Sensitivity to weighting in life cycle impact assessment (LCIA)

Prado, V.; Cinelli, M.; Haar, S.F. ter; Ravikumar, D.; Heijungs, R.; Guinée, J.B.; Seager, T.P.

The International Journal of Life Cycle Assessment (2020) 25:2393–2406
https://doi.org/10.1007/s11367-019-01718-3

IMPLICATIONS OF LCA CHOICES ON INTERPRETATION OF RESULTS AND ON DECISION SUPPORT

Sensitivity to weighting in life cycle impact assessment (LCIA)

Valentina Prado[1,2] · Marco Cinelli[4] · Sterre F. Ter Haar[5] · Dwarakanath Ravikumar[6] · Reinout Heijungs[5,7] · Jeroen Guinée[5] · Thomas P. Seager[3]

Received: 14 November 2018 / Accepted: 6 December 2019 / Published online: 17 December 2019

Abstract

Purpose Weighting in life cycle assessment (LCA) incorporates stakeholder preferences in the decision-making process of comparative LCAs. Research efforts on this topic are concerned with deriving weights according to different principles, but few studies have evaluated the relationship between normalization and weights and their effect on single scores. We evaluate the sensitivity of aggregation methods to weights in different life cycle impact assessment (LCIA) methods to provide insight on the receptiveness of single score results to value systems.

Methods Sensitivity to weights in two LCIA methods is assessed by exploring weight spaces stochastically and evaluating the rank of alternatives via the Rank Acceptability Index (RAI). We assess two aggregation methods: a weighted sum based on externally normalized scores and a method of internal normalization based on outranking across CML-IA and ReCipE midpoint impact assessment. The RAI represents the likelihood that an alternative occupies a certain rank given all possible weight spaces, and it can be used to compare the sensitivity of final ranks to weight values in each aggregation method and LCIA. Evaluation is based on a case study of a comparative LCA of five PV technologies whose inventory is readily available in Ecoinvent.

Results and discussion Influence of weights in single scores depend on the scaling/normalization step more than the value of the weight itself. In each LCIA, aggregated results from a weighted sum with external normalization references show a higher weight insensitivity in RAI than outranking-based aggregation because in the former, results are driven by a few dominant impact categories due to the normalization procedure. Differences in sensitivity are caused by the notable variety (up two orders of magnitude) in the scales of normalized values for the weighted sum with external normalization and intrinsic properties of the methods including compensation and a lack of accounting for mutual differences.

Conclusions Contrary to the belief that the choice of weights is decisive in aggregation of LCIA results, in this case study, it is shown that the normalization step has the greatest influence in the results. This point holds for EU and World references in ReCiPe and CML-IA alike. Aggregation consisting of outranking generates rank orderings with a more balanced contribution of impact categories and sensitivity to weights' values as opposed to weighted sum approaches that rely on external normalization references.

Responsible editor: Alexis Laurent

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s11367-019-01718-3) contains supplementary material, which is available to authorized users.

✉ Valentina Prado
v.prado@uniandes.edu.co

[1] School of Management, Los Andes University, Bogotá, Colombia

[2] EarthShift Global LLC, 37 Route 236, Suite 112, Kittery, ME 03904, USA

[3] Sustainable Engineering and the Built Environment, Arizona State University, Tempe, AZ, USA

[4] Institute of Computing Science, Poznań University of Technology, Piotrowo 2, 60-965 Poznań, Poland

[5] Department of Industrial Ecology, Institute of Environmental Sciences (CML), Leiden University, Einsteinweg 2, 2333 Leiden, CC, The Netherlands

[6] School for Environment and Sustainability, University of Michigan, Ann Arbor, MI, USA

[7] Department of Econometrics and Operations Research, Vrije Universiteit Amsterdam, De Boelelaan 1105, 1081 HV Amsterdam, The Netherlands

Recommendations Practitioners aiming to include stakeholder values in single scores for LCIA should be aware of how the weights are treated in the aggregation method as to ensure proper representation of values.

Keywords Aggregation in LCA · LCIA (Life Cycle Impact Assessment) · PV technologies · Weighting in LCA

1 Introduction

1.1 Weighting in LCA

In principle, weighting impact categories clarifies tradeoffs among competing alternatives by enabling aggregation of incommensurable environmental performances into a single score, thereby facilitating interpretation of results. Beyond this, the value of weighting lies in allowing for the distinct preferences of different decision-makers to be reflected or contrasted in the results, rather than a single point of view serving as a final verdict (Bengtsson and Steen 2000). Weighting and normalization are optional procedures of life cycle impact assessment (LCIA) (ISO 2006) that can aid in decision support for comparative life cycle assessments (LCAs) by aggregating characterized results to a single score and facilitate identification of a more favorable alternative. Due to this application, weighting can aid in the interpretation of comparative LCA results without technically being classified as an interpretation step. This study compares current aggregation of LCIA results with an alternative method of aggregation from Multi-Criteria Decision Analysis (MCDA) on the basis of weight sensitivity. This optional procedure of LCIA, aided by analytical aggregation operators, is independent from the inventory analysis and characterization stages. This study lies between LCA and MCDA, and to clarify terminology, a glossary is included in Table 1.

While weighting and aggregation in LCA has been a controversial topic due to its subjective nature, it is still recognized as an important and useful step in the communication of LCA results (Zanghelini et al. 2018). The issue of subjectivity in LCA modeling has been discussed in the literature where it is generally accepted that there are subjective choices in every stage of an LCA even before weighting takes place (e.g., Hertwich et al. 2000). Nevertheless, there is a distinction in LCA phases that are generally considered to be science- and value-based. This distinction is apparent in the ISO guidelines, by making weighting an optional step and preventing aggregated scores to be disclosed to the public in comparative assertions (ISO 2006). These guidelines aim to avoid imposing values from an organization or company that may not necessarily match those of the public and that may be biased towards favoring a certain option. In these situations, characterized and/or normalized profiles are disclosed instead.

In some applications, weighting and aggregation is needed to help resolve tradeoffs among competing alternatives and provide decision support (Laurin et al. 2016). Specially, environmental insight is becoming more important in decision support, and a single indicator (such as climate change) does not fully address all the environmental concerns. There are several approaches to weighting in the literature (Ahlroth et al. 2011; Huppes et al. 2012; Pizzol et al. 2015; Castellani et al. 2016a) that can be categorized as panel, valuation, or target based. Most advances in these approaches deal with the calculation of weight values to improve accuracy in assessments or to provide a "better" principle for deriving weights. For instance, Itsubo et al. (2015) expand weights based on public surveys for LIME2 in Japan to other G20 countries to gain broader applicability and understanding of options according to different regions of the world. Alternatively, in

Table 1 Glossary

Term	Definition
Weighting	Weighting consists of defining the relative priorities of the different impact categories (criteria/indicators in MCDA literature) to identify those of major concern for the interest group. It does not refer to the actual action of aggregating to generate a single score/ranking.
Weight	Numerical factor resulting from the weighting. It can be a coefficient of importance representing the relative importance of indicators, or a trade-off representing the exchange rate between indicators (Keeney 2002). Importance coefficient weights are referred to as weighting factors in ISO 14044 and defined as "numerical factor based on value choices."
Aggregation method	Method for aggregation of different criteria/indicators (impact categories in LCA) to solve a ranking, sorting, or choice decision-making challenge. Scaling and weighting are two common components of an MCDA and also components of the two aggregation approaches studied here. The aggregation methods used in this study include an additive weighted sum and an outranking algorithm called PROMETHEE.
Scaling	Transformation of the values of the criteria/indicators (impact categories in LCA) to make them comparable and hence in a form suitable for the aggregation for the target method. Scaling of values in LCA is known as normalization and it typically consists of division by a normalization reference. In our case study, normalization for the weighted sum is performed with external references while with PROMETHEE it is implemented internally with pairwise comparisons (outranking).

