

# UC Santa Barbara

## UC Santa Barbara Previously Published Works

### Title

What distribution function do life cycle inventories follow?

### Permalink

<https://escholarship.org/uc/item/98k8d81j>

### Journal

The International Journal of Life Cycle Assessment, 22(7)

### ISSN

0948-3349

### Authors

Qin, Yuwei  
Suh, Sangwon

### Publication Date

2017-07-01

### DOI

10.1007/s11367-016-1224-4

### Supplemental Material

<https://escholarship.org/uc/item/98k8d81j#supplemental>

Peer reviewed

## What Distribution Function Do LCIs Follow?

### Abstract

**Purpose.** LCI results are often assumed to follow a log-normal distribution, while a systematic study that identifies the distribution function that best describes LCIs has been lacking. This paper aims to find the distribution function that best describes LCIs using ecoinvent version 3.1 database using a statistical approach, called overlapping coefficient analysis.

**Methods.** Monte Carlo simulation is applied to characterize the distribution of aggregate LCIs. 1,000 times of simulated LCI results are generated based on the unit-process level parametric uncertainty information, from each of which 1,000 randomly chosen data points are extracted. The 1 million data points extracted undergo statistical analyses including Shapiro-Wilk normality test and the overlapping coefficient analysis. The overlapping coefficient is a measure used to determine the shared area between the distribution of the simulated LCI results and three possible distribution functions that can potentially be used to describe them including log-normal, gamma, and Weibull distributions.

**Results and discussion.** Shapiro-Wilk normality test for 1,000 samples shows that average p-value of log-transformed LCI results is 0.18 at 95% confidence level, accepting the null hypothesis that LCI results are lognormally distributed. The overlapping coefficient analysis shows that lognormal distribution best describes the distribution of LCI results. The average of overlapping coefficient (OVL) for lognormal distribution is 95%, while that for gamma and Weibull distributions are 92% and 86%, respectively.

**Conclusions.** This study represents the first attempt to calculate the stochastic distributions of the aggregate LCIs covering the entire ecoinvent 3.1 database. This study empirically shows that LCIs of ecoinvent 3.1 database indeed follow a log-normal distribution. This finding can facilitate more efficient storage and use of uncertainty information in LCIs, and can reduce the demand for computational power to run Monte Carlo simulation, which currently relies on unit process-level uncertainty data.

**Keywords:** Life cycle inventory; uncertainty analysis; Monte Carlo simulation; probability distribution; lognormal distribution; Ecoinvent 3.1

## 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). Traditionally, LCA studies only include deterministic values in results. However, a sound decision-making can benefit from the understandings of the stochastic distribution of LCA results (Geisler et al. 2004; Sugiyama et al. 2005). 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). 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).

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) illustrates 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 the typologies of uncertainty and the approaches to treat 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). Heijungs and Huijbregts (2004) presented a review of four general uncertainty treatments for stochastic uncertainty and Citroth and colleagues (2004) proposed a method for uncertainty calculation. Two types of techniques have emerged: sampling method and analytical approach (Ross et al. 2002; Heijungs and Frischknecht 2004; Clavreul and Guyonnet 2013; Jung et al. 2013). 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).

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. (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 distribution 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 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 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; Citroth 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. Method and Data

### 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. MCS is a common sampling technique used in uncertainty assessment to obtain randomly generated numbers (Lloyd and Ries 2008). With the help of advancement in computer hardware and software, MCS of large datasets, such as the ecoinvent v3.1, became viable (Gentle 2013). Our approach to MCS takes several 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. Fig. 1 demonstrates the procedure for the statistical analysis used in this study.

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,  $M = B A^{-1}$  (Heijungs and Suh 2002).

This process can be summarized as in Eq.1:

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

$\delta B_i$ : randomly sampled deviation matrix for the elementary flows

$B$ : deterministic elementary flow matrix

$\delta A_i$ : randomly sampled deviation matrix for the intermediate flows

$A$ : deterministic intermediate flow matrix

$i$ : number of simulation,  $i = 1, \dots, n$  ( $n = 1,000$ )

The resulting  $M$  matrix has the dimension of 1,869 (elementary flows)  $\times$  11,332 (processes), and we have generated 1,000 of them,  $\{M_1^{\delta}, M_2^{\delta}, \dots, M_{1000}^{\delta}\}$ . To ensure efficiency, we further sampled 1,000 data points from each  $M_i^{\delta}$ . 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  $\approx$  17 hours.

