Universiteit Leiden

TU Delft

Thesis for the degree MSc. Industrial Ecology

An analytical approach of Uncertainty Propagation for Sensitivity Analysis of Life Cycle Assessment

---

Author
Sietske Lensen

Supervisors
Stefano Cucurachi
Reinout Heijungs

April 20, 2021

TU Delft, Faculty of Technology, Policy and Management (TPM)
Leiden University, Institute of Environmental Sciences (CML)

Abstract

Life Cycle Assessment (LCA) models are inherently uncertain due to the model structure interacting with the model inputs and modeling choices. The methods of sensitivity analysis (SA) aim at retracing the causes of the uncertainty of the results of a model. This work takes apart the methods of uncertainty propagation, SA, and LCA, and identifies the requirements of appropriate SA methods for LCA. Global input space assessment and inclusion of correlation are identified as important factors which both analytical and sampling SA methods have issues addressing. An analytical expression for covariance is formulated that combines research on the uncertainty propagation methods to address the posed requirements. Its performance is tested and shown to be promising, but further manipulation is required for practical application in SA for LCA in the future.

Mathematical notation

Object	Notation	Example
Scalar	italic	c
Stochastic coefficient	bold & capital	$\boldsymbol{X}$
Vector	arrow	$\vec{x}$
Matrix	capital	A
Element of matrix (vector)	subscripts	A_{ab} (x_i)

Table 1: Reference table for mathematical conventions and notations used in this thesis

Measure	As parameter	Alternatively
Mean or expectation value of X	μ_X	$E[\boldsymbol{X}]$
Variance of Y	Σ_{YY}	$V[\boldsymbol{Y}]$
Standard deviation of X	σ_X	$\sqrt{V[\boldsymbol{X}]}$
Covariance between X,Y	Σ_{XY}	$Cov[\boldsymbol{X},\boldsymbol{Y}]$
Correlation between X,Y	$\Sigma_{XY}\sigma_X^{-1}\sigma_Y^{-1}$	$Cov[\boldsymbol{X},\boldsymbol{Y}]\sqrt{V[\boldsymbol{X}]^{-1}V[\boldsymbol{Y}]^{-1}}$

Table 2: Reference table for parameter conventions used in this thesis

LCA coefficient object	Name	Content	Size
Impact matrix	H	Impact categories times reference flows	$(m \times q)$
Scaling matrix	S	Products times reference flows	$(p \times q)$
Characterization matrix	Q	Impact categories times externalities	$(m \times n)$
Externalities matrix	B	Externalities times processes	$(n \times p)$
Technology matrix	A	Processes times products	$(p \times p)$
Reference flow matrix	F	Products times reference flows	$(p \times q)$

Table 3: Reference table for the objects present in the LCA model calculation

Glossary

Black-box: A type of model of which nothing is known about the inner function. Only simulation can be used to assess its behavior

Coefficients: The input or output variables to a model

Global: Type of model assessment in which the whole input space is considered

Input space: All possible combinations of values the coefficients can take on. The combinations form a multidimensional space in which one point is one possible model state

Linearity: A type of model behavior where coefficients act independently upon the output. They do not multiply or influence each other otherwise.

Local: Type of model assessment in which a limited area of the input space is considered

Parameters: The moments (such as mean or covariance) of coefficients

Sensitivity analysis (SA): Methods used for apportioning uncertainty in the output to different sources of uncertainty in the input

Stochastic: Term that indicates a coefficient is uncertain

Uncertainty analysis (UA): Methods used to evaluate how uncertain a model output is

Uncertainty propagation (UP): The act of finding the uncertainty of the output of a model (for one set of input parameters)

Contents

1 Introduction to Sensitive Life Cycle Assessment

Life Cycle Assessment (LCA) is a method used in environmental science, policy-making, and industry to assess and compare the environmental impacts of products over the course of their life cycle. Typical applications of LCA are in strategic planning, capital investment, eco-design, product marketing, eco-labeling, municipal waste management, and the development or design of new technologies. LCA use is widespread and its conclusions are used in many types of decisions. It is therefore vital that the method produces reliable and credible results.

Unfortunately, outcomes LCA studies for similar product systems can have large differences, even if the studies comply with the same methodological guidelines because, in any LCA, context-specific assumptions have to be made by the practitioner (Cucurachi et al., 2016; Heijungs et al., 2019). Ignoring this aspect of LCA modeling is undesirable, because it results in contradicting conclusions, wrong decisions, and reduces the credibility of LCA; for example when suspicion of manipulation with modeling choices arises (Igos et al., 2019).

Heijungs (2002) states that the LCA model is inherently *sensitive* due to the non-linearity and reciprocity of the LCA calculation. These aspects interacting with the *uncertainty* of the modeling assumptions create *uncertain* results. Data variability is for example another type of input uncertainty that leads to uncertainty in the model output. The standard practice of deterministic LCA presents its conclusions as point-values, which are not able to illustrate uncertain model results (Mendoza Beltran et al., 2018). Extra steps must be undertaken to systemically address such results.

There are two parts to the assessment of LCA model uncertainties. First, uncertainty in the inputs of the model must be translated to uncertainty in the model output. This is **uncertainty propagation** (UP). Second, to gain an understanding of the causes of the uncertainty, we retrace which input coefficients have a profound or instead trivial impact on the output uncertainty. The study of apportioning uncertainty in the output to different sources of uncertainty in the input is called **sensitivity analysis** (SA) (Saltelli et al., 2019). Igos et al. (2019) identify four distinct phases of the SA procedure. They are (1) identification of input uncertainties, (2) UP, (3) evaluation of coefficient influence, or sensitivity characterization, and (4) communication of results.

The ISO 'Principles and framework' for LCA (International Standardization Organization, 2006) recommends performing an analysis of sensitivity, but does not present the LCA practitioner with guidelines and does not direct to particular approaches on performing such an analysis (Cucurachi et al., 2016). Ross et al. (2002) found that only 3 % of LCA studies reported quantitative uncertainty characteristics.

We identify two distinctions in the SA calculation approach; analytical and sampling. Analytical approaches define uncertainty and sensitivity in algebraic expressions. Exact equations for LCA can not be defined and thus approximations and simplifications are used. This way, model attributes such as large uncertainties, correlations, and non-

linear coefficient interactions are improperly included. Calculation of these analytical expressions is however nearly instant, and the relationships between the input and output uncertainties are shown explicitly. Therefore sensitivity can be retraced easily from uncertainty. Due to the limitations, analytical assessment is used mainly as a filtering strategy. For example, the sensitivity multipliers of Heijungs (2010) are for this use. The goal of filtering strategies is to simplify sampling SA.

SA sampling approaches can cover model attributes that the analytical strategies struggle with. The LCA model is treated by the sampling strategies as a *black-box*. This means forgetting the internal mathematical structure, and repeatedly calculating the output while applying small variations to the inputs. For sampling UP, large amounts of simulations are required, which is computationally intensive due to the size of LCA systems and the requirements of the calculation (Heijungs & Lenzen, 2014). Subsequently, performing SA using a sampling strategy involves performing many UP's. The combined computational time makes for an unfeasible endeavor (Pianosi & Wagener, 2015). The aforementioned filtering strategies can be applied beforehand to screen for possible sensitive coefficients and reduce the computational load.

This work proposes to improve upon the quality of the results of the analytical approach and make them more useful for SA in a standalone application; not just as a stepping-stone for a sampling strategy. The analytical SA strategies and their explicit statement of the relationships between the uncertainty parameters can provide the LCA practitioner with a profound insight into the LCA model workings that a black-box approach simply can not. To do this, the validity of the analytical approach must be extended and its limitations addressed. This thesis explores the theories of SA, LCA, and mathematical uncertainty manipulation for ways in which this can be achieved.

2 Methodology and thesis structure

In this section, the research questions, methodology, and structure of the work are laid out, but first, the collected literature is briefly introduced.

2.1 Literature collection

Literature collection started with a selection of relevant papers on sensitivity and uncertainty characterization in the field of LCA. A snowballing approach was followed next to gain a better understanding of the context of the research field, pursuing references within the papers and searching for other work by the same authors. Two books were consulted as starting points for information about the general mathematical field of SA (Saltelli et al., 2008) and on matrix calculus theory (Magnus & Neudecker, 1999). Furthermore, research for papers was done using Google Scholar and the Leiden University Library Catalogue, with search words as *'Sensitivity Analysis'* and *'Uncertainty'* both associated with *'Life Cycle Assessment'*. Additional literature was used as it became relevant during the research process. This mainly consisted of literature on the general mathematical understanding of uncertainty distributions, on the relationships between distribution moments, and on theories for analytical propagation of uncertainties. All literature can be found in the references section.

2.2 Main research question

In Section 1, areas were introduced in which SA for LCA is lacking. They are (1) the expression of an analytical formula for sensitivity, which ensures to (2) account for non-linear coefficient interactions and higher uncertainty ranges, and while (3) making sure correlation effects are included. To summarize these points and guide the research, the main research question is formulated as

How can the relationships between the uncertainty of an LCA model output and the uncertainties of its inputs be expressed in analytical formulation, without disregarding non-linear model behavior, and without omitting coefficient correlations.

2.3 Sub research questions and methods

Sub research questions are defined to break down the main research question. The posed research gaps must be developed and explored in search of areas suitable for improvement. The sub-questions are presented next, followed by a reflection on the appropriate methods and resulting thesis structure, forming a guide to the reader.

(1) What are the goals, process, and concerns of SA?

(2) What are the structural components and properties of the LCA model?

First, the topics of SA and LCA must be explored to establish the attributes of the methods. The goal of this is to understand the setting in which the expression of the main research question must reside. Sub-questions 1 and 2 address this for SA in Section 3 and for LCA in Section 4 respectively. The questions are answered by conducting a literature review.

(3) How do we apply the established properties of SA to LCA?

Next, we must establish the requirements, possibilities, and properties of the desired expression from our main research question. The SA tools available for LCA are laid out. For this purpose, sub-question 3 asks to combine the findings of sub-questions 1 and 2 through analytical research. The answers to this question are found in Section 5.

(4) What is the expression of uncertainty satisfying the posed requirements?

Sub-question 4 asks what analytical expression is follows from the established desired properties and limitations. To answer it we combine the answers of sub-question 3 with additional literature on the mathematics of uncertainties. This question is thus addressed using a combination of literature and analytical research. Section 6 is dedicated to this research question.

(5) How does the uncertainty expression perform compared to uncertainty calculated using a random sampling method?

Lastly, sub-question 5 addresses the validity and performance of the proposed analytical expression of sub-question 4. This research follows a modeling approach, and the process can be found in Section 7.

3 Theory of Sensitivity Analysis

This section is intended to pick apart the topic of SA. What it is, why it is performed, how it is performed, and what its points of concern are.

3.1 What is SA

Sensitivity analysis is the study of ascribing responsibility for the presence of uncertainty in a model output(s) to the uncertainties of the model inputs (Igos et al., 2019). The model inputs are called *coefficients*, while their uncertain behavior can be captured in *parameters* such as mean and variance. A model in this sense can be any process or formula that transforms input data into an output answer and can for example be mathematical or computational. The steps of SA have been mentioned in Section 1.

