



HAL
open science

Consistent normalization approach for Life Cycle Assessment based on inventory databases

Arnaud Hélias, Antoine Esnouf, Matthias Finkbeiner

► To cite this version:

Arnaud Hélias, Antoine Esnouf, Matthias Finkbeiner. Consistent normalization approach for Life Cycle Assessment based on inventory databases. *Science of the Total Environment*, 2020, 703, pp.134583. 10.1016/j.scitotenv.2019.134583 . hal-02609777

HAL Id: hal-02609777

<https://hal.inrae.fr/hal-02609777>

Submitted on 21 Dec 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Copyright

23 reference values are determined from the geometric means of the datasets of the
24 inventory database, used for assessing the studied systems. The exemplary application
25 to the ecoinvent databases provides normalization references for 878 versions of the
26 impacts categories listed by ecoinvent and for the 2077 involved substances. For eight
27 impact assessment methods, the results are compared with 16 normalization sets from
28 the literature and point out highly significant correlations.

29 **Keywords**

30 Life Cycle Assessment; normalization; bottom-up reference; geometric mean;

31

32

33 **1. Introduction**

34 The purpose of Life Cycle Assessment (LCA) is to assess and quantify the environmental
35 interventions of human activities. All introduction chapters in books and lectures
36 concerning LCA present the well-known ISO standard (ISO, 2006a) diagram. This
37 schematic highlights the importance of interpretation: within the sequence of goal and
38 scope, inventory, impact assessment and interpretation, the latter visually takes up the
39 half of the diagram. The reason for this is that LCA results are often difficult to decipher
40 and interpret.

41 LCA is a relative and multicriteria approach (Finkbeiner et al., 2006). When the
42 environmental impacts of a given system are calculated, the output often comprises a
43 set of results expressed with different units (such as global warming in $\text{kgCO}_2\text{-Eq}$,
44 freshwater ecotoxicity in Comparative Toxic Unit, $\text{CTUe}\cdot\text{m}^3\cdot\text{yr}$, or resource depletion in
45 kgSb-Eq). To put the results into context and for understanding the magnitude of the
46 impacts compared with one another, the results are related to a common reference
47 system. This removes the different units and brings the quantitative results in similar
48 orders of magnitude. A normalization procedure, an optional step in the LCA, addresses
49 this issue.

50 Normalization is the calculation of the magnitude of the category indicator results,
51 relative to some reference information (ISO, 2006b). Typically, external normalization
52 references are chosen, i.e. the data for the reference system are independent from the
53 environmental profile of the study and as a consequence also do not vary according to
54 the system under study (Laurent and Hauschild, 2015). External normalization data
55 require ideally a complete inventory of emissions and resource consumptions for the
56 reference area during the reference time period (e.g. regional or world emissions over

57 one year). It allows for the plausibility check of the order of magnitude of results and
58 facilitates communication of results by expressing them with a commensurable unit,
59 such as person equivalents for example (Pizzol et al., 2017). Normalization is also
60 recommended as a preliminary step for a potential weighting of environmental impacts
61 and assembling all impacts into a single reference system.

62 The quantification of the reference system and associated data raises the issue of
63 system boundaries in terms of space (region, country or worldwide) and duration.
64 According to these limits, consumption- or production-based inventories can be used
65 (Laurent and Hauschild, 2015). In all situations, the reference is intended to be
66 exhaustive, but this is obviously never fully achieved and introduces bias. The use of
67 different background data sources for the study results and the normalization data
68 introduce inconsistencies. Especially for regional normalization, there is a tendency to
69 over-interpret the results. As an example, normalization to European reference values is
70 often interpreted as the share of the product under study to emissions in Europe,
71 whereas depending on the product systems a significant share of emissions occurs
72 outside Europe and as such are not part of the reference value.

73 The Life Cycle Initiative, hosted by UN environment, recently published a state of the art
74 of current works and issues with normalization in LCA (Pizzol et al., 2017). In this
75 document, the potential biases of an external normalization approach have been
76 highlighted as well as in other references (Benini and Sala, 2016; Cucurachi et al., 2017;
77 Heijungs et al., 2007). They result from the discrepancies between the life cycle
78 inventory (LCI) of a system studied and the LCI of normalization references, in terms of
79 involved processes (the perimeter), modelling choices and inventoried substances. It

80 recommends the definition of methods providing consistent and sufficiently complete
81 inventories of emissions and resource consumptions at a global scale.

82 In line with previous works (Esnouf et al., 2019, 2018) where the principle was used in
83 another context, the purpose of this paper is to define a generic approach for
84 establishing an appropriate normalization reference, using the same representation (i.e.
85 modelling) of the world, both for the system under study and the reference system. The
86 method can be applied based on the inventory databases used by LCA practitioners, i.e.
87 it does not rely on any other data sources than the ones which are used in the study
88 itself. The approach is exemplarily applied to ecoinvent (Wernet et al., 2016) to test its
89 applicability and to compare it with existing external normalization references.

90 **2. Material and methods**

91 For matters of convenience, all the notations in the article are detailed in Table 1. The
92 matrices are in bold and their elements in italic. See Heijungs and Suh (2002) for details
93 about matrix products to compute the Life Cycle Impact Assessment (LCIA) result. Note
94 that pre-calculated inventories $\tilde{\mathbf{B}}$ (Broadbent et al., 2011; Lesage et al., 2018) are used
95 (known as “system” datasets in ecoinvent).

96 [Insert Table 1 here]

97 *2.1. Normalization principle.*

98 A normalized result h_k^N is defined as follows:

$$h_k^N = \frac{h_k}{h_k^{\text{REF}}} \quad (1)$$

99 According to LCA principles, the normalization references h_k^{REF} result from the
100 multiplications of emissions (or resource consumptions) of the reference system by the

101 associated characterization factors. The elementary flows of the reference system are
 102 commonly determined from statistics provided by international agencies and
 103 institutions, with a structured collection and aggregation work (Crenna et al., 2019; Sala
 104 et al., 2015). This is defined here in a different way but similar to other LCA results, i.e.
 105 with the intervention matrix and the final demand:

$$h_k^{\text{REF}} = \sum_{j=1}^n q_j \sum_{i=1}^m \tilde{b}_{j,i} f_i^{\text{REF}} \quad (2)$$

106 The definition of an external normalization reference implies the estimation of the
 107 overall emissions. Considering $\tilde{\mathbf{B}}$ as representative of all the processes involved in the
 108 reference system (i.e. the inventory database is sufficiently complete to represent what
 109 is happening at a global scale), the computing of \mathbf{h}^{REF} only requires the determination of
 110 \mathbf{f}^{REF} (the f_i^{REF} values). The difficulty in defining normalization references resides in the
 111 quantification of the final demand of the reference system.

112 As the system under study is the part of the reference system, consistent normalization
 113 implies the use of the same intervention datasets for both. Therefore, only the final
 114 demands \mathbf{f} and \mathbf{f}^{REF} differ between h_k and h_k^{REF} . By defining the reference system in this
 115 manner, the same perimeter, data sources and list of inventoried elementary flows are
 116 used.

117 *2.2. External normalization as an average intervention.*

118 The overall emissions are proportional, with the factor m , to $M_1(\tilde{b}_{j,i} f_i^{\text{REF}}, \forall i) =$
 119 $\frac{1}{m} \sum_{i=1}^m \tilde{b}_{j,i} f_i^{\text{REF}}$, defined as the arithmetic mean of all $\tilde{b}_{j,i} f_i^{\text{REF}}$

$$h_k^{\text{REF}} = m \sum_{j=1}^n q_j M_1(\tilde{b}_{j,i} f_i^{\text{REF}}, \forall i) \quad (3)$$

120 m does not depend on j and is therefore identical for all impact categories. The use of
121 $\frac{h_k^{\text{REF}}}{m}$ instead of h_k^{REF} as normalization factor does not modify the ratios between
122 normalized impact categories. The conclusions on the results should be identical. A
123 parallel can be made with a normalization concerning either an average citizen or a
124 whole population: between both references, the normalized impacts differ by a factor
125 equal to the population size, but the interpretation of the outcome remains identical for
126 both normalizations.

127 Normalization references correspond to the sum of all associated interventions;
128 however, their average could be easily taken into account instead. The normalization
129 procedure could thus be defined by the expression of the result as a function of the
130 impact of the arithmetic mean of interventions in the reference system.

