The International Journal of Life Cycle Assessment (2019) 24:2238–2254
https://doi.org/10.1007/s11367-019-01635-5

UNCERTAINTIES IN LCA

Prioritizing regionalization efforts in life cycle assessment through global sensitivity analysis: a sector meta-analysis based on ecoinvent v3

Laure Patouillard[1,2,3] · Pierre Collet[2] · Pascal Lesage[1] · Pablo Tirado Seco[1] · Cécile Bulle[4] · Manuele Margni[1]

Received: 27 March 2018 / Accepted: 7 May 2019 / Published online: 18 May 2019
© Springer-Verlag GmbH Germany, part of Springer Nature 2019

Abstract

Purpose Regionalization in life cycle assessment (LCA) aims to increase the representativeness of LCA results and reduce the uncertainty due to spatial variability. It may refer to adapting processes to better account for regional technological specificities (*inventory regionalization*) or adding of spatial information to the elementary flows (*inventory spatialization*) which allow using more regionalized characterization factors. However, developing and integrating regionalization requires additional efforts for LCA practitioners and database developers that must be prioritized.

Methods We propose a stepwise methodology for LCA practitioners to prioritize data collection for regionalization based on global sensitivity analysis (GSA) using Sobol indices. It involves several GSA to select the impact categories (ICs) that require further inventory data collection (IC ranking), prioritize between inventory regionalization and inventory spatialization (LCA phase ranking), and target specific data to collect. Then we propose a method to derive sector-specific recommendations using statistical tests to prioritize inventory regionalization versus spatialization and the ICs on which to focus inventory data collection. These recommendations are meant to help LCA practitioners and database developers define their strategy for regional data collection by focusing on data that have the highest potential to reduce the uncertainty of the results.

Results and discussion The applicability of the methodology is illustrated through three case studies using the ecoinvent v3 database and the regionalized impact methodology IMPACT World+: one on prioritizing data collection in a single biofuel product system and two meta-analyses of all product systems in two distinct economic sectors (biofuel production and land passenger transport). Recommendations for regionalization can be derived for an economic sector and appear to be different from one economic sector to another. GSA seems to be more relevant to prioritize regionalization efforts than an impact contribution analysis (ICA) approach often used to prioritize data collection in LCA. However, further improvements, such as accounting for spatial correlations and better computational times for GSA, are required to implement it in LCA.

Conclusions We recommend using the methodology based on GSA to efficiently prioritize regionalization efforts between ICs and between inventory regionalization and inventory spatialization. We proved that the implementation of IC ranking and LCA phase ranking is computationally feasible and therefore invite current LCA software providers to unlock this new horizon in LCA interpretation. We also invite to expand the meta-analysis to all sectors in an LCA database.

Keywords Data collection · Economic sector · Global sensitivity analysis · Prioritization · Regionalization

Responsible editor: Andreas Ciroth

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s11367-019-01635-5) contains supplementary material, which is available to authorized users.

✉ Laure Patouillard
laure.patouillard@polymtl.ca

[1] CIRAIG, Polytechnique Montréal, P.O. Box 6079, Montreal, QC H3C 3A7, Canada

[2] IFP Energies nouvelles, 1-4 avenue de Bois-Préau, 92852 Rueil-Malmaison, France

[3] UMR 0210 INRA-AgroParisTech Economie publique, INRA, Thiverval-Grignon, France

[4] CIRAIG, ESG UQAM, C.P.8888, succ. Centre ville, Montréal, QC H3C 3P8, Canada

1 Introduction

Life cycle assessment (LCA) is a methodological framework to assess the potential environmental impacts of a product or service throughout its life cycle (International Organization for Standardization (ISO) 2006a, b). LCA is traditionally site generic, i.e., it disregards the spatial information. The regionalization of LCA started in the 1990s to reduce the inaccuracies associated with site-generic LCA and increase the discrimination power of LCA (Vigon et al. 1993; Udo de Haes et al. 1999; Ross and Evans 2002). Indeed, the characteristics of an activity used in the life cycle inventory (LCI) often vary depending on the geographic location of the activity, such as electricity production technology mix, yields, plant operations (Turconi et al. 2013), or agricultural processes (Yang et al. 2018). Similarly, in the life cycle impact assessment (LCIA) phase, the potential environmental impact of an elementary flow (EF) may depend on "the location of a given source and the conditions of its surroundings" (Potting and Hauschild 2006). To address this, spatially differentiated LCIA methods were developed, leading to the calculation of regionalized characterization factors (CFs) (Pfister et al. 2009; Boulay et al. 2011, 2018; Helmes et al. 2012; Roy et al. 2013; Chaudhary et al. 2015; Plouffe et al. 2015). In this paper, regionalization refers to any attempt to increase the representativeness of unit processes and environmental phenomena by accounting for its location (Patouillard et al. 2016, 2018). At the LCI level, it may refer to adapting inputs or outputs of unit processes to better account for regional technological specificities or recontextualizing unit processes to better account for the specific background of an activity (Lesage and Samson 2016), collectively referred to as *inventory regionalization* (Patouillard et al. 2018). At the LCI level, regionalization can also refer to the adding of spatial information to the activities and elementary flows, referred to as *inventory spatialization* (Patouillard et al. 2018), which makes it possible to use more regionalized CFs. At the LCIA level, developing spatially differentiated or regionalized LCIA methods is referred to as *impact regionalization*. Ultimately, regionalization in LCA aims to increase the representativeness and environmental relevance of LCA results and reduce the uncertainty due to spatial variability. However, developing and integrating regionalization in LCA requires additional work for LCA practitioners, LCI database developers, and LCIA method developers, specifically for data collection and integration into the LCA model (Baitz et al. 2012). LCA practitioners must, therefore, prioritize their regionalization efforts.

LCA practitioners and LCI database developers traditionally regionalize or spatialize the inventory, whereas LCIA method developers regionalize LCIA methods, often relying on years of research. In this article, we do not investigate data collection for impact regionalization and rather focus on data collection for inventory regionalization and spatialization. Inventory regionalization will affect the type and quantity of flows whereas inventory spatialization will affect the CF that will be used. Therefore, the way for LCA practitioners to integrate regionalization and spatialization in LCA model is different, even if the type of data to be collected may sometimes rely on the same data sources (Patouillard et al. 2018). Inventory regionalization requires the collection of more representative data on technological characteristics and the context of activities in a specific region. Inventory spatialization requires data on the geographic locations of the assessed activity within its spatial coverage (spatial distribution of the activity), which must then be associated with the unit process and its elementary flows. It may involve the use of a geographic information system (GIS). The type of data and modeling required for inventory regionalization and inventory spatialization also depends on the assessed impact category (IC), i.e., the EF classified in the IC and the spatial resolution of the IC. For instance, when assessing water use IC, data collection will be focused on the water consumption of processes (for regionalization) and the spatial distribution of water EFs among watersheds (for spatialization). As the type of data to be collected and the modeling is different, efforts should be prioritized: (i) across ICs to determine which ICs require further priority data collection for inventory regionalization or inventory spatialization and (ii) between inventory regionalization or inventory spatialization. In addition, the requirements for data collection and its prioritization for inventory regionalization and inventory spatialization may vary from one economic sector to another. Therefore, the priority for data collection is product system specific or specific to economic sectors.

As proposed by Clavreul et al. (2013), there are different types of uncertainty in LCA: stochastic uncertainty, often referred to as variability in LCA (spatial, temporal, technological), and epistemic uncertainty related to the lack of knowledge on reality, often simply referred to as uncertainty in LCA (Huijbregts 1998). Here, we use the word *uncertainty* to refer to both stochastic and epistemic uncertainty. Here we focus on regionalization, so we aim at reducing the uncertainty due to spatial variability. However, LCA practitioner not only aims at reducing spatial variability but also the overall uncertainty. Consequently, in our case studies, we account for all uncertainty sources currently available at the operationalization level, i.e., LCI uncertainty covered by the pedigree matrix and spatial variability of CFs for LCIA uncertainty. It means that LCA practitioner may have to prioritize their efforts between inventory data refinement (including regional data) and spatialization.

An uncertainty analysis aims to describe the uncertainty of the results, whereas a sensitivity analysis aims to describe the contribution of input data to the uncertainty of the results (sometimes referred as to uncertainty contribution analysis) (Igos et al. 2015). There are two types of sensitivity analysis: local sensitivity analysis (LSA) and global sensitivity analysis (GSA). LSA assesses the effects of small variations of uncertain inputs on the results (Huijbregts et al. 2001; Sakai and Yokoyama 2002). It is the most widely used sensitivity analysis and is generally less time consuming than GSA. However, it only provides a partial view of the sensitivity as it only tests a very small variation of the input parameter around its deterministic value (first-order derivative). Therefore, it ignores the overall range of variation of input variables and also ignores the correlation between parameters (Mutel et al. 2013). However, the overall range of variation could be large for parameters in LCA, especially when accounting for spatial variability of LCI data like in the agricultural sector (Yang et al. 2018) and spatial variability of regionalized CFs at the global scale (Roy et al. 2013; Boulay et al. 2018). GSA explores the effects of the overall range of variation of input variables on the uncertainty of the results. Therefore, we focus on the use of GSA to perform sensitivity analysis.

The most sensitive data holds the highest potential to reduce the uncertainty of the results, whenever possible. Heijungs (1996) recommends prioritizing data collection by focusing on the data that most contribute to the impact scores and which are also the most uncertain. In other words, data collection efforts should focus on the most sensitive data, i.e., the data that contribute the most to the uncertainty. However, in practice, data collection is mostly prioritized according to the data that most contribute to the impact scores without taking uncertainty into account. This is referred to as impact contribution analysis (ICA). For instance, different authors propose methods based on ICA to prioritize LCI database improvements (Reinhard et al. 2016) or spatial dimension integration in LCA studies (Hernández-Padilla et al. 2017; Patouillard et al. 2018). ICA was also used to prioritize the recontextualization effort in the Québec LCI database (Lesage and Samson 2016). Indeed, ICA is already available in all LCA software and does not require additional computational time. Sensitivity analysis is rarely used to prioritize data collection in LCA (Collet et al. 2014; Pfister and Scherer 2015; Gregory et al. 2016; Wender et al. 2018). While not performing it themselves, Reinhard et al. (2016) suggest that sensitivity analysis would be a more efficient way to prioritize data collection for LCI database improvements than ICA. Furthermore, only two methodologies based on GSA exist to prioritize data collection efforts in LCA (Mutel et al. 2013; Gregory et al. 2016), and one of them was specifically designed for regionalized LCA model (Mutel et al. 2013). Both methodologies rely on the Spearman rank correlation coefficients as GSA indicators. However, the validity of these coefficients was being questioned, as they tend to overestimate the main sensitivity for a model with a high number of input parameters and with interactions (Saltelli and Sobol' 1995; Pfister and Scherer 2015). The LCA model is a model with a high number of uncertain parameters (economic flows, elementary flows, and CFs could be uncertain) and potentially a high degree of interactions due to the economic matrix inversion and matrix multiplications (Wei et al. 2015). Therefore, this article aims to fill this gap by proposing a methodology to prioritize data collection based on GSA indicators that are adapted to the LCA model characteristics.

