Meta-comparisons: how to compare methods for LCA

Heijungs, R.; Dekker, E.

The International Journal of Life Cycle Assessment (2022) 27:993–1015
https://doi.org/10.1007/s11367-022-02075-4

UNCERTAINTIES IN LCA

Meta-comparisons: how to compare methods for LCA?

Reinout Heijungs[1,2] · **Erik Dekker**[3]

Received: 28 September 2021 / Accepted: 20 June 2022 / Published online: 8 July 2022

Abstract

Introduction Many methodological papers report a comparison of methods for LCA, for instance comparing different impact assessment systems, or developing streamlined methods. A popular way to do so is by studying the differences of results for a number of products. We refer to such studies as quasi-empirical meta-comparisons.

Review of existing approaches A scan of the literature reveals that many different methods and indicators are employed: contribution analyses, Pearson correlations, Spearman correlations, regression, significance tests, neural networks, etc.

Critical discussion We critically examine the current practice and conclude that some of the widely used methods are associated with important deficits.

A new approach Inspired by the critical analysis, we develop a new approach for meta-comparative LCA, based on directional statistics. We apply it to several real-world test cases, and analyze its performance vis-à-vis traditional regression-based approaches.

Conclusion The method on the basis of directional statistics withstands the tests of changing the scale and unit of the training data. As such, it holds a promise for improved method comparisons.

Keywords Comparative LCA · Simplified LCA · Streamlined LCA · Proxy indicators · Regression · Correlation · Directional statistics

1 Introduction

The majority of the life cycle assessment (LCA) studies is relative in the sense of involving a comparison (Heijungs et al. 2019). Comparative LCA studies are usually dealing with the comparison of alternative products that fulfill a similar function (such as an electric car and a gasoline car) or the comparison of alternative production processes that produce the same product (such as coal-based electricity and nuclear electricity). They are done to take decisions regarding the best-performing (minimum-impact) product or process. But there is another class of comparisons which is more methodological. Such comparisons focus on alternative methods for calculating LCA results. Here we refer to such studies as "meta-comparisons," because they take a more abstract and higher vantage point.

This article provides a critical analysis of papers that engage in meta-comparative LCA, comparing methods for LCA. Because we will compare methods for meta-comparison, our paper may be classified as meta-meta-comparative (see https://xkcd.com/1447/). We believe that ours is the first paper studying meta-comparisons (although we acknowledge that Pizzol et al. (2011) observed that "it is not straightforward to compare the methods," and that Dong et al. (2016) wrote that "there is a lack of an agreed approach that can differentiate various [life cycle impact assessment] methods"), and that even the term "meta-comparison" has not been used before in the context of LCA.

As a first defining feature, we emphasize that our paper is not about so-called "meta-analysis of LCA" (Brandão et al. 2012) or meta-regression of LCA (Menten et al. 2013). Such studies try to draw generic lessons for product groups from a

Communicated by Mark Huijbregts

✉ Reinout Heijungs
r.heijungs@vu.nl

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

[2] Institute of Environmental Sciences, Leiden University, Leiden, The Netherlands

[3] National Institute for Public Health and the Environment, Bilthoven, The Netherlands

limited number of studies. Meta-comparative LCA, by contrast, aims to draw lessons on methods for LCA.

With the title's term "methods for LCA," we have several groups of studies in mind. Below, we briefly review the literature in a number of major topics:

- comparing life cycle impact assessment (LCIA) methods;
- comparing inventory (LCI) methods;
- developing streamlined LCA methods; and
- comparing software and databases.

As a sidenote, we emphasize that our meaning of "methods for LCA" is much broader than LCIA methods: it includes methods for the full LCA calculation.

A major group is provided by the studies that compare competing LCIA methods. An early example is Baumann and Rydberg (1994), who compare three LCIA methods that employ different principles. Most later analyses focus specifically on comparing characterization methods (Dreyer et al. 2003; Van der Werf and Petit 2002; Landis and Theis 2008; Weidema 2015; Chen et al. 2021), but some authors concentrate on normalization methods (Lautier et al. 2010; Myllyviita et al. 2014) or weighting methods (Huppes et al. 2012); Myllyviita et al. 2014), or address the full LCIA pathway (Notarnicola et al. 1998; Brent and Hietkamp 2003; Bovea and Gallardo 2006). This group also includes studies that compare LCA results of an established impact assessment with an updated version (Dekker et al. 2020).

A second group is formed by the studies that compare process-based, IO-based and hybrid inventories. Here we mention Hendrickson et al. (1997), Suh and Huppes (2005), Junnila (2006), Islam et al. (2016), and Crawford et al. (2018) as key representatives. In this group, we also include studies that investigate the effects of other inventory choices, such as allocation (Huijbregts 1998; Curran 2007) and algorithm (Heijungs et al. 2015).

A third group comprises the studies that use a streamlined method to approximate an LCA. For instance, Huijbregts et al. (2006) propose the use of cumulative energy demand (CED) of products as a proxy for addressing the impact scores for a host of other impact categories, including global warming, stratospheric ozone depletion, acidification, eutrophication, photochemical ozone formation, land use, resource depletion, and human toxicity. This idea has been further refined by, among others, Röös et al. (2013), Scipioni et al. (2013), and Steinmann et al. (2017a). A more abstract version of this type of analysis is the study on the use of a product property (such as life span or weight) as a predictor (Padey et al. 2013; Eddy et al. 2015). The idea has also been employed to predict characterization factors (e.g., Birkved and Heijungs 2011) or to predict entire LCA scores from chemical properties (e.g., Wernet et al. 2008; Eckelman 2016).

A final group of studies compares software and databases for LCA, using the same settings (system boundaries, LCIA methods, etc.). Examples include Speck et al. (2015), Herrmann and Moltesen (2015), and Iswara et al. (2020). It also includes studies that compare algorithms for LCA, like Peters (2007) and Heijungs et al. (2015). Notice that software is sometimes mixed up with the implemented data. An example is the study by Martínez et al. (2015), which sets out to compare software, but effectively compares different LCIA methods.

Some of these studies are of an analytical nature: they dissect the logical and/or mathematical structure of the contrasting methods and expose differences in assumptions, principles, and value choices. Examples are Van der Werf and Petit (2002), Amani et al. (2011), Núñez et al. (2016), and Crawford et al. (2018).

Other studies employ a quasi-empirical set-up. They use different LCA methods to calculate results for a number of products, and then check the degree of agreement between these results. Within this approach, there is quite some diversity in the details. One extreme is presented by Dreyer et al. (2003), who apply several LCA methods to just one product, and use a contribution analysis to assess the degree of correspondence. Another extreme is Laurent et al. (2012), who use up to 3954 products and calculate correlation coefficients and other statistics. Many studies are in-between: for instance, Cavalett et al. (2013) base their analysis on two products, and Röös et al. (2013) use 53 products. Notice that we speak of quasi-empirical. Empirical studies work with observed data, but in quasi-empirical, the data is constructed from an available unit process set or LCI database.

Our focus in this article is on these quasi-empirical studies. We will analyze a number of such studies and seek to find out the approaches used and the strengths and weaknesses of those approaches given the specific purpose of the analysis.

Our motivation for this purpose is that there is not only little methodological guidance for comparative LCA (the only sources we are aware of are Heijungs and Suh (2002); Jung et al. (2014)), but that the situation for meta-comparative LCA is even more obscure (Pizzol et al. 2011; Dong et al. 2016). For instance, Huijbregts et al. (2006) report regression coefficients and R^2 statistics while Dekker et al. (2020) report t statistics, root mean square errors (RMSEs), and Spearman correlation coefficients, and Simões et al. (2011) compare the LCIA results with a primarily verbal approach. In some cases, the set of products used as benchmark displays a large span of orders of magnitude for the scores, which then induces some researchers (e.g., Bösch et al. 2007; Steinmann et al. 2017a) to study logarithmic relationships. Some authors speak of "significant correlations" (e.g., Berger and Finkbeiner 2011) without using a hypothesis test; others explicitly set a

"significance alpha" (e.g., Pascual-González et al. 2016). A further complication is that the studies use different words and symbols for the techniques and indicators (for instance, the "correlation coefficient" is to Huijbregts et al. (2006) r^2, while it is r to Röös et al. (2013)), that even the same article sometimes is not consistent in its symbol use (e.g., Pascual-González et al. (2016) show several figures with an Ω-axis, which is in their text probably I_k), that equations are sometimes absent anyhow (e.g., in Sousa et al. 2000), that equations in a few cases contain mistakes (e.g., Kaufmann et al. (2010), show an Eq. (1) in which a term $\max_j \{E_i m_i\}$ occurs), and that many more things can go wrong (e.g., Ligthart and Ansems (2019) report cases with "$p < 0.00$".). Some authors do not specify equations but refer to specific software. For instance, Dekker et al. (2020) write that the "statistical analysis was done with R-studio version 3.4.0," which further complicates finding out the details, especially when no code is provided as supplementary information. Altogether, it appears that meta-comparative LCA has been practiced a lot, but that there is no guidance, let alone agreement on the methodological basis for carrying out such studies.

Some of the approaches have been criticized, for a variety of reasons. Hanes et al. (2013) criticize the use of log-transformed variables, and Heijungs (2017) comments on the absence of random sampling which would rule out the use of confidence intervals and p values. Valente et al. (2019) check the relationship between global warming and acidification for a number of hydrogen production systems, but they find a disappointing goodness-of-fit. Another possible critique is that many statistical techniques need assumptions, for instance, normal distributions or independence, and that such assumptions are often not mentioned or not checked for. Also the role of confounding variables (Pourhoseingholi et al. 2012) is in general not checked for.

Altogether, it appears that methodological guidance is needed to facilitate meta-comparative LCA, in order to eventually improve LCA and LCIA practices, reduce uncertainties, evaluate robustness of outcomes, and improve decision support. In Sect. 2, we will analyze a large number of such meta-comparative studies. Section 3 will examine the major techniques in terms of desirable and undesirable properties. Section 4 will then propose an innovative technique, which will be illustrated with the data set from Dekker et al. (2020). Section 5 summarizes and concludes.

2 Review of existing approaches

In this section, we analyze a number of meta-comparative LCA studies in order to extract the approach taken. In this process, we will focus on the quasi-empirical studies, which we define to be studies that calculate LCA results (LCI, midpoint LCIA, endpoint LCIA, weighting) for a number of products with two or more different LCA methods, which are next submitted to a quantitative analysis in order to draw conclusions on the agreement or disagreement between the methods.

2.1 Notation and terminology

Because every study employs its own notation and terminology, we will introduce a uniform set of principles here. The analysis is done on the basis of a sample of n products. The scores for one indicator will be denoted by x_i $(i = 1, \ldots, n)$ and for the other indicator, it will be y_i. For instance, the x scores may be the values for the predictor or streamlined or old characterization method, and the y scores the values for the predicted or full or new method.

Within this format, we discern five major purposes of the studies:

- streamlining;
- proxy;
- reduction;
- comparison; and
- sensitivity.

Streamlining includes those studies that attempt to mimic a full result (y_i) by means of another, more easily determined, result (x_i). The interesting question is then to what extent x_i resembles y_i. A particular characteristic of such studies is that x and y have the same unit. For instance, both are expressed in kg CO_2-equivalent. For an example, we refer to Frischknecht et al. (2007), who study to what extent the results of an LCA are influenced by ignoring capital goods. In those studies in which a value is predicted, we use a hat to indicate the predicted value. For instance, y_i is the observed value for product i, and $\hat{y}_i$ is the predicted value.

The group of proxy studies includes studies that attempt to establish or test a relationship between a proxy indicator (x_i) and the real indicator (y_i). In this case, the purpose is not necessarily to mimic y_i, but rather to find out to what extent choices ("product A is the best"), rankings ("product A is better than product B"), or subdivisions of scores ("60% of the score for product A is caused by transport") are stable when x scores are used instead of y scores. Here the x and y scores may have different units; for instance, x is in MJ of primary energy and y in kg CO_2-equivalent. An example is the paper by Huijbregts et al. (2006), where the cumulative energy demand is the predictor (x) and a variety of impact categories (global warming, stratospheric ozone depletion, acidification, etc.) is the variable-to-be-predicted (y).

With reduction studies, we embrace studies that seek to reduce the number of indicators to a smaller subset. A typical example is provided by Steinmann et al. (2016), who

attempt to reduce the "hundreds of indicators" to "a nonredundant key set of indicators representative of the overall environmental impact."

The group of comparison studies comprises studies that do not seek to predict or provide a proxy, but that merely try to find out how different the results are. Baumann and Rydberg (1994) provide a typical example here. This type also includes the study of updates. For instance, Dekker et al. (2020) use ReCiPe2008 for the x and ReCiPe 2016 for the y. In some cases, the two variables will have equal units, but there may also be situations in which this is not the case. A variation of this are studies like those by Junnila (2006), which compare process-based (x) and input–output-based (y) result, without necessarily declaring that one is better than the other one.

Finally, there are studies that primarily study one specific product and apply several methods (e.g., several LCIA methods) to study how robust the result is for methodological choices. We refer to these as sensitivity studies. A typical example is Cavalett et al. (2013), who compare gasoline and ethanol "using different LCIA methods."