The posed explanation of SA contains both the words sensitivity and uncertainty, which can be confusing. In fact, in literature, the terms tend to be conflated, where uncertainty analysis (UA) is incorrectly reported as SA or vice versa. UA is used to determine how uncertain a model output is. SA retraces the share that input parameters contribute to the uncertainty, and it characterizes their relative importance (Saltelli et al., 2019). Sensitivity may depend on the structure of a model; how coefficients interact and amplify one another, in addition to the individual coefficient uncertainties. SA can be seen as an extension of UA.

We can also express the distinction between UA and SA in terms of the SA steps of Igos et al. (2019). When only the uncertainty of a model is in question, and not its sensitivities, a single act of UP suffices, and performing steps (1), (2), and (4) would be a UA. In contrast, to obtain knowledge on the sensitivity of a model, a multitude of UP's under several different parameter conditions are required (Saltelli et al., 2019). Step (3) in which the different propagated uncertainties are compared, is in this case added to the process to form SA.

Saltelli et al. (2008) identify four rationales for SA, so-called *'settings'*. They are (1) to determine the robustness of a model, (2) to prioritize research on important coefficients, (3) to simplify the model by finding omissible coefficients, and (4) to determine interesting input space regions.

The input space of a model refers to all possible combinations of values the input coefficients can take on. The combinations form a multidimensional space in which one point is one possible model state.

3.2 Appropriate SA methods

Some SA methods are inappropriate for the analysis of models with certain properties. In this section, the properties of (non-)linearity and coefficient correlation in relation to SA are considered. Section 4 reviews these properties with respect to the LCA model.

Two matters must be addressed before diving into the next sections. First, Appendix A presents a brief overview of linear models. It includes an example surface plot and color plot of a non-linear function, which will be built upon here. Second, both the terms correlation and covariance appear and their distinction needs to be explained. Covariance between parameters $\boldsymbol{X}$ and $\boldsymbol{Y}$ equals the correlation between $\boldsymbol{X}$ and $\boldsymbol{Y}$, scaled by the variance of $\boldsymbol{X}$ times the variance of $\boldsymbol{Y}$. This definition in mathematical terms can be found in Table 2.

3.2.1 Local and global SA

One major distinction generally made between SA methods is that of local SA (LSA) and global SA (GSA). Literature on SA in LCA has had conflicting definitions of this distinction, therefore some clarification is required. Notably, both UP and sensitivity characterizations have ways of being local or global. These concepts overlap and yet have nuanced differences. Local UP can sometimes produce global SA, and global UP can sometimes result in local SA.

Uncertainty propagation is found to be local in two ways

(i) Only a small (local) input space range centered at a mean point is considered, applying little variations to multiple input coefficients (Jacques et al., 2006). Perturbation analysis is of this kind.

(ii) Coefficients are varied one by one, but possibly over a larger input space. An example of this is *one-at-a-time analysis* (Igos et al., 2019)

These two approaches to local UP are best illustrated graphically in a two-parameter model. Figure 1 is adapted from Saltelli et al. (2019) for this purpose. Model evaluations, marked in red, are projected on top of the non-linear function $\boldsymbol{Y} = \boldsymbol{X}_1\boldsymbol{X}_2$ (see Appendix A for an explanation of the color plot). The bold, capitalized characters signify stochastic coefficients. This mathematical notation and others can be found in Table 1. Figure 1a shows both types of local UP. For (i), only values of the model are considered near the mean, represented by the red dot. For (ii), values along the axes are considered, represented by the dotted lines. The values of Y are zero in all of the assessed locations of the input space.

If a given model is linear, local behavior can be extrapolated over the entire input space. Or put another way; assuming model linearity is equivalent to assuming that local model evaluations are representative of the global behavior of the model (Igos et al., 2019). When applying local UP to a non-linear model, the uncertainty results will only be valid around the evaluated point.

An example of a global approach to UP is given in Figure 1b. Here evaluation of the input space is spread out to ensure all model behavior is encountered. Evidently, for a non-linear model, uncertainty will be dependent on the reference point at which it is evaluated (Di Lullo et al., 2020).

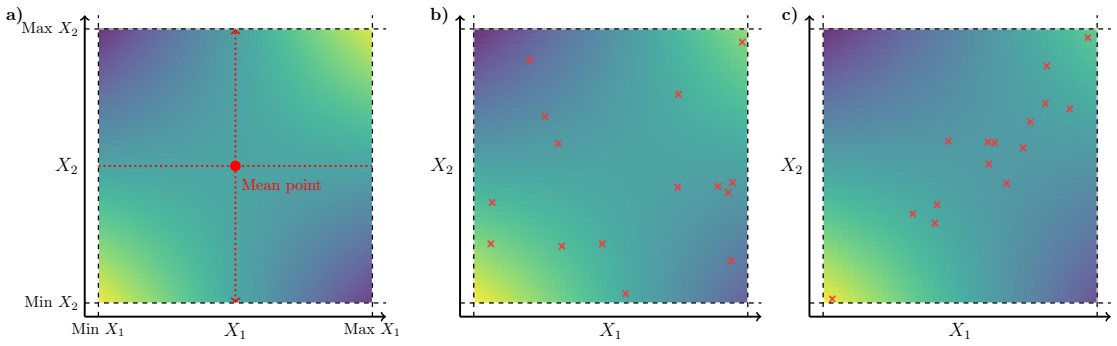


Figure 1: Comparison of evaluated input space when performing (a) local UP, (b) global UP and (c) correlated global UP. Approaches laid over a color plot of the non-linear function $\boldsymbol{Y} = \boldsymbol{X}_1\boldsymbol{X}_2$. Figures are adapted from Saltelli et al. (2019)

Sensitivity of a model to a particular coefficient is in part reflected by the first partial derivative of the model to that coefficient. A list of sensitivity coefficients using this premise is presented by Heijungs (2010). In a two-parameter model, such derivatives are equivalent to the output slope in the direction of the axis of the coefficient in question. If the model is non-linear with respect to the coefficient, the slope will depend on the point at which the model is evaluated. Interactions such as multiplication with other variables, cause the derivative to be dependent on the values of the other input coefficients also (Saltelli et al., 2019). In Figure 1a, we can see that the model slope in the direction of the input parameters is zero everywhere at the evaluated input space and therefore not representative of the true model sensitivity. Saltelli et al. (2008) solve this issue by evaluating the model variance globally, while fixing the coefficient in question at different values of its range, and averaging over these the found variances.

In an SA where multiple UP's with different starting input parameters are compared, globality can be included in three ways: (1) when the UP's themselves are global; (2) when UP parameter changes are introduced simultaneously, as opposed to one parameter at a time; (3) when the assessment is mindful of the fact that in a non-linear model, sensitivity itself is location dependent.

3.2.2 Correlation

Correlations arise, for example, when an engine that uses less fuel also produces fewer exhaust gasses, or when products share disposal pathways, or due to characterization indicators having common dependencies (Igos et al., 2019). Depending on the interactions of the model, uncertainty may turn out higher or lower when accounting for correlations (E. Groen & Heijungs, 2017). Assuming uncorrelated parameters, just like assuming linearity, is a way to simplify a model. For this reason, there is a lack of widespread and systematic inclusion of parameter correlation in UP.

Figure 1c demonstrates how considering correlation in the dummy model could affect UP. If $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ are positively correlated, and pairs of samples are generated with this in mind, higher values of $\boldsymbol{Y}$ (the yellow parts of the plot) are more likely to occur than in the uncorrelated sample set of Figure 1b. Ignoring correlations thus leads to a skewed evaluation of the input space.

In the SA that follows the UP, considerations of correlations are also important. Without them, we only have to assess which variances the model is sensitive to. With them, the covariances must be considered too. This can square the amount necessary UP's for SA, which further reduces the viability of sampling approaches. When using an analytical method only for filtering before applying sampling SA (as mentioned in Section 1), correlations should be addressed too.

4 Theory of Life Cycle Assessment

This section is dedicated to understanding the structure of the LCA model and formulating its mathematical notation.

4.1 The LCA components

The type of LCA calculation that the research centers on is matrix-based LCA, which allows for solving the model through algebraic matrix manipulation. Other approaches can be and have been used, such as Cramer's rule and sequential calculation (Heijungs & Lenzen, 2014).

Matrix-based LCA as described in Heijungs (2002) guides the model inputs into matrix structures and calculates the model output through matrix multiplication and inversion. It is considered the major method for solving LCA calculations, and it is implemented in various software tools such as CMLCA, SimaPro, and OpenLCA (Wei et al., 2016).

The components of an LCA calculation are

- A database of processes with corresponding externalities. These are called the background processes.
- A database of characterization factors
- Data on processes and corresponding externalities delivered by the LCA practitioner specific to their study. These are called the foreground processes.
- A reference flow matrix, describing the products that the model is intended to calculate impacts for

In literature, externalities are sometimes called *interventions*, characterization factors are sometimes called *impact factors*, and reference flows are sometimes called *alternatives* or *final demands*. All of the coefficients in the LCA input components can be uncertain and correlated. A more detailed characterization of uncertainties present in the LCA model will be given in Section 5.3 when it is related to performing SA.

4.2 LCA notation and calculation

The LCA matrix-based model calculation is of the following form:

$$H = QBA^{-1}F \tag{1}$$

where the symbols are defined in Table 4. The meaning of the LCA formula is as follows; solving the linear system $AS = F$, for $S = A^{-1}F$, determines the multipliers S that all unit processes in A must be scaled by to exactly produce the desired reference flows in F. These multipliers are then used to scale the externalities B. Finally, an impact is calculated by scaling by the characterization factors Q (Heijungs, 2002; Wei et al., 2016). F is a matrix when multiple reference flows are calculated, each represented in a column.

It however mostly appears in literature as a vector $\vec{f}$ when a single reference flow is in question. In this work, we prefer the matrix notation for it allows the expression of covariances between alternatives. Matrix S mirrors this shape of F. The dimensions of H depend on the number of reference flows and the number of impact categories. For an overview of the sizes of these objects, see Table 3.

LCA object	Symbol	Explanation
Impact matrix	H	Output of the LCA model. Reflects the impact of the reference flow
Scaling matrix	S	Scales unit processes to exactly produce the reference flows
Technology matrix	A	Contains the inputs and outputs of the unit processes
Externalities matrix	B	Contains the externalities emitted or consumed by each unit process
Characterization matrix	Q	Contains characterization factors for externalities. Scales externalities to impact categories
Reference flow matrix	F	Defines which products the LCA model calculates impacts for
Parameter object	Symbol	Explanation
Mean	μ	Used for representing the mean parameters of coefficients
Covariance	Σ	Used for representing the covariance parameters between pairs of coefficients

Table 4: Notation of the elements of the LCA calculation

4.3 Mathematical conventions

Baseline mathematical conventions have so far been assumed, and a summary of them can be found in Table 1. The coefficients in matrices H, S A, B, and Q are stochastic but not marked in bold, because that would lead to confusion with the also capitalized single coefficient stochastic object.

Instead, the stochastic properties of the matrices are compiled in objects μ and Σ, which are also mentioned in Table 4. For example, $\mu_{Q_{ab}}$ represents the mean of Q_{ab} and $\Sigma_{Q_{ab}B_{cd}}$ refers to the correlation between Q_{ab} and B_{cd}. See also Table 2. In practice, coefficients within the same column of matrices A and B may be correlated, and coefficients present within one of the matrices A, B, or Q may be correlated. The coefficients in F are deterministic and will not exhibit uncertainty, but for calculation sake, as will be useful in Section 6, their value is also captured in parameter object μ_F, while covari-

ance with itself and all other matrices is zero. The *Mathematical Notation* section on page 3 presents the complete overview.

