lca_algebraic: a library bringing symbolic calculus to LCA for comprehensive sensitivity analysis

Raphaël Jolivet¹ · Julie Clavreul² · Raphaël Brière² · Romain Besseau¹ · Anne Prieur Vernat² · Marie Sauze³ · Isabelle Blanc¹ · Mélanie Douziech¹ · Paula Pérez-López¹

Received: 17 June 2021 / Accepted: 23 October 2021 / Published online: 20 November 2021
© The Author(s) 2021

Abstract

Purpose In this paper, we present new tools to ease the analysis of the effect of variability and uncertainty on life cycle assessment (LCA) results.

Methods The tools consist of a standard protocol and an open-source library: *lca_algebraic*. This library, written in *Python* and based on the framework *Brightway2* (Mutel in J Open Source Softw 2(12):236, 2017) provides functions to support sensitivity analysis by bringing symbolic calculus to LCA. The use of symbolic calculus eases the definition of parametric inventories and enables a very fast evaluation of impacts by factorizing the background activities. Thanks to this processing speed, a large number of Monte Carlo simulations can be generated to evaluate the variation of the impacts and apply advanced statistic tools such as Sobol indices to quantify the contribution of each parameter to the final variance (Sobol in Math Comput Simul 55 (1–3):271–280, 2001). An additional algorithm uses the key parameters, identified from their high Sobol indices, to generate simplified arithmetic models for fast estimates of LCA results.

Results and discussion The protocol and library were validated through their application to the assessment of impacts of mono crystalline photovoltaic (PV) systems. A comprehensive sensitivity analysis was performed based on the protocol and the complementary functions provided by *lca_algebraic*. The proposed tools helped building a detailed parametric reference LCA model of the PV system to identify the range of variation of multi-criterion LCA results and the key foreground-related parameters explaining these variations. Based on these key parameters, we generated simplified arithmetic models for quick and simple multi-criteria environmental assessments to be used by non-expert LCA users. The resulting models are both compact and aligned with the reference parametric LCA model of crystalline silicon PV systems.

Conclusion This work brings powerful and practical tools to the LCA community to better understand, identify, and quantify the sources of variation of environmental impacts and produce simplified models to spread the use of LCA among non-experts. The library mainly explores the uncertainties of the foreground activities. Further work could also integrate the uncertainty of background activities, described, for example, by pedigree matrices.

Keywords Simplified LCA models · Parametrization · Variability · Uncertainty · Global sensitivity analysis · Symbolic calculus · Python · Multi-criteria

1 Introduction

Renewable energies are expected to develop rapidly in the next decades to help reduce greenhouse gas emissions and therefore limit global climate change (Eldenhofer et al. 2012). In that context, authorities and industrial stakeholders increasingly request accurate assessments of the potential environmental impacts of energy systems to avoid any burden-shifting when replacing existing energy systems by renewable ones and provide decision support (Parliament 2014). Such assessments typically rely on life cycle assessment (LCA), a

Communicated by Matthias Finkbeiner.

✉ Raphaël Jolivet
raphael.jolivet@mines-paristech.fr

¹ Centre Observation, Impacts, Energie (O.I.E.), MINES ParisTech, PSL Research University, Sophia Antipolis, France

² ENGIE Lab CRIGEN, Paris, France

³ ADEME, Sophia Antipolis, France

methodology estimating the potential environmental impacts of a system over all its life cycle stages (ISO 14040, 2006).

However, the important spread of LCA results complicates decision making and may affect the confidence in this analytical method. For example, for mono crystalline PV system, the Intergovernmental Panel on Climate Change (IPCC) reports a carbon footprint ranging from 30 to 215 g CO_{2eq}/kWh (Eldenhof et al. 2012).

Walker et al. (2003) distinguishes two types of uncertainty, influencing the spread in LCA results:

- Epistemic uncertainty is due to a lack of knowledge of the system or bad accuracy of the model. It can be reduced by acquiring more data and refining mathematical models (Clavreul et al. 2013; Igos et al. 2018).
- Stochastic uncertainty is inherent to the variability of the system and cannot be reduced with a single static model. It corresponds to various configurations and technical choices within a single sector.

Stochastic uncertainty can be accounted for by describing the system using well-defined parameters (Wei et al. 2014). Reference LCA models provide such a parametrized description of a system. They allow sensitivity analyses that quantify the effects of parameter variation on the LCA result (Pianosi et al. 2016). Several approaches are available for sensitivity analysis, varying in complexity and accuracy:

- Local analysis, which consists in varying one parameter at a time (OAT). It is simple and fast to carry out, but shows several limitations: It does not provide a quantitative assessment of the variance of the impacts and may hide the combined effect of several parameters.
- Global sensitivity analysis (GSA), on the contrary, is based on simultaneous variations of all parameters. It requires a good knowledge of the distributions of all input parameters and much more computing power.

The knowledge of the parameters influencing the LCA results the most, thus key parameters, is important to decision makers. It allows them to focus on a reduced number of parameters to (1) reduce the uncertainty of the environmental impacts by improving the parameters' characterization and (2) reduce the environmental impacts themselves by modifying these parameters at the design stage. The key parameters also allow to build simplified parametric models: These are short arithmetic formulas based on a few parameters, suitable for rapid environmental impact assessments by non-experts.