This article has two main objectives: (i) propose a stepwise methodology for LCA practitioners to prioritize data collection for regionalization (i.e., regional data collection) based on operational GSA tools and discuss the effectiveness of this approach as compared to an ICA approach and (ii) derive sector-specific recommendations to prioritize inventory data refinement versus spatialization and the ICs on which to focus data collection and modeling. These recommendations are meant to help LCA practitioners and LCI database developers define their strategy for regional data collection by focusing on data that have the highest potential to reduce the uncertainty of the results. The applicability of the methodology is illustrated through three case studies: one on prioritizing data collection in a single biofuel product system and two meta-analyses of all product systems in two distinct economic sectors (biofuel production and passenger land transport sectors), as defined in the ecoinvent database v3 cutoff (Wernet et al. 2016) applying the regionalized LCIA methodology IMPACT World+ (Bulle et al. 2019).

2 Methods

2.1 Overview of the proposed methodology

As shown in Fig. 1, a typical LCA model is an assembly of different models among which are (i) the LCI model with LCI input variables calculating EFs as output and (ii) the LCIA model with LCIA input variables calculating CFs as output. Uncertainty, and especially uncertainty due to spatial variability, may come from LCI input variables and LCIA input variables. However, in practice, LCA practitioners and LCI database developers do not influence LCIA input variables: they simply apply CFs resulting from LCIA models and provided by LCIA method developers. Consequently, their focus is on two groups of input variables: LCI variables, which are LCI input variables of the LCI model, and LCIA variables, which are CFs resulting from the LCIA model (Fig. 1). In this study, we adopt the point of view of LCA practitioners and LCI database developers, so both groups of variables will be referred to as input variables. A straightforward approach is to perform GSA on the LCA results to directly test the influence of each input variable seeing the LCA model as a black box (dashed line on Fig. 1). Unfortunately, this GSA approach can be computationally intensive and time-consuming because of the large number of variables, high order of interactions, and correlations between them. For instance, a GSA based on classical procedure to estimate Sobol indices has a computational cost of $N \times (k+2)$ where N is the number of samplings and k the number of uncertain variables (Saltelli et al. 2010; Wei et al. 2015). For the ecoinvent product system *ethanol production from sugarcane, BR* containing about 400,000 LCI variables, with IMPACT World+ which contains about 45,000

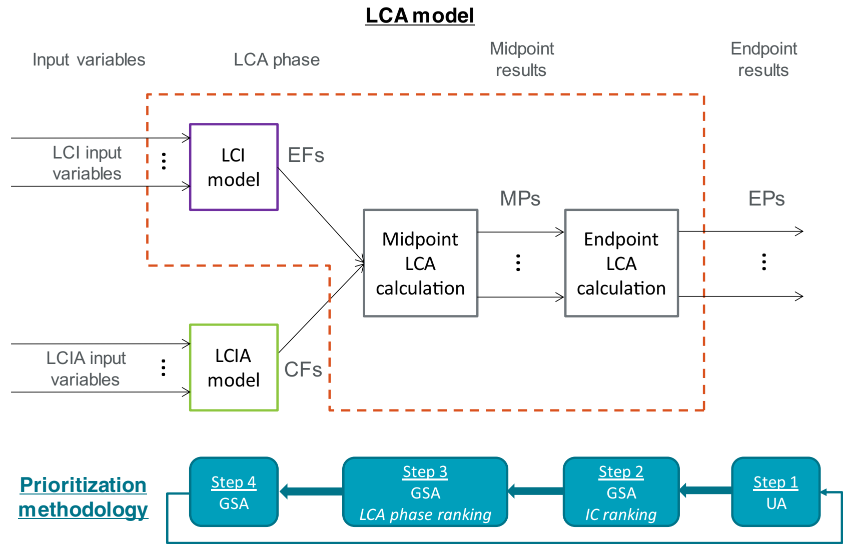

Fig. 1 From a black box approach of the LCA model (red dashed line as boundaries) to a stepwise approach to prioritize the regionalization effort in LCA based on UA and GSA. EFs: elementary flows, CFs: characterization factors, MPs: midpoint impact categories, EPs: endpoint impact categories, GSA: global sensitivity analysis, UA: uncertainty analysis

LCIA variables, a straightforward GSA would require about 230 days for calculation (assuming 5000 samplings using Brightway which is the fastest LCA software for Monte Carlo). Alternatively, we propose to decompose this complex problem into smaller elements and solve each one through a stepwise approach considering the structure of the LCA model and LCA practices (Fig. 1) to target a restricted number of relevant ICs and variables within LCI or LCIA group variables. This may be seen as an alternative to the two-step sensitivity analysis proposed by Mutel et al. (2013), who propose to perform a screening on the black box approach to eliminate insensitive variables and perform a GSA on the remaining ones. At first, we performed a GSA on endpoint results to identify the most sensitive midpoint ICs (step 2) after checking whether uncertainty must be reduced (step 1). Then, for most sensitive midpoint ICs, we performed a GSA to compare the sensitivity of the group of LCI variables to that of the group of LCIA variables (step 3). Finally, a final GSA tests the influence of each variable before matrix inversion among the ICs and the group of variables (LCI or LCIA) that were shown to be most sensitive (step 4). When using a midpoint LCIA methodology, step 2 should be skipped and step 3 should be applied to all midpoint ICs.

The stepwise and iterative methodology we propose to prioritize the regionalization effort in LCA (Fig. 1) is inspired from the scheme for the analysis of data inaccuracy in LCI proposed by Huijbregts et al. (2001). It aims to guide the LCA practitioner in prioritizing data collection to reduce the overall uncertainty of LCA results for one product system. It is made up of four steps, as summarized in Fig. 1 in Electronic Supplementary Materia (ESM).

- Step 1—determining whether the uncertainty of the results meets the target level of acceptable uncertainty. The confidence level of the decision-maker in the LCA results is reflected in the uncertainty level of the results (Refsgaard et al. 2007). If the confidence level is inadequately estimated, the decision-maker may make the wrong political or strategic decisions with potential negative impacts (Wardekker et al. 2008). Therefore, the target level of acceptable uncertainty should be set by the decision-maker to make the decision (Laurin et al. 2016). Considering that, in practice, some uncertainties can hardly be reduced in LCA (Weidema and Wesnæs 1996), the target level of acceptable uncertainty should be the outcome of a dialog between the LCA practitioner and the decision-maker. It should be in line with factors including the goal and scope of the study, assessed product system, available time and financial resources for the study, knowledge of the LCA practitioner, and data availability (Herrmann et al. 2014). It should at least be assessed qualitatively and compared with the uncertainty of LCA results, which addresses input uncertainty from both LCI variables and CFs, if available. If the uncertainty of the results does not meet the target level of acceptable uncertainty, efforts are needed to reduce the results uncertainty. This article does not aim to provide tools or procedures to determine what acceptable uncertainty levels would be in the context of a study.
- Step 2—selecting the impact category that requires further LCI data collection (IC ranking). If the uncertainty of the results does not meet the target level of acceptable uncertainty, the ICs that most contribute to the uncertainty of endpoint impact scores should be investigated first. Building on a midpoint-endpoint LCIA methodology framework, GSA is performed to rank the sensitivity of each midpoint IC to the corresponding endpoint IC, as described in Section 2.3. This step could be influenced by the number of EFs classified in the IC: a sensitive EF in a low sensitive IC with few EF classified could be hidden by the high sensitivity of an IC with many EFs.

Depending on available resources and the target level of uncertainty, one or more ICs, starting from the most sensitive, are selected for the next step. If using a midpoint LCIA methodology, this step should be skipped.

- Step 3—selecting inventory regionalization or inventory spatialization (LCA phase ranking). This step aims to test whether the main source of uncertainty lies among the group of LCI variables or the group of LCIA variables (i.e., CFs) to inform the LCA practitioner where to prioritize data collection. It refers to as *LCA phase ranking* (does the uncertainty mainly come from LCI or LCIA variables?) but it is used to prioritize data collection between inventory regionalization and spatialization. Inventory spatialization (e.g., collecting information on the spatial distribution of an EF) is meaningful only for spatially differentiated ICs by accounting for the spatial variability of the spatially aggregated CFs at a broader geographic scale. The sensitivity of the groups of LCI and LCIA variables is quantified using GSA, as described in Section 2.3. The LCA practitioner will, therefore, focus the data collection on inventory regionalization if the group of LCI variables is the main source of spatial uncertainty or on inventory spatialization if the group of LCIA variables is the main source of spatial uncertainty. If using a midpoint LCIA methodology, this step should be applied to all midpoint ICs.
- Step 4—collecting data for inventory regionalization or inventory spatialization. The most sensitive {LCI variable|unit process} couple for inventory regionalization or {elementary flow|unit process} couple for inventory spatialization should be identified for selected ICs based on a GSA and selected for additional data collection. This GSA may be performed following the framework defined by Wei et al. (2015), who recommend GSA tools based on the number of input variables by accounting for interactions and correlations in LCA. After refining the model with the newly collected data, the stepwise process starts over if the target level of acceptable uncertainty has not been reached. This article does not aim to provide additional tools or procedures for step 4.

