The International Journal of Life Cycle Assessment (2020) 25:2290–2308
https://doi.org/10.1007/s11367-020-01826-5

UNCERTAINTIES IN LCA

Designing hybrid life cycle assessment models based on uncertainty and complexity

Tapajyoti Ghosh[1] · **Bhavik R. Bakshi**[1]

Received: 25 March 2019 / Accepted: 27 September 2020 / Published online: 21 October 2020

Abstract

Purpose Despite the wide use of LCA for environmental profiling, the approach for determining the system boundary within LCA models continues to be subjective and lacking in mathematical rigor. As a result, life cycle models are often developed in an ad hoc manner, and are difficult to compare. Significant environmental impacts may be inadvertently left out. Overcoming this shortcoming can help elicit greater confidence in life cycle models and their use for decision making.

Methods This paper describes a framework for hybrid life cycle model generation by selecting activities based on their importance, parametric uncertainty, and contribution to network complexity. The importance of activities is determined by structural path analysis—which then guides the construction of life cycle models based on uncertainty and complexity indicators. Information about uncertainty is from the available life cycle inventory; complexity is quantified by cost or granularity. The life cycle model is developed in a hierarchical manner by adding the most important activities until error requirements are satisfied or network complexity exceeds user-specified constraints.

Results and Discussion The framework is applied to an illustrative example for building a hybrid LCA model. Since this is a constructed example, the results can be compared with the actual impact, to validate the approach. This application demonstrates how the algorithm sequentially develops a life cycle model of acceptable uncertainty and network complexity. Challenges in applying this framework to practical problems are discussed.

Conclusion The presented algorithm designs system boundaries between scales of hybrid LCA models, includes or omits activities from the system based on path analysis of environmental impact contribution at upstream network nodes, and provides model quality indicators that permit comparison between different LCA models.

Keywords Life cycle analysis · Uncertainty · System boundary · Network analysis

1 Introduction

Life cycle assessment has seen tremendous development over the last few decades. It is used for identifying environmental impacts of a wide spectrum of products, manufacturing activities, and services. This method considers activities from "cradle to grave" so that environmental impacts of a larger system may be assessed, which includes not only the primary activity itself but also upstream and downstream activities that are connected to it. The life cycle impact information should reduce the chance of shifting of impacts to other parts of the life cycle. LCA has also been used to incorporate elements of sustainability in engineering, including design of manufacturing processes, supply chain design (Ghosh and Bakshi 2019), and economic policy analysis (Searchinger et al. 2008; Samaras and Meisterling 2008) and environmental policy decisions across the world.

Despite the popularity and extensive application of LCA in different fields, drawbacks have plagued this method, thus reducing confidence in its results (Reap et al. 2008; Finnveden 2000). Any typical process incorporates multiple inputs which in turn have many other inputs for their

Communicated by: Andreas Ciroth

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s11367-020-01826-5) contains supplementary material, which is available to authorized users.

✉ Bhavik R. Bakshi
bakshi.2@osu.edu

[1] William G. Lowrie Department of Chemical and Biomolecular Engineering, The Ohio State University, 151 W. Woodruff Avenue, Columbus, OH 43210, USA

own production. Tracing these upstream activities until the fundamental raw materials presents considerable difficulty (Joshi 1999). System boundary selection is another major prevalent problem in LCA (Reap et al. 2008). Since the upstream life cycle of any product or manufacturing process is essentially infinite, modeling it requires differentiating between significant processes that are included within the system boundary and insignificant or background processes that are neglected. These decisions are subjective and difficult since it is not known in advance if the information is insignificant enough to be neglected (Finnveden et al. 2009). Selection of a highly truncated boundary can result in omission of a large part of life cycle emissions (Lenzen 2000), incorrect reflection of reality, and lower confidence in results.

1.1 Boundary selection problem

Even though ISO 14040 and ISO 14044 provide guidelines for determining the boundary, they are still quite subjective (Suh et al. 2004). Reap et al. (2008) explain that the cutoff criterion is difficult to implement since it requires "perfect, holistic knowledge of all the possible effects a decision might have on the product system and consequently on the impacts of interest." The cutoff is expressed as a percentage of the total life cycle emission to establish an acceptable truncation error. However, if the complete life cycle data has already been obtained, introducing a cutoff criterion becomes redundant. Several approaches have been proposed to solve the boundary selection problem in LCA, of which economic input–output-based environmental analysis has been popular (Hendrickson et al. 2006; Sharrard et al. 2008; Moriguchi et al. 1993). It involves using economic models proposed by Leontief to perform environmental analysis, known as economic input–output (EIO) models. It considers the complete upstream life cycle within a nation's or region's economy, thus providing a partial solution to the boundary selection problem. While explaining in detail the basic structure of input–output life cycle assessment(IOLCA), Suh and Huppes (2005) discuss its methodological limitations, especially errors arising due to aggregation of industries and commodities. Also, due to economic data generally being older than process LCA data and due to lack of information, it is difficult to study novel and developing technologies. Process-based and EIO-based LCA models were combined in hybrid LCA models that build on the strengths of both approaches while addressing each other's inadequacies (Haes et al. 2004; Joshi 1999). In hybrid analysis, the decision of modeling activities in the process LCA or the EIO LCA scale depends on several factors. Sometimes, foreground processes are modeled using process LCA inventories while the background processes are modeled using EIOLCA. Data availability also affects such decisions as well as practitioner's choice of the *importance* of an activity. This results in consideration of an expanded system boundary compared with conventional process LCA while incorporating more accurate information compared with just IOLCA (Treloar et al. 2004; Stokes and Horvath 2006).

1.2 Boundary delineation problem in hybrid LCA models

System boundaries are difficult to define for hybrid LCA models as seen from the blurred edges of the model in Fig. 1. Due to the presence of EIO models, the mathematical functions for performing LCA through the Leontief inverse equations make the system boundary "infinite." However, the downside is that data quality gets severely degraded due to high aggregation. For the sake of this article, if we temporarily consider that hybrid models do actually expand the system boundary and enable inclusion of far upstream processes, a different problem comes up. As pointed out by Suh et al. (2004), hybrid approaches introduce a new problem: the delineation of boundaries between the process system and the input output system as seen from the blurred edges of boundary between the process and economy scales of the LCA model in Fig. 1. They conjecture that a multiscale hybrid LCA model with few activities for the process LCA scale will have significant differences from another with many activities modeled in it. Along with that, adding more activities to the process part results in higher resolution of the hybrid model while increasing its data requirements (Graedel and Graedel 1998). This leads to the development of a model design problem with conflicting objectives—increasing resolution while decreasing data requirements. Looking into a wide range of hybrid LCA studies (Treloar et al. 2001; Treloar et al. 2004; Stokes and Horvath 2006; Wiedmann et al. 2011; Pairotti et al. 2015; Hou et al. 2014; Bilec et al. 2006; Rowley et al. 2009), it is seen that the decision of distributing life cycle activities between the scales of hybrid LCA model is subjective. While most studies decide the boundary based on experience, data availability, cost, and time requirements, Treloar et al. (2001) and Treloar et al. (2004) used structural path analysis (SPA) to extract the hotspots of environmental impacts in the life cycle of a product in terms of contribution to the total emission. However, it is difficult to judge based on the emission quantities if a certain activity is significant enough to be modeled at a particular scale. The path exchange method (Lenzen and Crawford 2009) is another approach to combine process LCA data with EIO level data. Even though this method claims lower data requirements and labor costs, it misses out on providing a quantitative measure of the "goodness" or correctness of models generated. Hondo and Sakai (2002) use sensitivity

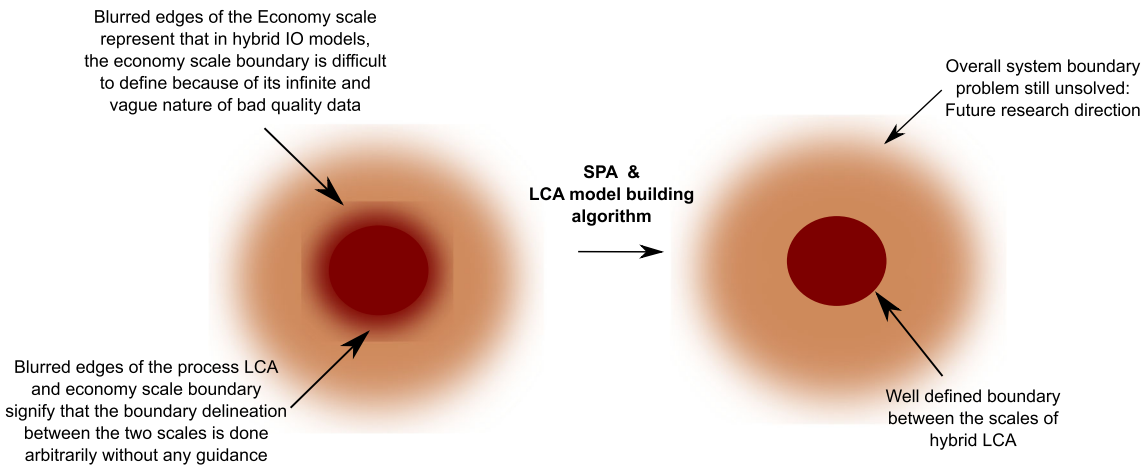

Fig. 1 The system boundary selection problem and system boundary delineation problem in hybrid LCA model. The hazy part signifies that it is difficult to determine the boundaries. The algorithm tries to solve the boundary delineation problem between scales

analysis to rank the economic flows in EIO analysis for delineating the interface between scales. This work does not explore the correctness of the models built. Suh et al. (2004) propose a stochastic framework to determine the boundary based on uncertainty analysis using the Monte Carlo method. They conclude by mentioning a general idea that replaces aggregated data in the EIO part by more accurate process LCA data continuously, until a required level of uncertainty is achieved. The data to be shifted between the scales would be for the "most important lower order." However, the decision making procedure for shifting scales and the uncertainty calculation process are not explained. Neither this method nor its variation seem to have been developed or presented in any form. Recently, Stephan et al. (2018) published the path exchange method for building hybrid life cycle models. It compares process LCA and EIO models and combines them to create a hybrid system. The searching and building processes are automated. However, it does not extend to determination of "goodness" of these hybrid models and how different models can be compared with each other. This work does not address the boundary delineation and selection problems.

From this literature review, it is quite evident that the boundary selection as well as delineation of boundaries between different scales in multi-scale LCA models has an impact on the validity of the final results and there is a compelling need to address this problem through a mathematically rigorous method directed toward removing subjectivity in these decisions.

