LCA is only relative – Experiences from the quantification of overall dispersions around aquaculture LCI results

Patrik Henriksson*, Jeroen Guinée

CML, Institute of Environmental Sciences, Leiden University
* Corresponding author. E-mail: *henriksson@cml.leidenuniv.nl*

ABSTRACT

As part of an evaluation of European seafood imports from Asia, overall dispersions were determined for a wide range of processes. Once propagated into LCI results using Monte Carlo (MC) simulations, large absolute uncertainties around results made conclusions difficult to reach. In response, the results were reevaluated, only considering relative uncertainties related to different hypotheses. This approach allowed for dispersions related to more than one of the products under study to be assumed the same, producing correlated samples. These results could, in turn, be tested using paired T-tests which allow for greater statistical power. Greater statistical power means that the resolution of comparisons is improved and the risk of reaching flawed conclusions reduced. Among other bias related to absolute results, the present research presents many contending arguments for only comparing LCI and LCA results in relative terms.

Keywords: LCI, uncertainty, dispersion, Monte Carlo, significance, statistical test

1. Introduction

Quantifying uncertainties related to life cycle inventory (LCI) results has long been desired in the field of life cycle assessment (LCA). To date, however, most estimates of uncertainties have been derived using a pedigree approach (Weidema and Wesnaes 1996), based upon expert judgment (Röös et al. 2010), or focused on a few key parameters (Middelaar et al. 2013), resulting in manageable ranges. In the few studies where parameters have been more extensively defined (e.g. Gregory et al. 2013), comparisons of products or services are often inconclusive. This is due to the large dispersions related to LCI results. Meanwhile, results are commonly communicated as point values, with little to no indication of the confidence behind any conclusive outcomes. Supporting LCI outcomes with information on its uncertainty and statistical testing is therefore needed.

The first step towards implementing statistical tests to LCI results is to identify all relevant sources of dispersion. In the present research, the *overall dispersions* defined in Henriksson et al. (2013) were adopted. Overall dispersions are the sum of inherent uncertainties (inaccuracies in measurements which may be reduced by additional research), spread (variability resulting from averaging) and unrepresentativeness (mismatch between the representativeness and application of data) (Henriksson et al. 2013). Secondly, data need to be collected and assembled following a representative sampling technique. However, as data collection commonly is the most resource demanding stage of any LCA, and LCA is data intensive, most studies source their data opportunistically. This results in data often representing very small sample sizes (*n*), surveys carried out using snowball sampling, geographically limited areas, temporally limited timespans or otherwise biased samples. Thirdly, a propagation method is needed to aggregate the unit process dataset into LCI results. In the field of LCA, Monte Carlo (MC) simulations with random virtual sampling is the most common propagation method (Lloyd and Ries 2007). While MC simulations of large datasets require relatively extensive computing, advances in hardware and software today allow for a great number of iterations to be produced from extensive LCI databases within a reasonable time using a standard personal computer.

When all sources of dispersion are taken into account, the resulting *absolute uncertainties* often become very large, making it difficult to draw conclusions in comparative studies (Henriksson et al. 2014). Common when comparing LCI results, however, is that many of the sources of dispersion are positively correlated with each other (Hong et al. 2010). In LCI studies this may refer to dispersions related to emission models (e.g. CH_4 estimates from diesel combustion), emissions from land-use change, or other processes which are shared amongst the production systems compared. Therefore, in comparative studies, only *relative uncertainties* are of importance and all shared dispersions can be assigned the same random samples. E.g. if an LCA evaluates two toasters, the uncertainties related to electricity generation can be assumed identical and only the uncertainty related to the amount of electricity used and the production of the toasters should be considered. Thus, only dispersions unique to either of the products under study, the relative uncertainties, are considered.

The types of statistical tests that can be implemented to test LCI results depend upon the characteristics of the data. Parametric statistics, for example, allow for more statistical power but assume equal variances and a type of probability distribution (Table 1). The probability distribution assumed by the most commonly used statistical tests (e.g. the T-test) is the normal distribution, but other test exist and distributions can also be transformed (e.g. from a lognormal to a normal distribution). In the meantime, non-parametric tests which do not require data to conform to a specific probability distribution offer more robustness, but an increased risk of committing a type II error (see Table 1). Implementing the wrong type of test can consequently result in flawed conclusions. For example, a novel biofuel might be deemed as equally polluting as diesel, while it in fact has significantly lower emissions. If only relative uncertainties are considered, the outcomes of each individual MC run for the product systems compared are correlated with each other (Wilcoxon 1945; Zimmerman and Zumbo 1993). This is referred to as correlated samples and implies that a paired statistical test should be implemented, where only the difference between product A and product B (A-B) is considered for each MC run. Paired statistical tests offers increased statistical power and are often used for e.g. testing groups of patients before and after a drug is administered (Zimmerman and Zumbo 1993).

