Implementation of uncertainty analysis and moment-independent global sensitivity analysis for full-scale life cycle assessment models

Cucurachi, S.; Blanco Rocha, C.F.; Steubing, B.R.P.; Heijungs, R.

DOI: 10.1111/jiec.13194

METHODS, TOOLS, DATA, AND SOFTWARE

JOURNAL OF INDUSTRIAL ECOLOGY WILEY

Implementation of uncertainty analysis and moment-independent global sensitivity analysis for full-scale life cycle assessment models

Stefano Cucurachi[1] | Carlos Felipe Blanco[1] | Bernhard Steubing[1] | Reinout Heijungs[1,2]

[1] Institute of Environmental Sciences (CML), Leiden University, Leiden, The Netherlands

[2] Department of Operations Analytics, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands

Correspondence
Stefano Cucurachi, Institute of Environmental Sciences (CML), Leiden University, Einsteinweg 2, 2333 CC Leiden, The Netherlands.
Email: s.cucurachi@cml.leidenuniv.nl

Funding information
Stefano Cucurachi and Carlos Felipe Blanco received funding from the European Union's Horizon 2020 research and innovation program within the project SiTaSol under grant agreement No 727497.

Editor Managing Review: Manfred Lenzen

Abstract

Life cycle assessment (LCA) models and databases have increased in size, resolution, and complexity, requiring analysts to rely on an ever-increasing number of uncertain model inputs. Such increased complexity calls for systematic approaches to assessing the uncertainty of the output results of LCA models and the sensitivity of LCA model outputs to the model's uncertain inputs. In this contribution, we provide a theoretical basis and present a practical software implementation that combines uncertainty analysis and moment-independent global sensitivity analysis, which can be readily applied to full-scale LCA models. We implemented our approach in the Activity-Browser open source LCA software and it is made available for use in LCA studies. We demonstrate the approach and software implementation with a case study of crystalline silicon photovoltaics.

KEYWORDS
global sensitivity analysis, industrial ecology, life cycle assessment, modeling, open-access software, uncertainty analysis

1 | INTRODUCTION

The complexity of models and the uncertainty of inputs require life cycle assessment (LCA) analysts to understand the impact of all potential sources of uncertainties on the model output if the model is to be used effectively and responsibly in any decision-making process. Various authors recommend the application of a combination of uncertainty and sensitivity analyses to assure the quality of any mathematical model that supports decision-making (Borgonovo et al., 2012; Risbey et al., 2005; Saltelli et al., 2019). In the context of LCA, ISO 14044, for instance, prescribes that the interpretation of LCA studies "shall include an assessment and a sensitivity check of the significant inputs, outputs, and methodological choices to understand the uncertainty of the results" (see clause 4.5.1.1; (ISO, 2006)).

In this paper, we will be adopting the following definitions. Uncertainty analysis (UA) is the quantification and propagation of input uncertainties to output uncertainties (Igos et al., 2019a). In LCA, uncertainty can be related to input data that are uncertain (for instance, the effects of a pesticide, such as chlorpyrifos, on human health are not fully known), to input data that are variable (for instance, the lifetime of two identical PV panels is not exactly equal), and to choices that must be made by the analyst (for instance, related to the choice of mass-based or economic allocation); see

Heijungs (2020). Besides, data may be missing and therefore estimated, introducing new uncertainties. Two types of sensitivity analysis can be distinguished: local and global (Razavi et al., 2021). Local approaches to sensitivity (LSA) refer to the variation of one input of an uncertain model around its reference point, keeping the others at their nominal values (Igos et al., 2019a). Conversely, Saltelli et al. (2008) define global sensitivity analysis (GSA) as the study of how the uncertainty in the output of a mathematical model can be apportioned to different sources of uncertainty of its inputs. We focus on GSA, which limits the risk of not fully understanding the behavior of a model on which an inference is based, that is, the so-called black-box effect (Borgonovo et al., 2012), and which can allow the analyst to best understand how uncertain inputs of a mathematical model interact, thus allowing drawing out the maximum capabilities of mathematical modelling (Rabitz, 1989).

UA and GSA are not new to the LCA community. A series of methodological contributions have formalized UA in LCA, although sometimes using different terminology (Heijungs, 1994; Lloyd & Ries, 2007; Heijungs & Lenzen, 2014; Groen et al., 2014; Huijbregts, 1998; Mutel et al., 2013; Igos et al., 2019a). Current and common standard practice for UA consists of applying computational algorithms that rely on repeated sampling using the Monte Carlo method to obtain numerical results for LCA models (Groen et al., 2014). The output of UA for LCA models is typically a distribution for the inventory results (e.g., emissions of CO_2 for gas-powered electricity production; see, e.g., Scherer & Pfister, 2016; von Pfingsten et al., 2017; Gavankar et al., 2015), or a distribution for the impact scores (e.g., kg of CO_2 equivalent for gas-powered electricity production; see e.g., Golsteijn et al., 2012; Douziech et al., 2019). Some authors concentrate on uncertainty in a comparative setting; see for example, Huijbregts (2003); Heijungs & Kleijn (2001); Mendoza Beltran et al. (2018); Heijungs (2021).

Proposals for GSA are scattered. The most common approaches to GSA combine probabilistic UA based on Monte Carlo simulations with GSA techniques, such as Sobol' indices, which allow to numerically apportion the variance of the output of any computational model to individual uncertain inputs (Sobol', 1993). In LCA, examples of the use of the variance-based approach of Sobol' include the work of Geisler et al. (2005), Mutel et al. (2013), Bisinella et al. (2016), and Lacirignola et al. (2017). We further refer the reader to the work of Groen et al. (2017), Igos et al. (2019b), and Michiels and Geeraerd (2020) for a review of GSA methods applied to LCA.

The above brief review of approaches shows that disparate proposals for UA and GSA in LCA exist, and yet no consensus seems to have emerged. Existing techniques have limitations that are computational, but also methodological. Also, most proposals have not been implemented in commercially and freely available software for LCA, thus further hampering the regular application of such techniques in LCA studies. Furthermore, no proposal to date provides a joint assessment at the level of a full LCA model, for example, including the impact assessment phase.

To improve on existing proposals and to support the alignment of methods, with this contribution we propose (i) an implementation of UA and GSA that suits the needs of standard LCA practice.

With this contribution we also aim to (ii) align for the first time the theory and notation of UA and GSA techniques to the computational structure of LCA, allowing for a streamlined collaboration across communities in the future (see Sections 2.2 and 4.3). We also translate theory and methods into (iii) an open-source LCA software developed in the Activity Browser (Steubing et al., 2020) that allows practitioners to directly apply the proposed implementation in LCA studies. We also provide solutions (e.g., contribution-based filtering; see Section 2.3) to existing issues with UA and GSA software tools, which are computationally demanding, and inadequate to assess full-scale real LCA models. Finally, we (iv) demonstrate the application of the theories, methods, and software in a realistic case study of crystalline silicon photovoltaics (see Section 3), which showcases the set of results and interpretation that can be obtained by applying the proposed approach.

2 | METHODS

2.1 | Choice of methods for uncertainty and global sensitivity analysis in LCA

We consider suitable for LCA the GSA methods that have the following properties identified by Saltelli et al. (2019, 2008):

- coping with the influence of scale and shape. The influence of the input should incorporate the effect of the range of input variation and the shape of its probability density function. This applies to the uncertain inputs of LCA models.
- including a multidimensional analysis. The chosen GSA method should evaluate the effect of a factor while all others are also varying, thus allowing to appreciate global effects in LCA models.
- being applicable to assess a full LCA model, independently of complexity. The method should work for a broad range of LCA models, also in the case of highly-interactive, non-linear, and non-additive models. This is the case of LCA models, in which factors interact with one another, and the effect of changing two factors is different from the sum of their individual effects.

Moment-independent (or density-based) methods are a class of sensitivity measures with the above properties. These measures consider the entire distribution without reference to a particular moment, such as the mean or the variance. Borgonovo and Plischke (2016) demonstrated that moment-independent measures: (1) are suitable to assess correlated inputs; (2) do not depend on a particular moment of the model output distribution, and 3) are equal to zero if and only if the output of a model is independent of a specific input (see also Baucells & Borgonovo, 2013). Within the

moment-independent class of methods, we select the so-called δ-measure (Borgonovo & Apostolakis, 2001). The interpretation of the δ-measure is unaffected by the presence of correlations, which affect, for instance, variance-based sensitivity indices, which do not hold in the presence of dependences (Borgonovo & Plischke, 2016). By looking at the shift of the actual output distribution, the approach intrinsically accommodates correlations that exist in LCA models among dependent processes. Insofar as the source of shared uncertainty in correlated coefficients is represented by a specific variable model input (e.g., the efficiency of a combustion engine), the total effects resulting from interactions among correlated coefficients are accounted for by the δ-measure. We further refer the reader to Plischke and Borgonovo (2020) for a review of these aspects for the δ-measure also in comparison to other GSA measures. Additionally, moment-independent methods can deal with multimodal output distributions, which could result from more complex LCA models (Blanco et al., 2020), and they are computationally more parsimonious, thus more efficiently implementable in LCA software.