1.3 Uncertainty in LCA

Lloyd and Ries (2007) provide a broad review of LCA uncertainty studies by classifying them into three categories—parameter, model, and scenario with parameter uncertainty being the most studied (Heijungs and Huijbregts 2004). It includes data uncertainties in system inputs, technology coefficients, and environmental impact factors. Among the popular uncertainty modeling approaches are stochastic modeling, scenario modeling, fuzzy datasets (Benetto et al. 2008), interval calculations (Chevalier and Téno 1996), and analytical uncertainty propagation. While many studies have used Monte Carlo techniques, they have been deemed to be computationally intensive when comparing multiple scenarios (Hong et al. 2010; Ross et al. 2002). Analytical error propagation (Morgan et al. 1992) based on Taylor series expansion is used to estimate the moments of a distribution. The model output variance is described by a function of variances of each uncertain input variable. This method has been applied in the context of LCA (Ciroth 2002; Heijungs et al. 2005; Hong et al. 2010) and various studies have compared it with the more prevalent Monte Carlo sampling methods for uncertainty analysis (Groen et al. 2014; Heijungs et al. 2005; Imbeault-Tétreault et al. 2013; Heijungs and Huijbregts 2004). The main advantages of this method are (1) it requires less information as the probability distribution of the input variables is not required (Groen et al. 2014; Heijungs 2010); (2) it is less computationally demanding than Monte Carlo sampling methods (Heijungs 2010; Heijungs and Lenzen 2014).

1.4 Contributions and vision

LCA practitioners face decisions such as how many upstream activities should be included within the system boundary. Evaluating the "best" model is yet another challenge that needs to be addressed by defining some quality criterion. The life cycle of any process is immense due to

interconnected flow linkages with other processes. Carving out a smaller system for LCA studies from this huge network requires information about the total model itself which is difficult to obtain. Considering such a problem for multi-scale frameworks, the difficulty is compounded by the presence of two or more scales. There will be certain combinations where a large number of activities needs to be modeled at a detailed scale, thus tremendously increasing the resolution, complexity, computational tractability, and data requirements of the model. This posits the need for an approach that can simultaneously bypass these challenges and help in the search for the best possible model. Solving such a problem is challenging due to high computational and data requirements.

This paper develops a framework for life cycle model construction that determines model quality based on uncertainty analysis using analytical error propagation. This approach addresses the boundary delineation problem as is commonly encountered in the development of hybrid LCA models as described in Section 1.2. As seen from Fig. 1, the boundary between EIO scale and process-based LCA scale in current hybrid LCA modeling is vague and determined arbitrarily. The goal of this article is to address the problem such that after the application of this algorithm, the boundary is determined using a mathematically rigorous procedure. The overall system boundary selection problem as described in Section 1.1 is not addressed in this study and is the subject of future work. Heijungs and Lenzen (2014) used the error propagation method to perform uncertainty analysis in IOLCA and process LCA separately. This work extends the analytical error propagation approach to any LCA model for performing uncertainty analysis for comparing between different life cycle models. Using this algorithm, the user can create a model of a system which provides LCA results with quantitative confidence bounds.

This paper has three major contributions. First, it develops an algorithm for designing the system boundary between the scales of hybrid life cycle assessment models by using information about nodal environmental impact from network analysis of activity upstream networks. Such a method reduces the subjectivity in LCA model building. Indicators are introduced to compare and choose between models with different system boundaries. Second, characterization indicators based on input data uncertainty and error propagation theory are developed to determine the variance of the model output result. Using such information, the confidence in LCA model results can be quantified. Third, to keep life cycle models within limits of computational and practical feasibility, a complexity indicator is developed based on data cost of modeling and information about the granularity of data sources. The indicator helps limit the generated model from becoming too complex or expensive. Only data or parameter uncertainty has been included in the present study. A simple LCA model with four activities and data at multiple scales is presented as a case study.

The vision put forth in this work is that if life cycle assessment is to make a lasting contribution to industrial ecology and sustainable engineering, it needs to become a more scientifically rigorous approach such that its results will reflect the effect of decisions about boundary selection, boundary delineation, and type of data used. This will allow direct comparison of results from various studies, and even combination of multiple studies to develop better models. This is already happening, as described earlier in this section. Such methods are already available in many other disciplines such as weather forecasting, climate change studies, image recognition, and process operation. Consider the evolution of weather forecasting. In its early days, weather prediction was highly subjective and uncertain, and relied on looking at clouds, wind, and principles drawn up by philosophers. As barometers, hygrometers, and other instruments were invented, vast amounts of weather data became available. In parallel, more sophisticated mathematical models of relevant systems were also developed. Scientifically rigorous frameworks now enable combination of data of various degrees of uncertainties and at multiple levels of resolution with results of model simulations. It is our view that LCA should also follow a similar path as uncertainty information of all types, model, and parameter continue to become available at multiple scales of equipment, processes, and economies.

The rest of the paper is organized as follows. The Section 2 explains SPA and LCA uncertainty analysis in brief. The general framework and algorithm for model generation are explained in the Section 3. Application to the LCA model is provided along with detailed description of each step within Section 3 itself. Insights, shortcomings of the algorithm, and future work are discussed in Section 4.

2 Background

2.1 Structural path analysis

Structural path analysis is a mathematical technique for exploring networks based on infinite series expansion. It can identify individual nodes for hotspot analysis and search for major activity hubs. It has been applied to environmentally-extended input-output (EEIO) models for identifying

hotspots of environmental emissions in an economy. EEIO analysis gives the total environmental impact as:

$$\overline{g} = \overline{B}(\overline{I} - \overline{A})^{-1}\overline{F} \tag{1}$$

where $\overline{B}$ is the vector of direct environmental interventions from all economic sectors and $\overline{F}$ is vector of final demands from the sectors. $\overline{A}$ is the direct requirements matrix that contains information about contribution of different sectors to the input flow of every other sector. $\overline{I}$ is an identity matrix. Overbars on the vectors and matrices represent that they belong to the economy scale. The equation derived from the Taylor series expansion of the Leontief inverse (1) is:

$$\overline{g}_i = b_i f_i + \sum_{j=1}^{n} b_j a_{ji} f_i + \sum_{j=1}^{n}\sum_{k=1}^{n} b_k a_{kj} a_{ji} f_i + \sum_{j=1}^{n}\sum_{k=1}^{n}\sum_{l=1}^{n} b_l a_{lk} a_{kj} a_{ji} f_i + \ldots\ldots \tag{2}$$

where i, j, k, and l are sector indices, n is the total number of sectors, and f_i is the final demand from the i^{th} sector. In this hierarchical disintegration, each term of Eq. 2 corresponds to a path order level determined by the number of linkages to the i^{th} sector final demand. Thus, the term $\sum_{j=1}^{n} b_j a_{ji} f_i$ contains n first-order paths as they are linked to the final demand directly. The first term itself is of the zeroth order. Similarly, $\sum_{j=1}^{n}\sum_{k=1}^{n} b_k a_{kj} a_{ji} f_i$ represents the sum of all n^2 second-order paths. Detailed equations and application to a two-sector economy are provided in Section 1.1 of the Supporting Information (SI).

2.2 Uncertainty propagation in matrix-based LCA

Analytical uncertainty propagation (AUP) (Ku 1966) is used to determine the variance of the output uncertainty based on the variance of each uncertain input variable. The method assumes first-order second-moment form of uncertainty analysis which estimates the second moment (standard deviation) of the result based on a first-order approximation of any linear or non-linear function. As described by Heijungs and Suh (2002), the technology matrix X is the network of economic flows of all unit processes included in the life cycle. The final demand vector F reflects society's requirement for the product. Environmental flows of all unit processes are provided in the intervention matrix B. The total environmental impact is calculated as:

$$g = BX^{-1}F \tag{3}$$

Using error propagation theory, Heijungs (2010) expressed uncertainty of the output as variance of the LCA result using:

$$\sigma^2(g) = \sum_{i,j}(s_j\lambda_i)^2\sigma^2(x_{ij}) + \sum_{j}(s_j)^2\sigma^2(x_j) + 2\sum_{i,j,l,m} s_j\lambda_i s_m\lambda_l\sigma^2(x_{ij}, x_{lm}) + 2\sum_{j,l} s_j s_l\sigma^2(b_j, b_l) + 2\sum_{i,j,l} s_j\lambda_i s_l\sigma^2(x_{ij}, b_l) \tag{4}$$

where s is BX^{-1} and λ is $X^{-1}f$. Equation 4 combines the variances of the different terms of Eq. 3 using the theory of error propagation. The last three terms of the equation represent the covariance between the different parameters. Detailed information of uncertainty error propagation methods in matrix-based LCA is provided in Section 1.2 in the SI.

3 Methodology: a framework for life cycle model generation

The approach for designing LCA models that is proposed in this work is summarized in Fig. 2. The entire algorithm can be divided into two parts. The first part uses a EIO model to perform SPA that provides hotspot and relative activity importance information. This information serves as guidance for building the life cycle models in the second part of the algorithm. The final model is a hybrid model. The initial linear part of this figure initializes the network construction algorithm and relies on SPA, while the circular loop denotes the sequential life cycle model construction. In this algorithm, we use SPA on EIO models to guide the life cycle model building algorithm. Goal and scope definition and inventory analysis, which include system boundary selection between the scales, determining cutoffs, and building the actual LCA model, are performed by this algorithm. For the goal and scope definition step, information such as prior experience, guidance from experts, or information about major activities or upstream environmental hotspots are used to build the life cycle model. SPA initializes with two major data sources: information about the process under study and linkages of this process with other activities. The detailed information of the processes can be obtained from engineering models with exact information about input and output flow rates or average representations of these material flows from standard life cycle inventory databases. EIO models provide an "umbrella" model that contains a snapshot of actual

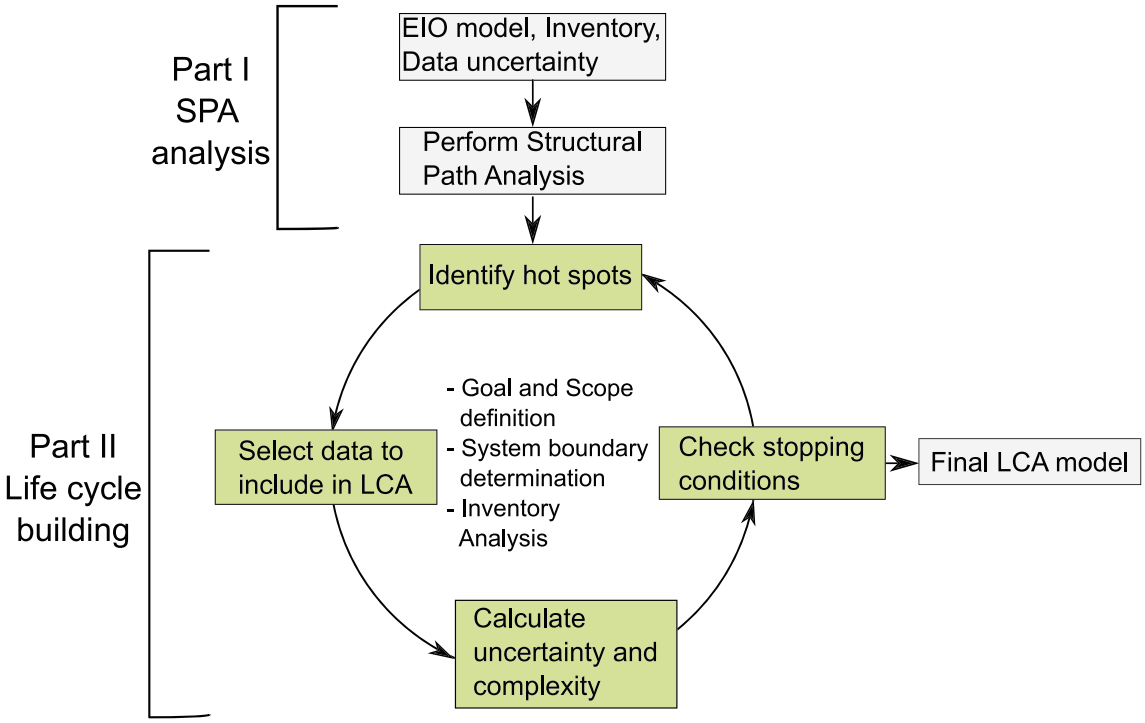

Fig. 2 Overview of methodology involving structural path analysis for guiding life cycle model building algorithm. The algorithm presented in this paper covers goal and scope analysis and inventory analysis steps of life cycle study

movement and consumption of materials and products in the economy at a certain point of time, albeit in monetary units. A network analysis of such information can help identify real-world emission hotspots in the upstream of any product or production technology. This information is used as guidance for sequentially choosing process data for inclusion in the LCA model, while monitoring the trade-off between uncertainty in the results and complexity of the model. The task of building the life cycle models occurs in the circular part of the flowchart in Fig. 2.