Table 1: Common statistical terms used throughout this manuscript

Parametric statistics	Statistics of data of an assumed type of probability distribution; offers more statistical power and less chance of type II errors
Type II error	Failure to reject a false null-hypothesis
Statistical power	The probability of rejecting the null-hypothesis when the alternative hypothesis is true
Non-parametric statistics	Statistics of data with no characteristic structure; offers more robustness and less chance of type I errors
Type I errors	An incorrect rejection of the null-hypothesis
Statistical robustness	The ability to evaluate data of no characteristic structure, data affected by outliers, or data where variances are not equal

In the present research we evaluated the advantages of only considering relative uncertainties when comparing LCIs and challenged the use of absolute values. In order to do so, we worked on an inventory dataset describing Asian aquaculture. The data were collected as part of the SEAT project (www.seatglobal.eu) for several important aquaculture systems in Asia. In the present study, however, we only used the example of carbon dioxide emissions from the production of one tonne of tilapia fillets from two different Chinese farming systems, conventional and integrated with pigs, as an example. Our hypothesis was that "*fish integrated with pigs would have lower emissions than conventional systems as part of their feed is provided by the pig manure*".

2. Methods

Primary data collection started in 2010-2011, building upon a random sample of 200 fish farms (Murray et al. 2013). Based upon the outcomes of this sample, additional primary data were collected in 2012-2013 on supporting processes, including feed mills, hatcheries, processing plants, etc. The primary dataset was also supported by an extensive review of secondary data, detailing many of the supporting services (e.g. electricity production, transportation, etc.). For these data, aggregation of dispersion parameters was conducted using the spreadsheet supplied as Online resource to Henriksson et al. (2013) (also available at cml.leiden.edu/software/software-quanlci.html). For all other supporting processes the ecoinvent v2.2 database was consulted, relying upon default inherent uncertainties and the pedigree approach presented by Frischknecht et al. (2007). A full review of the data used and modelling choices adopted is available in Henriksson et al. (2014). Modeling and propagation was conducted in the CMLCA software v5.2 (available at cmlca.eu). For each farming system alternative 1 000 iterations were conducted. Probability density functions were evaluated in EasyFit v5.5 (mathwave.com), using the Anderson-Darling test to determine the distribution of data. Statistical tests were conducted using Wilcoxon matched-pair signed-rank test in SPSS v.21.

3. Results

The outcomes from the correlated and the non-correlated sampling were largely identical, apart from some random sampling differences. Between the two, integrated farms performed slightly better than the conventional farms (Table 1). The relative differences between the two data ranges (conventional-integrated) were not normally distributed.

Table 2: LCI results for carbon dioxide emissions from the production of one tonne of tilapia in either of two different Chinese tilapia farming systems.

	Conventional	Integrated
Average	3027	2684
Median	2805	2530
Standard deviation	1226	953

When compared on a MC run basis, two products with equal emissions would each be expected to yield the higher emissions 50% of the time. Meanwhile, the integrated farms came out as having higher emissions than conventional farms only 43% of the runs when absolute results were regarded, and 39% when relative results were regarded. The difference in the dependently sampled results could also be determined as highly significant ($p<0.001$) using the non-parametric Wilcoxon matched-pair signed-rank test. When visualized for a small selection of MC runs (Figure 1), the correlation between the samples in the dependent sampling becomes evident. Where only relative uncertainties are considered (dependent sampling) the samples generally come out closer to each other than when absolute uncertainties are considered (independent sampling). This as the results rely upon the same LCI matrix, where e.g. the assumed emissions from electricity generation and transportation in China remain the same for both production chains in each MC run.