2.2 | Theoretical support to the proposed approach to UA and GSA

2.2.1 | NOTATION

We will adopt the following general conventions unless otherwise specified:

- scalars are written as non-bold, italic, letters;
- vectors are written as bold, roman, lowercase letters;
- matrices are written as bold, roman, capitals;
- parameters of probability distributions are written as Greek letters;
- stochastic variables are written as non-bold, italic, capitals;

In the LCA-part, we will connect, where possible, to the standard symbols used in Heijungs and Suh (2002). Further, we will be differentiating between random variables (written in capital) and their realized values (written as lower case).

2.2.2 | DEFINITION OF THE LCA MODEL

Let us consider the deterministic scalar function representing a reliable and fixed specification of the LCA algorithm, such that:

$$\mathbf{h} = \mathbf{QBA}^{-1}\mathbf{f}. \tag{1}$$

In Equation (1), $\mathbf{A}$ represents the square technology matrix of size $m \times m$. The intervention matrix $\mathbf{B}$ of size $p \times m$ reports the use of resources and emissions for each production process in the technology matrix. The vector $\mathbf{f}$ of size m represents the final demand vector, to which production processes are scaled to produce the desired amount of output to fulfill a certain functional unit. $\mathbf{Q}$ of size $l \times p$ represents the characterization matrix, that is, the ensemble of the characterization vectors available for several impact categories (e.g., global warming, toxicity; see Heijungs & Suh, 2002 and Heijungs et al., 2015). $\mathbf{h}$ is the output of the LCA model, that is, the impact score for the l impact categories considered in a LCA model (e.g., acidification, eutrophication, climate change, toxicity). If desired, the impacts may also be defined at endpoint level (human health, ecosystem quality, etc.).

We can indicate with $\mathbf{x}$ the vector of the k inputs to the LCA model introduced in Equation (1):

$$\mathbf{x} = \begin{pmatrix} \mathrm{vec}(\mathbf{A}) \\ \mathrm{vec}(\mathbf{B}) \\ \mathrm{vec}(\mathbf{Q}) \\ \mathbf{f} \end{pmatrix}, \tag{2}$$

where vec(·) is the vectorization operator of the matrix considered. The vector of inputs to an LCA model can also include allocation factors (Jung et al., 2014) and parameters (Mutel et al., 2013; Blanco et al., 2020); for simplicity, we do not include these in this demonstration.

We consider the output of the LCA model $\mathbf{h}$ (e.g., the impact score for climate change and other impacts), and a set of independent input variables $\mathbf{x}$ as:

$$\mathbf{h} = \mathbf{g}(\mathbf{x}), \quad \mathbf{g} : \mathbb{R}^k \to \mathbb{R}^l \tag{3}$$

with **g** representing the relationship between inputs and output, k denoting the size of the input space, that is, $k = \underbrace{m \times m}_{\mathbf{A}} + \underbrace{p \times m}_{\mathbf{B}} + \underbrace{l \times p}_{\mathbf{Q}} + \underbrace{m}_{\mathbf{f}}$, and l denoting the number of the model outputs considered in the LCA model under assessment.

2.2.3 | Uncertainty analysis

We consider that almost all input data to the LCA model are uncertain, that is, that they should be treated as stochastic variables. Moreover, we will treat them as a continuous variable, so that we can use probability density functions instead of discrete probability distribution functions for all inputs. This is of course not completely true (for instance, the number of passengers in a train is discrete), but it is a reasonable approximation. We only assume that we have perfect knowledge of the final demand, thus that the final demand vector accurately represents the demand for the product system under assessment, allowing us to ignore any uncertainty on **f**. We address the variation of the other model inputs across the full range defined by their probability distributions. Thus, we follow Cucurachi et al. (2016) and consider that the vector of model inputs **x** can be represented by a stochastic vector denoted as **X**. We further assume that the elements of **X** are represented by probability distributions. For example, we may consider that a particular element of **X**, for instance X_1, is normally distributed:

$$X_1 \sim N\left(\mu_{X_1}, \sigma_{X_1}\right), \tag{4}$$

and similar specifications of the other elements of **X**. The model output, then, is the stochastic vector **H**, related to **X** through:

$$\mathbf{H} = \mathbf{g}(\mathbf{X}). \tag{5}$$

We consider the full set of stochastic inputs of **X**, which are activated by the demand vector, and for which the following are available: (1) a probability distribution (or density) function; (2) the appropriate parameters to characterize the distribution (for instance a mean and a standard deviation in case of a normal distribution).

We further make the distinction in the system between background and foreground, modifying Equation (2) as:

$$\mathbf{x} = \begin{pmatrix} \text{vec}\left(\mathbf{A}_{fg}\right) \\ \text{vec}\left(\mathbf{A}_{bg}\right) \\ \text{vec}\left(\mathbf{B}_{fg}\right) \\ \text{vec}\left(\mathbf{B}_{bg}\right) \\ \text{vec}\left(\mathbf{Q}\right) \\ \mathbf{f} \end{pmatrix}. \tag{6}$$

We consider pertaining to the foreground all processes that are modeled for a specific system, with specific conditions determined by the modeler, as opposed to the average market activities in the background. We consider that these data are directly collected by the analyst and include realistic uncertainty information, if available. For the background system, we rely on the uncertainty information provided by the ecoinvent database (Wernet et al., 2016). We additionally consider uncertainty information for the LCIA phase (see Section 3.1.2).

We sample the input space, for example, using random Monte Carlo sampling (Hastings, 1970), with sample size N. Following this approach, we proceed with drawing random instances from the specified distribution functions of the relevant uncertain inputs of **X**. For each of the probability distributions defined by the inputs, we obtain N **X**-structures in the k-dimensional input space.

We evaluate the model output in correspondence with each realization of **X**. As a result, we obtain N realizations of the l-dimensional model outputs **H**. More precisely, we can define the outputs of the LCA model under consideration as the stochastic variable:

$$\mathbf{H} \sim ? \tag{7}$$

where ? indicates a distribution that is to be estimated from the N sampled values.

2.2.4 | Global sensitivity analysis

Let us denote by:

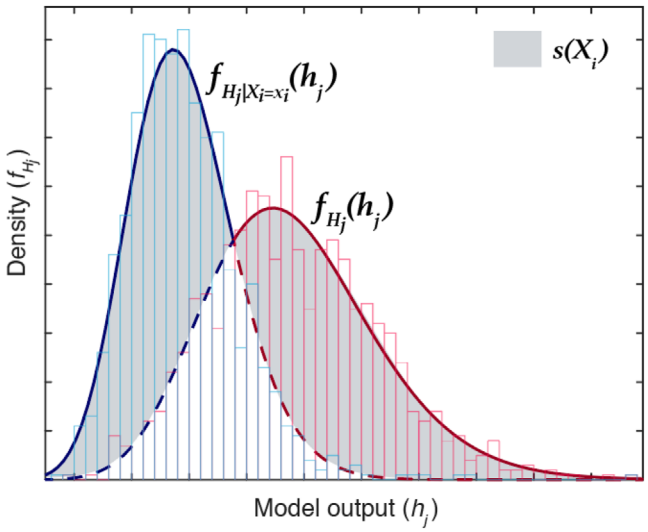

FIGURE 1 The shift $s(X_i)$ in the model output density curve (f_{H_j}) of an LCA impact score H_j is the area not enclosed between the output density curve conditional to a coefficient X_i fixed at value x_i ($f_{H_j|X_i=x}$, shown in blue) and the density curve when X_i is left to vary randomly (f_{H_j}, shown in red). Adapted from Borgonovo (2007)

- $F_{X_i}(x)$ and $F_{H_j}(h)$ the cumulative distributions functions (cdf) of the *i*th input and the *j*th output of the LCA model, respectively;
- and by $f_{X_i}(x)$ and $f_{H_j}(h)$ their probability densities (pdf).

Moment-independent measures of sensitivity follow the intuition that if we let all model inputs free to simultaneously vary per the assigned distributions, we then obtain the unconditional model output density f_{H_j}. The idea behind moment-independent measures is to measure how fixing the input X_i at a fixed value x modifies the entire distribution of the output H_j into a conditional distribution.

By fixing X_i at x_i, we obtain, thus, the conditional density of H_j given that X_i is fixed at x_i, that is, $f_{H|X_i=x_i}(h_j)$ (see Figure 1). Following Borgonovo et al. (2017) and Borgonovo & Plischke (2016), we can measure the shift between $f_{H_j}(h_j)$ and $f_{H_j|X_i=x_i}(h_j)$ as in:

$$s(x_i) = \int_{H_j} \left| f_{H_j}(h_j) - f_{H_j|X_i=x}(h_j) \right| dh_j. \tag{8}$$

The operator $\int |\cdot|$ represents an inner separation between probability distributions, thus measures the area enclosed between the conditional and unconditional model output densities obtained for $X_i = x_i$. The integral is taken over the full support of H_j. Notice that $s(X_i)$ is a stochastic value, because it is dependent on the random variable X_i.