3.1 Illustrative example

To effectively communicate how the algorithm works and can be applied for the generation of life cycle models, it is explained with an illustrative example depicted in Fig. 3. The activities shown in this figure represent the hypothetical real-world complete system without any boundary or cutoff error. Normally, for a real life case study, this information is unknown or very difficult to obtain, which ultimately results in boundary selection issues. However, for this illustrative problem, since we know the extent of the entire system, we can compare the life cycle models generated by this algorithm with the complete life cycle emissions and evaluate the performance of the proposed algorithm. P_C is the product under study, for which we need to perform LCA. Let P_C be biodiesel which is produced by the biorefinery process C. Process C, the biorefinery, consumes the products of processes R and T, P_R, and P_{T1}, respectively. P_R is soy oil: the product of soy oil processing depicted as R. Similarly, P_{T1} and P_{T2} both are electricity flows from the power production process T. R, soy oil processing, also consumes electricity from power production process T; the flow is depicted as P_{T2}. The total production of electricity T is thus the sum of P_{T1} and P_{T2}. R also consumes soybean from the farming activity D, where the flow of soybeans is depicted as P_D. The equations governing this system are provided in Eqs. 27–34 in Section 3 of SI. The functional unit for process C, the biorefinery that we are interested in doing a life cycle study for, has an output of 49 kg of P_C or biodiesel. The solution of the sets of equations in the SI for this quantity of biodiesel (P_C) production as well as different quantities is provided in Table 2 of the SI. These equations are used to generate an LCI database for this case study to build the final LCA model. Solving this system of equations will provide the correct value of emissions from the entire system. In this example, the final model generated by the algorithm should be as close as possible to the "true" value calculated from the equations. The process of building an LCI database, relevant assumptions, and requirements are provided in Section 3 of SI. The resulting averaged LCI database is shown in Table 1. The Input and Output columns contain the material flows to and from the processes. Environmental

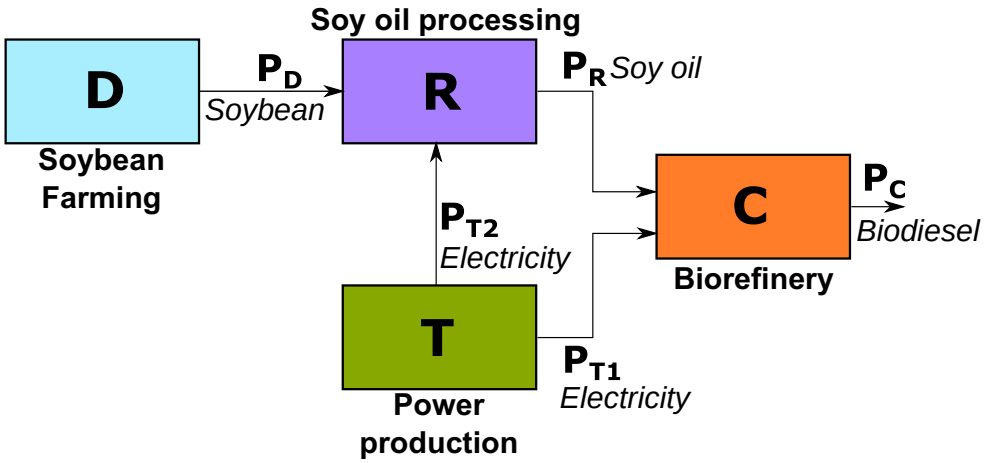

Fig. 3 Illustrative system for demonstration of process-based LCA model generation using the proposed algorithm. R, C, T, and D are activities while the production flow from x activity is denoted by P_x

interventions are listed in the Emission column. In practice, such information can be obtained from LCI such as GREET, USLCI, and EcoInvent.

3.2 Guiding model delineation by SPA

Initially, an EIO model for the region in which the LCA study is based on is required for the purposes of SPA calculations. Information required includes detailed input and output flows from the primary process of concern from either engineering models or process LCA database. The EIO model for a real-life case study should be available to the user from economic databases. These models are easily obtained from open-source national government databases (Bureau of Economic Analysis 2015). For this case study, an illustrative EIO model of the case study is used to explain the algorithm. EIO models are generally represented by $\overline{\mathbf{V}}$ and $\overline{\mathbf{U}}$, known as make and use matrices, respectively. Overbars on symbols refer to components at the economy scale. The make matrix contains total monetary production from every economic sector included in the model. The use matrix contains intersectoral flows. It maps the consumption of economic flows by an economic sector from all other sectors as well as from itself. For this illustrative example, the economy scale make $\overline{\mathbf{V}}$ and use $\overline{\mathbf{U}}$ matrices are defined as:

$$\overline{\mathbf{V}} = \begin{bmatrix} 120 & 0 & 0 & 0 \\ 0 & 189 & 0 & 0 \\ 0 & 0 & 756 & 0 \\ 0 & 0 & 0 & 115.2 \end{bmatrix} \tag{5}$$

$$\overline{\mathbf{U}} = \begin{bmatrix} 0 & 0 & 72 & 0 \\ 36 & 0 & 54 & 0 \\ 0 & 0 & 0 & 0 \\ 96 & 0 & 0 & 0 \end{bmatrix} \tag{6}$$

3.2.1 Setting up the economic model for SPA

The first step that is observed from Fig. 4 involves operations on an EIO model to make it suitable for applying SPA. The primary process of concern needs to be a separate economic sector in the economy model. This presents two cases.

- *Case 1* Primary process of concern is not present in the original EIO model. Then, the primary process information is transformed into an economic activity using price information and the new data are included in the make and use matrices of the EIO model by expanding rows and columns. The addition of new activity is depicted through the addition of a new row and column to EIO model matrices in Fig. 4. This situation often arises for novel technologies. Mostly, EIO models are not available for the current year and are often 4 to 7 years old. Thus, novel technologies that have been introduced in recent times do not appear in the economic data from previous years.
- *Case 2* Primary process of concern is contained in the original EIO model but is aggregated in a larger economic sector. The primary process information is transformed into an economic activity using price information. However, before including it in the EIO model matrices, the activity needs to be separated

Table 1 Life cycle inventory data for case study illustrative example

Processes	Emissions	Input	Output
R	96	$P_D = 3.43 P_{T2} = 1.71$	$P_R = 8.57$
T	27.84		$P_{T1} = 4.28$, $P_{T2} = 1.71$
C	183.48	$P_{T1} = 4.28$; $P_R = 8.57$	$P_C = 30$
D	15.36		$P_D = 3.43$

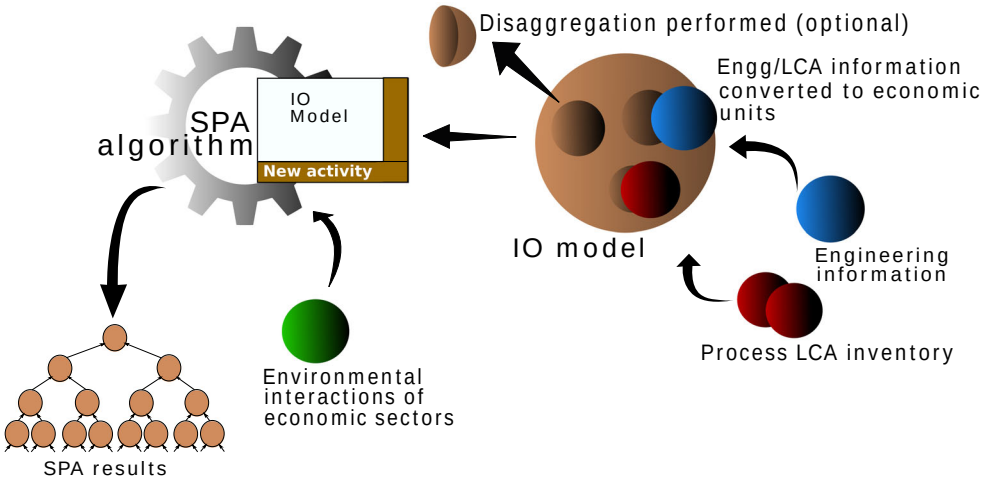

Fig. 4 Disaggregation of economic sectors and SPA of EIO model to obtain hotspot information and sector importance in upstream network of processes. It includes algorithm steps from start to structural path analysis in Fig. 5

out from its aggregated state from the original EIO model to prevent double counting (Suh 2004). For example, if the primary process whose life cycle model is to be generated is a corn ethanol manufacturing process, then the economic sector in the EIO model that contains this process is the other basic organic chemical manufacturing sector. The upstream linkages for this sector as calculated by SPA will include all pathways that are necessary to manufacture thousands of chemicals manufactured by this sector. Hence, the life cycle network obtained will be the aggregation of thousands of networks. The linkage of only corn ethanol manufacturing will be indiscernible within this aggregated life cycle network. To rectify this problem, the primary process is disaggregated from the EIO model. This situation arises for most established technologies, for example coal-based power generation, crude oil refining, copper ore mining, etc.

This step is expressed as a question in the flowchart in Fig. 5. Disaggregation of an economic sector, if needed, is accomplished by using Eqs. 26–35 in Hanes and Bakshi (2015) to obtain disaggregated make and use matrices.