Table 2 Main characteristics of aggregation methods included in this study

Characteristic	Weighted sum with external normalization (EU and World in ReCiPe and CML-IA)	Outranking
Aggregation	Aggregation based on additive weighted sum consisting on a scaling step dependent on external normalization and application of importance weights. It is context *independent*, where each alternative is evaluated separately (Cinelli et al. 2014). In a context independent evaluation, the score is independent from the alternatives in a set (this score is technically a rating) (Prado et al. 2012).	Aggregation based on pair wise comparisons as a function of the mutual differences and application of importance weights. It is context *dependent* and the evaluation of alternatives depends on the other alternatives in the set (Cinelli et al. 2014). Therefore, the analysis can be subject to rank reversal in the event that alternatives and/or criteria are removed or added (Figueira and Roy 2009; Cinelli et al. 2014). The consideration of rank reversal in outranking is discussed more extensively in earlier works (Prado et al. 2012).
Linearity and compensation	Linear aggregation that allows full compensation between impact categories. *Full* compensation has been linked to a weak sustainability perspective when it comes to environmental management (Cinelli et al. 2014, Pollesch and Dale 2015, 2016). Compensation relates to the property of an aggregation method to "compensate" a poor performance by a good performance. In fully compensatory methods, it is possible for a single good performance to compensate for multiple poor performances. This property is adequate when dealing with interchangeable aspects (like economic profit and loss), but when it comes to environmental management, full compensation can lead to burden shifting.	Applies a non-linear aggregation algorithm with *partial* compensation between impact categories. Classified as a semi-strong sustainability perspective (Rowley et al. 2012; Matarazzo et al. 2013).
Scaling	Scaling by external normalization can lead to scaled values differing by orders of magnitude (a difference factor of 100 is enough to mask most weights) prior to application of weights (White and Carty 2010; Castellani et al. 2016b; Prado et al. 2017).	Due to the nonlinear properties, scaling via outranking generates values within the same order of magnitude (between 0 and 1) prior to application of weights (Prado-Lopez et al. 2014). Scaling is done in a pair-wise manner where the mutual differences are evaluated against the uncertainty in the data to determine whether one alternative is superior, inferior, or indifferent to the other.
Meaning of weights	Weights should be interpreted as tradeoffs in a weighted sum (Riabacke et al. 2012). Application of relative importance weights as done here (representative of current practice) exemplifies a mathematical inconsistency known as the "typical weighting error" (Edwards and Barron 1994; Steele et al. 2009).	Outranking is compatible with importance weights because it takes into account the absolute relevance of the criteria, irrespective of their measurement scale (Riabacke et al. 2012).

search of a more holistic coverage of public opinion, Ji and Hong (2016) apply internet search volumes (such as Google Trends) to determine weight values. Other recent examples include the determination of weights given EU2020 targets (Castellani et al. 2016a), planetary operating spaces (Tuomisto et al. 2012), and MCDA panel elicitation techniques (Myllyviita et al. 2014). These developments however pertain to importance coefficients, which are incompatible with the typical aggregation approach of a weighted sum that requires trade-off weights (Keeney 2002; Dias et al. 2016). Application of importance coefficient type of weights with weighted sums as typically done in LCA can be problematic as it does not account for the effect of the range in the perceived importance of an aspect, known as the "range sensitivity principle" (Fischer 1995). In fact, this mismatch of the scaling and the weight type is known as the "typical weighting error" (Edwards and Barron 1994).

The discussion of which weight principle is the most appropriate for LCA is not the purpose of this study; rather, we call attention to the relationship that exists between *scaling* (typically consisting of external normalization in LCA) and the weight. Some recent studies touch upon this issue. Myllyviita et al. (2014) explores different MCDA weighting elicitation techniques and finds that external normalization is more influential than weights, and Wulf et al. (2017) evaluates three aggregation methods and various weight values but finds that the normalization method has the largest influence on results. Kalbar et al. (2017) and Sohn et al. (2017) also identify the issue of weight insensitivity in the current weighted sum but do not consider the issue of linearity and compensation when applying TOPSIS, a fully compensatory aggregation method, to generate a complete ranking (Seager and Prado 2017). Compensation is a property of the aggregation method, and it is a key aspect evaluated in this study (Table 2). The Kalbar et al. (2017) goes as far as applying a non-compensatory method, namely, the Hasse Diagram Technique, to generate a partial rank ordering and identify alternatives to eliminate prior to the complete ranking with TOPSIS. In the end, the ranking of alternatives in Kalbar et al. (2017) is based on a fully compensatory method. The present study differs from previous evaluations of aggregation methods in that it generates a complete rank ordering with a non-compensatory method; it includes uncertainty in characterized results and weights and communicates results via a probabilistic ranking.

Besides this, weighting developments do not explicitly address the relationship between weights and scaling and how the values relate to the quantified environmental performances (characterized results). This connection, described previously as a *relational claim* by Hertwich et al. (2000), is critical to LCIA as it dictates how the facts relate to our concerns, including relevancy and mathematical consistency. Here, the characterized results can be described as the "facts" (modeled as much as possible evidence-based) and the weights as the values (what we care about). Connecting the two is the aggregation method, the *relational claim*. Alternative aggregation approaches to the weighted sum appear in Prado-Lopez et al. (2014), which illustrates a case of outranking, a methodology developed in the field of MCDA, which can be used to aggregate impact category results with quantified uncertainty. To explore the sensitivity to weight values as a function of the scaling step, we apply stochastic weight values to two aggregation methods: (1) weighted sum with external normalization by a reference (as currently implemented in LCIA corresponding to a World and EU reference) and (2) outranking as in Stochastic multi-attribute analysis (SMAA in Prado and Heijungs 2018). Single scores resulting from each aggregation method can be used to rank alternatives, and the frequency of such ranks is evaluated via the Rank Acceptability Index (RAI), which quantifies the likelihood of each alternative to occupy a certain rank (Tervonen and Lahdelma 2007). Variation in ranks indicates sensitivity to weight values. Here, the scaling step becomes the *independent* variable and the ranking is the *dependent* variable. Table 2 provides a summary of the characteristics of the two aggregation methods used to derive the single scores, according to the underlying aggregation approach, linearity, scaling of criteria, and meaning of the weights, all key distinctive features of MCDA methods (Cinelli et al. 2014).

This paper is organized as follows. Section 2 presents the methodological approach, which includes introduction to the comparative LCA case study with uncertainty analysis (2.1), description of the weighted sum using external normalization references (2.2), description of the outranking method (2.3), description of the stochastic exploration of weights to be implemented in both aggregation methods (2.4), and a description of the RAI used to evaluate the results (2.5). Section 3 shows the results of the RAIs for the two aggregation methods in ReCiPe and CML-IA which corresponds to three probabilistic rankings per impact assessment (EU and World references and outranking). Results include the contribution of individual impact categories to the overall scores. Section 4 discusses the meaning and implication of results and provides closing remarks as well as recommendations for the analysts interested in developing single scores in LCA.

2 Methods

2.1 Comparative LCA illustration and systematic evaluation

This study utilizes life cycle inventory of five different PV technologies for the production of 1 MJ of electricity as compiled by Jungbluth et al. (2012) and implemented in Ecoinvent 3 with Simapro PhD version. Impact assessment is conducted

