

Disclosure of Product System Models in Life Cycle Assessment: Achieving Transparency and Privacy

Brandon Kuczenski
Institute for Social, Behavioral, and Economic Research
University of California, Santa Barbara

2018-02-12

Response to Reviewers

Thank you again to the reviewers for their robust reading of my revised manuscript. I appreciate that all reviewers found the article to be engaging and important. Their comments helped me to clarify the presentation of the methodology and to improve the connection to current scholarship, notably Majeau-Bettez et al (2014).

I have taken all of the reviewers' recommendations to heart. The major changes to the manuscript were in the mathematical introduction sections– I have rewritten “Mathematical Formulation of LCA” to centralize the supply-and-use framework, and to emphasize that the matrices developed in the text are coefficient matrices derived from supply and use tables. I have also rewritten “Foreground Study” to better illustrate how a foreground study is constructed with reference to a background database, and to better define the canonical functional unit. I added a new figure at the request of Reviewer #3, and a new discussion subsection dealing with the role of the community. I also made an assortment of clarifying changes throughout the document.

Finally, I added a section entitled “Recommendations to Advance LCA Practice” at the end of the discussion, to respond to reviewers #2 and #3's suggestions about the role of the community and next steps. This section has added about 400 words to the paper.

Below I respond to each comment individually.

Reviewer #1

In the response to the reviewers I read: "I have adopted the input-output notation exclusively in the revision". That is true. There are no longer formulas with A^{-1} , only with the IO-style $(I-A)^{-1}$. Unfortunately this has weakened the relevance of the paper quite a bit. Let me explain. There are two ways of doing LCA: process-based and IO-based. A major example of a process-based background database is Ecoinvent, a major example of an IO-based background database is Exiobase. The issue which this paper addresses (proprietary licensed data in the background) is the case for Ecoinvent, but not for Exiobase. So the paper makes a wrong choice in explaining how to protect non-proprietary data from IO but not explaining how to protect proprietary data from process-based DBs. I have no clue why you are making this choice, except one debatable statement that the IO-approach "is computationally more efficient" (which I personally doubt, but moreover is irrelevant for the present paper). As a matter of fact, most software for LCA (so page 7 line 49 is

definitely wrong) is process-based, and so are the standards from ISO (in which the "unit process" assumes a core role). I strongly urge to phrase the paper entirely in process-based style, using A^{-1} instead of $(I-A)^{-1}$. Editorially speaking it is a small change, but it increases the relevance of the paper tremendously.

I thank the reviewer for his/her deep insights into the paper, which have led to a complete rewrite of the section of the manuscript dealing with the core LCA computation. I now tie this strongly to the supply-and-use framework. (p.5-7 in the revision).

I note that the reviewer inaccurately conflates the *contents* of an inventory database with their mathematical *representation*. In fact, both Ecoinvent and Exiobase are best represented using the supply-and-use framework. It is only when preparing the LCA computation for matrix inversion that the supply and use values are placed into coefficient form. In Ecoinvent, each unit process is described in terms of distinct products (outputGroup) and consumption (inputGroup) separately. The un-allocated datasets are not suitable for computation, and the linked ecoinvent system models are stored as collections of normalized, single-output EcoSpold v2 files, which are trivial to compile in either technology-matrix or input-output form. I have emphasized this in the revision. The point of the article was to demonstrate how foreground models may be built on top of, and subsequently separated from, a background database, and the development applies equally well to a foreground model based on either Ecoinvent or Exiobase.

Regarding the computational efficiency of iteration vs inversion, I agree that the performance aspects are out of scope for the current paper. However, the reviewer should note that in fact the Leontief inverse is the standard. SimaPro, the leading worldwide software system for scientific LCA, converted to using an iterative Leontief inverse for Version 8, as disclosed in a whitepaper published in 2014 (<https://www.pre-sustainability.com/news/new-calculation-engine-simapro-8>), realizing significant speed and memory usage improvements. OpenLCA, the open-source alternative, also reported a dramatic reduction in memory usage and computation time around the release of version 1.4 in 2014, coincident with a revised algorithm "optimized for huge databases" (<http://www.openlca.org/openlca-beta-1-4-released-optimised-for-huge-databases/>). This is highly suggestive of a switch to an iterative approach. GaBi, the more industrially oriented software, does not use a matrix formulation at all, but has always computed activity levels using iteration at the plan level.