For the illustrative example, the primary process of concern C is contained in the original EIO model aggregated with a larger economic sector, making this a case 2 situation as explained. The inventory information for production of 30 kg of P_C is transformed into an economic activity using price information. However, before incorporating it in the EIO model matrices, the activity needs to be disaggregated from the original EIO model. This is done by using Eqs. 7 and 11. Permutation matrices $\mathbf{P}_P^E$ and $\mathbf{P}_F^E$ contain information on the relationships between parent sectors and their commodities, and constituent activities and processes and their products. These matrices along with price data vectors $\hat{\mathbf{p}}_\mathbf{r}$ are used to "translate" flows in the various make and use matrices from one scale to another. $\mathbf{X^E}$ is known as the cutoff matrix from economy scale to primary process P_C. It contains the monetary flow of materials from the economy scale to P_C. Detailed steps for calculation are provided in Section 2 in the SI. As P_C is being disaggregated out, these flows also need to be removed.

$$\overline{\mathbf{V}}^* = \overline{\mathbf{V}} - \hat{\mathbf{p}}_\mathbf{r}(\mathbf{P}_P^E)^T\mathbf{V}(\mathbf{P}_F^E)^T \tag{7}$$

$$(\mathbf{P}_P^E)^T = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix} \tag{8}$$

$$(\mathbf{P}_F^E)^T = \begin{bmatrix} 0 & 0 & 1 & 0 \end{bmatrix} \tag{9}$$

$$\overline{\mathbf{V}}^* = \begin{bmatrix} 120 & 0 & 0 & 0 \\ 0 & 189 & 0 & 0 \\ 0 & 0 & 576 & 0 \\ 0 & 0 & 0 & 115.2 \end{bmatrix} \tag{10}$$

$$\overline{\mathbf{U}}^* = \overline{\mathbf{U}} - \mathbf{X^E}\mathbf{P}_P^E \tag{11}$$

$$\mathbf{X^E} = \begin{bmatrix} 17.2 \\ 12.9 \\ 0 \\ 0 \end{bmatrix} \tag{12}$$

$$\overline{\mathbf{U}}^* = \begin{bmatrix} 0 & 0 & 54.8 & 0 \\ 36 & 0 & 41.1 & 0 \\ 0 & 0 & 0 & 0 \\ 96 & 0 & 0 & 0 \end{bmatrix} \tag{13}$$

The primary process which was subtracted out from the original EIO model is now included as a separate sector to create the following augmented make and use matrices. The augmented make and use matrices are constituted of the old make and use matrices along with the primary process

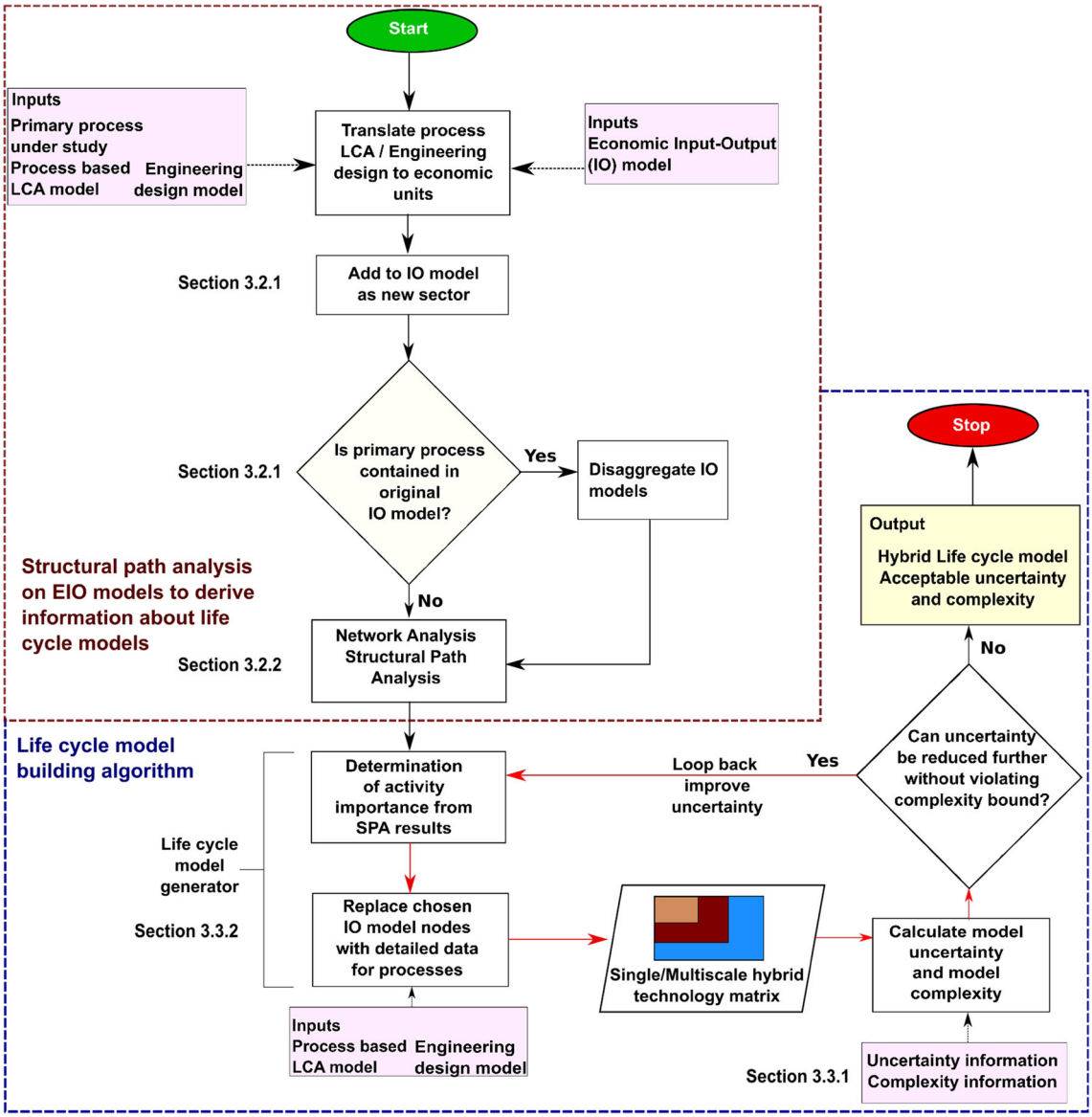

Fig. 5 Flowchart showing overall progression of model generation algorithm from initialization to obtaining final LCA model of reasonable uncertainty and complexity

separately added in economic terms in the fourth column and the row as shown in Eq. 17.

$$\overline{\mathbf{V}}_n^* = \begin{bmatrix} 120 & 0 & 0 & 0 & 0 \\ 0 & 189 & 0 & 0 & 0 \\ 0 & 0 & 576 & 0 & 0 \\ 0 & 0 & 0 & 115.2 & 0 \\ 0 & 0 & 0 & 0 & 180 \end{bmatrix} \tag{14}$$

$$\overline{\mathbf{U}}_n^* = \begin{bmatrix} 0 & 0 & 54.8 & 0 & 17.2 \\ 36 & 0 & 41.1 & 0 & 12.9 \\ 0 & 0 & 0 & 0 & 0 \\ 96 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \tag{15}$$

The new modified direct requirements matrix is calculated using:

$$\overline{\mathbf{A}}_n^* = \overline{\mathbf{U}}_n^* ((\overline{\mathbf{V}}_n^*)^T)^{-1} \tag{16}$$

$$\overline{\mathbf{A}}_n^* = \begin{array}{c} \begin{matrix} \overline{S}_R & \overline{S}_T & \overline{S}_C & \overline{S}_D & \overline{S}_{CE} \end{matrix} \\ \begin{bmatrix} 0 & 0 & 0.0951 & 0 & 0.0956 \\ 0.3 & 0 & 0.0714 & 0 & 0.0717 \\ 0 & 0 & 0 & 0 & 0 \\ 0.8 & 0 & 0 & 0 & 0 \\ 0.8 & 0 & 0 & 0 & 0 \end{bmatrix} \end{array} \begin{matrix} \overline{S}_R \\ \overline{S}_T \\ \overline{S}_C \\ \overline{S}_D \\ \overline{S}_{CE} \end{matrix} \tag{17}$$

The columns and rows of these matrices refer to the economic sectors that include the activities in the system in Fig. 3. The first column refers to the economic sector

$\overline{S}_R$ which includes the activity R. The second column refers to the economic sector $\overline{S}_T$ which includes the activity T. Similarly, the third and fourth columns/rows refer to economic sectors $\overline{S}_C$ and $\overline{S}_D$ that contain respective activities. For example, if R is a bioethanol-producing industry, then $\overline{S}_R$ is the basic organic chemical manufacturing economic sector. $\overline{S_{CE}}$ is the economic sector that has been created through disaggregation of $\overline{S_C}$. This can be described as removing corn ethanol manufacturing from the basic organic chemical manufacturing economic sector $\overline{S}_C$ and modelling it separately as one activity.

3.2.2 SPA applied to primary economic sector

Structural path analysis is performed as explained in Section 2.1 on the separate sector built specifically from the primary process as shown in Fig. 4. In Fig. 2, this step relates to the determination of activity importance and hotspots from SPA results. SPA is applied to this augmented economy model with final demand for ($\overline{S}_{CE}$) and paths extracted from the network and ranked based on their percent contribution to total emissions. Performing structural path analysis by application of Eq. 2, the upstream life cycle network information is obtained. Table 2 shows the first few results of many network linkages as computed by SPA. The linkages are sorted and ranked according to their percentage contribution to the environmental impact. From that information, the model generator in Fig. 5 determines which sectors need to be sequentially considered for inclusion in the life cycle model, the rule being activity having the highest contribution to total environmental impact should be included first.

3.3 Life cycle model generation

The life cycle model building algorithm constitutes the next step as shown in Fig. 2. This algorithm uses data as guidance from the SPA results and enables construction of the LCA models. Measurement of model quality and ranking of models are accomplished through the development of characterization indicators based on uncertainty and complexity of LCA models. Using such indicators, different LCA models can be compared against each other and chosen based on user's preferences. A second decision maker as shown in Fig. 5 is employed to determine if the built LCA model has satisfied the desired values of the indicators. Improvements over the uncertainty indicator are performed through the iteration loop (formed using the red arrows in Fig. 5) by shifting command back to the LCA model generation steps. If improvement is not possible any more or user constraints fail, the algorithm stops and final LCA model with acceptable uncertainty and complexity is obtained.

Table 2 SPA results obtained by applying (2) to EIO model for illustrative example

Rank	Path	Order	% Contribution
1	$\overline{S}_{CE}$	Zero	65.1
2	$\overline{S}_R \rightarrow \overline{S}_{CE}$	First	24.3
3	$\overline{S}_T \rightarrow \overline{S}_{CE}$	First	5
4	$\overline{S}_D \rightarrow \overline{S}_R \rightarrow \overline{S}_{CE}$	Second	3.2
5	$\overline{S}_T \rightarrow \overline{S}_R \rightarrow \overline{S}_{CE}$	Second	2

3.3.1 Indicators for life cycle model

Two parameters are selected for determination of model quality and serve as a basis for comparison between built models. The quality of LCA models to give results with low error probability was captured using an uncertainty indicator while size of model that directly translates to time and data requirements is captured using a complexity indicator.

Uncertainty analysis The quality of the generated models as guided by the SPA results is evaluated through uncertainty analysis of the results. As mentioned in Section 1, only parameter uncertainty has been considered. The main goal of this analysis is to gather the uncertainty in the input data, propagate it according to the model structure, and determine uncertainty in the output from the model. For this study, we employ analytical uncertainty propagation method due to its simplicity, ease of use, and suitability to our model. The uncertainty analysis operation is done simultaneously within the model generation loop as shown in Fig. 5.

Using the analytical error propagation method explained in Section 2, the variance information of input data for building LCA models can be translated into variance information for the final result from the LCA model. The life cycle environmental impact g on any matrix-based LCA procedure is calculated using Eq. 3. The uncertainty analysis operation can be visualized as depicted in Fig. 6. Essentially, the error propagation equation is combining uncertainty data for the environmental intervention matrix B, and the technology matrix X to calculate the uncertainty of the final life cycle impact g.