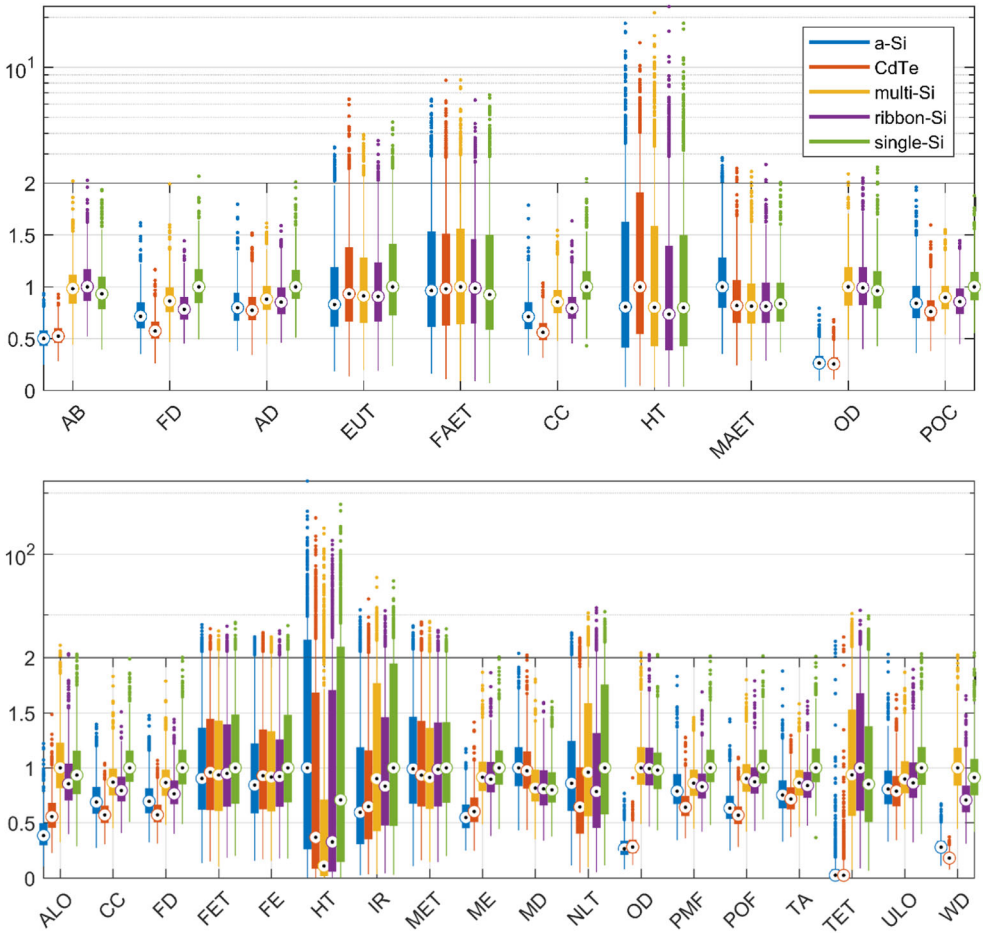


Fig. 1 Characterized results with uncertainty of PV alternatives according to CML-IA (top) and ReCiPe (bottom). The x-axis corresponds to the impact categories and the y-axis represents the relative characterized performances scaled to the largest median in each impact category. The axis is linear from 0 to 2 and logarithmic for values above 2. The bottom (q_1) and top (q_3) edges of the box represent the 25th and 75th percentiles, respectively. The whiskers extend to the most extreme data points not considered outliers. Outliers are points greater than $q_3 + 1.5 \times (q_3 - q_1)$ or less than $q_1 - 1.5 \times (q_3 - q_1)$. Underlying data can be found in the SI

via two common LCIA methods: ReCiPe (H) midpoint (Goedkoop et al. 2009) and CML-IA 2001 baseline (Guinée et al. 2002). Alternatives pertain to a 3-kWp slanted-roof installation and they include single-crystalline silicon cells (single-Si), multi crystalline silicon cells (multi-Si), thin film cadmium telluride (CdTe), amorphous cells (a-Si), and ribbon silicon (ribbon-Si). Comparative results include inventory uncertainty based on the pedigree matrix and 1000 Monte Carlo runs performed in Simapro PhD version (Muller et al. 2014). Given the purpose is to compare the results of the aggregation methods, stability of the stochastic results which would call for a higher number of Monte Carlo runs, is not of concern. Characterized results for ReCiPe and CML-IA are shown with a box plot in Fig. 1. For illustration purposes, we include ten impact categories from CML-IA, namely, abiotic depletion (elements) (AB), abiotic depletion (fossil fuels) (FD), acidification (AD), eutrophication (EUT), fresh water aquatic ecotoxicity (FAET), global warming (CC), human toxicity (HT), marine aquatic ecotoxicity (MAET),[1] ozone layer depletion (OD), and photochemical oxidation (POC). Terrestrial exotoxicity was excluded because the uncertainty analysis in Simapro generated negative mean values—consequence of an error in the uncertainty propagation of the software as lognormal distributions do not have negative means. From ReCiPe, we include all 17 impact categories, namely, agricultural land occupation (ALO), climate change (CC), fossil depletion (FD), freshwater ecotoxicity (FET), freshwater eutrophication (FE), human toxicity (HT), ionizing radiation (IR), marine ecotoxicity (MET), marine eutrophication (ME), metal depletion (MD), natural land transformation (NLT), ozone depletion (OD), particulate matter formation (PMF), photochemical oxidant formation (POF), terrestrial acidification (TA), terrestrial ecotoxicity (TET), and urban land occupation (ULO). Water depletion (WD) from ReCiPe while it is shown in Fig. 1 is excluded from further aggregation because it does not have a normalization reference and cannot be included in the weighted sum aggregation method. To keep the impact categories consistent between aggregation methods, we exclude WD from both aggregation methods. It must be stressed that

[1] Due to constraints related to modeling metals on longer time horizons in multi-media models, MAET is often excluded from baseline categories (Heijungs et al. 2004)

the goal of this study is to evaluate how two aggregation methods (described in Table 2) are affected by weights. The PV example is only for illustrative purposes with the aim of experimenting with representative comparative LCA results. Evaluation of the effect of weights is performed separately within each impact assessment method, so variations in coverage of impacts between CML-IA and ReCiPe do not affect findings.

2.2 Weighted sum with external normalization

A weighted sum with external normalization represents the most common form of aggregation in LCA. This method consists of scaling via external normalization and application of importance weights (Eq. 1). It is important to highlight however that external normalization has different purposes in LCA besides scaling. The LCA Handbook describes the other uses of external normalization as error checking and hotspot or relevance analysis (Guinée et al. 2002). This paper focuses on application of normalization references for scaling purposes:

$$\sum \frac{CI_i}{NR_i} \times w_i = \text{Single score} \quad (1)$$

where

CI_i is the category indicator result; NR_i is the normalization reference (in the units of the CI_i per year), which can represent a global or a regional community; and w_i is the weight as it pertains to impact category *i*. The single score is referred to as "points," but it has units of "year." Since this study includes uncertainty in the characterized results (as shown in Fig. 1) and weights, each value of *CI* and *w* is based on deterministic sampling within the boundaries of the impact category with a lognormal distribution and weights with a beta distribution. The parameters of the corresponding lognormal distributions for CI are included in the SI along with the weight values for each run. Normalization references are deterministic, but a probabilistic treatment is possible.

Reports on how external normalization references can lead to biases in aggregated results already exist in the literature (Heijungs et al. 2007; White and Carty 2010; Myllyviita et al. 2014; Castellani et al. 2016b) and solutions fall for the most part in the "repair" category (Kim et al. 2013) where recent recommendations call for the use of global as opposed to regional references to overcome biases (Verones et al. 2017). Others argue that the issues of external normalization go beyond data repair efforts or choice of geographical reference because the problem lies in the linear aggregation approach and the neglect of mutual differences in the assessment (Prado et al. 2017; Cucurachi et al. 2017; Ravikumar et al. 2018). A recurrent finding is that the use of external normalization can have a dominant effect in the interpretation of results (White and Carty 2010; Myllyviita et al. 2014; Castellani et al. 2016b; Wulf et al. 2017). In this study, we explore how external normalization as a scaling step affects the representation of weights in the results.

It must be noted that the weighted sum as shown in Eq. 1 can also be structured with a variety of normalization by division schemes where instead of referring to a particular community, it can represent one of the alternatives of the set in which case it is deemed internal normalization (Norris 2001). For instance, on each impact category, the best alternative can attain a normalization score of 1 and the worst alternative, a normalized score of 0 or vice versa (Nzila et al. 2012; Du et al. 2019). Other strategies for normalization include a "status quo" using as a reference the average scenario (Domingues et al. 2015) or the most common scenario (Dias et al. 2016). Furthermore, rank, percentile rank, categorical, distance to target, and logistic approaches are further options for normalization (Nardo et al. 2008; Cao et al. 2016). This "normalization by division," however, remains fully linear and compensatory in nature. This study focuses on a weighted sum using external normalization references provided by already established LCIA methods.

