Title: Where that emissions went? Characterising the uncertainty of characterisation factors of underspecified emissions.

Authors: Miguel F. Astudillo

Corresponding Author: miguel.astudillo@lca-net.com

Conflict of Interest Statement: The author declares no conflict of interest

Keywords: industrial ecology, LCA, archetype, subcompartment, uncertainty, characterization factor

Abstract: Life cycle impact assessment (LCIA) methods use unspecified archetypes to model situations of imperfect knowledge. There is an inherent uncertainty in characterisation factors to unspecified archetypes, and this study proposes an estimation of this uncertainty using continuous and discrete probability distributions. The extent of “archetype uncertainty” is analysed for several methods and the impact on LCIA scores is quantified for all activities in the database ecoinvent. Results indicate that this source of uncertainty can be very large, introducing systematic as well as random errors in LCIA scores. Based on this research we recommend to use undefined archetypes only when needed and quantify this source of uncertainty by default. The continuous and discrete approaches to model uncertainty give similar results, but the continuous approach is easier to implement.

1. INTRODUCTION

Life cycle impact assessment (LCIA) methods model different exposure situations using archetypal scenarios such as emissions to *indoor air* or *air in urban area*. The characterisation factors (CFs) can vary widely between different archetypal scenarios of release, given the very different fate and exposure factors. For this reason archetypes are an effective way to account for the variability of CFs (Mutel et al. 2018). For example, the carcinogenic impacts of cadmium emitted to soils can vary three orders of magnitude depending on whether it is emitted to agricultural or industrial soils (Fazio et al. 2018). Regionalised models would need a very high spatial resolution to capture the differences between e.g. rural and urban settings, therefore archetypes are considered a pragmatic approach to capture CF variability (Bulle et al. 2019). The archetypal scenarios are sometimes called “context”, dividing them into primary context and secondary context (Edelen et al. 2019). The primary and secondary contexts are commonly referred to as compartment and subcompartment in LCA software.

LCA software allows to have “unspecified” archetypes, and the CFs for these cases are sometimes defined by modeller developers other times by those who implement the model, but the reasoning behind the choice is rarely explained and approaches diverge. In OpenLCA and Simapro the CFs to unspecified archetypes are used to fill data gaps (Acero et al. 2017; Pré 2019). Brightway2 uses the implementation of LCIA methods maintained by ecoinvent. The ecoinvent implementation uses different approaches depending on the method (Bourgault 2019). For example, for ILCD method, the CFs for unspecified archetypes are sometimes used as a proxy to fill data gaps, but the documentation does not clarify how values for unspecified archetypes are chosen when not provided by the method developer (Fazio et al. 2018). Overall, there is no consensus on how to assign CFs to unspecified archetypes.

Even under the assumption that the chosen CFs for unspecified archetypes are a reasonable approximation, these CFs are inherently uncertain. The uncertainty associated with the use of unspecified archetypes is to my knowledge not accounted for in LCA studies, potentially leading to misleading conclusions. Several studies have analysed the variability of CFs. Bulle et al (2019) compared the variability of CFs between archetypes and regions for the method Impact World +. Impacts per kg of PM_{2.5} in the urban environment were found to be a factor of 30 higher than in the rural environment, and the archetype was found to be more important than the continent in which the emission took place (Bulle et al. 2019). Some studies have included CF uncertainty in the calculation of LCA scores. Henriksson et al. (2015) included uncertainty on global warming and freshwater ecosystem impacts. Recently, Patouillard et al. (2019) included spatial variability of regionalized CFs in their approach to prioritise regionalization efforts. Their study used the regionalised method impact world + (Bulle et al. 2019) and modelled the spatial variability with a four-parameter beta distribution, using proxies of probability for an elementary flow to occur in each region (Patouillard et al. 2019). However, no study seems to have addressed yet what is the uncertainty of unspecified subcompartments and how it affects LCA scores.

In this study, I propose a generic procedure to define the CFs for undefined archetypal scenarios as a probability distribution. The uncertainty associated with uncertain archetypes can be considered to be a source of epistemic uncertainty (i.e. uncertainty associated to lack of knowledge). The study could be framed as a study of uncertainty due to archetypal aggregation (UNEP-SETAC 2019). The procedure is implemented using the open-source LCA framework Brightway 2 (Mutel 2017). The procedure is applicable to all methods that use archetypes in their modelling. The modified impact assessment methods are then tested with activities of the LCA database ecoinvent 3.6 (Wernet et al. 2015) to assess the effects on widely used LCA databases.