This work expands the approach of Heijungs explained in Section 2 so that uncertainty propagation can be achieved for a multiscale life cycle model, such as hybrid LCA. The life cycle environmental impact g on any matrix-based LCA procedure is calculated using:

$$\underline{\overline{g}} = \underline{\overline{BX}^{-1}}\,\underline{\overline{F}} \tag{18}$$

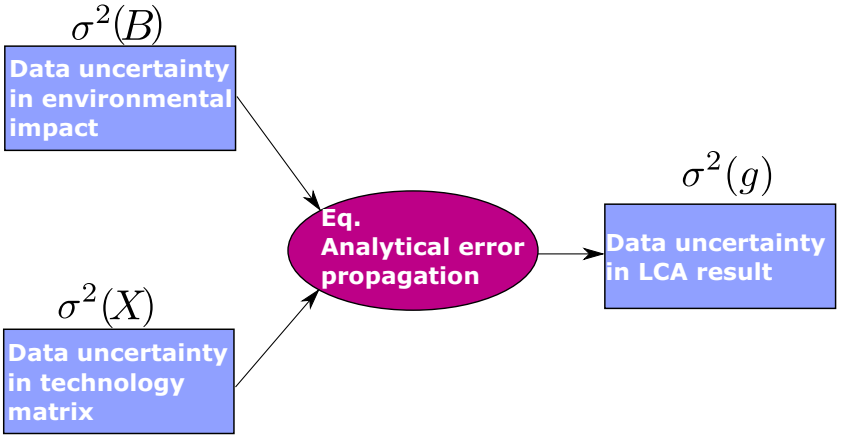

Fig. 6 Data uncertainty analysis of life cycle models using error propagation method

The double bars over the matrices and vectors are to emphasize the fact that they can be single scale or multi-scale/hybrid in nature. Heijungs used Eq. 4 to determine propagation of uncertainty in single-scale LCA models. Modifying Eq. 4 for encompassing multiple scales, propagation of the error, expressed as variance in the input data, to the output LCA result for any multiscale framework is obtained through:

$$\begin{aligned}\sigma^2(\underline{\overline{g}}) = &\sum_j (\underline{\overline{X}}^{-1} f)_j^2 \upsilon + \sum_{i,j} \{(\underline{\overline{BX}}^{-1})_i (\underline{\overline{X}}^{-1} f)_j\}^2 \psi \\ &+2\sum_{i,j,l} (\underline{\overline{X}}^{-1} f)_j (\underline{\overline{BX}}^{-1})_i (\underline{\overline{X}}^{-1} f)_l \tau \\ &+2\sum_{i,j,l,m} (\underline{\overline{X}}^{-1} f)_j (\underline{\overline{BX}}^{-1})_i (\underline{\overline{X}}^{-1} f)_m (\underline{\overline{BX}}^{-1})_l \gamma \\ &+2\sum_{j,l} (\underline{\overline{X}}^{-1} f)_j (\underline{\overline{X}}^{-1} f)_l \omega \end{aligned} \quad (19)$$

where,

$$\upsilon = \begin{cases} \sigma_1^2(b_j) \text{ if } j \le n_1 \\ \sigma_2^2(b_j) \text{ if } j \le n_1 + n_2 \\ \vdots \\ \sigma_k^2(b_j) \text{ if } j \le n_1 + n_2 + n_3 + \cdots + n_k \end{cases}$$

$$\psi = \begin{cases} \sigma_1^2(x_{ij}) \text{ if } j \le n_1 \\ \sigma_2^2(x_{ij}) \text{ if } j \le n_1 + n_2 \\ \vdots \\ \sigma_k^2(x_{ij}) \text{ if } j \le n_1 + n_2 + n_3 + \cdots + n_k \end{cases}$$

$$\gamma = \begin{cases} \text{cov}_1(x_{ij} x_{lm}) \text{ if } j \le n_1 \\ \text{cov}_2(x_{ij} x_{lm}) \text{ if } j \le n_1 + n_2 \\ \vdots \\ \text{cov}_k(x_{ij} x_{lm}) \text{ if } j \le n_1 + n_2 + n_3 + \cdots + n_k \end{cases}$$

$$\tau = \begin{cases} \text{cov}_1(x_{ij} b_l) \text{ if } j \le n_1 \\ \text{cov}_2(x_{ij} b_l) \text{ if } j \le n_1 + n_2 \\ \vdots \\ \text{cov}_k(x_{ij} b_l) \text{ if } j \le n_1 + n_2 + n_3 + \cdots + n_k \end{cases}$$

$$\omega = \begin{cases} \text{cov}_1(b_j b_l) \quad \text{if } j \le n_1 \\ \text{cov}_2(b_j b_l) \quad \text{if } j \le n_1 + n_2 \\ \vdots \\ \text{cov}_k(b_j b_l) \quad \text{if } j \le n_1 + n_2 + n_3 + \cdots + n_k \end{cases} \quad (20)$$

and,

$\underline{\overline{g}}$ Total environmental impact from model
$\underline{\overline{X}}$ Multiscale/single-scale technology matrix
$\underline{\overline{F}}$ Multiscale/single-scale final demand matrix
x Individual values of the $\underline{\overline{X}}$ matrix.
$\underline{\overline{B}}_j$ Multiscale/single-scale environmental interventions matrix
b Individual values of the $\underline{\overline{B}}$ matrix.
1, 2, 3.....k Represents the different scales that are included in the hybrid life cycle model
n_k Number of activities in the k^{th} scale

Equation 19 computes variance of output result σ^2(g) from the uncertainty information of the input data of life cycle matrices. The parameters $\upsilon, \mu, \omega, \tau, \gamma,$ and ψ are used to substitute variance and covariance information in the equation. Thus, using Eq. 19, the uncertainty of the input data in matrices $\underline{\overline{X}}$ and $\underline{\overline{B}}$, considered as variance of the parameters, is propagated to variance of the final LCA result $\underline{\overline{g}}$. However, variance of the result alone does not provide a good estimate of the quality. To incorporate both the standard deviation and mean of LCA result, the final parameter used as a measure of uncertainty from the generated model is the relative standard deviation (Everitt 2006) expressed as:

$$RSD = \frac{\sigma(g)}{g} \quad (21)$$

When comparing different models, the one with the lower RSD is considered to be better in terms of quality as it has lower uncertainty.

Obtaining uncertainty information Obtaining uncertainty information for life cycle inventories and economic data can prove to be difficult. The multi-regional EORA EIO model (Lenzen et al. 2013) is one source of uncertainty data for EIO models. The USEEIO model (Yang et al. 2017) describes a formal method of data quality evaluation for US economic model interventions. Uncertainty information for US-based life cycle inventories can be sourced from EcoInvent. If unavailable, they can be substituted with similar information for the European subcontinent (Ecoinvent Database 2015). If information is not available, the user can run the algorithm by considering reasonable assumptions. It is expected that as activities are shifted from the aggregated economy scale to a detailed process-based LCA scale, uncertainty decreases due to the greater resolution and more details of the models used. Thus, by fixing one scale as a basis and setting the uncertainty values of other scales based on that, the user can temporarily bypass the challenge of missing uncertainty data as explained in the case study. This approach currently deals with only data or parameter uncertainty. Among other uncertainties arising from temporal variations, spatial variations can be incorporated as they are just subsets of parameter uncertainty values. However, different classes of uncertainties in LCA, such as scenario uncertainty and model uncertainty, have not been incorporated in this proposed approach. The proposed approach is also capable of using pedigree matrix uncertainty data to perform the error propagation calculations and derive an uncertainty indicator based on quality values from the pedigree matrices. As pedigree matrices (Ciroth et al. 2016) for LCIs are becoming more widely available, this approach will thus be useful to determine uncertainty indicator for the algorithm.

Application to illustrative example For variance calculation, uncertainty information about the different scales are required. For the purpose of the case study, at the economy scale, variance ($\overline{b}_j$) and variance ($\overline{x}_{ij}$) are assigned to be 0.1. Variance terms for value-chain scale, variance ($\underline{b}_j$), and variance ($\underline{x}_{ij}$) are assigned to be 0.01. The economy scale is expected to have the highest uncertainty due to its high aggregation of data as well as extensive range of sources from which EIO data are collected. The process LCA scale is not as aggregated as EIO models and hence is likely to have lower values of variance. Covariance terms are assumed to be zero.

Model complexity analysis Complexity of a model is defined as a measure of the size of a model, inter-connectedness of the modules, or the computational demand of building and solving a model. Trade-off between complexity and accuracy is understood by studying a system governed by a high-order equation but modeled with a lower order one. The results obtained from the use of the lower order function are bound to be less accurate. Complexity is usually inversely related to the uncertainty of the model: a model with low uncertainty tends to be more complex. However, it is not always true. The complexity parameter is represented as M_c, and various approaches may be used to quantify it. As seen from Fig. 5, complexity analysis is done simultaneously with uncertainty analysis. The two parameters, uncertainty and complexity, are used to determine the usefulness of the model.

Cost-based complexity The monetary costs of inventory data and of building a model are convenient indicators of model complexity, since usually a more complex model will have a higher cost of generation. The cost will depend on the number of units that are included in the model. For a multiscale model, different scales will have different costs of data and modeling. The economy-scale information is available to be used directly in a matrix format. Its modeling cost is lowest since such data are usually freely available from government agencies. The complexity and costs of process LCA data are higher than economy scale because it has to be obtained from LCI inventories, requires more time investment, and sometimes needs to be bought. The total model price will depend on the relative size of the scales within the model. Inclusion of sectors at scales that cost more to build increases the total cost of the model and vice versa. Thus, cost of modeling at different scales is proportional to their complexity. The total model generation cost is calculated as:

$$M_c = \sum_i n_i p_i \tag{22}$$

where n_i is the number of units at the i^{th} scale and p_i is the cost of modeling one unit at the i^{th} scale. It is evident from Eq. 22 that as n_i with high generation cost increases, total model generation cost rises. During the model generation algorithm, model generation cost needs to be limited within a certain maximum value, indicated by $M_{c,max}$ to keep the model complexity under check.

Information for this cost-based complexity may be obtained in different ways. If quantitative data are not available, the cost can be set based on the amount of effort that the user spends in developing the model. The cost of modeling at a certain scale is taken as the basis and if other scales require more effort, their cost of modeling is adjusted accordingly.

Granularity-based complexity parameter The granularity-based model complexity parameter depends on the number

of information sources that have been aggregated to determine data at a certain scale. For example, life cycle process inventories of a product are obtained by aggregating information from a number of individual activities that produce the same product. Each of these individual sources is considered to be a *grain* of information that is agglomerated with other grains to get the final data. This parameter is obtained as:

$$M_c = \sum_{k=1}^{n_i} z_i^{-1} \tag{23}$$

where z_i represents the number of information sources that have been aggregated to determine the data at the i^{th} scale. n_i is the number of sectors in i^{th} scale. From Eq. 23, it is observed that a large increase in the number of units at a scale with high aggregation causes equivalent increase of M_c compared with a smaller increase in the number of units at a scale with low aggregation or *granularity*. Generally, as information is aggregated, it is expected that error in the data increases, thus increasing the uncertainty. Hence, to reduce data error in the model, it is necessary to include more units at scales with low aggregation thus increasing M_c. Complexity and uncertainty analysis shown in Fig. 5 forms the last step of the model builder loop.

