April 20th, Module 2 session 2.

- Q about the mass balances in MC simulation – is it preserved?

Massimo: No it is not. In principle, and from a theoretical perspective this is a problem. Because we want a mass balance model to keep it realistic. However, in practice, its unclear how much this is pragmatically a problem, for various reasons.
1) for the central limit theorem values tend to stay within the center of distribution. Mass differences in input output should tend to be small, the question is, will they be in the same size as the general "noise" in the data? No clear empirical evidence on this issue yet.
2) the impression is that many datasets are not mass balances anyway or at least not at single substance level.
3) what is the alternative? Parametrize everything (huge model hardly manageable) or repeated sampling in all flow of all activities?
4) The stochastic simulation is an artificial procedure so must always be taken with a substantial grain of salt.

- Q about how to separate the uncertainty in the "model" (intended as foreground system in this case) and that in the "database". How to separate?

You can run the MC simulation only on the foreground. Need to specify uncertainty though, using pedigree matrix or other method, either experimental (repeated samplings if available) or expert-based (assumptions about range or estimated uncertainty). See example here: https://github.com/massimopizzol/seaweed-bioplastic-lca for this paper: https://doi.org/10.1016/j.algal.2023.103036

- Q about significance in comparative context – how to say when two alternatives are different? The simulation creates always very overlapping distributions.

Massimo: the problem is that the simulation is often not done using paired samples. One should first sample the tech matrix, then compare alternatives, then sample again. Then look at how many times one alterative is better than the other across simulations. If one samples first n times one alternative and then n times another alternative the comparison of the two distributions is largely overestimating the uncertainty. This is clearly explained in the notebook *6-Comparative-Monte-Carlo.ipynb*

- Q about how to proceed to do uncertainty analysis?

Massimo: there are two ways. Normal way, do the LCA then in the end remember that "oh there is also uncertainty" and then do some uncertainty analysis to nuance interpretation. Works in many cases and fine. Second way is useful for situations where uncertainty is expected to be large. E.g. studying new technologies or products at early stage or processes where data gaps are known to be substantial (e.g. chemicals) or data quality low. This is what I call a "uncertainty first" paradigm. Start having the uncertainty in mind since the beginning and design the LCA accordingly. Use ranges for parameters, use scenarios to check the assumptions, apply simulation. Do not focus on the accuracy of the final result (impact indicator, e.g. carbon footprint) but focus on the range of results obtainable and the reasons.

This for a nuanced decision support in situations of large expected uncertainty. Avoid providing a false sense of accuracy (tendency of many LCAs unfortunately, especially in product labelling, number with two significant digits…) where accuracy can not reasonably be achieved.

- Q about uncertainty in consequential models, which assumptions should be discussed?

Massimo: the soundness of assumptions should be discusses in all LCA models not only consequential ones. For the consequential ones usually a very sensitive assumption is about the activities to be substituted in cases of joint co-production (using substitution method). Sometimes also the marginal mix assumptions deserve a sensitivity analysis.

- Q about difference contribution analysis and GSA?

Massimo: The latter is a contribution analysis on steroids in the sense that it's like a contribution analysis but performed under several different versions of the product systems. They can be interpreted in similar way and sometimes results are logically similar (results are very sensitive to e.g. the input of an activity which is a large contributor to the total impact).

- Q about difference uncertainty analysis and sensitivity analysis, and between OAT and GSA.

Massimo: Uncertainty analysis is about quantifying the range of the model output based on the range of model inputs. It's a quantification (how large or narrow is this range?). Sensitivity instead is about tracing back a change in output to a change in input. In particular then the GSA is about explaining hoe much of the variance (range) in output is explained by/due to each input parameter.

- Q about the LCA. Method in bw and difference with LCA.lci()

Massimo: unclear, need to read the source code. If you can run the LCA.activity_dict() function already before running LCA.lci() then my guess is that the LCA. Method is the one creating the technology matrix in bw, while LCA.lci() is about solving the inventory.

- Q about the use of pedigree matrix, opinions?

Massimo: semi-quantitative method, which I have nothing against. Qualitative considerations transformed into quantitative estimates. Could be defined as an "expert-based assessment". Some argue can not be sued for simulation, based on the rationale that it was not designed for that, but I still have to read a strong argument for not using it. Also, very pragmatic method to assign uncertainties at database scale (easily applicable to thousands of parameters). However, my impression is that it tends to overestimate uncertainties and return too wide distributions. And again, what alternative do we have at database scale? More empirical research needed on this.

- Q about UUID for datasets, where does the code originates from?

Massimo: brighway codes are generated automatically starting from the metadata of each ecoinvent dataset. They work across installations and this is what allows me to make a code that you can run in your computer and reproduce the same result, i.e. it is essential for reproducibility. But the bw codes are a product of bw and exist only in bw. Ecoinvent has other UUIDs to identify their datasets (two of them I think). When shifting a project to a new database version one has to change all codes manually usually. It's a pain of course but right now inevitable.

- Q about UUID for databases, isn't there an unique identifier for ecoinvent database versions? Why do we have to call it "ecoinvent 3.9 conseq".

No there isn't. I have actually asked them. You can call the database with any name you want in bw but use informative ones (e.g. specifying the db version and system model).

- Q about using the notebooks, can we do it?

Massimo: you can use the notebooks with attribution as they are in CC-BY license. You can copy paste parts and use them in your research, feel free to do that, but if you do please add a line that this part is taken from the notebook so you give credit to the original developers.

- Discussion of notebook 8.1.

Massimo: the whole point is to modify the background database but without touching the files and writing on It, so that we maintain speed and do not mess up with the original database files, and also it goes faster. In the notebook you can access the technology matrix using the .activity_dict method and basically operate on the huge technology matrix (foreground + background). This is very useful to perform simulations and other types of modifications of the database that are reflected on all calculations within a project but do not affect other projects.