Journal of
Industrial Ecology

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

Journal:	*Journal of Industrial Ecology*
Manuscript ID	Draft
Wiley - Manuscript type:	Research & Analysis
Date Submitted by the Author:	n/a
Complete List of Authors:	Kuczenski, Brandon; University of California, Santa Barbara, Institute for Social, Behavioral, and Economic Research
Keywords:	life cycle assessment (LCA), transparency
User-Supplied Keywords:	critical review, product system models, confidential information
Abstract:	Many of the challenges facing knowledge synthesis from life cycle assessment (LCA) studies stem from the inability of study authors and readers to formally agree on the structure and content of the product system models used to perform LCA computations. This article presents a framework for formally disclosing the foreground of an LCA study in a way that permits the computations to be inspected, verified, and reproduced by a reader, provided that the reader has access to the same life cycle inventory and impact characterization resources as the author. The framework can also be used to partition a study into public and private portions, allowing both portions to be critically reviewed but omitting the private information from the disclosure. A disclosure is made up of six components, including three lists of entities in the model and three sparse matrices describing their interconnections. The entity lists make reference to previously-published resources, including background inventory databases and characterized elementary flows, and the disclosure framework requires both author and reader to agree on the meaning of each of these references. The framework contributes to ongoing efforts within and beyond industrial ecology to improve the reproducibility and verifiability of scholarly works, and if implemented, plots a course toward distributed, platform-independent computation and validation of LCA results.

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

2017-11-16

Submitted after revision to the Journal of Industrial Ecology; November 2017

Abstract

Many of the challenges facing knowledge synthesis from life cycle assessment (LCA) studies stem from the inability of study authors and readers to formally agree on the structure and content of the product system models used to perform LCA computations. This article presents a framework for formally disclosing the foreground of an LCA study in a way that permits the computations to be inspected, verified, and reproduced by a reader, provided that the reader has access to the same life cycle inventory and impact characterization resources as the author. The framework can also be used to partition a study into public and private portions, allowing both portions to be critically reviewed but omitting the private information from the disclosure. A disclosure is made up of six components, including three lists of entities in the model and three sparse matrices describing their interconnections. The entity lists make reference to previously-published resources, including background inventory databases and characterized elementary flows, and the disclosure framework requires both author and reader to agree on the meaning of each of these references. The framework contributes to ongoing efforts within and beyond industrial ecology to improve the reproducibility and verifiability of scholarly works, and if implemented, plots a course toward distributed, platform-independent computation and validation of LCA results.

1 Introduction

Life cycle assessment (LCA) is a standard methodology to estimate the potential environmental impacts of products or services by modeling the network of industrial processes that must occur to deliver them to a consumer. The technique is well-established, widely practiced, and draws on

an extensive body of standardization and scholarship. Increasingly, the results of LCA studies are appearing in environmental declarations, corporate sustainability reports, and product marketing information. LCA is also gaining prominence as a tool for developing and evaluating environmental policy. At the core of an LCA study is a model of a "product system," which represents a collection of "processes … performing one or more defined functions", that describes the life cycle of a product (ISO 14044, 2006). The product system model, or PSM, encodes information about how a product is manufactured, distributed, used by the consumer, and what happens to it after it is disposed. A product system can be divided into a foreground, which denotes the portions of the life cycle whose operations are directly modeled by the study, and a background, which represents the global industrial system (Clift *et al.*, 1998). While the foreground is often modeled from direct observation or other primary data, background processes are typically drawn from a life cycle inventory (LCI) database prepared by a third party.

Although there has been considerable effort to normalize and harmonize LCI database design methodology and data interchange (UNEP/SETAC, 2011; Recchioni *et al.*, 2013; Mila i Canals *et al.*, 2015; Ingwersen *et al.*, 2015), less attention has been paid to harmonizing PSMs. These models are often highly complex and include countless modeling decisions, approximations, and assumptions made by a study author (Lloyd & Ries, 2007; Reap *et al.*, 2008). Factors such as co-production strategy (Finnveden, 1999; Pelletier *et al.*, 2014) can also render study findings unreliable or highly contingent. Consequently, while comparative evaluations can be made regarding alternative cases or scenarios within a single study, comparisons across studies are much more challenging (Heath & Mann, 2012; Henriksson *et al.*, 2014). Even when relatively simple and homogeneous systems are considered, wide variations can be found in the results of studies from different authors (van der Harst & Potting, 2013; Turconi *et al.*, 2013). A lack of transparency in reporting, particularly regarding definitions of study scope and system boundary, is a crucial challenge to the interpretation of results in comparative analysis (Cleary, 2009; Laurent *et al.*, 2014). A further challenge to transparency can be found in the use of input data that are confidential or proprietary (Kuczenski *et al.*, 2016). One consequence of all these challenges is the high cost

and uneven rigor of critical review, in which the complexities of LCA come face to face with the limitations of current practice (Curran & Young, 2014).

As LCA gains prominence, particularly in the policy realm, these problems become more acute. If LCA is to be used in policy, it will necessarily have the effect of recommending certain product systems, technologies, or approaches at the expense of others; yet in that event it is critically important that there be a consensus among stakeholders that the quantitative results are credible and well-supported (Rainville *et al.*, 2015; McManus *et al.*, 2015). As noted above, this level of consensus is hard to realize, in part because of the conflicting implications inherent in different analytic modes. Plevin et al. famously observed in the strongest terms that careless reporting of attributional LCA study results can distort the significance of findings in a policy context (Plevin *et al.*, 2013), a shortcoming that can be found in other modes as well (Brandão *et al.*, 2014). In the scope of Environmental Product Declarations (EPDs), studies within an industry are supposed to be rendered comparable by adhering to a common product category rule (PCR) (Fet & Skaar, 2006). However, this prescription is insufficient to ensure the comparability of different declarations, even if they use the same PCR (Modahl *et al.*, 2012). In addition, PCRs are often themselves not unique and can include wide variations in scope and boundary requirements (Subramanian *et al.*, 2012), leading to ambiguity.

An LCA study result is ultimately an assertion that for some constructed PSM, the delivery of a particular reference flow is associated with a certain amount of environmental impact or potential impact. But if the PSM itself is ambiguously stated, then the significance of the result is ambiguous as well. For a computational model to have scientific significance, its results need to be both verifiable and replicable by other scientists (Fomel & Claerbout, 2009; Mesirov, 2010). But in the case of LCA, the structure and contents of the model are never stated precisely; instead they are *described* in a written report, and the description must be interpreted by a reader. When a study is reviewed, it is often the written report and *not* the PSM itself that is the object of review. ISO 14071 (2014), which provides guidelines on critical review, emphasizes that review of the

PSM or individual datasets is optional.