2. Materials and methods

2.1. Uncertainty characterization

The functions developed for this study take an impact assessment method as an input, and return a version of the method where the elementary flows to uncertain archetypes have their CFs defined as a probability distribution instead of a scalar value.

The CF for a given substance and “compartment” and undefined archetype is logically a value within the range of CFs for defined archetypes. This range is used to define the minimum and maximum of the probability distribution. If the minimum and maximum have a relative difference of less than 0.1 %, we attribute the difference to rounding errors and assume a range of 0 (i.e. no uncertainty). The CFs used in the original implementation of the method is used as an estimator of the mode. The functions allow choosing between the uniform and the triangular probability distribution to model the archetype uncertainty. Using the uniform distribution we take the position that we do not know anything about the CFs of unspecified archetypes apart of their upper and lower bounds. In the implementation of the triangular distribution, the mode is equal to the value of the CF for the unspecified archetype. The triangular modelling gives weight to the choice of the CF for uncertain archetypes given in the original implementation of the impact category. The uniform and triangular distributions are

chosen because, in the absence of more information, they are the simplest and introduce fewer assumptions than other distributions, following thus the principle of insufficient reason.

Another option is to describe archetype uncertainty as a discrete (categorical) probability distribution. For example, if a substance has CF for air, in rural and urban archetypes of 1 and 10 respectively, the CF to the uncertain archetype would have possible values of 1 or 10. Different weights could be given to different CFs, depending on the likelihood of occurrence. Furthermore, CFs from specified archetypal scenarios could be uncertain themselves, compounding different sources of uncertainty such as spatial variability. Discrete probability distributions are not natively supported by Brightway2 but can be handled with the package *presamples* (Lesage et al. 2018) to modify the matrix of characterisation factors. With this approach, the CFs of substances to underspecified archetypes are taken from a list of possible values in each iteration. As for the continuous case, The modelling follows the principle of insufficient reason, and assign equal probability to each CF. With more information about the likelihood of occurrence, this approach can calculate CFs more accurately.

2.2. Assessing the uncertainty of existing impact assessment methods

The uncertainty of CFs to unspecified archetypes is analysed for several impact assessment methods. The analysis explores how CFs to uncertain archetypes compare to the values of specified archetypes. The analysis is only done for the methods currently maintained by ecoinvent and for which the concept of archetype is relevant. The latter criterion excludes methods where CFs do not change with the archetype (e.g. the current implementation of global warming potentials of greenhouse gases). For illustrative purposes, the methods ReCiPe 2008 (Goedkoop et al. 2013), Environmental Footprint 2.0 (Fazio et al. 2018) and ecological scarcity 2013 (Frischknecht and Büsser Knöpfel 2014) are analysed.

2.3. Testing the methods on background LCA databases

The methods are tested with activities of ecoinvent 3.6 (consequential), in an attempt to identify activities affected by *archetype uncertainty* and the extent of these effects. For simplicity, only the impact assessment method Environmental Footprint 2.0 is used, as it is one of the few maintained by ecoinvent and recently updated.

The analysis only evaluates uncertainty of the CFs, ignoring uncertainty coming from the inventory (i.e. flows in technology and intervention matrices). This simplification is done because analysing both sources of uncertainty for all activities of the database would be excessively computationally demanding from a normal laptop. We use the concept of intensity matrix proposed by (Heijungs and Suh 2002) defined as $\Lambda = BA^{-1}$ where A is the technology matrix and B is the intervention matrix, each column of A and B represent an activity in ecoinvent. The intensity matrix is calculated only once while the matrix of CFs is updated in each iteration. Calculating the inverse of A for databases such as ecoinvent, with more than 15000 activities is computationally demanding, and with this approach, we only do it once. Uncertainty is propagated using the Monte Carlo technique.

For a limited number of cases, archetype uncertainty was combined with inventory uncertainty. In each iteration, the technology, intervention, and intensity matrix are calculated, and differences come from the matrix of CFs. This is done to avoid spurious uncertainty due to correlation.

3. RESULTS

The results first explore to what extent impact assessment methods are affected by archetype uncertainty, using as example three case studies. The results follow with the impact of archetype uncertainty in the LCA database ecoinvent.

3.1. Characterization factors of undefined archetypes

Fig.1 illustrates the normalised CFs of flows to unspecified archetypes relative to the range of values of the same substance and compartment to specified archetypes. The normalization bounds the CFs in the interval [0,1], being 1 the maximum observed CF for the defined archetypes and 0 the minimum. The value in the label is the maximum range observed for that impact category normalised by the median CF in that impact category. The larger the range the more important is the definition of subcompartment.