4.4 Non-linearity of the LCA model

We can see that the LCA model is distinctly *not* linear in two ways: (1) coefficients are multiplied, and (2) technology matrix A is inverted which is a non-linear mathematical operation. Applying LSA to LCA will lead to results of limited use. An analysis that leaves most of the input space unexplored, or whose approximations are invalid for larger perturbations, draws an incomplete picture of the LCA model behavior. As will be discussed in Section 5.2, SA's of the LCA model often assume linear behavior or will try to project the model onto a linear approximation.

The non-linearity of the LCA model can be simplified by the application of the geometric series for matrix inverses, also called the *Neumann expansion*. The problem turns into just an issue of multiplication and summation. This version of the LCA equation reads

$$H = \sum_{m}^{\infty} QB(I - A)^m F \tag{2}$$

where I is the identity matrix. The simplification applies when A is invertible and $\lim_{m\to\infty}(I - A)^m = 0$. The trick of using the geometric series in place of the matrix inverse is used for example in Input-Output analysis. In that kind of analysis, each instance of the power sum is equated to a *'production layer'* of the processes (Peters, 2007). For an exact result, the sum must go to infinity, but often the expression is evaluated up to a certain *order*. The geometric series will be used in Section 6.

5 Applying SA to LCA

In this section, the theory of SA as identified in Section 3 is applied to LCA as established in Section 4. First, the goals and appropriate methods from Section 3.1 and 3.2 respectively are considered. After this, the steps of SA as defined in Section 1 are filled out. Step 4, communication of results, is not a focus of this thesis and will therefore not be evaluated.

5.1 The goals of SA in LCA context

Applying SA to LCA can be done with different intentions in mind. We mirror the rationales identified in Section 3.1. In a comparative LCA, SA may focus on understanding how the variation in results affects the conclusions about which products have the lowest impacts. Attributional LCA may use SA to identify critical coefficients to focus future research on to reduce the model uncertainty (Bisinella et al., 2016). Ex-ante LCA could be interested in diagnosing processes that, when improved, could bring great reductions in ecological impact (Wei et al., 2016).

5.2 Appropriate SA methods currently in use in LCA

This section reflects on the use in LCA literature of appropriate SA methods as identified in Section 3.2. It concerns the issues of using a global SA perspective and of addressing correlations.

An extensive literature review by Saltelli et al. (2019) showed 42% of highly cited scientific papers do not properly explore a wider area of the input space rather than one-dimensional corridors. Generally, LCA studies focus on comparing worst- and best-case scenarios, doing one-at-a-time UP, or considering coefficient variation around a nominal value (Padey et al., 2013). Sometimes modeling choices are analyzed by applying scenarios, for example by comparing energy mix or allocation choices (Igos et al., 2019). These practices all fall under local UP.

When it comes to SA implementation in LCA software, Igos et al. (2019) note that today, LCA software tools most commonly offer the option of (global) Monte Carlo simulation for UA, but only the (local) one-at-a-time approach for SA.

Including the effects of correlations between coefficients is also not standard practice. Determining correlations of all input parameters is a major task in itself and data on covariance is therefore of limited availability. GSA methods tend to not account for correlations, and software tools lack functionality in this department (Heijungs et al., 2019; Igos et al., 2019). The traditional GSA method of Sobol sensitivity indices assumes coefficient independence (Wei et al., 2015). The Sobol indices are perhaps the most commonly used measures for global sensitivity. In brief, they reflect on changes in global model output uncertainty when re-propagating uncertainty while fixing one (or multiple, up to all but one) input coefficient.

Working with an uncorrelated model is considered easier. Saltelli et al. (2008) suggest working with uncorrelated models as much as possible and treating dependent coefficients *"as explicit relationships with a noise term."*(p.41). The effect of including correlations compared to excluding correlations in LCA models was tested by E. Groen & Heijungs (2017), assessing a small case study with only normally distributed parameters. It was shown that the effect of ignoring calculations in this system could be predicted with a second-order variance approximation. Additionally, they concluded that the effect is most pronounced for models with high ratios of covariances to means.

5.3 Characterizing the input uncertainties

The uncertainties present in an LCA study can be categorized in several ways. Igos et al. (2019) classify

1. Modeling assumptions; the structure of relationships between coefficients, or assumed scenarios. Under this fall choices on boundaries, allocation methods, and functional units (Mendoza Beltran et al., 2016)
2. Process data uncertainty; due to imperfect measurements of processes
3. Process data variability; due to the natural fluctuations of quantities in processes. An example of this is the energy use of transportation being affected by seasonal temperatures.
4. Uncertainties in characterization factors. These are derived from other types of models that describe the (ecological) impacts of externalities, hence they can carry uncertainties too.

Uncertainty can also be expressed in multiple ways, such as by probability distributions, moments (such as variance and mean), fuzzy sets, or scenarios (Igos et al., 2019). Uncertainties due to different categories may naturally have a particular expression, such as modeling assumptions being best represented as scenarios. Process data uncertainty can be represented by moments, but probability distributions are often preferred because some properties of the uncertain behavior are lost when reduced to a set of moments. Fuzzy sets are used in UP methods beyond the scope of this thesis.

For this thesis, information about the probability distributions of the input coefficients is assumed to be available, although in practice, as mentioned in Section 5.2, data for this may be lacking at present. The analysis in this thesis is focused on process data and characterization uncertainties. Such uncertainties can be described as continuous probability distributions. In LCA most probability distributions, such as those found in the Ecoinvent database, are lognormal (Igos et al., 2019). Lognormal distributions can be fully characterized by just two parameters: their mean and (co)variance. These assumptions will allow for deriving an explicit expression of the model variance in Section 6.

5.4 The uncertainty propagation

Lee & Chen (2009) distinguish five categories of UP approaches. The categories and some examples are shown in Table 5. Igos et al. (2019) show how three of these categories are typically applied to the LCA model. They first distinguish *'analytical'* and *'sampling'* approaches, which respectively correspond to Lee & Chen's local expansion- and simulation-based methods. *'Fuzzy logic'* is identified as a third method category. It however does not fall under the five presented categories, for it is a so-called *possibilistic* approach and possibilistic approaches are beyond the scope of this thesis. The first three categories are analyzed in the next sections. The second to last category (functional expansion-based methods) has not notably been applied to LCA as of yet. The Neumann expansion mentioned in Section 4.4 appears in this category. The last category (most probable point-based methods) is also beyond the scope of this thesis.

UP category	***Examples***
Numerical integration-based methods	*Full factorial numerical integration, dimension reduction*
Local expansion-based methods	*Taylor series, perturbation method*
Simulation-based methods	*Monte Carlo simulations, importance sampling, adaptive sampling*
Functional expansion-based methods	*Neumann expansion, orthogonal or Karhunen–Loeve expansion, polynomial chaos expansion, wavelet expansion)*
Most probable point-based methods	*First- and second-order reliability method*

Table 5: Categories of uncertainty propagation as distinguished by Lee & Chen (2009)

5.4.1 Numerical integration-based methods

Ideally one could calculate the entire distribution of the model output analytically without approximation or compromise. This would be a full factorial numerical integration. Uncertainties are then propagated by mapping probability distribution functions of the input data analytically to the distribution of the output. The advantage of this is that it could provide the LCA practitioner with an explicit expression for the relationships between the model input and output uncertainty.

The equation for mapping the distribution of a random n-dimensional vector onto the m-dimensional distribution of a function of that vector is as follows (Wikibooks, 2021):

$$\rho_{\vec{\boldsymbol{Y}}}(\vec{y}) = \frac{\partial}{\partial y_1} \cdots \frac{\partial}{\partial y_m} \int_{\{\vec{x} \in \mathbb{R}^n | \vec{f}(\vec{x}) \leq \vec{y}\}} \rho_{\vec{\boldsymbol{X}}}(\vec{x}) d^n x \tag{3}$$

Where $\rho_{\vec{\boldsymbol{X}}}(\vec{x})$ is the correlated probability density function of the model input coefficients, which are compiled in multivariate stochastic variable $\vec{\boldsymbol{X}}$. $\rho_{\vec{\boldsymbol{Y}}}(\vec{y})$ the probability density function of the outputs $\vec{\boldsymbol{Y}}$.

Borgonovo et al. (2011) show several kinds of models with particular input coefficient distributions that can be evaluated analytically with this method. The answer to a direct probability density transformation exists only under specific circumstances. For example, the sum of coefficients with a normal distribution, or the product of coefficients with a lognormal distribution. Since the LCA model is a combination of multiplication and addition, an exact result does not exist for the evaluation of Equation 3. The integral does not evaluate at all, even when complicating factors such as correlations are ignored.

The lack of an exact solution to the probability distribution of a sum of (correlated) lognormal coefficients is a well-known problem in science, with many possible applications. Not much is known about such a distribution, but it is neither normal nor lognormal (Lo, 2012). However, there are several ways to approximate a sum of lognormal coefficients.

The Fenton-Wilkinson approximation (Fenton, 1960; Abu-Dayya & Beaulieu, 1994) is one of the oldest and most commonly used methods. It makes use of *'moment matching'*, which starts by choosing an uncertainty distribution shape resembling the output distribution. Then a set of moments of the actual output distribution is calculated analytically. The parameters of the predetermined distribution shape are chosen in such a way that its moments mirror the calculated moments.

For example, the distribution of a sum of independent lognormal variables $\boldsymbol{X}_i$ is fitted to the lognormal distribution $\boldsymbol{Z}$. The relationships between the parameters are

$$V[e^{\boldsymbol{Z}}] = \ln\left(\frac{\sum_i \exp(2\mu_i + \sigma_i^2)(\exp(\sigma_i^2) - 1)}{(\sum_i \exp(\mu_i + \sigma_i^2/2))^2} + 1\right)$$

$$E[e^{\boldsymbol{Z}}] = \ln\left(\sum_i \exp(\mu_i + \sigma_i^2/2\right) - V[e^{\boldsymbol{Z}}]/2$$

Where $\mu_i = E[e^{\boldsymbol{X}_i}]$ and $\sigma_i^2 = V[e^{\boldsymbol{X}_i}]$

The approximation makes use of the fact that the variance of a sum of uncertain coefficients is defined explicitly in terms of the input variances, no matter the shape of the coefficient distributions. This allows for an analytical matching of the expressions for the moments.

In the approximation, correlations between variables are not included. The structure of the LCA model is more complex than the sum of uncorrelated lognormal coefficients, but the Fenton-Wilkinson approach is nevertheless a good starting point for an analytical expression of the model uncertainty and it will be expanded on in Section 6.

5.4.2 Local expansion-based methods

Perturbation analysis is the first type of straightforward approximation of uncertain models. In non-linear models perturbation is not valid for extrapolation over a larger input space, it does not allow for the inclusion of correlation, and it does not implement the actual uncertainty of input coefficients. It does however give intuitive insight into the model behavior. An analytical expression for perturbation of the LCA model is

$$H + \Delta H = (Q + \Delta Q)(B + \Delta B)(A + \Delta A)^{-1}F \tag{4}$$

The issue with perturbing a non-linear model becomes evident when trying to isolate the expression for ΔH. This quickly becomes a cumbersome formula that also is dependent on the mean values of the input coefficients. Perturbation only gives a limited view of the model behavior.