These five purposes are summarized in Table 1.

Many quasi-empirical, meta-comparative studies employ overall descriptive statistics that are computed from the (x_i, y_i) data, such as a correlation coefficient or p values. The five types of studies may require different types of statistics. After all, for the streamlining group, we expect that the y_i is close to the x_i for the majority of products, but for the proxy group, we might be more interested in a robust ranking of the products. As such, there is no universally best meta-comparison indicator. Instead, a purpose-dependent result may appear to emerge.

Because ranking can be important for certain applications, we need to introduce the idea more precisely. Order statistics of a data vector refer to a rearrangement of the data vector, such that the elements are ordered from small to large. The i th order statistic of a data vector with elements $x_1, \ldots, x_n$ is indicated by $x_{(i)}$. Altogether, we have $x_{(1)} \le x_{(2)} \le \cdots \le x_{(n-1)} \le x_{(n)}$. Using this notation, we can easily indicate the smallest value of x by $x_{(1)}$ and the largest value by $x_{(n)}$. Ranks refer to the place of a particular value x_i in the vector of order statistics. Often, symbols like R_i are used to indicate ranks, but as we need to be able to distinguish the ranks of the x- and y-series, we prefer the notation R_{x_i} and R_{y_i}. In ranking, a choice has to be made about how to handle ties. Ties occur when two or more data points have the same value. We will adopt the midrank convention, in which all data with the same value will receive an average rank (Agresti 2002).

Ranking implies a preference. If $x_i > x_j$ and a lower value of x is preferable ("less is better"), we have $R_{x_i} < R_{x_j}$. We further write in that case that $i \prec j$, meaning that product i has a lower preference than product j. The symbol $\sim$ indicates indifference.

In some cases, we will need to work with the average value of x or y, over the entire set of products. For this. we use the bar-notation:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{1}$$

and similar for $\bar{y}$. Likewise, the standard deviation will be indicated by s, with possible subscripts for x and y:

$$s_x = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2} \tag{2}$$

and similar for s_y. The variances are then simply the squared standard deviations: s_x^2 and s_y^2.

Some authors do not analyze the raw data, but use the logarithm of the values of x_i and y_i. In the analysis below, we will pay particular attention to this aspect. On the other hand, we will, at some places, ignore the use of logarithms, and just provide formulas with x and y for which, if needed, $\log(x)$ and $\log(y)$ may be inserted, which may indicate 10-log or natural logarithm.

In some studies, there are multiple x variables. We will then write k for the number of x variables. We will indicate the values for product i as $x_{i1}, x_{i2}, \ldots, x_{ik}$. The data can then be conceived as building a data matrix $\mathbf{X}$.

In some of the studies analyzed, hypothesis tests are used. The test statistic will be indicated by symbols like z, t, and F for the standard normal, Student t, and Fisher F distribution, and where needed degrees of freedom will be indicated by df, df_1, etc. The resulting p values will be

Table 1 Proposed differentiated use of statistical techniques per purpose

Purpose	Main characteristic	Example
Streamlining	Finding a short-cut in doing LCA	Ignoring capital goods
Proxy	Using an easy indicator	Using cumulative energy demand to predict usual impact categories
Reduction	Reducing the number of indicators	Leaving out impacts that yield no extra information
Comparison	Comparing two competing methods	Using IO-based or process-based LCI
Sensitivity	Testing for robust conclusions	Checking if USEtox and ReCiPe yield the same preferences

indicated by p, and the critical value for significance by α. The null hypothesis is indicated by H_0 and the alternative hypothesis by H_1. We use the convention that H_0 and H_1 are complementary (see, e.g., Ott and Longnecker 2015), for instance $H_0\colon\ \mu \geq 0$ versus $H_1\colon\ \mu < 0$. In this, we deviate from some other texts (e.g., Agresti and Franklin 2013), who use $H_0\colon\ \mu = 0$ versus $H_1\colon\ \mu < 0$.

In a situation of sampling, we should distinguish population parameters and sample statistics. In general, we will use Greek letters, like σ and β, for parameters, and their Roman equivalents, like s and b for their realized values in a sample. An exception is the mean, for which the parameter is μ and the statistic $\bar{x}$. The sample statistic as a random variable will be denoted by Roman capitals, such as $\overline{X}$, S, and B.

In the following sections, we will analyze the approaches by the major approaches from literature. We will often, without notice, change the original symbols to agree with the uniform principle outlined above. We will also sometimes add, or remove, some other details, such as indices and summation symbols.

2.2 Review of studies

There is no objective bibliometric way to identify meta-comparative LCA studies. For the purpose of our review, we selected studies on the basis of our private knowledge of the literature, including the references in and to those papers. This resulted in a collection of around 100 papers, most of which were published in peer-reviewed journals. With a focus on quasi-empirical methods, this number slightly reduces. Table 2 provides an overview of the selected articles, with an indication of their main characteristics.

The table reveals that meta-comparative LCA in fact has been done quite often, by many authors, on different topics, and using an array of techniques. Nevertheless, we can discern a number of trends:

- LCIA, and in particular the characterization, is the most popular topic;
- comparison is the most popular purpose; and
- the most popular statistical techniques are correlation/regression and the presentation of differences or contribution analyses.

In the next few sections, we discuss the statistical techniques in more detail.

2.3 Individual measures of difference

If the score of product i for one method is indicated by x_i and for the other method by y_i, we can form various measures of difference. We first discuss the one-by-one measures, and then move to overall indicators.

Several studies (e.g., Junilla 2006; Weidema 2015) list the x and y scores without any further processing. Valente et al. (2018) look at the difference between the two scores:

$$d_i = x_i - y_i \tag{3}$$

A variation in the form of ratios is used by Herrmann and Moltesen (2015) as well as by Huijbregts et al. (2008):

$$r_i = \frac{x_i}{y_i} \tag{4}$$

Frischknecht et al. (2007) use an indicator of the type:

$$\delta_i = \frac{x_i - y_i}{x_i} \tag{5}$$

This indicator expresses the relative error of using y_i instead of x_i. Crawford (2008) also uses this indicator, giving it the name "GAP."

Several studies (e.g., Simões et al. 2011; Monteiro and Freire 2012; Cavalett et al. 2013) visualize the results with the largest indicator set to 100%:

$$x_{\mathrm{rel},i} = \frac{x_i}{\max_{i=1}^{n} x_i} \times 100\% \text{ and } y_{\mathrm{rel},i} = \frac{y_i}{\max_{i=1}^{n} y_i} \times 100\% \tag{6}$$

Valente et al. (2018) do a similar thing, but they use the largest of both methods as a reference, inserting $\max_{i=1}^{n}\left(x_i, y_i\right)$ in the denominator for both expressions.

There are also studies where one product is used as a reference. For instance, Notarnicola et al. (1998) use the score for steel in the denominator.

Peters (2007) compares two algorithms for solving an IO-based LCI. He calculates an "error," which compares the two methods, as well as a "tolerance". Unfortunately, the precise details are not specified. We guess that the error is defined as $\frac{x_i - y_i}{x_i}$, but what exactly is used here for x_i (sector outputs, emissions) is unclear.

2.4 Aggregated measures of difference

The indicators above express differences per product. As such, they are less suitable for studies that address a large number of products, such as Huijbregts et al. (2006) and Pascual-González et al. (2016). In this section, we discuss the overall indicators, in which some form of aggregation or averaging over all products ($i = 1, \ldots, n$) is made.

Dekker et al. (2020) use a number of indicators. These include the root mean square error (RMSE), defined as follows:

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(x_i - y_i\right)^2} \tag{7}$$

Table 2 Overview of articles on meta-comparative LCA that apply a quasi-empirical procedure

Reference	Topic	Purpose	Statistical techniques
Balugani et al. (2021)	LCIA	Reduction	PCA, robust ordinal regression
Baumann and Rydberg (1994)	LCIA	Comparison	Contribution analysis
Berger and Finkbeiner (2011)	LCIA	Reduction	Correlation, regression
Birkved and Heijungs (2011)	LCIA	Streamlining	PLSR
Bösch et al. (2007)	LCIA	Proxy	Correlation, regression, contribution analysis
Bovea and Gallardo (2006)	LCIA	Comparison	Contribution analysis
Brent and Hietkamp (2003)	LCIA	Comparison	Contribution analysis
Bueno et al. (2016)	LCIA	Sensitivity	Contribution analysis
Cavalett et al. (2013)	LCIA	Sensitivity	Differences
Chen et al. (2021)	LCIA	Sensitivity	Contribution analysis
Cherubini et al. (2018)	LCI	Sensitivity	Contribution analysis, rankings
Crawford (2008)	LCI	Comparison	Differences, correlation
Curran (2007)	LCI	Comparison	Differences
Curzons et al. (2007)	LCIA	Reduction	Regression, PCA, cluster analysis
Dekker et al. (2020)	LCIA	Comparison	Differences, correlations
De Rosa et al. (2018)	LCIA	Comparison	Contribution analysis
Dewulf et al. (2007)	LCIA	Comparison	Contribution analysis
Dong et al. (2016)	LCIA	Comparison	Correlation
Dreyer et al. (2003)	LCIA	Comparison	Differences, contribution analysis
Eckelman (2016)	LCIA	Streamlining	Correlation
Eddy et al. (2015)	LCI, LCIA	Proxy	Kriging
Emami et al. (2019)	LCI, LCIA	Comparison	Differences, contribution analysis
Frischknecht et al. (2007)	LCI	Streamlining	Differences
Gutiérrez et al. (2009)	LCIA	Reduction	Correlation, MDS
Gutiérrez et al. (2010a)	LCIA	Reduction	Correlation, PCA, cluster analysis
Gutiérrez et al. (2010b)	LCIA	Reduction	Correlation, PCA, cluster analysis
Halleux et al. (2006)	LCIA	Comparison	Contribution analysis
Hanes et al. (2013)	LCI	Comparison	Differences
Heijungs et al. (2015)	LCI	Comparison	Differences
Hendrickson et al. (1997)	LCI	Comparison	Differences
Herrmann and Moltesen (2015)	Software	Comparison	Differences
Hochschorner and Finnveden (2003)	LCI, LCIA	Comparison	Verbal
Hou et al. (2020)	LCIA	Streamlining	Neural network and other ML techniques
Huijbregts (1998)	LCI	Sensitivity	*
Huijbregts et al. (2005)	LCIA	Comparison	Differences
Huijbregts et al. (2006)	LCIA	Proxy	Regression
Huijbregts et al. (2008)	LCIA	Comparison	Differences
Huijbregts et al. (2010)	LCIA	Proxy	Regression, contribution analysis
Huppes et al. (2012)	LCIA	Comparison	Contribution analysis
Iswara et al. (2020)	Software	Comparison	Differences, contribution analysis
Joyce et al. (in press)	LCI	Comparison	Differences
Junnila (2006)	LCI	Comparison	Differences
Kaufman et al. (2010)	LCIA	Proxy	Regression
Kalbar et al. (2017)	LCIA	Proxy	Correlation, regression
Kounina et al. (2014)	LCIA	Comparison	Differences
Laleman et al. (2013)	LCIA	Comparison	Differences
Landis and Theis (2008)	LCIA	Comparison	Contribution analysis
Lasvaux et al. (2016)	LCIA	Reduction	PCA
Laurent et al. (2012)	LCIA	Proxy	Correlation
Lautier et al. (2010)	LCIA	Comparison	Differences

Table 2 (continued)

Reference	Topic	Purpose	Statistical techniques
Ligthart and Ansems (2019)	LCIA	Comparison	Correlation
Martínez et al. (2015)	LCIA	Comparison	Differences
Marvuglia et al. (2014)	LCIA	Streamlining	Neural network
Marvuglia et al. (2015)	LCIA	Streamlining	Neural network and other ML techniques
Masnadi et al. (2020)	LCI	Streamlining	Regression
Mendoza Beltrán et al. (2016)	LCI	Comparison	*
Monteiro and Freire (2012)	LCIA	Comparison	Differences, contribution analysis
Myllyviita et al. (2014)	LCIA	Comparison	Differences
Notarnicola et al. (1998)	LCIA	Comparison	Differences, contribution analysis
Owsianiak et al. (2014)	LCIA	Comparison	Differences
Padey et al. (2013)	LCI, LCIA	Proxy	Regression
Pant et al. (2004)	LCIA	Comparison	Contribution analysis
Park et al. (2001)	LCI, LCIA	Proxy	Regression, neural network
Park and Seo (2003)	LCI, LCIA	Proxy	Regression, neural network
Park and Seo (2006)	LCI, LCIA	Proxy	Neural network
Park et al. (2015)	LCIA	Reduction	PCA
Pascual-González et al. (2015)	LCIA	Reduction	Multiple regression
Pascual-González et al. (2016)	LCIA	Reduction	Correlation, regression
Peters (2007)	LCI	Streamlining	Differences
Pizzol et al. (2011)	LCIA	Comparison	Contribution analysis
Pozo et al. (2012)	LCIA	Reduction	PCA
Renou et al. (2008)	LCIA	Comparison	Contribution analysis
Röös et al. (2013)	LCIA	Proxy	Correlation
Schulze et al. (2001)	LCIA	Comparison	Differences, contribution analysis
Scipioni et al. (2013)	LCIA	Proxy	Contribution analysis
Shariar Hossain et al. (2014)	LCI, LCIA	Proxy	Cluster analysis
Simões et al. (2011)	LCIA	Comparison	Contribution analysis
Song et al. (2017)	LCIA	Streamlining	Regression
Sousa et al. (2000)	LCI, LCIA	Streamlining	Neural network
Speck et al. (2015)	LCI, LCIA	Comparison	Differences, contribution analysis
Steinmann et al. (2016)	LCIA	Reduction	Multiple regression, PCA
Steinmann et al. (2017a)	LCIA	Proxy	Multiple regression
Valente et al. (2018)	LCIA	Streamlining	Difference
Valente et al. (2019)	LCIA	Proxy	Correlation
Weidema (2015)	LCIA	Comparison	Contribution analysis
Wernet et al. (2008)	LCI, LCIA	Streamlining	Regression, neural network
Wernet et al. (2011)	LCI, LCIA	Streamlining	Contribution analysis
Zhang and Bakshi (2007)	LCI, LCIA	Streamlining	Regression

