simulations in the literature (Sonnemann et al. 2003; Muller et al. 2014). The shape of the distributions in Figure 2 can be visually identified as lognormal, gamma, or Weibull distributions (Holland and Fitz-Simons 1982). To further determine the type of probability distributions for these results, normality statistical test and overlapping coefficient method are applied.

Figure 2. Histograms of 9 random points in 1,000 iterations of LCI results



By definition, if the logarithm of the data is in normal distribution, then the data has a lognormal distribution. The QQ-plots of log-transformed LCI results in the Supplementary Information indicate the majority of LCI results are very close to lognormal distribution. The normality of the data can also be assessed through a variety of statistical tests. One of the most common tests is Shapiro-Wilk normality test, which is known to be the most powerful approach to normality test (Razali and Wah 2011). The results of Shapiro-Wilk normality test of simulated LCI are provided in Table 2.

Table 2. Shapiro-Wilk Normality Test results of simulated LCIs (p-value)

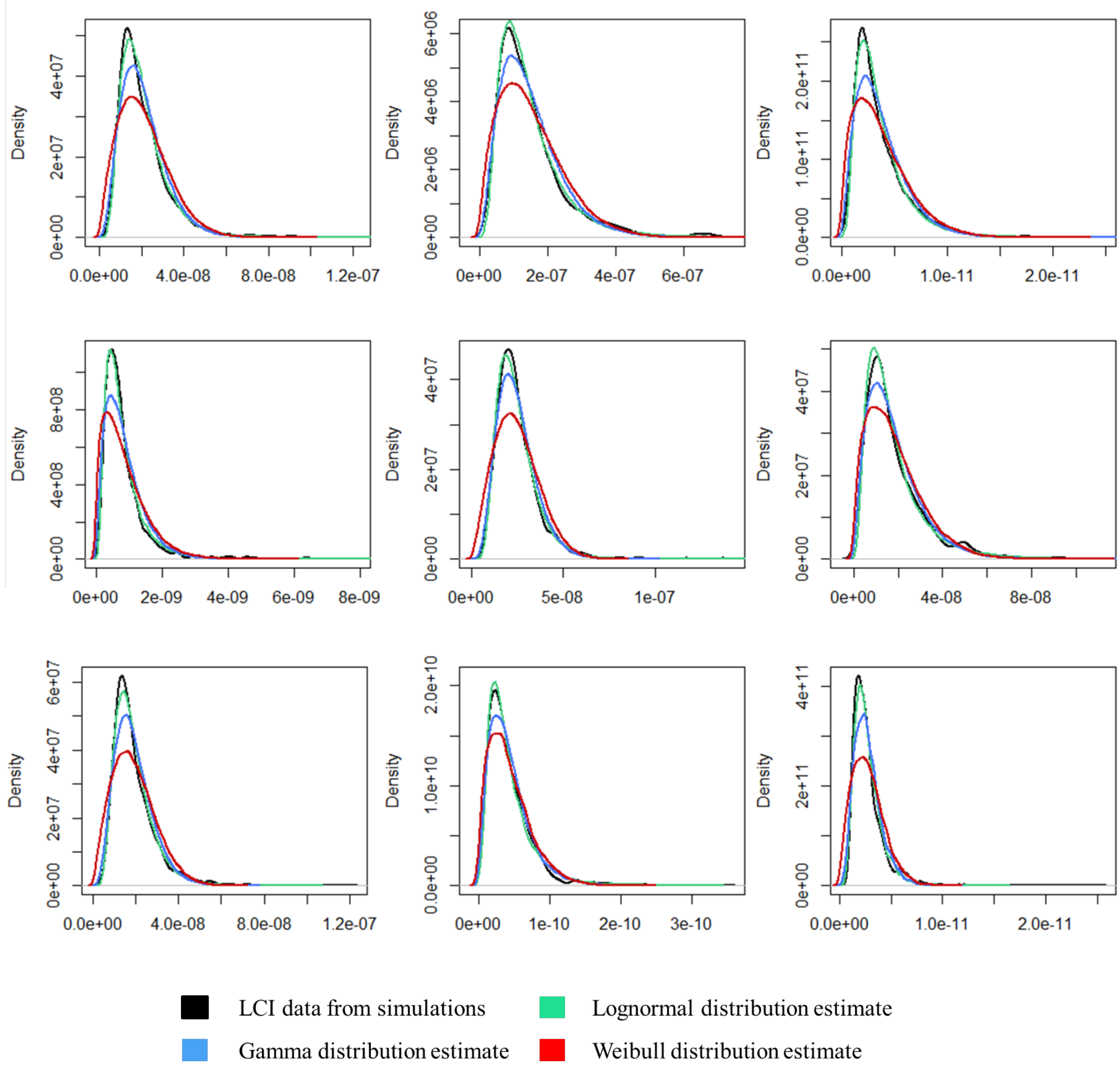
1,000 samples		100 samples	
X	log(X)	X	log(X)
Average *p* -values			
0.00	0.18	0.02	0.38
Percentage of *p* -value > 0.05			
0%	43%	5%	81%

The results of normality test for the 1,000 random elementary flow-process pairs are presented in Table 2. At 95% confidence level, p-value less than 0.05 means we reject the null hypothesis that the probability distribution of the data is normal. About 99.8% of the simulated LCI results showed p-values greater than 0.05, meaning that nearly all of the simulated LCI results are not normally distributed.

After we log-transformed the LCI outputs, the share of the simulated LCIs that passed the test increased to 43% (Table 2), indicating that they more likely to be lognormally distributed than normally distributed. At 95% confidence level, average p-value of log-transformed LCI results is 0.18, accepting the null hypothesis that LCI results are lognormally distributed. Still 57% of the 1,000 samples of LCI results did not passed the normality test after log-transformation. This can be explained by the well-known observation that the power of Shapiro-Wilk test diminishes as the size of log-normally distributed sample increases (Yazici and Yolacan 2007). Therefore, we performed the Shapiro-Wilk normality test for only 100 randomly chosen samples of simulated LCIs. The results show that 81% of the simulated LCIs passed the normality test in this case, confirming that simulated LCIs generally follow lognormal distribution.

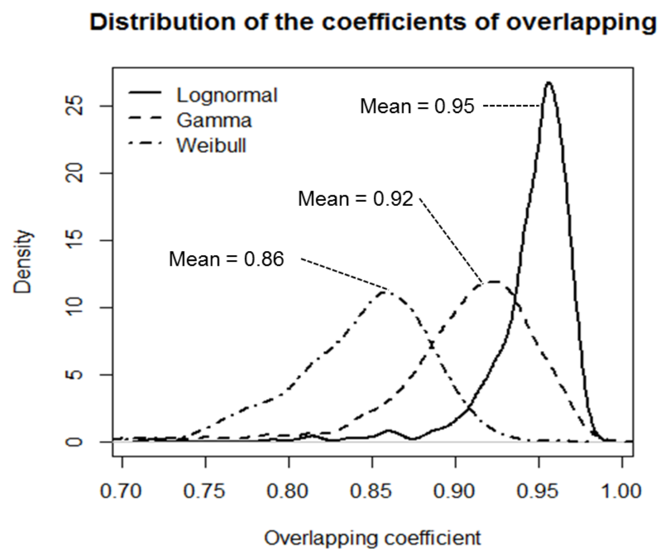
The next step of fitting the distribution is to test how well a lognormal distribution or other possible distributions actually fit LCI simulations. As mentioned before, according to the shape of the curves in histograms, some possible distributions of LCI results include lognormal, gamma and Weibull distributions. The results are fitted by those distributions, and the coefficients of overlapping (OVL) are calculated to find the closeness of the results to those distributions. The three types of distributions are generated based on the corresponding distribution parameters of simulated LCI results as described in the 'method and data' section. Detailed description about the probability density functions for the three distributions is included in the Supplementary information. Figure 3 represents 9 typical comparisons among the results and the estimates of lognormal, gamma and Weibull distributions of random elementary flow-process pairs.

Figure 3. Density plots of LCI data, lognormal, gamma and Weibull distribution estimates



In the plots of the distribution comparisons (Fig. 3), lognormal distribution estimates have the larger shared area with simulated LCI data than gamma or Weibull distribution. Figure 4 illustrates the distributions of OVL results from the LCI results verses lognormal, gamma and Weibull distributions. For example, the solid line in Figure 4 shows the OVL probability density of expected lognormal distribution and LCI simulations. The average overlapping coefficient (OVL) for lognormal distribution and LCI result is 95%, while that for gamma and Weibull distributions are 92% and 86%, respectively. The result shows that LCI samples are closest to a lognormal distribution compared to other distribution types based on the coefficients of overlapping approach.

Figure 4. The coefficients of overlapping (OVL) of 1,000 samples of LCI results and lognormal, gamma, and Weibull distribution estimates.



Graphically and numerically, therefore, we could conclude that LCI results of ecoinvent v3.1 are lognormally distributed. This observation allows us to characterize the distribution of aggregate LCI results more efficiently using GSD and median. In other words, individual users do not need to perform a MCS using unit process-level data, which can be highly time-consuming given the dimensions of matrices involved.

2.4. Conclusions

In this study, the probability distribution type for aggregate LCIs of the ecoinvent v3.1 database is identified by comparing the simulated LCIs to three possible distributions. The results show that lognormal distribution has the highest overlapping coefficient (average

95%) with simulated LCIs as compared to gamma (average 92%) or Weibull distribution (average 86%). Our normality test results also confirm that 43% of aggregate LCIs follow lognormal distribution. Therefore, aggregate LCIs can be presented efficiently as lognormal distribution (i.e. median and GSD).

Though the current database has uncertainty values for unit process inventory, conducting uncertainty analysis starting from the unit process level is neither time-efficient nor necessary for most studies. Therefore, the determination of the distribution that best fits the aggregate LCIs is needed. It would help improve the efficiency of storing uncertainty data and performing uncertainty analysis in LCA by saving computation time and storage of LCI data.

By way of an example, 1,000 times of LCI simulation using unit process-level distribution information for a product system that involves 30 inputs from ecoinvent v3.1 would take 1,000 mins for a modern, average desktop computer (7 core computer, 16 GB ram, 3.4 GHz). By using pre-calculated distribution function for LCIs, this can be reduced to 15 seconds, which is 1/4000th of the time needed for the unit process-level computation.

Our study only considers the uncertainty information from unit process data from ecoinvent 3.1, which is mostly based on the pedigree matrix. Pedigree method is a pragmatic approach to uncertainty in the absence of better uncertainty information. However, the theoretical and empirical grounds of applying pedigree approach to quantify uncertainty itself are questionable (Ciroth et al. 2013). The validity of pedigree approach was not within the scope of our paper; the methodology presented in this paper can be applied to any uncertainty data regardless of how they are derived in the first place. Though the majority of the unit process data in A matrix include uncertainty values in the current database, there is still part of them lacking uncertainty information. The problem is more severe when it comes to B matrix, where only about 60% of the data contains uncertainty values in ecoinvent v3.1. The aggregate LCI results that we have calculated, therefore, does not reflect all the uncertainties, because some of the uncertainty data, especially those in B matrix, were not considered. However, for the purpose of this study, adding additional uncertainty information for those that are missing in the original data is unlikely to change the conclusions of our study.

Aggregate LCI uncertainty is only one step in the analysis of LCA uncertainty. Not only LCI uncertainty, but also the uncertainty from impact assessment should be assessed in order to achieve the overall uncertainty of the final LCA results. Additional research is needed to understand the uncertainties in LCA encompassing both LCI and LCIA.

2.5. Appendix

Provided in the supporting information are the QQ-plot of 9 random log-transformed LCI results and mathematical notations for 5 distribution types that are relevant to our study.

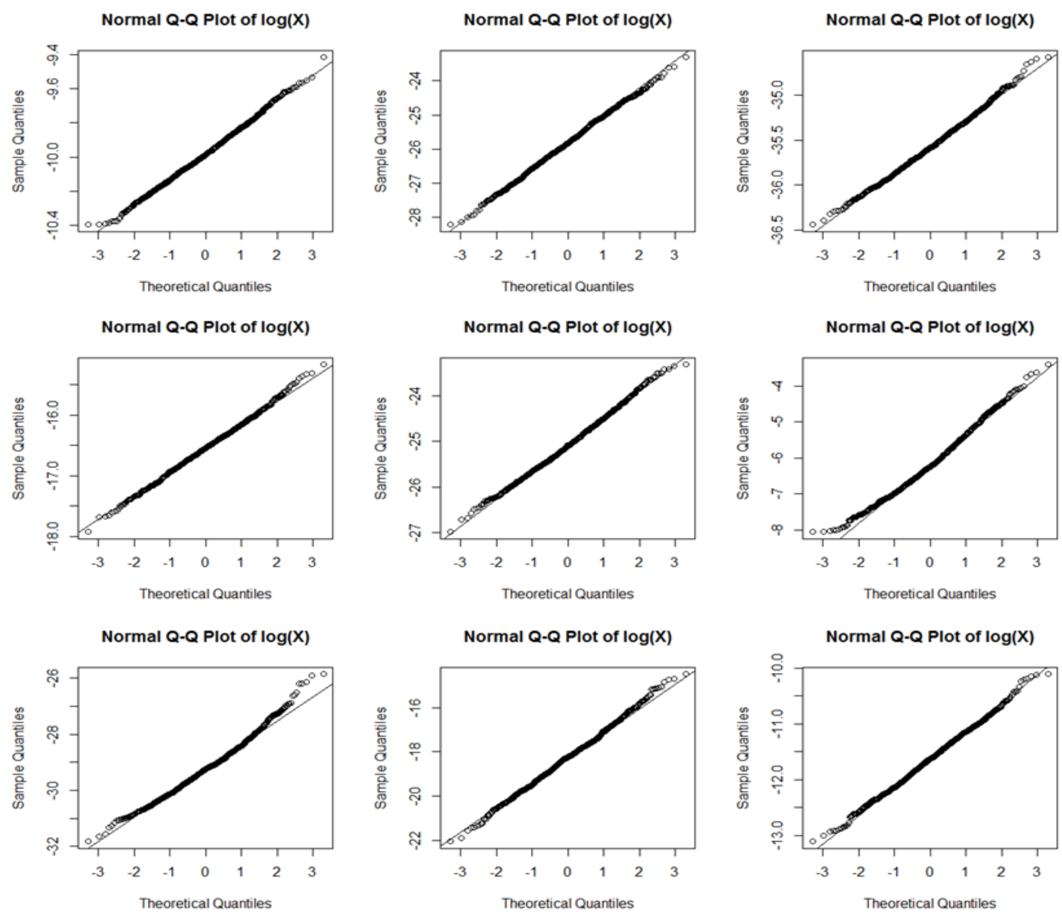
2.5.1. Uncertainty values for 1,000 LCIs

Lognormal distribution can be expressed by two parameters: median and geometric standard deviation (GSD), which are shown in the file. To protect the original ecoinvent v.3.1 data, median values of these results are not shown in absolute value but only in the percentage of the deterministic LCI results. For example, 120% in the spreadsheet means the median of simulated LCI results is 120% of the deterministic LCI value shown in the original ecoinvent v. 3.1 data. Also included are two tabs that show intermediate and elementary flow names used in the study.

2.5.2. QQ-plot of 9 random log-transformed LCI results

The QQ-plots are used to test the normality of the log-transformed LCI results. The results in Figure S1 indicate the majority of LCI results are very close to lognormal distribution.

Figure S1. QQ-plot of 9 log-transformed LCI results



2.6.3. Description of distribution functions used in the study

Description of five major distributions from ecoinvent data and for the distribution analysis is presented in the following. Among the five distributions, normal, lognormal, and triangular distributions are used in ecoinvent unit process data, and lognormal, Weibull and gamma distributions are used in distribution fitting of the LCI results based on distribution shape.

Normal distribution

If the probability distribution of X is a bell-shaped curve and symmetric to its mean value, X has a normal distribution (Gaussian distribution). Normal distribution is the most common and important probability distribution and it is often used in science to represent random variables (Limpert et al. 2001). The distribution can be characterized by arithmetic mean μ and the standard deviation σ in the equation:

$$f\ (x)\ =\ \frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Lognormal distribution

If X is lognormally distributed, Y = *ln*(X) is normally distributed. The probability distribution function for lognormal distribution is:

$$f\ (x) = \frac{1}{x\sigma\sqrt{2\pi}}e^{-\frac{(ln\,x-\mu)^2}{2\sigma^2}}$$

with μ is mean of the normal distribution and σ is standard deviation of the normal distribution. In ecoinvent data, values of representing lognormal distribution are median which is geometric mean and variance with pedigree uncertainty which used to calculate geometric standard deviation (Weidema et al. 2013).

Triangular distribution

The triangular distribution is a probability distribution in a triangular shape with lower bound a, upper bound b and mode c. The probability density function is defined by the following function:

$$f(x) = \begin{cases} \dfrac{2(x-a)}{(b-a)(c-a)} & a < x < c \\ \dfrac{2(b-x)}{(b-a)(b-c)} & c \le x < b \end{cases}$$

Triangular distribution is used to estimate the distribution if only limited sample data is available because this distribution is based on the mode, minimum and maximum. More detailed explanation of normal, lognormal and triangular distributions and their presentation in ecoinvent database can be found in Heijungs and Frischknecht 's paper (Heijungs and Frischknecht 2004).

Weibull distribution

The Weibull distribution is used to add flexibility of exponential distribution, and it has lighter tails than lognormal (Holland and Fitz-Simons 1982). The Weibull distribution can describe distribution with positive or negative skewness while lognormal and gamma can only describe positive skewed distribution. It has a distribution function where k is the shape parameter and λ is the scale parameter of the distribution:

$$f(x) = \frac{k}{\lambda}(\frac{x}{\lambda})^{k-1}e^{-(x/\lambda)^k}$$

Gamma distribution

The gamma distribution is often selected as distribution type for representing ecological and physical data (Dennis and Patil 1984). The gamma distribution provide population model, and chi-square and exponential distributions are special cases of the gamma distribution (Holland and Fitz-Simons 1982). The probability density function is in the following formula with shape k and scale θ:

$$f(x) = \frac{x^{k-1}e^{-\frac{x}{\theta}}}{\theta^k \Gamma(k)}$$

Chapter 3. Does the Use of Pre-calculated Uncertainty Values Change the Conclusions of Comparative Life Cycle Assessments? – An Empirical Analysis[2]

3.1. Introduction

Life cycle assessment (LCA) is a tool to evaluate the environmental performance of a product (Guinée 2002; Normalización 2006). LCA results often support corporate and public policy decisions (Azapagic 1999; Burgess and Brennan 2001; de Bruijn et al. 2002; Cederberg and Stadig 2003; Kloepffer 2008; Finnveden et al. 2009). When using LCA results for decision-support, it is crucial to understand the uncertainty in them (Sugiyama et al. 2005; Geisler et al. 2005; Finnveden et al. 2009), because lack of understanding of the uncertainty behind them may materially mislead the decisions (Huijbregts et al. 2001; Heijungs and Huijbregts 2004; Ciroth et al. 2013).

In general, uncertainty analysis in LCA is performed using sampling methods or analytical approaches, and the most commonly used approach is the Monte Carlo simulation (MCS) (Huijbregts 1998; Sonnemann et al. 2003; Peters 2007; Hung and Ma 2009; Imbeault-Tétreault et al. 2013; Heijungs and Lenzen 2014; Prado-Lopez et al. 2014; Vinodh and Rathod 2014; von Pfingsten et al. 2017). MCS uses random samples of input parameters following their stochastic characteristics, and runs the model repeatedly for a sufficiently large number of times to allow statistical analysis of the results (Helton and Davis 2002; Heijungs and Frischknecht 2005; Bojacá and Schrevens 2010; Castaings et al. 2012). For

[2] This chapter was published in PLoS ONE. Qin, Y., Suh, S. (2018). Does the Use of Pre-calculated Uncertainty Values Change the Conclusions of Comparative Life Cycle Assessments? – An Empirical Analysis. *PLoS ONE*, 13(12), e0209474.

example, Noshadravan et al and Gregory et al performed MCS to compare two pavement designs using the distributions of expected LCA results (Noshadravan et al. 2013; Gregory et al. 2016). These studies considered parameter uncertainty using a fully dependent sampling approach. Imbeault-Tetreault et al performed MCS for an LCA case study with nearly 900 unit processes using fully dependent sampling and compared two scenarios around the use of Global Position System (GPS) (Imbeault-Tétreault et al. 2013). The fully dependent sampling used by Imbeault-Tetreault et al required several hours to complete the MCS. Henriksson et al conducted 1,000 Monte Carlo simulations with fully dependent sampling for a comparative LCA of Asian aquaculture products (Henriksson et al. 2015). Ren et al also performed fully dependent sampling using OpenLCA, which took the team 16 hours for 1,000 Monte Carlo simulation runs on a personal computer (Ren et al.). Existing MCS packages in professional LCA software tools including SimaPro and OpenLCA can sample parameters from foreground processes and from the underlying life cycle inventory (LCI) databases (SimaPro 2016; OpenLCA 2018).