2.2 Data and settings used for the case studies

We applied the proposed methodology to three case studies: one based on the assessment of a single product system and two others based on the meta-analysis of two economic sectors. Our case studies relied on the ecoinvent v3.3 database (cutoff by classification version) for the LCI data (Wernet et al. 2016) and the IMPACT World+ regionalized LCIA methodology for the LCIA data (Bulle et al. 2019). The single product system that was selected is *ethanol production from sugarcane*, *BR*. We chose two economic sectors, as defined in ecoinvent v3.3, which use the ISIC rev. 4 classification (United Nations Statistics Division 2008): biofuel production sector (19a—liquid and gaseous fuels from biomass) and passenger land transport sector (4922—other passenger land transport). They include 95 and 127 product systems, respectively. The implemented version of IMPACT World+ has two endpoint ICs, ecosystem quality (EQ) and human health (HH), and several midpoint ICs. We used the Brightway 2 (Mutel 2017) LCA software to perform the LCA calculation (conventional LCA calculation, not regionalized LCA calculation) and the Monte Carlo simulations with independent sampling (5000 iterations). The code used for the case studies and a visual description of the sampling procedure for the case study is available in the ESM.

The uncertainty of the LCI data is based on the pedigree matrix approach to estimate the probability distribution parameters of input data using a lognormal distribution (Muller et al. 2016) most of the time. Several sources of uncertainty are accounted using the pedigree matrix approach, including an estimation of the spatial variability of the LCI data. The version of ecoinvent 3 implemented in Brightway (and in all other LCA software) does not consider correlations between input parameters, which may bias the results when performing an uncertainty or sensitivity analysis (Groen and Heijungs 2017). For instance, mass conservation in a unit process is not respected between input and output water flows when performing a Monte Carlo analysis. In the same way, the amount of land transformation elementary flows (from/to) in a unit process are not balanced when each amount is independently sampled. Therefore, we developed a methodology to preserve physical correlation, such as mass conservation, between uncertain input and output flows within a unit process when performing an uncertainty analysis. It only applies to input and output flows that are represented as random variables and if the distribution used to describe the uncertainty of the flow quantity is a lognormal distribution. Doing so, inputs are no longer variables but are recalculated based on the output variables. The detail of this methodology is described in the ESM. We applied it to all land transformation and water elementary flows.

The uncertainty of LCIA variables can be decomposed into the basic uncertainty of CFs from the LCIA model itself and uncertainty due to the variability of CFs over space and time (if spatially or time-differentiated). For the LCIA variables, we only considered the spatial variability of regionalized CFs as the source of uncertainty because it is the only type of uncertainty that the LCA practitioner can reduce himself and it is the only uncertainty information currently available in a format that can be integrated into an uncertainty analysis. Spatial variability is accounted for in spatially differentiated ICs in IMPACT World: freshwater acidification, terrestrial acidification, freshwater eutrophication, land occupation, and land transformation for EQ and water availability for HH. Global

CFs are used to run LCA calculations for regionalized ICs, even when we knew the process location, in order to test if more regionalized CFs are needed or not. Global CFs are calculated using a weighted average of native CFs based on different proxies representing the probability for an elementary flow to occur in each native region (Bulle et al. 2019). The spatial variability of global CFs is represented with a four-parameter beta distribution, which has definite lower and upper bounds and fits a variety of shapes. We used the moment method described in Riggs (1989) to estimate the four parameters of each CF distribution. During the Monte Carlo sampling, we also accounted for the LCIA spatial correlation in a unit process for land transformation IC only and only for some EFs, i.e., sampled value for CF *from* one type of land use is consistent with the sampled value for CF *to* one type of land use. Other spatial LCIA correlations for regionalized ICs are not considered as data and tools are not ready to implement it in a reasonable amount of time. More specifically, we are sampling the same CF for all EFs and disregarding the EF location, i.e., implicitly assuming all activities are occurring in the same location.

2.3 GSA methodology used for IC ranking and LCA phase ranking

GSA may be performed using analytic (Imbeault-Tétreault 2010; Heijungs 2010) or sampling methods. The former is less computationally intensive but involve certain limitations, among which the fact that the results are only robust for slight input variations and do not account for the correlation between input variables (Heijungs 2010). However, in the LCA framework, probability density functions of LCA input variables may span several orders of magnitude (especially regionalized CFs) and are likely to be correlated with each other. GSA using sampling methods are more computationally intensive, but they can account for large input variations and the correlation between variables. Therefore, they are better adapted to the LCA framework. There are different approaches to performing a GSA using sampling methods. These may be divided into two groups: (a) screening approaches (e.g., elementary effects method), which aim to simplify a model by identifying non-influential variables before performing a more targeted GSA, and (b) importance measures (e.g., correlation coefficient, Sobol indices) that quantify the importance of input variables on the results (Iooss and Lemaître 2015). In this article, because we were interested in identifying most sensitive variables, we focused on importance measures. Importance measures based on regression techniques, such as the Spearman correlation coefficient, are often used in LCA (Mutel et al. 2013; Pfister and Scherer 2015) but perform poorly for models with interaction (nonlinear) (Saltelli and Sobol' 1995; Borgonovo and Plischke 2016). Since LCA is a model with interactions, importance measures based on regression techniques should be avoided. Sobol indices (also called sensitivity indices) (Sobol 1993) are based on variance decomposition and are recommended for sensitivity analysis in LCA as they account for interactions and correlations (Padey et al. 2013; Wei et al. 2015). Therefore, the sensitivity indices are chosen as sensitivity measurements for GSA for IC ranking and LCA phase ranking.

Performing a GSA to obtain sensitivity indices may be divided into two steps: (1) uncertainty analysis and (2) estimation of sensitivity indices based on the uncertainty analysis results. Here, uncertainty analysis is performed using a Monte Carlo simulation. The estimation of sensitivity indices to rank the ICs and LCA phase, respectively, is described in the following sections.

2.3.1 Estimation of sensitivity indices from GSA

Sensitivity indices are non-parametric, i.e., no assumption on the form of the probability distribution for the input variables or the result is required to calculate it. They could be calculated for any kind of model, including non-linear ones. Two important measures may be derived from Sobol variance decomposition: (1) the first-order sensitivity index $SI1_{X_i}$, which measures the main effect of the variable X_i on the results, and (2) the total sensitivity index SIT_{X_i}, which measures the total effects of the variable X_i on the results, i.e., the main effect and the interaction effects with the other variables (Saltelli et al. 2010). Each sensitivity index has its own purpose (Saltelli and Tarantola 2002; Saltelli 2017): (1) first-order sensitivity indices are designed for factor prioritization, i.e., to identify the most sensitive variables; (2) total sensitivity indices are designed for factor fixing, i.e., to identify non-sensitive variables in order to fix them for model reduction purposes to build a simplified model, for instance (Padey et al. 2013). However, both provide complementary information on the influence of a variable in the model. Indeed, a variable may have no main effect but still influence the model through the interaction effects. Estimating first-order sensitivity indices is relatively easy, whereas estimating total sensitivity indices is more computationally intensive (Saltelli et al. 2010). However, total sensitivity indices may be calculated easily for simple models, as demonstrated in Sections 2.3.2 and 2.3.3. Since the purpose of this article is to prioritize the data collection effort, we selected the first-order sensitivity index as an indicator for prioritization but always check higher order sensitivity indexes that reflect the interaction effects to remain critical with regard to the total effect of the variables on the model.

For the model $Y = f(X_1, X_2, \ldots X_K)$ where X_i are the uncertain (or random) variables of the model and Y is the output, the first-order sensitivity index ($SI1_{X_i}$) and total sensitivity index (SIT_{X_i}) for X_i are defined as follows (Saltelli et al. 2010):

$$SI1_{X_i} = \frac{Var(E(Y|X_i))}{Var(Y)} \quad (1)$$

$$SIT_{X_i} = SI1_{X_i} + \sum_{i \neq j} SI2_{X_i X_j} + \ldots + SIk_{X_i \ldots X_k} \quad (2)$$

where

- $Var(Y)$ is the variance of Y;
- $Var(E(Y|X_i)) = E((E(Y|X_i) - E(E(Y|X_i)))^2)$. $Var(E(Y|X_i))$ is the variance of $E(Y|X_i)$ (the expectation of Y conditional on X_i). It represents "the expected reduction in variance that would be obtained if X_i could be fixed" (Saltelli et al. 2010).
- $SIk_{X_i \ldots X_k}$ is the kth-order sensitivity index which represents the sensitivity due to interactions between variables $X_i \ldots X_k$.

If $X_1, X_2, \ldots X_k$ are uncorrelated, the following equation can be written (Saltelli et al. 2010):

$$\sum_i SI1_{X_i} + \sum_i \sum_{j>i} SI2_{X_i X_j} + \ldots + SIk_{X_1 \ldots X_k} = 1 \quad (3)$$

It is important to point out that, when X_i are correlated, sensitivity indices ($SI1_{X_i}$ and SIT_{X_i}) may still be calculated and interpreted in the same way as in the uncorrelated case (Most 2012). When X_i are correlated, sensitivity indices contain two parts: "the correlated contribution (by the correlated variations, i.e. variations of a parameter which are correlated with other parameters) and the uncorrelated contribution (by the uncorrelated variations, i.e. the unique variations of a parameter which cannot be explained by any other parameters)" (Xu and Gertner 2008). However, the correlated and uncorrelated contributions to sensitivity indices cannot be distinguished, unless a dedicated procedure described by Xu and Gertner (2008) is applied. The first-order sensitivity index, which contains both parts, still identifies the most sensitive variables and may thus be used for factor prioritization (Most 2012).

To estimate the first-order sensitivity index, as suggested by Most (2012), we performed a Monte Carlo sampling to obtain 5000 samples $\hat{x}_{ij}$ for the set of input random variables $X = \{X_1, X_2, \ldots X_I\}$ and then computed the 5000 associated values of model output $\hat{y}_j$. The model output Y values may be plotted against the values obtained for a single variable X_i, called scatter plot, that represents $(Y|X_i)$ (see the ESM for an example of a scatter plot). We then sorted the sampled values $\hat{x}_{ij}$ and divided them into 100 subsets. For each subset, we calculated the average value of $\hat{y}_j$ to obtain an estimation of $E(Y|X_i)$, also called smoothed curve. Finally, we obtained an estimation of $Var(E(Y|X_i))$ by calculating the variance of the estimated smoothed curve, which was then used to estimate $SI1_{X_i}$. The estimation of the first-order sensitivity indices for a simple model is provided as examples in the ESM.