Although the general structure for life cycle assessment can be stated formally, there is no current practice for formally describing the PSM and thus, no automatic mechanisms for interpretation or review of published LCA studies. Moreover, there is no current practice that permits a reviewer to verify the correctness of the computations inherent in an LCA study if access to the PSM itself is not available. The purpose of this paper is to advance a framework for the formal description of product system models that would provide robust transparency and reproducibility of computed results.

2 Disclosing the Product System Model

2.1 Objectives

The task of computing life cycle impact assessment (LCIA) results from a product system model can be described as an "aggregation," in which the data in large, sparse models are combined through linear algebra operations into a small set of numeric scores. Although several equivalent methods for computing aggregated LCI and/or LCIA results have been described in the literature (Suh & Huppes, 2005; Peters, 2007), the computational representation of and exchange of product system models, and the generation of aggregated results, remains software-specific. In the context of this paper, an "LCA Study" should be understood as any collection of inventory data under consideration as a group, from a single-operation unit process to a full ISO-style product LCA.

Reporting the results of an LCA study implies the existence of two parties with different information: the study author has full knowledge of the product system model; and the reader only has knowledge of what the author has disclosed. Obviously, the level of detail of the disclosure will determine how useful it is to the reader. In this context, the reader will want to know the answers to various review objectives. At the same time the author may have constraints that limit what can

be disclosed. The main goal of the framework is to give form to the disclosure so that the requirements of various review objectives, and restrictions associated with various disclosure contraints, can be evaluated clearly.

On the basis of transparency, the framework should meet the following objectives:

- **Computability**. The result should be computed using only the disclosure, plus resources available independently to the reader. The computation should follow the consensus understanding of process LCA computation.
- **Completeness / Minimality**. The computation should use all of the information provided in the disclosure.
- **Reproducibility**. The result obtained by the reader should match the result provided by the author.

One key objective of disclosure is to inspect the work of the author. Consequently, the disclosure should be made up of the author's own work product and should exclude, for instance, data that were drawn from a reference database or other independent source and used in an unmodified form. The purpose of this is twofold: to focus the efforts of the reader on the work of the author, and to enable the reader to obtain and inspect the third-party data independently, thus ensuring its integrity.

- **Authority**. The contents of the disclosure should authentically represent the outcomes of the author's labor in constructing the PSM.
- **Primacy**. The disclosure should exclude information that is made available by a third party. (It must, however, still include previously-published data that were re-implemented by the author in the study.)

In addition, because of the ubiquity of confidential and sensitive information in LCA, any operational disclosure framework must also meet the following:

- **Privacy**. The disclosure should protect information that is held in confidence by the author.

2.2 Mathematical Formulation of LCA

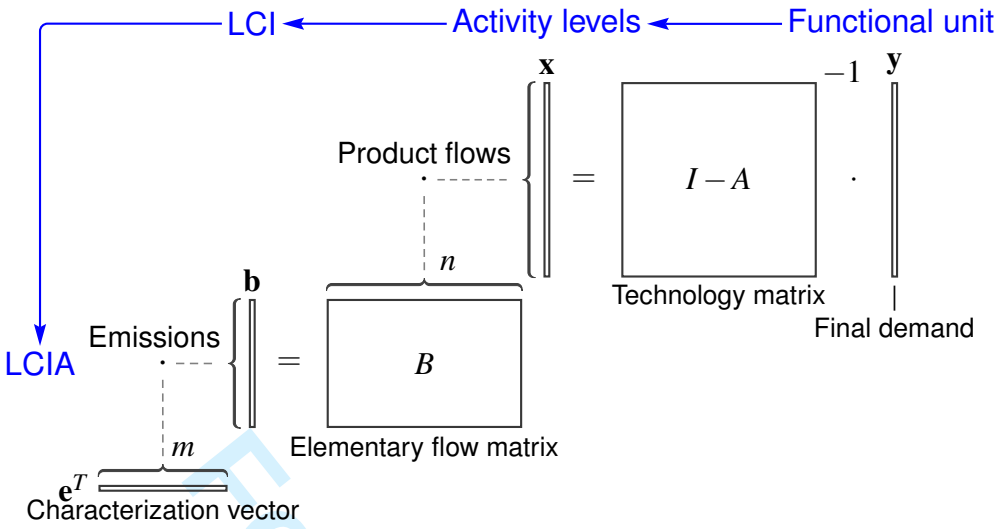
We begin by providing a mathematical formulation of the LCA problem. An LCA computation is performed over a collection of linked unit processes, each of which describes a single operation or reference product. The task of procedure can be described in two steps: making an inventory of environmental interventions; and determining their environmental impacts. If the processes are assumed to be linear and time-invariant, the problem can be written in terms of matrix algebra. There are two common formulations: the "technology matrix" approach and the "input-output" or Leontief approach (Heijungs & Suh, 2002; Pauliuk *et al.*, 2015). Although the two models are understood to be equivalent, the input-output approach is computationally more efficient in some applications because it lends itself to iterative solution (Peters, 2007). That formulation may be stated:

$$s = \mathbf{e}^T \cdot B \cdot (I - A)^{-1} \cdot \mathbf{y} \tag{1}$$

Here A represents the *direct requirements matrix*, where each column of values reports the necessary inputs (positive values) and generated non-reference outputs (negative values)[1] per unit of a process's reference output; B is the environmental intervention or emission matrix; and $\mathbf{e}$ is a column vector of characterization factors for the environmental emissions with respect to a given impact assessment method. In this model, $\mathbf{y}$ is the externally specified final demand, which is the product or service whose delivery is to be assessed, and s is the numerical impact score or category indicator. For simplicity, we will consider the computation of a single impact result. However, it is straightforward to imagine the more typical case, in which E is a matrix of characterization vectors, and $\mathbf{s}$ is a vector of results.

The term $(I - A)$ is equivalent to the normalized technology matrix, i.e. the difference between the

[1] Whether negative values are present or allowed depends on the particular construct or co-production method being used. Majeau-Bettez *et al.* (2014)



$$s = \mathbf{e}^T \cdot \mathbf{b} = \mathbf{e}^T \cdot B \cdot \mathbf{x} = \mathbf{e}^T \cdot B \cdot (I - A)^{-1} \cdot \mathbf{y}$$

Figure 1: Traditional matrix structure of an LCA computation.

supply and use tables in classical input-output analysis, $(V' - U)$ (Suh *et al.*, 2010). The conditions under which this matrix has an inverse, particularly with respect to different strategies for handling co-production, have been discussed extensively (Majeau-Bettez *et al.*, 2014). Eq. 1 is visualized in Figure 1.