We can further measure the expected value of this shift as follows:

$$E_{X_i}[s(X_i)] = \int f_{X_i}(x_i) \left[\int_{H_j} \left| f_{H_j}(h_j) - f_{H_j|X_i=x}(h_j) \right| \right] d\,x_i = \int f_{X_i}(x_i)\, s(X_i)\, dx_i. \tag{9}$$

We obtain the δ sensitivity measure as proposed by Borgonovo (2007) setting:

$$\delta_i = \frac{1}{2} E_{X_i}[s(X_i)]. \tag{10}$$

Equations (9) and (10) show that δ_i is proportional to the average of the difference between the product of the marginal distributions of H_j and X_i and their joint density. The estimation of δ_i is not related to an explicitly defined model $g(\cdot)$, but only to the quasi-empirical input-output mapping obtained through the Monte Carlo sampling implied by $g(\cdot)$. We refer the reader to Borgonovo (2007) for a description of the properties of the δ measure (see also Section 2.1).

We proceed by estimating δ_i following Plischke et al. (2013), and estimate the measures δ_i considering the input-output mapping in the LCA model $H_j = g(X_i)$. Here, we focus on using the dataset containing the N realizations of (X_i, H_j), calculated using a Monte Carlo simulation as described in Section 2.2.3. By this sample, it is possible to obtain the empirical pdfs $f_{H_j}(h)$ and $f_{H_j|X_i=x}$. We refer the reader to the work of Plischke and co-authors (Plischke et al., 2013) for further details on the estimation of δ_i, and to Section 2.3 for the implementation of UA and GSA in LCA software.

2.3 | Software implementation

We implemented the proposed joint approach to UA and GSA in Python, building upon the sensitivity analysis library SALib (Herman & Usher, 2017), and the brightway LCA framework (Mutel, 2017). Our implementation is available at the following repository: github.com/bsteubing/lca-global-sensitivity-analysis. Additionally, we implemented the proposed approach in the Activity Browser open source LCA software (Steubing et al., 2020; available from github.com/LCA-ActivityBrowser/activity-browser)[1].

The general steps of the software implementation are the following:

1. Monte Carlo simulation;
2. Filtering of input data to GSA (see Section 2.3.1);
3. Formulation of the input data for GSA as in Equation (6);
4. GSA and display of results.

It may be helpful to further specify the nature of **A** and **B** in the context of the software implementation. As described, **B** contains what is often referred to as "environmental flows" or "interventions" and describes the flows between human activities and the environment, which can thus be seen as "causers" of environmental impact. **A** contains what is often referred to as "economic exchanges" or "intermediate flows" and describes the flows between human activities, which can be thus be seen as "connectors" between the functional unit and the environmental exchanges in **B** (Reinhard et al., 2016). GSA allows identifying those elements that are associated with high uncertainties from the UA and either cause important environmental impacts within a product system (e.g., a high, but uncertain, emission in steel production) or connect these to the functional unit.

We perform a Monte Carlo simulation for a reference flow as described in Section 2.4, using a random Monte Carlo sampling design with 1000 runs (see Heijungs, 2020). The number of MC runs can be adjusted by the user in the Activity Browser software. Alternative and more efficient sampling strategies (e.g., quasi-Monte Carlo or Latin Hypercube) can also be considered, without major changes in the software. The computational efficiency of sampling strategies is outside of the scope of this paper, and we further refer the reader to Groen et al. (2014) for a review of approaches that can be readily implemented in combination with our proposal using existing software packages. The LCA case study we use in this contribution is non-comparative. Should the LCA analyst be comparing two alternatives, the software implementation considers a dependent sampling approach (Henriksson et al., 2015; Mendoza Beltran et al., 2018), thus ensuring that all systems under comparison are assessed using the same **A**, **B**, and **Q** matrices.

2.3.1 | Implementation of variable-filtering

When applying traditional approaches to GSA in practice with LCA, a major challenge for both computation and interpretation is the size of the vector of inputs **x**. In the ecoinvent 3.6 (cutoff) database (Wernet et al., 2016), which was used here, a typical product system is described by more than 200,000 elements in **A**, more than 400,000 elements of **B**, as well as thousands of elements of **Q**. Yet, the vast majority of these elements have a negligible influence on the environmental impacts of a specific product system. Feeding such large numbers of variables to GSA would lead to very long computation times and potentially limiting memory requirements, thus requiring some form of filtering (see also Blanco et al. (2020).

In the software implementation here proposed, we tackle these challenges by applying *contribution-based filtering*. Contribution-based filtering exploits the fact that most elements in **A** and **B** have a negligible influence on the environmental impact score. A cut-off value could thus be defined to exclude those elements that do not cause contributions to an environmental impact score (H_j) below a certain threshold; directly in the case of **B** and indirectly in the case of **A** (see below). In most cases, this approach should reduce the number of inputs $\mathbf{X_i}$ sufficiently to well overcome limitations posed by the computational intensity of the algorithms.

For environmental flows, contribution-filtering is applied as follows. First, we calculate the scaling factors **s** for the static product system (no uncertainty data used), as in:

$$\mathbf{s} = \mathbf{A}^{-1}\mathbf{f}. \tag{11}$$

Then, we calculate the scaled intervention matrix $\tilde{\mathbf{B}}$ by

$$\tilde{\mathbf{B}} = \mathbf{B}\mathrm{diag}\left(\mathbf{s}\right), \tag{12}$$

[1] A user manual of the implementation in the Activity Browser is available at: github.com/LCA-ActivityBrowser/activity-browser/wiki/Global-Sensitivity-Analysis
[Corrections added on Sep 28, 2021 after first online publication: Equation 12 was corrected.]

such that the row sum of $\tilde{\mathbf{B}}$ yields the inventory result **g**:

$$\mathbf{g} = \tilde{\mathbf{B}}\mathbf{1}, \tag{13}$$

where **1** is a vector of ones. Then, we characterize the environmental flows by multiplying each element of $\tilde{\mathbf{B}}$ with its corresponding characterization factor for a given impact category. We first select the row from **Q** that contains the characterization factor for a given impact category, *k*, by selecting the *k*th row of **Q**. This row vector is indicated by $\mathbf{q'}_k$, where the prime indicates transposition, turning the column vector $\mathbf{q}_k$ into a row vector. As a result, we have

$$\mathbf{Q} = \begin{pmatrix} \mathbf{q'}_1 \\ \mathbf{q'}_2 \\ \cdots \end{pmatrix}. \tag{14}$$

Next, we create:

$$\mathbf{H}_k = \text{diag}\left(\mathbf{q'}_k\right)\tilde{\mathbf{B}} \tag{15}$$

$\mathbf{H}_k$ is a matrix that represents the characterization results split by environmental flows (rows) and processes (columns). The sum of all elements of $\mathbf{H}_k$ is, thus, equal to the *k*th element of the impact vector **h** in Equation (1):

$$h_k = 1'\mathbf{H}_k\mathbf{1}. \tag{16}$$

The cut-off value, *c*, is then applied to identify the largest elements of $\mathbf{H}_k$, that is, those elements that contribute to the specific impact category by more than the set cut-off value. More precisely, we identify the elements in row *i* and column *j* of $\mathbf{H}_k$, such that their contribution to the total impact, h_k, is more than the specified value, *c*. So, we find all (*i*, *j*) such that

$$\frac{(\mathbf{H}_k)_{i,j}}{h_k} > c. \tag{17}$$

We repeat this procedure for every impact category, *k*. If for any *k*, the above condition holds, the entry (*i*, *j*) of **H** and therefore of **B** is maintained. Otherwise, that entry of **B** is removed (or changed into a zero) before being an entry into Equation (6). More formally, we replace **B** in Equation (6) by a matrix $\mathbf{B}^*$, such that

$$(\mathbf{B}^*)_{i,j} = \begin{cases} (\mathbf{B})_{i,j} & \text{if for any } k: \frac{(\mathbf{H}_k)_{i,j}}{h_k} > c \\ 0 & \text{otherwise} \end{cases}, \tag{18}$$

and pass $\mathbf{B}^*$ to Equation (6). For a cut-off value $c = 0.001$, the number of elements in **B** is typically reduced to a few dozens to hundreds.