The highest inter-subcompartment CF variability” for ReCiPe 2008 is found for human toxicity, where the difference between minimum and maximum can be up to eight orders of magnitude higher than the median CF, illustrating the relevance of a well-defined subcompartment. Overall, it seems toxicity impact categories are the most sensible to archetypes.

In some impact categories, flows to unspecified subcompartments have a CF equal to the maximum or the minimum of CFs of the same substance and compartment, but in other cases is an intermediate value. Given the diversity of assumptions, it is only possible to say whether underspecification of archetypes results in an overestimate or underestimate impacts for certain impact categories. For example, CFs of uncertain archetypes for the impact category *respiratory inorganics* in *Environmental Footprint 2.0* is equal to the maximum of the observed range of values (Fig.1). Therefore, for this category using unspecified archetypes should overestimate impacts. It is also apparent that methods do not use a coherent approach to quantify the CF of undefined archetypes for all their impact categories. Fig.1 only shows impact categories with uncertainty associated to archetypes, the rest of categories do not have any archetype uncertainty. Furthermore, it was also found that the list of possible elementary flows and emissions scenario is slightly inconsistent ¹.

¹ There seems to be an incomplete definition of unspecified archetypes in the list of possible triples (elementary flow, compartment, subcompartment) in Brightway2 (which derives this list from ecoinvent). Many pairs “elementary flow – compartment” do not have an “undefined archetype” case. In most of the cases where the unspecified case is not defined, these pairs only have one subcompartment defined, which stands to reason. However, in some cases, several archetypal emission scenario exist, without an “underspecified” archetypal option.

Fig. 1 Normalized CF of elementary flows to unspecified archetypes, for impact categories with archetype uncertainty. Each point in the figure represents a CF. The maximum observed range is added in parenthesis.

3.2. Effects of archetype uncertainty on LCIA scores

This section assesses the effects of introducing uncertainty from underspecified archetypes in the LCA scores of activities in the database ecoinvent. The mean absolute percentage deviation (MAPD) is used as a metric to compare the effects across all activities in the database. To compare different approaches, the case with no uncertainty is used as a reference score (S_{ref}) while the cases with uncertainty are considered as alternative scores (S_{alt}). The MAPD in this context should not be interpreted as a measure of error (it is unclear which approach provides scores closer to the *true value*) but as a measure of deviation from the reference case.

$$\frac{1}{n} \sum \left| \frac{S_{ref} - S_{alt}}{S_{ref}} \right|$$

Archetype uncertainty can increase the overall uncertainty of LCA scores without introducing systematic errors but introducing random errors instead. An increase in uncertainty would make scores less discernible. In other words, it would be more difficult to ascertain if two comparable products have “significantly” different scores. The “random errors” introduced by archetype uncertainty are assessed using the normalised standard deviation, also known as coefficient of variation, using the absolute value of the mean as normalising factor.

3.2.1. systematic and random errors

The sensitivity of ecoinvent to archetype uncertainty are represented in Fig. 2 and 3. Fig 2 displays the MAPD with a cut-off of 0.1% (i.e. deviations below 0.1% are not displayed). Results are skewed, the MAPD is below 1% for around 64% of the activities and scores, while for 0.5% of activities deviations are higher than 100%. Higher systematic errors are consistently observed for the category of *respiratory effects, inorganics*. The categories of *ionising radiation*, *freshwater and terrestrial acidification* and *terrestrial eutrophication* do not have significant uncertainty and are not displayed.

The MAPD is lower when the archetype uncertainty is assumed to follow a triangular distribution. This is an expected result as MAPD measures differences with the reference case, and the reference case sets the mode of the triangular distribution.

Fig. 2: Systematic errors due to archetype uncertainty: mean absolute percentage deviation for all the activities in Ecoinvent 3.6 consequential. MAPDs are ordered by the “uniform” case. Each element in the x axis represents an activity in the database.

The coefficient of variation is presented in Fig. 3 for the cases of uniform, triangular and discrete distribution. Fig. 3 has a cut-off of 0.1%, meaning that coefficients of variation below that value are considered not relevant and not displayed. *Respiratory inorganics* has the most consistent random errors (i.e. highest median coefficient of variation), followed by *non-carcinogenic effects*. The coefficient of variation is low for most of the activities but high for some. The coefficient of variation is always lower if we assume uncertainty follows a triangular distribution. This is also expected because for distributions with the same range, standard deviations of uniform distributions are always higher than for triangular distributions.

Fig.3 Uncertainty introduced by unspecified archetypes on LCA scores measured as relative standard deviation for Ecoinvent 3.6 consequential. The scores are ordered by the uniform case.