### 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 lognormal distribution (Table 1). Normal and triangular distributions are also considered as the input parameter distributions, though they are less common as compared to 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 SI.

### 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),  $\Delta$  is the OVL that represents the common area under both density curves. If the two density functions are  $f(x)$  and  $g(x)$ , then

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

(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).

### 2.4 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 non-invertible. 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).

### 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. Fig. 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 Fig. 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.

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.

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 pass 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. Fig. 3 represents 9 typical comparisons among the results and the estimates of lognormal, gamma and Weibull distributions of random elementary flow-process pairs.

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. Fig. 4 illustrates the distributions of OVL results from the LCI results versus lognormal, gamma and Weibull distributions. For example, the solid line in Fig. 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.

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.

## **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 1/4000<sup>th</sup> 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.

### **Acknowledgement**

We are thankful to Guillaume Bourgault, Ecoinvent Project Manager, for providing data and assisting us to better understanding the database. His help and support is invaluable to the research. We would like to acknowledge US Environmental Protection Agency (EPA) for its funding of the publication (8-442650-22870-2 RFEP70).

## Supplementary information

The results of lognormal distributions of 1,000 LCIs are presented in the supporting material. Also 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.

## Reference

- 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.
- Björklund AE (2002) Survey of approaches to improve reliability in LCA. *Int J Life Cycle Assess* 7:64–72.
- 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.
- 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.
- Ciroth A, Srocka M (2008) How to obtain a precise and representative estimate for parameters in LCA. *Int J Life Cycle Assess* 13:265–277.
- Clavreul J, Guyonnet D (2013) Stochastic and epistemic uncertainty propagation in LCA.
- 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.
- Cucurachi S, Heijungs R (2014) Characterisation factors for life cycle impact assessment of sound emissions. *Sci Total Environ* 468:280–291.
- Dennis B, Patil GP (1984) The gamma distribution and weighted multimodal gamma distributions as models of population abundance. *Math Biosci* 68:187–212.
- Fava J (1994) Life-cycle Assessment Data Quality: A Conceptual Framework: Workshop Report.
- Finnveden G, Hauschild MZ, Ekvall T, et al (2009) Recent developments in life cycle assessment. *J Environ Manage* 91:1–21.
- Frischknecht R, Jungbluth N, Althaus HJ, et al (2004) Overview and Methodology. Final report ecoinvent 2000 No. 1. Swiss Centre for Life Cycle Inventories, Dübendorf, CH.

- 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 (2004) 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. doi: 10.1065/lca2004.09.178
- Gentle J (2013) Random number generation and Monte Carlo methods. Springer Science & Business Media.
- 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, Huijbregts M (2004) A review of approaches to treat uncertainty in LCA.
- 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. Springer Science & Business Media
- 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.
- Hofstetter P (1998) Perspectives in Life Cycle Impact Assessment: A Structured Approach to Combine Models of the Technosphere, Ecosphere and Valuesphere. Kluwer Academic Publishers
- Holland DM, Fitz-Simons T (1982) Fitting statistical distributions to air quality data by the maximum likelihood method. *Atmospheric Environ* 1967 16:1071–1076.
- Hong J, Shaked S, Rosenbaum RK, Jolliet O (2010) Analytical uncertainty propagation in life cycle inventory and impact assessment: application to an automobile front panel. *Int J Life Cycle Assess* 15:499–510.
- Huijbregts MA, Gilijamse W, Ragas AM, 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.
- 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-Tétréault 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.
- 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
- Laner D, Rechberger H, Astrup T (2014) Systematic evaluation of uncertainty in material flow analysis. *J Ind Ecol* 18:859–870.

- 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
- Maurice B, Frischknecht R, Coelho-Schwartz 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
- 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
- Muller S, Lesage P, Cirotto A, et al (2014) The application of the pedigree approach to the distributions foreseen inecoinvent v3. *Int J Life Cycle Assess* 1–11.
- 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.
- Razali NM, Wah YB (2011) Power comparisons of shapiro-wilk, kolmogorov-smirnov, lilliefors and anderson-darling tests. *J Stat Model Anal* 2:21–33.
- 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.
- Ross S, Evans D, Webber M (2002) How LCA studies deal with uncertainty. *Int J Life Cycle Assess* 7:47–52.
- Singh A, Singh A, Engelhardt M (1997) The lognormal distribution in environmental applications. Technology Support Center Issue Paper.
- 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. doi: 10.1016/S0959-6526(02)00028-8
- 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.
- Weidema BP, Bauer C, Hischer R, et al (2013) Overview and methodology: Data quality guideline for theecoinvent 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.
- Yazici B, Yolacan S (2007) A comparison of various tests of normality. *J Stat Comput Simul* 77:175–183.