For elements in **A**, the objective is to identify the most important "connectors", that is, those economic flows that connect the environmental flows throughout the production system to the reference flow (Reinhard et al., 2016). We used the graph traversal (Rodriguez & Neubauer, 2012) function of brightway (Mutel, 2020)[2] to identify all activities that cause or transmit environmental impacts above a user-defined cut-off value. The graph traversal approach follows a "contribution-first" logic, that is, the supply chain is traversed starting from the reference flow based on the magnitude of environmental impact that is either connected through or caused by supply chain activities. Note that this is the same approach that can be applied to generate a Sankey diagram to show the cumulative contribution of processes across supply chains for a given reference product. By including only those activities that contribute or connect environmental impacts above a certain cut-off value, the graph traversal can be used to filter economic flows. For a cut-off value of 0.001 of the total impact score, the input to the GSA in terms of number of economic exchanges can be typically reduced to a few hundreds.

Finally, elements in **A**, **B** or **Q** that have either a zero value or no uncertainty information, are filtered out as well. For example, for the impact category of global warming, only several hundred elements of **q** are non-zero, only a few have uncertainty information, while the total number of columns in **Q** is several thousand.

[2] A description of the function is available here: https://2.docs.brightway.dev/lca.html#graph-traversal

3 | CASE STUDY

3.1 | Description of the case study

For illustration, we applied the proposed approach to investigate the environmental implications of recent process optimizations in the supply chain of crystalline silicon photovoltaic (PV) installations. The studied product system, which we henceforth refer to as updated PV, is largely based on a slanted roof-mounted mono-crystalline (single-Si) PV installation, as modeled in the LCA database ecoinvent v3.6 (cut-off version (Wernet et al., 2016)). However, the data in ecoinvent represents PV installations manufactured before the year 2010. In the updated PV system, the amounts of materials and energy consumed by several upstream processes as well as PV module performance are adjusted to reflect more improvements reported by various authors (Stamford & Azapagic, 2018; Lunardi et al., 2018; Woodhouse et al., 2020). We do not know with precision to what extent these improvements have been implemented and in several cases ranges rather than precise values have been reported. Therefore, we reflected these uncertainties in the corresponding coefficients in **A** and propagated them by running 1000 Monte Carlo simulations. We then used GSA to understand the influence of these uncertainties on the distribution of the updated PV system's LCA impact scores **H**.

The key components of the updated PV system are the panels, which contain the cells within an aluminum and glass frame, an aluminum mounting system used to attach the panels to the roof, an AC/DC inverter, and an electrical installation which mostly consists of cabling. The cells are based on a silicon wafer, which is sliced from mono-crystalline silicon ingots grown via the energy-intensive Czochralski process (CZ). We defined the functional unit of the updated PV system as 1 kWh of electricity and focused on the assessment of the climate change and freshwater ecotoxicity impacts using the IPCC-2013 GWP100 (IPCC, 2014), and USEtox (Rosenbaum et al., 2008) impact assessment methods, respectively.

3.1.1 | Uncertain coefficients in the background and foreground system

For the foreground system, the coefficients we used to represent optimizations in the updated PV supply chain are presented in Table 1. For the background system, we used the uncertainty parameters described in the ecoinvent v3.6 database.

3.1.2 | Uncertain characterization factors

The models behind characterization factors may be prone to large uncertainties, given that the conditions for transport, exposure, and effects of substances can vary considerably across time and space, as well as across species. Furthermore, characterization methods are subject to model uncertainties such as choice of system boundaries and impact pathways or mechanisms. Even though the importance of uncertainty has been recognized for impact assessment models (see, e.g., Cucurachi et al., 2016), impact assessment models do not typically come with uncertainty information and commercial LCA database do not include uncertainty information for CFs. We investigated the potential influence of uncertainties in **Q** in comparison to those in **A** and **B** by introducing uncertainty in the global warming potential of CH_4 following Boucher (2012), who proposed a normal distribution with a standard deviation of 2.8. For ecotoxicity, USEtox suggests a lognormal distribution with a squared geometric standard deviation of 18 and 176 to represent model uncertainties in freshwater and rural air CFs, respectively (Rosenbaum et al., 2008). We applied these uncertainties to the highest contributors to the freshwater ecotoxicity impact category, namely the CFs of chromium (to surface water), silver (to urban air), copper (to non-urban air or from the high stack), arsenic (to surface water), and zinc (to groundwater). In the Section 4, we further explore how the current exploration can be extended to the entire **Q** matrix once more comprehensive CF uncertainty information becomes available, and how this would affect the GSA.

3.2 | Case study results

3.2.1 | Climate change

Three of the uncertain coefficients in the foreground resulted as the most sensitive (i.e., having the highest δ), while the rest had a small influence on the climate change impact score distribution ($\delta \leq 0.05$). The δ-values for these coefficients are shown in Table 2 (we refer the reader to the Supporting Information for complete results for the case study).

The most sensitive coefficient in the updated PV system is coefficient #15, associated with the electricity output from the updated PV system. This coefficient, which depends on the solar panel's conversion efficiency, triggers all other processes in the PV supply chain including ancillary infrastructures such as inverter, cabling, and mounting structures. Therefore, this coefficient is influencing many other coefficients in matrices **A**

TABLE 1 Updated A and B coefficients in the foreground supply chain of electricity production with a single-Si PV slanted roof installation

Coeff ID	Process	Input (I)/Output (O)	Units	Original distribution type (from ecoinvent)	Original distribution parameters (from ecoinvent)	Updated distribution type	Updated parameters	Data sources and assumptions
1	Silicon production, solar grade, Siemens process	I: electricity, high voltage	kWh	Lognormal	Mean: 110 SD: 1	Uniform	Min: 34.1 max: 44.3	Total consumption of 50–65 kWh/kg for electricity and steam (Woodhouse et al., 2020). (Note: we apply this reduction proportionally to electricity and heat inputs in ecoinvent)
2	Silicon production, solar grade, Siemens process	I: heat, industrial natural gas	MJ	Lognormal	Mean: 185 SD: 1	Uniform	Min: 57.24 max: 74.52	Total consumption of 50–65 kWh/kg for electricity and steam (Woodhouse et al., 2020). (Note: we apply this reduction proportionally to electricity and heat inputs in ecoinvent)
3	silicon production, electronic grade, Siemens process	I: electricity, high voltage	kWh	Lognormal	Mean: 168.8 SD: 1.56	Uniform	Min: 48.9 max: 60.2	Total consumption of 65–80 kWh/kg for electricity and steam (Woodhouse et al., 2020). (Note: we apply this reduction proportionally to electricity and heat inputs in ecoinvent)
4	Silicon production, electronic grade, Siemens process	I: heat, industrial natural gas	MJ	Lognormal	Mean: 180 SD: 1.56	Uniform	Min: 52.2 max: 64.1	Total consumption of 65–80 kWh/kg for electricity and steam (Woodhouse et al., 2020). (Note: we apply this reduction proportionally to electricity and heat inputs in ecoinvent)
5	Silicon production, single crystal, Czochralski process	I: electricity, medium voltage	kWh	Lognormal	Mean: 85.6 SD: 1.22	Uniform	Min: 43.4 max: 69.3	NREL estimates 0.7-1.1 kWh per wafer, where 1 kg of polysilicon yields 62–63 wafers. We multiply these quantities to obtain min/max energy consumption per kg of CZ silicon produced (Woodhouse et al., 2020)

(Continues)

TABLE 1 (Continued)

Coeff ID	Process	Input (I)/Output (O)	Units	Original distribution type (from ecoinvent)	Original distribution parameters (from ecoinvent)	Updated distribution type	Updated parameters	Data sources and assumptions
6	Single-Si wafer production	I: silicon, single crystal, Czochralski process	kg	Lognormal	Mean: 1.07 SD: 1	Triangular	Min: 0.40 mode: 0.66 max: 0.75	Silicon wafer thickness has decreased from 270 to 180 μm (Stamford & Azapagic, 2018) and even newer technologies proposing thicknesses as low as 100 μm (Sai et al., 2018). We adjust the masses proportionally for the min/max values. For the mode, we use the typical consumption reported by NREL of 16 g per wafer which gives 0.66 kg per m^2 wafer ().
7	Photovoltaic cell production	I: electricity, medium voltage	kWh	Lognormal	Mean: 30.24 SD: 1	Uniform	Min: 16.39 max: 24.59	Based on the NREL estimate of 0.4–0.6 kWh per 244cm^2 monocrystalline PERC cell (Woodhouse et al., 2020).
8	Photovoltaic cell production	I: metallization paste, front side (silver)	kg	Lognormal	Mean: 7.39E-3 SD: 1	Uniform	Min: 3.28E-3 max: 4.51E-3	Based on NREL estimates for a 244 cm^2 cell, the industry standard is currently around 80–110 mg (Woodhouse et al., 2020).
9	Photovoltaic cell production	I: metallization paste, back side aluminum	kg	Lognormal	Mean: 7.19E-2 SD: 1	Uniform	Min: 5.33E-2 max: 6.15E-2	Based on NREL estimates for a 244 cm^2 cell, the industry standard is currently around 1.3–1.5 g of aluminum paste for the backside (Woodhouse et al., 2020).
10	Photovoltaic cell production	I: metallization paste, back side (silver)	kg	Lognormal	Mean: 4.93E-3 SD: 1	Uniform	Min: 8.2E-4 max: 2.05E-3	Based on NREL estimates for a 244 cm^2 cell, the industry standard is currently around 20–50 mg of silver paste for the backside (Woodhouse et al., 2020).
11	Photovoltaic cell production	O: silver, in air, urban, close to the ground	kg	Lognormal	Mean: 7.73E-4 SD: 5.02	Uniform	Min: 2.46E-04 max: 3.94E-04	Silver emissions are adjusted proportionally to the reductions in silver paste consumption reported by NREL (Woodhouse et al., 2020).