131 *2.3. From the arithmetic mean to the geometric mean.*

132 The arithmetic mean is the most commonly used mean. It is the expected value of data
133 distributed according to the well-known normal law. For the normalization process, the
134 distribution of $\tilde{b}_{j,i} f_i^{\text{REF}}$ cannot be described because, unlike the $\tilde{b}_{j,i}$ values, the values of
135 f_i^{REF} are unknown. Let us focus on $\tilde{b}_{j,i}$ values. The shapes of flow distributions in
136 inventory databases rather seem to resemble log-distribution laws, with values
137 spreading over several orders of magnitude. This has been demonstrated by Qin and
138 Suh (2017) who observed that the main part of the technology and intervention flows in
139ecoinvent follows lognormal laws. Log-distributions are not well described by the
140 arithmetic mean but rather by the geometric mean (i.e. the exponential of the arithmetic
141 mean of logarithm-transformed values)

$$M_0(\tilde{\mathbf{B}}_j \mathbf{f}^{\text{REF}}) = \exp\left(\frac{1}{m} \sum_{i=1}^m \log(\tilde{b}_{j,i} f_i^{\text{REF}})\right) = \sqrt[m]{\prod_{i=1}^m \tilde{b}_{j,i} f_i^{\text{REF}}} \quad (4)$$

142 It is therefore relevant to use it in eq (3) with the hypothesis of a constant proportion
 143 between the means

$$c M_0(\tilde{b}_{j,i} f_i^{\text{REF}}, \forall i) \approx M_1(\tilde{b}_{j,i} f_i^{\text{REF}}, \forall i) \quad (5)$$

144 then

$$h_k^{\text{REF}} \approx m c \sum_{j=1}^n q_j M_0(\tilde{b}_{j,i} f_i^{\text{REF}}, \forall i) \quad (6)$$

145 $M_0(\tilde{b}_{j,i} f_i^{\text{REF}}, \forall i)$ can be split into $M_0(\tilde{b}_{j,i}, \forall i) = M_0(\tilde{\mathbf{B}}_j)$ and $M_0(f_i^{\text{REF}}, \forall i) = M_0(\mathbf{f}^{\text{REF}})$

$$M_0(\tilde{b}_{j,i} f_i^{\text{REF}}, \forall i) = M_0(\tilde{\mathbf{B}}_j) \times M_0(\mathbf{f}^{\text{REF}}) = \sqrt[m]{\prod_{i=1}^m \tilde{b}_{j,i}} \times \sqrt[m]{\prod_{i=1}^m f_i^{\text{REF}}} \quad (7)$$

146 Then, with eq (6)

$$h_k^{\text{REF}} \approx m c M_0(\mathbf{f}^{\text{REF}}) \sum_{j=1}^n q_j M_0(\tilde{\mathbf{B}}_j) \quad (8)$$

147 2.3.1. Inventory-based normalization references

148 As previously stated for the term m in eq (3), the constant term $m c M_0(\mathbf{f}^{\text{REF}})$ can be
 149 neglected without consequences on the interpretation of results. The normalized
 150 reference is defined as

$$h_k^{\text{REF}, \tilde{\mathbf{B}}} = \sum_{j=1}^n q_j M_0(\tilde{\mathbf{B}}_j) \quad (9)$$

$$h_k^{\text{REF},\tilde{\mathbf{B}}} \approx \frac{1}{m c M_0(\mathbf{f}^{\text{REF}})} h_k^{\text{REF}} \quad (10)$$

151 which should produce the same results, in terms of interpretation, as h_k^{REF} , without the
 152 need to know \mathbf{f}^{REF} .

153 The obtained normalization reference is based on characterization factors and on the
 154 inventory database only, without necessary external data collection. The approach
 155 implies two assumptions: (1) The reference system is defined by the inventories listed
 156 in the used LCI database. These inventories ensure consistency with the system under
 157 study. (2) The geometric mean is used instead of the arithmetic mean. This is a coherent
 158 choice for representing a variability spread over several orders of magnitude.

159 Similar reasoning can be carried out with the intervention matrix \mathbf{B} instead of $\tilde{\mathbf{B}}$ and so
 160 with $M_0(\mathbf{B}_j) \times M_0(\mathbf{A}^{-1}\mathbf{f}^{\text{REF}})$ instead of $M_0(\tilde{\mathbf{B}}_j) \times M_0(\mathbf{f}^{\text{REF}})$. However, the geometric
 161 mean estimations with \mathbf{B} are more difficult (see below).

162 2.3.2. Zero and negative values for interventions

163 While the arithmetic mean can be calculated with all true values, the geometric mean
 164 strictly requires positive values. This constraint has to be considered for eqs (4) to (10),
 165 for $M_0(\mathbf{f}^{\text{REF}})$ to allow the simplification (note that the value of $M_0(\mathbf{f}^{\text{REF}})$ does not need
 166 to be calculated, but only its existence should be taken into account), and for eq (4)
 167 to (9) to compute $M_0(\tilde{\mathbf{B}}_j)$.

168 For $M_0(\mathbf{f}^{\text{REF}})$, this seems quite simple to achieve. If $f_i^{\text{REF}} < 0$, the avoided quantity is
 169 higher than its production in the final demand. This case should not be encountered
 170 when the whole system is taken into account for the reference. It is possible to have
 171 $f_i^{\text{REF}} = 0$; however, for a given process, it indicates that its use is exactly equal to the

172 sum of its calls by the other processes, without any additional final demands for the
173 normalization reference. To ensure that $f_i^{\text{REF}} > 0, \forall i$, the non-existence of zero values
174 can be considered by assuming that there will always be a very small amount (e.g. 1 μg
175 of a product or 1 μWh of electricity) missing in calls and thus at least required for the
176 final demand of the reference.

177 For $M_0(\tilde{\mathbf{B}}_j)$, the issue is more challenging. Most processes directly induce a few
178 numbers of elementary flows, usually less than one hundred, while the total number of
179 elementary flows amounts to a thousand. Therefore, matrix \mathbf{B} is particularly sparse, but
180 this appears less problematic for $\tilde{\mathbf{B}}$: a high number of elementary flows are involved
181 because a pre-calculated intervention encompasses the whole system associated to the
182 corresponding process (the complete process tree). Not many zeros are observed in $\tilde{\mathbf{B}}$,
183 although some do obviously occur.

184 The computation of geometric means using zero values is an issue often encountered in
185 data processing, as for example, in biology with microbial counting or in chemistry, with
186 concentrations below detection limits. The most common approach is to add 1 to the
187 values, compute the product and remove 1 afterwards. In this work, a recent algorithm,
188 based on this principle, was used (de la Cruz and Kreft, 2018). It determines the
189 smallest value to be added and then removed in order to compute a consistent
190 estimation of geometric means involving zero values.

191 The intervention matrix $\tilde{\mathbf{B}}$ contains negative values. This can be due to the intrinsic
192 properties of the represented process, such as plant growth removing CO_2 from
193 atmosphere through photosynthesis. Negative emissions are found in system models
194 involving substitution (i.e. avoided processes, then avoided elementary flows) or

195 partition/allocation (e.g. negative flows ensuring the carbon balance consistency
 196 between products and coproducts, see for example Weidema (2018) for details).