Derivatives and Taylor approximations The variance of the LCA output is often approximated with a Taylor approximation to the first order, or in a couple of papers to the second order. The Taylor approximation makes use of partial derivatives. A second-order approximation can express quadratic effects in addition to the linear effects of the first-order approximation. Low-order approximations can only account for small uncertainty ranges, but higher-order approximations quickly lead to complicated formulas.

The next equation shows the first and second-order approximations of the variance of the LCA model output (Ciroth et al., 2004). The equation assumes that the input coefficients are normally distributed and uncorrelated. This assumption is not ideal for LCA, as has been discussed in Section 4.1 and 5.3, and the low order approximations are only valid locally. This is reflected in the fact that Ciroth et al. (2004) determine the approximation breaks down when uncertainty becomes too great. The expression reads

$$V[H] \approx \overbrace{\sum_i \left(\frac{\partial H}{\partial X_i}\right)^2 V[X_i]}^{\text{first order}} + \underbrace{\sum_{i,j}\left[\frac{1}{4}\frac{\partial^2 H}{(\partial X_i)^2}\frac{\partial^2 H}{(\partial X_j)^2} + \frac{1}{2}\left(\frac{\partial^2 H}{\partial X_i \partial X_j}\right)^2 + \frac{\partial H}{\partial X_i}\frac{\partial^3 H}{\partial X_i(\partial X_j)^2}\right] V[X_i]V[X_j]}_{\text{second order}}$$

Where $X_i \in \{A, B, Q | A \in \mathbb{R}^{p\times p}, B \in \mathbb{R}^{n\times p}, Q \in \mathbb{R}^{m\times n}\}$, or put into words $\vec{X}$ is an object representing all of the model input coefficients.

A different approach that does include covariance and is not restricted to coefficients with normal distributions, is presented by Heijungs (2010). The formula is a first-order

approximation of the variance

$$V[H] \approx \sum_i \left(\frac{\partial H}{\partial X_i}\right)^2 V[X_i] + 2\sum_{ij} \frac{\partial^2 H}{\partial X_i \partial X_j} Cov[X_i, X_j]$$

Another use of the first-order Taylor approximation was proposed by Hong et al. (2010). Their approximation assumes lognormal behavior of the input coefficients, just like we decided on, but assumes it also of the output, which is not given. It can not address negative lognormal coefficients that may appear in avoided processes and such. Therefore this approach is also of limited use.

Imbeault-Tétreault et al. (2013) state that there is a need to derive analytical expressions of model uncertainty valid for higher output uncertainties for example by using higher-order approximations that do account for both the additive and multiplicative properties of the models.

5.4.3 Simulation-based methods

Simulation-based methods present a much more unambiguous type of UP for LCA. Random samples are generated using the characterized input uncertainty properties. The model is evaluated for each sample, creating an output distribution. Abilities not always taken advantage of is that simulation-based methods can sample the entire input space, and account for correlations as well. Notably, the covariance of different LCA reference flow impacts can only be calculated when comparing output sets for which the same data set of input samples is used.

Simulation-based methods ignore information available about the structure of the model. Uncertainty is found by recalculating the model multiple times. This black-box approach to UP is useful for highly complex models for which more sophisticated analysis is not an option. High numbers of simulations are required for representative results, up to 10.000 runs (Heijungs & Lenzen, 2014). This may cause the calculation of a single UP the take up to several hours. Results may not even always converge. Purely random sampling is called Monte Carlo simulation. For a more effective sampling of the input space, strategies such as Quasi-Monte Carlo and Latin Hypercube sampling are available (Igos et al., 2019). With these methods, attempts are made at making sampling a viable strategy for SA on large amounts of uncertain parameters.

5.5 The characterization of sensitivity

In this step of SA, the by UP acquired uncertainty data for several combinations of input parameters is plugged into a sensitivity measure. Different characterizations of UP results are again only suitable for particular SA measures. Therefore UP methods and SA measures tend to come together. Commonly, uncertainty in terms of variance is used as a basis for sensitivity (Pianosi & Wagener, 2015), but distribution-based methods also exist (Borgonovo et al., 2012; Cucurachi et al., 2016).

Wei et al. (2016) describe that GSA is rarely applied to LCA, despite publications on why it is important and how to use it (Di Lullo et al., 2020). Examples of strategies are *analysis of variance*, the *elementary effects method*, *global derivative-based measures*, *moment-independent methods* (Cucurachi et al., 2016), *variogram-based approaches* and many others (Saltelli et al., 2019). Noteworthy is that the elementary effects and global derivative-based methods both use first-order partial derivatives. This is a local kind of SA that cannot address coefficient interactions and correlations by itself. The methods are classified as global nonetheless because the derivatives are sampled globally throughout the input space.

Some GSA methods specifically researched for use in LCA are the Sobol indices, the standard regression coefficients, and random balance design (E. A. Groen et al., 2017). These methods all operate on variance which is acquired from global simulation-based UP. An analytically acquired expression of variance could also be plugged into these methods. They however are not equipped for assessing the effects of correlation. Appropriate measures must be part of the comprehensive analytical approach to SA.

6 Analytical sensitivity expression

With all the compiled information of the previous sections, we can proceed with constructing an analytical UP expression for LCA. Using further mathematical manipulation, such an expression can be translated to sensitivity.

As established in Section 5.4.1, the probability density of the output of the LCA model cannot be propagated exactly. The next best route would be to use the also aforementioned moment matching, described in that same section. Performing moment matching does not add to the knowledge of the analyst (it only helps with interpretation), so we will simply start with applying its logic to calculate the model covariance. This section will follow the same logic as the Fenton-Wilkinson approximation, expanded for the additive properties and the potential correlations of the LCA model. Combining this with the Neumann expansion, or geometric series, for the matrix inverse, makes our approach a nod to the functional expansion-based methods.

6.1 Mathematical basis

To be able to construct a formula for the covariance of the model, some mathematical relationships need first to be defined and explained. For an overview of the basic mathematical notations see Tables 1 and 2.

6.1.1 Lognormal distributions

We establish vector $\vec{\boldsymbol{X}}$, containing lognormally distributed and correlated coefficients. $\vec{\boldsymbol{X}}$ is called a *lognormal multivariate*. Then, using an underline, we define $\underline{\vec{\boldsymbol{X}}} = \ln(\vec{\boldsymbol{X}})$, which is by definition a *normal multivariate*. All lognormal distributions have a related normal distribution like this. In mathematical terms

$$\begin{aligned}
\vec{\boldsymbol{X}} &= \exp(\underline{\vec{\boldsymbol{X}}}) \\
\vec{\boldsymbol{X}} &\sim \mathcal{LN}(\mu_X, \Sigma_{XX}) \\
\underline{\vec{\boldsymbol{X}}} &\sim \mathcal{N}(\mu_{\underline{X}}, \Sigma_{\underline{X}\underline{X}})
\end{aligned}$$

The relationships between the parameters (mean and covariance) of $\vec{\boldsymbol{X}}$ and $\underline{\vec{\boldsymbol{X}}}$ are defined explicitly (Halliwell, 2015). First, the mean and covariance of $\vec{\boldsymbol{X}}$ in terms of the mean and covariance of $\underline{\vec{\boldsymbol{X}}}$ are

$$\begin{aligned}
\mu_{X_i} &= \exp\left(\mu_{\underline{X}_i} + \frac{1}{2}\Sigma_{\underline{X}_i\underline{X}_i}\right) \\
\Sigma_{X_iX_j} &= \exp\left(\mu_{\underline{X}_i} + \frac{1}{2}\Sigma_{\underline{X}_i\underline{X}_i} + \mu_{\underline{X}_j} + \frac{1}{2}\Sigma_{\underline{X}_j\underline{X}_j}\right)(\exp\left(\Sigma_{\underline{X}_i\underline{X}_j}\right) - 1) \\
&= \mu_{X_i}\mu_{X_j}(\exp\left(\Sigma_{\underline{X}_i\underline{X}_j}\right) - 1)
\end{aligned} \tag{5}$$

These equations can be used to derive the parameters of $\vec{\underline{\boldsymbol{X}}}$ in terms of the mean and covariance of $\vec{\boldsymbol{X}}$

$$\begin{aligned}\Sigma_{\underline{X}_i\underline{X}_j} &= \ln\left(\frac{\Sigma_{X_iX_j}}{\mu_{X_i}\mu_{X_j}}+1\right)\\ \mu_{\underline{X}_i} &= \ln(\mu_{X_i}) - \frac{1}{2}\ln\left(\frac{\Sigma_{X_iX_i}}{\mu_{X_i}\mu_{X_i}}+1\right)\\ &= \ln(\mu_{X_i}) - \frac{1}{2}\Sigma_{\underline{X}_i\underline{X}_i}\end{aligned} \tag{6}$$

Not all parameters of $\vec{\boldsymbol{X}}$ can be translated to parameters for $\vec{\underline{\boldsymbol{X}}}$ without a second thought. This stems from the fact that (1) lognormal distributions are not defined for variables with a mean smaller than or equal to zero, and in that (2) any covariance matrix has to be positive semi-definite.

Setting $\kappa = \frac{\Sigma_{X_iX_j}}{\mu_{X_i}\mu_{X_j}}$ and imagining different values for μ, then issue (1) is reflected in the equations as follows;

a) the logarithms return complex numbers whenever $\kappa < -1$ or $\mu < 0$;

b) the logarithms are not defined at $\kappa = -1$ or $\mu = 0$

c) κ itself will not be defined if either μ is zero;

Point (a) can be solved by allowing for complex-valued parameters of $\vec{\underline{\boldsymbol{X}}}$. As for (b) and (c), consider that $\lim_{x\downarrow 0}\ln(x) = -\infty$ and $\exp(-\infty) = 0$. A lognormal distribution of mean zero present in an LCA could for example mean a product not partaking in a certain process, a process not having a certain externality, or an externality not having a certain impact. The variance of these absent coefficients is zero by definition. The technical problem is mended by overriding the division, setting $\kappa = 0$, $\Sigma_{\underline{X}_i\underline{X}_j} = 0$ and allowing $\mu_{\underline{X}_i}$ to equal $-\infty$. When the parameters are transformed back and forth using these solutions, the properties of the distributions will be preserved.

Issue (2) is that of the structure of $\Sigma_{\underline{X}\underline{X}}$ and Σ_{XX}. Not all matrices are allowed to fulfill the role of covariance matrix. They must be positive semi-definite. Because the covariances of an implemented LCA model are given, assumed is that they will be correct in both lognormal and normal multivariate representation. This problem will however be of relevance again in Section 7.

6.1.2 The sum of correlated coefficients

The parameters of a sum of correlated coefficients are defined in terms of the parameters of the summed coefficients. This is irrespective of the distributions of the coefficients. The Fenton-Wilkinson approximation of Section 5.4.1 relies on the same principle. The

relationships are as follows:

$$\begin{aligned} E\left[\sum_m \boldsymbol{X}_m\right] &= \sum_m E\left[\boldsymbol{X}_m\right] \\ V\left[\sum_m \boldsymbol{X}_m\right] &= \sum_m \sum_n Cov\left[\boldsymbol{X}_m, \boldsymbol{X}_n\right] \end{aligned} \tag{7}$$

For example, when correlated coefficients expressed in the normal multivariate $\boldsymbol{X} \sim \mathcal{N}(\mu_X, \Sigma_{XX})$ are summed, the parameters of this sum become

$$\sum_m \boldsymbol{X}_m \sim \mathcal{N}\left(\sum_m \mu_{X_m}, \sum_{m,n} \Sigma_{X_m X_n}\right)$$

Note that the elementary relationship between covariance and expectation values is

$$Cov[\boldsymbol{X}, \boldsymbol{Y}] = E[\boldsymbol{XY}] - E[\boldsymbol{X}]E[\boldsymbol{Y}] \tag{8}$$

It is perhaps useful to interject the following consequence. When the covariance of dummy matrix $\Psi = QBA^{-1}$ (without F) is found in whatever way, the covariances between reference flows become very easy to calculate. Because F contains only deterministic coefficients, the impacts of the reference flows are different linear combinations of the elements of Ψ. This simply makes the output covariances linear combinations of the covariances of Ψ too.