2.3 Outranking

Outranking algorithms originate from the MCDA literature, and they aggregate multiple criteria based on pairwise comparisons (Roy 1985; Behzadian et al. 2010; Greco et al. 2016). Here, we apply a version of outranking known as PROMETHEE II (Preference Ranking Organization METHod for Enrichment of Evaluations), which results in a full rank ordering of alternatives. PROMETHEE II consists of a scaling step via outranking and application of importance weights in a mathematically meaningful manner (Cinelli et al. 2014; Munda 2016). PROMETHEE II algorithm used in this study allows accounting for uncertainty in two ways: (1) via the inclusion of uncertainty in the characterized performances by sampling the values propagated when calculating mutual differences and (2) by assigning a preference and indifference thresholds (P and Q respectively in Fig. 2) as a function of the average standard deviation of performances within each impact category (specifically, in this work, P was set to be the average standard deviation for each impact category, and Q was set to be half of P). This implies that in impact categories with larger standard deviations, alternatives need to outperform each other by a greater margin to achieve the largest outranking score of 1. While the preference thresholds are static, they will change given any refinement in the data so that the results are subject to change even if median results do not. Note that while these thresholds are called "preference thresholds" in this application, these are based on the data

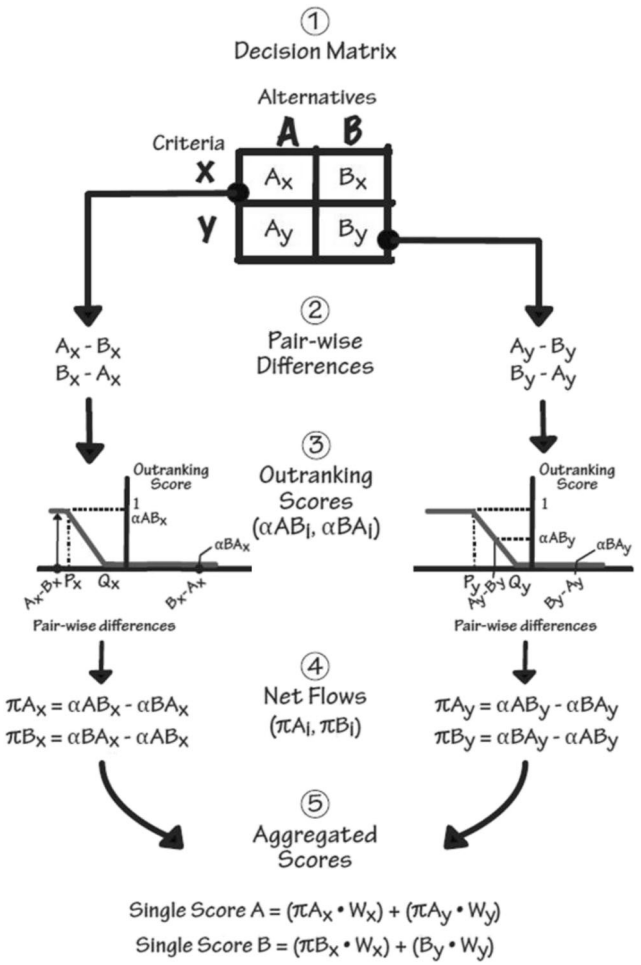


Fig. 2 PROMETHEE outranking procedure illustrated with a decision matrix consisting of two alternatives (A and B) and two criteria (x and y). For each criterion (x on the left and y on the right), there is a process of pair wise comparisons. The pair wise differences in the respective units of the criteria (in this case, impact categories) are harmonized to unitless outranking scores, α, ranging from 0 to 1 as shown in step 3. As shown in the illustration, when dealing with environmental impact, a lower value is preferred. Then, the outranking scores for both directions are converted into "net flows," π (step 4). Finally, in step 5, weight factors ($w_x$and w_y) are applied to the net flows to generate a single score

and not expert elicitation as done in other contexts (Rogers and Bruen 1998).

PROMETHEE II also limits compensation which is useful in environmental manage as it can avoid few extreme performances to dominate the assessment. Outranking consists of evaluating mutual differences with respect to a defined value function where it generates unweighted scores (referred to as outranking score) depending on how each alternative performs against each other in every single criterion (in this case impact category). Pairwise evaluation is executed in both directions, leading to two measures called positive and negative flows for each alternative. The former indicates the strength of the target alternative against the other one, while the latter is the weakness of the target alternative in comparison with the other one. The combination of the two flows leads to the net flow. We provide a brief explanation and visualization of the outranking procedure with the use of a generic example in Fig. 2, leaving the detailed step-by-step explanation of calculation procedures of PROMETHEE II as applied to LCA to Prado and Heijungs (2018).

2.4 Stochastic exploration of weights

Sampling of weights is implemented stochastically so that the weight values are equally distributed among all the impact categories, the sum of the weights equal 100, and each individual weight ranges between 0 and 100 (Tervonen and Lahdelma 2007). Stochastic weights can also be applied including specific constraints from decision makers (Rogers and Seager 2009; Du et al. 2019), but in this study, we apply the same weight distribution in impact categories. We follow the weights calculation according to the pseudo Markov Chain by Tylock et al. (2012) where each weight value distribution follows a beta distribution as a function of the number of criteria. We apply a beta distribution as a function of the number of impact categories (n) where $\alpha = 1$ and $\beta = n - 1$.These distributions are calculated consecutively where the result of the first one affects the second one (hence, a pseudo Markov Chain procedure) so that when $n = 1$, meaning the last impact category, is calculated as the remaining possible value so $\sum w = 100$. Resulting weight value distributions for ReCiPe and CML-IA are equally distributed (Fig. 3). Note that in the case of ReCiPe, WD was excluded because it cannot be incorporated in the weighted sum with external normalization; therefore, the aggregated score for both aggregation methods will be a function of the remaining 17 impact categories.

2.5 Rank acceptability index

When aggregating stochastic weights and performances, each combination generates a single score per alternative that can be used to generate a rank for each of the 1000 Monte Carlo runs. Rank frequency is then evaluated via the Rank Acceptability Index (RAI) (Tervonen and Lahdelma 2007). The use of RAI is widely accepted and confirmed as a useful measure to assess ranking robustness and stability (Tervonen et al. 2011; Corrente et al. 2014; Greco et al. 2018a, b). It allows to easily visualize the trend of the stochastic results by accounting for a multitude of stakeholders' perspectives, which has the added value of making the decision makers more comfortable when exerting their decisions, as the variability in the results is visible (Bertola et al. 2019).

It should be pointed out that Kalbar et al. (2017) discuss the issue of rankings comparisons by illustrating the overall extent of difference between the rankings is assessed, while in this paper, ranking variability is analyzed at the distribution level. More specifically, the RAI is used to study the variability of

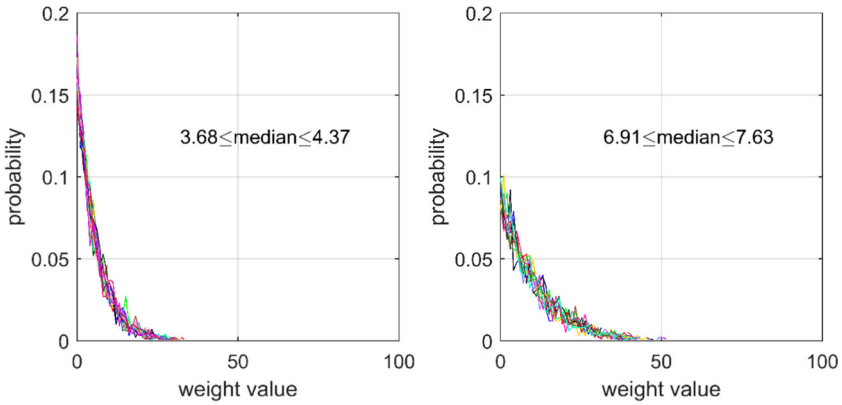


Fig. 3 Weight distributions consisting of 1000 Monte Carlo runs for the 17 impact categories in ReCiPe (left) and 10 impact categories in CML-IA (right)

the rankings based on stochastic simulation of the input, which is not part of the work in the mentioned study.

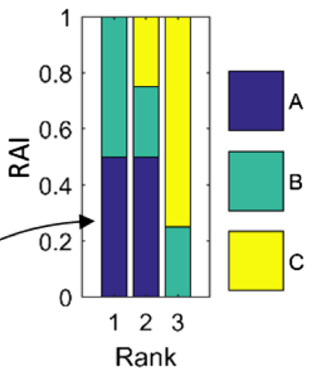
The RAI, R_x^r, is calculated per alternative (x) and rank position (r), and it evaluates the likelihood of an alternative to occupy a certain rank. For each run, the scores (u_x) of each alternative is ranked against each other. The frequency of a given rank throughout the runs for a certain alternative is divided by the number of runs to give the RAI for that rank. Figure 4 provides an illustration of an example consisting of three alternatives and four runs. The best alternatives have a higher RAI in the first rank and the poor alternatives tend to dominate in the lower ranks (such as alternative C in Fig. 4). The RAI allows identifying good and poor alternatives, and the level of competitiveness between them (for example, alternatives A and B appear to be competitive in the first rank in Fig. 4). We compute the RAI numerically over the final score in each aggregation method as illustrated in Fig. 4. Comparing the RAI associated to each aggregation method identifies those approaches that are most and least sensitive to weight ranges. Rank orderings with larger RAIs (dominant in a position) are more weight insensitive because results remain the same given most weight values.