In LCA, performing an MCS using fully dependent sampling typically involves repeated inversion of a technology matrix for each run. As the dimension of the technology matrices used in LCA databases grows, however, MCS is rapidly becoming a computational burden to lay practitioners. The ecoinvent database, which is one of the most widely used LCA databases, used to have about 5,000 processes, while the most recent version of the database, ver. 3.4 contains over 14,000 processes (Frischknecht et al. 2005; Verbeeck and Hens 2010; Weidema et al. 2013; Wernet et al. 2016; Moreno Ruiz et al. 2017). A Monte Carlo simulation using ecoinvent ver. 3.1 takes about 1 day for 1,000 runs in a personal computer environment using a Python solution for inversion based on Gaussian elimination algorithm with 16GB random access memory (RAM) and 1TB solid-state drive (SSD) (Qin and Suh 2017).

The time required for each matrix inversion in a modern computer is known to have an order of $n^{2.73}$ time complexity, where n is the dimension of a irreducible, invertible square matrix (Stothers 2010; Williams 2012; Wu et al. 2014), which is generally the case in LCAs (Suh and Heijungs 2007). This means that doubling the dimension of a technology matrix increases the computational time at least 4.8 times. Given that the number of processes in LCI databases continues to grow, running full MCSs will increasingly become a challenge.

In 2016, the current authors published pre-calculated uncertainty values for the entire ecoinvent ver. 3.1 LCI database for the purpose of saving computation time of running a full MCS by individual users (Qin and Suh 2016). Using pre-calculated uncertainty values for LCIs, the users of LCI database do not need to invert the entire ecoinvent database, while there still is a need to invert the technology matrix for the foreground system, which is generally much smaller in dimension. In a commentary to our paper, Heijungs et al. (Heijungs et al. 2017) raised a concern that the use of pre-calculated uncertainty values in comparative studies ignores the dependence among background processes, leading to a large overestimation of uncertainty due to independent sampling. In our response (Suh and Qin 2017), we empirically tested the argument by Heijungs et al, and found that (1) the difference in overall uncertainty characteristics in the results between fully dependent sampling and the use of pre-calculated uncertainty is small; and that (2) the use of pre-calculated uncertainty tends to underestimate, rather than overestimate, the uncertainty measured using the distribution of Geometric Standard Deviations (GSDs).

However, it remains as a question whether the additional errors due to the use of pre-calculated uncertainty values are small enough to maintain the conclusions of a comparative study, and, if not, what is the odds of misinterpreting a comparative LCA results due to the use of pre-calculated uncertainty values. In particular, the use of pre-calculated uncertainty values does ignore the presence of internal dependency within a technology matrix (Heijungs and Lenzen 2014), Henriksson and colleagues highlighted the importance of dependent sampling in understanding the distribution of comparative LCA results (2015). There are two main issues to consider. First, when performing an MCS, a data point of the same process commonly used by the two products under comparison can be perturbed independently. In principle, however, they should be perturbed in the same direction and magnitude, which is referred to as 'dependent sampling.' Second, in a comparative LCA setting, the distribution of the difference between the results by the two product systems being compared helps distinguish the real difference of the two results.

We agree with Hendriksson and colleagues on the theoretical superiority of fully dependent sampling, while the computational requirements for performing fully dependent sampling remains as a concern. Therefore, the objective of this paper is to empirically test the hypothesis that the use of partially independent sampling using pre-calculated uncertainty values in a life cycle inventory alters the conclusion that would have been drawn if the uncertainty values are sampled dependently.

3.2. Materials and methods

3.2.1. Two types of sampling methods

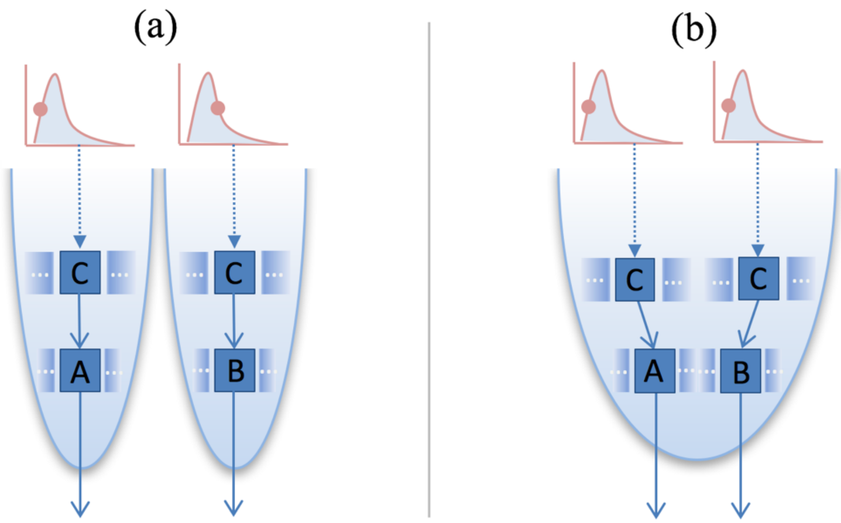
This study compared two types of sampling methods used in calculating LCIs. The first approach, partially independent sampling (PIS), used pre-calculated uncertainty characteristics that are derived using fully dependent sampling (FDS) (Qin and Suh 2016). Although these pre-calculated uncertainty characteristics such as GSDs were derived using dependent sampling, when they are applied to an comparative LCA between products A and B, each set of parameters applied to A and B are sampled independently in such a way that the same parameter that is commonly used by both A and B can be sampled at two different points within the pre-calculated distribution (Qin and Suh 2016; Suh and Qin 2017).

For example, suppose that two products produced from processes A and B are being compared. Both processes receive inputs from process C (see Figure 5(a)). When using PIS for an LCI item, e.g., CO2 emission for processes A and B, the randomly sampled value may be based on two different points of underlying CO2 emissions distribution of process C. In principle, however, the two processes should draw the same value from the distribution, if A and B are receiving the same input from the exactly same facility at the same time. Therefore, the second approach, FDS, draws the same value from process C for each run (see Figure 5 (b)).

Figure 5. Illustrative example of a comparative LCA between A and B involving a common input, C[3]. (a) partially independent sampling of the parameters involving C (use of

[3] Following the terminologies used in our previous paper, we are comparing (1) PIS (inter-input dependence with inter-product system independence; IID+IPI), which is represented in case (b) of Figure 4 in [2], with (2) FDS (inter-input dependence with inter-product system dependence; IID+IPD), which is represented in case (c) of Figure 4 in [2]. Under PIS, all parameters within each

pre-calculated uncertainty values), (b) fully dependent sampling of the parameters involving C (full Monte Carlo simulation); modified from (Suh and Qin 2017).



In reality, however, the parameters for process C may be derived by averaging multiple processes of different locations, and processes A and B may be using inputs from two different processes that are best represented by C in the database. In that case, the use of PIS depicted in Figure 5 (a) may represent the true underlying variability in the data and can thus be justified. Conceptually, however, FDS is the ideal method used in comparing two products' LCAs if the computation time and cost of running full Monte Carlo simulation is not considered as a barrier to LCA practitioners.

In this study, we compared the same elementary flow, i, in the two LCI results for A and B, which are denoted as a_i and b_i, respectively. Under the FDS approach, the distribution of the difference between the two, or $a_i - b_i$, was generated using fully dependent sampling. Under the PIS approach, we used pre-calculated GSDs that were

---

product system that produces A or B in Figure 5 are sampled dependently, while between the two product systems, a parameter commonly used by both A and B, may be sampled independently. Under FDS, all parameters of the two product systems are sampled dependently.

generated from FDS approach of LCIs for processes A and B. The GSDs of elementary flows were generated by sampling all processes simultaneously in the entire ecoinvent database. The use of pre-calculated uncertainty values is considered neither fully independent—because the way of generating the pre-calculated values for the two products are fully dependently sampled, nor fully dependently sampled—because the direct inputs and emissions of the two products are not dependently sampled. Under the PIS approach, we used the pre-calculated uncertainty values, more precisely GSDs, for sampling a_i and b_i, and examined the distribution of the difference between the two.

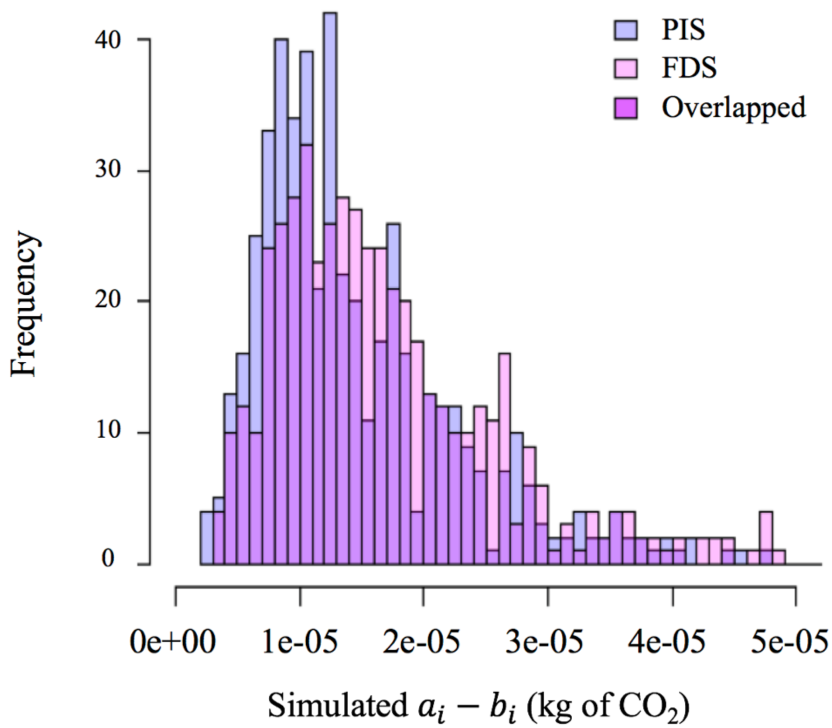
3.2.2. Distribution similarity analysis

After we ran the simulations for the comparative analysis by PIS and FDS, the distributions of LCIs from the two methods were obtained. In order to measure the similarity of the distributions of the two approaches, we used overlapping coefficient (OVL) analysis and determined the shared area between the two distributions of the difference between a_i and b_i. For the given density functions $f(x)$ and $g(x)$, the OVL is represented in the following equation:

$$OVL(f, g) = \int min\{f(x), g(x)\}\, dx \quad (3)$$

One example of overlapping coefficient is presented in Figure 6. The blue histogram represents the distribution of $a_i - b_i$ generated from the PIS approach, and the pink histogram shows the distribution of $a_i - b_i$ generated from the FDS approach. The purple area is the shared area of the two distributions, and the overlapped area can be calculated as a ratio, an overlapping coefficient. A high ratio of overlapping coefficient means the two distributions are similar to each other. The calculation of OVL for the distributions was completed in R program (Ridout and Linkie 2009).

Figure 6. An example of overlapped histograms of one pair of elementary flows of LCIs ($a_i - b_i$) using PIS and FDS.

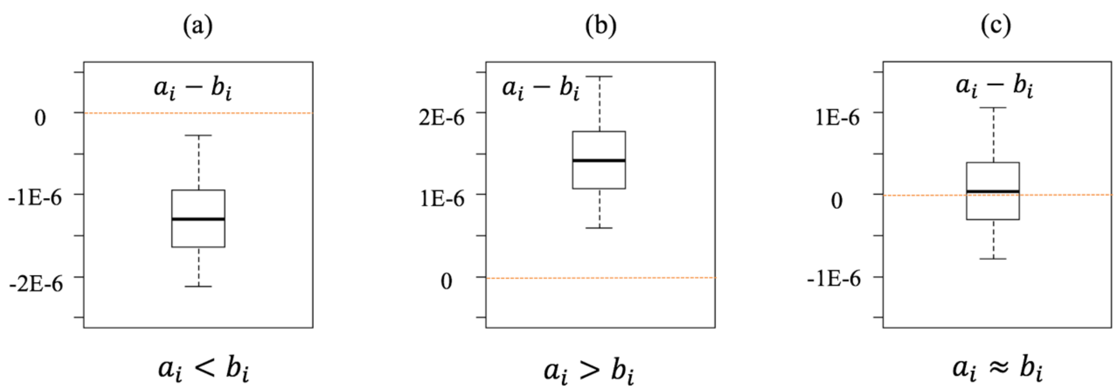


3.2.3. Decision context

In addition to analyzing the similarity of the distributions from the two approaches, we examined the potential outcomes of comparing A and B based on the inventory item, i (a_i and b_i). In practice, a single elementary flow is rarely, if at all, used as the basis of a comparative LCA. As we will discuss later, the use of characterized impact is likely to dampen the differences between the two sampling approaches, and therefore our use of elementary flow in this analysis should be considered as a more conservative approach; i.e., the frequency of reversing the conclusion due to the use of PIS instead of FDS would be lower if characterized results are used as the basis of a comparative LCA.

Figure 7 shows the three possible outcomes from comparative studies. The boxplots represent the distributions of the comparative LCI results of processes A and B for the elementary flow i (a_i and b_i).

Figure 7. Comparative results of LCI results for two processes A and B in the same elementary flow. (a) A is better. (b) B is better. (c) Inconclusive conclusion.



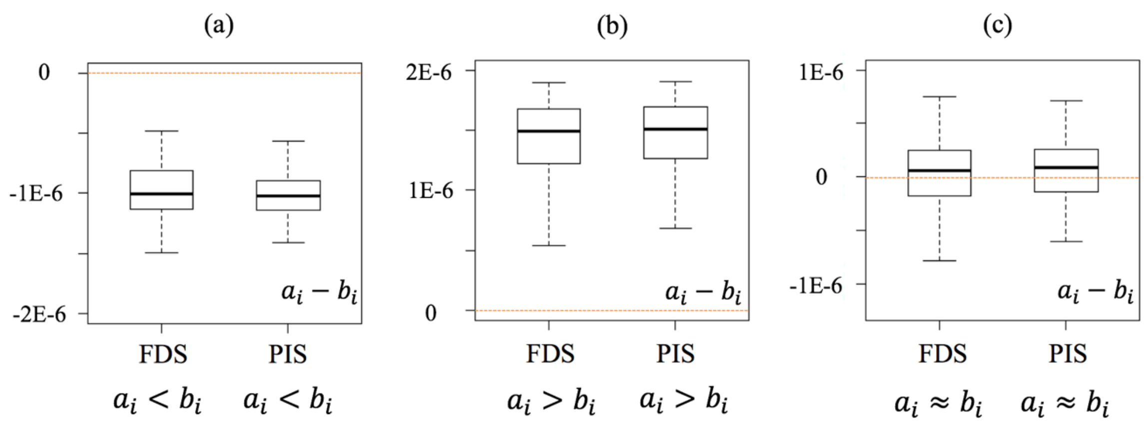
After 1,000 random samplings and calculations for each random pair of randomly selected LCIs under MCS, we analyzed the frequency that a_i is smaller than b_i. If the frequency exceeds the set threshold (70%, 80%, or 90% of the 1,000 runs), then we determined that A is better than B in terms of the elementary flow i (Figure 7 (a)). In other words, we determine that A is better than B in terms of elementary flow i if $a_i - b_i$ is smaller than 0 for at least 700 runs out of 1,000 under the 70% threshold case. If the opposite is true, we determined that B is better than A (Figure 7 (b)) with regard to the elementary flow. For all other cases, we determined that the comparison is inconclusive under the set threshold (Figure 7 (c)).

This concept is overlaid to the use of PIS and FDS as explained in the following sections.

The cases that the conclusions are identical

This is the case when the outcome of the comparative LCA using FDS and PIS is the same (Figure 8). In Figure 8 (a), for example, the results of $a_i - b_i$ of both FDS and PIS show A is better than B within the set threshold. In this case, there is no penalty for an LCA practitioner to use the computationally lighter approach, i.e., PIS, in a comparative LCA context.

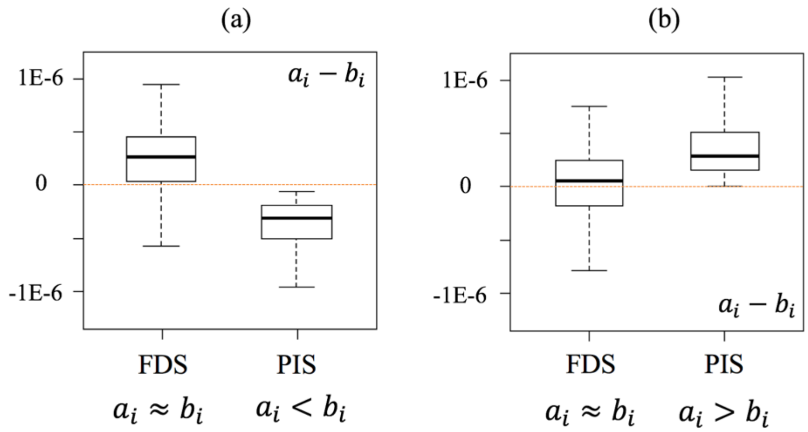
Figure 8. Identical conclusions of the comparison of A and B using FDS and PIS. (a) A is better. (b) B is better. (c) Moderated conclusion.



The cases that conclusions are moderated

This is the case when one of the two approaches (PIS or FDS) concludes that A or B is better, while the other approach concludes that the comparative outcome is inconclusive. Figure 9 shows the two cases where the conclusion is moderated by the use of PIS.

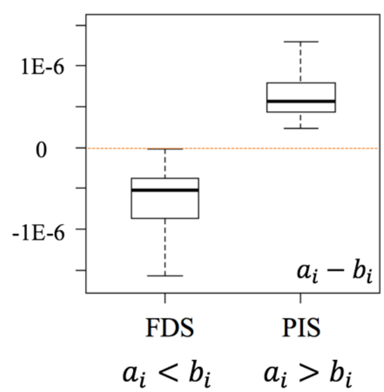
Figure 9. Moderated conclusions of the comparison of A and B using FDS and PIS. (a) One method shows A is better, and the other indicates inconclusive conclusion. (a) One method shows B is better, and the other indicates inconclusive conclusion.



The cases that conclusions are reversed

The third case is that the comparative outcome obtained from FDS is reversed when PIS is used instead. For example, the results from one approach conclude that A is better than B, while the results from the other approach indicate that B is better than A within the set threshold (Figure 10).

Figure 10. Reversed conclusions of the comparison of A and B using FDS and PIS. One method shows A is better, and the other indicates B is better.



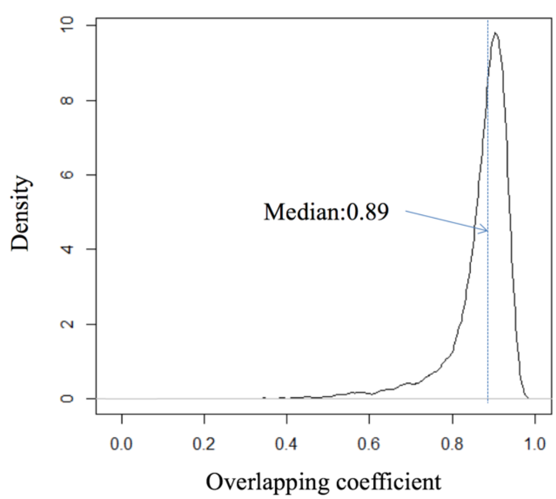
Using the framework outlined in this section, we conducted an empirical analysis using the ecoinvent database, and the results are discussed in the next section.

3.3. Results and discussion

3.3.1. Overlapping coefficient analysis

10,000 pairs of elementary flows of LCIs, a_i and b_i were randomly selected from ecoinvent v3.1, and we simulated 1,000 times of each pair of elementary flows for both PIS and FDS approaches. Therefore, the total number of data points used for the statistical analysis was 40 million (10,000 elementary flows × 2 processes × 2 approaches × 1,000 runs). The distribution of the overlapping coefficients for 10,000 pairs of comparison is shown in Figure 11. Most overlapping coefficients (86.8%) of the distributions of FDS and PIS approaches are above 0.80, and the median is 0.89, indicating that the two methods generate similar distributions of $a_i - b_i$.