Complexity data source For the illustrative example, the granularity information needs to be assumed. As described in the mock LCI database building exercise in Section 3 in SI, seven different systems were aggregated to generate the LCI information. Thus, the granularity of every data source for this case study is assumed to be 7. For the economic IO model, each sector has a different granularity value. It was assumed for purposes of this illustrative IO model that sector $\overline{R}$ was aggregated from 10 sources, $\overline{T}$ from 15, $\overline{C}$ from 6, and $\overline{D}$ from 4 data points. However, for practical examples, such information may be found from detailed information about the source of the LCI. Detailed documentation of most LCI databases contains this information. Data for granularity of EIO models can be obtained by exploring the census database that contains the number of facilities for every economic sector that participated in providing information for economic surveys. Using this parameter, a clear trade-off between uncertainty and complexity can be established for the models being generated. The maximum value of M_c, $M_{c,max}$ may be chosen by the algorithm user, and reflects the budgetary allowance for building the model.

3.3.2 Life cycle model building algorithm steps

Life cycle model generator builds life cycle model matrices by reading information from the SPA results and using information from engineering processes, economic models, process life cycle inventory databases, etc. Its functioning is depicted in Fig. 5 as a life cycle model generator. Mathematical representation of the algorithm is provided in the SI in Algorithm 1. It operates in the following manner.

- Uncertainty and complexity analysis relies on calculations of RSD and M_c. Algorithmic iterations can be expressed as:

$$RSD_m = \frac{\sigma(\overline{\underline{BX}}^{-1}\overline{\underline{F}})_m}{\overline{\underline{BX}}^{-1}\overline{\underline{F}}_m} \tag{24}$$

$$M_{c,m} \leq M_{c,max} \tag{25}$$

$$\mid RSD_m - RSD_{m-1} \mid \geq \xi \tag{26}$$

$$RSD_m - RSD_{m-1} \leq 0 \tag{27}$$

  where RSD_m is the RSD value of the model in the current iteration m and RSD_{m-1} is that of the previous iteration. The algorithm loop is shown in Fig. 5. The life cycle generator is programmed in a way such that it always tries to create models with the least possible output uncertainty. As seen in Fig. 5, the first step in this loop is the creation of an output model in the form of a technology matrix by the generator. Then, using the uncertainty and complexity data, the RSD and M_c values are calculated and passed on to the parameter checker where the constraints are checked. Increase in model complexity is resisted by the constraint on M_c which does not allow the model to increase beyond a pre-decided complexity $M_{c,max}$. Another constraint based on tolerance limits is also employed. For this constraint, the output RSD of the life cycle model from the previous iteration of the algorithm loop is compared with the output RSD of the model of the current iteration. If this value is lower than a certain tolerance ξ, then the algorithm loop stops. A third constraint monitors if the RSD of the current life cycle model is lower than the RSD of the model from a previous iteration. The entire algorithm is graphically described in Fig. 5 using red arrows.
- For hybrid life cycle models, the generator starts with modeling the simplest scale or the one with lowest complexity as shown in Fig. 7. The EIO model, being the simplest to apply, is chosen as the starting point of the algorithm. The EIO model contains all linkages shown in SPA results in Table 2. Subsequently, in the generation algorithm, the activities are shifted from the economy scale to the process-based LCA scale and model parameters monitored for improvement.
- When activities are shifted across scales, they need to be disaggregated from the source scale and added to the destination scale. It is achieved by first modeling the activity at the desired scale and then using

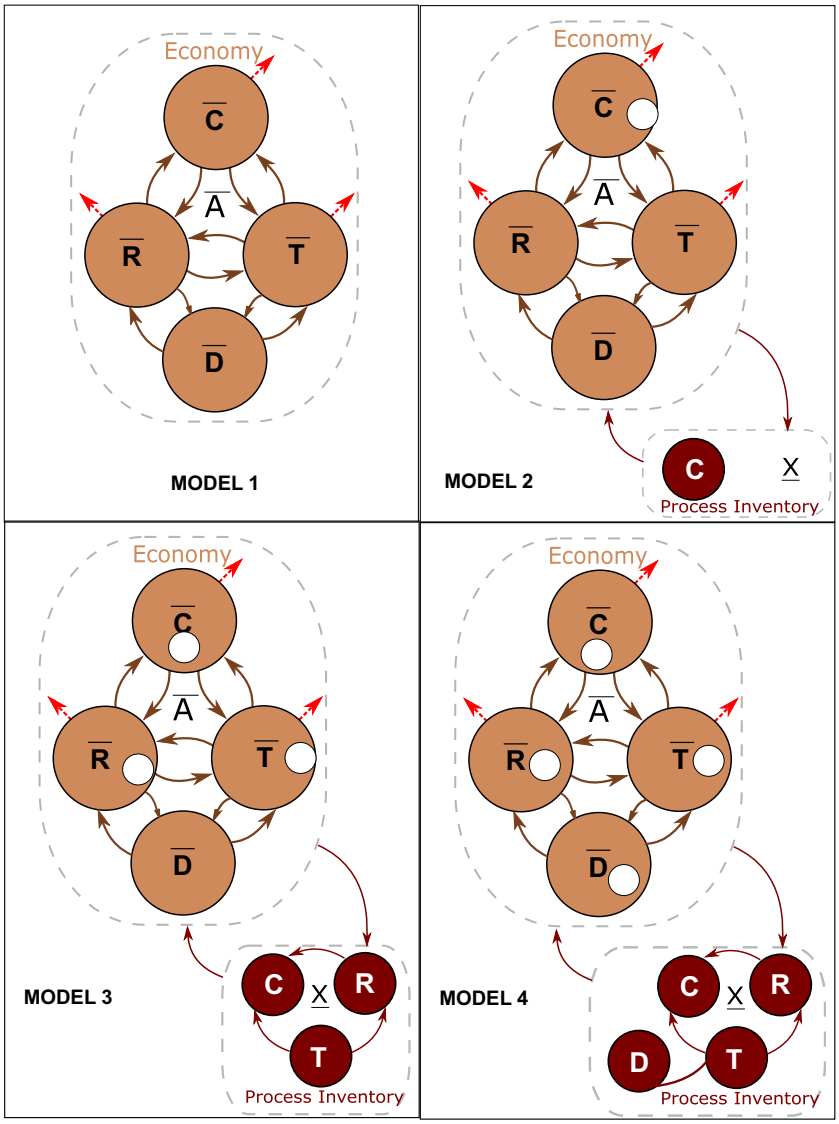

Fig. 7 The initial *Model* 1 started with only the economy scale with sectors $\overline{R}$, $\overline{C}$, $\overline{T}$, and $\overline{R}$. To reduce RSD, life cycle activities were removed from the economy scale as depicted by white blank spaces. Main process C was removed from economy scale sector $\overline{C}$ in *Model* 2. In *Model* 3, processes R and T are removed from their respective economic sectors and modeled in the process-based LCA scale. In the final model, R, T, C, and D are at the process-based LCA scale

disaggregation equations to subtract it from the source scale. Equations 7–11 from this paper and equations 26–35 in Hanes and Bakshi (2015) are the relevant equations for performing such operations.

- For every new model that is built by performing the listed operations, the goal is to decrease uncertainty or RSD. The generator uses an assumption that models using the highest detail or resolution have the lowest uncertainty. For example, if the same manufacturing process is represented using a nonlinear engineering model and a linear life cycle inventory model, mostly, the engineering model will have a lower error in its results which translates into lower uncertainty. Thus, the inherent property of the generator is to always create models with a lower uncertainty, hence the constraints of Eqs. 26 and 27. However, this increases model complexity. Therefore, model development involves a trade-off between uncertainty and complexity as shown in Fig. 8. Sector linkages which have high impact contribution as shown by SPA are chosen to be shifted. Moving such activities or linkages across scales causes a larger change in overall model variance compared with linkage with lower emission contribution. This is because both $\sigma^2(g)$ in Eq. 4 and impact contribution in SPA calculations depend upon b and s.
- The algorithm loops back as shown in Fig. 5 to check if modifications can be performed to reduce RSD of the model. During this operation, the generator for hybrid models removes the activity from the original scale and remodels it at other scales, if available, to reduce RSD and while satisfying (26) and (27). If remodeling is not possible due to violation of constraints, the model from the previous iteration with acceptable constraints is chosen as the final one and the algorithm quits.

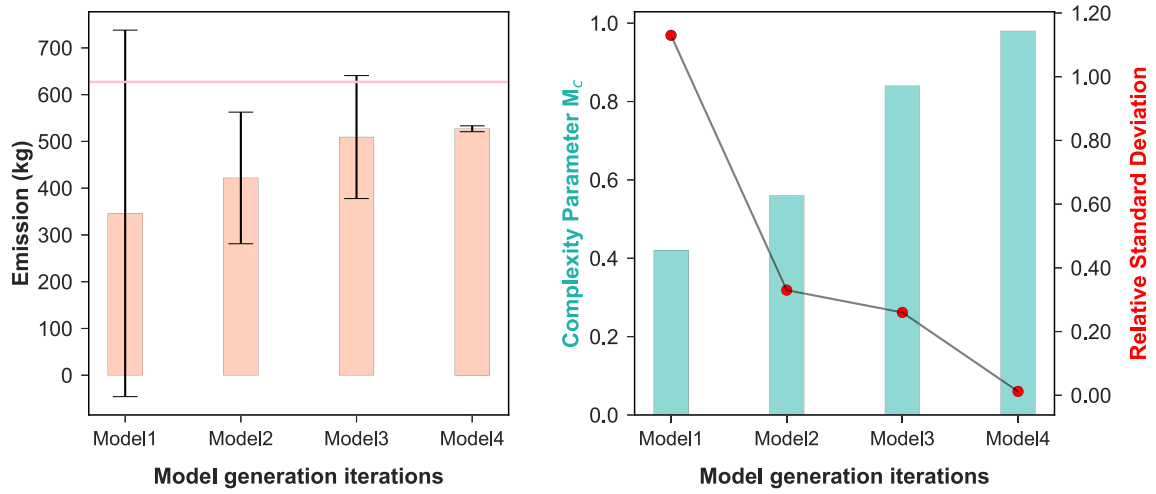

Fig. 8 Life cycle environmental intervention results along with standard errors for generated models shown on the left. Pink line shows the true emission at 627 kg. Progression of RSD and complexity parameter M_c along with model generation iterations shown on the right

3.3.3 Life cycle model generation for an illustrative example

The proposed model generation algorithm starts with obtaining guidance information from the SPA results present in Table 2. Detailed information for derivation of the technology matrices are provided in Section 4 of the SI for each iteration. Hybrid LCA model generation starts by first modeling the scale with lowest complexity, or in other words, the simplest one to model. The EIO model is all encompassing. It contains all linkages shown in SPA results in Table 2. At this step, the EIO model is equivalent to the LCA model. The EIO model for this problem is described in Section 3.2. Thus, reusing information from that model, the $\overline{X}$ matrix is obtained as:

$$\overline{X}_1 = \overline{I} - \overline{A} = \begin{bmatrix} 1 & 0 & -0.0952 & 0 \\ -0.3 & 1 & -0.714 & 0 \\ 0 & 0 & 1 & 0 \\ -0.8 & 0 & 0 & 1 \end{bmatrix} \tag{28}$$