(Continues)

TABLE 1 (Continued)

Coeff ID	Process	Input (I)/Output (O)	Units	Original distribution type (from ecoinvent)	Original distribution parameters (from ecoinvent)	Updated distribution type	Updated parameters	Data sources and assumptions
12	Photovoltaic panel production, single-Si wafer	I: aluminum alloy, AlMg3	kg	Lognormal	Mean: 2.63 SD: 1	Uniform	Min: 1.6 max: 2.0	Thinner frames have been incorporated (Stamford and Azapagic, 2018), (ITRPV, 2013).
13	Photovoltaic panel production, single-Si wafer	I: solar glass, low-iron	kg	Lognormal	Mean: 10.08 SD: 1.22	Uniform	Min: 5.04 max:7.56	Glass thickness has been decreased to reduce weight and cost, the industry now aiming for between 2 and 3 mm, whereas the original ecoinvent amount reflected 4 mm (ITRPV, 2020).
14	Photovoltaic panel production, single-Si wafer	I: electricity, medium voltage	kWh	Lognormal	Mean: 4.71 SD: 1	Uniform	Min: 12.22 max: 15.27	Based on NREL estimates of 20–25 kWh per 60-cell module with an area of $1.650 \times 0.992\ m^2$. Note (we removed the heat input from the original ecoinvent process).
15	Electricity yield from updated PV system	O: electricity, low voltage	kWh	None	Value: 1	Triangular	Min: 1.40 mode: 1.47 max: 1.62	Panel conversion efficiencies have improved from 13.6% to 19% (Lunardi et al., 2018) and up to 22%. We use a conversion efficiency of 20% to determine the mode, adjusting the output proportionally (20/13.6 = 1.47 kWh produced by the more efficient updated PV system). The expected (mean) electricity production would be 1.49 kWh.

TABLE 2 δ-sensitivity measures for climate change impacts of the updated PV system, foreground system only

Coeff ID	Process	Input (I)/ Output (O)	Description	δ estimate
15	Electricity yield from the updated PV system	O: electricity, low voltage	Amount of electricity that can be produced by the updated PV system (yield), which is a result of the uncertain increase in panel conversion efficiency between 19% and 22%.	0.27
6	Single-Si wafer production	I: silicon, single crystal, Czochralski process	Amount of single-Si Czochralski crystal that goes into single-Si wafer production. Varies depending on the uncertain wafer thickness between 100 and 180 μm.	0.21
5	Silicon production, single crystal, Czochralski process	I: electricity, medium voltage	The amount of electricity consumed by the Czochralski process for growing single-Si ingot; varies between 0.7 and 1.1 kWh per wafer according.	0.11

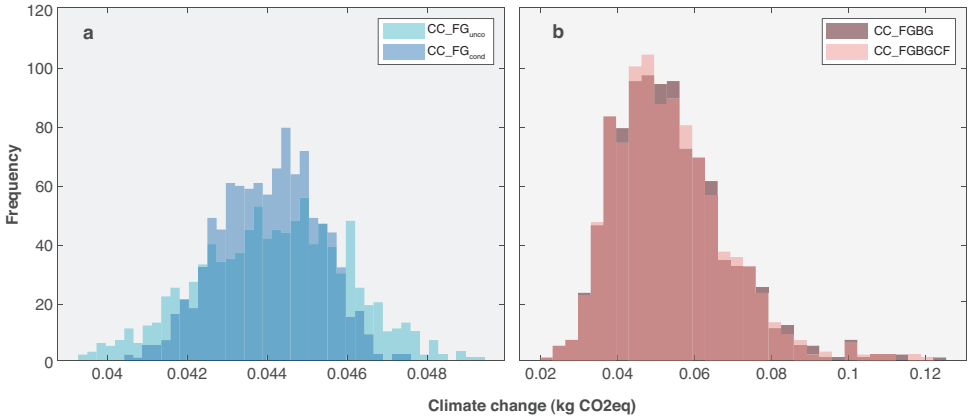

FIGURE 2 (a) Climate change impact score of updated PV system with uncertain foreground, unconditional (FG_{unco}) and conditional to coefficient #15 fixed at 1.49 kWh (FG_{cond}). (b) Climate change impact score of updated PV system with uncertainty in the foreground and background systems (FGBG) as well as in the characterization factors (FGBGCF). Data available as Supporting Information

and **B** of the product system. Its uncertainty can be expected to propagate throughout the whole supply chain, increasing the resulting influence on the dispersion of the impact score distribution. Coefficient #15 is followed in importance by coefficient #6, which corresponds to the amount of silicon contained in the wafer. This coefficient varies over a wide range (in proportion to the wafer thickness), while also having a considerable impact contribution due to its energy intensity.

To validate these results, we fixed the value of coefficient #15 at its expected (mean) value of 1.49 kWh, reran the Monte Carlo simulation, and plotted the new results overlapping the previous results (Figure 2a). Fixing the quantity of coefficient #15 and comparing the result to the previous impact score distribution confirmed that this parameter can considerably reduce dispersion, making the curve narrower.

Figure 2b shows the distribution of the climate change impact scores of the same system, this time with uncertainty in the background system, based on the uncertainty information available by default in the ecoinvent database (FGBG) and the characterization factors (FGBGCF; see Section 3.1.2). The ecoinvent database provides background uncertainty information by the application of the so-called Pedigree approach (Frischknecht et al., 2004; Ciroth et al., 2016; Heijungs, 2020). It can be seen that with default background uncertainty considered, the characterization factors do not introduce large changes in the distribution of the output.

Including uncertainty in the background **A** and **B** coefficients as well as in the characterization factors (**Q** coefficients) modified the sensitivity ranking considerably, as shown in Table 3. It can be seen that the ranking is now dominated by background coefficients. These background-uncertainties "mask" the sensitivity of foreground coefficients, which scored low δ-values ($\delta \leq 0.05$). The most sensitive coefficients are again associated with mostly downstream processes, which trigger many other processes in the supply chain. The characterization factors scored a $\delta < 0.04$, suggesting negligible sensitivity importance.

TABLE 3 δ-sensitivity measures for climate change impacts of the updated PV system, with uncertain foreground, background systems, and characterization factors

Coeff ID	Process	Input (I)/Output (O)	Description	δ estimate
BG	Photovoltaic slanted-roof installation, 3kWp, single-Si, panel, mounted, on roof	(I) photovoltaic panel, single-Si wafer	Quantity of photovoltaic panels used in the photovoltaic slanted-roof installation, in m^2.	0.22
BG	Electricity production, photovoltaic, 3kWp slanted-roof installation, single-Si, panel, mounted	(I) photovoltaic slanted-roof installation, 3 kWp, single-Si, panel, mounted, on roof	Quantity of photovoltaic slanted-roof installations required to generate electricity, in units (each installation contains 22.072 m^2 of panels).	0.19
BG	Photovoltaic panel production, single-Si wafer	(I) photovoltaic cell, single-Si wafer	Quantity of photovoltaic cells required used in the photovoltaic panel, in m^2.	0.13
BG	Photovoltaic cell production, single-Si wafer	(I) single-Si wafer, photovoltaic	Quantity of single-Si wafers used for photovoltaic cell production, in m^2.	0.09
BG	Photovoltaic mounting system production, for slanted-roof installation	(I) aluminum, wrought alloy	Quantity of aluminum, the wrought alloy used in photovoltaic mounting system production, in kg.	0.08

Abbreviation: BG, background coefficient.