197 When the whole dataset is negative, the computation of a geometric mean is generally
 198 carried out with absolute values after multiplication by minus one. Unfortunately, this is
 199 not feasible when both negative and positive values are in the dataset. To overcome this
 200 issue, two new matrixes $\tilde{\mathbf{B}}^+$ and $\tilde{\mathbf{B}}^-$ are defined, where the elements $\tilde{b}_{j,i}^+$ and $\tilde{b}_{j,i}^-$ are
 201 described as follows

$$\tilde{b}_{j,i}^+ = \begin{cases} \tilde{b}_{j,i}, & \tilde{b}_{j,i} > 0 \\ 0, & \tilde{b}_{j,i} \leq 0 \end{cases}, \quad \tilde{b}_{j,i}^- = \begin{cases} 0, & \tilde{b}_{j,i} \geq 0 \\ -\tilde{b}_{j,i}, & \tilde{b}_{j,i} < 0 \end{cases} \quad (11)$$

202 The LCIA result does not change with the following equation

$$\mathbf{h} = \mathbf{Q}\tilde{\mathbf{B}}\mathbf{f} = [\mathbf{Q} \quad -\mathbf{Q}] \begin{bmatrix} \tilde{\mathbf{B}}^+ \\ \tilde{\mathbf{B}}^- \end{bmatrix} \mathbf{f} \quad (12)$$

203 $\tilde{\mathbf{B}}^+$ and $\tilde{\mathbf{B}}^-$ only contain positive or null values, as the negative sign is represented by
 204 $-\mathbf{Q}$. It is therefore possible to compute the corresponding estimation of geometric
 205 means $M_0(\tilde{\mathbf{B}}_j^+)$ and $M_0(\tilde{\mathbf{B}}_j^-)$, $\forall j$. The normalization reference $h_k^{\text{REF},\tilde{\mathbf{B}}}$ is thus

$$h_k^{\text{REF},\tilde{\mathbf{B}}} \approx \sum_{j=1}^n q_j M_0(\tilde{\mathbf{B}}_j^+) - \sum_{j=1}^n q_j M_0(\tilde{\mathbf{B}}_j^-) = \sum_{j=1}^n q_j (M_0(\tilde{\mathbf{B}}_j^+) - M_0(\tilde{\mathbf{B}}_j^-)) \quad (13)$$

206 *2.4. Intervention databases and associated impacts.*

207 Inventories \mathbf{B} and pre-calculated inventories $\tilde{\mathbf{B}}$ are used. They are provided by ecoinvent
 208 (Wernet et al., 2016) 3.5 version released in august 2018. Ecoinvent edits three system
 209 models: At Point Of Substitution (APOS), Cut-off and Consequential; all three were run
 210 in the present work to assess the effect of modelling choices at the inventory level.

211 Ecoinvent also provides characterization factors (\mathbf{Q}) for several LCIA methods. Among
212 these, the LCIA methods are selected when normalization references are easily
213 available: CML 2000 (Guinée et al., 2002) (named CML 2001 by ecoinvent), EDIP 2003
214 (Hauschild and Potting, 2005), ILCD 2016 (EC-JRC, 2011), EF 2.0 (Fazio et al., 2018)
215 (European Environmental Footprint, named ILCD2018 by ecoinvent), ReCiPe midpoint
216 2008 (Goedkoop et al., 2013) (Egalitarian (E), Hierarchist (H) and Individualist (I)
217 versions, named ReCiPe V1.13 by ecoinvent, unfortunately 2016 version is not available
218 on the ecoinvent website and the use of the Simapro® software values would require a
219 parser to match the substance names) and TRACI (Bare, 2002) (first version). For CML
220 2000, EDIP 2003 and ILCD 2016, and ReCiPe 2008, normalization reference factors are
221 extracted from Simapro software (Pré Consultant, 8.5.2 version). EF 2.0 factors are
222 downloaded from JRC website (EC-JRC, 2018) and TRACI factors from the dedicated
223 article (Bare et al., 2006).

224 Ecoinvent provides characterization factors with and without long term emission
225 (Hellweg and Frischknecht, 2004) for CML 2000, EDIP 2003, ILCD 2016 and EF 2.0 and
226 ReCiPe (E and H versions). Both sets were used here for comparison.

227 Several normalization sets can be provided for some methods, according to the
228 perimeters of the inventories. Thus, normalization reference factors could be compared,
229 based on ecoinvent ($h_k^{\text{REF},\tilde{\mathbf{B}}}$), with those found in the literature (h_k^{REF}) for six methods
230 (eight methods considering ReCiPe I, H, and E versions as different methods) with a
231 total of 12 normalization sets. Four are available for CML 2000 (The Netherlands 1997,
232 Europe 1995, World 1995, World 1990) one for EDIP 2003, three for ILCD 2016 (EU-27
233 2010, EC-JRC Global, PROSUITE Global), one for EF 2.0 (only global factors are provided

234 on JRC website), two for ReCiPe (Eur 2000, World 2000) and one for TRACI.
235 Normalization set names are those used in Simapro.

236 **3. Results and discussions**

237 *3.1. Overview of intervention matrices*

238 An overview of intervention matrices is provided in supplementary information (see
239 Table SI.1). The zeros values are spread over a large number of elementary flows for
240 both \mathbf{B} and $\tilde{\mathbf{B}}$. Matrix \mathbf{B} is sparse, with 98 or 99% zero values, according to the system
241 model. This amount substantially decreases with the pre-calculated intervention matrix
242 $\tilde{\mathbf{B}}$. The smallest percentage of zero values is observed for the APOS system model
243 (3.6%). These observations justify the choice for dealing with $\tilde{\mathbf{B}}$ instead of \mathbf{B} for the
244 approach, as the estimations of geometric means are more accurate, with fewer zero
245 values.

246 Negative values remain an exception for \mathbf{B} and $\tilde{\mathbf{B}}$, except for the pre-calculated matrix of
247 the Consequential model where 15% of the interventions are negative. Note that
248 negative values mainly concern certain elementary flows for APOS and Cut-off models
249 (between 7 and 12 for 95% of negative values to be attained) but the Consequential
250 model has many, whether for \mathbf{B} or for $\tilde{\mathbf{B}}$ (1403 and 1547 to have 95% negative values
251 respectively). Hence, for the both first system models, $\tilde{\mathbf{B}}$ is almost completely positive
252 with $M_0(\tilde{\mathbf{B}}_j) \approx M_0(\tilde{\mathbf{B}}_j^+)$ and $M_0(\tilde{\mathbf{B}}_j^-) \approx 0$. Nevertheless, this is less applicable to the
253 Consequential system model where $M_0(\tilde{\mathbf{B}}_j^-)$ cannot be neglected.

254 *3.2. Geometric means of interventions.*

255 The estimation of geometric means of the pre-calculated intervention of ecoinvent
256 databases are all available online for download from

257 <https://doi.org/10.5281/zenodo.2598037> [NOTE TO THE REVIEWERS: this link will be
258 activated for the accepted version of the article, the file is available for the review process
259 here <https://www.dropbox.com/s/9wfgmmqo60i9elj/mu0%28Btild%29.xlsx?dl=0>
260 (anonymous download)]. Scatterplots are provided in supplementary material of the
261 article (see Figure SI.1). The estimations of geometric means from the three system
262 models reveal significant correlations. The highest is found between APOS and Cut-off
263 models with log-log Pearson coefficients of positive values equal to 0.91 (p-value \ll
264 0.001). The Consequential model is also correlated but to a lesser extent: 0.79 (p-value
265 \ll 0.001) with the APOS model and 0.89 (p-value \ll 0.001) with the Cut-off model.
266 These correlations make sense with respect to modelling principles, as APOS and Cut-
267 off models are both based on allocation procedures whereas the Consequential model
268 involves substitutions.

269 3.3. Ecoinvent-based normalization reference

270 Using the characterization factors provided by ecoinvent, 878 impacts have been
271 assessed. The corresponding normalization references $h_k^{\text{REF},\tilde{\mathbf{B}}}$ are downloadable at
272 <https://doi.org/10.5281/zenodo.2598035> [NOTE TO THE REVIEWERS: this link will be
273 activated for the accepted version of the article, the file is available for the review process
274 here <https://www.dropbox.com/s/wx7dt43ds7abi90/allHref.xlsx?dl=0> (anonymous
275 download).] From this table, entries associated to the selected methods with
276 normalization references provided from the literature (i.e. h_k^{REF}) have been extracted to
277 determine normalization references based on ecoinvent (i.e. $h_k^{\text{REF},\tilde{\mathbf{B}}}$) according to system
278 models, involving long-term emissions or not.

279 *3.4. Relevance of the estimates*

280 In a first approach, the log-log Pearson coefficients are provided in the supplementary
281 materials (see Figure SI.2). Generally, high correlations between normalization sets
282 fromecoinvent and from literature were observed. The correlation coefficients vary
283 between 0.77 (TRACI method versus APOS or Cut-off models) and 0.98 (various
284 methods). All are significant with p-values < 0.001, except for ReCiPe E and I vs the
285 Consequential system model that have p-values < 0.01 and TRACI (all system models)
286 where p-values < 0.05.