An exploratory analysis indicates that some regions have higher archetype uncertainty than others. The UN-Oceania region has the highest followed by Indian regions (see supplementary information for details). As for MAPD, it is difficult to explain why some categories or regions have larger uncertainty than others. The uncertainty depends on how uncertain CFs in a given impact category are, and how often inventories use underspecified archetypes. The latter depends on editorial guidelines and how much it is known about the scenario under which these emissions occur.

For the cases with highest archetype uncertainty a Monte Carlo simulation was done, compounding CF uncertainty with inventory uncertainty. The addition of CF uncertainty increases the MAPD with respect to the deterministic calculation, and in the tested cases the MAPD increased up to 43 % (see supplementary information for details). We note that cases with the highest archetype uncertainty do not necessarily mean cases where archetype uncertainty is more relevant than inventory uncertainty.

3.2.2. Effects of distribution choice

It can be observed that the continuous modelling of uncertainty and the discrete approach render similar results. As previously stated, the uniform distribution assumption deviates more from the reference case than the triangular but no clear pattern is observed for the discrete case. The discrete case is as expected “less smooth”. Overall, the three approaches are broadly comparable, at least under a scenario where no additional information is available.

4. DISCUSSION

Unspecified archetypes are a powerful mental construct when the modeller is aware of its implications but have potentially disruptive side-effects. Thus, it would be better if studies systematically account for this uncertainty. This study proposes an implementation in the open-source LCA software Brightway2 (Mutel 2017). The implementation is licensed with a permissive license and can be reused and improved by the research community. Although the consideration of uncertainty in CFs is uncommon, software providers should consider uncertainty in CFs.

With regards to the modelling choice between continuous, uniform, and triangular distributions, the discrete case introduces fewer assumptions than the continuous approaches, since the continuous cases model CFs that were not part of the reported values. However, all the CFs used have an uncertainty that is rarely reported (UNEP-SETAC 2019); therefore, in

our opinion, it is not unreasonable to model CFs as a continuous probability distribution. Moreover, the three approaches are in relatively close agreement, and the continuous cases can be used as an approximation. On the one hand, the continuous cases have the advantage of being easier to handle, as they do not require the additional package *presamples* and the time to learn how to use it. On the other hand, the use of *presamples* allows to model scenarios with a better characterisation of uncertainty. If CFs are uncertain, this uncertainty can be easily compounded with archetype uncertainty. If there is information of the likelihood of one archetypical scenario being more probable than other (e.g. emissions from industries being outside urban areas) these likelihoods can be incorporated as weighting factors when the samples are calculated.

On the inventory side, modellers and database developers should specify the archetype scenario when known and leave the “unspecified” option as a default in the absence of better information. A more thorough specification of archetypes would avoid the introduction of systematic errors and unnecessary uncertainty. In many cases the specification of archetypes is straight-forward. For instance, most of the field emissions of agriculture are likely to occur outside urban environments. Therefore, archetypes scenarios can be better specified without much effort. If the CFs to underspecified archetypes is calculated in a reasoned way (e.g. as a weighted average of the most likely scenario) it would make more sense to capture this information using the triangular distribution instead of the uniform.

With regards to method implementation, the assumptions around the characterisation of underspecified archetypes need to be better documented and reasoned. This study shows how these assumptions can have a large impact on LCA results, and that current implementations seem to be inconsistent and seldom documented. A commonly agreed inventory and LCIA implementation that all LCA software solutions can use is long overdue, and would help to have a more consistent set of assumptions.

In the three modelled approaches, measures of central tendency can be very different for the case where uncertainty is not considered (i.e. the mean or median score of the Monte Carlo simulation being very different to the deterministic score). This is not only true for archetype uncertainty, but for uncertainty estimates in general. Since the measures of central tendency incorporate more information than the deterministic case, it would make sense to consider these as a more accurate estimation of the true value of LCA scores, instead of the score that ignores all the uncertainty information.

Further work could systematically compare the CF uncertainty with the uncertainty coming from the inventory. This work only does it for a few cases. Fast methods to identify the main sources of uncertainty in an LCA calculation would also be helpful. As current techniques, are very computationally demanding (Patouillard et al. 2019).

ACKNOWLEDGEMENTS

The author thanks Pascal Lesage for his help with the *presamples* package.

FUNDING INFORMATION

This work was done without external funding.