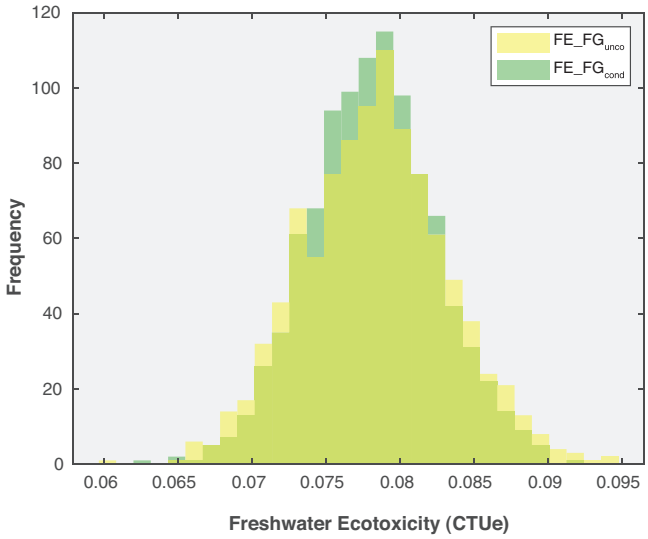

FIGURE 3 Freshwater ecotoxicity impact score of updated PV system with uncertain foreground, unconditional (FG_{unco}) and conditional to coefficient #15 fixed at 1.49 kWh (FG_{cond}). Data available as Supporting Information

3.2.2 | Freshwater ecotoxicity

Figure 3 shows the result of the Monte Carlo analysis for the freshwater ecotoxicity impacts of the updated PV system, considering uncertainty only in the foreground). As with climate change, the most sensitive coefficient is #15, related to the panel conversion efficiency. In this impact category, the δ-value for coefficient #15 was 0.15, while all other coefficients ranked low with $\delta \leq 0.07$. The change in the impact distribution curve resulting from fixing this coefficient at its expected (mean value) of 1.49 kWh is also shown in Figure 3.

Finally, we incorporated uncertainty in the background and characterization factors as earlier described and reran the simulation and GSA. All characterization factors ranked highest (Table 4), followed by the emission of chromium VI during the treatment of red mud from bauxite digestion (associated with the aluminum parts I in the PV system). None of the foreground coefficients ranked high in terms of sensitivity importance (i.e., all has a value smaller than 0.06).

TABLE 4 δ-sensitivity measures for freshwater ecotoxicity impacts of the updated PV system, with uncertain foreground and background systems (BG) and characterization factors (CF)

Coeff ID	Description	δ estimate
CF	Characterization factor for copper emissions to compartment 'air', 'non-urban air or from high stacks'	0.21
CF	Characterization factor for chromium VI emissions to compartment 'water', 'surface water'	0.11
CF	Characterization factor for silver emissions to compartment 'air', 'urban air close to the ground'	0.11
CF	Characterization factor for zinc, ion emissions to compartment 'water', 'ground-'	0.08
BG	Chromium VI emissions (output) from ecoinvent process 'treatment of red mud from bauxite digestion, residual material landfill (red mud from bauxite digestion)'	0.06
CF	Characterization factor for arsenic, ion emissions to compartment 'water', 'surface water'	0.06

4 | DISCUSSION

4.1 | Insights obtained for the updated PV system

The proposed approach to UA and GSA allowed interpreting the results of the updated PV system under assessment beyond what would have been possible to extract using existing methodologies in standard LCA software. The GSA tells us to focus first on the highest ranking coefficient in Table 2 (#15, electricity yield). As a function of panel conversion efficiency (19–22%), the electricity yield of the updated PV system varies only across a relatively small range. However, the influence of the electricity yield on the uncertainty of the system's impact scores is largest because the impacts are embedded in the materials in the panel, cell, and other installation (balance of system). The quantities of materials vary with PV area, which in turn varies with yield and as a result this parameter drives a large part of the uncertainty in the model output.

In absence of comprehensive uncertainty/variability data, we chose a triangular distribution to express uncertainty in coefficient 15, following that fewer panels can be expected to fall in the extremes (higher and lower conversion efficiency ranges). If we had chosen an even less informative distribution to represent this uncertainty (e.g., uniform), the influence of this coefficient would likely be even higher. The analysis calls for LCA studies of PV technologies -especially comparative ones- to pay special attention to yield models and the characterization of uncertainties and variabilities within them. Relative to other aspects of the PV system, errors, poor data, or inconsistencies in the yield models may have much larger chances to introduce error in the LCA impact scores and in the conclusions that can be derived.

A similar logic can be applied to the analysis that included uncertainty in the background coefficients (Table 3). Here, we used the default uncertainty data available in the ecoinvent database, which are generated following the Pedigree approach (Muller et al., 2016). Again in this case, uncertain coefficients that relate to the PV installation and panel area required to produce 1 kWh ranked highest in sensitivity importance. It is the installation/panel area that is driving the material demands and therefore most of the impacts. In the case of panel area (Table 3, row 1), however, the sensitivity importance may be seen as a contrived result of the Pedigree approach. In practice, the uncertainty around what area of panels is fitted in a fixed-size PV installation of 22.07 m^2 should not be large, as only small variations can be expected from occasional panel breakage and replacement. Yet the Pedigree approach in ecoinvent 3.6 applies a log-normal uncertainty to this coefficient with a standard deviation of 1.3269. Since the panel includes numerous upstream components, a variation in this quantity will propagate throughout most of the system, resulting in the high δ values observed. An important takeaway from this analysis is that special attention must be placed on the model input uncertainties that are incorporated, and whether pedigree-type uncertainties should be investigated under the GSA lens.

The case study also highlighted important aspects in the uncertainty of characterization models. The uncertainty of CFs for climate change is smaller than the uncertainty for ecotoxicity CFs, which spans several orders of magnitude. In such cases, the influence of uncertainties in **Q** can dwarf the influence of uncertainties in **A** and **B**, a potential issue that had been flagged by Cucurachi et al. (2016). In our case study, we only applied uncertainty to selected CFs for illustration purposes. Extending the uncertainty to the entire **Q** matrix would be equally straightforward and computationally efficient following the general approach we have presented here. However, further consideration should be given to how model uncertainties are incorporated. In the case of GWP, the IPCC reports specific uncertainties per CF (Myhre et al., 2013). In the case of USEtox, *model* uncertainties are applied to entire groups of CFs depending on the compartment (e.g., a squared geometric standard deviation of 18 for all emissions to freshwater in the freshwater ecotoxicity category, and 103 for all emissions to soil in the same category). In such cases, it may be that the *model* uncertainty in all the CFs within a group can be attributed to shared sources of uncertainty. If a model parameter can be used to represent this shared source of uncertainty, then it would likely outrank other coefficients significantly, as its influence would propagate throughout numerous CFs. The interpretation of such results may be more appropriate, as the GSA would call for a better refinement of the parameter rather than the specific CFs.

4.2 | Potential of the approach in broader applications

The practical approach we propose in this contribution is immediately available in software and applies to LCA studies. We also dedicated some attention to fully aligning the theory and notation of various disciplinary fields. We started from the standard notation used by Heijungs and Suh (2002) for the computational structure of LCA, further merged into LCA the theory and the standard notation of UA and GSA, and finally also presented in mathematical expressions the proposed filtering approach developed in software. While the expression of LCA, UA, and GSA in mathematical notation may be overlooked by the standard LCA practitioner, they are of fundamental importance to improve the collaboration between scientists and across disciplines.

Beyond a theoretical and a practical approach, we also strengthened the interpretation toolbox of the LCA analyst, thus allowing including in any LCA study an assessment and a sensitivity check of the significant inputs, as also prescribed by the ISO standard. The availability of sensitivity results allows the LCA analyst to pinpoint areas in the full LCA system that may require improvement of supporting standard databases (e.g., background inventory data), additional direct data collection (e.g., foreground data), improvement of models (e.g., characterization factors). With GSA, low scores are also of high value, that is, uncertain elements of the model that are found to rank low on the scale of sensitivity importance. Having a clear and grounded identification of these parameters of low importance helps to clear the way for the interpretation and communication of the modeling results. Fixing non-important parameters can reduce the degrees of freedom in the models and help focus efforts—both in further modeling, in their subsequent communication, interpretation, and decision-making that follows from the analysis. As a result, we are now able to provide a better understanding of the relationship between uncertain inputs and outputs of LCA models.

The results also highlight that care should be taken in blindly incorporating in UA the imprecise estimate of the input parameters. The analyst can approximate foreground uncertainty information using uncertainty information based on an understanding of the mechanisms driving it in each parameter. For the background, uncertainty information may come from the Pedigree-style approaches commonly used and propagated in LCA. The limitation of using Pedigree-style information for UA are discussed by Heijungs (2020). The application of GSA on the same Monte Carlo sample that we propose here mitigates potential challenges of Pedigree-information, as the root of the uncertainty of the output is highlighted allowing the LCA practitioner to better explore the system under assessment. An iterative approach, also involving experts on the specific product system under assessment, would allow, if resources are available, to improve the quality of the data or models for those parts of the system that result as top contributors (see Section 3.2).

Our approach can be readily used to apply UA and GSA to complex LCA systems and with reduced computational intensity (~minutes). The filtering solutions applied to reduce the complexity of the Monte Carlo simulation of the input space provided an additional avenue to improve standard UA in LCA software. While contribution-based filtering could be also applied at the level of the stochastic LCA results (i.e., for each Monte Carlo simulation run), we applied it at the level of the non-stochastic LCA results in this contribution. The advantage of this is that the input variables to the GSA are pre-defined and are the same for each simulation run. A potential disadvantage of this approach is that there is a risk that input variables that increase in relevance above the cut-off value as a result of the sampling, would be excluded at this stage although they might become relevant as a result of the sampling.