Classic LCA software does not provide, to our knowledge, advanced statistical tools for the application of GSA methods. To overcome these limitations, a growing community of LCA experts is now using programming languages, such as *Python* or *R* together with libraries

dedicated to LCA, such as *Brightway2* (Mutel 2017). While they provide higher flexibility for custom analysis, these tools require learning a programming language and specific syntax of custom libraries.

In that context, the INCER-ACV project, funded by the French Agency ADEME, provided practical methodology and tools for the LCA community to build reference LCA models, apply sensitivity analyses, and generate simplified models. INCER-ACV was built upon previous work by Padey et al. (2013), who first provided guidelines to develop parametrized inventories and explored how GSA could be used to identify key parameters.

In practice, we wrote a protocol and developed a new library, *lca_algebraic*, to generalize and ease the analysis of uncertainties and variability in LCA by bringing symbolic calculus to *Brightway2*. Traditional LCA software uses a procedural approach with direct numerical processing to calculate impacts. Symbolic calculation uses a declarative definition and manipulates full arithmetic formulas for deferred evaluation.

This novel approach constitutes the innovation of this article and adds several benefits:

- The use of symbolic calculus enables simple and compact declarative definition of parametric inventories, by using standard *Python* expressions and without learning any new technical interface.
- It permits the factorization of the impact of background activities, allowing fast computation of environmental impacts. This processing speed is key in the processing of GSA, requiring very large number of Monte Carlo simulations.
- The usage of symbolic calculus enables to automatically generate simplified models, based on the key parameters identified by the GSA. This method is automatic and systematic and does not require custom development or regression like simplified models developed by Padey et al. (2013) or Lacirignola et al. (2014).
- Our approach enables a fine tuning of the balance between the simplicity of the model and its accuracy, driven by the choice of the share of variability to be explained and the overlap between the reference LCA model and the simplified models.
- Finally, we propose a practical tool, ready to be used by the practitioners.

This protocol and the library have been used successfully in three case studies, carried out by academic and industrial actors, namely, photovoltaic systems (Besseau 2019; Pérez-López et al. 2020), floating wind turbines (Pérez-López et al. 2020), and geothermal energy (Douziech et al. 2021). The application to all the case studies made it possible to refine these tools by integrating the feedback of users.

Contrary to the above mentioned studies already published, our aim with this paper is to highlight the strengths of the newly developed library and explore how this novel approach enables to thoroughly analyze the effect of uncertainties and variability of input parameters in LCA results, and to develop simplified LCA models. (Besseau 2019; Pérez-López et al. 2020; Pérez-López et al. 2020; Douziech et al. 2021) focused on developing parameterized LCA models, and sometimes simplified models, without explaining the *lca_algebraic* library itself nor exploring the different methodologies for sensitivity analysis.

2 Methods

In this section, we first provide an overview of the protocol. Then we present the *lca_algebraic* library and its main features. Finally, we detail each step of the protocol and the implementation with the library.

2.1 Overview of the protocol

In this article, we propose a protocol to conduct comprehensive sensitivity analysis of LCA studies based on five steps, from the definition of the scope of the study, to the generation of simplified models. This protocol extends the work of Padey et al. (2013), providing generic methodology to build simplified models. It is fully compatible with the four steps of the application of LCA according to ISO 14040 (2006), namely, (i) goal and scope definition, (ii) inventory analysis, (iii) life cycle impact assessment, and (iv) interpretation.

Figure 1 gives an overview of the protocol detailed in the technical report of the project INCER-ACV (ADEME, grant no. 1705C0045) (Pérez-López et al. 2020). The protocol starts by defining the scope of the study (1) which requires setting the boundaries of the system under study and the environmental impacts to be considered. Then, a reference LCA model is built (2). The distribution of the input parameters of the model are characterized (3). Local and global sensitivity analyses are conducted on this reference LCA model (4). The GSA provides a quantification of the influence of each parameter's variation on the total variability of the environmental impact results. We chose Sobol indices as the method for GSA, as it provides accurate results, even for nonlinear models and large uncertainty (Groen et al. 2016). Finally, we propose an algorithm to select the key parameters and transform the reference parametrized model into a simplified arithmetic expression (5).

The application of this protocol provides three main outcomes: (a) an accurate estimate of the environmental impacts and their variability described by a statistical distribution; (b) a selection of the key input parameters, explaining most of the variability; and (c) simplified models for

rapid estimation of environmental impacts of a specific system within the domain defined in step (1).

2.2 lca_algebraic library

We developed the *Python* library *lca_algebraic* to support each step of the protocol. This library is published under the open-source BSD license. It is freely available on the internet, together with instructions for its installation and an example script (https://github.com/oie-mines-paristech/lca_algebraic/). It is based on two other open source libraries: *Brightway2*, a common tool for LCA, and *Sympy*, a library for symbolic calculus.

The *lca_algebraic* library uses algebraic computation instead of numerical evaluation. This choice provides several benefits for the implementation of the protocol. The functions provided by *lca_algebraic* serve to build the models needed to carry out the inventory analysis and impact assessment phases of LCA application, as well as to facilitate sensitivity analyses required in the interpretation phase.

2.2.1 Construction of the reference parametric model

The construction of the reference parametric model (step 2 in Fig. 1) of the life cycle inventory is eased by algebraic calculus. The parameters are defined as *Sympy* symbols. Unlike standard *Python* variables, expressions involving *Sympy* symbols will result in the definition of a symbolic formula, instead of a direct numerical evaluation. The complete formulas are stored in the inventory database. This enables a seamless and compact definition of parametric inventories for users, by writing standard *Python* expressions and without the need for additional technical syntax.