LCI life cycle inventory analysis, *LCIA* life cycle impact assessment, *MDS* multidimensional scaling, *PLSR* partial least squares regression, *PCA* principal component analysis, *ML* machine learning

*These authors combine method comparison with data uncertainty in a probabilistic treatment

and its normalized version:

$$\mathrm{RMSE}_n = \frac{\mathrm{RMSE}}{\bar{x}} \tag{8}$$

In these formulas, we have used Dekker's reference to Timsina and Humphreys (2006), correcting a typo. Birkved and Heijungs use a "root mean square error of prediction" (RMSEP), for which they give in their appendix C a formula that is probably wrong (e.g., it contains no root and no square). Given the general idea of a RMSE, we correct it here as follows:

$$\mathrm{RMSEP} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\hat{y}_i - y_i\right)^2} \tag{9}$$

Wernet et al. (2008) argue that RMSEP is less suitable when y varies over an order of magnitude or more, and prefer to use the mean of the absolute values of the relative prediction error:

$$\mathrm{MRE} = \frac{1}{n}\sum_{i=1}^{n}\frac{\left|\widehat{y_i} - y_i\right|}{y_i} \tag{10}$$

Interestingly, they apply this for 30 test sets, and report the mean MRE (which is therefore a mean of means), the median RME, as well as the standard deviation MRE. Despite their reservations, they do report RMSE in their Supplementary material, in a similar way.

A primitive form of statistical analysis is performed by Hochschorner and Finnveden (2003), who study the differences between x_i and y_i, defining these as "significantly better," "probably better," etc. A more sophisticated form is presented by Dekker et al. (2020), who use a "two-sided t test," for which no further details are provided. A study of their R code (supplied by the authors) reveals that the independent samples t test for equality of the mean, without assuming equality of variance, was used. The null hypothesis tested is as follows:

$$H_0 : \mu_x = \mu_y \tag{11}$$

where μ indicates the population mean, and the computational details are provided in the Supplementary Information of this article. When the p value is smaller than a pre-determined significance level (such as 5% or 1%), the test results in a "significant" result, in this case, a significant difference between the two means. Dekker et al. (2020) choose 5% for this.

Visual presentations of the difference take different forms. Dekker et al. (2020) show box plots of x and y next to each other. Huijbregts et al. (2008) also use box plots, but now of the ratio $\frac{x_i}{y_i}$. Chen et al. (2021) use scatter plots, using the horizontal axis for i, showing x and y with different colors on the vertical axis. Mendoza Beltrán et al. (2016) show box plots and partly overlapping histograms. Several authors (Huijbregts 1998; Cherubini et al. (2018) construct a "comparison indicator," which combines a comparison of methods with a stochastic treatment of the numerical data, and present a histogram of the comparison indicator.

2.5 Contribution analysis

Several studies (e.g., Junnila 2006; Bovea and Gallardo 2006; Dewulf et al. 2007; Weidema 2015) address a few products and concentrate on the contributions made by different parts, without constructing an indicator. Most studies do this in a quantitative manner, but a few papers (e.g., Brent and Hietkamp 2003) use a more qualitative approach. The results are often presented in tables and/or bar graphs; see Bueno et al. (2016) for a good example.

Contribution analysis can proceed in different ways. Monteiro and Freire (2012) split the scores by life cycle stage (materials, transport, maintenance, etc.). Pizzol et al. (2011) by contrast specify the contribution from different stressors (aluminum, antimony, etc.) to an overall impact score (human health). Weidema (2015) shows how aggregated impact categories (such as ecotoxicity) are built-up of subscores (such as freshwater, marine, and terrestrial). Halleux et al. (2006) go even further, and show how a single index is built-up in terms of endpoint impacts (resources, ecosystem quality, human health).

2.6 Measures of correlation

In meta-comparative LCA, we expect that a product with a relatively low x value will also have a relatively low y value. The degree to which the x and y values run together can be expressed in various ways. Below, we discuss different types of correlation coefficients and regression analysis. In this section, we discuss approaches using correlation, and in the next one, regression. Note that in many of the reviewed articles, the word "correlation" is used in an overall sense, also embracing regression. For instance, Bösch et al. (2007), Kaufman et al. (2010), and Berger and Finkbeiner (2011) speak of "correlations" which they determine using regression analysis.

The Pearson product-moment correlation coefficient, or correlation coefficient for short, indicated by r, is given in the Supplementary Information of this article. It measures the degree of linear correlation. If $r = 1$, there is a perfect linear positive correlation, indicating a perfect match with a straight line with positive slope b:

$$y_i = a + bx_i \tag{12}$$

with $b > 0$. If $0 > r > 1$, there is a certain scatter around this line, the closer r is to 1, the better the agreement with the straight line. A negative value of r represents a case of anti-correlation, reflecting a straight line with a negative slope. Notice that a correlation coefficient reveals little (only its sign, positive or negative) of the value of b. Correlation coefficients have been reported by, among others, Laurent et al. (2012), Röös et al. (2013), and Dong et al. (2016).

Several authors (e.g., Bösch et al. 2007; Wernet et al. 2008; Berger and Finkbeiner 2011; Valente et al. 2018) prefer to report the square of the correlation coefficient, indicated by R^2, and known as the coefficient of determination:

$$R^2 = r^2 \tag{13}$$

The reason is probably that these studies use regression analysis (see below) as a way to determine correlations. We will discuss R^2 in more detail below. Note that Huijbregts et al. (2006) refer to r^2 as the "correlation coefficient", and that Wernet et al. (2008) are somewhat vague on this.

Laurent et al. (2012), Kalbar et al. (2017), and Dekker et al. (2020) use the Spearman correlation coefficient, or rank correlation coefficient, to check for the consistency of ranking (note that Kalbar et al. (2017) use the term "nonlinear" correlation coefficient, which is a bit misleading). It is based on the Pearson correlation of the ranked variables, and sometimes indicated by ρ or r_S (see Supplementary Information). Like the Pearson correlation, the Spearman correlation is between −1 and 1; however, its interpretation is slightly different. While the Pearson correlation indicates the agreement with a straight line, the Spearman correlation indicates consistency of ranking. If for all products the rank according to x agrees with the rank according to y, we have $r_S = 1$. Otherwise, the value will be less than 1. A Spearman correlation of 1 can be interpreted as a ranking-preserving signal. If for at least one pair of products (i,j), the ranking according to x differs from the ranking according to y (for instance, $R_{x_i} < R_{x_j}$ but $R_{y_i} > R_{y_j}$) then the Spearman correlation coefficient is less than 1.

Because, as observed, several texts prefer to use the square of the Pearson correlation instead of the raw version, there are also authors who square the Spearman correlation. An example can be found in the paper by Wernet et al. (2008).

It is important to note that several textbooks supply a simplified version of the formula to calculate the Spearman correlation (see Supplementary Information). It is, however, only valid when there are no ties, i.e., when all x values and y values occur only once. We do not know which of the two formulas for calculating the Spearman correlation has been used by the meta-comparative studies that use this as an indicator.

A third type of correlation coefficient is known as Kendall's τ. For reasons of consistency, we will use the symbol r_K here; see Supplementary Information for details. We are aware of only one paper using it, namely Kalbar et al. (2017).

Correlations can be visually supported by scatter plots. Laurent et al. (2012) provide examples of such plots. Note that the straight line indicated is not a regression line, but a "45 degree" line, indicating equality. Dekker et al. (2020) show two lines: the equality line and a regression line. We discuss the regression line in more detail in the next section.

All these types of correlations can be subject to a hypothesis test, testing the null hypothesis that the population value of the correlation coefficient is 0; see the Supplementary Information for details. Röös et al. (2013) and Dong et al. (2016) test Pearson correlation coefficients, using $\alpha = 0.05$ as a criterion for significance, and Pascual-González et al. (2016) use $\alpha = 0.001$. Berger and Finkbeiner (2011) also identify "significant correlations," but do not inform the reader about their criterion for significance.

A variation is the use of confidence intervals. Laurent et al. (2012) are the only case we found where confidence intervals for correlation coefficients are used.

Kalbar et al. (2017) use the Spearman correlation with a hypothesis test, but they do not specify the precise form taken. Wernet et al. (2008) indicate the critical value of their (squared) Spearman correlation, using $\alpha = 0.01$, but they also do not indicate the precise procedure (z or t).

For Kendall's correlation coefficient, there are various forms available (see Supplementary Information). Kalbar et al. (2017) use a significance test, although precise details are not presented, except for a general reference to Matlab.

Spearman's and Kendall's correlation coefficient are examples of indicators that focus on the ranking of products according to the x- and y-scales. Rankings play also a prominent in the analysis by Heijungs (2017).

2.7 Simple regression

Closely related to, but different in a number ways from correlation, is regression analysis. Here, we restrict the discussion to simple regression, with only one x variable, which is the type of analysis used by a number of authors (e.g., Huijbregts et al. 2006; Curzons et al. 2007; Berger and Finkbeiner 2011), although in some cases a logarithmic transformation has been applied prior to the analysis (see below).

In a simple regression analysis, the data (x and y) is used to estimate:

$$y_i = a + bx_i + e_i \tag{14}$$

where a is the intercept (or constant), b is the slope (or regression coefficient), and e_i is a residual (or error) term that indicates the deviation of y_i from the regression line. With such a regression line, we predict with a given x_i :

$$\hat{y}_i = a + bx_i \tag{15}$$

which deviates from the observed value y_i by an error:

$$e_i = y_i - \hat{y}_i = y_i - \left(a + bx_i\right) \tag{16}$$

Details on the estimation procedure are in the Supplementary Information. The goodness-of-fit of a regression line is usually reported as the coefficient of determination, R^2 (see Supplementary Information), which can be interpreted as a fraction of explained variance. For instance, if $R^2 = 0.9$, the x variable is accountable for 90% of the variance in the y variable, the remaining 10% is due to random (unexplained) variation.

The standard error of the regression, also known as residual standard error (see Supplementary Information), is another measure of the goodness-of-fit. It is used by Huijbregts et al. (2006) for calculating an "uncertainty factor," k:

$$k = \frac{97.5p}{2.5p} \tag{17}$$

Although not described as such, we think that $97.5p$ refers to the 97.5 percentile of the distribution of residuals e_i, which is further assumed to be log-normally distributed. The k values have been reported in their Table 4, with values ranging between 1.2 and 42000.

Pascual-González et al. (2016) employ another measure of the quality of the fit, namely the relative error (see Supplementary Information), which they further express as a percentage. Pascual-González et al. (2015) use a variant of this, called the average relative error, which is a generalization for multiple y variables.

Birkved and Heijungs (2011) use, besides R^2, a related statistic, which is indicated by Q^2 (see Supplementary Information), and which is based on leave-one-out cross validation (LOOCV). In general, cross validation is a technique in which part of the sample (let us say, $m < n$ data points) is used to "train" the model (i.e., to estimate the coefficients), and the rest of the data (the remaining $n - m$ data points) is used to compute a goodness-of-fit measure. In LOOCV, we use $m = n - 1$, and loop over all n data points to find Q^2 that is averaged over the entire sample. For a more detailed description, we refer to James et al. (2015).

Also Wernet et al. (2008) mention the leave-one-out principle for cross validation, but they give a result (indicated as q^2) with the name "coefficient of determination," and their Supplementary Information provides a formula for q^2 which indeed looks more like the usual R^2.

Regression models can also be the subject of a significance test, in several ways. For a simple regression, this can take the form a t test or an F test, both of which test the null hypothesis H_0: $\beta = 0$, and which give identical p values. The details are in the Supplementary Information. As far as we know, this test has not been carried out before in the context of meta-comparative LCA. Although Zhang and Bakshi (2007) write that "statistical regression and hypothesis testing is used to determine whether a statistical correlation exists," they do not report t or F or p values.