** page 8 line 32 "The simplest conceivable foreground is simply the final demand y , in other words the list of background processes invoked by the study and their activity levels." I have no idea what this sentence means. The foreground is a collection of processes, but the final demand is not a process, so how can the final demand be a foreground? And as to the second part: what is the relation between the foreground and the list of background processes?*

** On page 8 line 48 I see a matrix A with a 0 in the top-left corner. Page 9 generalizes this, but suddenly has a matrix (A_f) at that place. This shift from page 8 to page 9 should be justified, because this is a major change.*

The passage introducing the foreground study has been revised extensively, and the shift from a final demand vector to a model for a process, and from a collection of processes to a complete foreground, has been rewritten in a narrative format to illustrate the workflow (p.9). Hopefully the revision is clear.

** The same matrix A on page 8 line 48 contains a bold 0, which suggests a column of zeros, but actually is a row of zeros. Adding the subscript T would help here.*

Thank you for the observation. This has been fixed.

** Page 10 line 12 says "The augmented A and B , makes up an LCA foreground study". I am wondering if something went wrong. Two grammatical errors (the spurious comma and the singular verb) suggest an*

editorial mistake. But the meaning is also unclear: A and B contain foreground and background, so why is this a "foreground study"? What is a foreground study anyhow?

The grammatical error has been corrected. A foreground study is defined on p.9 as the two augmented \tilde{A} and \tilde{B} matrices, containing foreground information described with reference to a background database.

** Page 10 line 12 also speaks of "the canonical functional unit". Can you define the term canonical?*

Some puzzles are best left as an exercise for the reader.

** Page 10 line 37 "the final demand vector has been replaced with a canonical functional unit and no longer needs to be explicitly reported". Of course my critique depends on the missing definition of "canonical", but even then, do I understand that you write that case studies do no longer need to report a functional unit? That would be a revolutionary statement, so please expand or revise.*

The *canonical functional unit* is now defined at the top of p.10 as a unit of the reference flow of the process described by the *first* column of the \tilde{A} matrix. It is introduced for the purposes of simplifying notation later on. While the use of a canonical functional unit means that the functional unit of a study no longer needs to be explicitly reported, it does still need to be defined as part of the disclosure. This is mentioned on p.12 in disclosure item *d-i*.

** Page 10 line 27 rewrites A as A_{flat}. What is the relation between these two symbols? Of course you may define a new symbol A_{flat}, but I even don't know what you intend do with A_{flat}: I see no equation in which it is used (like for instance A is used in eq (2)).*

The equation formerly known as Eq. 4 is now removed from the manuscript. A_{flat} still makes an appearance in the supporting information, but it is only an intermediate step toward deriving the main foreground study equation (old Eq. 5; new Eq. 4).

Reviewer #2

The resubmitted version prepared by the author represents a major improvement compared to the previous version in terms of clarity and conveyance of the message. I had a good time while reading the manuscript. The derivation of the canonical LCA computation equation, the simplifications and data arrangements, the listing of the six data components d-i to d-vi, and the data privacy arrangement are convincing and clear.

Thank you for the positive comments.

+ I have only one major issue with the framework presented: It starts with the A-matrix, which already includes a major modelling step: the conversion of raw unallocated processes data, including production volumes, etc. to 1:1 product-to-industry coefficients. What about the documentation of the raw, unallocated process data, which are the most valuable? Including this part of the data work into the disclosure framework is crucial to its eventual success in my (and many others') view. It may be beyond the scope of the current paper, but I still think the paper here must position itself. I therefore

+ request to comment on the issue in the manuscript, e.g., in the discussion, and

+ suggest to spend a couple of pages in the SI describing the issue.

I agree with the reviewer that omitting the conversion of unallocated processes to single-output processes leaves out an important modeling step. I first revised the introduction of Eq.1 by making strong ties to the supply-and-use framework of input-output analysis (beginning on p.5), as mentioned for the first reviewer. I then undertook a close review of Majeau-Bettez et al (2014) upon the reviewer's recommendation, and

found that the article's taxonomy of allocation strategies could be captured within the foreground disclosure framework.