In classical LCA, the matrices A (or $I - A$) and B are called a life cycle inventory database (LCIDB), and the functional unit of a study is given by the final demand vector $\mathbf{y}$. Intermediate results of this computation include:

- $\mathbf{x} = (I - A)^{-1} \cdot \mathbf{y}$, the activity level vector;
- $\mathbf{b} = B \cdot \mathbf{x}$, the life cycle inventory;
- $s = \mathbf{e}^T \cdot \mathbf{b}$, the life cycle impact category indicator, the study result.

This formulation implemented in all major LCA software systems, some of which provide the capacity to export the A and B matrices for a given study model. However, the contents of commercial LCIDBs are proprietary and subject to licensing restrictions, and moreover are not usually modified by study authors. Thus exporting A and B does not meet our authority and primacy objectives.

Additionally, replicating the entire LCIDB is cumbersome because of the large size of the matrices. As a consequence, the matrix formulation in Eq. 1 is unsuitable as a disclosure framework.

2.3 Foreground Study

In practice, most LCA studies contain foreground information that is not present in a background database, and LCA software systems generally allow users to augment the standard LCIDBs with foreground models. To meet our disclosure objectives, it is important to be able to distinguish foreground content from background content in the computation. Background databases are prepared independently of any study and do not typically vary across the studies in which they are used. However, the foreground systems themselves cannot be computed without the background. Put another way, the foreground depends on the background but the background does not depend on the foreground. This condition allows the processes in an LCIDB to be partially ordered to group foreground processes together (Kuczenski, 2015).

The simplest conceivable foreground is simply the final demand $\mathbf{y}$, in other words the list of background processes invoked by the study and their activity levels. For an LCA computation as expressed in Eq. 1, the same result can be obtained by relocating the $\mathbf{y}$ vector into an augmented A matrix, augmenting the B matrix with a vector of zeros, and replacing the final demand with a vector $\tilde{\mathbf{y}}$ in which the first entry is 1 and all successive entries are 0:

$$\begin{aligned} s = &\quad \mathbf{e}^T \cdot B \cdot (I - A)^{-1} \cdot \mathbf{y} \\ = &\quad \mathbf{e}^T \cdot \tilde{B} \cdot \left(I - \tilde{A}\right)^{-1} \cdot \tilde{\mathbf{y}} \end{aligned} \tag{2}$$

where $\tilde{A} = \left[\begin{smallmatrix} 0 & \mathbf{0} \\ \mathbf{y} & A \end{smallmatrix}\right]$, $\tilde{B} = [\mathbf{0}, B]$, and $\tilde{\mathbf{y}} = [1, 0, 0, \ldots, 0]^T$. It can easily be shown that Eq. 2 holds for any $\mathbf{y}$. *author's note: a proof for this equality is provided in the electronic supplementary information.* This is intuitively true because the $\tilde{\mathbf{y}}$ selects the first column of the augmented $\tilde{A}$ matrix as the "final demand," but the "direct requirements" of that column are precisely the final demand from the prior, un-augmented A matrix.

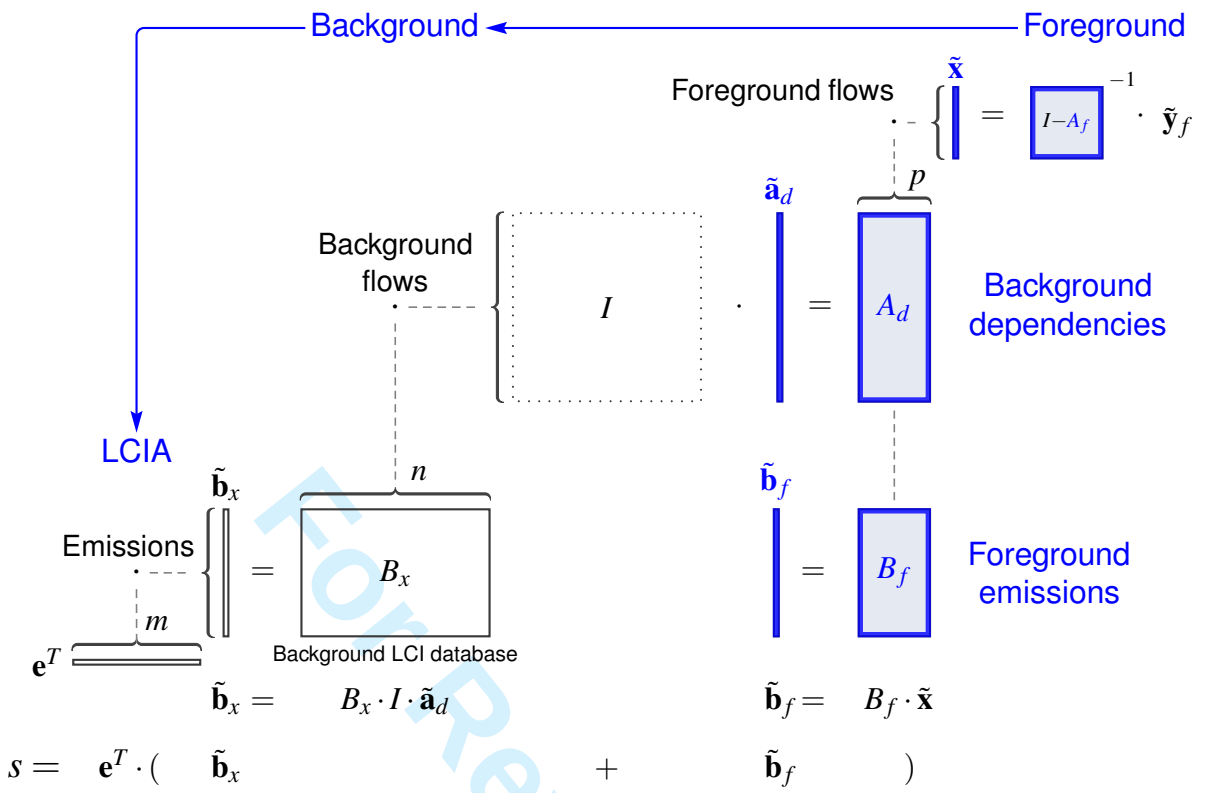


Figure 2: Matrix structure of an LCA foreground study. The background technology matrix has been incorporated into a background LCI database B_x, and replaced in the figure with a placeholder identity matrix I. The foreground and aggregated foreground elements are highlighted.