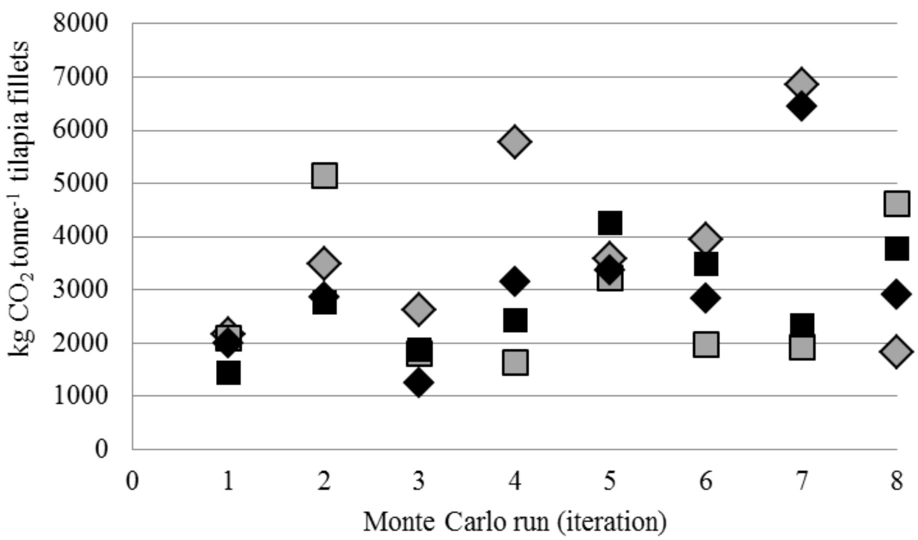

Figure 1: Illustrative example of a selection of outcomes using correlated (black) and non-correlated (gray) sampling for carbon dioxide emissions from conventional (diamonds) and integrated (squares) tilapia farming in China.

4. Discussion

Given the definition of the overall dispersions evaluated here, the ranges of results give an indication of the *absolute uncertainties* underlying results. Meanwhile, if two production chains which experience many shared sources of dispersions are to be compared, only relative uncertainties need to be considered. Only considering relative uncertainties improve the resolution of comparisons and also allow for the implementation of more powerful paired statistical tests. The field of LCA should therefore amend the practices of most other sciences, where results are only compared to an alternative or control (benchmark), working towards a predefined hypothesis.

Although often used in LCA studies instead of great or large, the word significant in the scientific realm implies that a statistical test has been performed. Meanwhile, only a handful of LCA studies have so far implemented statistical testing to support their conclusions (Canter et al. 2002) and the role of statistical tests in the field of LCA is still vaguely defined. Embracing statistical concepts would, however, contribute towards more solid LCA practices. Essential for this will be the understanding that statistical tests are only ever as strong as the sampling framework they build upon. Sampling frameworks therefore need to promote random unbiased samples, where any remaining bias is communicated so that it can be corrected for. Moreover, the probability density function of LCI outcomes should be characterized and transformed when needed, in order to allow for parametric tests. Otherwise there is a risk that type II errors are committed.

If the approach proposed here is adopted, modeling options may also need to be considered. For example, if a generic process for “Combustion of diesel in generator” is used, rather than defining its emissions as part of each process using a generator, dependent sampling becomes possible. Similarly for waste flows, where e.g. a waste processes may be created for each kg of nitrogen applied to agricultural fields. While the uncertainties for these more generic processes might be larger, they are less relevant from a relative point-of-view. Dependent sampling is also beneficial when uncertainties around characterization factors are considered. In some cases, such as freshwater ecotoxicity, uncertainties may range an order of magnitude (Henderson et al. 2011). In such cases, adopting dependent sampling could greatly improve the resolution of comparative studies.

Further efforts are needed to develop the ideas presented here. For example, the strong influence of sample size in most statistical tests. When results are resampled, as in MC simulations, the sample size can be changed at the click of a button. This means that a null-hypothesis might not be rejected (no significant difference) at 100 iterations, but at 1 000 iterations. This rather arbitrary feature has even been addressed by implementing sequential stopping boundaries, where MC tests will stop once a sufficient number of runs have been produced to achieve a certain significance level (Fay et al. 2007). Future research should therefore explore how to deal with such resampling risks in the field of LCA (Gandy 2009) and evaluate statistical tests building upon the cumulative distribution function, such as the Kolmogorov–Smirnov test.

5. Conclusion

LCAs presenting results as point values may be flawed when concluding that one product is better than another while no significant difference exists (type I error), or to conclude that two results are indifferent from each other while they actually significantly differ (type II error) when dispersions are quantified as yet. Either of these false outcomes may result in flawed decisions, and only by quantifying the dispersions around LCI results can statistical tests be implemented and the confidence behind conclusions communicated. In doing so, dependent sampling only regarding relative uncertainties offers many advantages. This, however, requires studies to work towards a hypothesis, in order to avoid multiple comparison problems. Multiple comparison problems refer to the elevated chance of finding significant trends by chance if a large number of comparisons are made. Amending these recommendations will allow LCA to become a much more scientifically robust tool, rather than an exploratory framework. Conclusively, we argue that LCI and LCA results should only ever be seen as relative, and discourage comparisons of absolute results across studies.