4.3 | Further work

We proposed an approach in open source LCA software, thus allowing other developers to implement UA and GSA also in other LCA software. The results of the GSA are currently presented in software in a tabular format, ranked based on the δ scores of input parameters. Additional efforts could also further expand our proposal toward a better graphical representation of results.

The filtering approach we proposed will be key to future rapid implementation of advanced techniques of UA and GSA in LCA. Filtering could also be performed following other rationales. For instance, one could consider *attribute-based filtering*, where any attribute or meta-data available could be used to filter exchanges (e.g., the product system could be divided into the foreground and background systems, environmental exchanges could be filtered by emission compartment, or economic exchanges could be filtered based on a sectoral classification). Naturally, if attribute-based filtering is applied, this means that by definition certain variables are excluded from the GSA and their relevance is thus not assessed. Similarly, if contribution-based filtering is applied there is a theoretical risk that input variables are missed as they may only become relevant as a combination of value changes during sampling; this risk can be minimized by using a lower cut-off value, with the drawback of also including more variables into the GSA.

Of particular importance toward the further implementation of the filtering is also the handling of multiple impact categories. The Monte Carlo simulation we performed is for all reference flows and impact categories at once. For the GSA, we currently repeat the filtering for every impact category, and we approach the GSA one impact category at a time (see Section 2.3). The automation of this process and the simultaneous calculation of all impact categories at once would be possible, but it would also likely lead to longer calculation times (e.g., with 10 reference flows and 10 impact

categories the proposed approach will likely be a factor of 100 slower). To this end, additional research on the filtering should also explore potential synergies across impact categories.

Aside from filtering, further work should also go into improving the estimation of the δ function in open source software applications, such as the SAlib library. Here, the algorithm used requires partitioning the output distribution into a given number of bins; see also Plischke and co-authors (Plischke et al., 2013; Plischke & Borgonovo, 2020). In the case study considered, some negative values suggested that the δ estimation protocol was not immediately applicable for the impact category under consideration. Output distributions for freshwater ecotoxicity tend to be extremely pronounced with narrow peaks and long tails. To accommodate for this, we log-transformed the output vector to obtain a curve that could be more easily partitioned. Thanks to transformation invariance, the sensitivity rankings were preserved. The workaround allows using the SALib library in its current format until a better estimate of the δ measure will be implemented.

Recent improvements in the GSA literature will be able to improve on such matters and the computational capabilities of GSA for LCA. For instance, Derennes et al. (2019) and Plishke and Borgonovo (2020) have recently proposed updated estimation protocols, which should be considered for future work as they could allow estimating the δ-values without requiring partitioning and log-transformation, and with using approaches that might be overall computationally less expensive. Such updates should be also adequately developed in software packages that are compatible with open source LCA software. Other updates include the use of more stable estimation algorithms other than the kernel-based estimation used here and the use of improved sampling strategies of the input space (see, e.g., Derennes et al., 2019). To exploit such recent updates and to keep LCA practice up to date, regular close collaboration between LCA scientists and GSA developers is advised.

ACKNOWLEDGMENTS

The authors would like to thank Emanuele Borgonovo, Universitá Bocconi, and Elmar Plischke, TU Clausthal, for their recommendations and feedback on earlier versions of the manuscript. They further thank Daniel de Koning, CML, for supporting us with the software implementation in the Activity Browser. Stefano Cucurachi is thankful to Valentina Bisinella, DTU Environment, for the important discussions on dealing with uncertainty and implementing global sensitivity analysis in LCA.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ORCID

Stefano Cucurachi https://orcid.org/0000-0001-9763-2669
Carlos Felipe Blanco https://orcid.org/0000-0001-8199-8420
Bernhard Steubing https://orcid.org/0000-0002-1307-6376
Reinout Heijungs https://orcid.org/0000-0002-0724-5962

REFERENCES

Baucells, M., & Borgonovo, E. (2013). Invariant probabilistic sensitivity analysis. *Management Science*, *59*(11), 2536–2549.

Bisinella, V., Conradsen, K., Christensen, T. H., & Astrup, T. F. (2016). A global approach for sparse representation of uncertainty in life cycle assessments of waste management systems. *International Journal of Life Cycle Assessment*, *21*(3), 378–394.

Blanco, C. F., Cucurachi, S., Guinée, J. B., Vijver, M. G., Peijnenburg, W. J. G. M., Trattnig, R., & Heijungs, R. (2020). Assessing the sustainability of emerging technologies: A probabilistic LCA method applied to advanced photovoltaics. *Journal of Cleaner Production*, *259*, 120968.259. https://doi.org/10.1016/j.jclepro.2020.120968

Borgonovo, E. (2007). A new uncertainty importance measure. *Reliability Engineering and System Safety*, *92*(6), 771–784.

Borgonovo, E., & Apostolakis, G. E. (2001). A new importance measure for risk-informed decision making. *Reliability Engineering & System Safety*, *72*(2), 193–212.

Borgonovo, E., Castaings, W., & Tarantola, S. (2012). Model emulation and moment-independent sensitivity analysis: An application to environmental modelling. *Environmental Modelling and Software*, *34*, 105–115.

Borgonovo, E., Lu, X., Plischke, E., Rakovec, O., & Hill, M. C. (2017). Making the most out of a hydrological model data set: Sensitivity analyses to open the model black-box. *Water Resources Research*, *53*(9), 7933–7950.

Borgonovo, E., & Plischke, E. (2016). Sensitivity analysis: A review of recent advances. *European Journal of Operational Research*, *248*(3), 869–887.

Boucher, O. (2012). Comparison of physically- and economically-based CO2- equivalences for methane. *Earth System Dynamics*, *3*(1), 49–61.

Ciroth, A., Muller, S., Weidema, B., & Lesage, P. (2016). Empirically based uncertainty factors for the pedigree matrix in ecoinvent. *The International Journal of Life Cycle Assessment*, *21*(9), 1338–1348.

Cucurachi, S., Borgonovo, E., & Heijungs, R. (2016). A protocol for the global sensitivity analysis of impact assessment models in life cycle assessment. *Risk Analysis*, *36*(2), 357–377.

Derennes, P., Morio, J., & Simatos, F. (2019). A nonparametric importance sampling estimator for moment independent importance measures. *Reliability Engineering & System Safety*, *187*, 3–16.

Douziech, M., Oldenkamp, R., van Zelm, R., King, H., Hendriks, A. J., Ficheux, A.-S., & Huijbregts, M. A. J. (2019). Confronting variability with uncertainty in the ecotoxicological impact assessment of down-the-drain products. *Environment International*, *126*, 37–45.

Frischknecht, R., Jungbluth, N., Althaus, H.-J., Doka, G., Dones, R., Heck, T., Hellweg, S., Hischier, R., Nemecek, T., Rebitzer, G., & Spielmann, M. (2004). The ecoinvent database: Overview and methodological framework (7 pp). *The International Journal of Life Cycle Assessment*, *10*(1), 3–9.

Gavankar, S., Anderson, S., & Keller, A. A. (2015). Critical components of uncertainty communication in life cycle assessments of emerging technologies. *Journal of Industrial Ecology*, *19*(3), 468–479.

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

Golsteijn, L., Hendriks, H. W. M., van Zelm, R., Ragas, A. M. J., & Huijbregts, M. A. J. (2012). Do interspecies correlation estimations increase the reliability of toxicity estimates for wildlife? *Ecotoxicology and Environmental Safety*, *80*, 238–243.

Groen, E. A., Bokkers, E. A. M., Heijungs, R., & de Boer, I. J. M. (2017). Methods for global sensitivity analysis in life cycle assessment. *International Journal of Life Cycle Assessment*, *22*(7), 1125–1137.

Groen, E. A., Heijungs, R., Bokkers, E. A. M., & de Boer, I. J. M. (2014). Methods for uncertainty propagation in life cycle assessment. *Environmental Modelling & Software*, *62*, 316–325.

Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, *57*(1), 97–109.

Heijungs, R. (1994). A generic method for the identification of options for cleaner products. *Ecological Economics*, *10*(1), 69–81.

Heijungs, R. (2020). On the number of Monte Carlo runs in comparative probabilistic LCA. *The International Journal of Life Cycle Assessment*, *25*(2), 394–402.

Heijungs, R. (2021). Selecting the best product alternative in a sea of uncertainty. *The International Journal of Life Cycle Assessment*, *26*(3), 616–632.

Heijungs, R., & Kleijn, R. (2001). Numerical approaches towards life cycle interpretation five examples. *The International Journal of Life Cycle Assessment*, *6*(3), 141–148.

Heijungs, R., de Koning, A., & Wegener Sleeswijk, A. (2015). Sustainability analysis and systems of linear equations in the era of data abundance. *Journal of Environmental Accounting and Management*, *3*(2), 109–122.

Heijungs, R., & Lenzen, M. (2014). Error propagation methods for LCA—A comparison. *The International Journal of Life Cycle Assessment*, *19*(7), 1445–1461.