Of course, most studies have a more elaborate foreground than simply the demand for processes in the background database, and in these cases the A and B matrices can be further augmented with additional foreground content. An LCA study augmented with foreground information can be written in block-triangular form:

$$\tilde{A} = \begin{bmatrix} A_f & 0 \\ A_d & A \end{bmatrix}; \quad \tilde{B} = \begin{bmatrix} B_f & B \end{bmatrix} \tag{3}$$

Here, the submatrix A_f represents the foreground (the first column of which delivers the functional unit of the study); A represents the background, and the rectangular matrix A_d represents the dependency of the foreground on the background. The ordered $\tilde{B}$ matrix is similarly partitioned into B_f, which includes foreground emissions, and B which includes background emissions. The constitutive characteristic of this formulation is that the submatrix in the top right corner of $\tilde{A}$ is

zero. As long as this is the case, all computations regarding the background are invariant with respect to any foreground, and they can be computed in advance. However, it does not apply for so-called "integrated hybrid" models, in which all quadrants contain non-zero cells (for instance, Gibon *et al.* (2015)).

The augmented $\tilde{A}$ and $\tilde{B}$, makes up an *LCA foreground study*, for which $\tilde{\mathbf{y}}$ is the canonical functional unit. This formulation can be applied to the vast majority of LCA case studes, and very likely to all case studies prepared using commercial LCA software.

It is currently common practice for background database maintainers to pre-compute the life cycle inventory results for their databases. Using the notation $B_x = B \cdot (I - A)^{-1}$ to represent the pre-computed background LCI database, Eq.s 3 can be rewritten as:

$$\tilde{A}_{flat} = \begin{bmatrix} A_f & 0 \\ A_d & 0 \end{bmatrix}; \quad \tilde{B}_{flat} = \begin{bmatrix} B_f & B_x \end{bmatrix} \tag{4}$$

The proof of the equivalency of results derived from Eq.s 3 and 4 is provided in the supplementary materials. In this form, the main content of the LCIDB has been removed from the $\tilde{A}$ matrix, and the only information remaining in $\tilde{A}_{flat}$ pertains to the study foreground. Furthermore, the final demand vector has been replaced with a canonical functional unit and no longer needs to be explicitly reported.

Using the ordered matrices, Eq. 1 can be written so that the computationally costly foreground and background inversions can be performed separately:

$$s = \mathbf{e}^T \cdot (B_f + B_x \cdot A_d) \cdot (I - A_f)^{-1} \cdot \tilde{\mathbf{y}}_f \tag{5}$$

where $\tilde{\mathbf{y}}_f$ is a canonical functional unit vector having the same rank as A_f. Eq. 5 is derived in the supplementary materials. Eq. 5 is visualized in Figure 2, with the background technology matrix A, having been subsumed into B_x, represented by an identity matrix I.

The activity levels of the foreground nodes are represented by $\tilde{\mathbf{x}}$, which is the result of $\tilde{\mathbf{y}}_f$ selecting the first column of $(I - A_f)^{-1}$. This vector can be used to describe an *aggregated foreground*:

$$\begin{aligned} \tilde{\mathbf{b}}_f &= B_f \cdot \tilde{\mathbf{x}} \\ \tilde{\mathbf{a}}_d &= A_d \cdot \tilde{\mathbf{x}} \\ \tilde{\mathbf{b}}_x &= B_x \cdot \tilde{\mathbf{a}}_d \end{aligned} \tag{6}$$

The aggregated result can also be computed from the aggregated foreground:

$$\begin{aligned} s &= \mathbf{e}^T \cdot (B_f + B_x \cdot A_d) \cdot \tilde{\mathbf{x}} \\ &= \mathbf{e}^T \cdot (\tilde{\mathbf{b}}_f + B_x \cdot \tilde{\mathbf{a}}_d) \\ &= \mathbf{e}^T \cdot (\tilde{\mathbf{b}}_f + \tilde{\mathbf{b}}_x) \end{aligned} \tag{7}$$

2.4 Contents of a Disclosure

The foreground matrix A_f, dependency matrix A_d, and foreground emission matrix B_f now contain all study-specific information. Assuming that the appropriate background LCIDB B_x, and the vector of characterization factors $\mathbf{e}$ (or their product $\mathbf{e} \cdot B_x$) are available and held in common between the author and the reader, then a disclosure that precisely describes A_f, A_d and B_f is sufficient to reproduce the study result, as formulated in Eq. 5. These three submatrices together can be considered a complete representation of the PSM.

Thus, the problem of how to disclose an LCA study design is reduced to the problem of how to communicate the contents of these submatrices accurately. Recognizing that all three matrices are likely to be sparse (that is, that most of the entries are zero), the most efficient disclosure would be to only report non-zero entries. A sparse matrix can be represented by a list of 3-tuples indicating (row, column, value) for each non-zero entry. Thus, a formal disclosure of the PSM can be stated in six parts:

d-i. An ordered list of foreground nodes (rows/columns of A_f);

d-ii. An ordered list of background flows, making reference to a particular background database (rows of A_d, mapping to columns of B_x);

d-iii. An ordered list of exchanges with environment (rows of B_f, mapping to entries in **e**);

d-iv. A list of 3-tuples for nonzero elements of A_f, in which the row and column are indices into Item *d-i*;

d-v. a list of 3-tuples for nonzero elements of A_d, in which the rows are indices into Item *d-ii* and the columns are indices into Item *d-i*;

d-vi. a list of 3-tuples for nonzero elements of B_f, in which the rows are indices into Item *d-iii* and the columns are indices into Item *d-i*.

A disclosure containing these six elements satisfies most of the requirements established at the top of this section (the exception is ensuring the privacy of confidential data, which is treated later in the article):

- The result is **computable** using Eq. 5;
- the computation is **complete** and **minimal** because all components are required and no superfluous components are included;
- The results are **reproducible** if the reader is able to accurately interpret the author's references in *d-ii* and *d-iii*;
- The disclosure includes only information for which the author can claim **authority**;
- The disclosure includes only information for which the study is the **primary** source.

As mentioned, the proper interpretation of the disclosure requires that the reader accurately interprets the author's references to background databases and elementary flows, including mutually consistent implementation of the LCIA vector **e**. These aspects are out of scope for the purposes of this paper, and we will assume that a shared understanding between author and reader can be reached. However, the the challenges of operationalizing this requirement are complex (see Dis-

cussion and Supplementary materials).