6.1.3 The product of lognormal correlated coefficients

Next, equations are established for the parameters of a product of correlated lognormal coefficients $\boldsymbol{Z} = \prod_m \boldsymbol{X}_m \sim \mathcal{LN}(\mu_Z, \sigma_Z^2)$ where $\vec{\boldsymbol{X}} \sim \mathcal{LN}(\mu_X, \Sigma_{XX})$.

$$\begin{aligned} \text{Let } \Gamma &= \prod_{m_1 \neq m_2} \left(\frac{\Sigma_{X_{m_1} X_{m_2}}}{\mu_{X_{m_1}} \mu_{X_{m_2}}} + 1\right) \\ \text{then } \mu_Z &= \prod_m \mu_{X_m} \sqrt{\Gamma} \\ \text{and } \sigma_Z^2 &= \prod_m (\Sigma_{X_m X_m} + \mu_{X_m}^2)\Gamma^2 - \mu_Z^2 \end{aligned} \tag{9}$$

Derivation of these equations can be found in Appendix B. In a different, perhaps more general mathematical notation, the expectation value of a product of correlated lognormal coefficients can be written as

$$E[\boldsymbol{X}_1 ... \boldsymbol{X}_N] = \prod_m E[\boldsymbol{X}_m] \prod_{i \neq j} \sqrt{\frac{Cov[\boldsymbol{X}_i, \boldsymbol{X}_j]}{E[\boldsymbol{X}_i]E[\boldsymbol{X}_j]} + 1} \tag{10}$$

This formula can be seen as a version of the Isserlis theorem, which expresses the expectation of a product of normal correlated variables (Isserlis, 1918; Leonov & Shiryaev, 1959).

6.1.4 The LCA model in product sum form

By applying the geometric series formula (Equation 2) for the inverse of A as we did in Section 4.4, we assume it is possible to write the LCA calculation as

$$H_{ij} = \sum_{abcm} Q_{ia} B_{ab} (I - A)^m_{bc} F_{cj} \tag{11}$$

This formulation of the model as a sum of products allows us to find an exact expression for the output covariance using the equations from the previous sections. When applying Equation 7, the model covariance can be written as

$$Cov[H_{ij}, H_{kl}] = \sum_{abcm} \sum_{defn} Cov\left[Q_{ia} B_{ab} (I - A)^m_{bc} F_{cj}, Q_{kd} B_{de} (I - A)^n_{ef} F_{fl}\right] \tag{12}$$

For brevity $I - A$ is called W from these equations on. Before we can continue, three issues with this model covariance formulation must be addressed:

(1) The distribution of the diagonal coefficients in matrix W needs extra thought; a constant minus a lognormal variate results not in a lognormal distribution. The problem is mended by starting with a normalized technology matrix (of which the diagonal elements are 1 with variance 0). In which case
$\mu_{W_{ij}} = \begin{cases} -\mu_{A_{ij}} & \text{if } i \neq j. \\ 0 & \text{if } i = j. \end{cases}$ and $\Sigma_{WW} = \Sigma_{AA}$ and $\Sigma_{AB} = \Sigma_{WB}$

(2) By applying the geometric series expansion we assume the inverse of A exists and is calculable by this series, something which may not always be the case.

(3) Additional summations arise because of raising W to powers m and n. For example for the element of the sum where $m = 3$ and $n = 2$, $Cov\left[Q_{ia} B_{ab} W^3_{bc} F_{cj}, Q_{kd} B_{de} W^2_{ef} F_{fl}\right]$ becomes

$$\begin{aligned} &Cov\left[\sum_{xy} Q_{ia} B_{ab} W_{bx} W_{xy} W_{yc} F_{cj}, \sum_{z} Q_{kd} B_{de} W_{ez} W_{zf} F_{fj}\right] \\ &= \sum_{xyz} Cov\left[Q_{ia} B_{ab} W_{bx} W_{xy} W_{yc} F_{cj}, Q_{kd} B_{de} W_{ez} W_{zf} F_{fj}\right] \end{aligned} \tag{13}$$

6.2 Uncertainty propagation

Let the following objects represent sets of input coefficients of the LCA model, as present in single elements of the sum in Equation 11

$$\Omega_{(ij)(abx_1 \ldots x_m c)} = \{Q_{ia}, B_{ab}, W_{bx_1}, ..., W_{x_m c}, F_{cj}\} = \Omega_p$$
$$\Omega_{(kl)(dey_1 \ldots y_n f)} = \{Q_{kd}, B_{de}, W_{ey_1}, ..., W_{y_n f}, F_{fl}\} = \Omega_q$$

so that, using this set notation, the LCA model can be written as

$$H_{ij} = \sum_{mabx_1 \ldots x_m c} \left(\prod_{X \in \Omega_{(ij)(abx_1 \ldots x_m c)}} X \right) = \sum_p \left(\prod_{X \in \Omega_p} X \right) \tag{14}$$

Note that $\sum_{mabx_1 \ldots x_m c}$ is shortened to $\sum_p$. This turns the model covariance equation into the following expression

$$\begin{aligned} Cov[H_{ij}, H_{kl}] &= \sum_{mn} \sum_{\substack{abx_1 \ldots x_m c \\ dey_1 \ldots y_n f}} Cov \left[\prod_{X \in \Omega_{(ij)(abx_1 \ldots x_m c)}} X, \prod_{Y \in \Omega_{(kl)(dey_1 \ldots y_n f)}} Y \right] \\ &= \sum_{mn} \sum_{pq} Cov \left[\prod_{X \in \Omega_p} X, \prod_{Y \in \Omega_q} Y \right] \end{aligned}$$

Using the definition of covariance from Equation 8 this becomes

$$Cov[H_{ij}, H_{kl}] = \sum_{mn} \sum_{pq} E \left[\prod_{Z \in \Omega_p \cup \Omega_q} Z \right] - \sum_{mp} E \left[\prod_{X \in \Omega_p} X \right] \sum_{nq} E \left[\prod_{Y \in \Omega_q} Y \right] \tag{15}$$

Equations 9 are used to define

$$\begin{aligned} \Gamma_p &= \prod_{\{X_s, X_t\} \in \Omega_p^2} \left(\frac{\Sigma_{X_s X_t}}{\mu_{X_s} \mu_{X_t}} + 1 \right) \\ \Gamma_q &= \prod_{\{Y_s, Y_t\} \in \Omega_q^2} \left(\frac{\Sigma_{Y_s Y_t}}{\mu_{Y_s} \mu_{Y_t}} + 1 \right) \\ \Gamma_{p \cup q} &= \prod_{\{Z_s, Z_t\} \in (\Omega_p \cup \Omega_q)^2} \left(\frac{\Sigma_{Z_s Z_t}}{\mu_{Z_s} \mu_{Z_t}} + 1 \right) \end{aligned}$$

Where the product is overall possible off-diagonal pairings $\{(.)_s, (.)_t\}$ of elements of the set product $\Omega^2_{(.)}$. Expression 15 becomes

$$\begin{aligned} Cov[H_{ij}, H_{kl}] &= \sum_{mn} \sum_{pq} \prod_{Z \in \Omega_p \cup \Omega_q} \mu_Z \sqrt{\Gamma_{p \cup q}} - \sum_{mp} \prod_{X \in \Omega_p} \mu_X \sqrt{\Gamma_p} \sum_{nq} \prod_{Y \in \Omega_q} \mu_Y \sqrt{\Gamma_q} \\ &= \sum_{mn} \sum_{pq} \left(\prod_{Z \in \Omega_p \cup \Omega_q} \mu_Z \sqrt{\Gamma_{p \cup q}} - \prod_{X \in \Omega_p} \mu_X \sqrt{\Gamma_p} \prod_{Y \in \Omega_q} \mu_Y \sqrt{\Gamma_q} \right) \end{aligned}$$

Since $\prod_{Z\in\Omega_p\cup\Omega_q}\mu_Z = \prod_{X\in\Omega_p}\mu_X\prod_{Y\in\Omega_q}\mu_Y$, the next step towards simplification is

$$Cov[H_{ij},H_{kl}] = \sum_{mn}\sum_{pq}\left(\prod_{Z\in\Omega_p\cup\Omega_q}\mu_Z\right)\left(\sqrt{\Gamma_{p\cup q}}-\sqrt{\Gamma_p\Gamma_q}\right) \tag{16}$$

This is our final expression for the covariance of the geometric series approximated LCA model. It is (1) inclusive of the non-linear model behavior, (2) includes correlations, and (3) shows explicit relationships between in- and output variance.

6.3 Sensitivity characterization

Now the expression for model covariance is defined, it can be used to calculate sensitivities.

The first option of GSA is to evaluate the entire equation various times, changing one or more input covariance parameters. The results could be applied to any SA measure that is based on variance. As the UP incorporates covariance relationships, the SA measure should be able to reflect on those too. This course of action mirrors the process of a sampling GSA. But another way to reflect on GSA is enabled by the analytical expression too. Since the goal of SA is to ascribe responsibility for the output uncertainty to the input coefficients, this could simply be determined by closely examining Equation 16. This way it would not be necessary to evaluate every part of it repeatedly.

Such close examination starts as follows. If any Σ_{XY} is set to zero, the unit $\left(\frac{\Sigma_{XY}}{\mu_X\mu_Y}+1\right)$ of which it is part turns to 1. Since these units appear in a *product* to constructs a certain Γ, it can be concluded that the influence of Σ is removed, which recognizes that the formula performs as it should. Further manipulation of Equation 16 is required to fully understand such interactions.

Partial derivatives of the equation with respect to a covariance can say something about the sensitivity of the model to changes of said covariance. These partial derivatives also follow directly from Equation 16. The sensitivity measures can be constructed from this.

7 Verification by comparison with simulation

The performance of Equation 16 is verified by comparing its results to a sampling strategy. This verification is sufficient to show that the analytically calculated covariance is also suitable for use in sensitivity measures because it simply proves the validity of the equation. This section will detail the implementation of the code. The written Python code is included in Appendix C.

The verification process is split into two parts; (1) verification of the parameters of a model that is purely matrix multiplicative; modeled as a multiplication of three input matrices $H = ABC$. Part (2) is verification of the formula for the inversion of a matrix $H = A^{-1}$, utilizing the geometric series as demonstrated in Equations 2 and 11. In both parts (1) and (2) the covariance of the output matrix is calculated analytically and compared to a simulated covariance. The division of verification is chosen to separate the performance of the equation for matrix multiplication and inversion.