The $\overline{A}$ matrix is obtained from the initial $\overline{V}$ and $\overline{U}$ matrices from Eqs. 5 and 6. As mentioned earlier, the entire analysis is being done for 49 kg of P_C; a final demand of a similar amount needs to be attached to this EIO model. The third column and row of this matrix contain the sector S_C. The final demand vector is given as:

$$\overline{F}_1 = \begin{bmatrix} 0 \\ 0 \\ 294 \\ 0 \end{bmatrix} \tag{29}$$

where the number 294 is the monetary value of 49 kg of P_C (6$/unit). The prices of individual products are provided in Table 1 in the SI. Derivations of the $\overline{V}$ and $\overline{U}$ matrices as well as the $\overline{B_1}$ interventions matrices are beyond the scope of the paper and are not relevant because these are information or data that are expected to be available to the user for the particular region or case study. The interventions vector in units of kilogram of emission per dollar is obtained as:

$$\overline{B_1} = [3, 0.83, 0.76, 0.60] \tag{30}$$

g for this model during the first iteration is calculated as:

$$\overline{g_1} = \overline{B}_1 \overline{X}_1^{-1} \overline{F}_1 = 346.12 \tag{31}$$

Relative standard deviation (RSD) is calculated using:

$$RSD = \frac{\sigma(BX^{-1}F)}{(BX^{-1}F)} \tag{32}$$

$\sigma(BX^{-1}f)$ is calculated using Eq. 4. A Matlab code is used for performing this calculation which is provided with the supplementary files. Thus, RSD for the model in the first iteration is obtained as:

$$RSD_1 = \frac{391.86}{346.12} = 1.13 \tag{33}$$

The model can be visualized as model 1 in Fig. 7. From the description and assumptions taken for building EIO model of the illustrative example, the number of activities that have been aggregated to create the make and use matrix information is known. These derivation and data came from assumptions for building the illustrative EIO model and are beyond the scope of the paper. The number of activities assumed to have been aggregated for obtaining the economic sectors of $\overline{S}_R$, $\overline{S}_T$, $\overline{S}_C$, and $\overline{S}_D$ are 10, 15 and 6 and 12 respectively. Thus, as all the economic sectors are included in the EIO model used for this step, the model complexity is calculated as:

$$M_{c,1} = 1/10 + 1/15 + 1/6 + 1/12 = 0.42 \tag{34}$$

In iteration 2, a new model is generated with lower RSD as described. From the SPA results in Table 2, it is observed

that economic sector $\overline{S}_{CE}$ has the largest contribution to environmental impact. $\overline{S}_{CE}$ is the economic sector built by converting process C to monetary input–output units using price information. It is the disaggregated $\overline{S}_C$ sector representing only one process C. Thus, for model 2, P_C needs to be included at the process-based LCA scale using information directly from the LCI database as shown in Fig. 7. That is achieved easily by disaggregating the third row and column in the previous $\underline{\overline{X}}$ matrix. The information for the activity C is obtained from the LCI database, provided in Table 1. Model 2 now becomes a hybrid multiscale model. Thus, matrix terms are represented using both over- and underbars. It contains both the economy and process LCA scales. The technology matrix is obtained as:

$$\underline{\overline{X}}_2 = \begin{bmatrix} 1 & 0 & -0.0951 & 0 & -17.2 \\ -0.3 & 1 & -0.714 & 0 & -12.9 \\ 0 & 0 & 1 & 0 & 0 \\ -0.8 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 30 \end{bmatrix} \tag{35}$$

Detailed disaggregation calculations are shown under iteration 2 in Section 4 of the SI. The number 30 in the last cell of the matrix is directly taken from Table 1 for the flow P_C. From the same table, we see that the flows of P_R and P_{T1} into C for production of 30 units of P_C are 8.6 and 4.3, respectively. However, these flows need to be converted from physical to monetary units because activities R and T are still in the economy scale. Thus, multiplying with monetary information provided in Table 1 in the SI, the flows change to dollar amounts of 17.2 and 12.9 respectively. The final demand for this hybrid multiscale model also changes. The main process C is now in the process-based LCA scale and modeled with physical units from the LCI database in Table 1. Thus, the final demand for 49 kg of P_C now becomes:

$$\underline{\overline{F}}_2 = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 49 \end{bmatrix} \tag{36}$$

Direct physical values now show up in the final demand matrix. The environmental interventions matrix also expands to a multiscale vector.

$$\underline{\overline{B}}_2 = \begin{bmatrix} 3 & 0.8285 & 0.763 & 0.6 & 183.5 \end{bmatrix} \tag{37}$$

The last number 183.4 is again obtained from the LCI database in Table 1 for the production of 30 units of P_C. g for this model during the second iteration is calculated as:

$$\underline{\overline{g_2}} = \underline{\overline{B}}_2 \underline{\overline{X}}_2^{-1} \underline{\overline{F}}_2 = 421.89 \tag{38}$$

RSD is calculated using Eq. 32. Thus, RSD for the model is obtained as:

$$RSD_2 = \frac{140.71}{421.89} = 0.33 \tag{39}$$

The model can be visualized as model 2 in Fig. 7. The white cutout denotes the disaggregation of C from the economy scale. Activity C is obtained from the LCI database. According to the description of building the LCI database in the SI in Section 3, it has been mentioned that 7 different systems are aggregated to derive the LCI information in Table 1. Thus, the complexity of this new model is calculated as:

$$M_{c,2} = 0.42 + 1/7 = 0.56. \tag{40}$$

For the next step, as seen from Fig. 5, we go back to the SPA results in Table 2. We see that S_R is the next most important activity that needs to be modeled at the process-based LCA scale. Thus, we need to disaggregate this activity from the economy model sector $\overline{S}_R$ and then include it again using life cycle inventory data. In the next iteration, similar operations are required for activity T. For brevity, we combined these two steps and both activities R and T are simultaneously modeled in the process LCA scale. However the flow of material from T to R is still kept at the economy scale. This is because using SPA results, we are modeling each pathway as a new addition to the LCA model rather than the complete activity. The flow from T to R is at the bottom of Table 2, so we are not modeling it at this step. In this iteration, we obtain:

$$\underline{\overline{X}}_3 = \begin{bmatrix} 1 & 0 & -0.095 & 0 & 0 & 0 & 0 \\ -0.30 & 1 & -0.07 & 0 & 0 & -5.1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ -0.80 & 0 & 0 & 1 & 0 & -13.6 & 0 \\ 0 & 0 & 0 & 0 & 6 & 0 & -4.3 \\ 0 & 0 & 0 & 0 & 0 & 8.57 & -8.57 \\ 0 & 0 & 0 & 0 & 0 & 0 & 30 \end{bmatrix} \tag{41}$$

All the data for the process LCA scale have been obtained from the LCI database provided in Table 1. The flows of P_R and P_{T1} are extracted out from economy scale and included in the process LCA scale as seen from the numbers 8.57 and 4.3 in the last column. The sixth column which belongs to the activity R is consuming 13.6 monetary units of P_D and 5.1 monetary units of P_{T2} from the respective sectors in the economy scale; 13.6 is obtained by multiplying flow of 3.4 units of P_D from Table 1 with the price (4$/unit), and 5.1 is obtained similarly by multiplying 1.7, P_{T2} from Table 1 with price information (3$/unit). The interventions and final demand vectors are also redefined as:

$$\underline{\overline{B_3}} = \begin{bmatrix} 3 & 0.8285 & 0.763 & 0.6 & 27.84 & 96 & 183.48 \end{bmatrix} \tag{42}$$

The final demand from this model is defined as:

$$\underline{\overline{F_3}} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 49 \end{bmatrix} \tag{43}$$

In the seventh row, 49 is present because the process P_C is modeled in the seventh row and column of the technology matrix $\underline{\overline{X_3}}$. Total environmental impact $\underline{\overline{g_3}}$ for this is calculated using Eq. 3 as 509.30. The model is shown as model 3 in Fig. 7. R and T are modeled in the process LCA scale in this model. Complexity of the model is calculated using Eq. 23. RSD is calculated using Eq. 32.

$$RSD_3 = \frac{131.49}{509.30} = 0.26 \tag{44}$$

M_c is calculated as:

$$M_{c,3} = 0.56 + 1/7 + 1/7 = 0.84 \tag{45}$$

With the model complexity value of the LCA model from the previous LCA step, we add the complexity of the two new activities being added in this iteration. As explained in the LCI building exercise in Section 3 in the SI, the LCI database was created by aggregating 7 different systems. Thus, for the two new activities, we add two terms to obtain model complexity of this iteration.

Similarly, the last two pathways of Table 2 are added to the LCA model. For the fourth pathway, D requires to be disaggregated from the economy scale and modeled in the process LCA scale using the LCI information in Table 1. For the last pathway, the operation will be a little different. T has already been modeled as a separate activity in the third iteration. Thus, it is not required to be disaggregated again. Thus, all that is required is to shift the specific flow from T to R from the economy scale to within the process LCA scale. These changes can be clearly seen in the final multiscale technology matrix obtained as:

$$\underline{\overline{X}}_4 = \begin{bmatrix} 1 & 0 & -0.095 & 0 & 0 & 0 & 0 & 0 \\ -0.30 & 1 & -0.071 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ -0.80 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 3.40 & 0 & -3.40 & 0 \\ 0 & 0 & 0 & 0 & 0 & 6 & -1.70 & -4.30 \\ 0 & 0 & 0 & 0 & 0 & 0 & 8.57 & -8.57 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 30 \end{bmatrix} \tag{46}$$

The first change in this new matrix from the previous model $\underline{\overline{X_3}}$ is the addition of the activity D in the fifth column and row. The other major change is that the monetary flows of 5.1 and 13.6 from the economic sectors to the activity R have been replaced with physical flows of 1.7 and 3.4 from the process LCA activities of T and D. Thus, at this final step, we have completed modeling all the steps as seen from the SPA results in Table 2. For this iteration, environmental impact is obtained as:

$$\underline{\overline{g_4}} = \underline{\overline{B}}_4 \underline{\overline{X}}_4^{-1} \underline{\overline{F}}_4 = 527.04 \tag{47}$$

RSD is calculated using Eq. 32.

$$RSD_4 = \frac{6.27}{527.04} = 0.012 \tag{48}$$

The model complexity is calculated in the same way as the previous iterations. One new activity has been added in this LCA model. Thus, we obtain the final model complexity as:

$$M_{c,4} = 0.84 + 1/7 = 0.98 \tag{49}$$

This is the best possible model as the upper limit on M_c does not allow any other shifts. Along with that, all the pathways as listed in Table 2 have been included at the best possible data scale. The final model is shown in Fig. 7. The white cutouts depict the disaggregation of processes from their respective economic sectors. The following important observations are obtained from Fig. 8 :

- As we already know the entire system, detailed information of the activities, and solved the system for numerous cases using the governing fundamental equations as described in the SI, the *true* environmental impact for 49 kg of P_C can be calculated easily. On calculation, this value is found to be 627 kg. However, the final model determined by this algorithm provides an environmental impact of 527 kg. This difference is due to multiple factors. First, the LCI information is built by averaging a nonlinear industrial process as seen from the governing equations 27 to 32 in the SI. Thus, on averaging, the values obtained for production of 30 kg of P_C and listed in Table 1 are actually incorrect. The solution to the governing equations in the SI for the system for 30 kg of P_C and its corresponding environmental impact will be very different from the values listed in Table 1. Secondly, when the LCI information is used in the LCA model and scaled to calculate the life cycle emissions for 49 kg of C, the results digress more from the true value. This is because the scaling occurs linearly whereas the system description in SI, particularly the environmental impact equations, are nonlinear. However, this does not mean that the algorithm fails. In fact, on the left plot in Fig. 8, we see that the emissions slowly push toward the "true" environmental impact of 627 kg line with successive model iterations. The difference exists due to the *erroneous* averaged data supplied by the life cycle inventory.