3 Results

3.1 Weigh sensitivity in CML-IA

Figure 5 shows the RAIs for the PV comparative LCA using CML-IA baseline characterization and two aggregation methods: a weighed sum with World 2000 and EU 25 external normalization and outranking. From all three results, World 2000 shows the greatest weight insensitivity as alternatives have the highest probability of remaining in a single rank. The ranking, in the order from first to fifth, shows ribbon-Si (79.4%), single-Si (81.4%), multi-Si (87.1%), CdTe (87.1%), and a-Si (98%), respectively. This means that for much of the weight space, the rank ordering of alternatives remains the same. The RAI results for EU 25 show slightly more variation of ranks, but still, the position of alternatives stays the same in most of the sampled weight space: probabilities of ranking first to fifth shows ribbon-Si (62.1%), single-Si (70.1%), multi-Si (63.7%), CdTe (63.9%), and a-Si (93.6), respectively. Contrary to the weighted sum with World 2000 and EU 25, outranking shows a greater distribution of the ranks where the largest RAI does not exceed 65.2% (for R^1_{CdTe}). Rank orderings generated with outranking show alternatives CdTe (65.2%), a-Si (43.5%), ribbon-Si (36.2%), multi-Si (35%),

run	Score			Rank		
	u_A	u_B	u_C	r_A	r_B	r_C
1	6	5	8	2	1	3
2	2	3	4	1	2	3
3	4	7	6	1	3	2
4	5	3	9	2	1	3



$$R_A^1 = \frac{count\ if\ r_A = 1}{number\ of\ runs} = \frac{2}{4} = 0.5$$

Fig. 4 Illustration of numerical computation of the RAI. In a per run basis, the lowest score (u_x), obtains the first rank. Here, alternative A obtains the first rank twice (runs 2 and 3). Alternative B also ranks first twice (runs 1 and 4) and therefore obtains the same RAI as alternative A—illustrated by the figure to the right

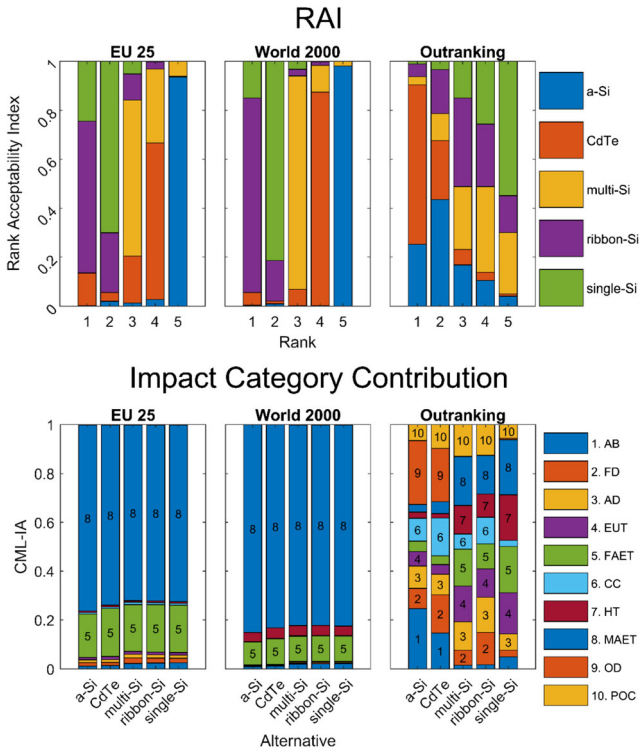


Fig. 5 Top: RAIs of the Comparative LCA of PV using the different aggregation methods in CML-IA baseline. From left to right: Weighted sum with EU 25 external normalization, weighted sum with World 2000 external normalization and outranking. The x-axis represents the rank ordering and the y-axis represents the rank acceptability index (RAI). Bottom: Impact category contributions per alternative and per aggregation method. The numbers denote the impact category (as shown in the legend to the right of the graph). The contributions correspond to the share of the overall score that is attributed to each impact category. For weighted sum approaches, the contribution is based on each impact category weighted performance relative to the total score and averaged over the 1000 MC simulations. For outranking, this contribution was calculated by the share, according to the average over the MC simulations, to the single score from each individual weighted net flow. Weighted net flows are shown in step 5 of Fig. 2. Values can be found in the Supplementary Information

and single-Si (54.9%) ranking first to fifth, respectively. Overall, RAIs in outranking are smaller than a weighted sum with World 2000 and EU 25, indicating that the rank ordering of alternatives generated by outranking is more sensitive to weight values than any of the weighted sum methods as applied in CML-IA.

When comparing the rank ordering of alternatives, we see that the weighted sum methods (EU 25 and World 2000) generate the same ranking. Ribbon-Si is ranked as the most preferred alternative followed by single-Si, multi-Si, CdTe, and a-Si. The ordering in outranking is quite different since CdTe is ranked first followed by a-Si; ribbon-Si and multi-Si compete for the third and fourth place, and single-Si ranks last. It is notable how a-Si moves from ranking last in EU 25 and World 2000 to a competitive second place in outranking. The reason for the difference in rank orderings between the weighted sum methods and outranking is the dominance of one impact category, marine aquatic ecotoxicity (MAET) in EU 25 and World 2000 (Fig. 5). Here, MAET contributes the majority of the total score for the externally normalized values (EU 25 and World 2000), followed by fresh aquatic ecotoxicity (FAET).

Alternatively, outranking score shows contribution of multiple impact categories where there is not a single or a few that override other performances (Fig. 5). Comparing lowest-ranked A-si to highest-ranked alternative Ribbon-si in both the EU25 and World 2000, A-si outperforms Ribbon-si in seven impact categories, is nearly equivalent twice, and is environmentally more burdensome only once for the MAET category (Fig. 1). Therefore, the ranking in EU 25 and World 2000 is determined by the performance in this single impact

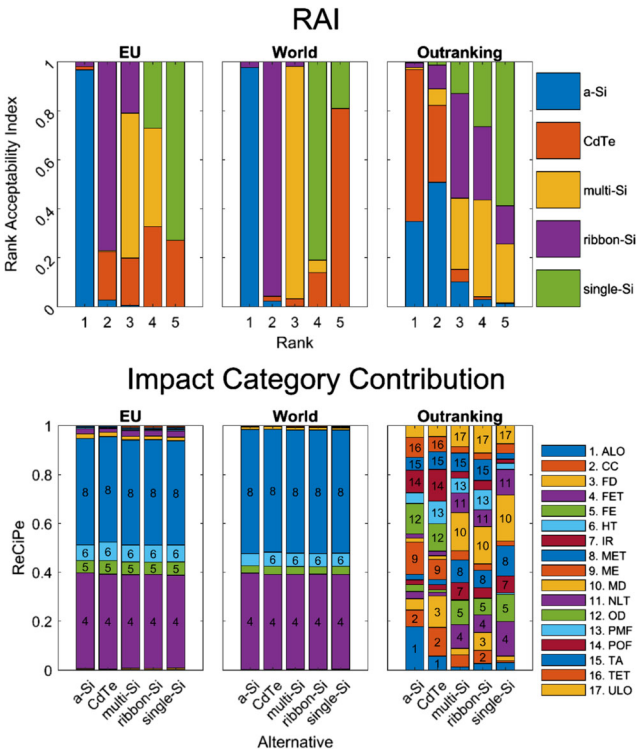


Fig. 6 Rank acceptability indices of the Comparative LCA of PV using two aggregation methods inReCiPe H midpoint. From left to right: weighted sum with Europe (EU) external normalization, weighted sum with World external normalization and outranking. The x-axis represents the rank ordering and the y-axis represents the RAI. Individual alternatives are denoted by color as shown in the legend to the right. Bottom: Impact category contributions per alternative and per aggregation approach. The numbers denote the impact category (as shown in the legend to the right of the graph). The contributions correspond to the share of the overall score that is attributed to each impact category. For weighted sum approaches, the contribution is based on each impact category weighted performance relative to the total score and averaged over the 1000 MC simulations. For outranking, this contribution was calculated by the share, according to the average over the MC simulations, to the single score from each individual weighted net flow. Weighted net flows are shown in step 5 of Fig. 2. Values can be found in the Supplementary Information

category. A favorable performance in MAET will compensate poor performances as is the case of Ribbon-si. Outranking however, where partial compensation is allowed, allows A-si to achieve better ranks under certain weighting conditions. Overall, the contribution plots in Fig. 5 show a different pattern between the composition of scores of a weighted sum and outranking.