7.1 The process

Step-wise, the procedure that both programs follow is

1. Initialization
    a) Define model size and other parameters, such as the number of simulations and the order of the geometric series
    b) Randomly select model parameters
2. Simulation
    a) Generate lognormal multivariate samples
    b) For each sample, calculate model output
    c) Calculate model covariances
3. Perform the analytical calculation of covariance
4. Compare simulated and calculated covariances

Each program run will generate a different set of model parameters for which the simulation and covariance calculation will be performed. This is done in order to verify the formula in the most general sense, and not for a single dummy model. Caution must be taken when randomly selecting the covariance matrix. As mentioned in 6.1.1, such matrices need to be positive-semidefinite, and this is not always given when transforming between normal and lognormal model parameters. Additional caution is required when selecting the mean parameters of the matrix that is to be inverted because not all matrices are invertible via geometric series.

The lognormal multivariate coefficients are sampled taking the exponential of a complex-valued normal multivariate. This uses a transformation of the lognormal parameters into appropriate normal ones following Equation 5. A library for complex normal multivari-

ate sampling is not included in normal Python packages, therefore functions that enable such sampling come as an additional deliverable with the two verification programs. The complex-valued normal multivariate approach allows for the inclusion of lognormal coefficients with a mean smaller than or equal to zero.

7.2 Discussion of the results

Sampling of the lognormal input coefficients and subsequent calculation of the sampling output covariance runs smoothly in both programs. The analytical calculation of the covariance is a much slower process. This can be ascribed to the fact that the procedure is not optimized the way that the established sampling and matrix multiplication functions are. Especially the calculation time of the inverted matrix covariance is significant and grows exponentially with the geometric series order. Therefore increasing accuracy by choosing a higher order is limited in this small-scale testing application.

Figure 2 displays the ratios of analytically calculated to sampled output covariances for one run of each of the programs. The covariance ratios are color-coded so that ratios differing from 1 stand out, which indicates that the sampling and analytically calculated covariance do not match. Note that the color scaling is different for the individual figures.

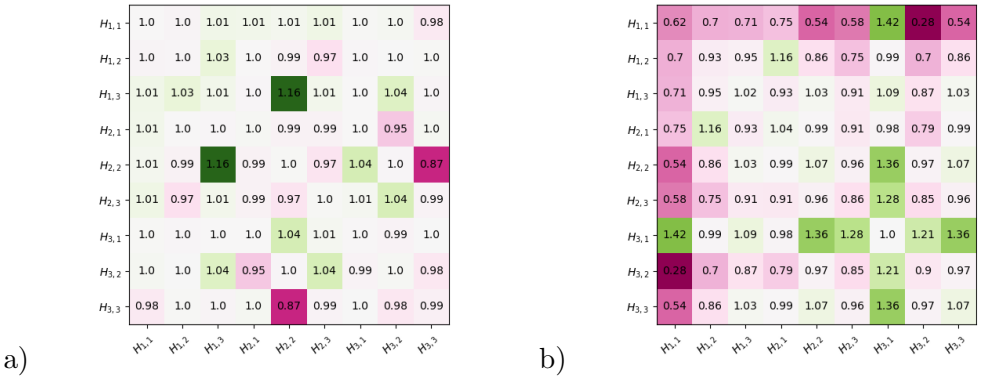


Figure 2: Ratio of analytically calculated to sampled covariance of (a) the product of three matrices and (b) the inverse of a matrix. The input coefficients are lognormally distributed and correlated. The output matrix H is in both cases of size 3×3. The correlation ratios are between elements of H as indicated by the labels on the left and on the bottom.

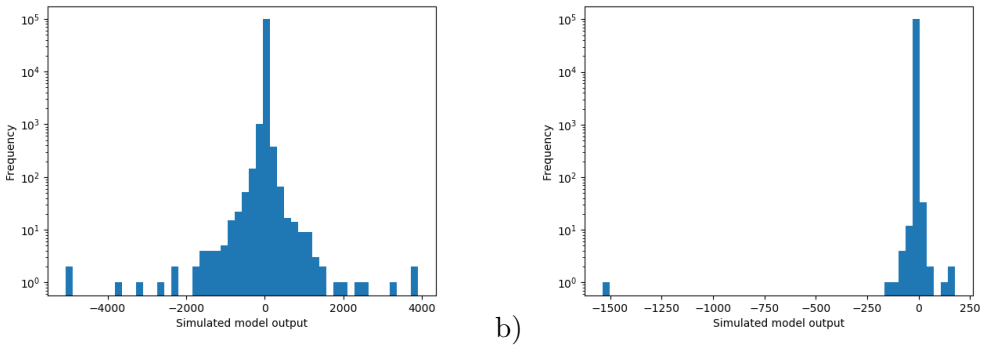
The matrix product results show great overlap, which can be seen in the program run displayed in Figure 2a. The matrix inverse procedure is less accurate (Figure 2b), and these results can differ a lot between program runs. Not every program run generates a random matrix for which the geometric series converges quickly, and with limits on the maximal order, this leads to less accurate results. Some program runs even generate erratic results. To give a sense of this beforehand, the inverse program starts with giving an indication of the convergence by printing the ratio A/B of the mean matrix

as inverted by geometric series (A) to the mean matrix inverted algebraically (B). Since in an actual LCA there is no random sampling of coefficient parameters, this should not be as big of an issue.

Some problems are identified when larger covariances are simulated, or when the inverted matrix samples behave unexpectedly even when the geometric series should have converged.

- When the randomly selected model input covariances are larger, the coefficient interactions can sometimes result in extreme outliers
- Sometimes in an overall stable matrix inverse simulation, also a couple of extreme outliers appear.

To illustrate the issues, examples of output distributions exhibiting the behavior are presented in Figure 3. Both are based on simulation sizes of 100.000. Mind that the vertical scaling of these figures is lognormal. The lognormal scaling is chosen to bring attention to the outliers, whose bins are stacked with only one or two results. Otherwise, the distributions would look like single tall pillars, and their tails would be invisible.



a) b)

Figure 3: Histogram of a simulated output of (a) the product of three matrices and (b) the matrix inverse, of which the input coefficients are correlated and lognormally distributed. Only a single of the 9 possible outputs was chosen for these figures

The outliers have a profound influence on the model covariance. The characterization of these distributions through variance alone gives odd results. The variance of the distribution in Figure 3a is **3264.1**, while the analytically calculated variance was **8900.7**, a factor **2.7** higher. That factor is relatively small considering the improbably large covariance numbers. As has been mentioned in 5.4.3, LCA sampling results do not always converge. Larger amounts of samples do not necessarily lead to more stable results, which is likely what has happened here.

The variance of the distribution in Figure 3b is **25.6**, but when the single outlier at **-1500** is eliminated, it reduces to only **1.94**. The outlier of the matrix inversion is due to the extreme behavior of near-singular matrices. Heijungs (2002) shows an example

of such a matrix. In large amounts of matrix samples, sometimes such matrices will appear and throw off the results. Nonetheless, if a near singular matrix can appear in the results, a sensitivity measure should be able to identify its risk.

The output variance being a sensible measure of uncertainty is implicitly assumed by variance-based SA. But clearly, the measure of covariance is not always great for describing systems like these. This is a limitation of the proposed analytical equation (Pianosi & Wagener, 2015). Overall, no definitive conclusions can be drawn on the performance of Equation 16 and more in-depth research is necessary. Not only to verify in which uncertainty domains it performs best, but also to experiment with how well the geometric series approach behaves.

8 Discussion

This section forms the resolution of the thesis. The research sub-questions are revisited and explicitly answered, after which the benefits and limitations of this work are summarized, and suggestions for further research are presented.

8.1 Answers to the research questions

(1) What are the goals, process, and concerns of SA? Sensitivity analysis determines which coefficients are responsible for the uncertainty in the outputs of models. Care must be taken in using appropriate SA methods for models that exhibit non-linear behavior, or which have correlated coefficients. A global SA approach is essential.

(2) What are the structural components and properties of the LCA model? The LCA model calculates impacts by combining various matrices filled with correlated coefficients through a non-linear process of addition, multiplication, and matrix inversion. The non-linearity can be simplified by substituting the matrix inversion with an expansion.

(3) How do we apply the established properties of SA to LCA? SA methods for LCA must be global and must be able to include the correlation. Current methods often lack these attributes. The LCA inputs can be characterized as lognormally distributed and correlated. Established methods do not adequately assess the LCA model properties and an opening for improvement is found by combining elements from functional expansion, elements from numerical integration, and elements from local expansion-based methods.

(4) What is the expression of uncertainty satisfying the posed requirements? The qualities of the lognormal inputs of the LCA model can be combined with its product-sum structure and the identified knowledge on UP methods. This comes together in an exact analytical expression (Equation 16) for the model covariance which includes correlations and should be valid for larger uncertainties.

(5) How does the found expression perform compared to uncertainty calculated using a random sampling method? The equation performs well when applied to a simple matrix multiplication model. The matrix inverse is computationally intensive to calculate, but the expression can perform decently for it. Issues arise when it comes to characterizing the output uncertainty when the model shows erratic behavior due to large uncertainties, or near-singular matrices.

8.2 Benefits and limitations

This work has proposed an analytical formulation of the output uncertainty of the LCA model. It is different from previous approaches in that (1) correlations are allowed for, (2) it takes into account the multiplicative *and* additive properties of the model, and (3) it is an exact expression. The expression assumes the invertibility of the technology matrix through the geometric series. It also assumes the inputs of the LCA model can be described by lognormal distributions.

Application of the equation to everyday LCA SA practice is, for now, not attractive. UP with it is more time-consuming than performing simulations, thought there is room for optimization. The expression does open up possibilities of making SA much easier in the future, because it shows the explicit relations between input and output covariance.

One limitation is that, as we have also observed ourselves in Section 7.2, variance is not always a proper indicator of uncertain behavior. Since the proposed equation can not produce a probability distribution of the output, this is a point of concern. Additionally, the expression is only exact when the geometric series order is infinite. This is not feasible for a straightforward evaluation of the equation. At lower orders, it is only a good approximation, depending on the convergence of the inverted matrix.

8.3 Future research recommendations

Since the results of this work are not usable for immediate application to LCA, this section will lay out the fronts on which improvements can be made, and where further work is necessary.

First of all, there is room for further simplification of formula 16. Making it more readable and therefore more transparent will aid in the understanding of all the interactions taking place. First of all, this can be done by introducing tensor-notation, which is the higher-dimensional cousin of matrix notation. As a second option, I suggest decomposing covariance parameters Σ. This involves writing Σ as AA^T, which is a decomposition that is available to positive-semidefinite matrices such as covariance objects. The symmetry may allow to further pick the equation apart and simplify it. Simplification may also aid in increasing the calculation speed of the equation when using it to explicitly calculate covariances as we did in Section 7.

The last step of SA steps as defined by Igos et al. (2019) was the communication of results. This point has not been addressed but is of course relevant when it comes to the actual use of the proposed expression for SA. Further work must be done on this topic.

Miscellaneous areas of interest Because SA is such a large topic, not all aspects have been explored in this thesis; there are plenty more methods to adapt for LCA if so desired. There are also other avenues of interest that this work was forced to gloss over due to scope limitations. Suggestions for future research are compiled in this section.

More UP methods of the categories *'functional expansion-based methods'* and *'most probable point-based methods'* (Section 5.4) have the potential for being solutions to address LCA sensitivities. Research could focus on these propagation categories. The Neumann expansion has only been used in this work to simplify the matrix inversion, but more literature exists on the topic (Yamazaki et al., 1988).

Expressions for higher output distribution moments (such as skewness and kurtosis) of the LCA model can be attained with the same procedure as performed in Section 6. Just like with covariance, the higher-order moments of products of lognormal multivariates have definitive expressions, as do the higher-order moments of sums of multivariates (e.g. $Skewness\left[\sum_i X_i\right] = \sum_{ijk} Coskewness\left[X_i, X_j, X_k\right]$). The expressions may be cumbersome to fully flesh out, but there is potential for better characterization of the model output distribution.