Heijungs, R., & Suh, S. (2002). *The computational structure of life cycle assessment* (Vol. 11). Springer Science and Business Media.

Henriksson, P. J. G., Heijungs, R., Dao, H. M., Phan, L. T., De Snoo, G. R., & Guinée, J. B. (2015). Product carbon footprints and their uncertainties in comparative decision contexts. *PLoS ONE*, 10(3), 1–11.

Herman, J., & Usher, W. (2017). SALib: An open-source python library for sensitivity analysis. *Journal of Open Source Software*, *2*(9), 97.

Huijbregts, M. (1998). Application of uncertainty and variability in LCA. *The International Journal of Life Cycle Assessment*, *3*(5), 273.

Huijbregts, M. (2003). Geographical scenario uncertainty in generic fate and exposure factors of toxic pollutants for life-cycle impact assessment. *Chemosphere*, *51*(6), 501–508.

Igos, E., Benetto, E., Meyer, R., Baustert, P., & Othoniel, B. (2019a). How to treat uncertainties in life cycle assessment studies? *The International Journal of Life Cycle Assessment*, *24*(4), 794–807.

Igos, E., Benetto, E., Meyer, R., Baustert, P., & Othoniel, B. (2019b). How to treat uncertainties in life cycle assessment studies? *International Journal of Life Cycle Assessment*, *24*(4), 794–807.

IPCC. (2014). Climate change 2013—The physical science basis. In T. F. Stocker, D. Qin, G. K. Plattner, M. M. B. Tignor, S. K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex, & P. M. Midgley (Eds.), *Climate change 2013: The physical science basis. Working group I contribution to the fifth assessment report of the Intergovernmental Panel on Climate Change* (Vol. 9781107057). Cambridge University Press.

ISO. (2006). ISO 14044:2006 environmental management—Life cycle assessment—Requirements and guidelines.

Jung, J., von der Assen, N., & Bardow, A. (2014). Sensitivity coefficient-based uncertainty analysis for multi-functionality in LCA. *The International Journal of Life Cycle Assessment*, *19*(3), 661–676.

Lacirignola, M., Blanc, P., Girard, R., Pérez-López, P., & Blanc, I. (2017). LCA of emerging technologies: Addressing high uncertainty on inputs' variability when performing global sensitivity analysis. *Science of The Total Environment*, *578*, 268–280.

Lloyd, S. M., & Ries, R. (2007). Characterizing, propagating, and analyzing uncertainty in life-cycle assessment. *Journal of Industrial Ecology*, *11*(1), 161–179.

Lunardi, M. M., Alvarez-Gaitan, J. P., Chang, N. L., & Corkish, R. (2018). Life cycle assessment on PERC solar modules. *Solar Energy Materials and Solar Cells*, *187*, 154–159.

Mendoza Beltran, A., Prado, V., Font Vivanco, D., Henriksson, P. J., Guinée, J. B., & Heijungs, R. (2018). Quantified uncertainties in comparative life cycle assessment: What can be concluded? *Environmental Science & Technology*, *52*(4), 2152–2161.

Michiels, F., & Geeraerd, A. (2020). How to decide and visualize whether uncertainty or variability is dominating in life cycle assessment results: A systematic review. *Environmental Modelling & Software*, *113*, 104841.

Muller, S., Lesage, P., Ciroth, A., Mutel, C., Weidema, B. P., & Samson, R. (2016). The application of the pedigree approach to the distributions foreseen in ecoinvent v3. *The International Journal of Life Cycle Assessment*, *21*(9), 1327–1337.

Mutel, C. (2017). Brightway: An open source framework for Life Cycle Assessment. *Journal of Open Source Software*, *2*(12), 236.

Mutel, C. (2020). Graph traversal function for brightspace. https://2.docs.brightway.dev/lca.html#graph-traversal

Mutel, C. L., de Baan, L., & Hellweg, S. (2013). Two-step sensitivity testing of parametrized and regionalized life cycle assessments: methodology and case study. *Environmental Science & Technology*, *47*(11), 5660–5667.

Myhre, G., Shindell, D., & Pongratz, J. Intergovernmental Panel on Climate Change. (2014). Anthropogenic and natural radiative forcing. In *Climate Change 2013 – The Physical Science Basis: Working Group I Contribution to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (pp. 659-740). Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781107415324.018

von Pfingsten, S., Broll, D. O., von der Assen, N., & Bardow, A. (2017). Second-order analytical uncertainty analysis in life cycle assessment. *Environmental Science & Technology*, *51*(22), 13199–13204.

Plischke, E., & Borgonovo, E. (2020). Fighting the curse of sparsity: Probabilistic sensitivity measures from cumulative distribution functions. *Risk Analysis*, *40*(12), 2639–2660.

Plischke, E., Borgonovo, E., & Smith, C. L. (2013). Global sensitivity measures from given data. *European Journal of Operational Research*, *226*(3), 536–550.

Rabitz, H. (1989). Systems analysis at the molecular scale. *Science*, *246*(4927), 221–226.

Razavi, S., Jakeman, A., Saltelli, A., Prieur, C., Iooss, B., Borgonovo, E., Plischke, E., Lo Piano, S., Iwanaga, T., Becker, W., Tarantola, S., Guillaume, J. H.A., Jakeman, J., Gupta, H., Melillo, N., Rabitti, G., Chabridon, V., Duan, Q., … Maier, H. R., (2021). The future of sensitivity analysis: An essential discipline for systems modeling and policy support. *Environmental Modelling and Software*, *137*, 104954.

Reinhard, J., Mutel, C., Wernet, G., Zah, L., & Hilty, M. (2016). Contribution-based prioritization of LCI database improvements: Method design, demonstration, and evaluation. *Environmental Modelling & Software*, *86*, 204–218.

Risbey, J., Van Der Sluijs, J., Kloprogge, P., Ravetz, J., Funtowicz, S., & Quintana, S. C. (2005). Application of a checklist for quality assistance in environmental modelling to an energy model. *Environmental Modeling and Assessment*, *10*(1), 63–79.

Rodriguez, M. A., & Neubauer, P. (2012). The graph traversal pattern. In *Graph data management: Techniques and applications* (pp. 29–46). IGI Global.

Rosenbaum, R. K., Bachmann, T. M., Gold, L. S., Huijbregts, M. A. J., Jolliet, O., Juraske, R., Koehler, A., Larsen, H. F., MacLeod, M., Margni, M., McKone, T. E., Payet, J., Schuhmacher, M., van de Meent, D., Hauschild, M. Z. (2008). USEtox - The UNEP-SETAC toxicity model: Recommended characterisation factors for human toxicity and freshwater ecotoxicity in life cycle impact assessment. *International Journal of Life Cycle Assessment*, *13*(7), 532–546.

Sai, H., Umishio, H., Matsui, T., Nunomura, S., Kawatsu, T., Takato, H., & Matsubara, K. (2018). Impact of silicon wafer thickness on photovoltaic performance of crystalline silicon heterojunction solar cells. *Japanese Journal of Applied Physics*, *57*(8S3), 08RB10-1–08RB10-6.

Saltelli, A., Aleksankina, K., & Becker, W. (2019). Why so many published sensitivity analyses are false: A systematic review of sensitivity analysis practices. *Environmental Modelling & Software*, *114*, 29–39.

Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., Saisana, M., & Tarantola, S. (2008). *Global sensitivity analysis. The primer*. John Wiley & Sons.

Scherer, L., & Pfister, S. (2016). Dealing with uncertainty in water scarcity footprints. *Environmental Research Letters*, *11*(5), 054008.

Sobol', I. M. (1993). Sensitivity analysis for non-linear mathematical models. *Mathematical Modelling and Computational Experiment*, *1*, 407–414.

Stamford, L., & Azapagic, A. (2018). Environmental impacts of photovoltaics: The effects of technological improvements and transfer of manufacturing from Europe to China. *Energy Technology*, *6*(6), 1148–1160.

Steubing, B., Koning, D., Haas, M., & Mutel, C. (2020). The Activity Browser—An open source LCA software building on top of the brightway framework. *Software Impacts*, *3*, 100012.

Wernet, G., Bauer, C., Steubing, B., Reinhard, J., Moreno-Ruiz, E., & Weidema, B. (2016). The ecoinvent database version 3 (part I): overview and methodology. *The International Journal of Life Cycle Assessment*, *21*(9), 1218–1230.

Woodhouse, M., Smith, B., Ramdas, A., & Margolis, R. (2020). *Crystalline silicon photovoltaic module manufacturing costs and sustainable pricing: 1H 2018 benchmark and cost reduction road map*.Golden, CO (United States): National Renewable Energy Lab (NREL).

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

How to cite this article: Cucurachi S, Blanco CF, Steubing B, Heijungs R. Implementation of uncertainty analysis and moment-independent global sensitivity analysis for full-scale life cycle assessment models. *J Ind Ecol*. 2022;*26*:374–391. https://doi.org/10.1111/jiec.13194