2.3.2 Impact categories' ranking

To perform the IC ranking for a product system, we applied a GSA on the following model f_1 called *IC ranking*, which represents the aggregation of midpoint impact categories expressed in endpoint units in the LCIA phase of an LCA. Their contribution to an endpoint impact category is assessed.

$$f_1(Y_{MP_1}, \ldots, Y_{MP_N}) = \sum_i Y_{MP_i} = Y_{EP_j} \quad (4)$$

- Y_{EP_j}: total impact score for one endpoint impact category (EP_j).
- Y_{MP_i}: impact score contribution to the EP_j of the ith midpoint impact category (MP_i).
- N: number of MP_i contributing to an EP_j.

This model is a one-order additive model, so there is no interaction between variables. Inputs Y_{MP_i} are correlated since the same LCI model is used to calculate them and they may have common EFs, so Eq. (3) does not apply. For a product system and for each EP_j, we estimated SI1 for each Y_{MP_i}, as described Section 2.3.1, and ranked them. The higher the $SI1_{Y_{MP_i}}$, the more sensitive MP_i is to EP_j.

2.3.3 LCA phase ranking

Here, we present the method used to estimate the sensitivity of the two groups of LCI and LCIA variables, respectively. LCI input variables may be correlated with each other, as may LCIA input variables. But, in the LCA framework, it is generally assumed that LCI variables are not correlated with LCIA variables. As recommended by Wei et al. (2015), to account for correlation in LCA when performing GSA, we defined uncorrelated groups of correlated variables to estimate the sensitivity indices of each group. This procedure refers to group sensitivities in order to better understand the sensitivity of a group of variables, as described by Jacques et al. (2006). We, therefore, created two uncorrelated groups of variables: X_{LCI} gathering all LCI input variables and X_{LCIA} gathering all CFs. To perform the LCA phase ranking for a product system, we applied a GSA on the following model f_2 called *LCA phase ranking*, which represents the midpoint LCA calculation:

$$f_2(X_{LCI}, X_{LCIA}) = [Y_{MP_i}] \quad (5)$$

Due to the structure of LCA calculation, this model is a two-order model. As input variables are assumed to be uncorrelated, Eq. (3) is true: $SI1_{X_{LCI}} + SI1_{X_{LCIA}} + SI2_{X_{LCI,LCIA}} = 1$ (Saltelli et al. 2010). $SI2_{X_{LCI,LCIA}}$ is the second-order sensitivity index that represents the sensitivity of the model to the interactions between X_{LCI} and X_{LCIA}, i.e., "the part of the variance of Y due to Xi and Xj which is not included in the individual effects of Xi and Xj" (Jacques et al. 2006). Here, if $SI2_{X_{LCI,LCIA}}$ is higher than $SI1_{X_{LCI}} + SI1_{X_{LCIA}}$, no prioritization between X_{LCI} and X_{LCIA} can be made. We will consider that X_{LCI} and X_{LCIA} are equally important and that data collection should be enhanced for both.

As suggested by Jacques et al. (2006), we created two mutated models based on the reference model $Y = f_2(X_{LCI}, X_{LCIA})$ to estimate the first sensitivity index for X_{LCI} and X_{LCIA}: $Y' = f_{2'}(X_{LCI}, X_{LCIA} = fixed)$ is a mutated model where only LCI variables are uncertain because LCIA variables are set to their deterministic value; $Y'' = f_{2''}(X_{LCI} = fixed, X_{LCIA})$ is a mutated model where only LCIA variables are uncertain because LCI variables are set to their deterministic value. By construction, first-order sensitivity indices for each mutated model are equal to 1: $SI1'_{X_{LCI}} = 1$ and $SI1''_{X_{LCIA}} = 1$.

Based on Jacques et al. (2006), first-order (and higher order) sensitivity indices may be estimated for the mutated models based on the value of sensitivity index of the reference model SI_{X_i} and the ratio of the variance between the reference model and the mutated models: $SI'_{X_{LCI}} = SI_{X_{LCI}} * \frac{Var(Y)}{Var(Y')}$ and $SI''_{X_{LCIA}} = SI_{X_{LCIA}} * \frac{Var(Y)}{Var(Y'')}$.

So, for a product system and the selected midpoint ICs, the LCA phase ranking is determined by comparing the values for $SI1_{X_{LCI}}$ and $SI1_{X_{LCIA}}$, which may be estimated by:

$$SI1_{X_{LCI}} = \frac{Var(Y')}{Var(Y)} \tag{6}$$

$$SI1_{X_{LCIA}} = \frac{Var(Y'')}{Var(Y)} \tag{7}$$

with $Var(Y)$ the variance of the Y for the reference model, $Var(Y')$ the variance of the Y' for the corresponding mutated model, and $Var(Y'')$ the variance of the Y'' for the corresponding mutated model. $Var(Y')$ is estimated based on a Monte Carlo simulation addressing only LCI uncertainty, setting LCIA variables to their deterministic value (see Section 2.2 for Monte Carlo settings in case studies). $Var(Y'')$ is estimated based on a Monte Carlo simulation addressing only LCIA uncertainty, setting LCI variables to their deterministic value. $Var(Y)$ is estimated based on a Monte Carlo simulation addressing LCI and LCIA uncertainty simultaneously. In addition, from Eq. (3), we calculated:

$$SI2_{X_{LCI,LCIA}} = 1 - SI1_{X_{LCI}} - SI1_{X_{LCIA}} \tag{8}$$

2.4 Methodology for the meta-analysis per sector

The procedure described in Section 2.1 applies to a product system in a specific study. To assess another product system, the same procedure should be applied from the beginning. Here, we propose a method to test whether the stepwise approach may be generalized to all product systems in a specific economic sector and whether the trends in IC ranking and LCA phase ranking are maintained. If trends are observed and statistically significant, specific recommendations may be formulated for an economic sector regarding the data collection prioritization for regionalization. LCA practitioners and LCI database developers would be able to use the requirements without performing additional sensitivity analyses unless the product system they are interested in contains disruptive technologies or technologies that are poorly represented in the database that describes the economic sectors.

The method we use to perform a meta-analysis of all the product systems in a given sector in an LCI database to test sectorial trends in IC ranking and LCA phase ranking contains four steps.

1. *Uncertainty analysis of all product systems in a sector*: We performed three Monte Carlo simulations per product system. All three simulations rely on the same sampled values at each iteration for LCI variables and CFs: (a) solely addressing LCI uncertainty, setting LCIA variables to their deterministic value; (b) solely addressing LCIA uncertainty, setting LCI variables to their deterministic value; and (c) simultaneously addressing both LCI and LCIA uncertainty. For each simulation and each iteration, we obtained impact scores for all ICs.
2. *GSA to estimate the first-order sensitivity indices of all product systems in a sector*: Using impact scores from (c), we derived SI1 to rank midpoint ICs for each endpoint IC, as described in Section 2.3.2. Using samplings from (a) and (b), we derived SI1 for the groups of LCI variables and LCIA variables to identify the most sensitive LCA phase (LCA or LCIA), as described in Section 2.3.3. Therefore, for each product system, we obtained the sensitivity-based ranking of midpoint ICs and LCA phases.
3. *Statistical tests for IC ranking in a sector*: We used the one-tailed and non-parametric Page's trend test for each endpoint IC to test whether the mean trend of the midpoint IC ranking is statistically significant (Page 1963). We tested the null hypothesis $H_0 : SI1_{MP_k} = SI1_{MP_l}$ for k and l, referring all midpoint ICs contributing to the selected endpoint. If H_0

may be rejected with a level of significance < 0.05, then we assume that the trend for the IC ranking for the sector is $H_1 : Mean(SI1_{MP_1}) > \ldots > Mean(SI1_{MP_k})$ with $Mean(SI1_{MP_k})$ representing the mean value of $SI1_{MP_k}$ among all the product systems in the sector.

4. *Statistical tests for LCA phase ranking in a sector*: We used the one-tailed and non-parametric Wilcoxon signed-rank test for paired data to test whether a trend statistically exists for the LCA phase ranking among all the product systems for the sector for each midpoint IC (Wilcoxon 1945). We tested the null hypothesis $H_0 : SI1_{X_{LCI}} = SI1_{X_{LCIA}}$ for each midpoint IC for the paired data $(SI1_{X_{LCI}}, SI1_{X_{LCIA}})$ of each product system. If H_0 may be rejected with a level of significance < 0.05, then we assume that the trend for the LCA phase ranking for each midpoint IC for the sector is $H_1 : SI1_{X_{LCI}} > SI1_{X_{LCIA}}$ *or* $SI1_{X_{LCI}} < SI1_{X_{LCIA}}$ depending on the sign of the statistics of the test.

3 Results

The results presented in this section illustrate how to apply the methodology proposed in Section 2.1, specifically focusing on step 2 *IC ranking* and step 3 *LCA phase ranking*. For certain cases studies, we (a) identified the most sensitive ICs to be investigated in priority and (b) determined whether data collection should prioritize inventory regionalization or inventory spatialization. First, we applied the methodology to a single product system and then present the results of the meta-analysis of two economic sectors.

3.1 Operationalization of the methodology through a single product system

We selected the *ethanol production from sugarcane*, *BR* product system from the ecoinvent v3.3 database (cutoff version). It belongs to the biofuel production sector that is studied in the next section. Figure 2 illustrates the first-order sensitivity indices (SI1) for the IC ranking model for all midpoint ICs for the two endpoint ICs (EQ and HH). The SI1 calculation considers the uncertainty from LCI and LCIA variables simultaneously. The higher the SI1, the more sensitive the midpoint IC. The results in Fig. 1 clearly show that land transformation IC (EQ) and water availability IC (HH) should be prioritized for regionalization. However, these results should be considered with caution given the way we sampled CFs for the two ICs, i.e., without accounting for spatial correlation (see Section 4.2.3 for the discussion).

Figure 3 shows the sensitivity indices for the LCA phase ranking model for all midpoint ICs for the two endpoint ICs (EQ and HH) for the selected product system. SI1_LCI represents the share of the IC sensitivity due to the group of LCI variables only. SI1_LCIA represents the share of the IC sensitivity due to the group of LCIA variables only. SI2_LCI,LCIA represents the share of the IC sensitivity due to the interactions between LCI and LCIA variables, i.e., sensitivity which is not purely due to LCI variables only or LCIA variables only but due to the term where both are multiplied. The sum of all sensitivity indices for each IC is equal to one since the groups of LCI variables and the group of LCIA variables are uncorrelated. Results show that for both land transformation IC (EQ) and water availability IC (HH), the group of LCIA variables is the most sensitive. Therefore, the regionalization effort should focus on inventory spatialization for both ICs. Again, these results should be considered with caution (see Section 4.2.3 for the discussion). For all regionalized ICs except land occupation IC (EQ), the group of LCIA variables is the most sensitive, and regionalization effort should focus on inventory spatialization. For other ICs, efforts should focus on inventory refinement.