Moment matching to a lognormal distribution shape using the proposed expression is not recommended. This approach would finalize the Fenton-Wilkinson approximation. The output of the LCA model likely does not match the lognormal shape (as we have seen in Figure 3), but research could possibly focus on finding another distribution shape that is suitable for that purpose. Imbeault-Tétreault et al. (2013) too suggests searching for a type of probability distribution that can represent the simulated uncertainty of the LCA model with lognormal coefficients.

The Wishart distribution is a probability distribution that characterizes the uncertainty of a particular type of matrix. If a matrix uncertainty can be characterized as such, its inverse also has a defined distribution, called the Wishart Inverse. These distributions might be used for a better understanding of the distributions of the LCA model.

Partitioning the LCA matrices can simplify the model too. Encountered but not implemented was a formulation of the LCA model equations that separated the LCA background from the foreground processes and externalities. This separation may allow for a reduction of computational complexity by isolating parts of the process that can be pre-calculated. Such an expression would read

$$
\begin{aligned}
H = QBA^{-1}F &= Q\begin{pmatrix} B_1 & B_2 \end{pmatrix}\begin{pmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{pmatrix}^{-1}\begin{pmatrix} F_1 \\ F_2 \end{pmatrix} \\
&= Q\begin{pmatrix} B_1 & B_2 \end{pmatrix}\begin{pmatrix} A_{11}^{-1} & -A_{11}^{-1}A_{12}A_{22}^{-1} \\ 0 & A_{22}^{-1} \end{pmatrix}\begin{pmatrix} F_1 \\ F_2 \end{pmatrix} \\
&= \underbrace{QB_1A_{11}^{-1}F_1}_{\text{All background coefficients}} \underbrace{- QB_1A_{11}^{-1}A_{12}A_{22}^{-1}F_2}_{\text{Mixed coefficients}} + \underbrace{QB_2A_{22}^{-1}F_2}_{\text{All foreground coefficients}}
\end{aligned}
$$

Where Q has remained the same, B_1 and A_{11} are matrices containing the background externalities and processes, B_2 and $\begin{pmatrix} A_{12} \\ A_{22} \end{pmatrix}$ contain the foreground externalities and processes, F_1 contains background reference flows and F_2 contains foreground reference flows.

Singular matrix behavior Lastly, it may prove useful if further research is done on the likelyhood of the technology matrix becoming singular during sampling. Such a measure could indicate also the sensitivity of the LCA model, since the inversion of near singular matrices leads to extreme outliers.

References

Abu-Dayya, A., & Beaulieu, N. (1994). Outage probabilities in the presence of correlated lognormal interferers. *IEEE Transactions on Vehicular Technology*, *43*(1). doi: 10.1109/25.282277

Bisinella, V., Conradsen, K., Christensen, T. H., & Astrup, T. F. (2016, 3). A global approach for sparse representation of uncertainty in Life Cycle Assessments of waste management systems. *The International Journal of Life Cycle Assessment*, *21*(3), 378–394. Retrieved from http://link.springer.com/10.1007/s11367-015-1014-4 doi: 10.1007/s11367-015-1014-4

Borgonovo, E., Castaings, W., & Tarantola, S. (2011, 3). Moment Independent Importance Measures: New Results and Analytical Test Cases. *Risk Analysis*, *31*(3), 404–428. Retrieved from http://doi.wiley.com/10.1111/j.1539-6924.2010.01519.x doi: 10.1111/j.1539-6924.2010.01519.x

Borgonovo, E., Castaings, W., & Tarantola, S. (2012, 6). Model emulation and moment-independent sensitivity analysis: An application to environmental modelling. *Environmental Modelling & Software*, *34*, 105–115. Retrieved from https://linkinghub.elsevier.com/retrieve/pii/S1364815211001617 doi: 10.1016/j.envsoft.2011.06.006

Ciroth, A., Fleischer, G., & Steinbach, J. (2004, 7). Uncertainty calculation in life cycle assessments. *The International Journal of Life Cycle Assessment*, *9*(4), 216. Retrieved from http://link.springer.com/10.1007/BF02978597 doi: 10.1007/BF02978597

Cucurachi, S., Borgonovo, E., & Heijungs, R. (2016, 2). A Protocol for the Global Sensitivity Analysis of Impact Assessment Models in Life Cycle Assessment. *Risk Analysis*, *36*(2), 357–377. Retrieved from http://doi.wiley.com/10.1111/risa.12443 doi: 10.1111/risa.12443

Di Lullo, G., Gemechu, E., Oni, A. O., & Kumar, A. (2020, 2). Extending sensitivity analysis using regression to effectively disseminate life cycle assessment results. *The International Journal of Life Cycle Assessment*, *25*(2), 222–239. Retrieved from http://link.springer.com/10.1007/s11367-019-01674-y doi: 10.1007/s11367-019-01674-y

Fenton, L. (1960). The Sum of Log-Normal Probability Distributions in Scatter Transmission Systems. *IEEE Transactions on Communications*, *8*(1). doi: 10.1109/TCOM.1960.1097606

Groen, E., & Heijungs, R. (2017, 1). Ignoring correlation in uncertainty and sensitivity analysis in life cycle assessment: what is the risk? *Environmental Impact Assessment Review*, *62*, 98–109. Retrieved from https://linkinghub.elsevier.com/retrieve/pii/S0195925516300907 doi: 10.1016/j.eiar.2016.10.006

Groen, E. A., Bokkers, E. A. M., Heijungs, R., & de Boer, I. J. M. (2017, 7). Methods for global sensitivity analysis in life cycle assessment. *The International Journal of Life Cycle Assessment*, *22*(7), 1125–1137. Retrieved from http://link.springer.com/10.1007/s11367-016-1217-3 doi: 10.1007/s11367-016-1217-3

Halliwell, L. J. (2015). The lognormal random multivariate. In *Casualty actuarial society e-forum, spring* (Vol. 5).

Heijungs, R. (2002). The use of matrix perturbation theory for addressing sensitivity and uncertainty issues in LCA. *Proceedings of The Fith International Conference on EcoBalance - Practical tools and thoughtful principles for sustainability*, 6–8.

Heijungs, R. (2010, 6). Sensitivity coefficients for matrix-based LCA. *The International Journal of Life Cycle Assessment*, *15*(5), 511–520. Retrieved from http://link.springer.com/10.1007/s11367-010-0158-5 doi: 10.1007/s11367-010-0158-5

Heijungs, R., Guinée, J. B., Mendoza Beltrán, A., Henriksson, P. J. G., & Groen, E. (2019, 9). Everything is relative and nothing is certain. Toward a theory and practice of comparative probabilistic LCA. *The International Journal of Life Cycle Assessment*, *24*(9), 1573–1579. Retrieved from http://link.springer.com/10.1007/s11367-019-01666-y doi: 10.1007/s11367-019-01666-y

Heijungs, R., & Lenzen, M. (2014). Error propagation methods for LCA - A comparison. *International Journal of Life Cycle Assessment*, *19*(7), 1445–1461. doi: 10.1007/s11367-014-0751-0

Hong, J., Shaked, S., Rosenbaum, R. K., & Jolliet, O. (2010, 6). Analytical uncertainty propagation in life cycle inventory and impact assessment: application to an automobile front panel. *The International Journal of Life Cycle Assessment*, *15*(5), 499–510. Retrieved from http://link.springer.com/10.1007/s11367-010-0175-4 doi: 10.1007/s11367-010-0175-4

Igos, E., Benetto, E., Meyer, R., Baustert, P., & Othoniel, B. (2019, 4). How to treat uncertainties in life cycle assessment studies? *The International Journal of Life Cycle Assessment*, *24*(4), 794–807. Retrieved from http://link.springer.com/10.1007/s11367-018-1477-1 doi: 10.1007/s11367-018-1477-1

Imbeault-Tétreault, H., Jolliet, O., Deschênes, L., & Rosenbaum, R. K. (2013, 8). Analytical Propagation of Uncertainty in Life Cycle Assessment Using Matrix Formulation. *Journal of Industrial Ecology*, *17*(4), 485–492. Retrieved from http://doi.wiley.com/10.1111/jiec.12001 doi: 10.1111/jiec.12001

International Standardization Organization. (2006). *ISO 14044:2006 Environmental management — Life cycle assessment — Requirements and guidelines* (1st ed.). Geneva: ISO. Retrieved from https://www.iso.org/standard/38498.html

Isserlis, L. (1918, 11). On a formula for the Product-Moment Coefficient of any order of a Normal Frequency Distribution in any number of variables. *Biometrika*, *12*(1-2), 134–139. Retrieved from https://doi.org/10.1093/biomet/12.1-2.134 doi: 10.1093/biomet/12.1-2.134

Jacques, J., Lavergne, C., & Devictor, N. (2006, 10). Sensitivity analysis in presence of model uncertainty and correlated inputs. *Reliability Engineering & System Safety*, *91*(10-11), 1126–1134. Retrieved from https://linkinghub.elsevier.com/retrieve/pii/S0951832005002231 doi: 10.1016/j.ress.2005.11.047

Lee, S. H., & Chen, W. (2009, 1). A comparative study of uncertainty propagation methods for black-box-type problems. *Structural and Multidisciplinary Optimization*, *37*(3), 239–253. Retrieved from http://link.springer.com/10.1007/s00158-008-0234-7 doi: 10.1007/s00158-008-0234-7

Leonov, V. P., & Shiryaev, A. N. (1959, 1). On a Method of Calculation of Semi-Invariants. *Theory of Probability & Its Applications*, *4*(3). doi: 10.1137/1104031

Lo, C. F. (2012). The Sum and Difference of Two Lognormal Random Variables. *Journal of Applied Mathematics*, *2012*, 1–13. Retrieved from http://www.hindawi.com/journals/jam/2012/838397/ doi: 10.1155/2012/838397

Magnus, J. R., & Neudecker, H. (1999). *Matrix Differential Calculus with Applications in Statistics and Econometrics (Revised Edition)*. John Wiley & Sons Ltd. Retrieved from https://www.janmagnus.nl/misc/mdc-ch18.pdf

Mendoza Beltran, A., Heijungs, R., Guinée, J., & Tukker, A. (2016, 2). A pseudo-statistical approach to treat choice uncertainty: the example of partitioning allocation methods. *The International Journal of Life Cycle Assessment*, *21*(2), 252–264. Retrieved from http://link.springer.com/10.1007/s11367-015-0994-4 doi: 10.1007/s11367-015-0994-4

Mendoza Beltran, A., Prado, V., Font Vivanco, D., Henriksson, P. J. G., Guinée, J. B., & Heijungs, R. (2018, 2). Quantified Uncertainties in Comparative Life Cycle Assessment: What Can Be Concluded? *Environmental Science & Technology*, *52*(4), 2152–2161. Retrieved from https://pubs.acs.org/doi/10.1021/acs.est.7b06365 doi: 10.1021/acs.est.7b06365

Padey, P., Girard, R., le Boulch, D., & Blanc, I. (2013, 2). From LCAs to Simplified Models: A Generic Methodology Applied to Wind Power Electricity. *Environmental Science & Technology*, *47*(3), 1231–1238. Retrieved from https://pubs.acs.org/doi/abs/10.1021/es303435e doi: 10.1021/es303435e

Peters, G. P. (2007, 9). Efficient algorithms for Life Cycle Assessment, Input-Output Analysis, and Monte-Carlo Analysis. *The International Journal of Life Cycle Assessment*, *12*(6), 373–380. Retrieved from http://link.springer.com/10.1065/lca2006.06.254 doi: 10.1065/lca2006.06.254

Pianosi, F., & Wagener, T. (2015, 5). A simple and efficient method for global sensitivity analysis based on cumulative distribution functions. *Environmental Modelling & Software*, *67*, 1–11. Retrieved from https://linkinghub.elsevier.com/retrieve/pii/S1364815215000237 doi: 10.1016/j.envsoft.2015.01.004

Ross, S., Evans, D., & Webber, M. (2002, 1). How LCA studies deal with uncertainty. *The International Journal of Life Cycle Assessment*, *7*(1), 47. Retrieved from http://link.springer.com/10.1007/BF02978909 doi: 10.1007/BF02978909

Saltelli, A., Aleksankina, K., Becker, W., Fennell, P., Ferretti, F., Holst, N., ... Wu, Q. (2019, 4). Why so many published sensitivity analyses are false: A systematic review of sensitivity analysis practices. *Environmental Modelling & Software*, *114*, 29–39. Retrieved from https://linkinghub.elsevier.com/retrieve/pii/S1364815218302822 doi: 10.1016/j.envsoft.2019.01.012

Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., ... Tarantola, S. (2008). *Global Sensitivity Analysis. The Primer*. Chichester, UK: John Wiley & Sons, Ltd.