3.2 Weight sensitivity in ReCiPe

Similar to Fig. 5, Fig. 6 shows the RAIs for the PV comparative LCA using ReCiPe H midpoint characterization. For ReCiPe, a weighted sum with World normalization leads to the greatest weight insensitivity because rank shares tend to be larger than in a weighted sum with Europe normalization and outranking. ReciPe World shows alternatives a-Si (97.6%), ribbon-Si (95.7%), multi-Si (94.8%), single-Si (81%), and CdTe (81%) rank first to fifth, respectively. Rank orderings with EU normalization reference result in a-Si (96.7%), ribbon-Si (77.1%), multi-Si (59.3%), CdTe (32.7%), and single-Si (72.9%) ranking first to fifth, respectively. In outranking, rank orderings do not generate a RAI higher than 66% (for $R^5_{Single-si}$). Here, CdTe (62.0%), a-Si (50.8%), ribbon-Si (42.7%), multi-Si (39.5%), and single-Si (58.8%) rank first to fifth, respectively. Overall, there is a greater distribution of ranks among alternatives in outranking than in both weighted sum approaches (EU and World), indicating greater weight sensitivity like in the CML-IA case.

Rank orderings between weighted sum with external normalization (EU and World) and outranking align more in ReCipe than in CML-IA. For example, all three aggregation approaches in ReCiPe place single-Si in the lowest ranks and a-Si in the higher ranks. The greatest discrepancy in ranks occurs with CdTe, which according to outranking it is a much better alternative than in the weighted sum approaches (EU and World). The difference in rank can be explained by the corresponding contributions per impact category (Fig. 6). In the weighted sum approaches in ReCiPe, EU, and World, there are two dominating categories: freshwater ecotoxicity (FET) and marine ecotoxicity (MET). In outranking, individual contributions of impact categories are much more balanced similar to the results in CML-IA (Fig. 5). The ranks in ReCipe from a weighted sum with EU and World normalization, where a-Si and ribbon-Si are the most likely first and second alternatives, coincide with their corresponding impacts in FET (Fig. 1) where these are the alternatives with the lowest two mean impacts in this impact category. In essence, the dominance of these two impact categories dictates the final rank.

4 Discussion

When evaluating the weight sensitivity of aggregation approaches in CML-IA and ReCiPe, weighted sum approaches (with EU and World references) led to higher weight insensitivity than outranking. In both impact assessment methods, the global reference (World 2000 in CML-IA and World in ReCiPe) had the greatest weight insensitivity because of the higher RAI values, followed by the EU normalization references. That means that despite sampling all possible weight values (within the number of Monte Carlo runs and considering natural numbers), the ranking of alternatives remained the same.

Weight insensitivity in a weighted sum using external normalization is due to the large differences in magnitude at the point of scaling—a difference in magnitude that is much larger than typical differences found in weight values. While scaled results in external normalization can deviate for up to two orders of magnitude, scaled results via outranking range between 0 and 1. Weight values of impact categories are typically within the same order of magnitude (0 to 100 with most weight values around 5—Fig. 3) so that the influence of weights is limited when the measurement scale of impact categories differ by more than the orders of magnitude of the weights. Such setting then leads to systematic biases where the same impact categories drive the analysis regardless of the alternative's performance in other aspects, or the assigned weights (Figs. 5 and 6)—a result that agrees with Prado et al. (2017). For CML-IA, this meant that the normalized MAET category indicator result dominates the final aggregated score and hence the ranking, and in ReCiPe the normalized FET and MET indicator results. The case of bias was such that for instance, when applying a weighted sum in CML-IA with World 2000 and EU 25 normalization references (Fig. 5), a-Si ranked last and ribbon-Si ranked first despite a-Si outperforming ribbon-Si in seven impact categories, being nearly equivalent in two and *only* being more environmentally burdensome in MAET. The same trend occurs in ReCiPe with the alternative CdTe, which has a lower impact in all but two categories, FET and MET, and yet obtains a poor rank in weighted sum approaches (Fig. 6). Results showed that with a weighted sum with external normalization, it is possible for a single (or two) impact categories to determine the ranks for all alternatives despite uncertainty and differences in inventory. In fact, the dominant impact categories in the weighted sum approaches for CML-IA and ReCiPe as applied to the PV case study are those with some of the largest uncertainties at characterization (Fig. 1). Prado et al. (2017) document the same dominant impact categories in results using a weighted sum with external normalization in ReCiPe (seven published studies) and CML (six published studies and one including this finding in hundreds of processes as in White and Carty 2010) which shows a tendency in LCIA results. Anyhow, the

dominant impact categories in the weighted sum approaches included here are calculated *independently* of uncertainties at characterization. This means that when using a weighted sum with external normalization to aggregate results, there is the risk of making the weight irrelevant and thus insensitive to the underlying values that weights are intended to represent. Decisions based on such results will reflect the bias in the normalization reference and neglect the relative performance of the alternatives in highly weighted impact categories. The risk of masking weights is shown here with respect to a weighted sum with external normalization. Other studies find this phenomenon in weighted sum with internal normalization and AHP where the normalization dominates the outcome of single scores (Myllyviita et al. 2014). With TOPSIS, this effect was reduced by applying a partially compensatory method prior to the complete rank ordering (Kalbar et al. 2017). Further studies should be conducted to gain a general understanding of the relative weight sensitivity of different ranking methods.

In contrast, contributions from individual impact categories in outranking showed a greater balance, where there is no impact category dominating the composition of aggregated scores across alternatives. Consider the different visuals in the composition of scores between weighted sum approaches and outranking (Figs. 5 and 6). For the graphs depicting the outranking, there is no dominant pattern. This is due to the partially compensatory nature of the aggregation algorithm in outranking. The ranking generated by this method (Figs. 5 and 6) is defensible taking into account the characterized results (Fig. 1). In both impact assessment methods, CdTe was the most preferred alternative and Single-si was the least preferred alternative (Figs. 5 and 6). When inspecting characterized result in Fig. 1, CdTe appears to have a lower impact in most impact categories in CML-IA and ReCiPe, while single-Si tends to be at the higher end. Given these results, a ranking like the one generated with outranking is more defensible than the ranking generated with either weighted sum approaches using EU or World normalization references in CML-IA and ReCiPe, which result in ribbon-Si and a-Si ranked first, respectively. In fact, the ranking of alternatives using outranking is consistent between the two impact assessment methods because the alternatives show a similar profile at the point of characterization. The take away is that the ranking generated with outranking shows a greater sensitivity to weight values, accounts for uncertainty in the aggregation, does not rely upon normalization references which can hinder incorporation of certain aspects (as it was for WD in this case), does not have a few impact categories determining the ranking of all alternatives, and is congruent with characterized results.

Application of this outranking approach for a comparative LCA as illustrated here requires quantification of uncertainty which can be a limiting factor because uncertainty analysis is not a feature contained in all LCA software packages or versions, nor is something that LCA practitioners exercise widely, although the practice is advancing in this direction. Quantitative uncertainty information also enables calculation of preference thresholds, although these could also be based on genuine preferences from decision makers if such information is available.

When considering aggregation, it is fundamental to understand how the method deals with facts and values—the *relational* claim. A method that produces a result with limited to negligible sensitivity to weights is not a suitable method for decision support. This relational claim exits the realms of objectivity, but for science to be useful, to be applicable, it must relate appropriately to decision-maker values. We found that extent of sensitivity to the weights is variable in the aggregation methods used, where outranking confirms a higher sensitivity compared to a weighted sum with external normalization. The critique of external normalization in this study refers to aggregation purposes, and it does not apply to the other purposes of normalization such as error checking and hotspot identification in improvement assessment.

Findings of this study show that the scaling procedure (either outranking or external normalization) in the aggregation method is the most determining factor across all the feasible weight space and even beyond the performance of alternatives and uncertainty. These findings call for a reevaluation of the latest UNEP-SETAC recommendation (Pizzol et al. 2016; Verones et al. 2017) and the PEF guide (PEF 2013) to use external normalization in the aggregation of results. The recommendation also calls for using global references so as to minimize bias, but using global references does not minimize bias in aggregation as shown by this study. The bias as discussed by UNEP pertains to data gaps and discrepancy, but this committee does not acknowledge the role of compensation in aggregation which is an issue that remains beyond data repair issues. The effect of compensation still makes the use of a weighted sum with this type of external normalization questionable for aggregation of comparative LCA results.