287 The log-log Pearson coefficient indicates a linear correlation between $h_k^{\text{REF},\tilde{\mathbf{B}}}$ and h_k^{REF} ,
288 but eq (10) presents a more restricted relationship: $\log(h_k^{\text{REF},\tilde{\mathbf{B}}}) \approx \log(h_k^{\text{REF}}) -$
289 $\log(c M_0(\mathbf{f}^{\text{REF}})) - \log(m)$. The parameter m is known and constant for a given system
290 model, whatever the impact. $c M_0(\mathbf{f}^{\text{REF}})$ is also constant but unknown. To address the
291 latter, $c M_0(\mathbf{f}^{\text{REF}})$ average values are calculated in the log-space for each comparison.
292 Values are provided in supplementary materials (Table SI.2). The average values are
293 similar for APOS (4.9×10^{-9} with, and 2.7×10^{-9} without long-term emissions) and
294 Cut-off (5×10^{-9} and 2.7×10^{-9}) models, but the Consequential model presents values
295 ten-fold smaller (0.5×10^{-9} and 0.3×10^{-9}). Values with long-term emissions are about
296 twice higher than those without long-term emissions except for EF 2.0 where values are
297 comparable.

298 The coefficient of variation of the estimated $c M_0(\mathbf{f}^{\text{REF}})$ (see Figure 1) describes the
299 accuracy of eq (10). The lower the coefficient of variation, the better eq (10) is
300 validated. The match between normalization sets from the literature and those built
301 from datasets excluding long-term emissions seems improved: the coefficient of

302 variation decreases when long-term emissions are not taken into account (average
303 value 3%). Only an exception remains for EF 2.0 normalization set vs the Consequential
304 model, describing a slight 0.1% increase. Regarding the estimation of $c M_0(\mathbf{f}^{\text{REF}})$, results
305 are similar between APOS and Cut-off models. The coefficient of variation is higher for
306 Consequential models except for the CML 2000 and TRACI methods.

307 When normalization sets are available for methods at both European and global levels
308 (ILCD 2016, ReCiPe 2008, CML 2000) the best fit with ecoinvent-based references is
309 found at the European level. Although ecoinvent inventories now cover the whole
310 world, the database was originally essentially focused on the Swiss and European
311 contexts. The global level is being increasingly addressed thanks to the addition of new
312 inventories; nevertheless, it is still frequently represented by extrapolations of
313 European technologies. Consequently, the large proportion of European inventories in
314 ecoinvent accounts for the better match amongst European normalization references.

315 [Insert Figure 1 here]

316 For a more detailed analysis of the adequacy between $h_k^{\text{REF},\tilde{\mathbf{B}}}$ and h_k^{REF} , a log-log
317 scatterplot can be drawn for each comparison. This was done for EF 2.0 without long-
318 term emissions in Figure 2. Scatterplots for EF 2.0 with long-term emissions and for the
319 other methods are available in supplementary materials. The reader can refer to them
320 during the following discussion.

321 These scatterplots indicate that a perfect adequacy between the reference system
322 traced out in the normalization literature and the reference determined by the
323 inventory database would be an alignment of all impact results (circles in Figure 2)

324 along a single line. This line represents the relationship with the average value for
325 $c M_0(\mathbf{f}^{\text{REF}})$ estimation.

326 Beyond the gap between $c M_0(\mathbf{f}^{\text{REF}})$ and $M_1(\mathbf{f}^{\text{REF}})$ and its consequences on the
327 relationship, a higher value either implies that the normalization literature
328 underestimates the impact or that the associated flows are overestimated in the
329 inventory database. Intuitively, this second possibility seems unlikely (since the main
330 issue with inventory data rather concerns data gaps) but can occur when obsolete
331 processes with high impacts are represented in the database. A lower value reflects the
332 opposite situation: either an overestimation of the impact in the normalization
333 literature or an underestimation of the involved intervention flows in the inventory
334 database. This second explanation appears to be more realistic, as the inventory data
335 may well be incomplete.

336 [Insert Figure 2 here]

337 3.4.1. EF 2.0 method

338 Most of the impacts are in the vicinity of the average line. According to Figure 2, the
339 APOS and Cut-off models have same result, whereas the Consequential model is
340 different. No substantial changes occur between system models involving or not long-
341 term emissions (see supplementary materials Figure SI.2) while the values of the
342 coefficients of variation are similar. This is a particular situation, as the other methods
343 present an increasing quality of the relationship excluding long-term emissions but it is
344 easily explained because EF 2.0 method is stated without long-term emission for human
345 toxicity and ecotoxicity impacts (Fazio et al., 2018). Between ILCD 2016 and EF 2.0 (the
346 latter being an update of the former), the impact characterization model for human
347 toxicities (both cancer and non-cancer effects) and for freshwater ecotoxicity is the

348 same USEtox® model (Rosenbaum et al., 2008). In ILCD 2016, two types of
349 characterization factors are considered: inclusion and exclusion of long-term emissions.
350 This directly affects the results (see below). In EF 2.0, values without long-term
351 emissions are also used for assessment with long-term emissions. There is therefore no
352 difference between them.

353 For APOS and Cut-off, the two most distant impacts situated beneath the line are Ozone
354 depletion and Land use. The first could be explained by incomplete inventories in the
355 databases rather than by an overestimation of the reference in the literature: ozone
356 depleting gases have been quantified for decades and estimations of country level
357 emissions seem robust. It is noteworthy that a similar result is reached for all other
358 methods (see below). For land use, there is most likely a lack of hindsight regarding the
359 application of the LANCA method (Bos et al., 2016) aggregated indicator to ecoinvent
360 datasets. This indicator has been selected for the latest version of the environmental
361 footprint. However, the normalization set for this impact is recent (Faragò et al., 2019)
362 and is characterized by global and spatialized data. Consequently, incomplete
363 inventories could influence this impact rather than an overestimation of the global
364 flows of the normalization reference (see discussion below about land transformation
365 in the ReCiPe 2008 method).

366 For the Consequential model, three other impacts deviate from the line: “ecotoxicity,
367 freshwater” and “ionizing radiation human health” are situated beneath the line;
368 “resource use, minerals and metals” are significantly above. The latter explains the high
369 coefficient of variation value for this system model (see Figure 1). Similar results for
370 resources with the Consequential model are found for ReCiPe methods (I, H and E)
371 where fossil and mineral resources are assessed separately. This similarity is not

372 observed when they are not differentiated (ILCD 2016 and CML 2000 methods). Indeed,
373 as the fossil resources seem less affected by the system model, they probably conceal
374 the effect of the latter on the mineral resources.

375 The “Ionizing radiation human health” impact is strongly correlated to the electricity
376 mix. For a rising market (i.e. electricity), products issued from modern and competitive
377 technologies, instead of established ones, are to be avoided in the Consequential model.
378 The Consequential electricity demand (having a low nuclear contribution) thus strongly
379 differs from the current electrical mix, which uses normalization sets from the
380 literature. It is therefore obvious that the match betweenecoinvent-based and
381 literature-based normalization is not satisfactory with this system model, as it reveals
382 an overestimation due to the literature-based reference.

383 3.4.2. ILCD 2016

384 The comparison of database-based normalization references with sets from literature
385 for the ILCD 2016 method show, as a general tendency, similar results as the EF 2.0
386 method. The main difference is a decrease in the coefficient of variation of about 3%
387 without long-term emissions. This is particularly significant for human toxicity (both
388 cancer, non-cancer effects) and freshwater ecotoxicity impacts: they do not deviate in
389 large proportion from the line without long-term emissions, whereas normalization sets
390 from the literature underestimate them when all emissions are considered. The global
391 inventories proposed in the literature thus seem to reflect best the inventories without
392 long-term emissions.

393 3.4.3. ReCiPe 2008.

394 Three versions of the ReCiPe 2008 method are available, the egalitarian (E) with a
395 1000-year time horizon for the impacts, the hierarchist (H) with a 100-year time

396 horizon and the individualist (I) with 20 years. The short time horizon of the latter
397 justifies the absence of characterization sets without long-term emissions. However, the
398 results for the I version are closest to those of the E and H versions which are assessed
399 with long term emissions. The I version only takes into account short-term impacts,
400 although these are also related to short- and long-term emissions. Consequently, the
401 ReCiPe I cannot be interpreted as a method without long-term emissions for
402 normalization.

403 Some results are similar to those of previously discussed methods: a decrease in the
404 coefficient of variation excluding long-term emissions, an ozone depletion beneath the
405 average line and underestimations of toxicity and ecotoxicity impacts by literature
406 when all emissions are considered. A further underestimation by the normalization
407 references has been observed for freshwater eutrophication.