3.2 Meta-analysis of two economic sectors: biofuel production and passenger land transport

Results for the sector meta-analysis are illustrated in Fig. 4 for the biofuel production (left) and passenger land transport (right) sectors for EQ (figures for HH are in the ESM). It represents midpoint ICs sorted in decreasing order of their respective mean values for the sector of first sensitivity index from the IC ranking model, as per Fig. 2. The length of the histograms represents the mean for all product systems within each sector of the first-order sensitivity index from the IC ranking model. Stacked portions in the histograms represent the share of sensitivity due to the groups of LCI variables, the group of LCIA variables and the interactions between LCI and LCIA variables, as per Fig. 3. Each portion is calculated based on the mean for all product systems within each sector of corresponding sensitivity index from the LCA phase ranking model.

Statistical test results for each sector, as described in Section 2.4, are statistically significant with a confidence level of 95 to > 99%. They indicate that there are similar IC and LCA phase rankings across all product systems in each sector. For HH, water availability IC (HH) is the most sensitive midpoint IC for both studied sectors. As the sensitivity mainly stems from the group of LCIA variables, efforts should first focus on spatializing the inventory for this midpoint IC for product systems pertaining to both sectors. For EQ endpoint, the efforts should focus on inventory spatialization of land transformation IC for the biofuel production sector. For the passenger land transport sector, efforts should focus on inventory refinement of global warming and marine acidification ICs. In both sectors, the group of LCIA variables is generally the most sensitive for almost all regionalized ICs, except for land occupation IC for both sectors and land transformation IC for passenger land transport only.

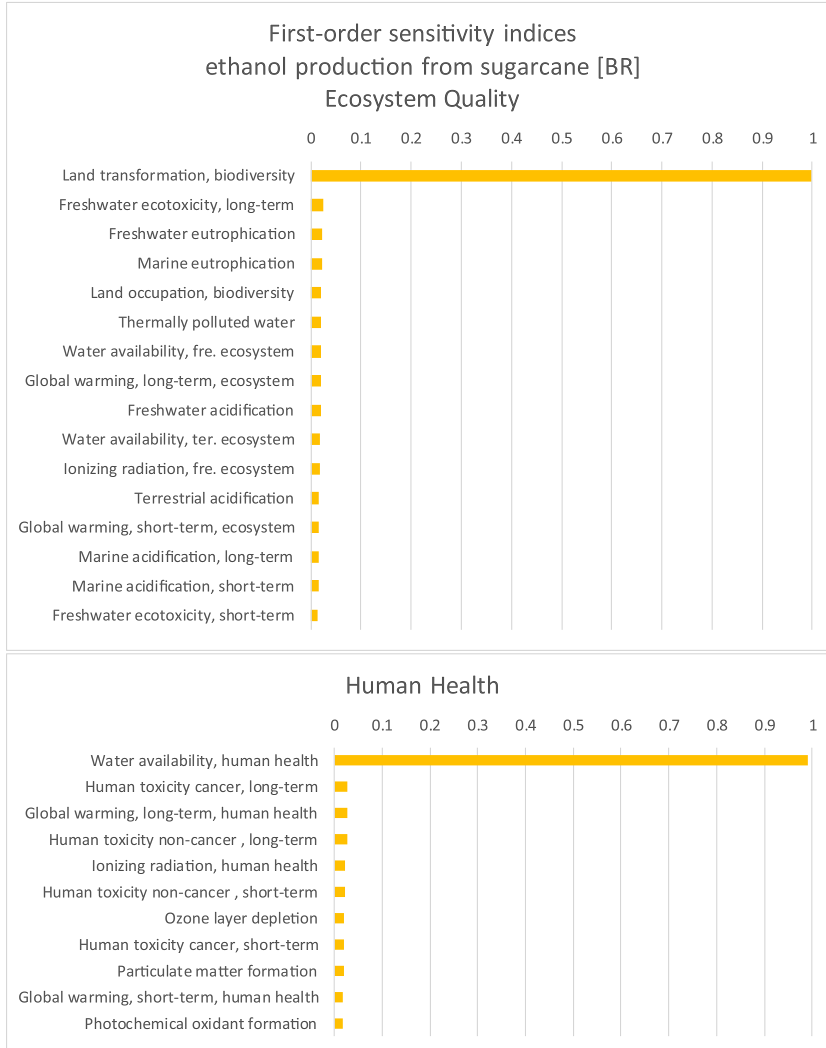

Fig. 2 First-order sensitivity indices for the IC ranking model for midpoint ICs pertaining to EQ endpoint (left) and HH endpoint (right) for the ethanol production from sugarcane, BR product system, sorted in decreasing order

4 Discussion

4.1 Strengths of the methodology

Impact contribution analysis (ICA) is often used to prioritize data collection in an LCA study even though Heijungs (1996) stated that data prioritization should also account for sensitivity in addition to ICA. Figure 5 presents the comparison of IC ranking based on ICA versus GSA for the ethanol production from sugarcane, BR product system. It highlights different rankings depending on the prioritization methodology. The most sensitive midpoint IC is land transformation, whereas this IC is the least contributing midpoint to EQ impact score. In addition, the three most impacting ICs are not ranked as the most sensitive ICs. This figure underlines that prioritizing efforts based on ICA can mislead LCA practitioners by leading them to focus their data collection efforts on ICs that have a low sensitivity index and thus a low potential for uncertainty reduction. As a reminder, only uncertainty due to spatial variability for LCIA variables is accounted here during the GSA. GSA integrating all types of uncertainty will provide different results. In general, there is no reason to expect that the ICA and the GSA provided to same rankings. ICA is a way to decompose the impact score (deterministic results) to identify the origins of the impacts. GSA is a way to decompose the uncertainty (probabilistic results) to identify the origins of the uncertainty. A deterministic value could be low but the associated uncertainty (its sensitivity index for instance) could be high, or vice versa. The deterministic value of a parameter does not reflect the spreading of this parameter. Moreover, ICA does not make it possible to rank LCA phases to prioritize between inventory regionalization and spatialization. GSA appears to be more relevant than ICA to IC ranking and LCA phase ranking for this specific example. The relevance of GSA as compared to ICA for IC ranking should be studied for all the product systems of the different economic sectors.

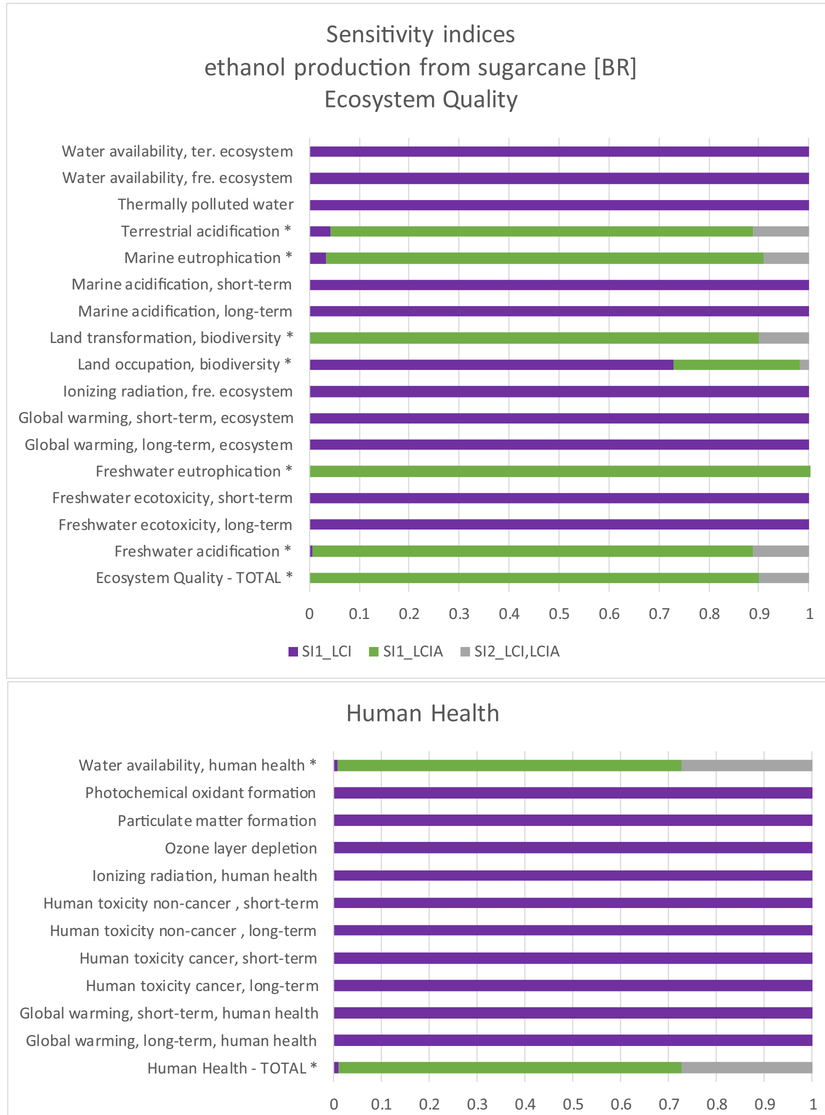

Fig. 3 Sensitivity indices for the LCA phase ranking model for midpoint and endpoint ICs pertaining to EQ endpoint (top) and HH endpoint (bottom) for the ethanol production from sugarcane, BR product system. SI1_LCI: first-order sensitivity index for LCI variables; SI1_LCIA: first-order sensitivity index for LCIA variables; SI2_LCI,LCIA: second-order sensitivity index due to the interactions between SI1_LCI: first-order sensitivity index for LCI and LCIA variables. Regionalized ICs are identified with the asterisk symbol

4.2 Challenges in the operationalization of the methodology

4.2.1 Setting the acceptable level of uncertainty

The first step of the proposed methodology is to set the target level of acceptable uncertainty for decision-makers. It may be challenging to identify and quantify the targets, and this may be related to (i) the confidence level between two scenarios in the comparison or (ii) the distribution of the results themselves when the purpose is not to compare scenarios. (i) may be quantified by using statistical tests between scenarios, taken into account correlations between scenarios if relevant, to assess the robustness of the conclusion and compare the confidence level set by the decision-maker (Henriksson et al. 2015). For instance, scenario A would be considered better than scenario B if the confidence level is higher than 75%. (ii) may be quantified by calculating a statistic measure of the dispersion of the results as, for instance, the coefficient of variation (standard deviation divided by mean value), interquartile ratio, or any metric reflecting the results' uncertainty. The choice of statistic measure of dispersion should be consistent with the properties of the result (underlying distribution, negative results, null mean, etc.).