5 Conclusions

A weighted sum approach with a linear aggregation algorithm is questionable to environmental management as it allows for a single impact category to dominate results. As opposed to what the committee from the UNEP-SETAC explains, bias reduction is not gained solely by expanding the geographical boundaries of the normalization reference but mainly by evaluating the mathematics of aggregation as it pertains to aspects of compensation (Rowley et al. 2012; Cinelli et al. 2014; Pollesch and Dale 2015, 2016; Seager and Prado 2017).

It is also necessary to re-asses the ISO guidelines 14044 (ISO 2006) as far as normalization and weighting are concerned. Although ISO 14044 states that normalization could

be helpful in "preparing for additional procedures, such as grouping, weighting, or life cycle interpretation," in practice, it is shown that it can lead to biased aggregated results. Moreover, the ISO 14044 states that weighted scores "shall not be used in LCA studies intended to be used in comparative assertions intended to be disclosed to the public". Contrary to this recommendation, sharing weighted results does not result in imposing values to an audience because it has been shown that weights have limited influence. It is in fact the scaling step (normalization step) that has the most influence. This has already been documented in the literature in and out of the LCA field (Stewart 2008; Rogers and Seager 2009; Myllyviita et al. 2014; Pollesch and Dale 2015; Wulf et al. 2017). Given the biases in external normalization, adopting externally normalized results for comparative assertions should be re-evaluated as it might bias the interpretation of overall environmental performance to one or a few dominating impact categories.

Equally problematic is the fact that even in the event of legitimate weights consisting of importance coefficients, these can have little effect on the results of the weighted sum with external normalization, which would be already determined by the normalization reference. It is thus recommended that practitioners and researchers become more aware of the methodological implications of aggregation methods to advance the interpretation of LCIA results. As LCA becomes more important in decision-making and given efforts in expanding the scope to all sustainability dimensions, interpretation methods that mask the role of value systems can be detrimental to decision making. Aggregation methods based on outranking algorithms, as shown here, are partially compensatory, account for uncertainty, operate with importance coefficient weights, and do not rely upon external normalization references, which makes them appropriate candidates for dealing with aggregation of multiple sustainability dimensions.

Funding information Marco Cinelli declares that this project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 743553.

References

Ahlroth S, Nilsson M, Finnveden G et al (2011) Weighting and valuation in selected environmental systems analysis tools – suggestions for further developments. J Clean Prod 19:145–156. https://doi.org/10.1016/j.jclepro.2010.04.016

Behzadian M, Kazemzadeh RB, Albadvi A, Aghdasi M (2010) PROMETHEE: a comprehensive literature review on methodologies and applications. Eur J Oper Res 200:198–215. https://doi.org/10.1016/j.ejor.2009.01.021

Bengtsson M, Steen B (2000) Weighting in LCA – approaches and applications. Environ Prog 19:101–109. https://doi.org/10.1002/ep.670190208

Bertola NJ, Cinelli M, Casset S et al (2019) A multi-criteria decision framework to support measurement-system design for bridge load testing. Adv Eng Inform 39:186–202. https://doi.org/10.1016/j.aei.2019.01.004

Cao XH, Stojkovic I, Obradovic Z (2016) A robust data scaling algorithm to improve classification accuracies in biomedical data. BMC Bioinforma 17:1–10. https://doi.org/10.1186/s12859-016-1236-x

Castellani V, Benini L, Sala S, Pant R (2016a) A distance-to-target weighting method for Europe 2020. Int J Life Cycle Assess 21:1159–11669. https://doi.org/10.1007/s11367-016-1079-8

Castellani V, Sala S, Benini L (2016b) Hotspots analysis and critical interpretation of food life cycle assessment studies for selecting eco-innovation options and for policy support. J Clean Prod. https://doi.org/10.1016/j.jclepro.2016.05.078

Cinelli M, Coles SR, Kirwan K (2014) Analysis of the potentials of multi criteria decision analysis methods to conduct sustainability assessment. Ecol Indic 46:138–148. https://doi.org/10.1016/j.ecolind.2014.06.011

Corrente S, Figueira JR, Greco S (2014) The SMAA-PROMETHEE method. Eur J Oper Res 239:514–522. https://doi.org/10.1016/j.ejor.2014.05.026

Cucurachi S, Seager TP, Prado V (2017) Normalization in comparative life cycle assessment to support environmental decision making. J Ind Ecol 21:242–243. https://doi.org/10.1111/jiec.12549

Dias LC, Passeira C, Malça J, Freire F (2016) Integrating life-cycle assessment and multi-criteria decision analysis to compare alternative biodiesel chains. Ann Oper Res:1–16. https://doi.org/10.1007/s10479-016-2329-7

Domingues AR, Marques P, Garcia R et al (2015) Applying multi-criteria decision analysis to the life-cycle assessment of vehicles. J Clean Prod 107:749–759. https://doi.org/10.1016/j.jclepro.2015.05.086

Du C, Dias LC, Freire F (2019) Robust multi-criteria weighting in comparative LCA and S-LCA: a case study of sugarcane production in Brazil. J Clean Prod 218:708–717. https://doi.org/10.1016/J.JCLEPRO.2019.02.035

Edwards W, Barron FH (1994) Smarts and smarter: improved simple methods for multiattribute utility measurement. Organ Behav Hum Decis Process 60:306–325

Figueira JR, Roy B (2009) A note on the paper, "ranking irregularities when evaluating alternatives by using some ELECTRE methods", by Wang and Triantaphyllou, Omega (2008). Omega 37:731–733. https://doi.org/10.1016/j.omega.2008.05.001

Fischer G (1995) Range sensitivity of attribute weights in multiattribute value models. Organ Behav Hum Decis Process 62:252–266

Goedkoop M, Heijungs R, Huijbergts M et al (2009) ReCiPe 2008. Report 1: Characterisation

Greco S, Ehrgott M, Rui Figueira J (2016) Multiple criteria decision analysis: state of the art surveys, 2nd edn. Springer New York Heidelberg Dordrecht London

Greco S, Ishizaka A, Matarazzo B, Torrisi G (2018a) Stochastic multiattribute acceptability analysis (SMAA): an application to the ranking of Italian regions. Reg Stud 52:585–600. https://doi.org/10.1080/00343404.2017.1347612

Greco S, Ishizaka A, Tasiou M, Torrisi G (2018b) On the methodological framework of composite indices: a review of the issues of weighting, aggregation, and robustness. Soc Indic Res 141:1–34. https://doi.org/10.1007/s11205-017-1832-9

Guinée JB, Gorree M, Heijungs R, et al (2002) Handbook on life cycle assessment - operational guide to the ISO standards. In: Guinée JB (ed) Handbook on life cycle assessment: operational guide to the ISO standards Series: Eco-Efficiency in Industry and Science. Springer, Dordrecht

Heijungs R, de Koning A, Ligthart T, Korenromp R (2004) Improvement of LCA characterization factors and LCA practice for metals. Apeldoorn

Heijungs R, Guinée J, Kleijn R, Rovers V (2007) Bias in normalization: causes, consequences, detection and remedies. Int J Life Cycle Assess 12:211–216

Hertwich EG, Hammitt JK, Pease WS (2000) A theoretical foundation for life-cycle assessment. J Ind Ecol 4:13–28. https://doi.org/10.1162/108819800569267

Huppes G, van Oers L, Pretato U, Pennington D (2012) Weighting environmental effects: analytic survey with operational evaluation methods and a meta-method. Int J Life Cycle Assess:1–16. https://doi.org/10.1007/s11367-012-0415-x

ISO (2006) ISO 14044: environmental management — life cycle assessment — requirements and guidelines. Environ Manag 3:54

Itsubo N, Murakami K, Kuriyama K et al (2015) Development of weighting factors for G20 countries—explore the difference in environmental awareness between developed and emerging countries. Int J Life Cycle Assess:1–16. https://doi.org/10.1007/s11367-015-0881-z

Ji C, Hong T (2016) New internet search volume-based weighting method for integrating various environmental impacts. Environ Impact Assess Rev 56:128–138. https://doi.org/10.1016/j.eiar.2015.09.008

Jungbluth N, Stucki M, Flury K, Frischknecht R (2012) Life Cycle Inventories of Photovoltaics, ESU-services Ltd.: Uster, CH, 2012.