Like with the correlation coefficient, the appropriateness of a two-tailed test can be doubted. More fundamentally, if y is supposed to mimic x, a test for a unit regression coefficient seems even more appropriate. Such a test could take the form:

$$H_0: \beta = 1 \tag{18}$$

which looks more like the unit root test of time series econometrics (Gujarati 2003; Hill et al. 2011).

Regression analyses (and correlations; see 2.6) are often supported by scatter plots, one variable showing the x values and the y variable at the other axis. In many cases, the regression line ($y = a + bx$) is shown in addition. Examples can be found in, among others, Berger and Finkbeiner (2011) and Curzons et al. (2007). It is sometimes due to the presence of such regression lines that it becomes clear that the authors indeed apply regression, while their paper uses the term correlation (see, for instance, Kaufman et al. 2010). The paper by Berger and Finkbeiner (2011) is further a good piece of evidence in showing the extent to what extent correlation and regression can be mixed up, using phrases like "strong linear regressions" in a paper which has just "correlation analysis" in the title.

2.8 Multivariate analyses

Correlation and simple regression are bivariate techniques, investigating the relationship between two variables, indicated here with x and y. Several extensions have been employed for meta-comparative LCA.

Steinmann et al. (2017a) use multiple regression. Although they do not specify the precise details, we can make some educated guesses here. The regression model in this case is as follows:

$$y_i = a + b_1 x_{i1} + b_2 x_{i2} + \cdots + b_k x_{ik} + e_i \tag{19}$$

with k slope coefficients $b_1, b_2, \ldots, b_k$. There exist standard matrix-based techniques to find the optimal values of these coefficients, as well as their standard errors. Multiple regression also yields R^2 values. The slope coefficients have no interpretation of a correlation, but their difference from 0 can still be tested with a t test. As these b-coefficients have different units and scales, they cannot be compared with each other. One way to allow for such comparisons is by transforming them into standardized regression coefficients (see Supplementary Information). Steinmann et al. (2017a) indeed use such standardized regression coefficients to express the relative importance of the different x variables in contributing to y.

Multiple regression is also used by Park et al. (2001) and Park and Seo (2003). These authors also report significance tests on the basis of the F-statistic (see Supplementary Information).

The multiple regression model requires that the x variables are independent. One way to test for dependence among the x variables is through variance inflation factors (VIFs; see Supplementary Information). All VIFs should be 1 for full independence, although values up to 5 can be argued to be still reasonable. Steinmann et al. (2017a) use the VIF to remove redundant x variables.

Another approach to study mutual dependence and to control for redundance is by principal component analysis (PCA). The purpose of a PCA differs in an important way, as it does not predict a y from one or more x variables, but rather studies the degree to which different x

variables provide added value. Examples of such studies include Le Téno (1999), Curzons et al. (2007), Gutiérrez et al. (2010a), Pozo et al. (2012), Steinmann et al. (2016), Lasvaux et al. (2016), and Balugani et al. (2021). Because their aim is not to compare but to reduce, we will exclude those studies from our analysis. However, we will discuss one aspect, because it resembles the previous discussions. The PCA technique proposes a rotated, orthogonal, coordinate system, in which the first principal components (PCs) describe a large fraction of the variance. For instance, Steinmann et al. (2016) show a scree plot in which the first PC explains 83.3% of the variance, and the second PC adds another 3.1%. Such numbers can be interpreted similar to the R^2 of a regression, and can therefore suggest to support the idea of a proxy indicator. However, the PCs are themselves weighted combinations of the original x variables, and therefore even a proxy by only one PC needs in general information from all x variables.

A useful distinction of ways of analysis has been made by Cattell (1952). For our purpose, we restrict the discussion to Q and R techniques (also: Q and R analyses; Legendre and Legendre 1998):

- the Q technique addresses similarities between "objects" (products), for instance to find out which products are comparable; and
- the R technique addresses similarities between "descriptors" (variables), for instance to reduce the number of impact categories.

The PCA studies mentioned are examples of R analyses. There are also a few meta-comparative LCA studies that use Q techniques. For instance, Gutiérrez et al. (2009) use multidimensional scaling (MDS), and Gutiérrez et al. (2010a) use cluster analysis to group similar products. We do not further discuss these Q techniques, because their aim falls outside the scope of this article.

Several more advanced variations on regression analysis have been used. We mention Birkved and Heijungs (2011), who use partial least squares regression (PLSR), which is a multivariate technique that is based on the combination of PCA and regression. We also mention Balugani et al. (2021), who use robust ordinal regression, a technique that focuses on ordinal rankings instead of the numerical values. Pascual-González et al. (2015) combine multiple regression and mixed integer linear programming (MILP). Eddy et al. (2015) apply kriging, which can also be regarded as a variation to regression. The advanced nature of these methods, combined with their only occasional use, forces us to keep these further undiscussed.

In the context of multiple regression, Steinmann et al. (2017a) use the Akaike information criterion (AIC) to assess the goodness-of-fit. AIC, like R^2, is a measure of the quality of the model, but it penalizes the use of an excessive number of x variables. In that respect, it resembles the more familiar adjusted R^2, R^2_{adj}. Both AIC and adjusted R^2 are described in the Supplementary Information. A difference is that R^2, and by extension R^2_{adj}, has a stand-alone interpretation, while AIC makes only sense in a comparison of regression models.

Pascual-González et al. (2016) investigate the correlation between multiple x variables, defining a "correlation index" as the relative number of variables correlated with a specific variable. We interpret this in our notation as follows:

$$I_l = \frac{1}{k} \sum_{\substack{j=1 \\ j \neq l}}^{k} \Theta\left(p\left(r_{jl}\right) - 0.001\right) \tag{20}$$

where $p\left(r_{jl}\right)$ is the p value associated with the correlation coefficient of variables x_j and x_l and $\Theta(x)$ is the Heaviside step function.

Finally, Kalbar et al. (2017) mention the use of partial correlation coefficients (see Supplementary Information). The result $r_{12\cdot3}$ measures the correlation between variables 1 and 2, corrected for a confounding variable 3 that is correlated with both 1 and 2.

2.9 Machine learning techniques

Modern developments in machine learning and artificial intelligence have enriched the toolbox of predictions by computationally intensive techniques. Here we briefly describe a few approaches that have been used in the context of meta-comparative LCA, without providing the full details.

Several authors have used neural networks (also called artificial neural networks, ANN) to establish relationships between predictors and results. Marvuglia et al. (2015) do this for the relation between chemical properties and characterization factors, and Park and Seo (2006) use ANN to streamline the design of products using simple product characteristics, such as mass, percentage of plastics, and lifetime. A similar approach is taken by Sousa et al. (2000).

Wernet et al. (2008) and Park and Seo (2003) use both regression analysis and neural networks, and can therefore be interpreted as a meta-meta-comparative LCA.

Shariar Hossain et al. (2014) use several clustering techniques. Such analyses can be interpreted as Q-mode analysis (see previous section), and therefore can be seen as answering a different type of question.

Also Hou et al. (2020) apply a number of ML techniques, ranging from neural networks to nearest neighbor methods.

3 Critical discussion

The previous section gave a neutral overview of the different indicators that have been used for meta-comparative LCA. It is clear that there is a tremendous choice of methods: differences, regression, correlation, neural networks, t tests, and p values. Further choices, such as the use of logarithmic transformations, complicate the situation even more. In this section, we will add a few critical discussions.

3.1 Lack of detail, inconsistencies, and other issues

We start our critical section by pointing out that many of the cited papers are incomplete, unclear, or contain mistakes. This is a pity, because the approaches are often interesting, but they are insufficiently clearly described to reproduce, and therefore it is difficult to come to a full appreciation or adoption in software. Here we give a few examples, without claiming to be complete.

Pascual-González et al. (2016) show several figures with an "Ω" on the axis, without defining its meaning in the text. They also use the bar-symbol ($\bar{x}$ and $\bar{y}$) for the mean in their Eq. (1), while in their Eq. (4) the mean is μ. Dekker et al. (2020) use a "two-sided t test," but do not specify details like paired vs. independent-samples, or equal vs. unequal variance.

If p values are reported, the null hypothesis is hardly ever mentioned. The paper by Röös et al. (2013) is one of the few who actually does report it, but many other papers just list p values. When p values are used to decide if something is "significant," the significance level (α) is often not mentioned. A proper use of null hypothesis tests further necessitates the distinction between population parameters (such as ρ and β) and sample statistics (such as r and b). Such refinements are almost completely lacking.

We already commented on the confused use of the terms "correlation" and "regression". A simple regression analysis yields an R^2 and a Pearson correlation analysis an r, which are trivially related. But for multiple regression and Spearman correlation, the situation becomes harder. Despite their similarities, the two types of analysis are fundamentally different. Correlation is about the moving together of two variables, without any assumption of causality or priority. Regression, by contrast, assumes that one variable (y) depends on another variable (x), implying a causal structure. Regression also assumes that the x variable is not random and without error, while the y has random error. Correlation assumes that both x and y (or perhaps more appropriately written, x_1 and x_2) are both random. The difference between correlation and regression has repercussions for their applicability. Comparisons of methods (like Dekker et al. 2020) may benefit from a correlation analysis, while for streamlining and proxy studies (like Huijbregts et al. 2006), regression is more appropriate. A clear definition of the approach used is therefore a requirement for a correct judgment of the quality of the studies.

3.2 Measures of difference

The (undocumented) choice by Dekker et al. (2020) for an independent samples t test instead of a paired t test is remarkable, because there is a natural pairing of an x value and a y value for every product i. A paired t test first defines the following:

$$d_i = x_i - y_i \tag{21}$$

and then constructs the test statistic:

$$t = \frac{\bar{d}}{s_d/\sqrt{n}} \tag{22}$$

which is tested with $df = n - 1$ and yields in general a much smaller p value than an independent samples t test.

Both the independent samples t test and its paired version effectively result in a t value which scale with $\sqrt{n}$, and as such can result in highly significant differences, even when these are small, given a large sample size n. And because sample size can be increased arbitrarily by including more products in the test set, such significant measures in the end have little meaning (Heijungs et al. 2016).

3.3 Measures of association

In our discussion of correlation, we found quite a few of papers that report p values for correlation coefficients. Traditionally, a two-tailed test is used, but in fact, it makes sense to consider applying a one-tailed test, because the hypothesis can be argued to be about a positive correlation:

$$H_0\colon\ \rho \leq 0 \text{ versus } H_1\colon\ \rho > 0 \tag{23}$$

Note that the number of tails, and therefore the directionality of the null hypothesis, is not always mentioned by the cited articles.

Like with the t test for equality of means discussed, a test for zero correlation will tend to suggest a rejected null hypothesis when the sample size is large (Heijungs et al. 2016). The interpretation of such a rejected null hypothesis is that there is evidence of some relation between x and y, but in general, a significant result does not at all imply that the relation is strong. In that respect, a better practice is the one by Huijbregts et al. (2006) and Bösch et al. (2007), who put the emphasis on high values of R^2, even though no significance test is performed. Alternatively, we might propose to test for a correlation of 1, instead of 0:

$$H_0\colon\ \rho = 1 \tag{24}$$

The tests mentioned so far are only applicable to test for a zero correlation. The Fisher transformation allows for a conversion of r into another variable, r':

$$r' = \frac{1}{2}\ln\left(\frac{1+r}{1-r}\right) \tag{25}$$

This transformation is applied to the observed correlation coefficient and to the hypothesized correlation, ρ'. However, the Fisher transform is undefined for $\rho = 1$, so this procedure is of no help.

A much more useful test would be as follows:

$$H_0\colon \beta = 1 \tag{26}$$

where β is the population value of the regression slope coefficient. Where in a simple regression, the quantity $t_b = \frac{b}{SE_b}$ is used to assess the null hypothesis $H_0\colon \beta = 0$, we use the following:

$$t = \frac{b-1}{SE_b} \tag{27}$$

to asses this modified null hypothesis. We have not identified any study which applied this hypothesis test.

3.4 The use of statistical theory

Any data set can be used to compute a mean and a standard deviation, and any paired data set can be used to compute correlation coefficients and a regression line of best fit. Further analysis typically involves distribution theory, which poses several requirements:

- the underlying data generating process must satisfy certain characteristics (e.g., it must be normally distributed), or the sample must be sufficiently large to allow for an asymptotic result (e.g., n must be larger than 30); and
- the analyzed sample must be a random sample.

If these conditions are not satisfied, several results (in particular standard errors, t, F and p values, and results of significance tests) are not reliable.

Similar remarks can be made for other types of indicators that rely on distribution theory, including standard errors, confidence intervals, variance inflation factors, and the Akaike information criterion.

Heijungs (2017) commented on the unjustified use of distribution theory by Steinmann et al. (2017a), after which Steinmann et al. (2017b) analyzed their case in more detail. This remains, however, an exception, and the use of t, F, and p values in quasi-empirical meta-comparative LCA should be interpreted with caution.