6. References

Canter KG, Kennedy DJ, Montgomery DC, et al. (2002) Screening stochastic Life Cycle assessment inventory models. Int J Life Cycle Assess 7:18–26. doi: 10.1007/BF02978906

Fay M, Kim H, Hachey M (2007) On using truncated sequential probability ratio test boundaries for Monte Carlo implementation of hypothesis tests. J Comput Graph Stat 16:946–967.

Frischknecht R, Jungbluth N, Althaus H, et al. (2007) Overview and Methodology. ecoinvent report No. 1. 77.

Gandy A (2009) Sequential Implementation of Monte Carlo Tests With Uniformly Bounded Resampling Risk. J Am Stat Assoc 104:1504–1511. doi: 10.1198/jasa.2009.tm08368

Gregory JR, Montalbo TM, Kirchain RE (2013) Analyzing uncertainty in a comparative life cycle assessment of hand drying systems. Int J Life Cycle Assess 18:1605–1617. doi: 10.1007/s11367-013-0606-0

Henderson AD, Hauschild MZ, Meent D, et al. (2011) USEtox fate and ecotoxicity factors for comparative assessment of toxic emissions in life cycle analysis: sensitivity to key chemical properties. Int J Life Cycle Assess 16:701–709. doi: 10.1007/s11367-011-0294-6

Henriksson PJG, Guinée JB, Heijungs R, et al. (2013) A protocol for horizontal averaging of unit process data - including estimates for uncertainty. Int J Life Cycle Assess 19:429–436. doi: 10.1007/s11367-013-0647-4

Henriksson PJG, Zhang W, Nahid SAA, et al. (2014) Final LCA case study report - Primary data and literature sources adopted in the SEAT LCA studies. SEAT Deliverable D3.5 – Annex report. 132.

Hong J, Shaked S, Rosenbaum RK, Jolliet O (2010) Analytical uncertainty propagation in life cycle inventory and impact assessment: application to an automobile front panel. Int J Life Cycle Assess 15:499–510. doi: 10.1007/s11367-010-0175-4

Lloyd SM, Ries R (2007) Characterizing, Propagating, and Analyzing Uncertainty in Life-Cycle Assessment: A Survey of Quantitative Approaches. J Ind Ecol 11:161–179. doi: 10.1162/jiec.2007.1136

Middelaar C van, Cederberg C, Vellinga T V, et al. (2013) Exploring variability in methods and data sensitivity in carbon footprints of feed ingredients. Int J Life Cycle Assess 18:768–782. doi: 10.1007/s11367-012-0521-9

Murray FJ, Haque MM, Zhang W, et al. (2013) Defining boundaries towards understanding sustainable ethical aquaculture trade between Asia and Europe. SEAT Project Report 2.8. Stirling, UK

Röös E, Sundberg C, Hansson P-A (2010) Uncertainties in the carbon footprint of food products: a case study on table potatoes. Int J Life Cycle Assess 15:478–488. doi: 10.1007/s11367-010-0171-8

Weidema B, Wesnaes MS (1996) Data quality management for life cycle inventories—an example of using data quality indicators. J Clean Prod 4:167–174. doi: 10.1016/S0959-6526(96)00043-1

Wilcoxon F (1945) Individual Comparisons by Ranking Methods. Biometric Bull 1:80–83.

Zimmerman D, Zumbo B (1993) Relative power of the Wilcoxon test, the Friedman test, and repeated-measures ANOVA on ranks. J Exp Educ 62:75–86. doi: 10.1080/00220973.1993.9943832

This paper is from:

Proceedings of the 9th International Conference on Life Cycle Assessment in the Agri-Food Sector

LCA Food 2014

8-10 October 2014 - San Francisco

Rita Schenck and Douglas Huizenga, Editors
American Center for Life Cycle Assessment

The full proceedings document can be found here:
http://lcacenter.org/lcafood2014/proceedings/LCA_Food_2014_Proceedings.pdf

It should be cited as:

Schenck, R., Huizenga, D. (Eds.), 2014. Proceedings of the 9th International Conference on Life Cycle Assessment in the Agri-Food Sector (LCA Food 2014), 8-10 October 2014, San Francisco, USA. ACLCA, Vashon, WA, USA.

Questions and comments can be addressed to: staff@lcacenter.org

ISBN: 978-0-9882145-7-6