Resources, Conservation & Recycling 137 (2018) 316–325

Contents lists available at ScienceDirect

Resources, Conservation & Recycling

journal homepage: www.elsevier.com/locate/resconrec

Full length article

Representing and visualizing data uncertainty in input-output life cycle assessment models

Xiaoju Chen[a,*], W. Michael Griffin[b], H. Scott Matthews[c]

[a] Department of Civil and Environmental Engineering, Carnegie Mellon University, 5000 Forbes Ave., Pittsburgh, PA, 15213, USA
[b] Department of Engineering and Public Policy, Carnegie Mellon University, 5000 Forbes Ave., Pittsburgh, PA, 15213, USA
[c] Department of Civil and Environmental Engineering and Department of Engineering and Public Policy, Carnegie Mellon University, 5000 Forbes Ave., Pittsburgh, PA, 15213, USA

ARTICLE INFO

Keywords:
Life cycle assessment
Uncertainty
Range method
Input-output model
Visualization

ABSTRACT

Uncertainty assessment is crucial to the reliability of decisions made based on results of life cycle assessment (LCA) models. However, the most popular uncertainty analysis method in LCA models defines uncertainty as quantified empirical judgment of the sources of the inventory data, which lacks the deviations due to the measurements of the data. We create a modified range method that leverages an array of publicly available data to represent the uncertainty of the actual values used in input-output models as ranges. Using these ranges, we propagate and visualize results from the uncertainties. We demonstrate the utility of the method using the Economic Input-Output Life Cycle Assessment (EIO-LCA) model and focus on the uncertainty of estimated energy consumption. Results show that for energy consumption values in the model, average uncertainty ranges are within ± 40% with some outliers. Our method screens based on the magnitude of impacts and the relative uncertainty. Improved uncertainty assessment supports various types of decisions, such as product comparisons, hotspot analysis, and overall energy analyses. We used three case studies to demonstrate the implementations of our method. This method can be extended to additional types of flows, beyond energy, and to process-matrix-based LCA models.

1. Introduction

Life Cycle Assessment (LCA) is a method that computes and evaluates inputs, outputs and environmental impacts from design to disposition of a product or technology (Guinee, 2002; ISO, 2006). Three primary types of methodologies have been used to conduct LCA studies: process-based LCA, which analyzes the impacts at the process level; economic input-output LCA (IO-LCA), which uses economic exchange values to trace the total impacts through the supply chain; and hybrid LCA, which combines process-based and input-output-based LCA (Finnveden et al., 2009; Hendrickson et al., 1998; Hertwich, 2005; Pairotti et al., 2015). Compared with process-based models, IO-LCA models have more complete boundaries (Islam et al., 2016), and avoid truncation bias (Majeau-Bettez et al., 2011). An IO-LCA model evaluates the environmental impacts of industrial sectors based on all economic exchange activities within a defined region. The total economic exchanges (e.g., in dollars) of all industries can be included by using the Leontief inverse ($(\mathbf{I}-\mathbf{A})^{-1}$, see Eq. (1)), which was first introduced by Leontief, for estimating direct and indirect (total) transactions (Hendrickson et al., 1998; Leontief, 1970). Matrix $\mathbf{R}$ is the environmental matrix; it is used to convert economic values to environmental flows. Environmental flows can be resource inputs such as energy consumption and water usage, or environmental pollutants such as greenhouse gases and hazardous wastes (CMUGDI, 2008). Each value in the $\mathbf{R}$ matrix, represents the direct environmental flow to produce one dollar output in the industry and is denoted by flow/$. To calculate each value, IO-LCA model developers first use a publicly available data source to obtain an estimate of the industry's annual direct environmental flow, which is then divided by the annual economic output of the industry. When users use an IO-LCA model for LCA studies, they define a vector $\mathbf{y}$ to represent the final demand of the target sector, then the model uses Eq. (1) to calculate the environmental flows from all sectors defined in the system (vector $\mathbf{B}$) (Matthews et al., 2014).

$$\mathbf{B} = \mathbf{R}(\mathbf{I}-\mathbf{A})^{-1}\mathbf{y} \tag{1}$$

IO-LCA models provide straightforward results with well-scaled inventory data. Therefore, users often use IO-LCA models as a screening tool to identify the hotspots of environmental flows in the supply chain. However, current IO-LCA models suffer from a lack of information about uncertainty, which limits the users' ability to interpret the results

* Corresponding author.
E-mail addresses: xiaojuc@andrew.cmu.edu (X. Chen), wmickaelgriffin@cmu.edu (W.M. Griffin), hsm@cmu.edu (H.S. Matthews).

https://doi.org/10.1016/j.resconrec.2018.06.011
Received 28 March 2018; Received in revised form 9 June 2018; Accepted 10 June 2018
0921-3449/

given by the model.