2.2.2 Fast sensitivity analyses

The use of symbolic calculus enables the optimization of the computation of environmental impacts, by factorizing the background activities. The processing speed is key for the application of global sensitivity analyses, based on large Monte Carlo simulations.

Figure 2 illustrates the internal details of the calculation of the environmental impacts.

A parametrized life cycle inventory can be represented as a tree of technical activities. Each of them may consume and emit elementary substances and use products of other activities (material or energy). Each activity can thus be modeled as a linear combination of exchanges with other activities and elementary substances. The amounts of these exchanges may be quantified by fixed or parametrized values.

We distinguish between foreground (F_i) and background (B_j) activities. The foreground activities are those within the

Fig. 1 Overview of the protocol for sensitivity analysis in LCA

control of the designer of the system (Frischknecht 1998). They are the ones which may be parametrized. The background activities are the resources and sub-parts used by the foreground activities. The decision-maker has no control over them.

When the calculation of the environmental impacts is requested, the library single function of both the input parameters (P_i) and the background activities (B_j) as formalized in Eq. 1.

$$\text{Inventory} = f(P_i, B_j) \quad (1)$$

Second, the library computes, for each of the considered impact categories, the environmental impact of the background activities (B_j), by calling the chosen life cycle impact assessment (LCIA) method, as implemented in *Brightway2*,

once and for all. Each background activity B_j is then substituted in the function $f(P_i, B_j)$ by the value of the computed environmental impact. This leads to the definition of one formula f_k , per impact category k , as a function of the parameters P_i only, as shown in Eq. 2.

$$\text{Impact}_k = f_k(P_i) \quad (2)$$

Third, the formulas f_k are compiled by *Sympy* into low-level functions, capable of processing large vectors of data at once. Typically, the library can compute environmental impact results for a million scenarios in a few seconds, for a large model with 30 parameters. This is several thousands of times faster than the standard parametrization mechanism of *Brightway2*, capable of processing hundreds of samples per second. *Brightway2* itself is already an improvement over

Fig. 2 Internal details of the library: development of the parametric inventory into a single symbolic function

other software, which can usually perform only a few calculations per second (Heijungs 2019).

This processing speed allows to perform quickly a large number of Monte Carlo simulations and to apply global sensitivity analysis on these results, thus providing insights on the variance of the environmental impact results and the role of each parameter.

2.3 Detailed protocol and use of lca_algebraic

In this section, we detail each step of the protocol and illustrate how the library is used in practice.

2.3.1 Definition of the scope of the study

In this step of the protocol, the following aspects must be defined:

- The functional unit of the system. The estimation of the environmental impacts refers to this functional unit.
- The boundaries of the system: location of production and usage, size of the system, technologies considered, etc.
- The environmental impacts to be estimated and their characterization methods.
- The LCA database(s) selected to model the background activities in the study.

2.3.2 Construction of the reference LCA model

The definition of the reference LCA model can be achieved either by parametrizing an existing inventory, or by building a new inventory from scratch. One should choose which characteristics of the system may vary and parametrize the inventory accordingly. We recommend starting from a typical system, representative of the scope defined previously. All the sub-parts of the system and all the phases of the life cycle should be considered.

A first life cycle impact assessment (LCIA) of the typical system can be performed in order to break down the environmental impacts by system part and/or life cycle phase. The parts of the system showing the greatest impacts should be parametrized in priority.

lca_algebraic provides three types of parameters that can be used in the inventory:

- Float: decimal parameters,
- Boolean: discrete parameters: 0 or 1,
- Enumerated (enum): discrete parameters set to one value among a predefined list. They are modeled internally as a linear combination of a set of exclusive boolean parameters.

Boolean and enum parameters are used to model conditional choices. By using a declarative definition instead of an imperative construction (if ... then ... else), we build a

single static inventory for all cases, controlled by the values of the parameters. This allows the factorization of background activities and the gain of performance described in the Sec. 2.2.

lca_algebraic also provides a set of helper functions, to:

- (1) Select background activities from the background database;
- (2) Duplicate background activities in the foreground and modify them;
- (3) Create new activities referring to background activities.

Overall, the declarative syntax and help functions make it possible to write compact and understandable code that is easier to maintain. A sample code is provided in Fig. 4 and discussed in Sec. 3.2.

Dependency of parameters The sensitivity analyses performed later in the protocol require that all parameters are statistically independent (Groen and Heijungs 2017). The library takes care of it for the enumerated, mutually exclusive parameters. When processing the GSA, it derives them from a single random float parameter between 0 and 1. For other parameters, the practitioners should ensure that they are independent. If not, they may manually perform a change of basis to make them independent.