REFERENCES

- Acero, A.P., C. Rodríguez, and A. Ciroth. 2017. LCIA methods Impact assessment methods in Life Cycle Assessment and their impact categories. *OpenLCA*(February): 1–23.
- Bourgault, G. 2019. *Implementation of impact assessment methods in the ecoinvent database version 3.6*. Vol. 6.
- Bulle, C., M. Margni, L. Patouillard, A.-M. Boulay, G. Bourgault, V. De Bruille, V. Cao, et al. 2019. IMPACT World+: Globally Regionalized Life Cycle Impact Assessment. *The International Journal of Life Cycle Assessment*. <https://link.springer.com/article/10.1007/s11367-019-01583-0>.
- Edelen, A., T. Hottle, S. Cashman, and W. Ingwersen. 2019. *The federal LCA commons elementary flow list: background, approach, description and recommendations for use*.
- Fazio, S., V. Castellani, S. Sala, E.M. Schau, M. Secchi, L. Zampori, and E. Diaconu. 2018. *Supporting information to the characterisation factors of recommended EF Life Cycle Impact Assessment method. Supporting Information to the Characterisation Factors of Recommended EF Life Cycle Impact Assessment Methods*,. <https://ec.europa.eu/jrc>.
- Frischknecht, R. and S. Büsler Knöpfel. 2014. Ecological scarcity 2013 - new features and its application in industry and administration. *The International Journal of Life Cycle Assessment* 19: 1361–1366. <https://link.springer.com/article/10.1007/s11367-014-0744-z>.
- Goedkoop, M., R. Heijungs, M. Huijbregts, A. De Schryver, J. Struijs, and R. van Zelm. 2013. ReCiPe 2008. https://www.leidenuniv.nl/cml/ssp/publications/recipe_characterisation.pdf. Accessed December 22, 2017.
- Heijungs, R. and S. Suh. 2002. *The Computational Structure of Life Cycle Assessment*. Dordrecht: Springer Netherlands. <https://www.springer.com/gp/book/9781402006722>.
- Henriksson, P.J.G., A. Rico, W. Zhang, S. Ahmad-Al-Nahid, R. Newton, L.T. Phan, Z. Zhang, et al. 2015. Comparison of Asian Aquaculture Products by Use of Statistically Supported Life Cycle Assessment. *Environmental Science and Technology* 49(24): 14176–14183.
- Lesage, P., C. Mutel, U. Schenker, and M. Margni. 2018. Uncertainty analysis in LCA using precalculated aggregated datasets. *The International Journal of Life Cycle Assessment*: 1–18. <http://link.springer.com/10.1007/s11367-018-1444-x>.
- Mutel, C. 2017. Brightway: An open source framework for Life Cycle Assessment. *The Journal of Open Source Software* 2(12): 11–12. <http://joss.theoj.org/papers/10.21105/joss.00236>.
- Mutel, C., X. Liao, L. Patouillard, J. Bare, P. Fantke, R. Frischknecht, M. Hauschild, et al. 2018. Overview and recommendations for regionalized life cycle impact assessment. *The International Journal of Life Cycle Assessment*.
- Patouillard, L., P. Collet, P. Lesage, P. Tirado Seco, C. Bulle, and M. Margni. 2019. Prioritizing regionalization efforts in life cycle assessment through global sensitivity analysis: a sector meta-analysis based on ecoinvent v3. *The International Journal of Life Cycle Assessment*. <http://link.springer.com/10.1007/s11367-019-01635-5>.

Pré. 2019. *Simapro Database Manual, methods library*. <http://www.pre-sustainability.com/download/DatabaseManualMethods.pdf>.

UNEP-SETAC. 2019. *Global Guidance on Environmental Life Cycle Impact Assessment Indicators Volume 2*.

Wernet, G., C. Bauer, B. Steubing, J. Reinhard, E. Moreno Ruiz, and B. Weidema. 2015. The ecoinvent database version 3 (part I): overview and methodology. *The International Journal of Life Cycle Assessment* 3(9): 1280–1230. <http://dx.doi.org/10.1007/s11367-016-1087-8>.

SUPPORTING INFORMATION

The functions to calculate archetype uncertainty are available with an open license at <https://pypi.org/project/ulcarchetype/0.1/>. The calculations done in this article are available at xxx. To fully reproduce the results, access to the LCA database ecoinvent is needed.

Figure Legends

Fig. 1 Normalized CF of elementary flows to unspecified archetypes, for impact categories with archetype uncertainty. Each point in the figure represents a CF. The maximum observed range is added in parenthesis.

Fig. 2: Systematic errors due to archetype uncertainty: mean absolute percentage deviation for all the activities in Ecoinvent 3.6 consequential. MAPDs are ordered by the “uniform” case. Each element in the x axis represents an activity in the database.

Fig.3 Uncertainty introduced by unspecified archetypes on LCA scores measured as relative standard deviation for Ecoinvent 3.6 consequential. The scores are ordered by the uniform case.