In both cases, considering LCA practitioners' and decision-makers' lack of knowledge and experience working with statistical tools, setting the target level of acceptable uncertainty may be a challenging task. In addition, this target level must be realistic with regard to the expected level of uncertainty of the different LCA studies. For instance, consequential LCA and prospective LCA are expected to be more uncertain by nature (Herrmann et al. 2014). Setting the acceptable level of uncertainty can also be part of the iterative process in LCA: uncertainty may be reduced step by step, and the decision-

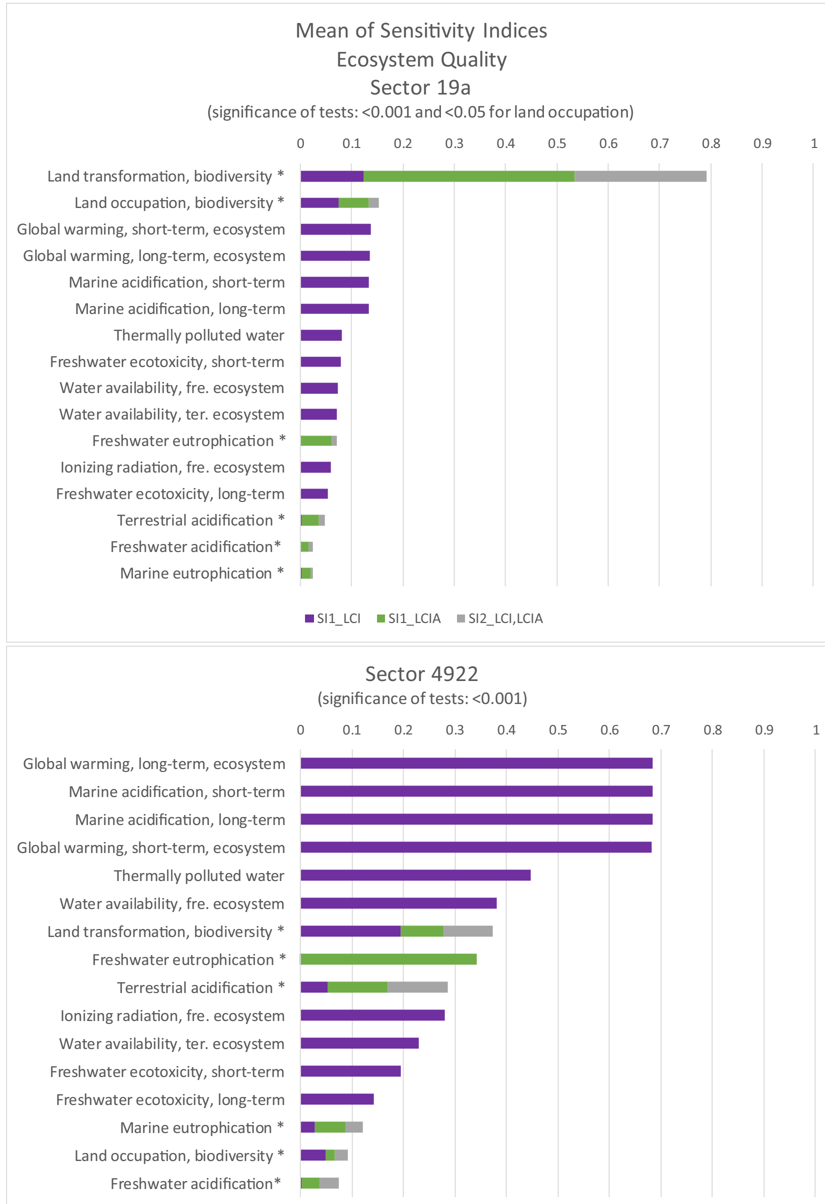

Fig. 4 Mean values of all product systems across the biofuel production sector (top) and the passenger land transport sector (bottom) of the first-order sensitivity index from the IC ranking model for each midpoint ICs pertaining to EQ endpoint. Stacked portions in the histograms represent the contribution of mean values of all product systems across the sector of sensitivity indices from the LCA phase ranking model: SI1_LCI: first-order sensitivity index for LCI variables; SI1_LCIA: first-order sensitivity index for LCIA variables; SI2_LCI,LCIA: second-order sensitivity index due to the interactions between SI1_LCI: first-order sensitivity index for LCI and LCIA variables. Regionalized ICs are identified with the asterisk symbol

maker may decide when he/she is comfortable with the uncertainty level. Further research and experiments on the meanings of acceptable uncertainty for decision-making and how to realistically assess it quantitatively must be conducted.

4.2.2 Uncertainty inputs in LCA

The proposed methodology fully relies on uncertainty assessment. Therefore, the results strongly depend on what type of uncertainty is included in the model and the way this probability is modeled. In this research, we only included parameter uncertainty from LCI and LCIA. We do not account for scenario or model uncertainty of the LCA model. The uncertainty of LCI variables, as defined in ecoinvent v3.3, includes an estimation of their variability and part of their epistemic uncertainty. The variability component of LCI uncertainty contains spatial variability, as well as other variability sources. Therefore, if LCI sensitivity is predominant, additional data collection may be required to obtain more representative data to better describe the spatial variability (inventory regionalization) or more representative data in general, not only focusing on spatial variability. The uncertainty of LCIA variables, as defined in our case studies, only accounts their spatial variability due to the loss of information when aggregating CFs from the native scale to a larger country, continent, or global scale.

To expand the use of our proposed methodology to prioritize data collection for any kind of data, not only spatial data,

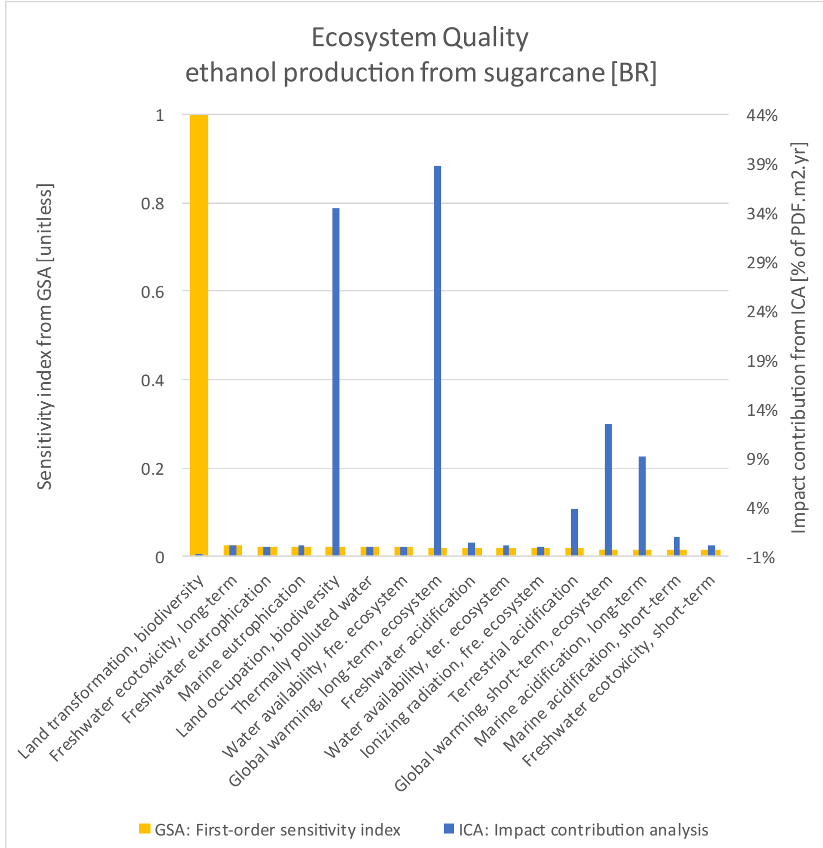

Fig. 5 Comparison of GSA and impact contribution analysis (ICA) for IC ranking for EQ endpoint IC for the ethanol production from sugarcane, BR product system. ICs are sorted in decreasing order for GSA

other sources of uncertainty for LCIA parameters should be added, such as the basic uncertainty of CFs from the LCIA model itself.

4.2.3 Correlations in LCI and LCIA

Correlations in LCA could affect the validity of an uncertainty assessment (Groen and Heijungs 2017). Here are some correlations in LCA.

- LCI correlation
- Physical correlation within the unit process itself, like mass balance. For instance, the CO_2 emissions from the combustion of carbon-based fuel in a vehicle should be correlated with the quantity of fuel.
- LCIA correlation for regionalized ICs
- Spatial correlation at the product system level: Each unit process should have its own location so the location of each unit process in a product system should be decorrelated. Therefore, different sets of CFs for each unit process in the product system should be used for each Monte Carlo iteration.
- Spatial correlation between ICs (inter-ICs): For a unit process, the chosen random region during one iteration should be consistent across all regionalized ICs, i.e., ICs should be spatially correlated.
- Spatial correlation between CFs within an IC (intra-CFs): For a unit process and for one IC, the chosen random region during one iteration should be consistent across all regionalized CFs, i.e., CFs should be spatially correlated.