In order to not clutter the text too much, I limit intensive discussion about this topic to the supporting information, where I demonstrate how an example multi-output process may be converted into single-output processes by means of the framework. I take care to note that each strategy preserves the raw, un-allocated process data in the disclosure. I draw attention to this fact a few times in the main text, first on p.14 bottom; again on p.17-18; and again in the conclusion.

Hopefully this work is convincing for the reviewer. I was quite happy to come to the determination that the multi-functionality problem fits well into the scope of the current manuscript, and I feel the potential impact of the paper has been greatly improved because of this revision.

Detailed comments

P4: "In the context of this paper, an "LCA Study" should be understood as ..." is not very clear, please rephrase.

The problematic sentence has been removed.

P5: "On the basis of transparency, the framework..." -> write 'product system model disclosure framework' or something else that is more explicit.

I accept the reviewer's suggestion.

P6: "The task of procedure can be described in two steps: making ..." Please link these steps to the ISO standard.

The sentence has been removed in the revision.

Equation 1: I recommend using c instead of e , since e is often used as summation vector, i.e., a vector that contains only 1 as elements.

This has been done throughout the text, figures, and supporting information.

Equation 1 and in general: Please be consistent throughout the entire manuscript which variables contain flows (at scale) and which contain coefficients. Any jargon in that regard must be avoided, e.g., referring to B as 'elementary flow matrix', where, in reality, it is a normalized (per unit) elementary flow matrix.

Assuming you mean "normalized (per unit) coefficient matrix," I have taken care to emphasize that all matrices in the disclosure framework are coefficient matrices. Notably p.6 middle; p.7; and p.11 top. No variables in the text refer to flows at scale.

Another jargon phrase to be clarified: "the foreground systems themselves cannot be computed without the background." (P8)

Revision: "However, the results of a study with foreground content cannot generally be computed without knowledge of the background."

P9: I think that at latest at this stage, many readers will wonder how the system you propose applies to cases where process descriptions in the background matrix are altered (also related to the comment of one of the reviewers regarding the zero-ness of the upper right corner of \tilde{A} . I think you should anticipate that question early on to keep your readers on board. In my opinion, ...) ...

This is discussed in Note #2, referenced on p.9 bottom.

Equation 4 is not a re-written version of equation 3! Please rephrase.

The equation formerly known as Eq. 4 has been removed from the manuscript.

P10: I have never heard of the rank of a vector. Do you simply mean its length?

Revision: “vector whose length is the same as the rank of A_f .”

P11: The term ‘aggregated foreground’ needs more introduction/explanation. At least the name, dimension, meaning, and unit (coefficient or flow??) of each of the newly introduced variables should be listed.

I have added text around Eq. 5 (formerly Eq. 6) to clarify these points.

It would be nice to see Equation 6 as part of a larger ‘master equation’, for example, adding a bit more detail at the bottom of Figure 2.

Eq. 4 (formerly Eq. 5) is the “master equation” for the foreground model. I have added it in a new row to the bottom of Fig.2. Hopefully it made the figure more helpful than confusing.

P10, “where $\sim y_f$ is a canonical functional unit vector . . .”, and P11: “The foreground matrix A_f , dependency matrix A_d , and foreground emission matrix B_f now contain all study-specific information.” That means that you are using the ‘canonical’ y vector all throughout your system, and that fact should be made more clear not that readers think that you forgot about y . Does ‘canonical’ here mean that y_f is always $[1,0,0,0, \dots, 0]$? If so, that should be made clear on P10.

I have improved the declaration of the canonical functional unit on p. 10.

P12: “The disclosure includes only information for which the study is the primary source.” To me that does not often seem to be the case, as also in the foreground, data from the literature and sometimes even recycled data fromecoinvent itself are used, e.g., when background inventories are copied into the foreground and modified there. I think that the statement on the foreground system description as ‘primary source’ should be toned down/modified.

Revision: “The disclosure includes all information specific to the study, or for which the study is the primary source, including re-implementations of already-published data.”

P13: “Objectives in critical review vary. . .” I recommend sharpening this paragraph. Reviewers should be held responsible for checking the model correctness and data/model/system structure as well, and a canonical PSM structure like the one proposed here would greatly facilitate the completion of this task. What is common now is clearly insufficient.