In both process-based and IO-based LCA studies, representative data are usually hard to obtain, which translates into data uncertainty. LCA researchers have long been trying to account for data uncertainty to provide more robust conclusions and policy-related decisions (Huijbregts and Huijbregts, 1998; Williams et al., 2009; Mullins et al., 2011; Venkatesh et al., 2011). Most frequently, LCA researchers use the pedigree approach (Huijbregts et al., 2001) to estimate data uncertainty. The pedigree approach, developed by Weidema and Wesnæs (1996), uses information related to the quality of data attributes to transform inventory data to a default lognormal distribution, which represents the uncertainty of the data. In the pedigree matrix, the quality of the data is described by six attributes: reliability, completeness, temporal correlation, geographical correlation, technological correlation, and sample size. The quality levels of these attributes are represented by indicator scores that vary between one (best quality) and five (worst quality). For example, for a U.S. life cycle study, an electricity consumption value provided by US Department of Energy three years prior to the study will be assigned score 1 for reliability, score 2 for temporal correlation, and score 1 for geographical correlation. To estimate uncertainty, these data indicator scores are converted to two factors; these factors are used as the parameters in the default lognormal distribution (Guo and Murphy, 2012). One disadvantage of the pedigree approach is that when multiple inventory data points are from data sources with similar quality, it is difficult to distinguish the variation in uncertainties between these data points. Given this limitation, the pedigree approach cannot effectively estimate the uncertainty in IO-LCA models, in which data are often from the same data source or with the same quality. Using the pedigree approach to represent uncertainty in IO-LCA models will result in one single uncertainty score for almost all data points, which is not representative of real life. Other methods, such as using data from multiple sources, have also been used to represent data uncertainty in LCA studies (Bawden et al., 2016; Deng et al., 2011; Huijbregts et al., 2001; Noshadravan et al., 2013). Compared to the pedigree approach, applying data from multiple data sources can represent uncertainty by including a variety of potential values.

In LCA studies, uncertainty in the data is commonly propagated using methods such as stochastic modeling, analytical modeling, and fuzzy theory (Heijungs and Lenzen, 2014; Groen et al., 2014). In stochastic modeling, simulation is used to propagate uncertainty represented by the distributions of the inventory data (Huijbregts et al., 2001); this approach is commonly used in both regional and multi-regional IO-LCA models (Lenzen et al., 2010; Moran and Wood, 2014; Temurshoev, 2015). In analytical methods, mathematical models are used to calculate uncertainty, which is represented by the inventory data's statistical parameters (Heijungs, 1996). The fuzzy theory method uses fuzzy logic to estimate uncertainty in the degree of plausibility of data (Tan, 2008). Range methods use a bound restricted by maximum and minimum values to represent uncertainty in the inventory data, and are commonly used to estimate uncertainty in LCA data and results. The maximum and minimum values of different inputs are treated as having equal likelihoods, and after calculation, the values are derived from all possible outcomes and used to present uncertainty in the LCA results (Chevalier and Le Téno, 1996; Wu and Chang, 2003). Range methods have two major advantages when applied to estimate the uncertainty in IO-LCA models. One advantage is the method's ability to include multiple data sources for one data point, which enables data collectors to consider different possible data inputs without defining another inventory. The other advantage is that the method's straightforward calculations limit the time and efforts of calculating uncertainty results with IO-LCA models' complex matrix structures.

To effectively include uncertainty information in IO-LCA models, we developed a modified range method to estimate uncertainty in LCA data and results. In contrast to the traditional range method that uses only maximum and minimum values, we used all available data points to create a range to represent uncertainty. All data points were derived from publicly available data sources such as government agencies, published research articles, and reports. We used various assumptions to convert the raw data to usable values for the model. We propagated the uncertainty represented by these values with our modified range method: the values were treated as having equal likelihood and used to estimate changes in results from the IO-LCA model's current values. The uncertainty results were presented using our visualization method, which displayed all possible results within the range. This method allows users to track the source(s) of uncertainty and use this information as a reference to focus on the important parts of their inventories or models.

We only focus on uncertainties in the environmental matrix (the **R** matrix) for two reasons. First, previous studies suggest that the uncertainty in the economic matrix (the **A** matrix) is relatively small (Bullard and Sebald, 1988; Karstensen et al., 2015). When working on the U.S. input-output table, Bullard and Sebald conclude that the uncertainty of more than 90% of the values in the Leontief matrix was less than 2% with a 95% confidence (Bullard and Sebald, 1988). Karstensen et al. (2015) found that at the national level, the uncertainty in the **A** matrix had less impact on the overall output than the uncertainty in the **R** matrix. At multi-regional levels, studies also confirmed the theory developed by Jaynes (1957) that the stochastic errors in an IO table can be canceled out and result in small uncertainty values in the Leontief inverse (Lenzen et al., 2010; Moran and Wood, 2014). Second, there is relatively little information related to the derivation of parameters of the economic matrix (the **A** matrix), making uncertainty assessment difficult. The **A** matrices in input-output models are the products of enormous data collection and preparation efforts by the creators, which are usually government economic agencies. Therefore, typically only a single data source for an **A** matrix for a particular economy exists, yielding single data points without uncertainty (deterministic values). This lack of information makes the range method inapplicable for current **A** matrices. When more information is available, such as the uncertainty of the **A** matrix values, this factor could be incorporated into the final results.

We use data from the U.S. 2002 Economic Input-Output Life Cycle Assessment (EIO-LCA) model to implement and demonstrate the utility of the uncertainty estimation method. We chose to use the 2002 model because in this model, the data in the **A** matrix and the **R** matrix were both from the same year; it is the only recent US IO-LCA model that has temporally matched matrices. We chose five energy resource inputs (coal, natural gas, petroleum, biomass, and non-fossil electricity) due to the high degree of data availability from multiple data sources, and because energy resource consumption is a fundamental component to estimate other environmental flows, such as emissions.

The following sections describe the methods used in developing these uncertainty-based estimations in the 2002 EIO-LCA model with respect to energy consumption. Detailed information about sectors in the EIO-LCA model are provided in Appendix D in Supplementary material.

2. Material and methods

2.1. Values in the R matrix

In a traditional IO-LCA model, energy consumption data in an **R** matrix are the deterministic values of each industry sector. Two steps are involved in estimating these deterministic values. First, an annual energy consumption value (in Joules) for each industry sector is estimated from a publicly available data source (such as government survey reports) with certain assumptions (such as assumptions on fuel prices), if necessary. Second, each energy value is divided by the industry's total annual economic output (in dollars). In this study, the first step was modified; deterministic energy consumption values were replaced by energy consumption ranges. For each sector, multiple energy