Let us consider for instance three positive float parameters a , b , and c , bound together by the constraint $a + b + c < s$. Those parameters are dependent. The developer of the model should introduce 3 parameters a' , b' , and c' , varying independently on $[0, s]$, and replace a , b , and c as follows:

$$a = \frac{a' \times s'}{a' + b' + c'}, b = \frac{b' \times s'}{a' + b' + c'}, c = \frac{c' \times s'}{a' + b' + c'}$$

2.3.3 Distribution of input parameters

Once a reference LCA model has been built, the distribution of all input parameters should be characterized. We recommend several sources of information for this purpose: (1) academic literature, (2) manufacturer's catalogs, (3) market studies, (4) atlas of resources for energy systems (sun, wind), or (5) open-data and crowd-sourcing.

lca_algebraic supports the modeling of seven types of probabilistic distributions:

- Fixed: for excluding parameters from the statistical study
- Uniform: uniform distribution within the range of definition
- Triangle: null probability at the boundaries and highest probability for a default value, to be defined within this range. This type of distribution is useful for parameters for which literature only provides a range of extreme values and a usual one.
- Normal: normal distribution, capped to minimum and maximum values
- Log-normal: log-normal distribution, capped to minimum and maximum values
- Beta: beta distribution, capped to minimum and maximum values
- Statistic weight: for discrete parameters: boolean and enum

These types of distribution are consistent with the possibilities offered by traditional LCA software and were sufficient to model all the input distributions for the three case studies.

2.3.4 Sensitivity analyses

Once the reference LCA model is defined and the probability distributions of input parameters characterized, two approaches of sensitivity analysis can be applied: local and global analyses.

Local analysis Local sensitivity analysis consists in varying one parameter at a time (OAT) on its domain of definition, while the other parameters are kept to their default values. This method is simple and fast and does not require a precise knowledge on the distribution of uncertainties of the input parameters. It may be useful as a first step, to provide an overview of the importance of each parameter.

lca_algebraic implements OAT and presents the results in two forms:

- a set of graphs showing the evolution of each impact according to the variation of each parameter
- a single heat-map of the relative variability (max–min/mean) for each impact and parameter

This method shows some limitations though: it does not provide a quantified assessment of the variance of the impacts, and it may hide the importance of some parameters, only revealed when combined with different setups of the other parameters.

Global analysis The limitations of a local analysis can be overcome with a global sensitivity analysis. A global analysis uses a Monte Carlo simulation of all parameters at once and provides quantified indications of the variability of the

impacts and the role of each parameter in this variability. It should be noted that the values of uncertainty computed at this step only account for the variability of parametrized activities and do not propagate the uncertainty of the other inventories, thus providing only a lower bound of the total uncertainty of the model.

The library starts by generating many input parameter values according to their distributions, as defined in Sec. 2.3.3. For this step, we use the Saltelli approach (Saltelli and Annoni 2010). This pseudo-random generation is optimized to accelerate the convergence of Sobol indices (Sobol 2001) with fewer samples. The number of samples required depends on the complexity of the model and the number of parameters. Several tests should be carried out with an increasing number of samples until all statistical indicators converge.

lca_algebraic then computes all environmental impacts for all the sample input parameters, harnessing the processing speed provided by the algebraic approach. Several statistical methods are then applied to the results in order to provide insights on the distributions of the environmental impacts and the importance of each parameter in the final variability.

Distribution of impacts *lca_algebraic* computes several outputs from the Monte Carlo simulation of the environmental impacts:

- A graph of the full statistical distribution of each environmental impact
- A set of statistical indicators: mean, median, standard deviation, coefficient of variation (standard deviation/mean) and percentiles

Validation of the reference model The outcome of the global sensitivity analysis can be used to validate the reference LCA model against specific LCA results of similar systems. The lower and upper bounds should be compared to typical bounds found in the literature. *lca_algebraic* also provides a function for computing the environmental impacts for a given configuration of input parameters. It can be used to simulate a particular system whose environmental impacts have been assessed and published.

Decomposition of variance and Sobol indices Based on the Monte Carlo analysis results, a global sensitivity analysis following the variance decomposition approach proposed by Sobol is performed (Sobol 2001). Sobol indices are scalar values from zero to one. They quantify the importance of each input parameter in the final variance of a model output, in this case, the LCA result for a given environmental impact category. They are based on the decomposition of variance:

$$S_i = \frac{\text{Var}[E[Y|P_i]]}{\text{Var}[Y]} \quad (3)$$

where S_i is the Sobol index of first order of the parameter P_i and Y is the impact.

The sum of all Sobol indices is equal to 1. Sobol indices of higher order quantify the combined effect of several parameters. We did not use higher-order Sobol indices in our case studies because the effect of interactions was marginal and accounting for them requires many more samples to converge. The Sobol indices are then shown as a heat-map, with a color associated with each parameter and impact category. This provides an overview of the importance of each parameter to the variance of each impact.

Sobol indices are also used for generating simplified models, as described in Sec. 2.3.5.

2.3.5 Generation of simplified models

The computation of Sobol indices allows the identification of key parameters with the highest contributions to the total variance. These parameters can be then used to obtain simplified expressions providing accurate estimates of LCA results.

The use of symbolic calculus enables automatic manipulation of formulas. The library implements an algorithm, described in Fig. 3, selecting the most significant parameters and generating simplified models of the environmental impacts.

To select the key parameters, a threshold needs to be selected first. It represents the share of variability to be explained by the simplified model. This threshold is a trade-off between the simplicity of the generated simplified model and its accuracy. We chose 80% for our case studies.

Then, for each category of impact k , the following four steps are automatically carried out to first identify the key parameters (step 1) and then transform the formulas of the reference model f_k , described in Sec. 2.2 (Eq. 2) (steps 2, 3, and 4) (Fig. 3):

- (1) The input parameters are sorted by their Sobol indices of first order. A subset of the most significant parameters is selected, until the sum of their Sobol indices reaches the chosen threshold. In the example shown in Fig. 3, the parameters P1 and P4 would be selected, since they are explaining more than 80% of the total variance.
- (2) Each minor parameter (not selected above) is replaced by the mean value (μ) of its statistical distribution. In the example shown on Fig. 3, the parameters P2 and P3 would be replaced by their mean value.