We only accounted for certain correlations in the uncertainty assessment of our case studies. Regarding LCI correlation, we only considered the physical correlation in LCI unit processes for water elementary flows and land transformation elementary flows. Regarding LCIA correlation, the same set of CFs is used for all unit processes in a product system during one Monte Carlo iteration because the inventory is aggregated before characterization (traditional LCA calculation). This is one of the limitations of our case study as we ignore the spatial correlation at the product system level. In addition, CFs are independently sampled between ICs and within IC during Monte Carlo analysis. Because we only considered spatial variability in the CF, the random sampling of a CF is basically identical to choosing a CF for a given region. Therefore, the location of the unit process is not consistent across ICs and across CFs for a given IC. Therefore, we ignore the spatial correlation inter-ICs and intra-ICs. The assumption of independent CF sampling may explain the high sensitivity of land transformation and water availability. Indeed, the respective elementary flows entering and leaving the same unit process are characterized by two different CFs. We only account for the intra-CF spatial correlation for the elementary flows of a unit process associated with the same land use type for land transformation because it is straightforward to implement (the same CF value is used at each iteration for the same land use type, we just switch its sign from *from* to *to*). Implementing intra-CF spatial correlation for other elementary flows for land transformation and water availability is more challenging. Indeed, water flows entering and leaving a unit process are not released in the same compartment as are characterized by different CFs. In the same way, the land is rarely transformed from one land use type to the same land use type. When there is a change in land use type, different CFs that are not from the same distribution should be used. We would have had to use a GIS software and invest in efforts beyond the resources available for this study. A way to solve this issue would be that LCI database developers provide net inventory flows.

One way to test whether intra-CF spatial correlation is important to implement would be to test the distribution of the LCIA sensitivity among {EF|unit process} for land transformation and water availability. To do so, it is possible to calculate an inequity measure like the Gini coefficient, as done by Reinhard et al. (2016), for the distribution of the impact contribution between unit processes. If the inequity measure is low, the sensitivity is equally spread between {EF|unit process} and intra-CF spatial correlation has a significant influence and should be implemented. This should be tested in the future.

The best way to account for LCIA spatial correlations is to perform Monte Carlo using regionalized LCA calculation (Mutel and Hellweg 2009; Mutel et al. 2012). However, this increases computational time. Further research and tool adaptation is needed to properly address correlations in LCA, especially regarding spatial correlation when performing regionalized LCA. It could be a way to reduce the model uncertainty in LCA. In addition, seeing as the same LCI data is used to calculate all ICs, regionalizing the inventory for one IC may lower the uncertainty of other ICs.

4.2.4 Prioritization of additional data collection for inventory regionalization and inventory spatialization

GSA performed in our case studies focuses on the second step (IC ranking) and the third step (LCA phase ranking) of the proposed stepwise methodology. In the fourth step of the proposed methodology, data that requires further collection for inventory regionalization and inventory spatialization should be identified. As explained before, we recommend carrying out the fourth step by performing a GSA. As the second and third steps are based on GSA, in the fourth step, the practitioner can focus on the most sensitive parts of the LCA model, i.e., on selected ICs and on LCI or LCIA variables. It may considerably reduce the number of input variables to test for sensitivity during the GSA in the fourth step. However, the model on which the fourth step GSA should be performed still has hundreds to thousands of input variables with a high order of interactions between them. Therefore, estimating sensitivity indices for this model could still be computationally intensive (Groen et al. 2017). In addition, LCA practitioners may not have the time or tools to perform a GSA. If it is the case and if the IC and LCA phases have been prioritized based on GSA, ICA could be used for the fourth step since it is less time consuming and already implemented in all LCA software and because the LCA practitioner would at least be focusing on the most sensitive part of the model.

4.2.5 Requirements for implementation in current practices

To operationalize our methodology, uncertainty sources from both LCI and LCIA variables should be considered. Most available LCA software makes it possible to account for the uncertainty from LCI variables during Monte Carlo sampling. However, to our knowledge, probability distributions for LCIA variables can currently only be set with openLCA and Brightway.

Besides, uncertainty information on CFs is often missing in LCIA methods or provided in an unusable format. When provided, often only the extrema, the mean, or the standard deviation are given. This type of information is not sufficient for Monte Carlo sampling as implemented in software since software requires probability distributions and is often restricted to limited standard probability distribution functions (uniform, triangular, normal, lognormal, etc.). In some cases, it is challenging to perfectly fit a probability distribution on the histogram of sampled CF. For instance, for some IC in the IMPACT World+ method, the histogram representing the sampled CF is multimodal. New computational approaches that directly sample the values from the CF histogram would solve this issue if LCIA developers provided the histogram with the originally sampled CF.

GSA for IC ranking (step 2) and LCA phase ranking (step 3) is divided into two tasks: (1) the uncertainty analysis based on Monte Carlo simulation and (2) the calculation of sensitivity indices, as described in Section 2.3.1. The computational time to analyze one product system with GSA for steps 2 and 3 only is less than a minute using Brightway, mainly due to the computational time for the Monte Carlo sampling. The estimation of sensitivity indices for the IC ranking and LCA phase ranking models is quick since the models are simple (one-order model and two-order uncorrelated model) and have very few input variables. Computationally speaking, we proved that steps 2 and 3 of our methodology could be implemented in standard LCA software to enhance the interpretation of LCA results with regard to sensitivity analysis and guide LCA practitioners in an efficient data collection effort. The fourth step of the methodology requires a GSA that may involve hundreds to thousands of variables to be tested. The time to run a GSA increases with the number of input variables and the level of interactions between them and further development may, therefore, be required to reduce computational time in GSA before implementation in standard LCA software.

Computational time for meta-analysis of a sector is longer, as Monte Carlo sampling should be performed for all product systems belonging to the same sector. Implementing other sampling techniques alternative to Monte Carlo samplings, like Latin hypercube or quasi-Monte Carlo sampling, could lower the computational time (Groen et al. 2014). Another avenue to improve the computational time to precompute once Monte Carlo impact scores for a database with an LCIA methodology allowing LCA practitioner to reuse Monte Carlo impact scores without performing the Monte Carlo simulation themselves (Lesage et al. 2018).

4.3 Discussion on the results of the case studies

4.3.1 Use and limits of results for the meta-analysis per sector

The results of the sector-specific meta-analyses aim to provide general recommendations for IC ranking and LCA phase ranking that may be used for any product system in the sector without performing a specific analysis for the product system. Pre-calculated results for all sectors would enable LCA practitioners to prioritize their regionalization effort for any product system in the analyzed sectors without needing to perform a GSA themselves. However, if the analyzed product system is poorly represented in the original database describing the sector or if the product system is a disruptive technology, the meta-analysis may not provide representative results.

The rank of the most sensitive IC and LCA phase depends on the LCI database and LCIA methodology. It would be relevant to apply the meta-analysis to different LCI databases and LCIA methodologies to test the robustness and/or complete the results per sector. Unfortunately, most LCI databases do not provide the uncertainty information required.

4.3.2 Importance of integration of LCIA regionalization

The meta-analysis of the two sectors shows that inventory spatialization should be prioritized for almost all regionalized ICs because CFs are more sensitive variables than LCI variables. This suggests the need for LCA software developers and LCIA method developers to implement and facilitate the use of regionalized LCIA methodology. It also highlights the need to include the spatial variability of regionalized CFs when performing uncertainty or sensitivity analysis in LCA. Nevertheless, this recommendation should be better supported by a systematic analysis of all sectors in an LCI database.

4.4 Potential adaptation for future work

The proposed methodology was designed to prioritize regionalization efforts by accounting for uncertainty related to spatial variability. This approach may be adapted to any type of uncertainty by developing a similar methodology based on GSA, including all uncertainty sources. Steps 1, 2, and 3 would remain relevant for this methodology. The methodology could also be adapted for normalization and weighting.

Here, we performed the meta-analysis differentiating product systems in their economic sectors. Performing a meta-analysis differentiating the product systems based on geographic location (e.g., their country) may help identify potential patterns regarding the need for data collection depending on geographic location. It may be a complementary avenue to provide specific recommendations for data collection.

5 Conclusions

We proposed a stepwise methodology for LCA practitioner to prioritize data collection for regionalization purposes based on global sensitivity analysis (GSA). It makes it possible to rank the most sensitive impact categories (ICs) within a given endpoint (step 2) and identify whether the LCI or LCIA group of variables is the most sensitive LCA phase (step 3). We recommend using the methodology to efficiently prioritize regionalization efforts between ICs and between inventory regionalization and inventory spatialization.

The use of GSA instead of impact contribution analysis (ICA) to prioritize the data collection effort enables modelers to focus on the most sensitive data with the highest potential for uncertainty reduction. It also makes it possible to prioritize efforts between inventory regionalization and inventory spatialization, which is not feasible with ICA. We proved that the implementation of steps 2 and 3 of our methodology is computationally feasible and therefore invite current LCA software providers to implement it. However, further improvements, such as accounting for spatial correlations and better computational times for GSA, are required to improve LCA calculation and interpretation.

We also demonstrated through a meta-analysis of two sectors that it is possible to derive sectorial recommendations regarding the ICs and LCA phases that should constitute regionalization priorities. By expanding the analysis to all sectors in an LCI database, one could derive recommendations to support LCA practitioners and LCI database developers in defining their strategies for regional data collection to lower the uncertainty of LCA results. The proposed methodology is primarily designed for regionalization purposes, i.e., to reduce the uncertainty related to spatial variability. Nevertheless, it may be adapted to other types of uncertainty to help prioritize data collection efforts.

Acknowledgements We acknowledge the financial and technical support of IFP Energies nouvelles and of the industrial partners of the International Chair in Life Cycle Assessment (a research unit of the CIRAIG): Arcelor-Mittal, Hydro-Québec, LVMH, Michelin, Nestlé, Solvay, Optel Vision, Total, Umicore.