3.5 Issues of scale and units

All quasi-empirical studies choose a certain unit of product as the basis of the x and y scores. For instance, Röös et al. (2013) calculate $n = 53$ sets of scores, each on the basis of 1 kg of product. Huijbregts et al. (2006) have a more mixed portfolio: these authors calculated results for 226 energy products per MJ, 750 materials per kg, etc. Such choices are pretty arbitrary. Because LCA results scale linearly with the amount of product, we would hope that the results of the meta-comparison are insensitive to the exact numerical choice. Would the results of Röös et al. (2013) change if they would choose 100 kg of product? And more subtly, would the results by Huijbregts et al. (2006) change if we would continue to use 1 kg for the materials, but switch to kWh for the energy products?

Clearly, there is no universal answer to this question. Some results will depend on a change of scale or units, but that does not mean that the final conclusion will change. A further complication is that the effect of a change of scale or unit will always affect the y variable (because it reflects the emission or impact per unit of product), but not always the x variable. For instance, in streamlining studies or comparisons, the x variable will also depend on the scale and unit of x. But for proxy studies, the situation may be different. Consider, for instance the case of a regression model:

$$y = a + b_1 x_1 + b_2 x_2 + e \tag{28}$$

where y is the carbon footprint, x_1 is the mass of the product, and x_2 the lifetime. y and x_1 are sensitive to changes of units and scale, but x_2 is not, and the precise way this affects a, b_1, and b_2 are not a priori clear. To facilitate our analysis, we will focus on situations in which both x and y (or all x and y variables) depend on the scale and unit in the same way.

Suppose we change for some of the products the LCA basis from 1 unit to k units. For instance, we change from 1 MJ to 1 GJ, so $k = 1000$. Or from 1 MJ to 1 kWh, so $k = 3.6$. For these products, we find:

$$\begin{cases} x_i' = kx_i \\ y_i' = ky_i \end{cases} \tag{29}$$

For this subset of products, we then also find that the difference between x and y scales with k:

$$d_i' = x_i' - y_i' = kx_i - ky_i = k\left(x_i - y_i\right) = kd_i \tag{30}$$

but the relative difference is not affected:

$$\delta_i' = \frac{x_i' - y_i'}{x_i'} = \delta_i \tag{31}$$

The overall scores, such as RMSE and correlation and regression coefficients, are more complicated to analyze. But Fig. 1 gives an illustration of the effect of rescaling one data point by a factor of 5, keeping all other points at their original position. The effect can be quite large, because rescaling can create or annihilate outliers at will. The figure shows that a neatly behaving data point can become an outlier through an essentially arbitrary change of scale or unit. When we inflate this point by a factor of 100, R^2 even becomes 0.9999, suggesting an extremely good approximation. As a concrete example of such inflation, we point to a database that contain LCA data for potatoes (in tonnes) as well as for potato harvesters (in units). These two are perhaps comparable in terms of impacts. But if we would deflate the potatoes to the scale of kg or even g, the harvester would suddenly turn into an outlier.

3.6 The use of logarithms

Some authors use logarithmically transformed data, at least in part. In several cases, the graphs have logarithmic axes, but it is often not clear if the statistical analyses (correlation coefficients, etc.) are based on the raw data or on their logarithms. For instance, Bösch et al. (2007) write in the caption of their figures "logarithmic scales," but they do not specify if their R^2 values are based on logarithmic scales as well. Similar remarks apply to Laurent et al. (2012) and Dekker et al. (2020). Huijbregts et al. (2008) are more explicit: they indicate that "the data… were log-transformed," and they provide a regression equation of the form which in our notation amounts to the following:

$$\hat{y} = 10^a x^b \tag{32}$$

The use of logarithms can create further issues. We mention the following:

- the base of the logarithms (10, e, etc.) is not always stated; and
- the terminology can be confusing.

Strictly speaking, the use of logarithms requires a specification of the base. But the precise choice matters little, because for any $b, x > 0$:

$$\log_b x = \frac{\ln(x)}{\ln(b)} \tag{33}$$

so that a change of logarithmic base leads to a change by a factor of $\frac{1}{\ln(b)}$, which has a similar interpretation as a change of unit.

For an example of a confusing terminology, we refer to Huijbregts et al. (2006), who applied "log-linear regression analysis," which might (see Hill et al. 2011) suggest a model of the type:

$$\log\left(y_i\right) = a + bx_i + e_i \tag{34}$$

However, their Fig. 1 contains equations of the type:

$$\log\left(y_i\right) = a + b\log\left(x_i\right) + e_i \tag{35}$$

which might be regarded as representing a log–log regression, and the horizontal and vertical axes are indeed both logarithmic. To increase the confusion, the caption of the figure mentions a "linear regression," which might suggest:

$$y_i = a + bx_i + e_i \tag{36}$$

Zhang and Bakshi (2007) are clearer in writing about a "linear regression of log transformed data," and they moreover provide a formula like:

$$\log\left(y_i\right) = a + b\ \log\left(x_i\right) + e_i \tag{37}$$

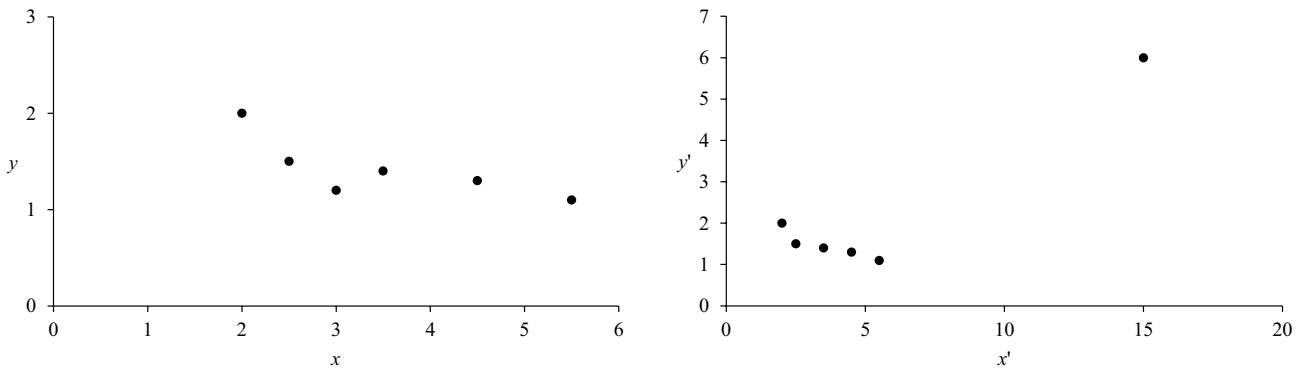

Fig. 1 Six data points with $R^2 = 0.59$ (and $r < 0$) (left) and the same data with one data point rescaled with $x' = 5x$ and $y' = 5y$, yielding $R^2 = 0.83$ (and $r > 0$)

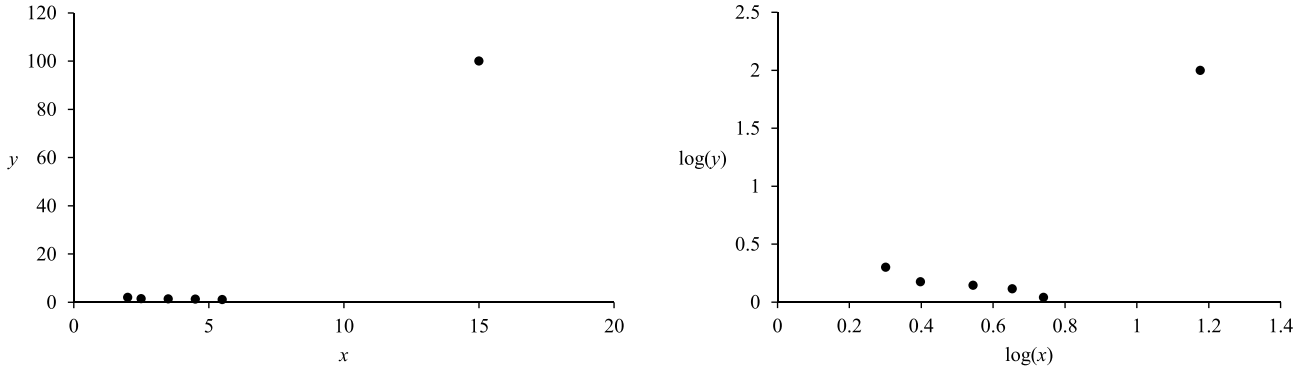

Fig. 2 Six data points with $R^2 = 0.93$ on a normal scale (left) and the same data on a logarithmic scale with $R^2 = 0.63$

But these authors show a mix of graphs: linear–linear (their Fig. 1), log–log (their Fig. 2) and linear-log (their Fig. 3), with the result that the details of their analysis are still confusing.

Part of the confusion is inherent in the terminology. Gujarati (2003) defines a number of terms in this respect, including log-linear, log–log, double-log, semilog, log-lin, and lin-log. But because other texts (e.g., Hill et al. 2011) use deviating terms, words alone cannot suffice, and specifying the relationship (as done by Zhang and Bakshi (2007) and Huijbregts et al. (2006)) seems imperative.

An import question is to what extent logarithms affect the results of the analysis. Obviously, logarithms can render a graph more convincingly. But they change some of the numerical indicators as well. In particular, outliers can make a large difference. Figure 2 gives an illustration of this phenomenon.

Reasons to use logarithms vary also. Huijbregts et al. (2006) introduce a logarithm "to account for [the] skewed distributions" and Steinmann et al. (2017a) do so "because the footprints varied up to 10 orders of magnitude". Bösch et al. just mention "logarithmic scales", without any reason.

The use of logarithms is in any case problematic in case negative or zero values of x or y occur. Laurent et al. (2012) mention this in their Supplementary Information and discard these data points. Probably, most impact scores are non-negative, but zeros certainly can occur, and also negative values may show up, for instance as an artifact of allocation.

3.7 The intercept of a regression line

The default linear regression model is based on the equation:

$$\hat{y} = a + bx \tag{38}$$

where a is the intercept and b the slope. This idea contradicts one of the basic principles of LCA, namely the proportionality of LCA results (emissions, impacts, etc.) with the quantity of product that is expressed by the functional unit. If the quantity of product is z, it follows that the LCA result x is given by the following:

$$x = pz \tag{39}$$

and that another LCA result y is given by the following:

$$y = qz \tag{40}$$

where p and q are the per-unit impacts of the product on variable x and y, respectively. As a consequence:

$$y = \frac{q}{p}x \tag{41}$$

which amounts to the following:

$$a = 0 \text{ and } b = \frac{q}{p} \tag{42}$$

Many quasi-empirical meta-comparative LCA studies use a sample of x and y results to estimate not only the slope, b, but they also estimate the intercept, a. For instance, Berger and Finkbeiner (2011) report $y = 1000000x + 256.16$, and Huijbregts et al. (2010) find $\log(\text{EF}) = 0.9\log(\text{CED}) - 0.6$ (notations slightly adapted).

For the linear case, the intercept should be 0. But a standard regression analysis estimates the intercept on the basis of the data. It is, however, possible to force the intercept to zero (see Supplementary Information). We have not identified regression studies that use a zero-intercept regression line for comparative LCA. That is remarkable, because there is quite unanimous agreement that LCA results scale proportionally (Heijungs 2020). The reason is probably that the default regression analysis includes the estimation of an intercept, and that turning off this feature requires a deliberate action.

For the logarithmic case, the situation is a bit more complicated. If $y = \frac{q}{p}x$, we have the following:

$$\log(y) = \log\left(\frac{q}{p}x\right) = \log\left(\frac{q}{p}\right) + \log(x) \tag{43}$$

In a log–log regression with the following:

$$\log\left(y_i\right) = a + b\ \log\left(\ x_i\right) + e_i \tag{44}$$

this would mean that a is to be estimated while $b = 1$ is given. Again, this type of analysis has not been found in our sample of studies.

The above critique was based on simple regression, but it also holds for the case of multiple regression.

3.8 The least-squares principle

Although the regression line $y = bx + e$ makes much more sense than the traditional $y = a + bx + e$, the estimation of b is still problematic. The reason is that the usual procedures rely on a least-squares principle, minimizing the following:

$$\sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} \left(y_i - \hat{y}_i\right)^2 \tag{45}$$

The fact that LCA results (x_i and y_i) can be moved with an arbitrary scale and unit creates a degree of arbitrariness in the optimal value of b. As an example, we refer to Fig. 2, in which the left panel yields $b = 0.34$ and the right panel $b = 0.38$, the only different being a shift of one data point by a factor of 5 in both x and y. Points at the far end will tend to dominate the sum of squares, and there is no unique way to define the scales.

4 A new approach

In this section, we introduce a novel approach for meta-comparative LCA. We also demonstrate its use on a real-world dataset.

4.1 Directional statistics

Summarizing the results so far, we postulate a proportional relationship between x and y of the form $y = bx$, and we also acknowledge that a sampled product i with coordinates $\left(x_i, y_i\right)$ might have been rescaled as $\left(kx_i, ky_i\right)$. For the first reason, a regression model of the form $y = a + bx$ is inappropriate, as the intercept a must be 0. For the second reason, a least-squares regression is inappropriate, as the sum of squares $\sum_{i=1}^{n} \left(\hat{y}_i - y_i\right)^2$ depends on the arbitrary rescaling of individual data points.