3 Reviewing Product System Models

Objectives in critical review vary with the study and the reviewer, but they generally include consistency with the ISO 14044 standard, the validity of methods, appropriateness of data used, correctness of interpretations, and transparency of reporting (ISO 14044, 2006, Section 6.1). Although the object of critical review is typically a written report, the use of a structured disclosure of a PSM introduces the possibility that reviewers can inspect the model itself and even verify the computations and reported results. This section discusses how a disclosure in the format described above can enable robust review of a product system model. Note that, for purposes of critical review, the objective for data privacy can be omitted because the reviewer may be granted privileged access to confidential data. We return to the data privacy requirement in the next section.

3.1 Interpreting the Foreground

The foreground of an LCA study is a collection of nodes that represent points in the inventory model where flows are exchanged. The foreground matrix A_f describes in precise terms what components are included in the product system and how they connect to one another. The foreground is modelled as a weighted directed graph in which there is a 1:1 correspondence between nodes and product flows. The foreground matrix A_f reports the adjacency and weights of edges in the graph, and thus describes the relationships among these nodes. These nodes typically represent unit processes, but they can also indicate exchanges across an implicit system boundary that encloses the foreground. Generally, the foreground matrix A_f can represent any process-flow model with a 1:1 correspondence between processes and flows, i.e. any allocated supply-and-use table, but some fundamental designs are common. Some simple foreground models are shown in the supplementary materials.

3.2 Dependencies, Emissions, and Cutoffs

The foreground matrix only describes the structure of the PSM; it does not provide any information about inventory or impacts. The inventory requirements and environmental burdens associated with the model are all described in the dependency matrix A_d and the foreground emission matrix B_f. The dependency matrix shows how foreground nodes depend on activities provided by processes in a background database; the emission matrix shows direct exchanges with the environment. In both cases, the respective columns of A_d and B_f correspond with columns in A_f and denote dependencies and/or emissions of the corresponding nodes. Review of the dependencies of foreground nodes must include review of the selection of background processes to which they are connected (i.e. the identities of the columns of B_x referred to in disclosure item *d-ii*, the list of background flows). Likewise, the identities of the emissions (entries in disclosure item *d-iii*) must also be reviewed.

Interpretation of the contents of these matrices thus hinges on the contents of A_f. If a column in A_f is nonempty, then the corresponding node has exchanges with other nodes in the foreground. If columns of A_d and/or B_f are empty, then the node is simply a pass-through process that serves to connect different foreground elements. However, if A_d and/or B_f have nonzero entries in that column, then that indicates some transformation process or emission is occurring at that point in the foreground.

On the other hand, if a column in A_f is empty, then that is an indication that the corresponding product flow is crossing the foreground system boundary. In order for this flow to contribute any impacts, it must either be "terminated" in the dependency matrix, or represented as an emission.

Flows whose entire column consists of 0s throughout A_f, A_d, and B_f are "cut-offs" that exit the model with zero burdens. Specifying cut-off criteria and reporting of cut-off flows is an important part of system boundary definition. Thus, reviewing cut-off flows (including evaluating their significance) is a crucial part of review that is facilitated by having a structured disclosure of the model.

4 Foreground Aggregation

When an LCA study result is computed, the foreground and background matrices are aggregated into a scalar indicator score (or a small vector of a few scores). The formulation in Eq. 5 and the illustration in Fig. 2 provides a framework for reporting results at varying levels of aggregation. In this section, I present several different aggregation computations that provide equivalent results for an LCA study. After that, I show how a foreground containing private data can be partitioned into public and private portions, each of which can be comprehensively checked by a reviewer having privileged access, and subsequently disclosed in a manner that protects the private data.

4.1 Forms of Aggregation

Figure 2 illustrates the LCA computation in various successive aggregation steps, from full matrices to vectors to a scalar value. The level of aggregation provided in a disclosure will constrain what review objectives can be met by the reader of a study. Different aggregation forms are illustrated in Figure 3.

4.1.1 Unit Process Inventory

A unit process makes up a single shared column of A_f, A_d, and B_f, as depicted in Figure 3a. Documentation of unit processes is specified in ISO 14048, and several compliant formats exist, the most well-known being the ILCD and Ecospold XML formats. ISO 14048 specifies a distinction between intermediate and elementary flows, but does not include the concept of a model foreground. In a matrix representation, each unit process must have exactly one reference flow, which corresponds to its row and column in A_f.

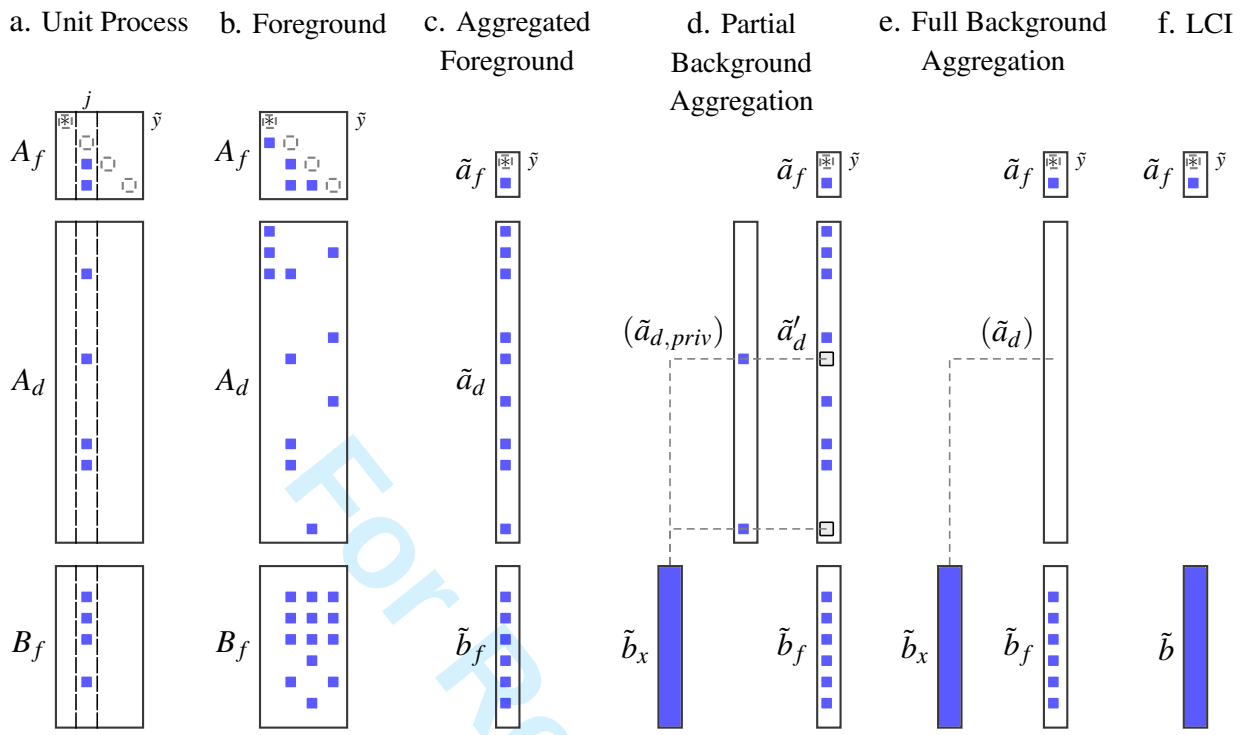


Figure 3: Matrix structures for different forms of foreground aggregation.