Figure 11. Distribution of overlapping coefficients for 10,000 pairs of elementary flows of LCIs ($a_i - b_i$) using FDS and PIS approaches.



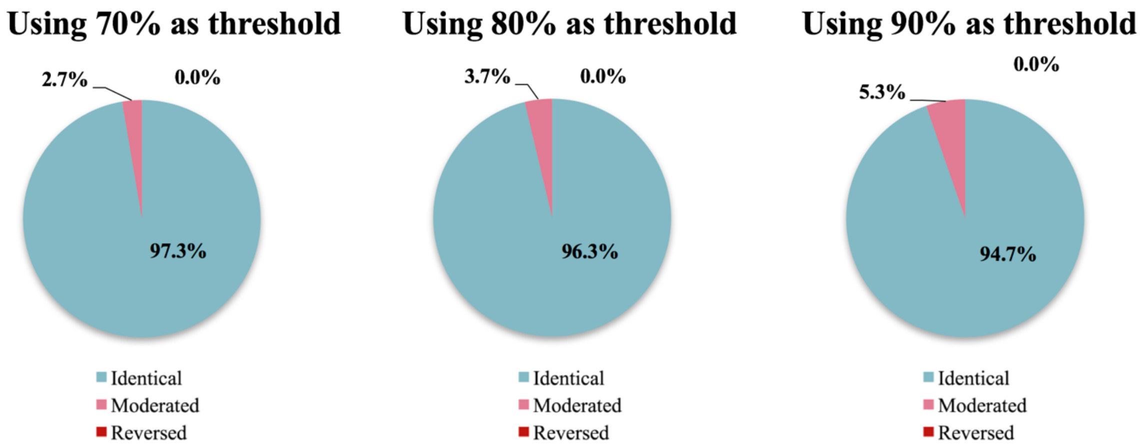
Over 74.9% of the cases showed the overlapping coefficient of 0.85 or higher indicating that the comparative results between PIS and FDS would be very similar. However, as much as 2.4% of the cases showed the overlapping coefficient of 0.6 or lower, and in those cases,

the risk of drawing a different conclusion by using PIS instead of FDS is more pronounced. While OVL analysis shows the general trend of similarity between the outcomes drawn using the two approaches, the frequency of drawing a different conclusion can only be tested empirically using random sampling of actual dataset. The following section presents the result of the empirical analysis.

3.3.2. Comparing randomly selected processes

Figure 12 shows the frequency of arriving at (1) an identical, (2) moderated, and (3) reversed conclusions by using PIS instead of FDS under three threshold conditions, 70%, 80%, and 90%. The chances that the conclusions are identical, moderated, and reversed were 94.7%, 5.3%, 0.0%, respectively when 90% was used as the threshold condition (i.e., $a_i - b_i$ should be smaller than 0 for 90% of the cases in order to determine that A is better than B). In other words, the use of pre-calculated uncertainty values generated the same results of FDS approach at about 95% of the time even when a very stringent threshold condition of 90% was employed. For the remaining 5.3%, the conclusion was moderated but not reversed.

Figure 12. Comparison results of FDS and PIS in 10,000 pairs of random processes using 70%, 80%, and 90% thresholds.



When the threshold condition was relaxed to 80% and 70%, as expected, the chance for PIS to arrive at a moderated conclusion than the case of using FDS was reduced to 3.7% and 2.7%, respectively. Irrespective of the threshold conditions, no case out of 10,000 pairs under each threshold condition arrived at a reversed conclusion.

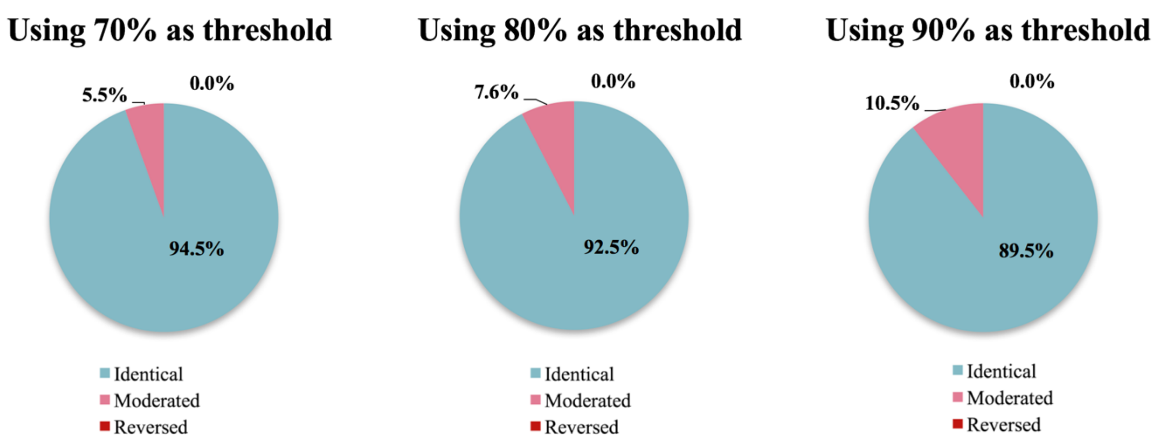
These results were drawn from the randomly selected processes regardless of their functional characteristics. In reality, comparative LCAs are more likely to be performed among the processes with the same or similar functional outputs. Functional equivalency of two process outputs is, however, case-dependent and often difficult to determine using only the intrinsic characteristics of the two processes. For example, polyethylene terephthalate (PET) and stainless steel are two different materials, while both of them can be used as a material for tumbler. In that sense, the results shown in Figure 12 is justifiable representation of the errors induced by PIS.

However, we also tested a more stringent case, where the processes to be compared produce the outputs of which not only the functions but also the intrinsic characteristics are equivalent. The following section follows the same procedure, while limiting the processes to be compared within electricity-producing processes, in order to see whether the same observation holds up.

3.3.3. Comparing the processes with identical functional output

This section quantifies the frequency of arriving at a different conclusion due to the use of PIS instead of FDS among the processes that produce electricity. The results showed that the chance of arriving at a moderated conclusion by using PIS was doubled as compared to the case of randomly selected processes. However, the results still showed that most (about 90%) of the conclusions from the two methods were identical, only about 10% of the conclusions were moderated. Again, not a single case showed a reversal of the comparative outcome (Figure 13).

Figure 13. Comparison results of FDS and PIS in 10,000 pairs of random electricity processes using 70%, 80%, and 90% thresholds.



As was the case for the randomly sampled processes, more relaxed threshold conditions generated fewer cases where the conclusion was moderated. Table 3 shows the numerical results of the comparison between the two methods for 10,000 pairs of random sampled processes and 10,000 pairs of random sampled electricity processes. Though the number of weakened conclusions increases in the random electricity processes, the overall identical conclusions are still about 90% in the total 10,000 pairs of electricity processes.

Table 3. Comparison results of LCIs generated from FDS and PIS approaches in 10,000 pairs of random processes and 10,000 pairs of random electricity processes using 70%, 80%, and 90% thresholds.

	Random processes			Random electricity processes		
Threshold	70%	80%	90%	70%	80%	90%
Identical	9,733	9,629	9,467	9,451	9,245	8,947
Moderated	267	371	533	549	755	1,053
Reversed	0	0	0	0	0	0

Regardless of the similarities in the functional outcome of the processes analyzed, PIS produced the identical comparative outcome for about 9 out of 10 times when an LCI was used as the basis of the comparison with the 90% threshold condition. In the remaining 1 out of 10 cases, the results from the use of PIS have been moderated. If the processes are selected randomly or if a more relaxed threshold condition can be used, the chance for PIS to produce a moderated conclusion is reduced to 2.7% -3.7%. If characterized or weighted results, instead of LCI, are used, the chances of moderating the conclusion by using PIS would be lower.

The question then becomes whether the benefits of using pre-calculated uncertainty values by reducing computational time and the costs associated with it outweighs the cost of added inconsistency. Certainly, this is a question that an analyst should consider given the circumstances where he or she is in, and one can hardly give a universally applicable answer to this question. For example, if an LCA practitioner is using an LCA result to claim the superiority of a produce to its competitor with a close margin, the use of PIS would not be a wise decision given the chance that it can introduce additional error in the analysis.

However, for less critical cases such as LCAs for internal purposes or with limited computational power, the added errors due to the use of PIS may be acceptable. If the computational requirement for FDS is a critical barrier for performing an uncertainty analysis, certainly the benefits of using PIS would outweighs the cost of not performing uncertainty analysis.

It is also notable that the results shown in Figure 12 may not be reproducible if applied to other products with functional equivalency.

3.4. Conclusions

Due to the growing size of LCA databases, fully dependent sampling is often a challenge to lay LCA practitioners when conducting an MCS. In this study, we evaluated the probability that an LCA practitioner will make an erroneous conclusion due to the use of pre-calculated uncertainty values or PIS instead of FDS in a comparative LCA setting. The results show that the distributions of the LCI results from the use of PIS and FDS are similar, as 86.8% of their overlapping coefficients are above 0.80. Furthermore, the chances for the use of PIS to moderate the outcomes (i.e., 'A is better than B' becomes 'A and B are indifferent', or vice versa) by ignoring the dependence in the upstream processes are less than 10.5% for the case of electricity-generating processes and less than 5.3% for randomly selected processes both at 90% threshold value. When the decision threshold is relaxed to 80% and 70%, the chances for the LCIs using PIS to moderate the conclusions become 3.7% and 2.7%, respectively, for randomly sampled processes and 7.6% and 5.5%, respectively, for electricity-producing processes. None of the 20,000 pairs of simulated LCIs, each of which took 1,000 runs of MCS, showed a reversal of the conclusion, which is defined in our study as the case where ‘A is better than B’ becomes ‘B is better than A,’ or vice versa, beyond the set thresholds (70%, 80% and 90% of the 1,000 runs).

These results are based on individual LCIs. If characterized or weighted results were used, we believe that the chances for PIS to produce erroneous conclusions may be even less pronounced that our results, given that over- and under-estimated LCIs due to the use of PIS are more likely to be cancelled out in the course of characterization and weighting.

In this paper we evaluated (1) comparisons between two randomly selected processes, and (2) comparisons between two electricity-producing processes. The latter case presents larger number of common processes in the background between the product systems being compared, therefore the errors due to independent sampling are more pronounced. Even more extreme case would be to compare two slight design changes for the same product. In that case, there will be much more significant overlap in the upstream processes. However, those overlaps would already occur at the direct inputs to the foreground process under study, in which case they can always excluded from the comparison, as those common inputs do not contribute to the difference between the two designs. By excluding them, an LCA practitioners are essentially practicing fully dependent sampling for those common inputs. Therefore, the dependence that this paper is concerned is that within the upstream processes modelled within LCA databases, not that in direct inputs to a foreground process, which is better be simply excluded from a comparison.

Our results indicate that pre-calculated uncertainty values can be used as a proxy for understanding the uncertainty and variability in a comparative LCA study especially when adequate computational resources are lacking. The number of unit processes is increasing for many LCI databases, adding to the challenge of running MCSs in a PC-environment in the future. LCA practitioners will need to evaluate whether the additional chances of altering the conclusion due to the use of pre-calculated uncertainty values is tolerable given the goal and scope of the study. The additional errors due to the use of pre-calculated uncertainty values shown in our study seem justifiable if the alternative is no uncertainty analysis due to the lack of computational resources needed for fully dependent sampling.

We believe that the concept of using pre-calculated distributions might be applicable to other related fields such as input-output analysis and material flow analysis, potentially saving computation times and costs.

Chapter 4. Perception of Uncertainty in Characterized Life Cycle Assessment Results

4.1. Introduction

Life cycle assessment (LCA) is a decision-support tool that quantifies the environmental impacts of products throughout their life cycles (ISO 2006). Conducting an LCA, however, often involves the use of uncertain data and models; measurement errors in input data, unrepresentive data, choices of system boundaries, underlying assumptions, and model incompleteness all contribute to the uncertainty of an LCA result (Lloyd and Ries 2008; Clavreul et al. 2012, 2013). Understanding the magnitude of uncertainty helps interpret LCA results for decision-making (Geisler et al. 2005; Sugiyama et al. 2005; Finnveden et al. 2009).

A growing number of LCA studies address uncertainty issues (Cooper et al. 2012; Sills et al. 2012; Groen et al. 2014). The majority of the uncertainty analyses in LCA, however, focuses on life cycle inventory (LCI) (Heijungs 1996; Maurice et al. 2000; Gavankar et al. 2014; Scherer and Pfister 2016; von Pfingsten et al. 2017). The most widely used LCI database, ecoinvent (Frischknecht and Rebitzer 2005), includes uncertainty values, e.g., the geometric standard deviation for lognormal distribution, for 62.7% of its unit process data in ver. 3.4. (Wernet et al. 2016). Professional LCA software tools including SimaPro and OpenLCA provide uncertainty analysis functionality using Monte Carlo simulations, again focusing only on LCI (SimaPro 2016; OpenLCA 2018).

However, not only LCI but also life cycle impact assessment (LCIA) phase of LCA are data- and calculation-intensive, involving many model and data assumptions that could cause errors in LCA results (Huijbregts 1998; Heijungs and Huijbregts 2004; Lloyd and Ries

2008) (Reap et al. 2008; Gavankar et al. 2014). Only few studies consider uncertainty from the characterization phase, and quantitative uncertainty assessments on characterization were mostly focusing on climate change impact category (Cellura et al. 2011; Hauschild et al. 2013). For example, Huijbregts (1998b) addressed characterization factors' contribution to the uncertainties in the global warming and acidification results of roof gutters. Huijbregts et al. (2003) further extended uncertainty analysis toward parameter, scenario, and model uncertainties of the two characterization models.

One of the challenges is that characterization models do not typically provide uncertainty information for the input parameters (Hung and Ma 2009; Noshadravan et al. 2013; Henriksson et al. 2015; Gregory et al. 2016). As a result, the influence of the uncertainty in characterization models on overall uncertainty of an LCA result is largely unknown (Hung and Ma 2009). Characterization, however, may dominate the overall uncertainty of an LCA study; characterization factors are calculated from simplified models of complex interacting physical and chemical systems, and do often require resorting to linearization of non-linear relationships (Cucurachi et al. 2016). Characterization model, thus, may carry larger uncertainties than LCI (Lloyd and Ries 2008).

Literature suggests that LCA practitioners tend to perceive that LCIA phase pertains higher uncertainty than LCI (Owens 1997; Huijbregts 1998; Clavreul et al. 2012). However, to our best knowledge, no studied attempted to compare perceived uncertainties between LCI and characterization. In this study, we collected the perceived uncertainty of LCA experts and compared the uncertainty LCI and characterization factors following the expert elicitation procedure. We also created a Pedigree matrix, which has been used for LCI data quality evaluation, for the characterization phase of LCIA. The survey design and the respondent demographics are presented in Section 4.2 Methods; the survey results and Pedigree matrix for LCI and LCIA are presented in Section 4.3 Results; and discussion and suggestions for future work are provided in Section 4.4 Conclusions and discussion.

4.2. Materials and methods

This study combines the Pedigree approach and expert elicitation approach using a survey.

4.2.1. Pedigree matrix

Uncertainty characterization in LCA using Monte Carlo simulation or global sensitivity analysis requires the information on ranges or distributions of underlying parameters. The most desirable source of information for such ranges and distributions would be empirical measurements, which are, unfortunately, often lacking in practice. In the absence of measurement data, the Pedigree method was often used in LCA to estimate the variability and quality associated with underlying parameters (Frischknecht and Rebitzer 2005).

The Pedigree approach—originally referred to as the Numerical Unit Spread Assessment Pedigree (NUSAP) system—was proposed by Funtowicz and Ravets (1990). The Pedigree approach is essentially a method to convert qualitative characteristics of a data set into quantitative uncertainties (Weidema and Wesnaes 1996; Weidema 1998). Van den Berg et al. (1999) is an early example of a Pedigree matrix, which uses 15 criteria for characterizing uncertainty. In the United State, the US Environmental Protection Agency (EPA) offers a guidance on the creation, management, and use of data quality information in LCA using a Pedigree matrix (Edelen and Ingwersen 2018). The ecoinvent database adopted the Pedigree method since its ver. 2.0 (Althaus et al. 2007; Weidema et al. 2013c). The

Pedigree method used in the ecoinvent database translates the qualitative uncertainty characteristics of a data using 5 criteria, “reliability”, “completeness”, “temporal correlation”, “geographical correlation”, and “further technological correlation”, into the estimated geometric standard deviation (GSD) of a lognormal distribution (Muller et al. 2014). GSD is a measure of the spread of a lognormally distributed data points. GSD of 1.8, for example, translates to one order of magnitude difference between the lower bound and the upper bound of a data set within the 95% range.

The Pedigree method enables quantitative uncertainty analysis in the absence of measured variability information, and it can assess not only parameter uncertainties but also non-parametric uncertainties associated with technical, methodological, and epistemic dimensions of a data set (Van Der Sluijs et al. 2005). Despite these strengths, however, the Pedigree approach fundamentally relies on experts’ subjective judgements, raising questions on its usefulness and validity. Ciroth et al. (2013) compared empirical observations and the uncertainty characteristics derived using the Pedigree approach of the ecoinvent database and found that the use Pedigree approach tend to underestimate underlying uncertainties. Yang et al. (2018) examined LCA results of major crops in the U.S. based on high-resolution spatial data, and concluded that the uncertainty values based on ecoinvent Pedigree method lead to a large underestimation.

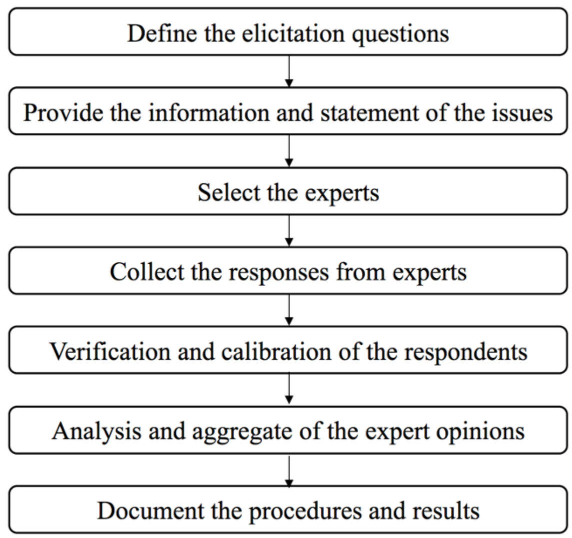
If nothing else, the Pedigree method helps gauge the perceived level of uncertainties in a data set when quantitative measurements are lacking. In this study, we employed the Pedigree approach with various modifications to compare perceived uncertainties in characterization relative to those in LCI. We sent two sets of survey questions, one for characterization and another for LCA, to each expert. For LCI, we modified the Pedigree matrix used in the ecoinvnet database. For characterization, we created a new Pedigree matrix based on the Numeral Unit Spread Assessment Pedigree (NUSAP) literature and environmental risk assessment literature (Funtowicz and Ravetz 1990; Jaworska and Bridges 2001; Van Der Sluijs et al. 2005; Ragas et al. 2009).

4.2.2. Expert elicitation

Expert elicitation is referred as the use of expert judgement on a subject which has insufficient data because of physical constraints or the lack of knowledge (de Franca Doria et al. 2009; Knol et al., 2010; McBride & Burgman, 2012; Morgan, 2014). Expert elicitation was first used in the Delphi method (Brown et al. 1969; Amara and Lipinski 1971; Rowe and Wright 1999). The use of knowledge and wisdom of experts can inform policies when scientific evidence is lacking and help address uncertainties when there is insufficient information. Elicitation of expert judgement has been implemented in various science-policy contexts such as the Intergovernmental Panel for Climate Change (IPCC) (Rypdal and Winiwarter 2001), European Environmental Agency (Meozzi and Iannucci 2006) and U.S. Environmental Protection Agency (2005).

The key steps of conducting an expert elicitation are summarized in the diagram (Figure 14) (Ayyub 2000; Knol et al. 2010). Under the expert elicitation process, experts receive a short description of the purpose of the expert elicitation, the conditions of the participation, an explanation of the performance measures, the uncertainties related to the studied problem, and the key literature substantiates the problem (Cooke & Goossens, 1990; Frey, 1998). Such information elicits the formation of responses to the questions. In our study, the purpose of expert elicitation is to create the Pedigree matrix for characterization factors, and we provided the background information of the Pedigree matrix and a graphical visualization of distributions with different GSDs, so that the experts can better conceptualize the relationship between GSDs and corresponding shapes of the distribution.

Figure 14. Flow chart of expert elicitation procedures.