1 Select key parameters

2 Replace minor parameters by mean values

$$\begin{aligned} P_2 &\Rightarrow \mu(P_2) \\ P_3 &\Rightarrow \mu(P_3) \\ f_k(P_1, P_2, P_3, P_4) &\Rightarrow f_k'(P_1, P_4) \end{aligned}$$

3 Round to 3 digits

$$\begin{aligned} m.nnnn &\Rightarrow m.nn \\ f_k'(P_1, P_4) &\Rightarrow f_k''(P_1, P_4) \end{aligned}$$

4 Remove residual sum elements

$$\begin{aligned} a + b + \varepsilon &\Rightarrow a + b \\ f_k''(P_1, P_4) &\Rightarrow f_k'''(P_1, P_4) \end{aligned}$$

Fig. 3 Algorithm generating simplified models

(3) Each numerical value is rounded to 3 significant digits. In Fig. 3, decimal values like $m.nnnn$ are replaced by $m.nn$. (4) For each sum in the formula, the terms contributing to less than 1% of the sum are removed. In Fig. 3, sums like $a + b + \varepsilon$ are replaced by $a + b$.

The resulting formula is a short arithmetic expression, involving only a few parameters, ready to be used for rough estimates of LCA results by non-experts. It should be noted that, for each category of environmental impact, a separate simplified model is generated, possibly involving a different subset of parameters. The validity of those simplified models is strictly limited to the framework defined in Sec. 2.3.1 and to the bounds identified for the parameters.

Validation of the simplified models The library provides an automatic function to assess the accuracy of the simplified models, compared to the reference one. It performs a paired Monte Carlo simulation (Suh and Qin, 2017) applying the same set of Monte Carlo samples to obtain the LCA results from the reference and the simplified models.

It should be noted that for the simplified model, only the uncertainty of the selected parameters is propagated. Performing paired Monte Carlo simulations ensures that we compare the models fairly (Brömssen and Rööös, 2020), only accounting for the differences of the two models, decoupled from the uncertainty of the input parameters themselves.

The resulting distributions and statistical indicators of both models are shown on a single graph. The function also

computes the coefficient of determination R^2 , quantifying the difference between the two distributions.

$$R^2 = 1 - \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i (y_i - \bar{y})^2} \quad (4)$$

y_i and $\hat{y}_i$ designate respectively the samples of the reference LCA model and the simplified model. $\bar{y}$ is the mean of the impact for the reference model.

The library also produces a separate graph of the residuals (y_i vs $\hat{y}_i$), enabling the user to explore in details how the simplified model fits to the reference one, and whether more parameters need to be taken into account, to locally increase the accuracy of the simplified models.

3 Results and discussion

We illustrate the application of the protocol by presenting the main results of the case study performed on photovoltaic systems (PV) in the INCER-ACV project. The purpose of this section is to discuss the benefits of the protocol and the library, rather than the values themselves.

3.1 Scope of the study

This case study covers crystalline silicon photovoltaic systems installed either on roof or on land, of any size: from small residential installations to large industrial ones.

Twelve environmental impact categories were identified as relevant for the analyzed technology according to the literature and selected in this case study. Out of these categories, we present here only the results for the climate change category of ILCD 2.0, (European Commission Joint Research Centre 2018). This impact is expressed in kilogram of CO₂ equivalent. All results for the other impact categories are provided separately as supporting information.

The production of one unit of energy by the PV system, expressed in kilowatt-hours (kWh), was selected as the functional unit. Thus, the total impacts over the PV system's lifetime were divided by the total energy produced over this period.

We used ecoinvent v3.4 as database of the background activities, due to its widespread use within the community. *lca_algebraic* should also support any database compatible with *Brightway2*.

3.2 Construction of the reference model

For this case study, we implemented the parametric model of mono-crystalline photovoltaic systems developed by Besseau (2019). This model includes 30 parameters. These parameters were identified by focusing on activities with significant relative contributions to the total impacts or activities that depend on highly variable parameters. This step of

the protocol is strongly dependent on the evaluated system and has to be, thus, carried out by an LCA expert with good knowledge on the sector of application.

Figure 4 shows a sample code of the definition of the reference model. New exchanges are defined as a standard *Python* dictionary. The keys of the dictionary refer to background activities selected from the database or to other foreground activities defined previously. The values of the dictionary are either static values or standard *Python* expressions (such as *P_module*, *surface*, ...) referring to parameters defined previously (such as *module_efficiency*). Note the usage of the boolean parameter *is_ground_system* conditioning the *diesel* activity required for ground PV installation.

This code shows mainly functional logic and little extra technical syntax. It is fairly easy to understand and maintain.