4.1.2 Foreground or Foreground Fragment

See Figure 3b. A foreground can be viewed as a linked set of unit processes. The collection of processes themselves make up the foreground of the study, and so exchanges among them are recorded in A_f, and exchanges outside the foreground are represented in A_d or B_f. A complete foreground represents the full disclosure of a model, and a reader can replicate and extend it independently. A fragment of the PSM can also be published as a complete foreground. 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.

4.1.3 Aggregated Foreground

See Figure 3c. Any collection of foreground nodes can also be expressed in aggregated form by performing the computations in Eq. 6. A publication of the aggregated foreground is identical to the publication of a unit process, except that the result was derived through aggregation. As in the unit process case, the aggregated foreground has exactly one reference flow, the canonical functional unit $\tilde{\mathbf{y}}$. The aggregated foreground still reports explicit links between the model and a background database and can thus be used to review data set selection.

4.1.4 Partial Background Aggregation

See Figure 3d. Using this approach, the aggregated dependency vector is split into two parts that sum to the original:

$$\tilde{\mathbf{a}}_d = \tilde{\mathbf{a}}_{d,priv} + \tilde{\mathbf{a}}'_d \tag{8}$$

The disclosed dependencies $\tilde{\mathbf{a}}'_d$ are reported, and the private dependencies $\tilde{\mathbf{a}}_{d,priv}$ are replaced with an aggregated background inventory derived from the background database, $\tilde{\mathbf{b}}_{x,agg}$. This approach can be used when a reader lacks access to the background inventory sets referred to in $\tilde{\mathbf{a}}_{d,priv}$, or when the author wishes not to disclose the private dependencies. Foreground emissions $\tilde{\mathbf{b}}_f$ can be reported separately from the aggregation result, but often in current practice they are not distinguished from the background flows.

4.1.5 Full Background Aggregation and LCI

See Figure 3e and f. At this level of aggregation, the entire dependency vector is replaced with an aggregated life cycle inventory vector $\tilde{\mathbf{b}}_x$ derived from the background LCIDB. The reader no longer requires access to any background database to perform the computation, but all dependency information is concealed. Again, foreground emissions are often not distinguished from the aggregation result (Figure 3f). The life cycle inventory $\tilde{\mathbf{b}}$ provides the most aggregated form of the study

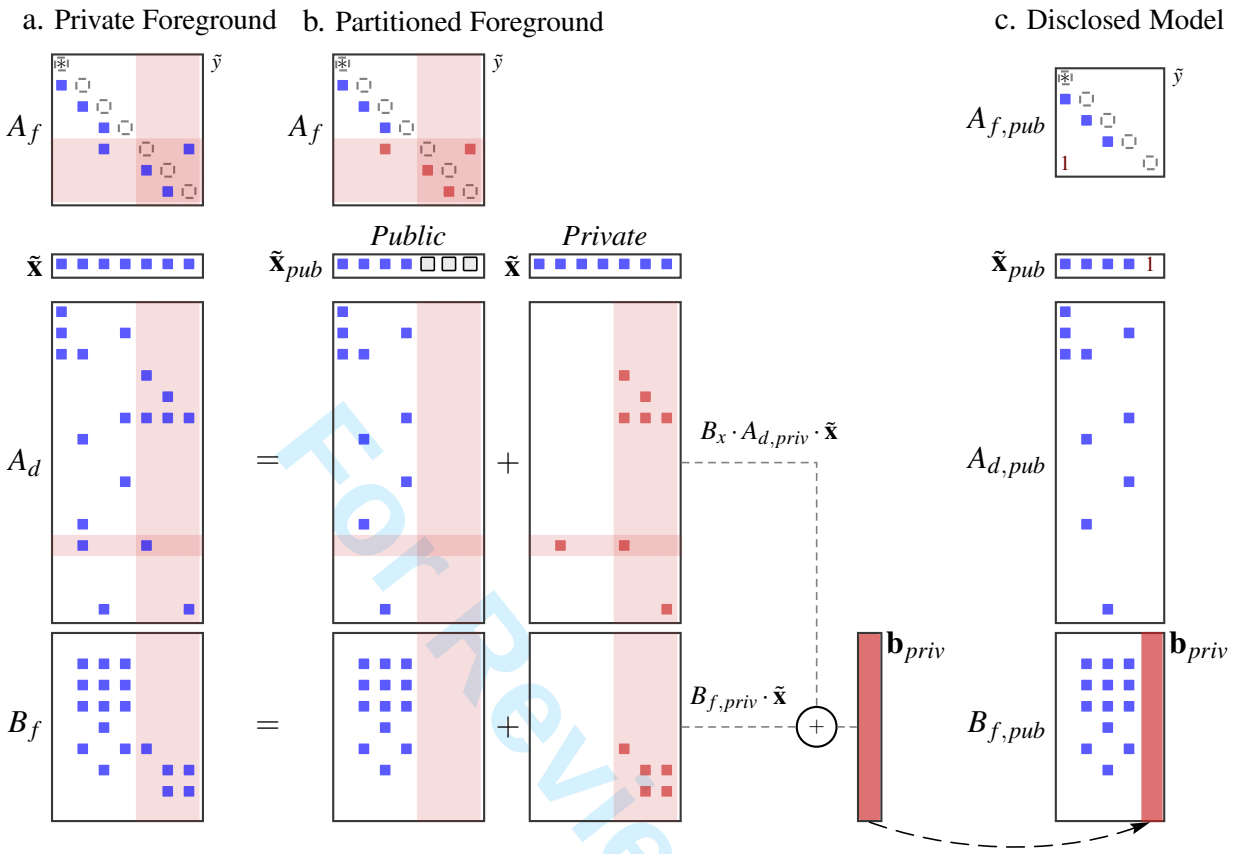


Figure 4: Partitioning the foreground into private and public segments for disclosure. The shaded sections in (a) and (b) indicate disclosure constraints, while in (c) the private data have been replaced with a unit-weighted aggregated column $\mathbf{b}_{priv}$ in B_f.

that can still be independently validated with an external characterization vector $\mathbf{e}$.