Wei, W., Larrey-Lassalle, P., Faure, T., Dumoulin, N., Roux, P., & Mathias, J.-D. (2015, 1). 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. *Environmental Science & Technology*, *49*(1), 377–385. Retrieved from https://pubs.acs.org/doi/10.1021/es502128k doi: 10.1021/es502128k

Wei, W., Larrey-Lassalle, P., Faure, T., Dumoulin, N., Roux, P., & Mathias, J.-D. (2016, 3). Using the Reliability Theory for Assessing the Decision Confidence Probability for Comparative Life Cycle Assessments. *Environmental Science & Technology*, *50*(5), 2272–2280. Retrieved from https://pubs.acs.org/doi/10.1021/acs.est.5b03683 doi: 10.1021/acs.est.5b03683

Wikibooks. (2021). *Probability/Transformation of Probability Densities — Wikibooks, The Free Textbook Project.* Retrieved from https://en.wikibooks.org/w/index.php?title=Probability/Transformation_of_Probability_Densities&oldid=3826066

Yamazaki, F., Member, A., Shinozuka, M., & Dasgupta, G. (1988, 8). Neumann Expansion for Stochastic Finite Element Analysis. *Journal of Engineering Mechanics*, *114*(8). doi: 10.1061/(ASCE)0733-9399(1988)114:8(1335)

Appendix A: Linearity

A linear model can be expressed as the (scaled) sum of all input parameters, or in mathematical terms $y(x_1, x_2, ..., x_n) = x_0 + \sum_{i=1}^{n} a_i x_i$. Here y is the output and x_i are the inputs. Any change in the output is again a linear combination of changes in the inputs. The model sensitivity to coefficient x_i is directly related to a_i. These simple relationships are why it is often desirable to fit a model in question to a linear approximation, by linear regression or a *first-order* (Taylor) polynomials, or perturbation (E. A. Groen et al., 2017). Under the assumption of linearity however, interactions between parameters - terms enhancing or reducing each others influence - are ignored.

Illustrated in Figure 4 is a plot of the function $\boldsymbol{Y} = \boldsymbol{X}_1\boldsymbol{X}_2$, which is multiplicative, polynomial, and *not* linear. The plot shows that the model output is not a flat plane and its slope depends on the point and direction of evaluation. This color plot is used in Section 3.2.1 to explain some issues of assuming linearity in LCA SA.

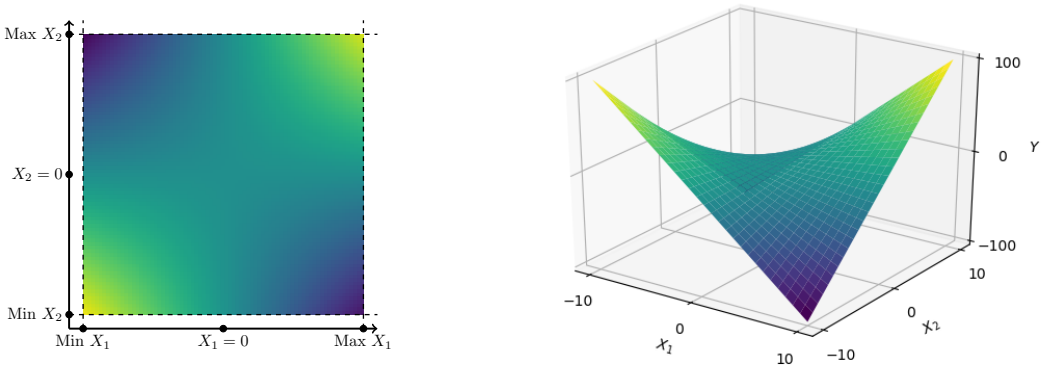


Figure 4: Flat color plot and corresponding 3d surface plot of the multiplicative function $\boldsymbol{Y} = \boldsymbol{X}_1\boldsymbol{X}_2$

Appendix B: Mathematical documentation

This appendix contains the omitted mathematical steps and explanations of Section 6. Normal multivariate $\underline{\boldsymbol{X}}$ and lognormal multivariate $\boldsymbol{X}$ have parameters and are related as follows

$$\begin{aligned}
\boldsymbol{X} &\sim \mathcal{LN}(\mu_{\boldsymbol{X}}, \Sigma_{\boldsymbol{XX}}) \\
\underline{\boldsymbol{X}} = \ln(\boldsymbol{X}) &\sim \mathcal{N}(\mu_{\underline{\boldsymbol{X}}}, \Sigma_{\underline{\boldsymbol{XX}}}) \\
\Sigma_{\underline{X}_i \underline{X}_j} &= \ln\left(\frac{\Sigma_{X_i X_j}}{\mu_{X_i}\mu_{X_j}} + 1\right) \\
\mu_{\underline{X}_i} &= \ln(\mu_{X_i}) - \frac{1}{2}\ln\left(\frac{\Sigma_{X_i X_i}}{\mu_{X_i}\mu_{X_i}} + 1\right)
\end{aligned}$$

A product of the coefficients in $\boldsymbol{X}$ is established as $\boldsymbol{Z}$

$$\begin{aligned}
\boldsymbol{Z} = \prod_i X_i &\sim \mathcal{LN}(\mu_Z, \sigma_Z^2) \\
\prod_i X_i &= \exp\left(\sum_i \underline{X}_i\right) \\
\boldsymbol{Z} &= \exp(\underline{\boldsymbol{Z}}) \\
\underline{\boldsymbol{Z}} = \sum_i \underline{X}_i &\sim \mathcal{N}(\mu_{\underline{Z}}, \sigma_{\underline{Z}}^2) \\
\mu_{\underline{Z}} &= \sum_i \mu_{\underline{X}} \\
\sigma_{\underline{Z}}^2 &= \sum_{ij} \Sigma_{\underline{X}_i \underline{X}_j}
\end{aligned}$$

Using the established relations, the mean of $\boldsymbol{Z}$ in terms of the input parameters is derived next

$$\begin{aligned}
\mu_Z &= \exp\left(\mu_{\underline{Z}} + \frac{1}{2}\sigma_{\underline{Z}}^2\right) \\
&= \exp\left(\sum_i \mu_{\underline{X}} + \frac{1}{2}\sum_{ij} \Sigma_{\underline{X}_i \underline{X}_j}\right) \\
&= \exp\left(\sum_i \left(\ln(\mu_{X_i}) - \frac{1}{2}\ln\left(\frac{\Sigma_{X_i X_i}}{\mu_{X_i}\mu_{X_i}} + 1\right)\right) + \frac{1}{2}\sum_{ij} \ln\left(\frac{\Sigma_{X_i X_j}}{\mu_{X_i}\mu_{X_j}} + 1\right)\right) \\
&= \prod_i \mu_{X_i} \prod_i \sqrt{\frac{\Sigma_{X_i X_i}}{\mu_{X_i}\mu_{X_i}} + 1}^{-1} \prod_{ij} \sqrt{\left(\frac{\Sigma_{X_i X_j}}{\mu_{X_i}\mu_{X_j}} + 1\right)} \\
&= \prod_i \mu_{X_i} \prod_{i \neq j} \sqrt{\left(\frac{\Sigma_{X_i X_j}}{\mu_{X_i}\mu_{X_j}} + 1\right)}
\end{aligned}$$

And the variance of $\boldsymbol{Z}$ in terms of the input parameters is

$$\begin{aligned}
\sigma_Z^2 &= \mu_Z^2(\exp\left(\sigma_{\underline{Z}}^2\right) - 1) \\
&= \mu_Z^2(\exp\left(\sum_{ij} \Sigma_{\underline{X}_i\underline{X}_j}\right) - 1) \\
&= \mu_Z^2\left(\exp\left(\sum_{ij} \ln\left(\frac{\Sigma_{X_iX_j}}{\mu_{X_i}\mu_{X_j}} + 1\right)\right) - 1\right) \\
&= \mu_Z^2 \prod_{ij}\left(\frac{\Sigma_{X_iX_j}}{\mu_{X_i}\mu_{X_j}} + 1\right) - \mu_Z^2
\end{aligned}$$

For brevity sake, the term Γ_Z is introduced, which makes the final equations for the parameters of $\boldsymbol{Z}$

$$\begin{aligned}
\Gamma_Z &= \prod_{i\neq j}\left(\frac{\Sigma_{X_iX_j}}{\mu_{X_i}\mu_{X_j}} + 1\right) \\
\mu_Z &= \prod_i \mu_{X_i}\sqrt{\Gamma_Z} \\
\sigma_Z^2 &= \prod_i (\Sigma_{X_iX_i} + \mu_{X_i}^2)\Gamma_Z^2 - \mu_Z^2
\end{aligned}$$

Note that Γ_Z is undefined when any involved μ is zero. This problem is mended by overriding the fractions containing a zero μ with $\frac{\Sigma_{X_iX_j}}{\mu_{X_i}\mu_{X_j}} = 0$.

An alternative notation is also possible. We define a new identity $\Theta_Z = \prod_{ij}\left(\frac{\Sigma_{X_iX_j}}{\mu_{X_i}\mu_{X_j}} + 1\right)$. Similarly to the approach of Γ_Z, undefined fractions are overridden.

$$\begin{aligned}
\mu_Z &= \prod_i \frac{\mu_{X_i}^2}{\sqrt{\Sigma_{X_iX_i} + \mu_{X_i}^2}}\sqrt{\Theta_Z} \\
\sigma_Z^2 &= \mu_Z^2(\Theta_Z - 1)
\end{aligned}$$

Appendix C: Verification code

The Python files can be found via the following link: **https://github.com/slensen/Thesis-SA-for-LCA-Python-code**. Included separately are the files

- `ThreeMatrixProduct.py` contains the code for a product of three matrices;
- `OneMatrixInverse.py` contains the code for the inverse of a matrix;
- `ComplexMultivariateNormal.py` contains functions used in both aforementioned files with which a complex, normally distributed multivariate can be sampled.