To overcome these problems, we propose an entirely different approach. The relation $y = bx$ with scalable x and y can be rewritten in a scale-independent form as follows:

$$b = \frac{y}{x} \tag{46}$$

Given a sample of data points $\left(x_i, y_i\right)$, we then analyze the sample of slope coefficients:

$$b_i = \frac{y_i}{x_i} \tag{47}$$

Changes of scale and unit of individual data points do not affect such ratios, because it trivially follows that

$$\frac{ky_i}{kx_i} = b_i \tag{48}$$

as well.

So, the main question is then: how to average a sample of b_i values? For this, we conceive these values as representing an angle θ_i, given by the following:

$$\theta_i = \arctan\left(b_i\right) \tag{49}$$

and turn to the field that is alternatively called directional statistics (Mardia and Jupp 2000; Ley and Verdebout 2017) and circular statistics (Batschelet 1981; Jammalamadaka and SenGupta 2001; Pewsey et al. 2013). For an accessible summary review, see Lee (2010). In the Supplementary Information, we review the basic concepts of directional statistics, and focus below on its application to meta-comparative LCA.

4.2 Example application

We reprocessed the dataset that was used by Dekker et al. (2020), which consists of the scores of $n = 154$ food products on different impact categories according to ReCiPe 2008 (x) and ReCiPe 2016 (y). The resulting directional statistics for the two endpoint indicators, human health and ecosystems, for the three perspectives (individualist, hierarchist, and egalitarian) are shown in Table 3.

Figure 3 shows how the data points are concentrated or dispersed over the unit circle. It also includes several of the descriptive statistics of Table 3.

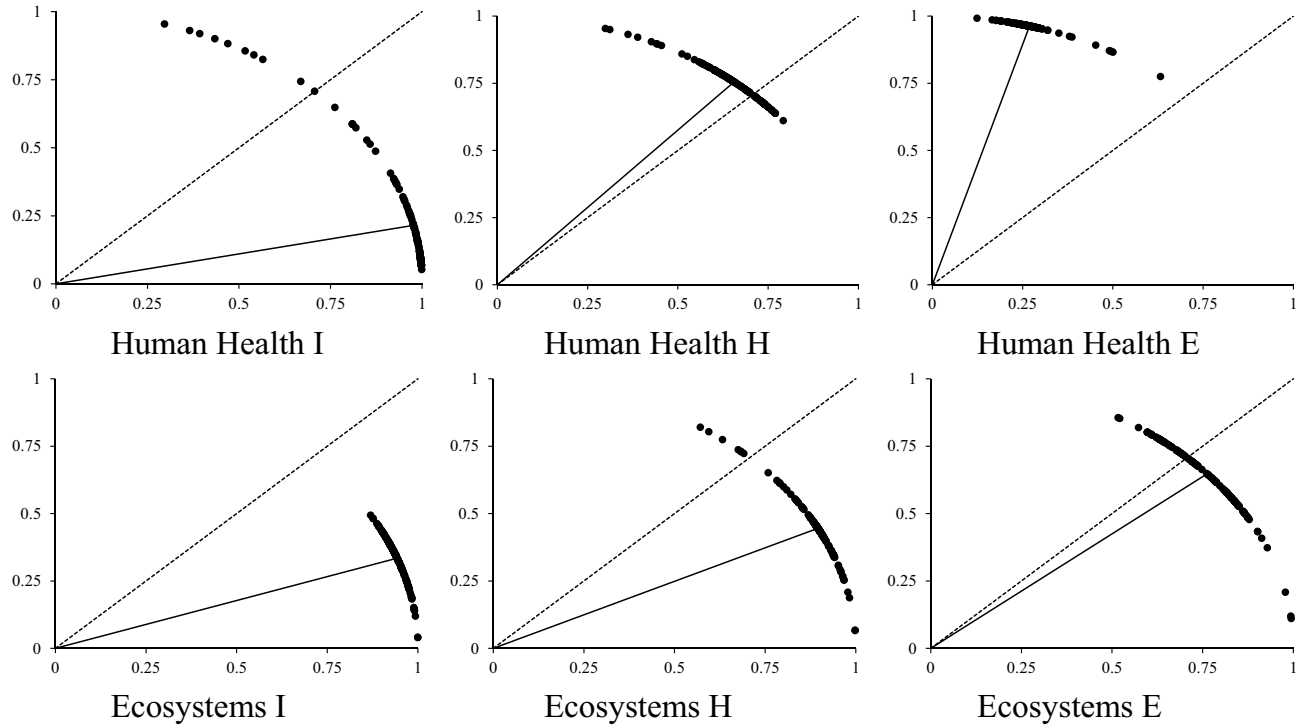

Fig. 3 Plots of the comparison of ReCiPe 2008 (horizontally) and ReCiPe 2016 (vertically) endpoints. The angle of the solid line indicates the mean slope, and its length the mean resultant length. The dashed line indicates the $y = x$ line. I = individualist perspective; H = hierarchist perspective; E = egalitarian perspective

These results should be compared with the (logarithmic) regressions by Dekker et al. (2020) (their Fig. 1). For instance, for human health, hierarchist perspective, Dekker et al. (2020) reported "no significant differences," which we can understand to mean that the dashed line is nearby the solid line. For the individualist perspective, the 2016 data (y) was "significantly smaller" than the 2008 data (x), which is confirmed by a solid line which is much flatter than the dashed diagonal. But the figures reveal much more, because the dispersion of data points over the unit circle is in some cases (e.g., human health, individualist) much larger than in other cases (e.g., ecosystems, individualist). Let us take an in-between case, human health, hierarchist. The R^2 of a linear regression is 0.98, for a logarithmic

Table 3 Results of using directional statistics on a test set of 154 food products (see also Fig. 3)

	Human health			Ecosystems		
Statistic	**I**	**H**	**E**	**I**	**H**	**E**
Circular mean* ($\overline{\theta}$)	0.217	0.856	1.298	0.341	0.461	0.704
Slope ($\tan\overline{\theta}$)	0.220	1.152	3.579	0.355	0.497	0.849
Mean resultant length ($\overline{R}$)	0.968	0.994	0.998	0.996	0.992	0.987
Circular variance (V)	0.032	0.006	0.002	0.004	0.008	0.013
Circular standard deviation (v)	0.256	0.107	0.065	0.090	0.129	0.162

*In radians (the $y = x$ diagonal in Fig. 3 corresponds to $\frac{\pi}{4} \approx 0.785$ radians)

I individualist perspective, H hierarchist perspective, E egalitarian perspective

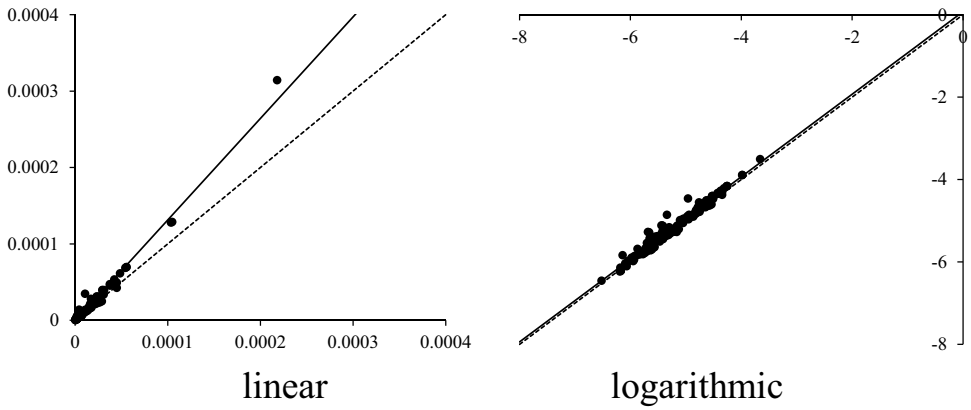

Fig. 4 Linear and logarithmic plots of the human health, hierarchist comparison. The solid line is the regression line, the dashed line the $y = x$ line

regression it is 0.97 (see Fig. 4). However, the directional plot of Fig. 3 shows a much more diverse picture, with a much larger variation than both R^2 values suggest, and in the logarithmic case, a much larger deviation between the solid and the dashed line.

4.3 Applicability and extensions

Directional statistics offers a method for meta-comparative LCA that is closer to the principles of LCA, in particular the arbitrary size, scale, and unit of the functional unit. It is also insensitive to the huge range of variation that is seen when we analyze a large number of very different products. But it is not a panacea to all problems in meta-comparative LCA.

In Table 1, we discerned five purposes:

- streamlining;
- proxy;
- reduction;
- comparison; and.
- sensitivity.

Some of these will, we believe, benefit from the use of directional statistics, while for others, the robustness of ranking, using for instance Spearman's or Kendall's correlation coefficient, will be more suitable. In Table 4, we present our ideas in this respect. We emphasize that these ideas are seminal and sometimes speculative. The field of meta-comparative LCA is, from a methodological side, still underexplored, and the present article should be considered as a first step.

Table 4 Proposed differentiated use of statistical techniques per purpose

Purpose	Main statistical analysis
Streamlining	Mix of directional statistics and regression analysis, probably extended with prediction intervals
Proxy	Directional statistics, probably extended with prediction intervals
Reduction	Principal component analysis, cluster analysis, etc
Comparison	Directional statistics, rank correlation
Sensitivity	Rank correlation

5 Conclusion

We have seen that using regression analysis, either linear or logarithmic, is incompatible with a basic axiom of LCA (namely that the impact of k units of product is equal to k times the impact of 1 unit of product), and it is vulnerable to arbitrary choices (namely of the unit and scale of the training set). In analyzing the cause of these problems, we have seen that fitting a best line through a number of data points introduces an unwanted dependence on scale and unit. By moving from a regression line to directional statistics, the deficits of the regression approach are resolved.

In other words, we have found a powerful recipe to compare methods for LCA on the basis of quasi-empirical sample of data:

- find for every product i the score on both methods (x_i and y_i);
- construct the average direction ($\tan\left(\overline{\theta}\right)$) according to the formulas in the Supplementary information (based on directional statistics);
- for comparisons and streamlining: assess if $\tan\left(\overline{\theta}\right)$ is close enough to 1; and.
- for proxies and streamlining, use $\hat{y} = \tan\left(\overline{\theta}\right)x$ to predict the y score from the x score.

A classical regression model returns, besides the estimates of the coefficients, supplementary statistics, such as the standard error of the estimates, R^2, and the AIC. For some of these, analogous concepts have been developed in the theory of directional statistics (Mardia and Jupp 2000; Jammalamadaka and SenGupta 2001). However, as discussed by Heijungs (2017), such statistics should be used with care, because quasi-empirical comparisons in LCA are typically not based on random samples.

In a completely different context, namely the comparison of methods for clinical measurements, others have observed that "the correct statistical approach is not obvious" and that popular methods like correlation and regression are inappropriate (Bland and Altman 2010). In fact, that critique went back as far as 1981, when Altman and Bland (1983) described the comparison of means, correlation, and regression as "incorrect methods of analysis." Because their topic markedly differs from ours, we cannot blindly copy the recommendations by these authors, but clearly, the comparison of methods has a wider history than merely LCA.

The subject of meta-comparative LCA is important, as is shown by our list of almost 100 articles. But the method to do meta-comparative LCA is underexplored, and deserves a more thorough investigation than one article can offer.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1007/s11367-022-02075-4.

Acknowledgements The reviewers provided very helpful comments.

Data availability All data generated or analyzed during this study are included in this published article and its supplementary information file.

Declarations

Conflict of interest The authors declare no competing interests.