The selection process is to identity what expertise is relevant to the elicitation and select a sample of experts who can best satisfy the requirements of expertise under the constraints of time and resource (Czembor and Vesk 2009; McBride and Burgman 2012). The quality of expert elicitation depends on the experts' knowledge, experience and practice (Hickey and Davis 2003; Slottje et al. 2008; Martin et al. 2012). It is important to include a diverse range of experts because a large sample of experts can not only represent the whole community but also reduce the influence of individual mistakes and biases (Clemen & Winkler, 1985; Armstrong, 2008). We selected experts based on publication records in the field of LCA and uncertainty analysis.

After the collection of expert judgements, verification and calibration of the expert responses were performed. This step is essential in the analysis of the expert opinions because it can not only check for errors and consistencies in the responses, but also compares the responses to other responses in the elicitation participation and other available information and sources (Cooke, 1991). The sources of bias and error include carelessness, misinterpretation, and overconfidence (Moore and Healy, 2008). Calibration can be used to control overconfidence and inconsistence (Murphy and Daan, 1984). Some methods involved in calibration process are probability theory, aggregation method, and analysis of bias (Clemen and Winkler, 1985). The purpose of the calibration is to reduce the influence of bias and overconfidence and make the expert's response consistent and close to expected true value (Winkler and Murphy, 1968; Alpert and Raiffa, 1982; Ferrell, 1994). In our study, we used weight and life expectancy at birth to calibrate experts' ability to relate perceived distribution to a GSD value (see Section 4.2.4.1).

4.2.3. Survey design and expert selection

We sent the survey information to nearly 200 potential respondents with varying experience levels in LCA. The web-based survey contained 12 questions. The full questionnaire and survey data can be found in the Supporting Information. Given the nature of the survey that involves human subjects, the survey was reviewed and approved by the Institutional Review Board at the University of California, Santa Barbara. The structure and the content of the survey are elaborated below.

4.2.3.1. Background questions

We asked the respondents about their level of experience in LCA, and assigned them into two groups. (1) Group 1: respondents who indicated that they have 6 years or more of experience in LCA and are familiar with the Pedigree approach and (2) Group 2: respondents who indicated that they have less than 6 years of experience or are not familiar with the Pedigree method (see Figure S3 in Appendix). We asked about their degree of approval regarding the use of the Pedigree approach in estimating uncertainties.

4.2.3.2. Pedigree matrix for LCI

In the survey, we asked respondents to provide their opinions about the importance of each criterion to be included in the Pedigree matrix for LCI (Table 4). For the LCI Pedigree matrix, we used the criteria that were provided in the previous versions of the Pedigree matrix of data quality, including geographical correlation, temporal correlation, further technological correlation, completeness, reliability and sample size (Weidema 1998; Wernet et al. 2016). Because the current Pedigree matrix that ecoinvent uses for data quality evaluation has 5 criteria, we used Likert scale for respondents to indicate their perceived importance of the mentioned criteria so that the most important 5 criteria will be selected in the final version of the Pedigree matrix. The Likert scale question in the survey used strongly disagree, disagree, neutral, agree and strongly agree to indicate how much respondents agree with the inclusion of the criterion in the Pedigree matrix.

We asked the respondents to provide their perceived GSDs for all the six criteria in the Pedigree matrix used for evaluating LCI data quality. We provided the criteria description of the original Pedigree matrix used for LCI data quality evaluation for each uncertainty level for each criterion without showing the actual GSDs (Weidema 1998; Wernet et al. 2016), and let respondents input their perceived GSD scores under each criteria description. In order for respondents to better link GSDs with their conceptual thinking regarding uncertainty, we provided the distribution of lognormal distribution with different GSDs.

4.2.3.3. Pedigree matrix for characterization factors

We developed the Pedigree matrix for characterization factors and let the respondents indicate the importance of each criterion to be included in the matrix. Similar to the Pedigree questions for LCI, we used the Likert scale question to ask their opinions on the importance of each criterion to be included in the Pedigree matrix for characterization factors. The 6 proposed criteria include level of consensus, model completeness, temporal specification, geographical specification, reliability of underlying science, and input data characteristics. In order to be consistent with the Pedigree matrix used in LCI, we let respondents rank each criterion and selected five of them to be included in the final version of the Pedigree matrix for characterization factors. We also asked their perceived GSDs for all the criteria we created for evaluating the uncertainty in characterization factors.

The criteria provided in the survey aimed to assess the uncertainties from LCI and characterization models at different and comprehensive aspects (Table 4).

Table 4. Pedigree matrix criteria for LCI and characterization factors.

Criteria for LCI	Purpose	Criteria for characterization factors	Purpose
Completeness	It examines the coverage of the elementary flows in life cycle inventory.	Model completeness	It examines the coverage of the characterization factors for the elementary flows in life cycle inventory.
Reliability	It concerns whether the data is based on measurement or assumptions.	Reliability of underlying science	It assesses the reliability of the underlying science of the method.
Temporal correlation	It evaluates the temporal difference between the data and the study.	Temporal specification	It evaluates whether the characterization model is dynamic which considers background environment and concentration.
Geographical correlation	It assesses the regional difference between the data and the study.	Geographical specification	It concerns the regional resolution of the results in the model.
Further technological correlation	It evaluates whether the technological difference between the data and the study.	Level of consensus	It evaluates whether the characterization method is accepted in the studied field.
Sample size	It assesses the sample size of the data.	Input data characteristics	It reviews the representation of the input parameters used in the model.

At the end of the survey, we also collected their suggestions and concerns regarding the use of the Pedigree matrix in LCA uncertainty estimation. More than half (53%) of the respondents submitted their suggestions as well as their concerns in the survey. The concerns and recommendations are summarized in the discussion section.

4.2.4. Survey analysis

A total of 47 experts from various countries and experiences responded to the survey. Among the 47 responses we received, 23 were in Group 1 who had at least 6 years of experience in LCA and were familiar with the Pedigree approach. The remaining 24 respondents were assigned to Group 2.

To evaluate the importance of the criterion to be included in the Pedigree matrix, we calculated the average scores that were translated from the Likert scales for the criteria, e.g. 1 means strongly disagree and 5 means strongly agree. In our version of the Pedigree matrices, we only selected the top five criteria based on the respondents' selections and included the criteria and the GSDs for the selected criteria into the Pedigree matrix for LCI and characterization factors.

4.2.4.1. Calibration method

To minimize personal biases in relating a perceived distribution to corresponding GSD value, we used a calibration method. First, we provided the GSD value of height of American adult males, which was 1.04 (Fryar et al. 2012). We then let the respondents provide the "best guess" of the distributions for (1) the weights of American adult males and (2) the life expectancy at birth of global population, which were 1.07 and 1.1, respectively (Fryar et al. 2012; CIA 2018). We assumed a linear relationship between actual GSD and the GSD in the response as shown in equation 1:

$$\widehat{\mathrm{GSD}} = \mathrm{a} * \mathrm{GSD}_{\mathrm{survey}} + \mathrm{b} \tag{4}$$

In addition, we explained—and assumed that the survey respondents understood—that GSD = 1 when there is no uncertainty, which provides the second equation to derive both a and b. For example, the GSD for the distribution of weights by American males is 1.07, but a respondent estimated it to be 1.1. Then we calibrate the respondent's GSD estimates by solving:

$$\begin{cases} 1 = \mathrm{a} * 1 + \mathrm{b} \\ 1.07 = \mathrm{a} * 1.1 + \mathrm{b}, \end{cases} \tag{5}$$

which results in:

$$\widehat{\mathrm{GSD}} = 0.7\mathrm{GSD}_{\mathrm{survey}} + 0.3$$

We calculated the expected GSD from both weight and life expectancy at birth for each respondent, and used the average of a and b as the coefficients for the expected GSD equation to calibrate all GSDs.

4.3. Results

We analyzed the survey data and created the Pedigree matrix based on the top five selected criteria in the matrix and GSDs for each uncertainty level for each criterion for both LCI and the characterization factors. The GSDs calibrated by weight and life expectancy at birth for the Pedigree matrices of LCI and characterization factors are shown in Table 7 and 8 (non-calibrated GSDs are in Table S2 and S3). For the sake of comparison, calibrated GSDs by ecoinvent Pedigree scores for characterization factors are given in Table S2 and S3 in the Appendix.

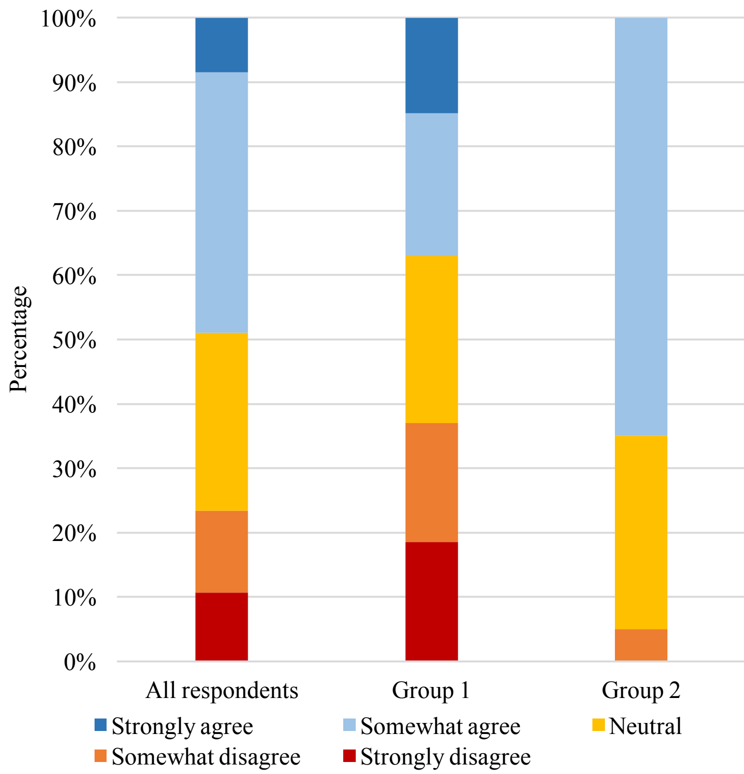
4.3.1. Survey demographics

Most (72%) respondents have been working in the LCA field for at least 6 years: 36% have worked more than 10 years and 26% of the respondents have been working in the field for 1 to 5 years. The majority of the respondents worked in academia (72%), 13% of them worked in a corporate environment, and the remainder came from a consulting firm operating in the field of LCA (9%) and from governmental organizations/research centers (6%). Most respondents came from North America (49%) and Europe (34%), and 13% and 4% came from Asia and South America, respectively. Details can be found in the Supplementary Information (Figure S3-S6).

4.3.2. Degree of approval of the use of Pedigree for uncertainty quantification in LCA data

Approximately half of all respondents expressed their approval to the use of the Pedigree matrix to estimate uncertainty in LCA data (Figure 15). However, the Group 1 (≥6 years of experience) was more likely to disagree with the use of the Pedigree matrix for estimating uncertainty than Group 2 (<6 years of experience) was. As much as 38% of the Group 1 respondents selected “disagree” or “strongly disagree” with the use of the Pedigree method for uncertainty estimation, while only 5% of the respondents in Group 2 chose “disagree”. No respondents from Group 2 selected “strongly disagree” or “strongly agree” to the use Pedigree for uncertainty quantification.

Figure 15. Survey results for the question of the use Pedigree for uncertainty quantification in LCA data.



We also received comments that address the level of acceptance for the use of the Pedigree matrix in characterizing uncertainties in LCA. Some of the respondents strongly disapproved the use of the Pedigree method, largely on the ground of the lack of empirical support to the approach, while others strongly supported the use of the Pedigree method given the lack of quantitative uncertainty information. One respondent commented that

"LCA practitioners do not have an accurate intuitive sense of what is the GSD of the Pedigree matrix". Some respondents found it difficult to provide uncertainties even when they had sufficient experience in this field, partly because the uncertainty characteristics would depend on the characterization models in question. For example, one respondent noted that "GWP and freshwater toxicity will express uncertainties at different orders of magnitude." Such responses are reasonable given that the characterization model for climate change is not regionally sensitive, but that for ecotoxicity is. Thus, applying the same GSDs for multiple impact categories is not desirable. One respondent recommended to use "the distribution coming from the characterization model directly" using empirical data instead of using the Pedigree approach.

However, some respondents commented that they support the use of Pedigree approach for the purpose of filling in the gaps in the uncertainty information in LCIA. One respondent commented that the method "would indeed be worthwhile to quantify the uncertainty of LCIA models". Another respondent noted that "the method could be useful in the absence of uncertainty data".

4.3.3. Criteria to be included in the Pedigree matrix

We asked respondents to what extent they agree or disagree with including each of the six criteria in the Pedigree matrices for LCI and characterization factors. We used numerical values to translate Likert scale responses that numbers from 1 to 5 represent strongly disagree, disagree, neutral, agree, to strongly agree. Table 5 and Table 6 show the rank and average scores of the six criteria used in our study.

4.3.3.1 Criteria for LCI

For LCI, both temporal correlation and geographical correlation were ranked as the top criteria to be included in the Pedigree matrix, and were followed by completeness, then by further technological correlation and reliability (Table 5). Group 1 tended to rank technological correlation higher than completeness and reliability, while respondents from the Group 2 ranked reliability and sample size higher than technological correlation. We included temporal correlation, geographical correlation, completeness, technological correlation, and reliability into the Pedigree matrix for LCI (Table 7).

Table 5. Pedigree matrix criteria selected for LCI and mean scores*.

Rank	All Respondents	Score	Group 1	Score	Group 2	Score
1	Geographical correlation	4.11	Geographical correlation	4.11	Geographical correlation	4.10
2	Temporal correlation	4.11	Temporal correlation	4.11	Temporal correlation	4.10
3	Completeness	3.91	Technological correlation	4.00	Completeness	4.05
4	Technological correlation	3.89	Completeness	3.81	Reliability	3.95
5	Reliability	3.83	Reliability	3.74	Sample size	3.90
6	Sample size	3.32	Sample size	2.89	Technological correlation	3.75

* 1 = strongly disagree, 2 = disagree, 3 = neutral, 4 = agree, and 5 = strongly agree.

4.3.3.1 Criteria for characterization factors

For characterization factors, both the Group 1 and Group 2 came up with the same ranking. Temporal specification was the most important criterion to be included in the Pedigree matrix for characterization factors, followed by geographical specification, model completeness, reliability of underlying science, input data characteristics, and level of consensus. The list of top 5 criteria is shown in Table 6.

Table 6. Pedigree matrix criteria for characterization factors and mean scores.*

Rank	All Respondents	Score	Group 1	Score	Group 2	Score
1	Temporal specification	4.05	Temporal specification	3.96	Temporal specification	4.19
2	Geographical specification	3.93	Geographical specification	3.81	Geographical specification	4.11
3	Model completeness	3.70	Model completeness	3.56	Model completeness	3.89
4	Reliability of underlying science	3.59	Reliability of underlying science	3.41	Reliability of underlying science	3.83
5	Input data characteristics	3.42	Input data characteristics	3.19	Input data characteristics	3.76
6	Level of consensus	3.09	Level of consensus	2.89	Level of consensus	3.39

* 1 = strongly disagree, 2 = disagree, 3 = neutral, 4 = agree, and 5 = strongly agree.

4.3.4. Pedigree matrix obtained from the survey

The respondents were asked to provide their best guesses of GSDs for each level of uncertainty for each criterion for LCI and characterization factor, as well as for weight and life expectancy where the uncertainty is known. Respondents tended to overestimate the GSDs for the distribution of weight and life expectancy at birth. The average ratios of the surveyed GSD to the actual GSD for distributions of weight and life expectancy at birth were 111% and 118%, respectively. Resulting average a and b of equation (4) were 0.60 and 0.40, respectively.

4.3.4.1. Pedigree matrix for LCI

Table 7 shows the Pedigree matrix generated by averaging the responses after the calibration using the distributions of weight and life expectancy at birth. Both the Group 1 and Group 2 gave similar GSD responses to LCI uncertainties. We performed a non-paired t-test for the two groups and found no significant difference between the average of the answers of the two groups to all the cell entries, as the p-value was much greater than 0.05, while the Group 1 tended to give slightly higher GSDs (3%) than the Group 2.

Table 7. Pedigree matrix for LCI from the survey results with GSDs calibrated using GSDs of distributions of weight and life expectancy at birth. The non-calibrated results of the GSDs that the respondents directly provided in the survey are presented in the

supplementary information (Table S2).

	Criteria	Score				
		1 (Low uncertainty)	**2 (Moderately low uncertainty)**	**3 (Moderate uncertainty)**	**4 (Moderately high uncertainty)**	**5 (High uncertainty)**
1	**Reliability**	Verified data based on measurement 1.00	Verified data partly based on assumptions or non-verified data based on measurements 1.09	Non-verified data partly based on assumptions 1.20	Qualified estimate (e.g. by industrial expert) 1.32	Non-qualified estimate 1.59
2	**Completeness**	Representative data from a sufficient sample of sites over an adequate period to even out normal fluctuations 1.00	Representative data from a smaller number of sites but for adequate periods 1.09	Representative data from an adequate number of sites but from shorter periods 1.18	Representative data but from a smaller number of sites and shorter periods or incomplete data from an adequate number of sites and 1.29	Representativeness unknown or incomplete data from a smaller number of sites and/or from shorter periods 1.55
3	**Temporal correlation**	Less than three years of difference to year of study 1.00	Less than six years difference 1.09	Less than 10 years difference 1.18	Less than 15 years difference 1.29	Age of data unknown or more than 15 years of difference 1.51
4	**Geographical correlation**	Data from area under study 1.00	Average data from larger area in which the area under study is included 1.09	Data from area with similar production conditions 1.16	Data from area with slightly similar production conditions 1.28	Data from unknown area or area with very different production conditions 1.57
5	**Technological correlation**	Data from enterprises, processes and materials under study 1.00	Data from processes and material under study but from different enterprises 1.08	Data from processes and materials under study but from different technology 1.22	Data on related processes or materials but same technology 1.33	Data on related processes or materials but different technology 1.63

We also compared the GSDs that the respondents provided for the LCI Pedigree matrix with the GSDs that the Pedigree matrix of ecoinvent uses (Figure 16). We found that respondents generally estimated higher GSDs for LCI than for ecoinvent. The average ratios of non-calibrated GSDs and calibrated GSDs to traditional GSDs were 1.19 and 1.06, respectively, which means that the GSDs after calibration were closer to the GSDs used by ecoinvent. When comparing the respondent GSDs to the empirical-based GSDs produced by Ciroth et al. (Ciroth et al. 2013), it showed that the respondents underestimated the GSDs more than the original GSD scores because GSDs in Ciroth's version are generally greater than the traditional version (Figure 17).

Figure 16. Comparison of the average GSDs in the response and the GSDs in the ecoinvent Pedigree matrix.