3.3 Distribution of input parameters

The selection of appropriate distributions for the input parameters has a strong influence on the results of the sensitivity analysis and the resulting deduction of simplified models, as already highlighted by Lacirignola et al. (2017). The definition of the type of variable input parameter and the distribution that represents it well depends on the type of parameter and ideally requires detailed data gathering from existing literature. When the information is scarce, certain

Fig. 4 Sample code of the definition of a parametrized LCA model with *lca_algebraic*

```
# Select background activity
diesel = findTechAct('market for diesel, burned in building machine')

# Define a new parameter
module_efficiency = newFloatParam("module_efficiency",
    default=0.175, min=0.15, max=0.22, distrib=DistributionType.TRIANGLE,
    description="peak power of the module per surface",
    unit="kWp/m²")

# Define shared expressions
P_module = module_efficiency * 1.6 * 1000
surface = P*1e3 / P_module * 1.6
inverter_weight = inverter_weight_per_p * P

# Add exchanges to the inventory
pvInstall.addExchanges({
    mounting_system_modified : surface,

    # Ground installation only : 7673 MJ per 570 kWp
    diesel : is_ground_system * P/570 * 7673,

    # Number of equivalent unit of 3kW elec install
    electrical_installation_3kW : unit_3kW_equiv,

    inverter_per_kg : inverter_weight * pv_lifetime / inverter_lifetime,
    panel_adjusted : surface
})
```

distributions can be used as an estimate (e.g., triangular, uniform), but further analyses of the sensitivity of the results to these choices might be needed.

In this case, we used all types of parameters provided by the library in the reference LCA model for PV:

We used a boolean parameter for the type of installation, with equiprobability between rooftop and ground installations.

We used an enum parameter as the choice of the electricity mix used for the manufacturing phase, with probability weights corresponding to the current reality of the market.

All other parameters are modeled as float values, with the following distributions:

For most parameters of the model, we used triangle distributions, with bounds corresponding to the minimum and maximum values found in the literature and manufacturer's catalogs, and the mode (higher probability) set as the most common value found in current installations. We used a crowd-sourced database of domestic PV production (BDPV) to assess the capacity of production (kWh/kWp/year) and modeled it with a normal distribution.

The complete list of parameters is provided as supporting information.

3.4 Sensitivity analysis

In this section, the results of the sensitivity analyses as processed automatically by the library are presented and discussed.

3.4.1 Local analysis

Figure 5 is an output of the local (OAT) analysis, as produced by the library. It shows the impact on *climate change* according to the yearly expected producible.

Although the library allows for quick exploration of the OAT of each parameter, it is best to focus on the parameters that show the most impact on the total variability. In this case study, the GSA provided in Sec. 3.4.2 shows the yearly producible as one key parameter.

The information provided by the OAT is complementary to that provided by the GSA. The GSA provides a quantification of the relative importance of a parameter. The OAT, on the other hand, provides insight into the direction of that contribution and the absolute bounds of that variation.

Here, the climate change impact decreases sharply (from 80 to 30 g of CO_{2eq}/kWh) as the annual producible increases. This result is expected, since the producible is used in the calculation of the functional unit (annual electricity production), which is the denominator of all impacts.

3.4.2 Global analysis

For this case study, with thirty parameters, the convergence of the indicators was observed starting at one million samples. This convergence depends of the complexity of the model and the number of parameters. The practitioners should repeat the global analysis with an increasing number of samples until the results converge.

The computation of LCA for the corresponding Monte Carlo simulations took around three seconds on a PC with Intel i7 CPU (1.80 GHz) and 16 Gb of RAM. It would

Fig. 5 Example OAT. Climate change according to electricity producible

have required around three hours of computation using the standard parametrization mechanism provided by Brightway2 (around hundred iterations per second).

Distribution of impacts Figure 6 shows the outcome of the Monte Carlo simulation in form of a distribution and the statistic indicators of the impact. The statistic indicators are shown on the right: mean (μ), median, standard deviation (σ), coefficient of variation (σ/μ), and percentiles (5% and 95%).

The graph here shows the shape of a lognormal distribution, typical of a statistical variable resulting from the combined contribution of many independent parameters.

The mean of this distribution is 37.6 g CO_{2eq}/kWh with a standard deviation of 9.8 g CO_{2eq}/kWh. The variability is significant, with a coefficient of variation (σ/μ) equal to 26%.

Validation of the reference model Those results are consistent with the existing literature. IPCC (Eldenhofer et al. 2012) reports climate change impacts ranging from 30 to 215 g CO_{2eq}/kWh. Our results appear in the lower bounds of those estimates as Besseau (2019) updated the inventories of photovoltaic systems with the latest improvements of the industry.

Sobol indices Figure 7 is a part of the heat map showing the values of Sobol indices for all the impacts considered.

Fig. 7 Heat map of Sobol indices

Climate Change is highlighted in blue. The full figure is provided as supportive information.

Fig. 6 Distribution of the impact climate change

Fig. 8 Validation of the simplified model for climate change. Comparison of the distributions

While most parameters are either significant or negligible for all impacts, some turn out to be important for some categories of impact only. For climate change, five parameters stand out: *Producible*, *PvLifetime*, *SiliconElecIntensity*, *ModuleEfficiency*, and *ElecSwitch*.

3.5 Simplified model

The highest Sobol indices of four parameters (*Producible*, *PvLifetime*, *SiliconElecIntensity*, *ModuleEfficiency*) sum up to 80% of the total variance. Eighty percent is an arbitrary

Fig. 9 Validation of the simplified model for climate change. Graph of residuals

threshold we chose in this study as a good compromise between the complexity of the simplified models and their accuracy. The practitioners may incrementally adapt this threshold to the complexity of their inventory and the accuracy of the resulting models.