- As detailed and better information are added successively to the models, the RSD decreased progressively as seen in Fig. 8 due to the use of better quality data and the final life cycle model is generated. The algorithm stops when parameter of uncertainty or complexity violates their constraints or all possible activities have been included in the life cycle model.

The model design framework described in this work has been written in Matlab and can be tailored easily to suit the requirements of the user. Copy of the code, uncertainty data, and support for using the model can be obtained by contacting the author or from associated Mendeley dataset by Ghosh and Bakshi (2020).

4 Conclusions

So far, building life cycle models has mostly been a case-specific exercise based on ISO 14044 and 14044 guidelines. The work described here provides a quantitative procedure for building hybrid life cycle models based on their uncertainty and complexity. Using this model, the LCA practitioner does not have to make decisions for including activities arbitrarily or based on subjective criteria. In fact, with this algorithm, the effect of including or excluding activities can be quantified and compared in terms of the effect on parametric uncertainty and network complexity. As observed from the illustrative example, the algorithm can reduce the subjectiveness involved while building a life cycle model. Moreover, with this method, the improvement in uncertainty obtained while shifting activities between scales can be quantified. This helps build greater confidence in the results obtained from a life cycle model. While it is currently challenging to obtain uncertainty data, complexity data, and other required information for applying this algorithm, it is hoped that this work will encourage further development of such information.

However, this framework does have some inadequacies that will be addressed in future work:

- The choice of complexity and cost parameters that guide the algorithm is still subjective and based on user preferences. Two identical studies could make different choices on these criteria and build completely different LCA models.
- The proposed approach includes only input data or parameter uncertainty and translates those into errors in the result. Model error and scenario uncertainties need to be considered as well.
- Obtaining uncertainty data can be a challenge, especially for input–output and life cycle models and the outcome of the algorithm depends on the availability of such data.
- System boundary determination, specifically the outer boundary in Fig. 1, is still a matter of great subjectivity and not addressed by this paper. We hope to employ model and scenario-based uncertainty parameters for providing a solution to this problem.
- The presented algorithm initializes with the models with the lowest complexity and highest uncertainty and successively replaces the uncertain data with more detailed data to improve the model. This operation also increases the model complexity. For hybrid LCA models, this approach seems to be appropriate as we start with the all-encompassing crude IO model and developing a multiscale model by including more accurate data at finer scales for activities that represent emission hotspots. Developing a similar approach for process-based LCA is currently part of on-going research.

Funding Partial financial support was provided by the National Science Foundation (CBET-1804943) and the Sustainable and Resilient Economy Program at The Ohio State University.

References

Benetto E, Dujet C, Rousseaux P (2008) Integrating fuzzy multicriteria analysis and uncertainty evaluation in life cycle assessment. J Environ Manag 23(12):1461–1467

Bilec M, Ries R, Matthews HS, Sharrard AL (2006) Example of a hybridlife-cycle assessment of construction processes. J Infrastruct Syst 12(4):207–215

Bureau of Economic Analysis (2015) US IO 2007 Economy model kernel description. https://www.bea.gov/index.htm. Accessed: 2015-09-30

Chevalier J-L, Téno J-F (1996) COis Le Life cycle analysis with ill-defined data and its application to building products. Int J Life Cycle Assess 1(2):90–96

Ciroth A (2002) Error calculation in life cycle assessments. Int J Life Cycle Assess 7(5):310–310

Ciroth A, Muller S, Weidema B, Lesage P (2016) Empirically based uncertainty factors for the pedigree matrix in ecoinvent. Int J Life Cycle Assess 21(9):1338–1348

Ecoinvent Database (2015) http://www.ecoinvent.org/home.html. Accessed: 2016-09-30

Everitt BS (2006) The Cambridge dictionary of statistics. Cambridge University Press, Cambridge

Finnveden G (2000) On the limitations of life cycle assessment and environmental systems analysis tools in general. Int J Life Cycle Assess 5(4):229–238

Finnveden G, Hauschild MZ, Ekvall T, Guinée J, Heijungs R, Hellweg S, Koehler A, Pennington D, Suh S (2009) Recent developments in life cycle assessment. J Environ Manag 91(1):1–21

Ghosh T, Bakshi BR (2020) Designing hybrid life cycle assessment models based on uncertainty and complexity. Mendeley Data v1. https://doi.org/10.17632/kjzzk5v932.1

Ghosh T, Bakshi BR (2019) Designing biofuel supply chains while mitigating harmful algal blooms with treatment wetlands. Comput Chem Eng 126:113–127

Graedel TE, Graedel TE (1998) Streamlined life-cycle assessment. Prentice Hall, Upper Saddle River

Groen EA, Heijungs R, Bokkers EAM, de Boer IJM (2014) Methods for uncertainty propagation in life cycle assessment. Environ Modell Softw 62:316–325

Haes HAU, Heijungs R, Suh S, Huppes G (2004) Three strategies to overcome the limitations of life-cycle assessment. J Ind Ecol 8(3):19–32

Hanes RJ, Bakshi BR (2015) Process to planet: a multiscale modeling framework toward sustainable engineering. AIChE J 61(10):3332–3352

Heijungs R (2010) Sensitivity coefficients for matrix-based LCA. Int J Life Cycle Assess 15(5):511–520

Heijungs R, Huijbregts MAJ (2004) A review of approaches to treat uncertainty in LCA. In: Proceedings of the IEMSS conference, Osnabruck

Heijungs R, Lenzen M (2014) Error propagation methods for LCA – a comparison. Int J Life Cycle Assess 19(7):1445–1461

Heijungs R, Suh S (2002) The basic model for inventory analysis. In: The computational structure of life cycle assessment. Springer, pp 11–31

Heijungs R, Suh S, Kleijn R (2005) Numerical approaches to life cycle interpretation-the case of the ecoinvent '96 database (10 pp). Int J Life Cycle Assess 10(2):103–112

Hendrickson CT, Lave LB, Matthews HS (2006) Environmental life cycle assessment of goods and services: an input-output approach. In: Resources for the Future

Hondo H, Sakai S (2002) Consistent method for system boundary definition in LCA. J Adv Sci 13(3):491–494

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(5):499–510

Hou D, Al-Tabbaa A, Guthrie P, Hellings J, Gu Q (2014) Using a hybrid LCA method to evaluate the sustainability of sediment remediation at the london olympic park. J Clean Prod 83:87–95

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(4):485–492

Joshi S (1999) Product environmental life-cycle assessment using input-output techniques. J Ind Ecol 3(2-3):95–120

Ku HH (1966) Notes on the use of propagation of error formulas. J Res Natl Bur Stand

Lenzen M (2000) Errors in conventional and input-output—based life—cycle inventories. J Ind Ecol 4(4):127–148

Lenzen M, Crawford R (2009) The path exchange method for hybrid LCA. Environ Sci Technol 43(21):8251–8256

Lenzen M, Moran D, Kanemoto K, Geschke A (2013) Building EORA: a global multi-region input–output database at high country and sector resolution. Econ Syst Res 25(1):20–49

Lloyd SM, Ries R (2007) Characterizing, propagating, and analyzing uncertainty in life-cycle assessment: a survey of quantitative approaches. J Ind Ecol 11(1):161–179

Morgan MG, Henrion M, Small M (1992) Uncertainty: a guide to dealing with uncertainty in quantitative risk and policy analysis. Cambridge University Press, Cambridge

Moriguchi Y, Kondo Y, Shimizu H (1993) Analysing the life cycle impacts of cars: the case of CO2. Ind Environ

Pairotti MB, Cerutti AK, Martini F, Vesce E, Padovan D, Beltramo R (2015) Energy consumption and GHG emission of the mediterranean diet: a systemic assessment using a hybrid LCA-IO method. J Clean Prod 103:507–516

Reap J, Roman F, Duncan S, Bras B (2008) A survey of unresolved problems in life cycle assessment. Int J Life Cycle Assess 13(5):374

Ross S, Evans D, Webber M (2002) How LCA studies deal with uncertainty. Int J Life Cycle Assess 7(1):47

Rowley HV, Lundie S, Peters GM (2009) A hybrid life cycle assessment model for comparison with conventional methodologies in Australia. Int J Life Cycle Assess 14(6):508–516

Samaras C, Meisterling K (2008) Life cycle assessment of greenhouse gas emissions from plug-in hybrid vehicles: implications for policy. Environ Sci Technol 42(9):3170–3176

Searchinger T, Heimlich R, Houghton RA, Dong F, Elobeid A, Fabiosa J, Tokgoz S, Hayes D, Yu T-H (2008) Use of US croplands for biofuels increases greenhouse gases through emissions from land-use change. Science 319(5867):1238–1240

Sharrard AL, Matthews HS, Ries R (2008) Estimating construction project environmental effects using an input-output-based hybrid life-cycle assessment model. J Infrastruct Syst 14(4):327–336

Stephan A, Crawford R, Bontinck P-A (2018) A model for streamlining and automating path exchange hybrid life cycle assessment. Int J Life Cycle Assess 1–16

Stokes J, Horvath A (2006) Life cycle energy assessment of alternative water supply systems. Int J Life Cycle Assess 11(5):335–343

Suh S (2004) Functions, commodities and environmental impacts in an ecological–economic model. Ecol Econ 48(4):451–467

Suh S, Huppes G (2005) Methods for life cycle inventory of a product. J Clean Prod 13(7):687–697

Suh S, Lenzen M, Treloar GJ, Hondo H, Horvath A, Huppes G, Jolliet O, Klann U, Krewitt W, Moriguchi Y et al (2004) System boundary selection in life-cycle inventories using hybrid approaches. Environ Sci Technol 38(3):657–664

Treloar GJ, Love PED, Crawford RH (2004) Hybrid life-cycle inventory for road construction and use. J Constr Eng Manag 130(1):43–49

Treloar GJ, Love PeterED, Holt GD (2001) Using national input/output data for embodied energy analysis of individual residential buildings. Constr Manag Econ 19(1):49–61

Wiedmann TO, Suh S, Feng K, Lenzen M, Acquaye A, Scott K, Barrett JR (2011) Application of hybrid life cycle approaches to emerging energy technologies–the case of wind power in the UK. Environ Sci Technol 45(13):5900–5907

Yang Y, Ingwersen WW, Hawkins TR, Srocka M, Meyer DE (2017) Useeio: A new and transparent United States environmentally-extended input-output model. J Clean Prod 158:308–318