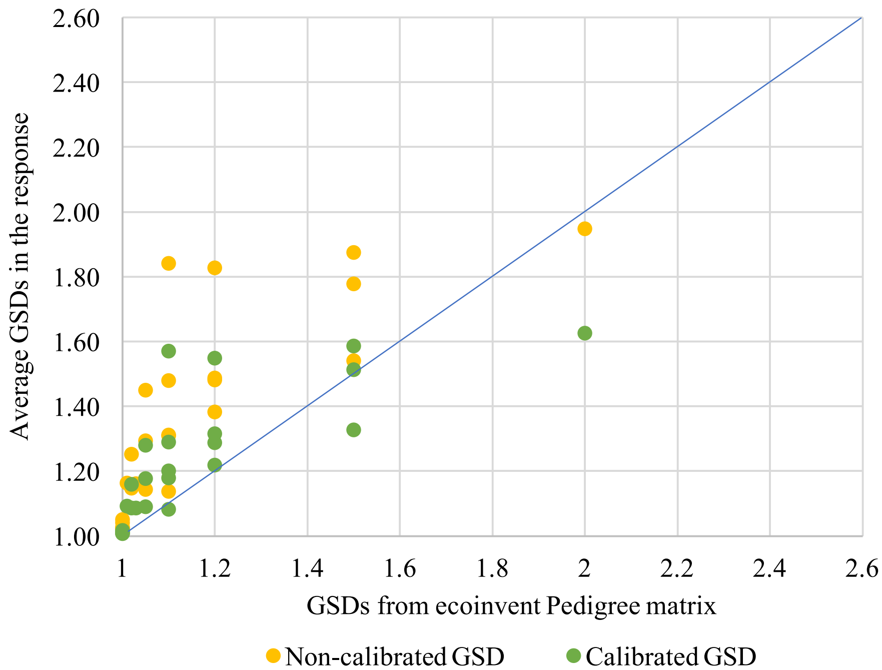
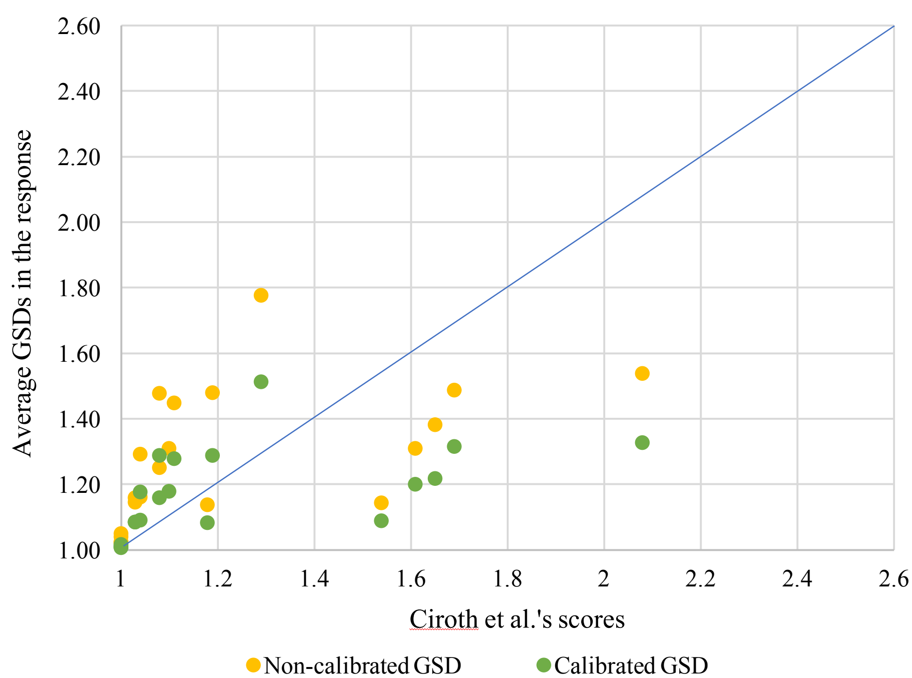


Figure 17. Comparison of the average GSDs in the response and the GSDs in Ciroth et al. (2013).



4.3.4.2. Pedigree matrix for characterization factors

Table 8 shows the Pedigree matrix of the calibrated GSDs for characterization factors from our survey results. Like the LCI results, the Group 1 gave higher GSDs than the Group 2 on average, and the average ratio of the Group 1's GSDs to the Group 2's GSDs was 1.08. We performed statistical non-paired t-test between the average of the answers of the two groups to find whether the two groups provided significantly different GSDs, and we found their responded GSD were not significantly different in general.

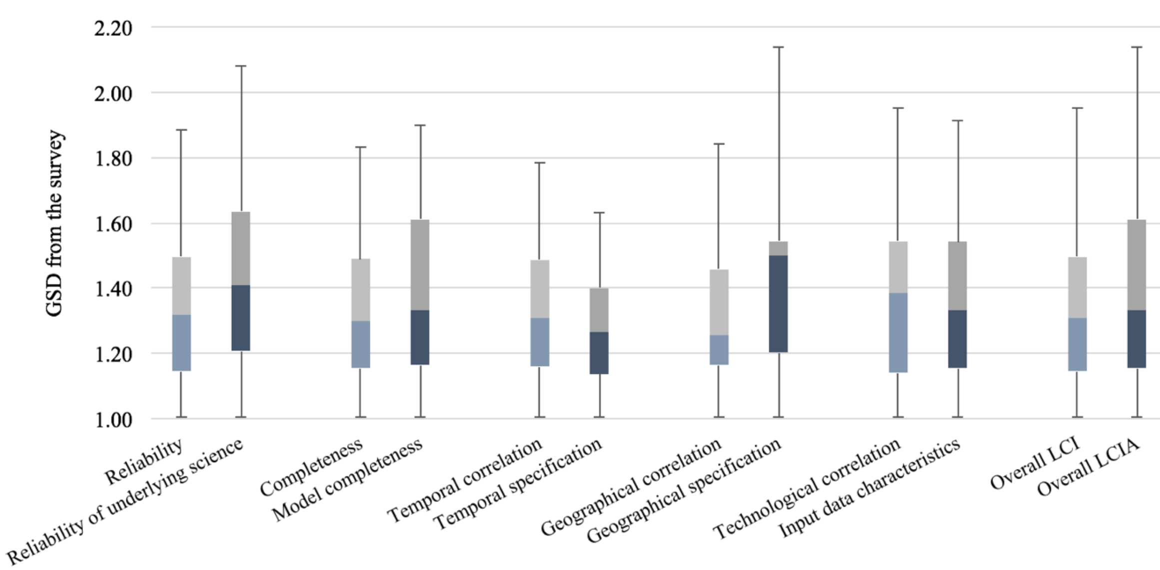
Table 8. Pedigree matrix for characterization factors from the survey results with GSDs calibrated using GSDs of distributions of weight and life expectancy at birth. *The non-calibrated results of the GSDs that the respondents directly provided in the survey are presented in the supplementary information (Table S3).

	Criteria	**Score**				
		1 (Low uncertainty)	**2 (Moderately low uncertainty)**	**3 (Moderate uncertainty)**	**4 (Moderately high uncertainty)**	**5 (High uncertainty)**
1	**Reliability of underlying science**	The model has been published in at least one peer-reviewed journal and has since been independently validated using observation 1.00	The model is based on peer-reviewed results 1.11	The model is based on non-peer-reviewed report 1.21	The model has been documented but has no indication of peer-review 1.38	The model has no documentation on its underlying science 1.70
2	**Model Completeness**	The results of the model have a full coverage of the characterization factors for all elementary flows in an LCI (100%) 1.00	The results of the model have a relatively high coverage of the characterization factors for all elementary flows in an LCI (over 80%) 1.08	The results of the model have a moderate coverage of the characterization factors for all elementary flows in an LCI (over 60%) 1.16	The results of the model have a relatively low coverage of the characterization factors for all elementary flows in an LCI (over 40%) 1.31	The model have a relatively low coverage of the characterization factors for all elementary flows in an LCI (over 40%) The results of the model have a low coverage of the characterization factors for all elementary flows in an LCI (equal to or less than 40%) 1.55
3	**Temporal specification**	The model is a fully dynamic model and considers background concentration and population change for receptors 1.00	The model is a fully dynamic model 1.06	The model is a non-steady-state model, which considers some dynamic components 1.12	The model is a steady-state model 1.18	The model has no indication of its temporal information 1.36
4	**Geographical specification**	The model is spatially explicit with a high level of spatial detail 1.00	The model is spatially explicit with a regional level of detail 1.09	The model provides continental level estimates of the 1.21	The model provides specific archetypes for generic locations 1.26	The model is not spatially explicit 1.58
5	**Input data characteristics**	The input parameters used in the characterization models are an exact measure of the desired quantity 1.00	The input parameters are statistically representative proxies 1.07	The input parameters are proxy values based on some statistical representativeness 1.14	The input parameters are proxies based on expert judgement 1.24	No indication on how input parameters were derived 1.46

4.3.5. Comparison of GSDs for LCI and characterization factors

We also compared the GSDs for LCI and characterization factors provided by the respondents to find which LCA phase has higher perceived uncertainty (Figure 17). In general, GSDs for characterization factors was slightly larger (3%) than those for LCI. The respondents gave much higher uncertainty scores for geographical correlation and reliability criteria and slightly higher uncertainty scores for the completeness criterion for characterization factors than those for LCI. For temporal correlation, the respondents gave lower uncertainty scores for characterization factors than that for LCI. The criterion for LCI, further technological correlation, and the criterion for characterization factors, input data characteristics, are not comparable, but the respondents provided similar GSDs for them.

Figure 17. Comparison between non-calibrated GSDs for LCI and characterization factor Pedigree matrix from the survey. Light blue and grey colors represents the GSDs for LCI criteria, and dark blue and grey colors represents characterization criteria. Each box plot presents the surveyed GSDs for the five uncertainty levels for each indicator.



4.4. Conclusions and discussion

In this study, we surveyed and analyzed perceive uncertainties in characterization factors relative to that in LCI using expert elicitation approach. We found that the perceived uncertainties were generally higher in the characterization factor than in LCI, which is consistent with the statements in the literature (Owens 1997; Huijbregts 1998; Clavreul et al. 2012). However, the difference in mean GSDs between LCI and characterization across all criteria was only marginal (3%). The differences in variations were also larger in characterization (coefficient of variance: 24.4%) than in LCI (coefficient of variance: 22.4%).

About half (49%) of the respondents were in favor of using the Pedigree method to characterize uncertainty in LCA, while 26% of the respondents disapproved the use of the approach. However, the opinions were sharply divided especially among the respondents with 6 years or more experience in LCA; 19% of them strongly approved while 15% strongly disapproved the approach. In general, more experienced group were much more skeptical about the use of Pedigree approach than those with less experience.

Among the criteria examined, the respondents perceived that model reliability and geographical correlation influence the variability in characterization more strongly than the two criteria do in LCI. The respondents generally perceived that temporal correlation is less important in characterizing the uncertainty in characterization than in LCI.

We found it challenging to apply the Pedigree approach to characterization. Our intention was to create different Pedigree matrix for each impact category and per each characterization model, but it didn’t take too long to realize how complicated the questionnaire should be to differentiate characterization models in our survey. Respondents’ time commitment to such survey was another major barrier to create characterization model-specific Pedigree matrix. We believe that the wide variability in responses observed for characterization can be explained in part by the lack of specificity in characterization model in our survey, which is a major limitation in our study.

Overall, our survey result shows that there is no strong consensus among LCA experts on the use of the Pedigree method in LCA, while there seems to be no alternatives available in the near future. The lack of appropriate methods to estimate underlying variability in LCA data is the main barrier to mainstreaming uncertainty analysis in LCA. Given that few disagrees the importance and need of uncertainty analysis in LCA, developing widely accepted methods to estimate underlying variability in LCA data is urgently needed, which would require not only continued research and development by individual researchers but also systematic efforts by international organizations to identify and build consensus on the best practices.

Our survey also confirms that uncertainties in characterization are perceived to be at least as large as those in LCI. Given the virtually nonexistent uncertainty measurements in characterization in today's LCA practices, our results indicate that existing uncertainty analyses in LCA are perceived to cover no more than half of the true uncertainties. Our results call for expediting the efforts to measure uncertainties in characterization and other steps in LCIA.

4.5. Appendix

Included in the Appendix are the survey questions, the demographics of the respondents and the Pedigree matrix for LCI and characterization factors by Group 1 and Group 2.

We collected information such as the time of the respondents have worked in the LCA field and the familiarity of the Pedigree matrix. We assigned them into two groups Group 1: respondents with more than 6 years' experience and familiar with the Pedigree method, and (2) Group 2: respondents with less than 6 years' experience or not familiar with the Pedigree method.

Figure S2. Survey questions.

Q1: How long have you been working in the field of LCA?

◯ < 1 year ◯ 1 - 5 years ◯ 6 - 10 years ◯ > 10 years

Q2: What is the primary sector that you are currently working in?

◯ Academic and Research ◯ Consulting ◯ Corporate ◯ Government ◯ Others

Q3: Where are you located?

◯ Asia ◯ Africa ◯ Australia ◯ Europe ◯ North America ◯ South America

Q4: How familiar are you with the Pedigree approach?

Haven't heard about it	Know very little about it	Know what it is	Familiar with it	Very familiar with it
◯	◯	◯	◯	◯

Q5: Do you believe that the Pedigree approach is an appropriate method to estimate the uncertainty in LCA data?

Strongly disagree	Somewhat disagree	Neutral	Somewhat agree	Strongly agree
◯	◯	◯	◯	◯

Optional question: If you disapprove of the use of the Pedigree approach in uncertainty quantification in LCA, please explain why:

Q6: Please enter your "best guess" of the geometric standard deviation (GSD) associated with the following distributions in each box.

Questions on LCI

Q7: To what extend do you agree or disagree of including the following indicators in a Pedigree matrix for LCI:

Indicator	**Strongly disagree**	**Somewhat disagree**	**Neutral**	**Somewhat agree**	**Strongly agree**
Reliability	◯	◯	◯	◯	◯
Completeness	◯	◯	◯	◯	◯
Temporal correlation	◯	◯	◯	◯	◯
Geographical correlation	◯	◯	◯	◯	◯
Further technological correlation	◯	◯	◯	◯	◯
Sample size	◯	◯	◯	◯	◯

Q8: Please enter the geometric standard deviation (GSD) for the following indicators for LCI based on your expert judgement.

Questions on LCIA

Q9: To what extend do you agree or disagree of including the following indicators in a Pedigree matrix for characterization in LCIA:

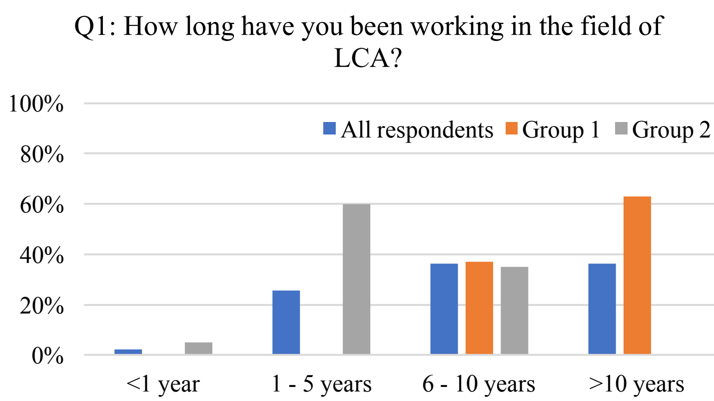
Indicator	**Strongly disagree**	**Somewhat disagree**	**Neutral**	**Somewhat agree**	**Strongly agree**
Level of consensus	◯	◯	◯	◯	◯
Model completeness	◯	◯	◯	◯	◯
Temporal specification	◯	◯	◯	◯	◯
Geographical specification	◯	◯	◯	◯	◯
Reliability of underlying science	◯	◯	◯	◯	◯
Input data characteristics	◯	◯	◯	◯	◯

Q10: Please enter the geometric standard deviation (GSD) for the following indicators for characterization in LCIA based on your expert judgement.

Q11: Please leave any comment that you deem useful for the creation of a Pedigree matrix for LCA characterization

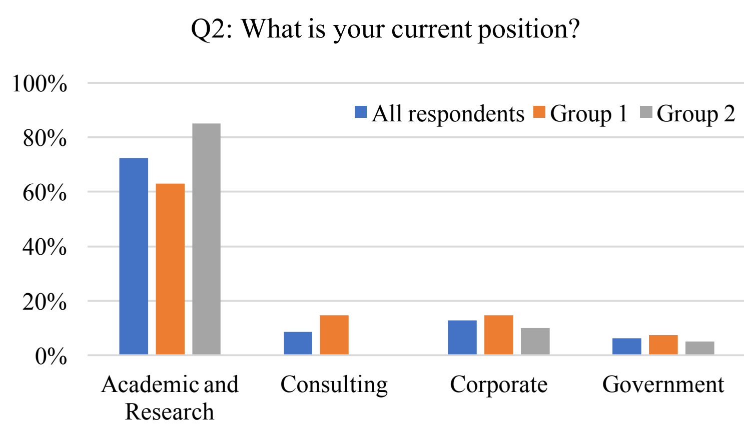
The results show that most (72%) of the respondents have worked more than 6 years in the field of LCA.

Figure S3. Respondents' experiences in LCA.



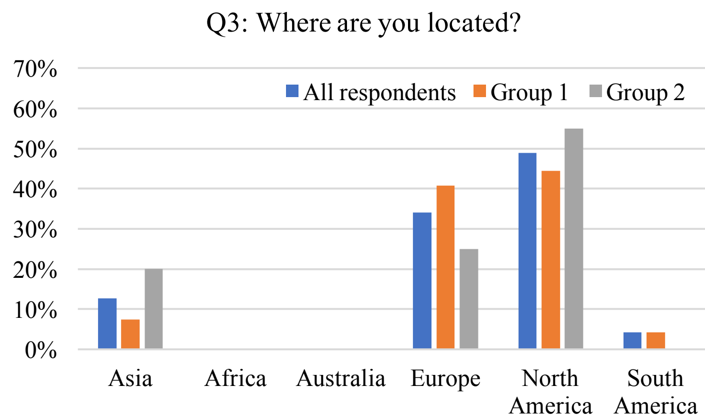
The majority (72%) of the respondents were from the academia and research-related area. Other respondents came from consulting, corporate and government, 9%, 13% and 6%, respectively.

Figure S4. Respondents' current positions.



Nearly half (49%) of the respondents were located in North America, and about one third (34%) of the respondents were from Europe. The rest of the respondents were from Asia and South America.

Figure S5. Respondents' locations.



Most of the respondents are familiar with the Pedigree matrix because we sent out the survey to LCA practitioners who contribute to the literature related to uncertainty in LCA.

Figure S6. Respondents' familiarity with Pedigree matrix.

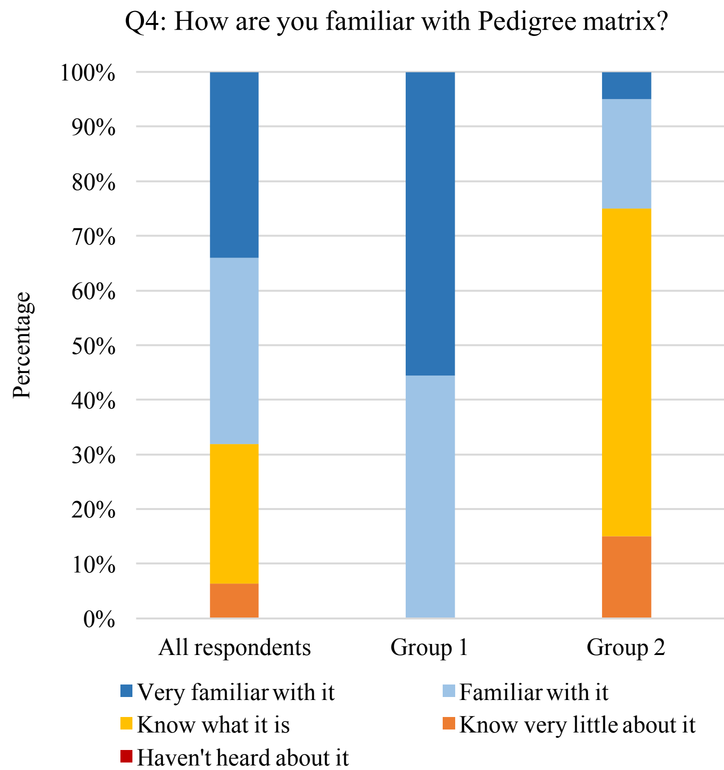


Table S2. Pedigree matrix for LCI from the survey results including non-calibrated and calibrated GSDs using GSDs of distributions of weight and life expectancy at birth (*). The responses from Group 1 and Group 2 are also provided in the table.

	Criteria	Score				
		1 (Low uncertainty)	**2 (Moderately low uncertainty)**	**3 (Moderate uncertainty)**	**4 (Moderately high uncertainty)**	**5 (High uncertainty)**
1	**Reliability**	Verified data based on measurement	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
		Total: Group 1: Group 2:				
	Non-calibrated:	1.03 1.02 1.03	1.14 1.11 1.18	1.31 1.32 1.30	1.49 1.56 1.43	1.88 2.00 1.78
	Calibrated by weight*:	1.01 1.01 1.01	1.09 1.05 1.12	1.20 1.17 1.22	1.32 1.30 1.33	1.59 1.52 1.64
2	**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
		1.02 1.01 1.03	1.15 1.14 1.15	1.29 1.28 1.30	1.48 1.50 1.46	1.83 1.90 1.77
		1.01 1.00 1.01	1.09 1.05 1.12	1.18 1.13 1.21	1.29 1.23 1.33	1.55 1.46 1.62
3	**Temporal correlation**	Less than three years of difference to year of study	Less than six years difference	Less than 10 years difference	Less than 15 years difference	Age of data unknown or more than 15 years of difference
		1.05 1.03 1.07	1.16 1.14 1.18	1.31 1.29 1.33	1.48 1.51 1.46	1.78 1.81 1.75
		1.02 1.01 1.02	1.09 1.06 1.11	1.18 1.13 1.22	1.29 1.23 1.33	1.51 1.38 1.62
4	**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
		1.04 1.01 1.06	1.16 1.14 1.18	1.25 1.24 1.26	1.45 1.47 1.43	1.84 2.00 1.72
		1.01 1.00 1.02	1.09 1.07 1.11	1.16 1.12 1.19	1.28 1.23 1.32	1.57 1.53 1.60
5	**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 same technology	Data on related processes or materials but different technology
		1.03 1.01 1.04	1.14 1.13 1.14	1.38 1.51 1.29	1.54 1.71 1.40	1.95 2.24 1.72
		1.01 1.00 1.01	1.08 1.06 1.10	1.22 1.24 1.20	1.33 1.35 1.31	1.63 1.66 1.60

Table S3. Pedigree matrix for characterization factor from the survey results including non-calibrated and calibrated GSDs using GSDs of distributions of weight and life expectancy at birth (*) and the LCI Pedigree matrix from ecoinvent (**). The responses from Group 1 and Group 2 are provided in the table.