## Tables and figures

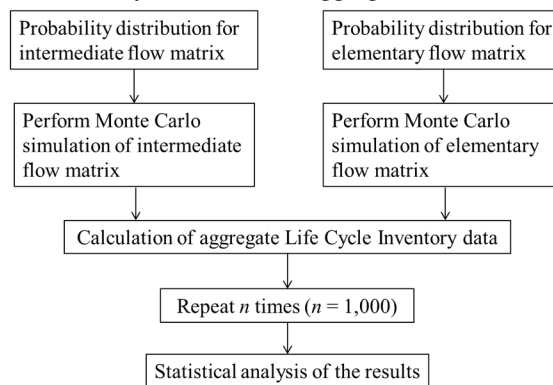
**Table 1** Summary of probability distribution in ecoinvent v3.1 unit process data

	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%

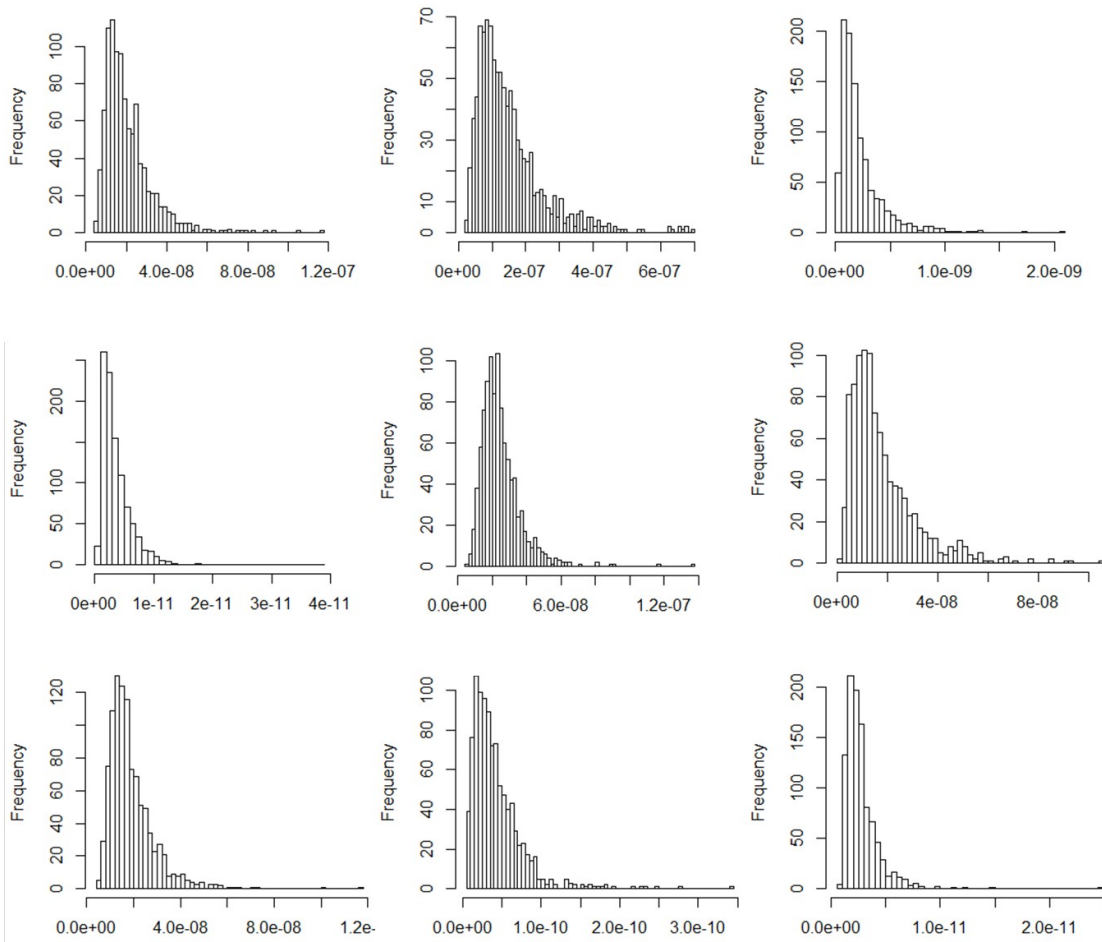
**Table 2** Shapiro-Wilk Normality Test results of simulated LCIs ( $p$ -value)