These four parameters are automatically selected by the algorithm described in Sec. 2.3.5, producing the simplified model shown in Eq. 5:

$$I_{\text{Climate}} = \frac{3.79 \cdot 10^{-6} \text{ModuleEfficiency} (1.58 \cdot 10^3 \text{PvLifetime} + 9.88 \cdot 10^3) + 4.48 \cdot 10^{-3} \text{SiliconElecIntensity} + 0.932}{\text{ModuleEfficiency} \cdot \text{Producibile} (1 - 0.994^{\text{PvLifetime}})} \quad (5)$$

I_{Climate} is expressed in kilogram of CO₂ equivalent; *ModuleEfficiency* is the expected peak power of the PV module, per surface, including inverter losses, expressed in kilowatt-peak per square meter; *Lifetime* is the expected duration of the installation, in years; *SiliconElecIntensity* is the amount of energy required to produce 1 kg of silicon, expressed in kilowatt-hour per kilogram; *Producibile* designates the expected yearly electricity production per installed power, at the location of installation, in kilowatt-hour per kilowatt-peak per year.

This arithmetic expression involves only four parameters out of the thirty of the reference LCA model.

3.5.1 Validation of the simplified model

The automatic validation of the simplified model is presented in Fig. 8. The figure provides the comparison between the distribution of this simplified model (in orange) and the one of the full reference LCA model (in blue). The simplified model reproduces the outcome of the reference LCA model well, with statistical indicators close to each other and an R^2 close to 1 (0.80).

The graph of residuals (Fig. 9) shows a uniform fit of the simplified model against the reference one, with no local deviation that would require to add more parameters to increase its accuracy.

This model is both simple and accurate compared to the reference LCA model. It is suitable to be used by non-experts for quick assessment of this environmental impact.

3.6 Conclusions

In this article, an operational protocol and the complementary library *lca_algebraic* were presented to help spread the application of sensitivity analysis approaches in LCA studies. The suitability of these tools to generalize sensitivity analysis was validated by both academic and industrial partners on distinct case studies exploring the uncertainties of three different energy systems: photovoltaic (presented here), floating wind turbines (Pérez-López et al. 2020), and

geothermal energy (Douziech et al. 2021). For each system, complex parametric models were built and accurate simplified models were generated for a given applicability range. Feedbacks gathered from the application to these case studies were integrated in updates of the protocol and the library.

It should be noted that building a parametric LCA model is only possible given a good knowledge of the studied system. This knowledge can be gathered through literature studies or exchanges with system's experts. However, once the reference LCA model is built, the sensitivity analysis is relatively flexible and can be applied even if precise knowledge of all parameters is lacking. In this case, the results of the sensitivity analysis could inform on the parts of the system that require more knowledge.

This work brings practical tools to the LCA community to build parametric inventories and explore the uncertainties of their models. Local and global sensitivity analyses provide complementary insights on the role of each parameter on the variation of environmental impacts.

We believe that this library can help LCA experts in the adoption of programming languages by providing a simple interface and helper functions. The use of symbolic calculus and the factorization of background impacts could also be integrated into standard LCA software to enable the application of advanced GSA methods. The proper characterization of the distributions of input parameters remains a complex task and is key in the outcome of this protocol.

Additional work may be conducted to combine these results with uncertainty of the full background, described by pedigree matrices. While this work focuses mainly on assessing the uncertainties of foreground activities, it may also be used for some background activities.

The results of applying the presented protocol to different systems are also useful for non-experts and help to disseminate LCA methods. The comprehensive evaluation of uncertainties and variability in LCA is essential to improving the transparency and reliability of the environmental impact results. The identification of key parameters and the simplified models allow industrial stakeholders to easily obtain fast estimates of environmental impacts by focusing on data collection for a limited number of parameters.

While originally developed for energy systems, the tools presented in this paper are generic enough to be used for any system.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s11367-021-01993-z>.

Acknowledgements The protocol and library presented in paper were also used by partners from GEOENVI, for assessing the environmental impacts of heat production from geothermal systems, in the scope of the GEOENVI project, part of the Horizon 2020 program of EU

(agreement no. 818242). This work has provided valuable feedback for the improvement of these tools.

Funding This work was conducted in the framework of the INCER-ACV project (contract 1705C0045). INCER-ACV, funded by ADEME in the framework of the call “Sustainable Energy” (APR-ED 2017), aims to contribute to the consolidation of quantification methods to account for the effects on the environmental impact results of possible parameter variations compared to average scenarios.