4.2 Reviewable Private Aggregation

Often, a study author may wish to make a partial disclosure of a PSM that retains the privacy of confidential data, while still permitting comprehensive critical review of the complete model. In this case it is necessary to grant a reviewer privileged access to the private data in order to meet the review objectives. The study formulation in Eq. 5 may be used to partition a foreground into public and private portions in order to comply with any applicable constraints on what may be disclosed. The approach is illustrated in Figure 4.

The first step is to identify the constraints on the disclosure. In the framework presented here, constraints may include the locations or values of any subset of nonzero entries in the A_d and B_f matrices, as well as any subset of nodes in A_f. Fig. 4a shows a PSM in which disclosure constraints are indicated by shaded regions of the matrices. In this case the nodes represented by the three rightmost columns of A_f, as well as one entire row of A_d, are to remain secret.

The A_d and B_f matrices are then partitioned into two components that sum to the original:

$$\begin{aligned} A_d &= A_{d,pub} + A_{d,priv} \\ B_f &= B_{f,pub} + B_{f,priv} \end{aligned} \tag{9}$$

This is illustrated in Fig. 4b. Substituted back into Eq. 5, the private portions are aggregated into a private life cycle inventory vector $\mathbf{b}_{priv}$, while the public portions remain disaggregated:

$$s = \mathbf{e}^T \cdot \left((B_{f,pub} + B_{f,priv}) + B_x \cdot (A_{d,pub} + A_{d,priv}) \right) \cdot \tilde{\mathbf{x}} \tag{10}$$

$$= \mathbf{e}^T \cdot \left((B_{f,pub} + B_x \cdot A_{d,pub}) \cdot \tilde{\mathbf{x}} + \mathbf{b}_{priv} \right) \tag{11}$$

$$\mathbf{b}_{priv} = (B_{f,priv} + B_x \cdot A_{d,priv}) \cdot \tilde{\mathbf{x}} \tag{12}$$

Any foreground nodes that are completely contained within the private partition can be omitted from the disclosed foreground. The aggregation result can then be included as a separate foreground node with a unit weight, and the aggregated inventory vector included as the corresponding column of B_f. This is illustrated in Fig. 4c.

The author may relax the disclosure constraints by reporting the locations of some or all non-zero entries in $A_{d,priv}$ and $B_{f,priv}$ but not disclosing their values, or disclosing a range that includes the actual value. This would enable a reader to understand *that* a certain background process or emission was included in the model (and checked by the critical reviewer) without knowing how much, which may jointly satisfy review objectives and disclosure constraints in some cases.

5 Discussion

5.1 Comparative Review: Formalizing Scope and System Boundary

The disclosure of a PSM contains six parts, including three lists and three tables of sparse matrix data. The graph structure encoded in the foreground matrices can be automatically rendered as a process-flow diagram, supplementing or supplanting the hand-crafted system diagrams that currently accompany ISO reports. While the numeric data are required to critically evaluate and reproduce the study results, substantial information about the study design can be found in the list of background dependencies (*d-ii*) and emissions (*d-iii*). These lists report all the distinct sources of environmental impacts in the model. The cutoff flows, which are foreground nodes that have no nonzero entries in A_f, A_d, or B_f, report on the other hand what aspects were explicitly excluded from the model.

Together, these lists can be considered a functional definition of the study scope that can be used to conduct a qualitative comparison of multiple studies by different authors. Background database choice, dataset selection, and version information is all made available in disclosure item *d-ii*, thereby streamlining what is presently an interpretive and study-specific process of extracting this information from written text. Similarly, the set of elementary flows explicitly modeled in the study foreground (item *d-iii*) expose whether any study-specific modeling of direct emissions was performed and what flows were included.

Moreover, because the disclosure of the PSM includes only references to background data without including the data themselves, it is trivial for item *d-ii* to include references to a variety of background databases without adding any additional requirements. For instance, a study that makes reference to a set of EPDs published by one firm, while using a background LCI database from another firm, can express both dependencies in an analogous way, even if the two sources are published independently.

5.2 Privacy and Partial Aggregation

Partial aggregation is already used in practice to conceal private data, but current methods do not provide a way for a reviewer to validate the result. If a study containing confidential information is split into public and private partitions as in Eq. 11 and Fig. 4, then the results of the computation can be validated directly by the reviewer. To achieve this, the reviewer must be granted access to both the public and private partitions in order to confirm that the partitioning is valid according to Eq. 9, and that $\mathbf{b}_{priv}$ was computed correctly. However, the private information used in the computation can still be withheld from the disclosure.

A further advantage to this approach is that the reader can evaluate easily what fraction of the overall impact score is accounted for by the public versus the private portions of the model. A "disclosure completeness" metric φ can be defined as the fraction of the overall score which is accounted for by the disclosed portion of the PSM:

$$\varphi = \frac{\mathbf{e}^T \cdot (\mathbf{b} - \mathbf{b}_{priv})}{\mathbf{e}^T \cdot \mathbf{b}} = 1 - \frac{\mathbf{e}^T \cdot \mathbf{b}_{priv}}{\mathbf{e}^T \cdot \mathbf{b}} \tag{13}$$

where $\mathbf{b}$ is the complete life cycle inventory. The value of this metric can provide an indication of the level of transparency of the disclosure with respect to a given impact category indicator.

5.3 Attacks on Privacy

It is not thought possible for a reader to discern any information about a disaggregated system solely from an aggregated disclosure such as $\mathbf{b}_{priv}$ (UNEP/SETAC, 2011, Ch. 3), although that assumption has never been rigorously tested. Compressed sensing is a signal processing technique that seeks solutions to an underdetermined linear system, in which there are more unknowns than equations, based on the assumption that the solution is sparse (Donoho, 2006). A form of linear optimization called "basis pursuit" can be used to try to detect the signal.

In the LCA case, if the private LCI result $\mathbf{b}_{priv}$ includes inventory data from a background database but no foreground emissions, then it can be written as:

$$\mathbf{b}_{priv} = B_x \cdot \mathbf{a}_{d,priv} \tag{14}$$

Typically this equation is highly underdetermined (In ecoinvent, for instance, B_x has roughly 1,800 rows and over 12,000 columns). If the private input $\mathbf{a}_{d,priv}$ contains only a few nonzero elements, and the identity of B_x is known, then $\mathbf{a}_{d,priv}$ may be vulnerable to a basis pursuit attack. Further research is required to evaluate whether this attack can be implemented in a real situation, and how it may be defended against.