	Criteria		Score				
			1 (Low uncertainty)	**2 (Moderately low uncertainty)**	**3 (Moderate uncertainty)**	**4 (Moderately high uncertainty)**	**5 (High uncertainty)**
1	**Reliability of underlying science**		The model has been published in at least one peer-reviewed journal and has since been independently validated using	The model is based on peer-reviewed results	The model is based on non-peer-reviewed report	The model has been documented but has no indication of peer-review	The model has no documentation on its underlying science
			Total: Group 1 Group 2				
		Non-calibrated:	1.00 1.08 1.09	1.20 1.23 1.18	1.40 1.48 1.34	1.63 1.82 1.48	2.08 2.46 1.77
		Calibrated by weight*:	1.00 1.04 1.03	1.11 1.13 1.08	1.21 1.29 1.15	1.38 1.54 1.24	1.70 1.95 1.48
		Calibrated by LCI**:	1.00 1.03 1.03	1.09 1.12 1.08	1.24 1.24 1.24	1.39 1.58 1.24	1.74 2.06 1.47
2	**Model Completeness**		The results of the model have a full coverage of the characterization factors for all elementary flows in an LCI (100%)	The results of the model have a relatively high coverage of the characterization factors for all elementary flows in an LCI (over 80%)	The results of the model have a moderate coverage of the characterization factors for all elementary flows in an LCI (over 60%)	The results of the model have a relatively low coverage of the characterization factors for all elementary flows in an LCI (over 40%)	The model have a relatively low coverage of the characterization factors for all elementary flows in an LCI (over 40%) The results of the model have a low coverage of the
			1.00 1.00 1.06	1.16 1.15 1.17	1.33 1.35 1.31	1.61 1.70 1.53	1.90 2.05 1.77
			1.00 1.00 1.02	1.07 1.07 1.06	1.16 1.18 1.14	1.31 1.40 1.23	1.55 1.65 1.47
			1.00 1.00 1.02	1.08 1.08 1.08	1.16 1.18 1.14	1.32 1.41 1.26	1.57 1.69 1.46
3	**Temporal specification**		The model is a fully dynamic model and considers background concentration and population change for receptors	The model is a fully dynamic model	The model is a non-steady-state model, which considers some dynamic components	The model is a steady-state model	The model has no indication of its temporal information
			1.00 1.02 1.07	1.13 1.10 1.16	1.26 1.22 1.29	1.40 1.35 1.43	1.63 1.59 1.66
			1.00 1.02 1.03	1.06 1.05 1.07	1.12 1.11 1.13	1.18 1.16 1.20	1.36 1.27 1.43
			1.00 1.01 1.03	1.06 1.05 1.07	1.22 1.30 1.16	1.19 1.16 1.20	1.35 1.28 1.41
4	**Geographical specification**		The model is spatially explicit with a high level of spatial detail	The model is spatially explicit with a regional level of detail	The model provides continental level estimates of the characterization factors	The model provides specific archetypes for generic locations	The model is not spatially explicit
			1.00 1.02 1.07	1.20 1.25 1.16	1.50 1.69 1.34	1.54 1.65 1.45	2.14 2.69 1.68
			1.00 1.01 1.03	1.09 1.11 1.07	1.21 1.28 1.14	1.26 1.30 1.22	1.58 1.74 1.44
			1.00 1.01 1.03	1.10 1.13 1.07	1.24 1.36 1.15	1.26 1.33 1.21	1.62 1.88 1.41
5	**Input data characteristics**		The input parameters used in the characterization models are an exact measure of the desired quantity	The input parameters are statistically representative proxies	The input parameters are proxy values based on some statistical representativeness	The input parameters are proxies based on expert judgement	No indication on how input parameters were derived
			1.00 1.00 1.04	1.15 1.15 1.15	1.33 1.36 1.30	1.54 1.66 1.45	1.91 2.18 1.71
			1.00 1.00 1.01	1.07 1.06 1.07	1.14 1.15 1.13	1.24 1.28 1.22	1.46 1.48 1.45
			1.00 1.00 1.02	1.07 1.07 1.06	1.16 1.18 1.14	1.26 1.33 1.22	1.51 1.60 1.44

Chapter 5. LMDI Approach to Decomposing LCA Uncertainty

5.1. Introduction

Understanding uncertainty in life cycle assessment (LCA) can not only help prioritize research efforts but also facilitate reasonable decisions (Geisler et al. 2005; Basson and Petrie 2007; Lloyd and Ries 2008). Statistical uncertainty analysis has been implemented in many LCA studies and two common forms are sampling methods and analytical approaches (Heijungs 1996; Heijungs and Huijbregts 2004). Among the various statistical methods, Monte Carlo simulation (MCS) is widely used to characterize the variability in LCA, and MCS relies on pre-defined probability distributions (Huijbregts 2002; Kollmuss and Agyeman 2002; Sonnemann et al. 2003; Beltran et al. 2018). LCA software now often supports Monte Carlo simulation and the most widely used data source, ecoinvent database, includes distribution information for 90% of the unit process data (Weidema et al. 2013; Qin and Suh 2017).

The ecoinvent database relies on the method called Pedigree matrix to evaluate data quality of inventory data and provide a distribution including the GSD for most unit process data. The Pedigree matrix was first introduced to the LCA field by Weidema and Wesnæs in 1996 from the literature of uncertainty analysis for environmental science (Funtowicz and Ravetz 1990). The Pedigree approach uses expert elicitation to translate the qualitative characteristics of the data into uncertainty factors which are aggregated to the geometric standard deviation (GSD) (Lewandowska et al. 2004). The distributions of flow data, which are usually assumed to be lognormal, can be estimated from the Pedigree matrix.

By running MCS using the pre-defined distributions such as GSDs, users can generate the distribution of the characterized result although the characterization factors do not contain uncertainty information. The distribution or the range of the LCA result gives the possible results of LCA considering the input, model, and scenario uncertainties, which helps the product designers or decision makers in decision-making.

However, the distributions do not indicate which factor, for example LCI or the characterization factor, contributes the most to the uncertainty so that LCA researchers or practitioners can focuses on the uncertainty reduction for that factor. Furthermore, some LCA scholars claimed that the impact assessment phase of LCA has larger influence on the uncertainty of a LCA result compared with inventory phase (Owens 1996, 1997; Clavreul et al. 2012). However, no empirical analysis or evidence confirmed such statement, and most of the uncertainty analyses only focus on the inventory phase of LCA (Maurice et al. 2000; Lloyd and Ries 2008).

Sensitivity analysis can be used to understand the relative importance of the parameters to uncertainty, but it cannot tell us the exact contribution of each parameter to the uncertainty. Unlike sensitivity analysis, where parameters are tested one at a time, parsing out the MSC results to contributing factor has been a challenge. To determine how much each factor within the LCA phases contributes to the overall uncertainty of the characterized LCA results, index decomposition analysis, a technique from economics, can be used.

5.1.1. The logarithmic mean Divisia index (LMDI) method

Index decomposition analysis (IDA) is used to decompose the influence of factors that contribute to the overall result, and it was first developed to study the impact of structural change on energy use in industry in the late 1970s (Jenne and Cattell 1983; Marlay 1984). IDA has been used to quantify the impact of different factors on the change of energy intensity and extended to many regions and various application areas such as transportation, electricity generation and environmental study (Ang et al. 1998; Paul and Bhattacharya 2004; Malla 2009; Al-Ghandoor et al. 2010). For example, Zhang et al. (2009) used the index decomposition method to decompose the influence of energy-related factors in CO2 emission reduction in China.

In the energy analysis field, many scholars use the logarithmic mean Divisia index (LMDI) method method to decompose the influence of energy factors in CO2 emission reduction (Boyd et al. 1987; Ang and Liu 2001). The LMDI leaves no residuals in the analysis and performs well where there is large variation of variables and zeros in the dataset (Ang 2004; Meng et al. 2018). Therefore, it is appropriate to use with LCA data because LCA data contain many zeros.

The LMDI method is used to decompose changes in the total results over time, but LCI items and characterization factors don't change over time. Thus, the methodology requires modification to decompose the uncertainty of the results. Similar to the problem of CO2 emission from energy use, which aggregates multiple key factors, LCA can be also regarded as an aggregation problem that involves several factors including LCI, characterization, normalization, and weighting, and the change of each factor, e.g. the consumption of iron ore in electric car production, would contribute to the change of the final LCA result, e.g. the ecosystem health impact from the production of an electric vehicle.

5.1.2. Aim of this study

This work aims to provide a methodology for quantifying the contribution of each factor within the LCA phases to the overall LCA uncertainty using LMDI method. This paper is the first attempt to apply the technique of decomposition analysis to decompose the uncertainty of LCA from LCI and impact assessment phases. A detailed methodology description of using LMDI in finding the contribution of intermediate LCA phase in overall LCA uncertainty is provided in the Methods section. The paper provides a practical guide for future LCA practitioners to use LMDI approach to decompose the effect of each intermediate LCA step to the uncertainty of final LCA output.

The method is demonstrated using a case study on natural gas focusing on two impact categories: global warming and USETox human health impacts. After each run of MCS, we calculated the difference in the simulated and deterministic category indicator results, and decomposed the difference using LMDI decomposition method, which returns the contribution of each factor to the difference of the run. Then, we analyzed the statistical properties of the contributions by each factor after 1,000 MCS runs.

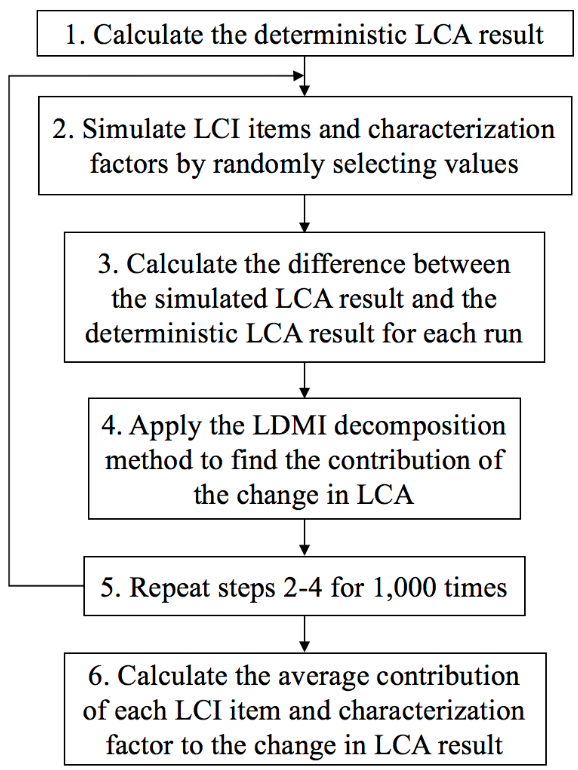
5.2. Methods

In this study, we developed a new method which incorporates Monte Carlo simulation with LMDI method to find the contribution of each intermediate LCA phase to the overall uncertainty in LCA results. Because characterization factors do not include uncertainty information or distributions, we need to estimate the uncertainty for the characterization factors. Our concurrent work provides an estimation for characterization factors using the Pedigree approach based on the survey results from LCA experts' judgement (Qin et al.). Therefore, we used the uncertainty estimates from the Pedigree matrix for characterization factors to generate the distributions used for the MCS in the analysis. The data used in the study are summarized in Section Methods 5.2.2 and the full dataset can be found in the Supplementary Information.

The LCI items and characterization factors don't change over time but they follow certain probability distributions. The LCI items and characterization factors can be considered as stochastic variables, and the changes can be simulated from the MCS. We simulated the changes of LCI items, characterization factors, and the characterized LCA results and calculated the contribution of each item/factor to the overall change in characterized results for each run.

To find the contribution of each LCA intermediate factor to the overall LCA uncertainty, the analysis involves 6 major steps (Figure 18).

Figure 18. Flow diagram of the use of LDMI method in decomposing the uncertainty of LCA result



The first step is to calculate the deterministic LCA result, h_i^0. The second step is to simulate LCI items and characterization factors by randomly selecting values from their specified distributions and store both simulated values and the calculated LCA result, h_i^K. The third step is to calculate the difference between the simulated LCA result and the deterministic LCA result for each simulation, and the difference is considered the change of the LCA results. The fourth step is to apply the LDMI decomposition method to find the contribution of the change in LCA result into each LCI item and characterization factor, and this step will be further explained in the next section. After repeating steps 2-4 1,000 times, the final step is to calculate the average contribution of each LCI item and characterization factor to the change in LCA result.

Steps 1, 2, 3, 5, and 6 follow the traditional Monte Carlo simulation procedures (Qin and Suh 2017), and this paper mainly focuses on step 4 and provides the explanation of how the LDMI is applied into the change in characterized LCA results in the following section.

5.2.1. LDMI method for characterized results

The characterized LCA result is calculated through

$$h_i = \sum_{i,j} h_{i,j} = \sum_{i,j} c_{i,j}\, m_j \qquad (6)$$

where, h_i is the characterized LCA result for characterization model i;

$h_{i,j}$ is the characterized LCA result from elementary flow j for characterization model i;

$c_{i,j}$ is the characterization factor for the elementary flow j in impact category i;

m_j is the inventory for the elementary flow j.

To analyze the influence of LCI and characterization factor, $c_{i,j}$ and m_j respectively, on the change of characterized LCA results, h_i, we introduced the simulation time, K, to calculate the change of h_i^0 and h_i^K, $(h_i^K - h_i^0)$, for each simulation.

LCA formula without normalization and weighting phases in simulation K, as follow:

$$h_i^K = \sum_{i,j} h_{i,j}^K = \sum_{i,j} c_{i,j}^K\, m_j^K \qquad (7)$$

where, h_i^K is characterized LCA result for characterization model *i* in simulation *K*;

$h_{i,j}^K$ the characterized LCA result from elementary flow *j* for characterization model *i* in simulation *K;*

$c_{i,j}^K$ is the characterization factor for the elementary flow *j* in impact category *i* in simulation *K*;

m_j^K is the inventory for the elementary flow *j* in simulation *K*.

The aggregate changes from h_i^0 in deterministic result to h_i^K in simulation *K* followed LMDI approach by Ang (2005, 2015). The multiplication decomposition suggests:

$$D_{h_i} = h_i^K / h_i^0 = D_{\mathrm{c}} D_m \tag{8}$$

In addictive decomposition method, the difference can be decomposed:

$$\Delta h_i = h_i^K - h_i^0 = \Delta h_{ic} + \Delta h_{im} \tag{9}$$

Using the logarithmic mean Divisia index (LMDI) approach, the effect of characterization factor *c* and inventory *m* for multiplication decomposition:

$$D_{\mathrm{c}} = \exp\left(\sum_j \frac{L\left(h_{i,j}^K, h_{i,j}^0\right)}{L(h^K, h^0)} \ln\left(\frac{c_{i,j}^K}{c_{i,j}^0}\right)\right) = \exp\left(\sum_j \frac{(h_{i,j}^K - h_{i,j}^0)/(\ln h_{i,j}^K - \ln h_{i,j}^0)}{(h_i^K - h_i^0)/(\ln h_i^K - \ln h_i^0)} \ln\left(\frac{c_{i,j}^K}{c_{i,j}^0}\right)\right) \tag{10}$$

$$D_m = \exp\left(\sum_j \frac{L\left(h_{i,j}^K, h_{i,j}^0\right)}{L(h^K, h^0)} \ln\left(\frac{m_j^K}{m_j^0}\right)\right) = \exp\left(\sum_j \frac{(h_{i,j}^K - h_{i,j}^0)/(\ln h_{i,j}^K - \ln h_{i,j}^0)}{(h_i^K - h_i^0)/(\ln h_i^K - \ln h_i^0)} \ln\left(\frac{m_j^K}{m_j^0}\right)\right) \tag{11}$$

For addictive decomposition:

$$\Delta h_{ic} = \sum_j L\left(h_{i,j}^K, h_{i,j}^0\right) \ln\left(\frac{c_{i,j}^K}{c_{i,j}^0}\right) = \sum_j \frac{h_{i,j}^K - h_{i,j}^0}{\ln h_{i,j}^K - \ln h_{i,j}^0} \ln\left(\frac{c_{i,j}^K}{c_{i,j}^0}\right) \tag{12}$$

$$\Delta h_{im} = \sum_j L\left(h_{i,j}^K, h_{i,j}^0\right) \ln\left(\frac{m_j^K}{m_j^0}\right) = \sum_j \frac{h_{i,j}^K - h_{i,j}^0}{\ln h_{i,j}^K - \ln h_{i,j}^0} \ln\left(\frac{m_j^K}{m_j^0}\right) \tag{13}$$

where $L(a, b) = (a - b)(\ln a - \ln b)$ is the logarithmic mean (Ang, 2004).

5.2.2. Case study

Natural gas was the top source (33%) of electricity generation in the U.S in 2018 and will remain primary energy source in the future (EIA 2019). We chose natural gas as an example to demonstrate the use of this method for analyzing the distribution of the entire LCA result and the relative contribution of LCI and characterization factor. We applied the Pedigree method to estimate the uncertainty for LCI and characterization factor in which the scores were obtained from the same group of experts (Qin et al.). Global warming potential and human health impact categories were chosen for the comparison because global warming impact is time and space-insensitive while the human health impact is time and space-sensitive.

From an LCA database, e.g. ecoinvent, it shows that producing 1 cubic meter of natural gas in the U.S. will generate 0.45 kg of CO2 equivalence and 1.27e-8 disability-adjusted life year. But, how certain are these values? In order to find the distributions of the results, we need to first understand and obtain the uncertainty of each LCI item and each characterization factor. We applied the same GSDs for LCI and impact factors using the Pedigree approach for the demonstrational purpose of this study.

The input data used for generating 1 cubic meter of natural gas in the U.S. and their distributions were presented in Table 8. Each value in Table 8 is the deterministic value from ecoinvent. The distribution type is lognormal because the Pedigree method assumes the data follow lognormal distribution. The Pedigree score is determined based on the characteristics of the data according to the criteria from the Pedigree matrix. The Pedigree tables used for LCIs and characterization factors are generated from a survey of experts in the previous chapter.

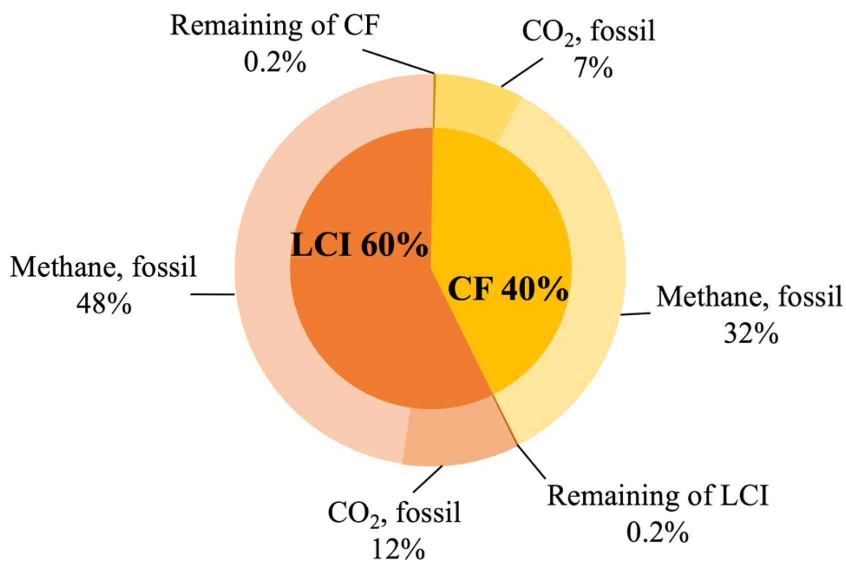
Table 8. Samples of LCI of natural gas production in the U.S. and characterization factors of GWP 100 from IPCC 2013 (Stocker 2014) and human health non-cancer impact from USEtox (Rosenbaum et al. 2008). The criteria for estimating the Pedigree score can be found in Chapter 4.