Data availability The library is distributed online under the BSD license, at the following address: https://github.com/oie-mines-paristech/lca_algebraic. The code of the PV model is available from the corresponding author on reasonable request. It requires a copy of the *ecoinvent* database v3.4.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Besseau R (2019) Analyse de cycle de vie de scénarios énergétiques intégrant la contrainte d'adéquation temporelle production-consommation. Theses, Université Paris sciences et lettres. <https://pastel.archives-ouvertes.fr/tel-02732972>
- Clavreul J, Guyonnet D, Tonini D, Christensen T (2013) Stochastic and epistemic uncertainty propagation in LCA. *Int J Life Cycle Assess* 18(7):1393–1403. <https://doi.org/10.1007/s11367-013-0572-6>
- Douziech M, Ravier G, Jolivet R, Pérez-López P, Blanc I (2021) How far can life cycle assessment be simplified? A protocol to generate simple and accurate models for the assessment of energy systems and its application to heat production from enhanced geothermal systems. *Environ Sci Technol* 55(11):7571–7582. <https://doi.org/10.1021/acs.est.0c06751>
- Eldenhofer O, Pichs-Madruga R, Sokona Y, Seyboth K, Marschoss P, Kadner S, Zwickel T et al (2012) Renewable energy sources and climate change mitigation: special report of the Intergovernmental Panel on Climate Change. *Choice* 49(11):49–6309–49–6309. <https://doi.org/10.5860/choice.49-6309>
- European Commission Joint Research Centre (2018) Supporting information to the characterisation factors of recommended EF life cycle impact assessment methods: new methods and differences with ILCD. LU: Publications Office. <https://data.europa.eu/doi/10.2760/671368>
- Frischknecht R (1998) Life cycle inventory analysis for decision-making. *Int J Life Cycle Assess* 3(2):67–67. <https://doi.org/10.1007/bf02978487>
- Groen EA, Bokkers EAM, Heijungs R, de Boer IJM (2016) Methods for global sensitivity analysis in life cycle assessment. *Int J Life Cycle Assess* 22(7):1125–1137. <https://doi.org/10.1007/s11367-016-1217-3>
- Groen, E. A., and R. Heijungs. (2017) Ignoring correlation in uncertainty and sensitivity analysis in life cycle assessment: what is the risk? *Environ Impact Assess Rev* 62:98–109. <https://doi.org/10.1016/j.eiar.2016.10.006>
- Heijungs R (2019) On the number of Monte Carlo runs in comparative probabilistic LCA. *The International Journal of Life Cycle Assessment* 25(2):394–402. <https://doi.org/10.1007/s11367-019-01698-4>
- Igos E, Benetto E, Meyer R, Baustert P, Othoniel B (2018) How to treat uncertainties in life cycle assessment studies? *Int J Life Cycle Assess* 24(4):794–807. <https://doi.org/10.1007/s11367-018-1477-1>
- ISO 14040 (2006) International Standard. Environmental management - life cycle assessment - principles and framework 2006. <https://www.iso.org/standard/37456.html>
- Lacirignola M, Meany BH, Padey P, Blanc I (2014) A simplified model for the estimation of life-cycle greenhouse gas emissions of enhanced geothermal systems. *Geotherm Energy* 2(1). <https://doi.org/10.1186/s40517-014-0008-y>
- Lacirignola M, Blanc P, Girard R et al (2017) LCA of emerging technologies: addressing high uncertainty on inputs' variability when performing global sensitivity analysis. *Sci Total Environ* 578:268–280. <https://doi.org/10.1016/j.scitotenv.2016.10.066>
- Mutel C (2017) Brightway: an open source framework for life cycle assessment. *J Open Source Softw* 2(12):236. <https://doi.org/10.21105/joss.00236>
- 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(3):1231–1238. <https://doi.org/10.1021/es303435e>
- Parliament European (2014) Directive 2014/52/EU of the European Parliament and of the Council of 16 April 2014 Amending Directive 2011/92/EU on the Assessment of the Effects of Certain Public and Private Projects on the Environment Text with EEA Relevance OJ L. <https://eur-lex.europa.eu/eli/dir/2014/52/oj>
- Pérez-López P, Jolivet R, Blanc I, Besseau R, Douziech M, Gschwind B, Scarlett Tannous J, Schlesinger, Brière R, Prieur-Vernat A, Clavreul J (2020) INCER-ACV: Incertitudes dans les méthodes d'évaluation des impacts environnementaux des filières de production énergétique par ACV. Final report of the study 1705C0045. 77. <https://bibliothèque.ademe.fr/cadic/5404/incer-acv-2021-rapport.pdf>
- Pianosi F, Beven K, Freer J, Hall H, Rougier J, Stephenson D, Wagener T (2016) Sensitivity analysis of environmental models: a systematic review with practical workflow. *Environ Model Softw* 79:214–232. <https://doi.org/10.1016/j.envsoft.2016.02.008>
- Saltelli A, Annoni P (2010) How to avoid a perfunctory sensitivity analysis. *Environ Model Softw* 25(12):1508–17. <https://doi.org/10.1016/j.envsoft.2010.04.012>
- Sobol IM (2001) Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates. *Math Comput Simul* 55(1–3):271–280. [https://doi.org/10.1016/s0378-4754\(00\)00270-6](https://doi.org/10.1016/s0378-4754(00)00270-6)
- Suh S, Qin Y (2017) Pre-calculated LCIs with uncertainties revisited. *Int J Life Cycle Assess* 22:827–831. <https://doi.org/10.1007/s11367-017-1287-x>
- von Brömssen C, Rös E (2020) Why statistical testing and confidence intervals should not be used in comparative life cycle assessments based on Monte Carlo simulations. *Int J Life Cycle Assess* 25:2101–2105. <https://doi.org/10.1007/s11367-020-01827-4>
- Walker WE, Harremoës P, Rotmans J, van der Sluijs JP, van Asselt MBA, Janssen P, Krayen von Krauss MP (2003) Defining uncertainty: a conceptual basis for uncertainty management in model-based decision support. *Integr Assess* 4(1):5–17. <https://doi.org/10.1076/iaij.4.1.5.16466>

Wei W, Larrey-Lassalle P, Faure T, Dumoulin N, Roux P, Mathias J-D (2014) 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(1):377–385. <https://doi.org/10.1021/es502128k>

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