408 Land transformation impacts are characterized by positive (from the previous
409 occupation) and negative (towards the subsequent occupation) characterization factors.
410 The ecoinvent based normalization references for natural land transformation are
411 negative with APOS and Cut-off models (this is never encountered elsewhere). They are
412 therefore not represented in the logscale scatterplots provided in supplementary
413 materials. As previously mentioned, ecoinvent has mainly been built using European
414 inventories. For this continent, natural land transformations are very low as they have
415 already been made during the previous millenniums, sometimes with re-naturalization,
416 with afforestation and with abandoned land (Fuchs et al., 2015). This situation is taken
417 into account in ecoinvent. The normalization references from literature are quite
418 different, with a world value two orders of magnitude higher than the European
419 reference. Therefore, concerning this type of impact, the database does not seem

420 representative at a global scale (in agreement with the conclusion on land use in the EF
421 2.0 method). Ecoinvent-based references are positive for this impact using the
422 Consequential model. The change in land use, introduced by agricultural avoided co-
423 products, leads to a good fit with the world reference set.

424 3.4.4. EDIP 2003

425 The list of impacts assessed here differs from previous methods but similar results are
426 found: The coefficients of variation decrease from the “with” to “without” long-term
427 emission situations, but within a lower range (-1.8%), with a better fit for toxicity and
428 ecotoxicity impacts. Ozone depletion sits beneath the line. With the Consequential
429 system model, radioactive wastes, associated to ionizing radiation impacts in the other
430 methods, are overestimated, as previously observed.

431 3.4.5. CML 2000

432 The main difference found between results from CML 2000 and from the other methods
433 is the increase in the coefficients of variation of the Consequential model. This can be
434 explained by (1) the association of mineral and fossil resource depletion into a unique
435 impact and (2) the lack of ionizing radiation in this method. The literature-based
436 normalization references for these two impacts poorly match the Consequential system
437 model for the other methods and increase the coefficients of variation; due to their
438 absence, this is not observed for CML 2000.

439 3.4.6. TRACI

440 The coefficients of variation are among the worst in Figure 1. Nevertheless, the
441 scatterplots available in supplementary materials provide insights for understanding
442 this result. Two subsets can be distinguished: the “carcinogenics”, “non-carcinogenics”
443 and “ecotoxicity” impacts are found well above the line with similar gap sizes, whereas

444 the other impacts apparently seem to follow the same line. As the TRACI reference
445 values are about 20 years old, toxic emissions are undoubtedly underestimated due to
446 lack of data at that time. In order to be aligned with the others, the normalization values
447 should be greater by almost three orders of magnitude. This range for deviations
448 corresponds to a proposed update for TRACI normalization (Kim et al., 2013), which
449 suggests values that are two (ecotoxicity) to four (carcinogenic) orders of magnitude
450 higher.

451 *3.5. Hypotheses needed for this approach*

452 3.5.1. Modelling and limits of the reference system

453 With eq (2), the reference system for the normalization is defined as a result of an LCA.
454 The functional unit would then be for example the one-year life for the world
455 population, or for a citizen. Conceptually, this is not an issue. The use of the same
456 modelling approach for both the studied system and the reference system can only be
457 expected. However, that differs from common normalisation practices where the
458 reference is established from external datasets, as national or regional statistics.

459 Consequently, the comparisons between the present approach and the references from
460 the literature (i.e. eq (10)) deal with the relevance of the approach, but also with the
461 completeness of the LCI database with respect to the exhaustivity of the method
462 reference sets. The first refers to the hypothesis on the means relationship (see below).
463 For the second aspect, the results show relevant matches and the discrepancies have
464 been discussed. Beyond proposing a new approach to normalization, this work is a way
465 of checking the exhaustiveness of the or the representativeness of the databases
466 relating to all human activities, at the same scale.

467 3.5.2. The geometric mean instead of the arithmetic mean

468 The constant difference between the geometric and the arithmetic means (eq. (5)) is the
469 main hypothesis of the approach. As f^{REF} is unknown, this cannot be verified but only
470 partially addressed looking at the arithmetic and geometric means of \tilde{B}_j . Figure SI.11
471 illustrates the high correlation between these means with the APOS system model (log-
472 log Pearson coefficients of positive values equal to 0.97 p-value <0.001), the Cut off
473 system model (0.89, p-value <0.001) and the Consequential system model, to a lesser
474 extent (0.70, p-value <0.001).

475 3.6. General discussion

476 The normalization step transforms results expressed at different scales to a common
477 one, pursuing two purposes.

478 The first is to free the study of the units of impacts. The aggregation and the comparison
479 of the result are then feasible. This aspect is not specific to LCA and can be found in
480 most of statistical treatments where values (samples) are positioned in relation to a
481 population (z-score, min-max normalisation, etc.). In LCA, the normalization is
482 conventionally done with external information. However, in data processing, it is
483 usually from the data that population descriptors are estimated. In line with this, the
484 present approach consists of using all available LCA data (the inventory database) and
485 describing the population in an adequate way (the geometric mean associated to the
486 log-normal distribution).

487 The second purpose is to compare the result with a reference quantity, checking its
488 consistency and facilitate its communication. The inventories-based normalization
489 obviously cannot be used directly for that, since the reference quantities do not have a
490 real meaning. This verification and communication work can be initiated by looking at

491 how the impacts are distributed in relation to others, but the present approach deals
492 more with the first aim.

493 This approach cannot produce relevant results if the database is not complete enough
494 and then does not accurately represent all human activities. But in this case the LCA
495 result itself, without normalization, is not relevant. The approach presented here makes
496 it possible to assess the quality of the involved database before the study, checking the
497 consistency with published normalization sets.

498 An LCA practitioner usually uses a pre-established inventory database for a given study.
499 He modifies and adds only a very small number of processes. The reference should
500 therefore not change significantly from one study to another and the same reference
501 should be able to be used, by proposing a normalization adapted to each system model
502 and corresponding to the same scope.

503 In addition to process-based LCA, Economic Input-Output LCA (EIO-LCA) allows an
504 assessment with respect to economic data. The final demand of a process is replaced by
505 the money spent in the associated economic sector, and the corresponding
506 environmental impact is attributed to the product under study. EXIOBASE 3 (Stadler et
507 al., 2018) is one of the most recent databases dedicated to. From a mathematical
508 perspective, the EIO-LCA result is determined in the same way as the process-based
509 LCA result. The approach can therefore be used. EIO databases contain only a few
510 substances, but it will be relevant to compare the reference values obtained under this
511 approach with cumulated impacts of all economic sectors, viewed as normalization
512 values.

513 4. Conclusion

514 Normalisation is typically done within the LCA framework through external
515 normalization references. They are related to the estimations of the sum of overall
516 process interventions which occur within the reference perimeter. Using an LCI
517 database to determine the reference values, the authors propose a normalization
518 approach for LCA that involves expressing the study result in relation to the geometric
519 average process. In this way normalization can be viewed as an approach for
520 determining whether the impacts of the system under study have a higher or lower
521 magnitude than for the other systems in the database. It compares the results with the
522 most expected values, i.e. the mean process of the database used.

523 Intervention flows in LCA do not follow normal laws, but are close to a log-normal
524 distribution. They spread out over several orders of magnitude and must be
525 investigated in the log-space. The choice of the geometric mean instead of the
526 arithmetic mean is therefore appropriate.

527 With this choice, it is not necessary to know how many times the processes are involved
528 to represent the overall impact: only the process interventions are required, and these
529 are currently described in LCI databases. This useful aspect of the geometric mean had
530 previously been highlighted in the LCA context for uncertainty assessments (Ciroth et
531 al., 2016).

532 Surprisingly good log-log matches are found between ecoinvent-based normalization
533 references according to the proposed approach and the external references currently
534 proposed in the literature. The main differences can be explained by examining the
535 features of the impacts. Therefore, thanks to these comparisons, plausibility checks can
536 be made to assess whether inventories are missing in the database or if the literature-

537 based normalization references underestimate the impact. These results confirm that
538 the proposed approach will not lead to completely different conclusions than the
539 existing practice, while enhancing data quality and consistency between study results
540 and normalization data.