References

Agresti A (2002) Categorical data analysis. Second edition. Wiley-Interscience

Agresti A, Franklin C (2013) Statistics. The art and science of learning from data. Third edition. Pearson

Altman DG, Bland JM (1983) Measurement in medicine. The analysis of method comparison studies. J Royal Stat Soc: Series D (The Statistician) 32:307–317. https://doi.org/10.2307/2987937

Amani P, Schiefer G (2011) Review on suitability of available LCIA methodologies for assessing environmental impact of the food sector. Int J Food Sys Dyn 2:194–206. https://doi.org/10.18461/ijfsd.v2i2.228

Balugani E, Lolli F, Pini M, Ferrari AM, Neri P, Gamberini R, Rimini B (2021) Dimensionality reduced robust ordinal regression applied to life cycle assessment. Exp Syst Appl 178:115021. https://doi.org/10.1016/j.eswa.2021.115021

Batschelet E (1981) Circular statistics in biology. Academic Press

Baumann H, Rydberg T (1994) Life cycle assessment. A comparison of three methods for impact analysis and evaluation. J Cleaner Prod 2:13–20. https://doi.org/10.1016/0959-6526(94)90020-5

Berger M, Finkbeiner M (2011) Correlation analysis of life cycle impact assessment indicators measuring resource use. Int J Life Cycle Assess 16:74–81. https://doi.org/10.1007/s11367-010-0237-7

Birkved M, Heijungs R (2011) Simplified fate modeling in respect to ecotoxicological and human toxicological characterisation of emissions of chemical compounds. Int J Life Cycle Assess 16:739–747. https://doi.org/10.1007/s11367-011-0281-y

Bland JM, Altman DG (2010) Statistical methods for assessing agreement between two methods of clinical measurement. Int J Nursing Stud 47:931–936. https://doi.org/10.1016/j.ijnurstu.2009.10.001

Bösch ME, Hellweg S, Huijbregts MAJ, Frischknecht R (2007) Applying cumulative exergy demand (CExD) indicators to the ecoinvent database. Int J Life Cycle Assess 12:181–190. https://doi.org/10.1065/lca2006.11.282

Bovea MD, Gallardo A (2006) The influence of impact assessment methods on materials selection for eco-design. Mat Design 27:209–215. https://doi.org/10.1016/j.matdes.2004.10.015

Brandão M, Heath G, Cooper J (2012) What can meta-analyses tell us about the reliability of life cycle assessment for decision support? J Ind Ecol 16:S3–S7. https://doi.org/10.1111/j.1530-9290.2012.00477.x

Brent AC, Hietkamp S (2003) Comparative evaluation of life cycle impact assessment methods with a South African case study. Int J Life Cycle Assess 8:27–38. https://doi.org/10.1007/BF02978746

Bueno C, Hauschild MZ, Rossignolo JA, Ometto AR, Mendes NC (2016) Sensitivity analysis of the use of life cycle impact assessment methods. A case study on building materials. J Cleaner Prod 112:2208–2220. https://doi.org/10.1016/j.jclepro.2015.10.006

Cattell RB (1952) The three basic factor-analytic research designs. Their Interrelations and Derivatives Psych Bull 49:499–520. https://doi.org/10.1037/h0054245

Cavalett O, Ferreira Chagas M, Seabra JEA, Bonomi A (2013) Comparative LCA of ethanol versus gasoline in Brazil using different LCIA methods. Int J Life Cycle Assess 18:647–658. https://doi.org/10.1007/s11367-012-0465-0

Chen X, Matthews HS, Griffin WM (2021) Uncertainty caused by life cycle impact assessment methods: case studies in process-based LCI databases. Res Cons Rec 172:105678. https://doi.org/10.1016/j.resconrec.2021.105678

Cherubini E, Franco D, Zanghelini GM, Soares SR (2018) Uncertainty in LCA case study due to allocation approaches and life cycle impact assessment methods. Int J Life Cycle Assess 23:2055–2070. https://doi.org/10.1007/s11367-017-1432-6

Crawford RH (2008) Validation of a hybrid life-cycle inventory analysis method. J Env Man 88:496–506. https://doi.org/10.1016/j.jenvman.2007.03.024

Crawford RH, Bontinck PA, Stephan A, Wiedmann T, Yu M (2018) Hybrid life cycle inventory methods. A Review J Cleaner Prod 172:1273–1288. https://doi.org/10.1016/j.jclepro.2017.10.176

Curran MA (2007) Studying the effect on system preference by varying coproduct allocation in creating life-cycle inventory. Env Sci Technol 41:7145–7151. https://doi.org/10.1021/es070033f

Curzons AD, Jiménez-González C, Duncan AL, Constable DJC, Cunningham VL (2007) Fast life cycle assessment of synthetic chemistry (FLASC™) tool. Int J Life Cycle Assess 12:272–280. https://doi.org/10.1065/lca2007.03.315

Dekker E, Zijp MC, van de Kamp ME, Temme EHM, van Zelm R (2020) A taste of the new ReCiPe for life cycle assessment. Consequences of the updated impact assessment method on food product LCAs. Int J Life Cycle Assess 25:2315–2324. https://doi.org/10.1007/s11367-019-01653-3

De Rosa M, Pizzol M, Schmidt J (2018) How methodological choices affect LCA climate impact results. The case of structural timber. Int J Life Cycle Assess 23:147–158. https://doi.org/10.1007/s11367-017-1312-0

Dewulf J, Bösch ME, de Meester B, van der Vorst G, van Langenhove H, Hellweg S, Huijbregts MAJ (2007) Cumulative exergy extraction from the natural environment (CEENE). A comprehensive life cycle impact assessment method for resource accounting. Env Sci Technol 41:8477–8483. https://doi.org/10.1021/es0711415

Dong YH, Ng ST, Kumaraswamy MM (2016) Critical analysis of the life cycle impact assessment methods. Env Eng Man J 15:879–890. https://doi.org/10.30638/eemj.2016.095

Dreyer LC, Niemann AL, Hauschild MZ (2003) Comparison of three different LCIA methods: EDIP97, CML2001 and Eco-indicator 99. Does it matter which one you choose? Int J Life Cycle Assess 8:191–200. https://doi.org/10.1007/BF02978471

Eckelman MJ (2016) Life cycle inherent toxicity. A novel LCA-based algorithm for evaluating chemical synthesis pathways. Green Chem 11:3257–3264. https://doi.org/10.1039/C5GC02768C

Eddy DC, Krishnamurty S, Grosse IR, Wileden JC, Lewis KE (2015) A predictive modelling-based material selection method for sustainable product design. J Eng Design 26:365–390. https://doi.org/10.1080/09544828.2015.1070258

Emami N, Heinonen J, Marteinsson B, Säynäjoki A, Junnonen J-M, Laine J, Junnila S (2019) A life cycle assessment of two residential buildings using two different LCA database-software combinations. Recognizing Uniformities and Inconsistencies Buildings 9:20. https://doi.org/10.3390/buildings9010020

Frischknecht R, Althaus H-J, Bauer C, Doka G, Heck T, Jungbluth N, Kellenberger D, Nemecek T (2007) The environmental relevance of capital goods in life cycle assessments of products and services. Int J Life Cycle Assess 12 (special issue):7–17

Gujarati DN (2003) Basic econometrics. Fourth edition. McGraw-Hill

Gutiérrez E, Adenso-Díaz B, Lozano S, Barba-Gutiérrez Y (2009) Visualisation of LCA environmental impacts of electrical and electronic products using multidimensional scaling. Int J Prod Lifecycle Man 4:166–185. https://doi.org/10.1504/ijplm.2009.031672

Gutiérrez E, Lozano S, Adenso-Díaz B (2010a) Dimensionality reduction and visualization of the environmental impacts of domestic appliances. J Ind Ecol 14:878–889. https://doi.org/10.1111/j.1530-9290.2010.00291.x

Gutiérrez E, Lozano S, Moreira MT, Feijoo G (2010b) Assessing relationships among life-cycle environmental impacts with dimension reduction techniques. J Env Man 91:1002–1011. https://doi.org/10.1016/j.jenvman.2009.12.009

Halleux H, Lassaux S, Germain A (2006) Comparison of life cycle assessment methods, application to a wastewater treatment plant. 13th CIRP International Conference on Life Cycle Engineering 93–96. URL: http://ww.seeds4green.org/sites/default/files/086_2.pdf

Hanes R, Bakshi BR, Goel PK (2013) The use of regression in streamlined life cycle assessment. Proc ISSST. https://doi.org/10.6084/m9.figshare.815891

Heijungs R (2017) Comment on "Resource footprints are good proxies of environmental damage." Env Sci Technol 51:13054–13055. https://doi.org/10.1021/acs.est.7b04253

Heijungs R (2020) Is mainstream LCA linear? Int J Life Cycle Assess 25:1872–1882. https://doi.org/10.1007/s11367-020-01810-z

Heijungs R, de Koning A, Wegener Sleeswijk A (2015) Sustainability analysis and systems of linear equations in the era of data abundance. J Env Acc Man 3:109–122. https://doi.org/10.5890/JEAM.2015.06.003

Heijungs R, Guinée JB, Henriksson PJG, Mendoza Beltrán MA, Groen EA (2019) Everything is relative and nothing is certain. Toward a theory and practice of comparative probabilistic LCA. Int J Life Cycle Assess 24 1573-1579 s11367-019-01666-y

Heijungs R, Henriksson PJG, Guinée JB (2016) Measures of difference and significance in the era of computer simulations, meta-analysis, and big data. Entropy 18:361. https://doi.org/10.3390/e18100361

Heijungs R, Suh S (2002) The computational structure of life cycle assessment. Kluwer, Dordrecht

Hendrickson CT, Horvath A, Joshi S, Klausner M, Lave LB, McMichael FC (1997) Comparing two life cycle assessment approaches. A process model- vs. economic input-output-based assessment. Proceedings of the 1997 IEEE International Symposium on Electronics and the Environment. https://doi.org/10.1109/ISEE.1997.605313

Herrmann IT, Moltesen A (2015) Does it matter which life cycle assessment (LCA) tool you choose? A comparative assessment of SimaPro and GaBi. J Cleaner Prod 86:163–169. https://doi.org/10.1016/j.jclepro.2014.08.004

Hill RC, Griffiths WE, Lim GC (2011) Principles of econometrics. Fourth edition. John Wiley & Sons

Hochschorner E, Finnveden G (2003) Evaluation of two simplified life cycle assessment methods. Int J Life Cycle Assess 8:119–128. https://doi.org/10.1007/BF02978456

Hou P, Jolliet O, Zhu J, Xu M (2020) Estimate ecotoxicity characterization factors for chemicals in life cycle assessment using machine learning models. Env Int 135:105393. https://doi.org/10.1016/j.envint.2019.105393

Huijbregts MAJ (1998) Application of uncertainty and variability in LCA. Part II: Dealing with parameter uncertainty and uncertainty due to choices in life cycle assessment. Int J Life Cycle Assess 3:343–351. https://doi.org/10.1007/BF02979345

Huijbregts MAJ, Geelen LMJ, Hertwich EG, McKone TE, Van de Meent D (2005) A comparison between the multimedia fate and exposure models CalTOX and uniform system for evaluation of substances adapted for life-cycle assessment based on the population intake fraction of toxic pollutants. Env Tox Chem 24:486–493. https://doi.org/10.1897/04-001R.1

Huijbregts MAJ, Rombouts LJA, Hellweg S, Frischknecht R, Hendriks AJ, van de Meent D, Ragas AMJ, Reijnders L, Struijs J (2006) Is cumulative fossil energy demand a useful indicator for the environmental performance of products? Env Sci Technol 40:641–648. https://doi.org/10.1021/es051689g

Huijbregts MAJ, Hellweg S, Frischknecht R, Hungerbühler K, Hendriks AJ (2008) Ecological footprint accounting in the life cycle assessment of products. Ecol Econ 64:798–807. https://doi.org/10.1016/j.ecolecon.2007.04.017

Huijbregts MAJ, Hellweg S, Frischknecht R, Hendriks HWM, Hungerbühler K, Hendriks AJ (2010) Cumulative energy demand as predictor for the environmental burden of commodity production. Env Sci Technol 44:2189–2196. https://doi.org/10.1021/es902870s

Huppes G, van Oers L, Pretato U, Pennington DW (2012) Weighting environmental effects. Analytic survey with operational evaluation methods and a meta-method. Int J Life Cycle Assess 17:876–891. https://doi.org/10.1007/s11367-012-0415-x

Islam S, Ponnambalam SG, Lam HL (2016) Review on life cycle inventory. Methods, examples and applications. J Cleaner Prod 136:266–278. https://doi.org/10.1016/j.jclepro.2016.05.144

Iswara AP, Farahdiba AU, Nadhifatin EN, Pirade F, Andhikaputra G, Muflihah I, Boedisantoso R (2020) A comparative study of life cycle impact assessment using different software programs. IOP Conf Ser Earth Env Sci 506:012002. https://doi.org/10.1088/1755-1315/506/1/012002

James G, Witten D, Hastie T, Tibshirani R (2015) An introduction to statistical learning. With applications in R. Springer

Jammalamadaka SR, SenGupta A (2001) Topics in circular statistics. World Scientific

Joyce PJ, Björklund A (in press) Futura. A new tool for transparent and shareable scenario analysis in prospective life cycle assessment. J Ind Ecol. https://doi.org/10.1111/jiec.13115

Jung J, Von der Assen N, Bardow A (2014) Sensitivity coefficient-based uncertainty analysis for multi-functionality in LCA. Int J Life Cycle Assess 19:661–676. https://doi.org/10.1007/s11367-013-0655-4

Junnila SI (2006) Empirical comparison of process and economic input-output life cycle assessment in service industries. Env Sci Technol 40:7070–7076. https://doi.org/10.1021/es0611902

Kalbar PP, Birkved M, Karmakar S, Elsborg Nygaard S, Hauschild M (2017) Can carbon footprint serve as proxy of the environmental burden from urban consumption patterns? Ecol Ind 74:109–118. https://doi.org/10.1016/j.ecolind.2016.11.0221

Kaufman SM, Krishnan N, Themelis NJ (2010) A screening life cycle metric to benchmark the environmental sustainability of waste management systems. Env Sci Technol 55:5949–5955. https://doi.org/10.1021/es100505u

Kounina A, Margni M, Shaked S, Bulle C, Jolliet O (2014) Spatial analysis of toxic emissions in LCA. A sub-continental nested USEtox model with freshwater archetypes. Env Int 69:67–89. https://doi.org/10.1016/j.envint.2014.04.004