LCI	Value	Distribution	Pedigree score	GSD
Methane (kg)	1.2	Lognormal	(2,2,3,3,2)	1.16
CO_2 (kg)	0.03	Lognormal	(2,2,3,3,2)	1.16
Chromium VI (kg)	0.0000067	Lognormal	(2,2,3,3,2)	1.16
…				
Characterization (GWP 100 CO_2 eq.)				
Methane (kg)	24	Lognormal	(1,1,3,1,1)	1.10
CO_2 (kg)	1	Lognormal	(1,1,3,1,1)	1.10
Chromium VI (kg)	0	-		
…				
Characterization (Human health non-cancer DALY)				
Methane (kg)	0	-		
CO_2 (kg)	0	-		
Chromium VI (kg)	258	Lognormal	(2,5,4,5,3)	1.43
…				

5.2 Results

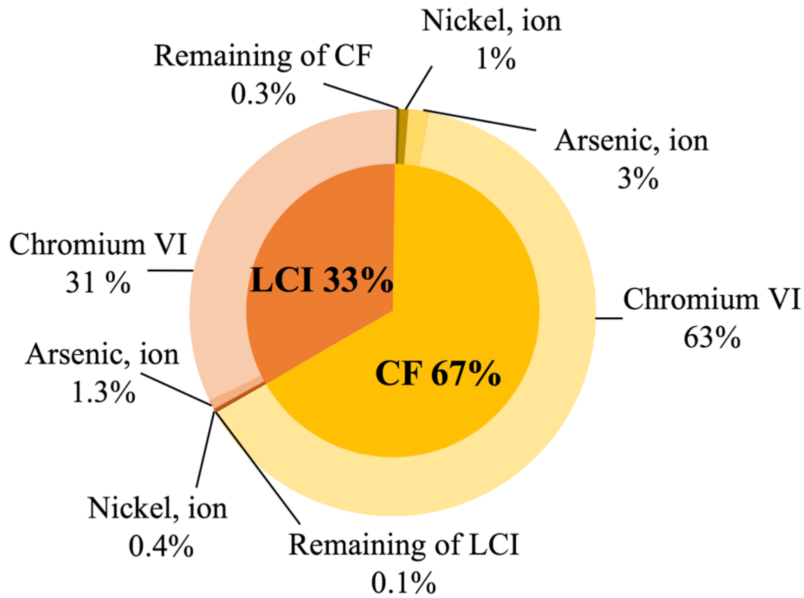
After running 1,000 MCSs of LDMI analysis, the results showed that LCI and characterization factors contribute 60% and 40%, respectively, of the uncertainty in the characterized results in GWP 100 IPCC 2013 (Figure 19). 48% of the uncertainty in climate change impact is contributed by the inventory of methane emissions, and CO2 emissions contribute 12%, and the remaining items in LCI only contribute to 0.2% of the overall uncertainty in characterized LCA result. The characterization factors of methane, CO2, and the remaining characterization factors contributed to 32%, 7%, and 0.2% respectively.

Figure 19. The LDMI decomposition result for natural gas production in GWP 100. The values are subject to the uncertainties based on our selections of the Pedigree scores in the Pedigree matrix in our study.



For the human health characterization model, we used USEtox in the case study. The results indicated that 33% and 67% of the uncertainty can be attributed to LCI and characterization factor respectively (Figure 20). Among the LCI items, Chromium VI, hexavalent chromium, which is toxic chemical that can cause cancer through the eyes, skin, and respiratory system, contributed to 31% of the overall uncertainty. Arsenic, nickel, and the remaining of the LCI contributed to 1.3%, 0.4%, and 0.1% respectively to the uncertainty in LCA result. Among the characterization factors, Chromium VI contributed 63% of the overall uncertainty. Arsenic, nickel, and the remaining of the LCI contributed to 3%, 1%, and 0.3% respectively to the uncertainty in LCA result.

Figure 20. The LDMI decomposition result for natural gas production in human health USEtox non-cancer model. The values are subject to the uncertainties that we defined in our study.



Both cases in GWP 100 and human health USEtox cancer suggested that the top 2 or 3 factors in LCI and characterization factors contributed to the majority (>99%) of the uncertainty, and the remaining of LCI and characterization factors only had little (<1%) influence of the overall uncertainty of the characterized results. Improving the reliability of those top contributors could largely reduce the uncertainty of the LCA result.

These results were based on the uncertainty estimates from the Pedigree matrix for both LCI and characterization factor, which was only used as a proxy in the absence of real uncertainty information. Therefore, the results were dependent on the scores in the Pedigree matrix and the characteristics we choose for LCI and characterization model. Though the uncertainty may be underestimated for the characterization factors, the main purpose of this paper is to demonstrate of the method of using the LDMI approach, and the uncertainty values, therefore, can be improved and replaced in the future study.

5.3 Discussion and conclusions

This paper introduced the mathematical solution to quantifying the contribution of each LCA phase to the overall uncertainty. Unlike sensitivity analysis which only provides the ranking of the importance of the factors to the results, the innovative approach proposed in this paper can generate the percentage of the contribution to the uncertainty in the result. The proposed method used LMDI decomposition method combined with Monte Carlo simulation, which does not require additional time from simulation if uncertainty analysis is also performed. This work can fill the gap in the literature of analyzing the uncertainty contribution in LCA, which provides statistics for all the contributors so that the uncertainty can be further investigated and reduced, especially for those top contributors.

In the case study, the results of decomposing uncertainties showed that LCI and characterization factors contributed 60% and 40%, respectively, to the uncertainties in the climate change impact of natural gas production in the U.S. LCI and characterization factors used in natural gas production also contributed 33% and 67%, respectively, to the uncertainties in the human health impact of USEtox. The proposed method cannot only provide the uncertainty decomposition of the uncertainties in LCA results into LCA phases but also into each LCI item and characterization factor. 48% of the uncertainty in climate change impact was contributed by the inventory of methane emissions, and 32% was due to the characterization factor of methane. In the human health impact category, the inventory and the characterization factor of Chromium VI contributes 31% and 63%, respectively, of the overall uncertainty. The top two factors in both impact categories contributed the majority ($\geq$ 80%) meaning that equal to or more than 80% of the uncertainty of the characterized LCA results can be reduced if we can find the accurate value for these two factors.

The approach presented in this study only considered parametric uncertainty, which is the most commonly addressed uncertainty in LCA studies (Lloyd and Ries 2008). Future studies can incorporate model and scenario uncertainties.

One area can be improved in the future would be more accurate uncertainty estimates for the LCI and characterization factors because all the values are derived from the Pedigree matrix which is developed from expert judgement. The uncertainty estimates from the Pedigree method use lognormal probability distributions, but the proposed LMDI combined with Monte Carlo simulation can also work with other distributions. More investigations are needed to determine uncertainty values for both LCI and characterization factors. Since most LCA studies only focus on the uncertainty analysis for the LCI part, more efforts need to be made into the uncertainty aspect for the characterization factors. Improved information can be used and replaced in this study when the uncertainty information is available for characterization models in the future.

The method can be used for other LCA calculations to analyze the influence of LCI, characterization, (normalization, and weighting, which can be found in the SI) to the final LCA uncertainty after the uncertainties for each LCA phase are defined. This approach can be also extended to other related fields such as material flow analysis to understand the contribution of the uncertainty. Using this approach, LCA practitioners can decompose the overall variability in the results to the underlying contributors, which can be used to prioritize the parameters that need further refinement to reduce overall uncertainty of an LCA result.

5.4 Appendix

Included in the Appendix is the LDMI method including the phases of LCI, characterization, normalization, and weighting in LCA. The equations LDMI method for normalized and weighted results are provided in the following.

LCA formula including normalization and weighting phases:

$$W = \sum_{i} w_i\,(h_i/n_i) = \sum_{i} w_i\,(h_i \frac{1}{n_i}) = \sum_{i} w_i\, h_i q_i = \sum_{i,j} w_i\, c_{i,j} m_j q_i$$

where, W is the normalized and weighted LCA result;

w_i is the weighting factor for impact category *i;*

h_i is characterized LCA result of characterization model *i;*

n_i is the reference impact factor for impact category *i*;

q_i is the divided reference impact factor for impact category *i;*

$c_{i,j}$ is the characterization factor for the elementary flow j in impact category *i*;

m_j is the inventory for the elementary flow j.

To find how the normalized and weighted result is influenced by the four factors including inventory, characterization factor, reference impact factor, and weighting factor, we have

$D_W = W^K/W^0 = D_{\mathrm{w}} D_c D_{\mathrm{m}} D_q$ (multiplication decomposition)

$\Delta W = W^K - W^0 = \Delta V_w + \Delta V_{\mathrm{c}} + \Delta V_m + \Delta V_q$ (addictive decomposition)

Multiplication decomposition:

$$D_{\mathrm{w}} = \exp\left(\sum_{i,j} \frac{L\left(W_{i,j}^K, W_{i,j}^0\right)}{L(W^K, W^0)} \ln\left(\frac{w_i^K}{w_i^0}\right)\right)$$

$$= \exp\left(\sum_{i,j} \frac{(W_{i,j}^K - W_{i,j}^0)/(\ln W_{i,j}^K - \ln W_{i,j}^0)}{(W_i^K - W)/(\ln W_i^K - \ln W_i^0)} \ln\left(\frac{w_i^K}{w_i^0}\right)\right)$$

$$D_{\mathrm{c}} = \exp\left(\sum_{i,j} \frac{L\left(W_{i,j}^K, W_{i,j}^0\right)}{L(W^K, W^0)} \ln\left(\frac{c_{i,j}^K}{c_{i,j}^0}\right)\right)$$

$$= \exp\left(\sum_{i,j} \frac{(W_{i,j}^k - W_{i,j}^0)/(\ln W_{i,j}^K - \ln W_{i,j}^0)}{(W_i^K - W)/(\ln W_i^K - \ln W_i^0)} \ln\left(\frac{c_{i,j}^K}{c_{i,j}^0}\right)\right)$$

$$D_{\mathrm{m}} = \exp\left(\sum_{i,j} \frac{L\left(W_{i,j}^K, W_{i,j}^0\right)}{L(W^K, W^0)} \ln\left(\frac{m_j^K}{m_j^0}\right)\right)$$

$$= \exp\left(\sum_{i,j} \frac{(W_{i,j}^K - W_{i,j}^0)/(\ln W_{i,j}^K - \ln W_{i,j}^0)}{(W_i^K - W)/(\ln W_i^K - \ln W_i^0)} \ln\left(\frac{m_j^K}{m_j^0}\right)\right)$$

$$D_{\mathrm{q}} = \exp\left(\sum_{i,j} \frac{L\left(W_{i,j}^K, W_{i,j}^0\right)}{L(W^K, W^0)} \ln\left(\frac{q_i^K}{q_i^0}\right)\right)$$

$$= \exp\left(\sum_{i,j} \frac{(W_{i,j}^K - W_{i,j}^0)/(\ln W_{i,j}^K - \ln W_{i,j}^0)}{(W_i^K - W)/(\ln W_i^K - \ln W_i^0)} \ln\left(\frac{q_i^K}{q_i^0}\right)\right)$$

Addictive decomposition:

$$\Delta W_{\mathrm{w}} = \sum_{i,j} L\left(W_{i,j}^K, W_{i,j}^0\right) \ln\left(\frac{w_i^K}{w_i^0}\right) = \sum_{i,j} \frac{W_{i,j}^K - W_{i,j}^0}{\ln W_{i,j}^K - \ln W_{i,j}^0} \ln\left(\frac{w_i^K}{w_i^K}\right)$$

$$\Delta W_{\mathrm{c}} = \sum_{i,j} L\left(W_{i,j}^K, W_{i,j}^0\right) \ln\left(\frac{c_{i,j}^K}{c_{i,j}^0}\right) = \sum_{i,j} \frac{W_{i,j}^K - W_{i,j}^0}{\ln W_{i,j}^K - \ln W_{i,j}^0} \ln\left(\frac{c_{i,j}^K}{c_{i,j}^0}\right)$$

$$\Delta W_{\mathrm{m}} = \sum_{i,j} L\left(W_{i,j}^{K}, W_{i,j}^{0}\right) \ln\left(\frac{m_j^K}{m_j^0}\right) = \sum_{i,j} \frac{W_{i,j}^{K} - W_{i,j}^{0}}{\ln W_{i,j}^{K} - \ln W_{i,j}^{0}} \ln\left(\frac{m_j^K}{m_j^0}\right)$$

$$\Delta W_{\mathrm{w}} = \sum_{i,j} L\left(W_{i,j}^{K}, W_{i,j}^{0}\right) \ln\left(\frac{q_i^K}{q_i^0}\right) = \sum_{i,j} \frac{W_{i,j}^{K} - W_{i,j}^{0}}{\ln W_{i,j}^{K} - \ln W_{i,j}^{0}} \ln\left(\frac{q_i^K}{q_i^0}\right)$$

Again, a step before applying the equations is to have the uncertainty of all the factors involved in the calculation.

Chapter 6. Conclusions

This dissertation developed a series of methodologies of improving the efficiency of running uncertainty analysis in LCA, filling the gap of the lack of uncertainty in the LCIA phase, and finding the contribution of LCI and characterization to the overall LCA uncertainty.

The first study proposed a method of storing distributions as uncertainty information for life cycle inventory for the purpose of saving computation time and cost for running uncertainty analysis in LCA. The study suggested that lognormal distribution (i.e. median and GSD) can efficiently present aggregate LCIs. Though the current database has uncertainty values for unit process inventory, conducting uncertainty analysis starting from the unit process level is neither time-efficient nor necessary for most studies. Therefore, the determination of the distribution that best fits the aggregate LCIs is needed. It would help improve the efficiency of storing uncertainty data and performing uncertainty analysis in LCA by saving computation time and storage of LCI data.

In Chapter 3, the results indicated that pre-calculated uncertainty values can be used as a proxy for understanding the uncertainty and variability in a comparative LCA study especially when adequate computational resources are lacking. The study evaluated the probability for an LCA practitioner to make an erroneous conclusion due to the use of pre-calculated uncertainty values instead of fully dependent sampling in a comparative LCA setting. The number of unit processes is increasing for many LCI databases, adding to the challenge of running MCSs in a PC-environment in the future. The additional errors due to the use of pre-calculated uncertainty values shown in the study seem justifiable if the alternative is no uncertainty analysis due to the lack of computational resources needed for fully dependent sampling.

Aggregate LCI uncertainty is only one of the steps in the analysis of LCA uncertainty. Not only LCI uncertainty, but also the uncertainty from impact assessment should be assessed in order to achieve the overall uncertainty of the final LCA results. Chapter 4 collected LCA practitioners' opinions on the use of Pedigree approach to estimate the uncertainty in LCI and characterization factors and their perceived uncertainty values for Pedigree matrix. The results showed that nearly half (49%) of the respondents strongly agree, 29% of them chose "neutral", and 26% disagree of using Pedigree matrix to estimate uncertainty in LCA data. The results showed that the experienced group is more skeptical about the use of Pedigree approach in estimating uncertainty in LCA, and the inexperienced group generally believes that Pedigree approach is a good method for uncertainty estimation. Pedigree approach is not a perfectly accurate method to quantify the uncertainties in LCI and LCIA phases. However, due to the lack of measurement data which causes burdens of validating the scores with the actual uncertainty, Pedigree approach, at least, provides an estimate about the uncertainty in the absence of better information and works when comparing uncertainties for multiple LCAs. Future work could focus on collecting measurement data to estimate more accurate GSDs for Pedigree matrix and run characterization model with actual uncertainties.

After the uncertainty values are determined for LCI and LCIA phases, another question is to find the influence of the uncertainty from each LCA step on the final LCA result. Chapter 5 demonstrated the application of LMDI method to quantify the contribution of each LCA phase to the overall LCA uncertainty. The mathematical solution was introduced combining LMDI method and Motne Carlo simulation to quantify the contribution of each LCA phase to the overall uncertainty. This work can fill the gap in the literature of analyzing the uncertainty contribution in LCA. Future study can follow the methodology provided in this paper to analyze the influence of LCI, characterization, normalization, and weighting to the final LCA uncertainty after the uncertainties for each LCA for the four phases are defined.

My research contributes to the improvement of the efficiency of performing uncertainty analysis in LCA and the understanding of the effects of LCA steps to the overall uncertainty of LCA. The approaches and the findings can be applied to all other problems outside of LCA that utilize MCS.

References

Al-Ghandoor A, Al-Hinti I, Mukattash A, Al-Abdallat Y (2010) Decomposition analysis of electricity use in the Jordanian industrial sector. Int J Sustain Energy 29:233–244

Alpert M, Raiffa H (1982) A progress report on the training of probability assessors

Althaus H, Doka G, Dones R, et al (2007) Overview and Methodology: Data v2. 0 (2007). Ecoinvent Cent Zurich CH Retrieved Httpwww Ecoinvent Orgdatabaseolder-Versionsecoinvent-Version-2methodology-Ofecoinvent-2methodology--Ecoinvent-2 Html

Amara RC, Lipinski AJ (1971) Some views on the use of expert judgment. Technol Forecast Soc Change 3:279–289

Ang BW (2015) LMDI decomposition approach: a guide for implementation. Energy Policy 86:233–238

Ang BW (2004) Decomposition analysis for policymaking in energy:: which is the preferred method? Energy Policy 32:1131–1139

Ang BW (2005) The LMDI approach to decomposition analysis: a practical guide. Energy Policy 33:867–871

Ang BW, Liu FL (2001) A new energy decomposition method: perfect in decomposition and consistent in aggregation. Energy 26:537–548

Ang BW, Zhang FQ, Choi K-H (1998) Factorizing changes in energy and environmental indicators through decomposition. Energy 23:489–495

Armstrong JS (2008) Methods to elicit forecasts from groups: Delphi and prediction markets compared

Ayyub BM (2000) Methods for Expert Opinion Elicitation of Probabilities and Consequences for Corps Facilities. US Army Corps Eng IWR Rep No 00-R-10

Azapagic A (1999) Life cycle assessment and its application to process selection, design and optimisation. Chem Eng J 73:1–21

Baker JW, Lepech MD (2009) Treatment of uncertainties in life cycle assessment. In: Proceedings of the 10th international congress on structural safety and reliability, Osaka, Japan. pp 13–17

Basson L, Petrie JG (2007) An integrated approach for the consideration of uncertainty in decision making supported by Life Cycle Assessment. Environ Model Softw 22:167–176

Beltran MM, Pomponi F, Guinée JB, Heijungs R (2018) Uncertainty Analysis in Embodied Carbon Assessments: What Are the Implications of Its Omission? In: Embodied Carbon in Buildings. Springer, pp 3–21

Björklund AE (2002) Survey of approaches to improve reliability in LCA. Int J Life Cycle Assess 7:64–72

Bojacá CR, Schrevens E (2010) Parameter uncertainty in LCA: stochastic sampling under correlation. Int J Life Cycle Assess 15:238–246

Boyd G, McDonald JF, Ross M, Hanson DA (1987) Separating the changing composition of US manufacturing production from energy efficiency improvements: a Divisia index approach. Energy J 8:77–96

Brown B, Cochran S, Dalkey N (1969) The Delphi method, II: structure of experiments. RAND CORP SANTA MONICA CALIF

Burgess AA, Brennan DJ (2001) Application of life cycle assessment to chemical processes. Chem Eng Sci 56:2589–2604

Castaings W, Borgonovo E, Morris MD, Tarantola S (2012) Sampling strategies in density-based sensitivity analysis. Environ Model Softw 38:13–26

Cederberg C, Stadig M (2003) System expansion and allocation in life cycle assessment of milk and beef production. Int J Life Cycle Assess 8:350–356

Cellura M, Longo S, Mistretta M (2011) Sensitivity analysis to quantify uncertainty in life cycle assessment: the case study of an Italian tile. Renew Sustain Energy Rev 15:4697–4705

CIA (2018) The World Factbook Life Expectancy At Birth. Central Intelligence Agency

Ciroth A, Fleischer G, Steinbach J (2004) Uncertainty calculation in life cycle assessments. Int J Life Cycle Assess 9:216–226

Ciroth A, Muller S, Weidema B, Lesage P (2013) Empirically based uncertainty factors for the pedigree matrix in ecoinvent. Int J Life Cycle Assess 1–11

Clavreul J, Guyonnet D (2013) Stochastic and epistemic uncertainty propagation in LCA. Int J Life Cycle Assess

Clavreul J, Guyonnet D, Christensen TH (2012) Quantifying uncertainty in LCA-modelling of waste management systems. Waste Manag 32:2482–2495

Clavreul J, Guyonnet D, Tonini D, Christensen TH (2013) Stochastic and epistemic uncertainty propagation in LCA. Int J Life Cycle Assess 18:1393–1403

Clemen RT, Winkler RL (1985) Limits for the precision and value of information from dependent sources. Oper Res 33:427–442

Cooke R, Goossens L (1990) The accident sequence precursor methodology for the European Post-Seveso era. Reliab Eng Syst Saf 27:117–130

Cooke RM (1991) Experts in uncertainty: opinion and subjective probability in science. Oxford University Press on Demand

Cooper JS, Noon M, Kahn E (2012) Parameterization in Life Cycle Assessment inventory data: review of current use and the representation of uncertainty. Int J Life Cycle Assess 17:689–695

Cucurachi S, Heijungs R (2014) Characterisation factors for life cycle impact assessment of sound emissions. Sci Total Environ

Cucurachi S, van der Giesen CC, Heijungs R, de Snoo G (2016) No Matter - How? Dealing with Matter-less Stressors in LCA of Wind Energy Systems. J Ind Ecol Early view:

Czembor CA, Vesk PA (2009) Incorporating between-expert uncertainty into state-and-transition simulation models for forest restoration. For Ecol Manag 259:165–175

de Bruijn H, van Duin R, Huijbregts MAJ, Guinee JB (2002) Handbook on Life Cycle Assessment. Springer Netherlands, Dordrecht

de Franca Doria M, Boyd E, Tompkins EL, Adger WN (2009) Using expert elicitation to define successful adaptation to climate change. Environ Sci Policy 12:810–819

Dennis B, Patil G (1984) The gamma distribution and weighted multimodal gamma distributions as models of population abundance. Math Biosci

Edelen A, Ingwersen WW (2018) The creation, management, and use of data quality information for life cycle assessment. Int J Life Cycle Assess 23:759–772

EIA (2019) Short-Term Energy Outlook. U.S. Energy Information Administration, Washington, D.C.