Kalbar PP, Birkved M, Nygaard SE, Hauschild M (2017) Weighting and aggregation in life cycle assessment: do present aggregated single scores provide correct decision support? J Ind Ecol 21:1591–1600. https://doi.org/10.1111/jiec.12520

Keeney RL (2002) Common mistakes in making value trade-offs. Oper Res 50:935–945. https://doi.org/10.1287/opre.50.6.935.357

Kim J, Yang Y, Bae J, Suh S (2013) The importance of normalization references in interpreting life cycle assessment results. J Ind Ecol 17:385–395. https://doi.org/10.1111/j.1530-9290.2012.00535.x

Laurin L, Amor B, Bachmann TM, Bare J, Koffler C, Genest S, Preiss P, Pierce J, Satterfield B, Vigon B (2016) Life cycle assessment capacity roadmap (section 1): decision-making support using LCA. Int J Life Cycle Assess 21:443–447. https://doi.org/10.1007/s11367-016-1031-y

Matarazzo A, Clasadonte MT, Ingrao C, Zerbo A (2013) Criteria interaction modelling in the framework of Lca analysis. Int J Eng Res Appl 3:523–530

Muller S, Lesage P, Ciroth A, et al (2014) The application of the pedigree approach to the distributions foreseen in ecoinvent v3. Int J Life Cycle Assess. https://doi.org/10.1007/s11367-014-0759-5

Munda G (2016) Multiple criteria decision analysis and sustainable development. In: Greco S, Ehrgott M, Figueira JR (eds) Multiple Criteria Decision Analysis. State of the Art Surveys, New York, pp 1235–1267

Myllyviita T, Leskinen P, Seppälä J (2014) Impact of normalisation, elicitation technique and background information on panel weighting results in life cycle assessment. Int J Life Cycle Assess 19:377–386. https://doi.org/10.1007/s11367-013-0645-6

Nardo M, Saisana M, Saltelli A et al (2008) Handbook on Constructing Composite Indicators: Methodology and User Guide. OECD, JRC European Commission

Norris G a (2001) The requirement for congruence in normalization. Int J Life Cycle Assess 6:85–88. https://doi.org/10.1007/BF02977843

Nzila C, Dewulf J, Spanjers H et al (2012) Multi criteria sustainability assessment of biogas production in Kenya. Appl Energy 93:496–506. https://doi.org/10.1016/j.apenergy.2011.12.020

PEF (2013) Recommendations on the use of common methods to measure and communicate the life cycle environmental performance of products and organizations

Pizzol M, Weidema B, Brandão M, Osset P (2015) Monetary valuation in life cycle assessment: a review. J Clean Prod 86:170–179. https://doi.org/10.1016/j.jclepro.2014.08.007

Pizzol M, Laurent A, Sala S et al (2016) Normalisation and weighting in life cycle assessment: quo vadis? Int J Life Cycle Assess. https://doi.org/10.1007/s11367-016-1199-1

Pollesch N, Dale VH (2015) Applications of aggregation theory to sustainability assessment. Ecol Econ 114:117–127. https://doi.org/10.1016/j.ecolecon.2015.03.011

Pollesch NL, Dale VH (2016) Normalization in sustainability assessment: methods and implications. Ecol Econ 130:195–208. https://doi.org/10.1016/j.ecolecon.2016.06.018

Prado V, Heijungs R (2018) Implementation of stochastic multi attribute analysis (SMAA) in comparative environmental assessments. Environ Model Softw 109:223–231. https://doi.org/10.1016/j.envsoft.2018.08.021

Prado V, Rogers K, Seager TP (2012) Integration of MCDA tools in valuation of comparative life cycle assessment. In: Curran MA (ed) Life cycle assessment handbook: a guide for environmentally sustainable products. Wiley, Hoboken, pp 413–431

Prado V, Wender BA, Seager TP (2017) Interpretation of comparative LCAs: external normalization and a method of mutual differences. Int J Life Cycle Assess 22:2018–2029. https://doi.org/10.1007/s11367-017-1281-3

Prado-Lopez V, Seager TP, Chester M et al (2014) Stochastic multi-attribute analysis (SMAA) as an interpretation method for comparative life-cycle assessment (LCA). Int J Life Cycle Assess 19:405–416. https://doi.org/10.1007/s11367-013-0641-x

Ravikumar D, Seager TP, Cucurachi S et al (2018) Novel method of sensitivity analysis improves the prioritization of research in anticipatory life cycle assessment of emerging technologies. Environ Sci Technol 52:6534–6543. https://doi.org/10.1021/acs.est.7b04517

Riabacke M, Danielson M, Ekenberg L (2012) State-of-the-art prescriptive criteria weight elicitation. Advances in Decision Sciences. https://doi.org/10.1155/2012/276584

Rogers M, Bruen M (1998) Choosing realistic values of indifference, preference and veto thresholds for use with environmental criteria within ELECTRE. Eur J Oper Res 107:542–551. https://doi.org/10.1016/S0377-2217(97)00175-6

Rogers K, Seager TP (2009) Environmental decision-making using life cycle impact assessment and stochastic multiattribute decision analysis: a case study on alternative transportation fuels. Environ Sci Technol 43:1718–1723. https://doi.org/10.1021/es801123h

Rowley HV, Peters GM, Lundie S, Moore SJ (2012) Aggregating sustainability indicators: beyond the weighted sum. J Environ Manag 111:24–33. https://doi.org/10.1016/j.jenvman.2012.05.004

Roy B (1985) Méthodologie multicritère d'aide à la décision. Economica, Paris

Seager TP, Prado V (2017) Letter to the editor on "weighting and aggregation in life cycle assessment: do present aggregated single scores provide correct decision support?". J Ind Ecol. https://doi.org/10.1111/jiec.12559

Sohn JL, Kalbar PP, Birkved M (2017) Life cycle based dynamic assessment coupled with multiple criteria decision analysis: a case study of determining an optimal building insulation level. J Clean Prod 162:449–457. https://doi.org/10.1016/j.jclepro.2017.06.058

Steele K, Carmel Y, Cross J, Wilcox C (2009) Uses and misuses of multicriteria decision analysis (MCDA) in environmental decision making. Risk Anal 29:26–33. https://doi.org/10.1111/j.1539-6924.2008.01130.x

Stewart TJ (2008) Robustness Analysis and MCDA. In: European Working Group Multiple Criteria Decision Aiding. Newsletter of the European Working Group "Multicriteria Aid for Decisions"

Tervonen T, Lahdelma R (2007) Implementing stochastic multicriteria acceptability analysis. Eur J Oper Res 178:500–513. https://doi.org/10.1016/j.ejor.2005.12.037

Tervonen T, van Valkenhoef G, Buskens E, Hillege HL, Postmus D (2011) A stochastic multicriteria model for evidence-based decision making in drug benefit-risk analysis. Stat Med 30:1419–1428. https://doi.org/10.1002/sim.4194

Tuomisto HL, Hodge ID, Riordan P, Macdonald DW (2012) Exploring a safe operating approach to weighting in life cycle impact assessment – a case study of organic, conventional and integrated farming systems. J Clean Prod 37:147–153. https://doi.org/10.1016/j.jclepro.2012.06.025

Tylock SM, Seager TP, Snell J et al (2012) Energy management under policy and technology uncertainty. Energy Policy 47:156–163. https://doi.org/10.1016/j.enpol.2012.04.040

Verones F, Bare J, Bulle C et al (2017) LCIA framework and cross-cutting issues guidance within the UNEP-SETAC life cycle initiative. J Clean Prod 161:957–967. https://doi.org/10.1016/j.jclepro.2017.05.206

White P, Carty M (2010) Reducing bias through process inventory dataset normalization. Int J Life Cycle Assess 15:994–1013. https://doi.org/10.1007/s11367-010-0215-0

Wulf C, Zapp P, Schreiber A et al (2017) Lessons learned from a life cycle sustainability assessment of rare earth permanent magnets. J Ind Ecol 00:1–13. https://doi.org/10.1111/jiec.12575

Zanghelini GM, Cherubini E, Soares SR (2018) How multi-criteria decision analysis (MCDA) is aiding life cycle assessment (LCA) in results interpretation. J Clean Prod 172:609–622. https://doi.org/10.1016/j.jclepro.2017.10.230