Laleman R, Albrecht J, Dewulf J (2013) Comparing various indicators for the LCA of residential photovoltaic systems. In: Singh A., Pant D., Olsen S. (eds). Life cycle assessment of renewable energy sources. Green energy and technology. Springer

Landis AE, Theis TL (2008) Comparison of life cycle impact assessment tools in the case of biofuels. 2008 IEEE International Symposium on Electronics and the Environment. https://doi.org/10.1109/ISEE.2008.4562869

Lasvaux S, Achim F, Garat P, Peuportier B, Chevalier J, Habert G (2016) Correlations in life cycle impact assessment methods (LCIA) and indicators for construction materials: What matters? Ecol Ind 67:174–182. https://doi.org/10.1016/j.ecolind.2016.01.056

Laurent A, Olsen SI, Hauschild MZ (2012) Limitations of carbon footprint as indicator of environmental sustainability. Env Sci Technol 46:4100–4108. https://doi.org/10.1021/es204163f

Lautier A, Rosenbaum RK, Margni M, Bare J, Roy P-O, Deschênes L (2010) Development of normalization factors for Canada and the United States and comparison with European factors. Sci Total Env 409:33–42. https://doi.org/10.1016/j.scitotenv.2010.09.016

Lee A (2010) Circular data. WIREs Comp. Stat 2:477–486. https://doi.org/10.1002/wics.98

Legendre P, Legendre L (1998) Numerical ecology. Elsevier, Second English edition

Le Téno JF (1999) Visual data analysis and decision support for non-deterministic LCA. Int J Life Cycle Assess 4:41–47. https://doi.org/10.1007/BF02979394

Ley C, Verdebout T (2017) Modern directional statistics. CRC Press

Ligthart TN, Ansems AMM (2019) EnvPack. An LCA-based tool for environmental assessment of packaging chains. Part 2: influence of assessment method on ranking of alternatives. Int J Life Cycle Assess 24:915–925. https://doi.org/10.1007/s11367-018-1531-z

Mardia KV, Jupp PE (2000) Directional statistics. John Wiley & Sons

Martínez E, Blanco J, Jiménez E, Saenz-Díez JC, Sanz F (2015) Comparative evaluation of life cycle impact assessment software tools through a wind turbine case study. Ren Energy 74:237–246. https://doi.org/10.1016/j.renene.2014.08.004

Marvuglia A, Kanevski M, Leuenberger M, Benetto E (2014) Variables selection for ecotoxicity and human toxicity characterization using gamma test. In: Murgante B, Misra S, Rocha AMAC, Torre C, Rocha JG, Falcão MI, Taniar D, Apduhan BO, Gervasi O (2014) Computational science and its applications. ICCSA 2014. Springer

Marvuglia A, Kanevski M, Benetto E (2015) Machine learning for toxicity characterization of organic chemical emissions using USEtox database. Learning the structure of the input space. Environ Int 83:72–85. https://doi.org/10.1016/j.envint.2015.05.011

Masnadi MS, Perrier PR, Wang J, Rutherford J, Brandt AR (2020) Statistical proxy modeling for life cycle assessment and energetic analysis. Energy 194:116882. https://doi.org/10.1016/j.energy.2019.116882

Mendoza Beltrán MA, Heijungs R, Guinée JB, Tukker A (2016) A pseudo-statistical approach to treat choice uncertainty. The example of partitioning allocation methods. Int J Life Cycle Assess 21:252–264. https://doi.org/10.1007/s11367-015-0994-4

Menten F, Chèze B, Patouillard L, Bouvart B (2013) A review of LCA greenhouse gas emissions results for advanced biofuels. The use of meta-regression analysis. Renew Sust En Rev 26:108–134. https://doi.org/10.1016/j.rser.2013.04.021

Monteiro H, Freire F (2012) Life-cycle assessment of a house with alternative exterior walls: comparison of three impact

assessment methods. Energy and Buildings 47:572–583. https://doi.org/10.1016/j.enbuild.2011.12.032
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
Notarnicola B, Huppes G, van den Berg NW (1998) Evaluating options in LCA. The emergence of conflicting paradigms for impact assessment and evaluation. Int J Life Cycle Assess 3:289–300. https://doi.org/10.1007/BF02979839
Núñez M, Bouchard CR, Bulle C, Boulay A-M, Margni M (2016) Critical analysis of life cycle impact assessment methods addressing consequences of freshwater use on ecosystems and recommendations for future method development. Int J Life Cycle Assess 21:1799–1815. https://doi.org/10.1007/s11367-016-1127-4
Ott RL, Longnecker MT (2015) An introduction to statistical methods and data analysis. Seventh edition. Cengage
Owsianiak M, Laurent A, Bjørn A, Hauschild MZ (2014) IMPACT 2002+, ReCiPe 2008 and ILCD's recommended practice for characterization modelling in life cycle impact assessment. A case study-based comparison. Int J Life Cycle Assess 19:1007–1021. https://doi.org/10.1007/s11367-014-0708-3
Padey P, Girard R, le Boulch D, Blanc I (2013) From LCAs to simplified models. A generic methodology applied to wind power electricity. Env Sci Technol 47:2131–1238. https://doi.org/10.1021/es303435e
Pant R, van Hoof G, Schowanek D, Feijtel TCJ, de Koning A, Hauschild M, Pennington DW, Olsen SI, Rosenbaum R (2004) Comparison between three different LCIA methods for aquatic ecotoxicity and a product environmental risk assessment. Int J Life Cycle Assess 9:1295–1306. https://doi.org/10.1007/BF02979419
Park J-H, Seo K-K (2003) Approximate life cycle assessment of product concepts using multiple regression analysis and artificial neural networks. KSME Int J 17:1969–1976. https://doi.org/10.1007/BF02982436
Park J-H, Seo K-K (2006) A knowledge-based approximate life cycle assessment system for evaluating environmental impacts of product design alternatives in a collaborative design environment. Adv Eng Informatics 20:147–154. https://doi.org/10.1016/j.aei.2005.09.003
Park J-H, Seo K-K, Wallace D. Approximate life cycle assessment of classified products using artificial neural network and statistical analysis in conceptual product design. Proceedings Second International Symposium on Environmentally Conscious Design and Inverse Manufacturing (2001), 321–326. https://doi.org/10.1109/ECODIM.2001.992373
Park YS, Egilmez G, Kucukvar M (2015) A novel life cycle-based principal component analysis framework for eco-efficiency analysis: case of the United States manufacturing and transportation nexus. J Cleaner Prod 92:327–342. https://doi.org/10.1016/j.jclepro.2014.12.057
Pascual-González J, Guillén-Gosálbez G, Mateo-Sanz JM, Jiménez-Esteller L (2016) Statistical analysis of the ecoinvent database to uncover relationships between life cycle impact assessment metrics. J Cleaner Prod 112:359–368. https://doi.org/10.1016/j.jclepro.2015.05.129
Pascual-González J, Pozo C, Guillén-Gosálbez G, Jiménez-Esteller L (2015) Combined use of MILP and multi-linear regression to simplify LCA studies. Comp Chem Eng 82:34–43. https://doi.org/10.1016/j.compchemeng.2015.06.002
Peters GP (2007) Efficient algorithms for life cycle assessment, input-output analysis, and Monte-Carlo analysis. Int J Life Cycle Assess 12:373–380. https://doi.org/10.1065/lca2006.06.254
Pewsey A, Neuhäuser M, Ruxton G.D (2013) Circular statistics in R. Oxford University Press
Pizzol M, Christensen P, Schmidt J, Thomsen M (2011) Impacts of "metals" on human health. A comparison between nine different methodologies for life cycle impact assessment (LCIA). J Cleaner Prod 19:646–656. https://doi.org/10.1016/j.jclepro.2010.05.007
Pourhoseingholi MA, Baghestani AR, Vahedi M (2012) How to control confounding effects by statistical analysis. Gastroenterol Hepatol Bed Bench 5:79–83
Pozo C, Ruíz-Femenia R, Caballero J, Guillén-Gosálbez G, Jiménez L (2012) On the use of principal component analysis for reducing the number of environmental objectives in multi-objective optimization. Application to the design of chemical supply chains. Chem Eng Sci 69:146–158. https://doi.org/10.1016/j.ces.2011.10.018
Renou S, Thomas JS, Aoustin E, Pons MN (2008) Influence of impact assessment methods in wastewater treatment LCA. J Cleaner Prod 16:1098–1105. https://doi.org/10.1016/j.jclepro.2007.06.003
Röös E, Sundberg C, Tidåker P, Strid I, Hansson P-A (2013) Can carbon footprint serve as an indicator of the environmental impact of meat production? Ecol Ind 24:573–581. https://doi.org/10.1016/j.ecolind.2012.08.004
Scipioni A, Niero M, Mazzi A, Manzardo A, Piubello S (2013) Significance of the use of non-renewable fossil CED as proxy indicator for screening LCA in the beverage packaging sector. Int J Life Cycle Assess 18:673–682. https://doi.org/10.1007/s11367-012-0484-x
Schulze C, Jödicke A, Scheringer M, Margni M, Jolliet O, Hungerbühler K, Matthies M (2001) Comparison of different life-cycle impact assessment methods for aquatic ecotoxicity. Env Tox Chem 20:2122–2132. https://doi.org/10.1002/etc.5620200936
Shariar Hossain M, Marwah M, Shah A, Watson LT, Ramakrishnan N (2014) AutoLCA. A framework for sustainable redesign and assessment of products. ACM Trans Intell Syst Tech 5:1–21. https://doi.org/10.1145/2505270
Simões CL, Xará SM, Bernardo CA (2011) Influence of the impact assessment method on the conclusions of a LCA study. Application to the case of a part made with virgin and recycled HDPE. Waste Man Res 29:1018–1026. https://doi.org/10.1177/0734242X11403799
Song R, Keller AA, Suh S (2017) Rapid life-cycle impact screening using artificial neural networks. Env Sci Technol 51:10777–10785. https://doi.org/10.1021/acs.est.7b02862
Sousa I, Wallace D, Eisenhard JL (2000) Approximate life-cycle assessment of product concepts using learning systems. J Ind Ecol 4:61–81. https://doi.org/10.1162/10881980052541954
Speck R, Selke S, Auras R, Fitzsimmons J (2015) Life cycle assessment software. Selection can impact results. J Ind Ecol 20:18–28. https://doi.org/10.1111/jiec.12245
Suh S, Huppes G (2005) Methods for life cycle inventory of a product. J Cleaner Prod 13:687–697. https://doi.org/10.1016/j.jclepro.2003.04.001
Steinmann ZJN, Schipper AM, Hauck M, Huijbregts MAJ (2016) How many environmental impact indicators are needed in the evaluation of product life cycles? Env Sci Technol 50:3913–3919. https://doi.org/10.1021/acs.est.5b05179
Steinmann ZJN, Schipper AM, Hauck M, Giljum S, Wernet G, Huijbregts MAJ (2017a) Resource footprints are good proxies of environmental damage. Env Sci Technol 51:6360–6366. https://doi.org/10.1021/acs.est.7b00698
Steinmann ZJN, Schipper AM, Hauck M, Giljum S, Wernet G, Huijbregts MAJ (2017b) Response to Comment on "Resource Footprints are Good Proxies of Environmental Damage." Env Sci Technol 51:13056–13057. https://doi.org/10.1021/acs.est.7b04926
Timsina J, Humphreys E (2006) Performance of CERES-Rice and CERES-Wheat models in rice-wheat systems. A Review Agr Syst 90:5–31. https://doi.org/10.1016/j.agsy.2005.11.007

Valente A, Iribarrena D, Dufour J (2018) Harmonising the cumulative energy demand of renewable hydrogen for robust comparative life-cycle studies. J Cleaner Prod 175:384–393. https://doi.org/10.1016/j.jclepro.2017.12.069

Valente A, Iribarrena D, Dufour J (2019) Harmonising methodological choices in life cycle assessment of hydrogen. A focus on acidification and renewable hydrogen. Int J Hydr Energy 44:19426–19433. https://doi.org/10.1016/j.ijhydene.2018.03.101

Van der Werf HMG, Petit J (2002) Evaluation of the environmental impact of agriculture at the farm level. A comparison and analysis of 12 indicator-based methods. Agr Ecosyst Env 93:131–145. https://doi.org/10.1016/S0167-8809(01)00354-1

Weidema BP (2015) Comparing three life cycle impact assessment methods from an endpoint perspective. J Ind Ecol 19:20–26. https://doi.org/10.1111/jiec.12162

Wernet G, Hellweg S, Fischer U, Papadokonstantakis S, Hungerbühler K (2008) Molecular-structure-based models of chemical inventories using neural networks. Env Sci Technol 42:6717–6722. https://doi.org/10.1021/es7022362

Wernet G, Mutel C, Hellweg S, Hungerbühler K (2011) The environmental importance of energy use in chemical production. J Ind Ecol 15:96–107. https://doi.org/10.1111/j.1530-9290.2010.00294.x

Zhang Y, Bakshi BR (2007) Statistical evaluation of input-side metrics for life cycle impact assessment of emerging technologies. Proceedings of the 2007 IEEE International Symposium on Electronics and the Environment 117–122. https://doi.org/10.1109/ISEE.2007.369378