Fava J (1994) Life-cycle Assessment Data Quality: A Conceptual Framework: Workshop Report

Ferrell WR (1994) Discrete subjective probabilities and decision analysis: Elicitation, calibration and combination.

Finnveden G, Hauschild MZ, Ekvall T, et al (2009) Recent developments in Life Cycle Assessment. J Environ Manage 91:1–21. doi: 10.1016/j.jenvman.2009.06.018

Frey C (1998) Briefing paper part 1: introduction to uncertainty analysis. Dep Civ Eng N C State Univ

Frischknecht R, Jungbluth N, Althaus H-J, et al (2005) The ecoinvent database: Overview and methodological framework (7 pp). Int J Life Cycle Assess 10:3–9

Frischknecht R, Rebitzer G (2005) The ecoinvent database system: a comprehensive web-based LCA database. J Clean Prod 13:1337–1343

Fryar CD, Gu Q, Ogden CL (2012) Anthropometric reference data for children and adults: United States, 2007-2010. Vital Health Stat 11 1–48

Funtowicz SO, Ravetz JR (1990) Uncertainty and quality in science for policy. Springer Science & Business Media

Gavankar S, Anderson S, Keller AA (2014) Critical components of uncertainty communication in life cycle assessments of emerging technologies. J Ind Ecol

Gavankar S, Suh S (2014) Fusion of conflicting information for improving representativeness of data used in LCAs. Int J Life Cycle Assess 19:480–490. doi: 10.1007/s11367-013-0673-2

Geisler G, Hellweg S, Hungerbühler K (2005) Uncertainty analysis in life cycle assessment (LCA): Case study on plant-protection products and implications for decision making (9 pp+ 3 pp). Int J Life Cycle Assess 10:184–192

Gentle J (2013) Random number generation and Monte Carlo methods

Gregory JR, Noshadravan A, Olivetti EA, Kirchain RE (2016) A Methodology for Robust Comparative Life Cycle Assessments Incorporating Uncertainty. Environ Sci Technol 50:6397–6405. doi: 10.1021/acs.est.5b04969

Groen EA, Heijungs R, Bokkers EAM, de Boer IJM (2014) Methods for uncertainty propagation in life cycle assessment. Environ Model Softw 62:316–325. doi: 10.1016/j.envsoft.2014.10.006

Guinée JB (2002) Handbook on life cycle assessment operational guide to the ISO standards. Int J Life Cycle Assess 7:311–313

Hauschild MZ, Goedkoop M, Guinée J, et al (2013) Identifying best existing practice for characterization modeling in life cycle impact assessment. Int J Life Cycle Assess 18:683–697

Heijungs R (1996) Identification of key issues for further investigation in improving the reliability of life-cycle assessments. J Clean Prod 4:159–166. doi: 10.1016/S0959-6526(96)00042-X

Heijungs R, Frischknecht R (2004) Representing Statistical Distributions for Uncertain Parameters in LCA. Relationships between mathematical forms, their representation in EcoSpold, and their representation in CMLCA (7 pp). Int J Life Cycle Assess 10:248–254. doi: 10.1065/lca2004.09.177

Heijungs R, Frischknecht R (2005) Representing statistical distributions for uncertain parameters in lca. relationships between mathematical forms, their representation in ecospold, and their representation in cmlca (7 pp). Int J Life Cycle Assess 10:248–254

Heijungs R, Henriksson PJ, Guinée JB (2017) Pre-calculated LCI systems with uncertainties cannot be used in comparative LCA. Int J Life Cycle Assess 22:461–461

Heijungs R, Huijbregts M (2004) A review of approaches to treat uncertainty in LCA. Orlando Fla Elsevier

Heijungs R, Kleijn R (2001) Numerical approaches towards life cycle interpretation five examples. Int J Life Cycle Assess 6:141–148

Heijungs R, Lenzen M (2014) Error propagation methods for LCA—a comparison. Int J Life Cycle Assess 19:1445–1461

Heijungs R, Suh S (2002) The computational structure of life cycle assessment

Helton JC, Davis FJ (2002) Illustration of sampling-based methods for uncertainty and sensitivity analysis. Risk Anal 22:591–622

Henriksson PJ, Heijungs R, Dao HM, et al (2015a) Product carbon footprints and their uncertainties in comparative decision contexts. PloS One 10:e0121221

Henriksson PJ, Rico A, Zhang W, et al (2015b) Comparison of Asian aquaculture products by use of statistically supported life cycle assessment. Environ Sci Technol 49:14176–14183

Hertwich EG, Hammitt JK (2001) A decision-analytic framework for impact assessment part I: LCA and decision analysis. Int J Life Cycle Assess 6:5–12

Hickey AM, Davis AM (2003) Elicitation technique selection: how do experts do it? In: Requirements engineering conference, 2003. proceedings. 11th ieee international. IEEE, pp 169–178

Holland D, Fitz-Simons T (1982) Fitting statistical distributions to air quality data by the maximum likelihood method. Atmospheric Environ 1967

Hong J, Shaked S, Rosenbaum R, Jolliet O (2010) Analytical uncertainty propagation in life cycle inventory and impact assessment: application to an automobile front panel. Int J …

Huijbregts M (2002) Uncertainty and variability in environmental life-cycle assessment. Int J Life Cycle Assess 7:173–173

Huijbregts M a J, Gilijamse W, Ragas AMJ, Reijnders L (2003) Evaluating uncertainty in environmental life-cycle assessment. A case study comparing two insulation options for a Dutch one-family dwelling. Environ Sci Technol 37:2600–2608. doi: 10.1021/es020971+

Huijbregts MA (1998a) Part II: Dealing with parameter uncertainty and uncertainty due to choices in life cycle assessment. Int J Life Cycle Assess 3:343–351

Huijbregts MA (1998b) Application of uncertainty and variability in LCA. Int J Life Cycle Assess 3:273–280

Huijbregts MA, Norris G, Bretz R, et al (2001) Framework for modelling data uncertainty in life cycle inventories. Int J Life Cycle Assess 6:127–132

Hung M-L, Ma H (2009) Quantifying system uncertainty of life cycle assessment based on Monte Carlo simulation. Int J Life Cycle Assess 14:19–27

Imbeault-Tétreault H, Jolliet O, Deschênes L, Rosenbaum RK (2013) Analytical propagation of uncertainty in life cycle assessment using matrix formulation. J Ind Ecol 17:485–492

ISO I (2006) 14040: Environmental management–life cycle assessment–principles and framework. Lond Br Stand Inst

Jaworska JS, Bridges TS (2001) Uncertainty in Environmental Risk Assessment. In: Assessment and Management of Environmental Risks. Springer, Dordrecht, pp 203–207

Jenne CA, Cattell RK (1983) Structural change and energy efficiency in industry. Energy Econ 5:114–123

Jung J, von der Assen N, Bardow A (2013) Sensitivity coefficient-based uncertainty analysis for multi-functionality in LCA. Int J Life Cycle Assess 19:661–676. doi: 10.1007/s11367-013-0655-4

Kloepffer W (2008) Life cycle sustainability assessment of products. Int J Life Cycle Assess 13:89

Knol AB, Slottje P, van der Sluijs JP, Lebret E (2010) The use of expert elicitation in environmental health impact assessment: a seven step procedure. Environ Health 9:19

Kollmuss A, Agyeman J (2002) Mind the Gap: Why do people act environmentally and what are the barriers to pro-environmental behavior? Environ Educ Res 8:239–260. doi: 10.1080/13504620220145401

Laner D, Rechberger H, Astrup T (2014) Systematic evaluation of uncertainty in material flow analysis. J Ind Ecol 18:859–870

Lesage P, Mutel C, Schenker U, Margni M (2018) Uncertainty analysis in LCA using precalculated aggregated datasets. Int J Life Cycle Assess 1–18

Lewandowska A, Foltynowicz Z, Podlesny A (2004) Comparative lca of industrial objects part 1: lca data quality assurance—sensitivity analysis and pedigree matrix. Int J Life Cycle Assess 9:86–89

Limpert E, Stahel WA, Abbt M (2001) Log-normal Distributions across the Sciences: Keys and Clues On the charms of statistics, and how mechanical models resembling gambling machines offer a link to a handy way to characterize log-normal distributions, which can provide deeper insight into variability and probability—normal or log-normal: That is the question. BioScience 51:341–352

Lloyd SM, Ries R (2008) Characterizing, Propagating, and Analyzing Uncertainty in Life-Cycle Assessment: A Survey of Quantitative Approaches. J Ind Ecol 11:161–179. doi: 10.1162/jiec.2007.1136

Lloyd SM, Ries R (2007) and Analyzing Uncertainty in Life-Cycle Assessment A Survey of Quantitative Approaches. J Ind Ecol 11:161–181. doi: 10.1162/jiec.2007.1136

Malla S (2009) CO2 emissions from electricity generation in seven Asia-Pacific and North American countries: A decomposition analysis. Energy Policy 37:1–9. doi: 10.1016/j.enpol.2008.08.010

Marlay RC (1984) Trends in industrial use of energy. Science 226:1277–1283

Martin TG, Burgman MA, Fidler F, et al (2012) Eliciting expert knowledge in conservation science. Conserv Biol 26:29–38

Maurice B, Frischknecht R, Coelho-Schwirtz V, Hungerbühler K (2000) Uncertainty analysis in life cycle inventory. Application to the production of electricity with French coal power plants. J Clean Prod 8:95–108. doi: 10.1016/S0959-6526(99)00324-8

McBride MF, Burgman MA (2012) What is expert knowledge, how is such knowledge gathered, and how do we use it to address questions in landscape ecology? In: Expert knowledge and its application in landscape ecology. Springer, pp 11–38

McCleese DL, LaPuma PT (2002) Using monte carlo simulation in life cycle assessment for electric and internal combustion vehicles. Int J Life Cycle Assess 7:230–236. doi: 10.1007/BF02978878

Mendoza Beltran A, Prado V, Font Vivanco D, et al (2018) Quantified Uncertainties in Comparative Life Cycle Assessment: What Can Be Concluded? Environ Sci Technol 52:2152–2161

Meng Z, Wang H, Wang B (2018) Empirical Analysis of Carbon Emission Accounting and Influencing Factors of Energy Consumption in China. Int J Environ Res Public Health 15:2467

Meozzi PG, Iannucci C (2006) Facilitating the Development of Environmental Information into Knowledge: Government Agency Perspective to Improve Policy Decision-making. In: 4th International Conference on Politics and Information Systems, Technologies and Applications

Moore DA, Healy PJ (2008) The trouble with overconfidence. Psychol Rev 115:502

Moreno Ruiz E, Valsasina L, Fitzgerald D, et al (2017) Documentation of changes implemented in the ecoinvent database v3. 4. ecoinvent. Zür Switz

Morgan MG (2014) Use (and abuse) of expert elicitation in support of decision making for public policy. Proc Natl Acad Sci 111:7176–7184

Muller S, Lesage P, Ciroth A, et al (2014) The application of the pedigree approach to the distributions foreseen in ecoinvent v3. Int J Life Cycle Assess 1–11

Murphy AH, Daan H (1984) Impacts of feedback and experience on the quality of subjective probability forecasts. Comparison of results from the first and second years of the zierikzee experiment. Mon Weather Rev 112:413–423

Normalización OI de (2006) ISO 14044: Environmental Management, Life Cycle Assessment, Requirements and Guidelines. ISO

Noshadravan A, Wildnauer M, Gregory J, Kirchain R (2013) Comparative pavement life cycle assessment with parameter uncertainty. Transp Res Part Transp Environ 25:131–138

OpenLCA (2018) User Manual. GreenDelta, Germany

Owens JW (1997) Life-cycle assessment: Constraints on moving from inventory to impact assessment. J Ind Ecol 1:37–49

Owens JW (1996) LCA impact assessment categories. Int J Life Cycle Assess 1:151–158

Paul S, Bhattacharya RN (2004) CO2 emission from energy use in India: a decomposition analysis. Energy Policy 32:585–593

Peters GP (2007) Efficient algorithms for Life Cycle Assessment, Input-Output Analysis, and Monte-Carlo Analysis. Int J Life Cycle Assess 12:373–380. doi: 10.1065/lca2006.06.254

Prado-Lopez V, Seager TP, Chester M, et al (2014) Stochastic multi-attribute analysis (SMAA) as an interpretation method for comparative life-cycle assessment (LCA). Int J Life Cycle Assess 19:405–416

Qin Y, Cucurachi S, Suh S Perceived Uncertainties of Characterization in Life Cycle Impact Assessment. Submitted

Qin Y, Suh S (2017) What distribution function do life cycle inventories follow? Int J Life Cycle Assess 22:1138–1145. doi: 10.1007/s11367-016-1224-4

Qin Y, Suh S (2016) What distribution function do life cycle inventories follow? Int J Life Cycle Assess 1–8. doi: 10.1007/s11367-016-1224-4

Ragas AM, Huijbregts MA, Henning-de Jong I, Leuven RS (2009) Uncertainty in environmental risk assessment: implications for risk-based management of river basins. Integr Environ Assess Manag 5:27–37

Razali N, Wah Y (2011) Power comparisons of shapiro-wilk, kolmogorov-smirnov, lilliefors and anderson-darling tests. J Stat Model Anal

Reap J, Roman F, Duncan S, Bras B (2008) A survey of unresolved problems in life cycle assessment. Int J Life Cycle Assess 13:374–388. doi: 10.1007/s11367-008-0009-9

Ren Y, Shi L, Bardow A, et al Environmental implications of China's ban on post-consumer plastics import. Submitted

Ridout MS, Linkie M (2009) Estimating overlap of daily activity patterns from camera trap data. J Agric Biol Environ Stat 14:322–337. doi: 10.1198/jabes.2009.08038

Rosenbaum R, Pennington DW, Jolliet O (2004) An implemented approach for estimating uncertainties for toxicological impact characterisation. In: 2nd Biennial Meeting of the International Environmental Modelling and Software Society, iEMSs

Rosenbaum RK, Bachmann TM, Gold LS, et al (2008) USEtox—the UNEP-SETAC toxicity model: recommended characterisation factors for human toxicity and freshwater ecotoxicity in life cycle impact assessment. Int J Life Cycle Assess 13:532–546

Ross S, Evans D, Webber M (2002) How LCA studies deal with uncertainty. Int J Life Cycle Assess

Rowe G, Wright G (1999) The Delphi technique as a forecasting tool: issues and analysis. Int J Forecast 15:353–375

Rypdal K, Winiwarter W (2001) Uncertainties in greenhouse gas emission inventories—evaluation, comparability and implications. Environ Sci Policy 4:107–116

Scherer L, Pfister S (2016) Dealing with uncertainty in water scarcity footprints. Environ Res Lett 11:54008

Sills DL, Paramita V, Franke MJ, et al (2012) Quantitative uncertainty analysis of life cycle assessment for algal biofuel production. Environ Sci Technol 47:687–694

SimaPro (2016) User Manual. PRé Consultants, Netherlands

Singh A, Singh A, Engelhardt M (1997) The lognormal distribution in environmental applications

Slottje P, van der Sluijs JP, Knol AB (2008) Expert Elicitation: Methodological suggestions for its use in environmental health impact assessments. National Institute for Public Health and the Environment

Sonnemann GW, Schuhmacher M, Castells F (2003) Uncertainty assessment by a Monte Carlo simulation in a life cycle inventory of electricity produced by a waste incinerator. J Clean Prod 11:279–292

Stocker T (2014) Climate change 2013: the physical science basis: Working Group I contribution to the Fifth assessment report of the Intergovernmental Panel on Climate Change. Cambridge University Press

Stothers AJ (2010) On the complexity of matrix multiplication

Sugiyama H, Fukushima Y, Hirao M, et al (2005) Using Standard Statistics to Consider Uncertainty in Industry-Based Life Cycle Inventory Databases (7 pp). Int J Life Cycle Assess 10:399–405

Suh S, Heijungs R (2007) Power series expansion and structural analysis for life cycle assessment. Int J Life Cycle Assess 12:381

Suh S, Qin Y (2017) Pre-calculated LCIs with uncertainties revisited. Int J Life Cycle Assess 22:827–831

US Environmental Protection Agency (2005) Guidelines for carcinogen risk assessment. In: Risk Assessment Forum. US EPA Washington, Dc

van den Berg NW, Huppes G, Lindeijer EW, et al (1999) Quality Assessment for LCA. Leiden: CML, Leiden University.(CML Report 152)

Van Der Sluijs JP, Craye M, Funtowicz S, et al (2005) Combining quantitative and qualitative measures of uncertainty in model-based environmental assessment: the NUSAP system. Risk Anal 25:481–492

Verbeeck G, Hens H (2010) Life cycle inventory of buildings: a calculation method. Build Environ 45:1037–1041

Vinodh S, Rathod G (2014) Application of life cycle assessment and Monte Carlo simulation for enabling sustainable product design. J Eng Des Technol 12:307–315

von Pfingsten S, Broll DO, von der Assen N, Bardow A (2017) Second-Order Analytical Uncertainty Analysis in Life Cycle Assessment. Environ Sci Technol 51:13199–13204

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

Weidema BP, Bauer C, Hischier R, et al (2013) Overview and methodology: Data quality guideline for the ecoinvent database version 3. Swiss Centre for Life Cycle Inventories

Weidema BP, Wesnaes MS (1996) Data quality management for life cycle inventories—an example of using data quality indicators. J Clean Prod 4:167–174

Wernet G, Bauer C, Steubing B, et al (2016) The ecoinvent database version 3 (part I): overview and methodology. Int J Life Cycle Assess 21:1218–1230. doi: 10.1007/s11367-016-1087-8

Williams VV (2012) Multiplying matrices faster than Coppersmith-Winograd. In: Proceedings of the forty-fourth annual ACM symposium on Theory of computing. ACM, pp 887–898

Winkler RL, Murphy AH (1968) “Good” probability assessors. J Appl Meteorol 7:751–758

Wu K, Soci C, Shum PP, Zheludev NI (2014) Computing matrix inversion with optical networks. Opt Express 22:295–304

Yang Y, Tao M, Suh S (2018) Geographic variability of agriculture requires sector-specific uncertainty characterization. Int J Life Cycle Assess 23:1581–1589

Yazici B, Yolacan S (2007) A comparison of various tests of normality. J Stat Comput …

Zhang M, Mu H, Ning Y, Song Y (2009) Decomposition of energy-related CO2 emission over 1991–2006 in China. Ecol Econ 68:2122–2128

[This page left blank on purpose]