References

Baitz M, Albrecht S, Brauner E et al (2012) LCA's theory and practice: like ebony and ivory living in perfect harmony? Int J Life Cycle Assess 18:5–13

Borgonovo E, Plischke E (2016) Sensitivity analysis: a review of recent advances. Eur J Oper Res 248:869–887

Boulay A, Bulle C, Bayart J-B et al (2011) Regional characterization of freshwater use in LCA: modeling direct impacts on human health. Environ Sci Technol 45:8948–8957

Boulay AM, Bare J, Benini L, Berger M, Lathuillière MJ, Manzardo A, Margni M, Motoshita M, Núñez M, Pastor AV, Ridoutt B, Oki T, Worbe S, Pfister S (2018) The WULCA consensus characterization model for water scarcity footprints: assessing impacts of water consumption based on available water remaining (AWARE). Int J Life Cycle Assess 23:368–378

Bulle C, Margni M, Patouillard L, Boulay AM, Bourgault G, de Bruille V, Cao V, Hauschild M, Henderson A, Humbert S, Kashef-Haghighi S, Kounina A, Laurent A, Levasseur A, Liard G, Rosenbaum RK, Roy PO, Shaked S, Fantke P, Jolliet O (2019) IMPACT World+: a globally regionalized life cycle impact assessment method. Int J Life Cycle Assess. https://doi.org/10.1007/s11367-019-01583-0

Chaudhary A, Verones F, de Baan L, Hellweg S (2015) Quantifying land use impacts on biodiversity: combining species-area models and vulnerability indicators. Environ Sci Technol 49:9987–9995

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

Collet P, Lardon L, Steyer J-P, Hélias A (2014) How to take time into account in the inventory step: a selective introduction based on sensitivity analysis. Int J Life Cycle Assess 19:320–330

Gregory JR, Noshadravan A, Olivetti EA, Kirchain RE (2016) A methodology for robust comparative life cycle assessments incorporating uncertainty. Environ Sci Technol 50:6397–6405

Groen EA, Heijungs R (2017) Ignoring correlation in uncertainty and sensitivity analysis in life cycle assessment: what is the risk? Environ Impact Assess Rev 62:98–109

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

Groen EA, Bokkers EAM, Heijungs R, de Boer IJM (2017) Methods for global sensitivity analysis in life cycle assessment. Int J Life Cycle Assess 22:1125–1137

Heijungs R (1996) Identification of key issues for further investigation in improving the reliability of life-cycle assessments. J Clean Prod 4: 159–166

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

Helmes RJK, Huijbregts MAJ, Henderson AD, Jolliet O (2012) Spatially explicit fate factors of phosphorous emissions to freshwater at the global scale. Int J Life Cycle Assess 17:646–654

Henriksson PJG, Heijungs R, Dao HM, Phan LT, de Snoo GR, Guinée JB (2015) Product carbon footprints and their uncertainties in comparative decision contexts. PLoS One 10:1–11. https://doi.org/10.1371/journal.pone.0121221

Hernández-Padilla F, Margni M, Noyola A, Guereca-Hernandez L, Bulle C (2017) Assessing wastewater treatment in Latin America and the Caribbean: enhancing life cycle assessment interpretation by regionalization and impact assessment sensibility. J Clean Prod 142:2140–2153

Herrmann IT, Hauschild MZ, Sohn MD, McKone TE (2014) Confronting uncertainty in life cycle assessment used for decision support. J Ind Ecol 18:366–379

Huijbregts MAJ (1998) Uncertainty in LCA LCA methodology application of uncertainty and variability in LCA part I: a general framework for the analysis of uncertainty and variability in life cycle assessment. Int J Life Cycle Assess 3:273–280

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

Igos E, Meyer R, Benetto E et al (2015) Uncertainty and sensitivity analyses in LCA: a review and application to noise characterization. In: SETAC Europe 25th Annual Meeting

Imbeault-Tétreault H (2010) Propagation analytique de l'incertitude à travers le calcul matriciel d'une analyse du cycle de vie

International Organization for Standardization (ISO) (2006a) ISO14040: 2006 Environmental management-life cycle assessment-principles and framework

International Organization for Standardization (ISO) (2006b) ISO14044: 2006 Environmental management—life cycle assessment—requirements and guidelines

Iooss B, Lemaître P (2015) A review on global sensitivity analysis methods. In: Dellino G., Meloni C. (eds) Uncertainty management in simulation-optimization of complex systems. Operations Research/Computer Science Interfaces Series, vol 59. Springer, Boston, MA, pp 101–122

Jacques J, Lavergne C, Devictor N (2006) Sensitivity analysis in presence of model uncertainty and correlated inputs. Reliab Eng Syst Saf 91: 1126–1134

Laurin L, Amor B, Bachmann TM, Bare J, Koffler C, Genest S, Preiss P, Pierce J, Satterfield B, Vigon B (2016) Life cycle assessment capacity roadmap (section 1): decision-making support using LCA. Int J Life Cycle Assess 21:443–447

Lesage P, Samson R (2016) The Quebec life cycle inventory database project. Int J Life Cycle Assess 21:1282–1289

Lesage P, Mutel C, Schenker U, Margni M (2018) Uncertainty analysis in LCA using precalculated aggregated datasets. Int J Life Cycle Assess 23:2248–2265

Most T (2012) Variance-based sensitivity analysis in the presence of correlated input variables. Conference: 5th International Conference on Reliable Engineering Computing, At Brno, Czech Republic

Muller S, Lesage P, Ciroth A, Mutel C, Weidema BP, Samson R (2016) The application of the pedigree approach to the distributions foreseen in ecoinvent v3. Int J Life Cycle Assess 21:1327–1337

Mutel C (2017) Brightway: an open source framework for life cycle assessment. J Open Source Softw 2:doi: https://doi.org/10.21105/joss.00236

Mutel CL, Hellweg S (2009) Regionalized life cycle assessment: computational methodology and application to inventory databases. Environ Sci Technol 43:5797–5803

Mutel CL, Pfister S, Hellweg S (2012) GIS-based regionalized life cycle assessment: how big is small enough? Methodology and case study of electricity generation. Environ Sci Technol 46:1096–1103
Mutel CL, de Baan L, Hellweg S (2013) Two-step sensitivity testing of parametrized and regionalized life cycle assessments: methodology and case study. Environ Sci Technol 47:5660–5667
Padey P, Girard R, le Boulch D, Blanc I (2013) From LCAs to simplified models: a generic methodology applied to wind power electricity. Environ Sci Technol 47:1231–1238
Page EB (1963) Ordered hypotheses for multiple treatments: a significance test for linear ranks. J Am Stat Assoc 58:216–230
Patouillard L, Bulle C, Margni M (2016) Ready-to-use and advanced methodologies to prioritise the regionalisation effort in LCA. Mater Tech 104:105
Patouillard L, Bulle C, Querleu C, Maxime D, Osset P, Margni M (2018) Critical review and practical recommendations to integrate the spatial dimension into life cycle assessment. J Clean Prod 177:398–412
Pfister S, Scherer L (2015) Uncertainty analysis of the environmental sustainability of biofuels. Energy Sustain Soc 5:1–12
Pfister S, Koehler A, Hellweg S (2009) Assessing the environmental impacts of freshwater consumption in LCA. Environ Sci Technol 43:4098–4104
Plouffe G, Bulle C, Deschênes L (2015) Assessing the variability of the bioavailable fraction of zinc at the global scale using geochemical modeling and soil archetypes. Int J Life Cycle Assess 20:527–540
Potting J, Hauschild M (2006) Spatial differentiation in life cycle impact assessment: a decade of method development to increase the environmental realism of LCIA. Int J Life Cycle Assess 11:11–13
Refsgaard JC, van der Sluijs JP, Højberg AL, Vanrolleghem PA (2007) Uncertainty in the environmental modelling process—a framework and guidance. Environ Model Softw 22:1543–1556
Reinhard J, Mutel CL, Wernet G, Zah R, Hilty LM (2016) Contribution-based prioritization of LCI database improvements: method design, demonstration, and evaluation. Environ Model Softw 86:204–218
Riggs LS (1989) Numerical approach for generating beta random variates. J Comput Civ Eng 3:183–191
Ross S, Evans D (2002) Excluding site-specific data from the LCA inventory: how this affects life cycle impact assessment. Int J Life Cycle Assess 7:141–150
Roy P-O, Deschênes L, Margni M (2013) Uncertainty and spatial variability in characterization factors for aquatic acidification at the global scale. Int J Life Cycle Assess 19:882–890
Sakai S, Yokoyama K (2002) Formulation of sensitivity analysis in life cycle assessment using a perturbation method. Clean Techn Environ Policy 4:72–78
Saltelli A (2017) Sensitivity analysis. Numbers policy Pract Probl Quantif
Saltelli A, Sobol' IM (1995) About the use of rank transformation in sensitivity analysis of model output. Reliab Eng Syst Saf 50:225–239
Saltelli A, Tarantola S (2002) On the relative importance of input factors in mathematical models. J Am Stat Assoc 97:702–709
Saltelli A, Annoni P, Azzini I, Campolongo F, Ratto M, Tarantola S (2010) Variance based sensitivity analysis of model output. Design and estimator for the total sensitivity index. Comput Phys Commun 181:259–270
Sobol IM (1993) Sensitivity estimates for nonlinear mathematical models. Math Model Comput Exp 1:407–414
Turconi R, Boldrin A, Astrup T (2013) Life cycle assessment (LCA) of electricity generation technologies: overview, comparability and limitations. Renew Sust Energ Rev 28:555–565
Udo de Haes HA, Jolliet O, Finnveden G et al (1999) Best available practice regarding impact categories and category indicators in life cycle impact assessment. Int J Life Cycle Assess 4:66–74
United Nations Statistics Division (2008) ISIC Rev.4 - Classifications Registry. In: 2008. https://unstats.un.org/unsd/cr/registry/isic-4.asp. Accessed 27 Feb 2018
Vigon BW, Tolle DA, Cornaby BW et al (1993) Life-cycle assessment: inventory guidelines and principles, EPA/600/R-. Washington, DC20460
Wardekker JA, van der Sluijs JP, Janssen PHM, Kloprogge P, Petersen AC (2008) Uncertainty communication in environmental assessments: views from the Dutch science-policy interface. Environ Sci Pol 11:627–641. https://doi.org/10.1016/j.envsci.2008.05.005
Wei W, Larrey-Lassalle P, Faure T, Dumoulin N, Roux P, Mathias JD (2015) How to conduct a proper sensitivity analysis in life cycle assessment: taking into account correlations within LCI data and interactions within the LCA calculation model. Environ Sci Technol 49:377–385
Weidema BP, Wesnæs MS (1996) Data quality management for life cycle inventories—an example of using data quality indicators. J Clean Prod 4:167–174
Wender BA, Prado V, Fantke P, Ravikumar D, Seager TP (2018) Sensitivity-based research prioritization through stochastic characterization modeling. Int J Life Cycle Assess 23:324–332
Wernet G, Bauer C, Steubing B, Reinhard J, Moreno-Ruiz E, Weidema B (2016) The ecoinvent database version 3 (part I): overview and methodology. Int J Life Cycle Assess 21:1218–1230
Wilcoxon F (1945) Individual comparisons by ranking methods. Biom Bull 1:80
Xu C, Gertner GZ (2008) Uncertainty and sensitivity analysis for models with correlated parameters. Reliab Eng Syst Saf 93:1563–1573
Yang Y, Tao M, Suh S (2018) Geographic variability of agriculture requires sector-specific uncertainty characterization. Int J Life Cycle Assess 23:1581–1589

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.