541 The present approach proposes to define a normalization procedure where the same
542 datasets are used to assess the system under study as well as the normalization
543 reference. It avoids discrepancies between both in relation to the system model, to the
544 system boundaries, to the data sources and to the considered environmental flows. It
545 therefore provides a solution to all the potential limitations of normalization listed by
546 Pizzol et al. (2017). However, the meaning of normalization in the present approach is
547 slightly different. The result is no longer expressed in relation to the sum of
548 interventions of the reference, as usually done, but according to the mean of
549 intervention of the reference elements. Although the approach is not completely
550 identical, the purpose remains nevertheless the same and the results are always
551 expressed in relation to the reference. This can be seen even as an advantage of the
552 more theoretically grounded proposal as over-interpretation of normalization results in
553 the sense of “share in a certain region” are implicitly avoided.

554 This LCI database-based normalization approach is fully operational for all LCIA
555 methods and databases. It allows therefore a broader coverage of studies than the
556 existing external normalization references. Normalization references are provided here
557 for ecoinvent, but this can be easily extended to other databases. The approach is
558 transparent and ensures consistency of the normalized result. Practitioners are
559 encouraged to test it and use it to compare, discuss and potentially aggregate
560 environmental impacts in LCA.

561 **Acknowledgment**

562 Authors are grateful to the ecoinvent Centre for providing the unit version of the
563 databases.

564 **References**

565 Bare, J., Gloria, T., Norris, G., 2006. Development of the Method and U.S. Normalization
566 Database for Life Cycle Impact Assessment and Sustainability Metrics. *Environ. Sci.*
567 *Technol.* 40, 5108–5115. <https://doi.org/10.1021/es052494b>

568 Bare, J.C., 2002. Traci. The Tool for the Reduction and Assessment of Chemical and
569 Other Environmental Impacts. *J. Ind. Ecol.* 6, 49–78.
570 <https://doi.org/10.1162/108819802766269539>

571 Benini, L., Sala, S., 2016. Uncertainty and sensitivity analysis of normalization factors to
572 methodological assumptions. *Int. J. Life Cycle Assess.* 21, 224–236.
573 <https://doi.org/10.1007/s11367-015-1013-5>

574 Bos, U., Horn, R., Beck, T., 2016. LANCA ® Characterization Factors for Life Cycle Impact
575 Assessment.

576 Broadbent, C., Stevenson, M., Caldeira-Pires, A., Cockburn, D., Lesage, P., Martchek, K.,
577 Réthoré, O., Frischknecht, R., 2011. Chapter 3. Aggregated data development, in:
578 Sonnemann, G., Vigon, B. (Eds.), *Global Guidance Principles for Life Cycle*
579 *Assessment Databases - A Basis for Greener Processes and Products 'Shonan*
580 *Guidance Principles.'* UNEP/SETAC Life Cycle Initiative, pp. 67–84.

581 Ciroth, A., Muller, S., Weidema, B., Lesage, P., 2016. Empirically based uncertainty
582 factors for the pedigree matrix in ecoinvent. *Int. J. Life Cycle Assess.* 21, 1338–1348.

583 <https://doi.org/10.1007/s11367-013-0670-5>

584 Crenna, E., Secchi, M., Benini, L., Sala, S., 2019. Global environmental impacts: data
585 sources and methodological choices for calculating normalization factors for LCA.
586 Int. J. Life Cycle Assess. <https://doi.org/10.1007/s11367-019-01604-y>

587 Cucurachi, S., Seager, T.P., Prado, V., 2017. Normalization in Comparative Life Cycle
588 Assessment to Support Environmental Decision Making. J. Ind. Ecol. 21, 242–243.
589 <https://doi.org/10.1111/jiec.12549>

590 de la Cruz, R., Kreft, J.-U., 2018. Geometric mean extension for data sets with zeros.
591 arXiv:1806.06403. <https://doi.org/http://arxiv.org/abs/1806.06403>

592 EC-JRC, 2018. Join Research Center - European plateforme on life cycle assessment
593 [WWW Document]. URL <http://eplca.jrc.ec.europa.eu> (accessed 1.16.19).

594 EC-JRC, 2011. International Reference Life Cycle Data System (ILCD) Handbook:
595 Recommendations for Life Cycle Impact Assessment in the European context EUR
596 24571 EN - 2011, European Commission. Publications Office of the European
597 Union, Luxembourg. <https://doi.org/10.278/33030>

598 Esnouf, A., Heijungs, R., Coste, G., Latrille, É., Steyer, J.-P., Hélias, A., 2019. A tool to guide
599 the selection of impact categories for LCA studies by using the representativeness
600 index. Sci. Total Environ. 658, 768–776.
601 <https://doi.org/10.1016/j.scitotenv.2018.12.194>

602 Esnouf, A., Latrille, É., Steyer, J.-P., Helias, A., 2018. Representativeness of environmental
603 impact assessment methods regarding Life Cycle Inventories. Sci. Total Environ.
604 621, 1264–1271. <https://doi.org/10.1016/j.scitotenv.2017.10.102>

605 Faragò, M., Benini, L., Sala, S., Secchi, M., Laurent, A., 2019. National inventories of land
606 occupation and transformation flows in the world for land use impact assessment.
607 Int. J. Life Cycle Assess. <https://doi.org/10.1007/s11367-018-01581-8>

608 Fazio, S., Castellani, V., Sala, S., Schau, E.M., Secchi, M., Zampori, L., Diaconu, E., 2018.
609 Supporting information to the characterisation factors of recommended EF Life
610 Cycle Impact Assessment method. EUR 28888 EN, European Commission,
611 JRC109369, Ispra. <https://doi.org/10.2760/671368>

612 Finkbeiner, M., Inaba, A., Tan, R., Christiansen, K., Klüppel, H.-J., 2006. The New
613 International Standards for Life Cycle Assessment: ISO 14040 and ISO 14044. Int. J.
614 Life Cycle Assess. 11, 80–85. <https://doi.org/10.1065/lca2006.02.002>

615 Fuchs, R., Herold, M., Verburg, P.H., Clevers, J.G.P.W., Eberle, J., 2015. Gross changes in
616 reconstructions of historic land cover/use for Europe between 1900 and 2010.
617 Glob. Chang. Biol. 21, 299–313. <https://doi.org/10.1111/gcb.12714>

618 Goedkoop, M., Heijungs, R., Huijbregts, M., De Schryver, A., Struijs, J., Van Zelm, R., 2013.
619 ReCiPe 2008 - First edition (revised) - Report I: Characterisation. The Hague,
620 Netherland.

621 Guinée, J., Gorrée, M., Heijungs, R., Huppel, G., Kleijn, R., de Koning, A., Oers, L., Wegener
622 Sleeswijk, A., Suh, S., de Haes, H., Bruijn, H., van Duin, R., Huijbregts, M. (Eds.), 2002.
623 Handbook on LCA, operational guide to the ISO standards. Kluwer Academic
624 Publishers, Dordrecht.

625 Hauschild, M.Z., Potting, J., 2005. Spatial differentiation in Life Cycle impact assessment -
626 The EDIP2003 methodology. Danish ministry of the environment, Environmental
627 protection agency.

628 Heijungs, R., Guinée, J., Kleijn, R., Rovers, V., 2007. Bias in normalization: Causes,
629 consequences, detection and remedies. *Int. J. Life Cycle Assess.* 12, 211–216.
630 <https://doi.org/10.1065/lca2006.07.260>

631 Heijungs, R., Suh, S., 2002. *The Computational Structure of Life Cycle Assessment, Eco-*
632 *Efficiency in Industry and Science.* Springer Netherlands, Dordrecht.
633 <https://doi.org/10.1007/978-94-015-9900-9>

634 Hellweg, S., Frischknecht, R., 2004. Evaluation of Long-Term Impacts in LCA. *Int. J. Life*
635 *Cycle Assess.* 9, 339–341. <https://doi.org/10.1007/BF02979427>

636 ISO, 2006a. 14040:2006, Environmental management — Life cycle assessment —
637 Principles and framework, CEN Central secretariat. International Organization for
638 Standardization.

639 ISO, 2006b. 14044:2006 Environmental management — Life cycle assessment —
640 Requirements and guidelines, CEN Central secretariat. International Organization
641 for Standardization.