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%

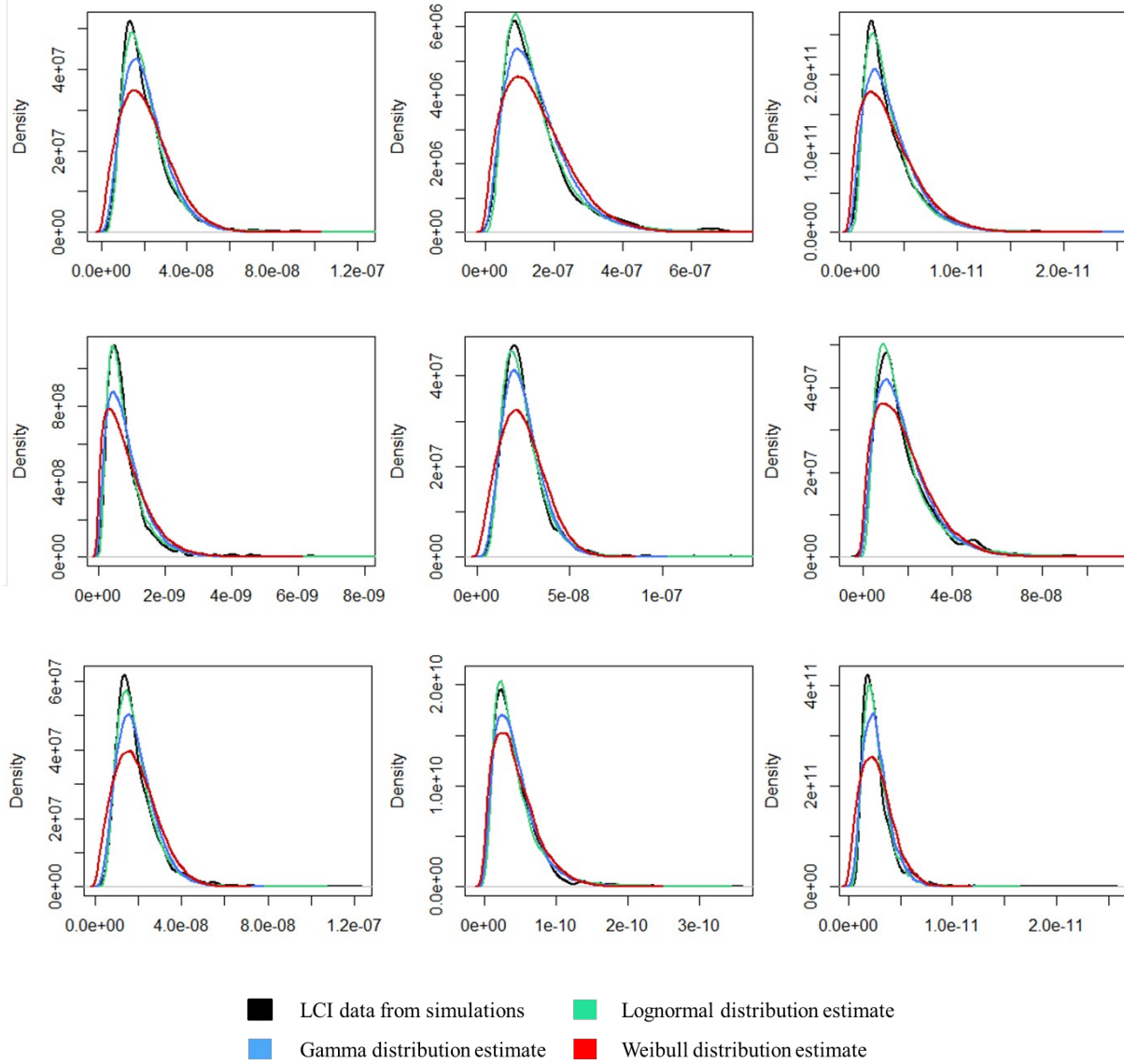
**Fig. 1** Monte Carlo procedure for uncertainty assessment of aggregate LCI



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



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



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