At your urging, I added the following to p.14: “Although the object of critical review is typically a written report, this level of review is not sufficient to ensure the correctness of quantitative results.” This is a statement of sentiment, but I do feel that it is accurate.

P13: “These nodes typically represent unit processes, but they can also indicate exchanges across an implicit system boundary that encloses the foreground.” I don’t understand the second part. What is an implicit system boundary here? How can A-matrix columns describe unit processes and exchanges at the same time? I thought the latter is coded in B_f ? Clarification needed.

I now explicitly introduce the notion of a “foreground system boundary” on p.15 and use it to define cut-off flows in a robust way.

P14: “. . . it must either be “terminated” in the dependency matrix. . .” What is the dependency matrix here? A_d ? Please specify.

I removed the term “terminated” since it is of my own coinage; and yes, the dependency matrix is defined as A_d in Eq.3 and so referenced on p.11 in “Contents of a disclosure.”

P15: “In a matrix representation, each unit process must have exactly one reference flow, which corresponds to its row and column in A_f .” That is a crucial LCA-matrix feature and should be described a bit more in detail: ... In short: Each A-matrix column refers to the making 1 reference product AND unit, and exactly this reference product and unit is then used by other processes.

These points are made with greater clarity in the supporting information.

In general, when writing about the classifications d-i, d-ii, and d-iii the importance of units should be stressed at some point in the manuscript.

This is now emphasized on p.13: “Each entity included in items ?? through ?? describes a flow of some kind, and its unit of measure must be reported in the disclosure.” It is also discussed in the supporting information.

Figure 3: A legend must be provided, explaining the meaning of the blue boxes, the dashed-line boxes, and the star. The variable \tilde{a}_f is new to me and I don’t understand its shape. Notation in Fig. 2, 3, and equation six must be checked for consistency. Else the figure is very informative and clear!

I added a legend and also a definition for the formerly un-defined \tilde{a}_f on p.18. I also rendered the vectors in bold face for consistency with the text.

P16: “A fragment of the PSM can also be published as a complete foreground.” I don’t understand, please explain.

The sentence has been removed, and the word “fragment” no longer appears in the manuscript. Along with “termination” of foreground flows, this is rather a concession for me!

P16: “Any co-production treatment, such as an allocation or system expansion, can only be fully expressed as a foreground fragment, because it intrinsically involves a collection of linked unit processes.” I don’t understand, please explain.

Revision: “Any modeling technique used to treat a multi-functional process, such as an allocation or system expansion, can only be fully expressed as a foreground with at least as many columns as outputs. For examples of multi-output processes represented as foreground models, please see the supporting information.”

P17: “A publication of the aggregated foreground is identical to the publication of a unit process, except that the result was derived through aggregation” I don’t understand, please explain.

Revision: “A disclosure of an aggregated foreground resembles the description of a unit process.”

Section 5.4. Consider citing DOI 10.1111/jiec.12245.

This has been added.

I suggest to include another brief section 5.5., containing a comment on the role of the ISIE (it’s their journal!) or the SETAC group in facilitating the establishment of science-wide classifications. What about the responsibility of the database and LCIA-method providers?

As noted in the introduction, I added a new section at the end of the Discussion entitled “Recommendations to advance LCA practice,” which addresses this.

Reviewer #3

First of all, the new title is much clearer and to the point.

The scope of the new version of the article is also clearer and more consistent with the objectives. I am satisfied with the responses to my comments on the earlier version as well as that of the other reviewers.

Thank you for your positive comments.

One additional comment in the form of a suggestion is the following. From my point of view, the core result in this article is presented in section 2.4 with the list of elements from the PSM which should be disclosed:

1. This list could be represented graphically, similar to a “disclosure card” such as a simplified versions of Figures S2 or S3.

New Figure 3 has been added to obtain this objective. I hope you like it!

2. A reference to the examples derived in the annexes would make the demonstration more concrete (there were several in the previous version of the article?).

I have added a mention of the examples to the section formerly identified as 2.4.

This would help in emphasizing the next steps, especially if the results draw on discussions within the SETAC working group, and possibly answer the question of whether this could become part of future LCA standards.

As noted in the introduction, I added a new section at the end of the Discussion entitled “Recommendations to advance LCA practice,” which addresses this.