642 Kim, J., Yang, Y., Bae, J., Suh, S., 2013. The Importance of Normalization References in
643 Interpreting Life Cycle Assessment Results. *J. Ind. Ecol.* 17, 385–395.
644 <https://doi.org/10.1111/j.1530-9290.2012.00535.x>

645 Laurent, A., Hauschild, M.Z., 2015. Normalisation, in: Hauschild, Michael Zwicky,
646 Huijbregts1, M.A.J. (Eds.), *Life Cycle Impact Assessment.* Springer Netherlands,
647 Dordrecht, pp. 271–300. https://doi.org/110.1007/978-94-017-9744-3_14

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

650 <https://doi.org/10.1007/s11367-018-1444-x>

651 Pizzol, M., Laurent, A., Sala, S., Weidema, B., Verones, F., Koffler, C., 2017. Normalisation
652 and weighting in life cycle assessment: quo vadis? *Int. J. Life Cycle Assess.* 22, 853–
653 866. <https://doi.org/10.1007/s11367-016-1199-1>

654 Qin, Y., Suh, S., 2017. What distribution function do life cycle inventories follow? *Int. J.*
655 *Life Cycle Assess.* 22, 1138–1145. <https://doi.org/10.1007/s11367-016-1224-4>

656 Rosenbaum, R.K., Bachmann, T.M., Gold, L.S., Huijbregts, M.A.J., Jolliet, O., Juraske, R.,
657 Koehler, A., Larsen, H.F., MacLeod, M., Margni, M., McKone, T.E., Payet, J.,
658 Schuhmacher, M., van de Meent, D., Hauschild, M.Z., 2008. USEtox—the UNEP-
659 SETAC toxicity model: recommended characterisation factors for human toxicity
660 and freshwater ecotoxicity in life cycle impact assessment. *Int. J. Life Cycle Assess.*
661 13, 532–546. <https://doi.org/10.1007/s11367-008-0038-4>

662 Sala, S., Benini, L., Mancini, L., Pant, R., 2015. Integrated assessment of environmental
663 impact of Europe in 2010: data sources and extrapolation strategies for calculating
664 normalisation factors. *Int. J. Life Cycle Assess.* [https://doi.org/10.1007/s11367-](https://doi.org/10.1007/s11367-015-0958-8)
665 015-0958-8

666 Stadler, K., Wood, R., Bulavskaya, T., Södersten, C.J., Simas, M., Schmidt, S., Usubiaga, A.,
667 Acosta-Fernández, J., Kuenen, J., Bruckner, M., Giljum, S., Lutter, S., Merciai, S.,
668 Schmidt, J.H., Theurl, M.C., Plutzar, C., Kastner, T., Eisenmenger, N., Erb, K.H., de
669 Koning, A., Tukker, A., 2018. EXIOBASE 3: Developing a Time Series of Detailed
670 Environmentally Extended Multi-Regional Input-Output Tables. *J. Ind. Ecol.* 00.
671 <https://doi.org/10.1111/jiec.12715>

672 Weidema, B.P., 2018. In Search of a Consistent Solution to Allocation of Joint Production.

673 J. Ind. Ecol. 22, 252–262. <https://doi.org/10.1111/jiec.12571>

674 Wernet, G., Bauer, C., Steubing, B., Reinhard, J., Moreno-Ruiz, E., Weidema, B., 2016. The
675 ecoinvent database version 3 (part I): overview and methodology. Int. J. Life Cycle
676 Assess. 21, 1218–1230. <https://doi.org/10.1007/s11367-016-1087-8>

677

678

679 **Table captions:**

680 Table 1. List of notations

681

682 **Figure Captions**

683 Figure 1. Coefficient of variation in the log-space of $c M_0(\mathbf{f}^{\text{REF}})$ estimations, determined
684 according to methods with an associated normalization set and to the ecoinvent system
685 models. Unaggregated values are depicted by blue histograms. The average coefficients
686 of variation are in green for ecoinvent system models and for method normalization
687 sets. The overall average is represented in red. When possible, differences in the
688 coefficient of variation between methods with (All) and without long term emission (No
689 LT) are indicated in square brackets.

690

691 Figure 2. Scatterplots for EF 2.0 method according to normalization sets from literature
692 (h_k^{REF} , x-axis) and system models ($h_k^{\text{REF},\bar{\mathbf{B}}}$, y-axis) without long-term emission. The lines
693 show the relationship $h_k^{\text{REF},\bar{\mathbf{B}}} = \frac{1}{m_{M_0(\mathbf{f}^{\text{REF}})}} h_k^{\text{REF}}$, see supplementary materials for
694 $c M_0(\mathbf{f}^{\text{REF}})$ values.

695

696

697

Notation	Description
m, i	Index: number of processes, i^{th} process ($i = 1, \dots, m$).
n, j	Index: number of elementary flows, j^{th} elementary flow ($j = 1, \dots, n$).
p, k	Index: number of impact categories, k^{th} impact category ($k = 1, \dots, p$).
\mathbf{f}, f_i	Final demand vector ($m \times 1$), final demand of the i^{th} process.
$\mathbf{f}^{\text{REF}}, f_i^{\text{REF}}$	Final demand vector for the normalization reference system ($m \times 1$), final demand of the i^{th} process for the normalization reference system.
\mathbf{A}	Technology matrix ($m \times m$).
$\mathbf{B}, \mathbf{B}_j, b_{j,i}$	Intervention matrix ($n \times m$), intervention vector for the j^{th} elementary flow ($1 \times m$), intervention value of the j^{th} elementary flow for the i^{th} process.
$\tilde{\mathbf{B}}, \tilde{\mathbf{B}}_j, \tilde{b}_{j,i}$	Pre-calculated intervention matrix ($n \times m$): $\tilde{\mathbf{B}} = \mathbf{B}\mathbf{A}^{-1}$, pre-calculated intervention vector for the j^{th} elementary flow ($1 \times m$): $\tilde{\mathbf{B}}_j = \mathbf{B}_j\mathbf{A}^{-1}$, pre-calculated intervention value of the j^{th} elementary flow for the i^{th} process.
$\tilde{\mathbf{B}}^+, \tilde{\mathbf{B}}_j^+, \tilde{b}_{j,i}^+$	Pre-calculated intervention matrix with only zeros and the positive values of $\tilde{\mathbf{B}}$ ($n \times m$), pre-calculated intervention vector with only zeros and the positive values of $\tilde{\mathbf{B}}_j$ ($1 \times m$), corresponding pre-calculated intervention value of the j^{th} elementary flow for the i^{th} process.

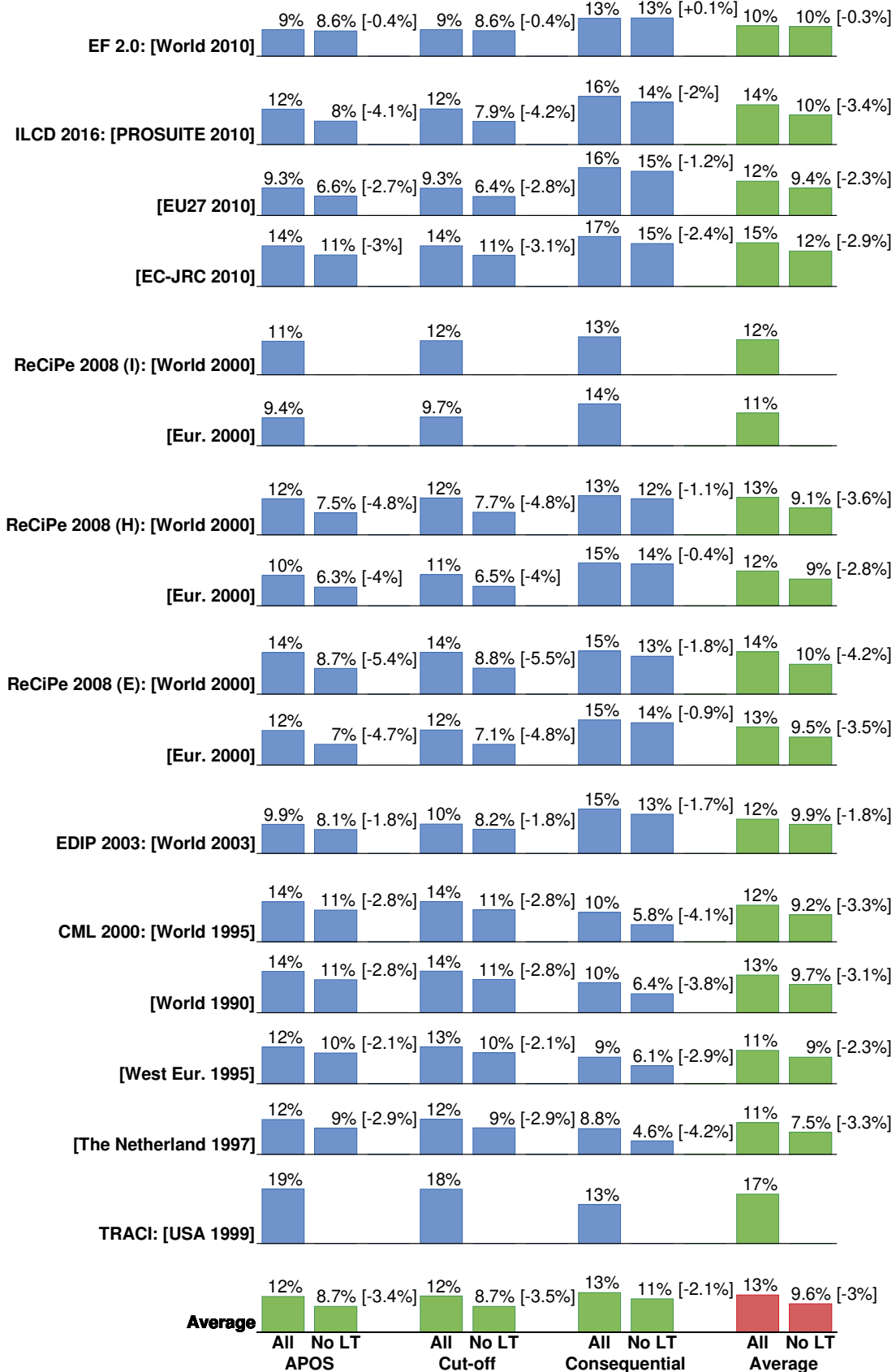
$\tilde{\mathbf{B}}^-, \tilde{\mathbf{B}}_j^-, \tilde{b}_{j,i}^-$	Pre-calculated intervention matrix with only zeros and the additive inverse of negative values of $\tilde{\mathbf{B}}$ ($n \times m$), pre-calculated intervention vector with only zeros and the additive inverse of negative values of $\tilde{\mathbf{B}}_j$ ($1 \times m$), corresponding pre-calculated intervention value of the j^{th} elementary flow for the i^{th} process.
$\mathbf{Q}, \mathbf{Q}_k, q_{k,j}$	Characterization matrix ($p \times n$), characterization vector of the k^{th} impact ($1 \times n$), characterization factor of the k^{th} impact category for the j^{th} elementary flow.
\mathbf{h}, h_k	LCIA result ($p \times 1$) $\mathbf{h} = \mathbf{Q}\tilde{\mathbf{B}}\mathbf{f}$, LCIA result on the k^{th} impact category.
$\mathbf{h}^{\text{REF}}, h_k^{\text{REF}}$	LCIA result of the normalization reference system ($p \times 1$) $\mathbf{h}^{\text{REF}} = \mathbf{Q}\tilde{\mathbf{B}}^{\text{REF}}\mathbf{f}^{\text{REF}}$, LCIA result of the normalization reference system on the k^{th} impact category.
$\mathbf{h}^{\text{REF},\tilde{\mathbf{B}}}, h_k^{\text{REF},\tilde{\mathbf{B}}}$	Normalization reference value based on $\tilde{\mathbf{B}}$ and \mathbf{Q} only ($p \times 1$), normalization reference value based on $\tilde{\mathbf{B}}$ and \mathbf{Q}_k only for the k^{th} impact category.
$\mathbf{h}^{\text{N}}, h_k^{\text{N}}$	Normalized LCIA result ($p \times 1$), normalized LCIA result of the k^{th} impact category.
$M_1(\dots)$	Arithmetic means of all elements in brackets, mean of a vector has to be interpreted as the mean of all elements of the vector.
$M_0(\dots)$	Geometric means of all elements in brackets, mean of a vector has to be interpreted as the mean of all elements of the vector.

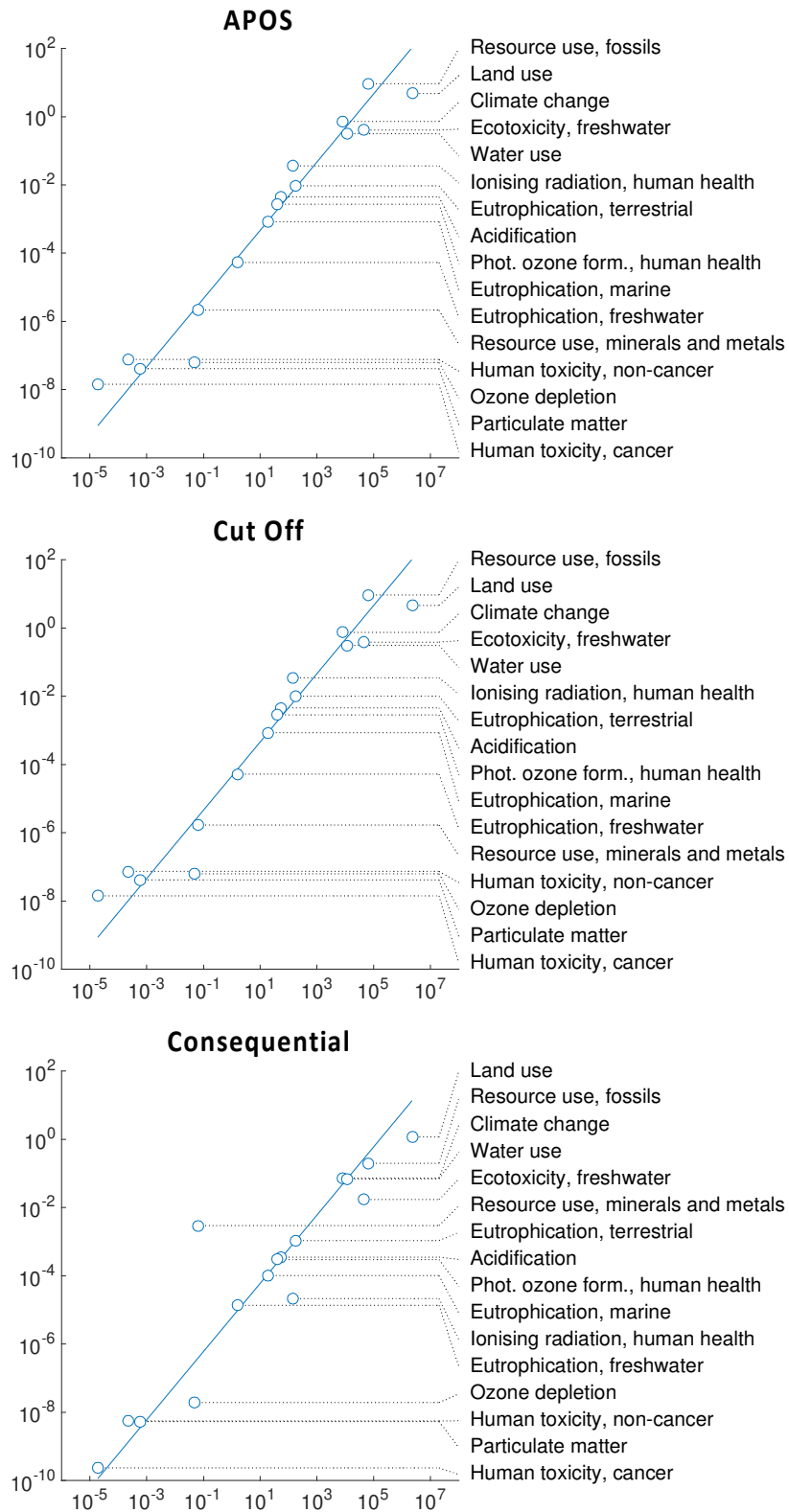
c Stated constant for the relationship between geometric mean and arithmetic mean.

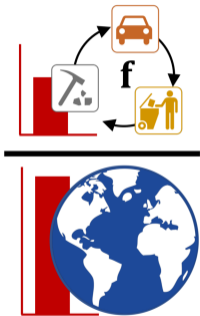
699

700

701 Figure 1







$$\frac{\sum_{i=1}^m \left(f_i \times \text{LCI} \right)}{\left(\prod_{i=1}^m \left(\text{LCI} \right) \right)^{\frac{1}{m}}}$$