5.4 Stable Semantic References to Study Elements

In practice, while it is easy to reproduce a set of sparse matrices, there is considerable potential for ambiguity in identifying the background datasets and emissions to which those matrix entries correspond. In order to achieve the transparency and reproducibility promised by this disclosure framework, it must be possible for study authors and readers to agree on the meaning of those references. In particular, the database configuration, dataset version, reference flow and dimension (quantitative unit of measure) must be unambiguous and easy to interpret.

Agreement about the meaning of concepts is often accomplished using linked semantic data (Bizer *et al.*, 2009), in which an object to be interpreted is signified by a link to a resource on the World Wide Web, typically referred to as a Uniform Resource Identifier (URI) or a hyperlink. The resource at the end of the link functions as a point of agreement: both study author and data user can follow it from anywhere on the Internet to obtain the same information. The content pointed to by the hyperlink, along with context provided by *other* references to the same content, gives meaning to the information and allows it to be curated by the study authors and other members of the community (Khan *et al.*, 2011).

Before the goal of easily reproducible foreground models can be realized, this linked data foundation must be laid. Background database providers should ensure that their users have access to *stable semantic references* for publicly available datasets that denote a particular activity in a particular database version and configuration. These references should take the form of URIs that can be accessed using any Internet browser and can provide both author and reader with documentary information describing the referred dataset, the available reference flows and their dimensions, and a mechanism for the reader to obtain access to the exchange data or LCI/LCIA computations for the purposes of model validation. A dependency reference in disclosure item *d-ii* should include both an activity and a reference flow.

Similarly, LCIA method providers should ensure that their methods make use of stable semantic references to elementary flows and contexts. Current research has revealed widespread challenges to finding consistency on the identities of elementary flows despite the community's longstanding awareness of the problem, and several mutually inconsistent reference flow sets exist (Edelen *et al.*, 2017). The conventional understanding of a "flow" as comprising a substance and a context together (e.g. methane, emissions to air) causes a combinatorial increase in the number of flows in a database, and also multiplies the potential points of disagreement across sources. I reiterate Edelen et al.'s (ibid.) suggestion to explicitly regard substances (which can easily include non-material "flowables" such as land occupation and transformation) and contexts as independent semantic classes. In this case, an elementary flow entry in the disclosure item *d-iii* would require both a flowable and a context.

6 Conclusion

The current practice of documenting a PSM with a written report sharply limits the ability of readers to interpret and reuse the data. It also results in the model being frozen to a particular configuration or set of configurations used to generate the published results. Any review objectives

involving sensitivity of the model to parameter variations and/or alternative scenarios are limited to the selections considered by the author. When a study result is published in an aggregated form, often it is beyond the capacity of the critical reviewer to validate the aggregation result, even if the reviewer has privileged access to confidential materials used in preparing the report.

This paper presents a possible solution by providing a mathematical formalization of the foreground of an LCA study and a functional specification for disclosing the PSM. A data user or critical reviewer with access to the six disclosure elements, and also adequate access to background datasets and elementary flows used in the model, can reproduce the LCA computation in Eq. 5, and ensure the reported result is correct. Once equipped with the PSM, the reviewer or data user can potentially go much further. Reviewers or data users who are able to reproduce the PSM from the disclosure can also modify it, adding or removing elements, altering dataset selections, and applying other impact assessment methods. Portions of the model which have been pre-aggregated can remain private in the disclosure. Further research is required to determine whether or under what conditions aggregation results can be reverse-engineered.

If data providers also make available stable semantic references to background data sets and elementary flow characterizations, then it becomes possible for a stand-alone description of a PSM to be used to reproduce an LCA result. The same description can be extended or incorporated into subsequent studies by other authors. From this prescription, a framework for distributed, platform-independent LCA computation can be imagined.

References

Bizer, Christian, Heath, Tom, & Berners-Lee, Tim. 2009. Linked Data - The Story So Far. *International Journal on Semantic Web and Information Systems*, **5**(3), 1–22. doi:10.4018/jswis.2009081901.

Brandão, Miguel, Clift, Roland, Cowie, Annette, & Greenhalgh, Suzie. 2014. The Use of Life Cycle Assessment in the Support of Robust (Climate) Policy Making: Comment on "Using Attributional Life Cycle Assessment to Estimate Climate-Change Mitigation …". *Journal of Industrial Ecology*, **18**(3), 461–463. doi:10.1111/jiec.12152.

Cleary, Julian. 2009. Life cycle assessments of municipal solid waste management systems: a

comparative analysis of selected peer-reviewed literature. *Environment international*, **35**(8), 1256–1266. PMID: 19682746. doi:10.1016/j.envint.2009.07.009.

Clift, R., Frischknecht, R., Huppes, G., Tillman, A.-M., & Weidema, B. 1998. *Toward a Coherent Approach to Life Cycle Inventory Analysis. Report of the Working Group on Inventory Enhancement*. Tech. rept. SETAC-Europe.

Curran, Mary Ann, & Young, Steven B. 2014. Critical review: a summary of the current state-of-practice. *Int. J. Life Cycle Assess.*, **19**(10), 1667–1673. doi:10.1007/s11367-014-0778-2.

Donoho, D.L. 2006. Compressed sensing. *IEEE Transactions on Information Theory*, **52**(4), 1289–1306. doi:10.1109/tit.2006.871582.

Edelen, Ashley, Ingwersen, Wesley W., Rodríguez, Cristina, Alvarenga, Rodrigo A. F., de Almeida, Artur Ribeiro, & Wernet, Gregor. 2017. Critical review of elementary flows in LCA data. *The International Journal of Life Cycle Assessment*, jul. doi:10.1007/s11367-017-1354-3.

Fet, Annik Magerholm, & Skaar, Christofer. 2006. Eco-labeling, Product Category Rules and Certification Procedures Based on ISO 14025 Requirements (6 pp). *The International Journal of Life Cycle Assessment*, **11**(1), 49–54. doi:10.1065/lca2006.01.237.

Finnveden, Göran. 1999. Methodological aspects of life cycle assessment of integrated solid waste management systems. *Resources, Conservation and Recycling*, **26**(3-4), 173–187. doi:10.1016/s0921-3449(99)00005-1.

Fomel, Sergey, & Claerbout, Jon F. 2009. Guest Editors' Introduction: Reproducible Research. *Computing in Science & Engineering*, **11**(1), 5–7. doi:10.1109/mcse.2009.14.

Gibon, Thomas, Wood, Richard, Arvesen, Anders, Bergesen, Joseph D., Suh, Sangwon, & Hertwich, Edgar G. 2015. A Methodology for Integrated, Multiregional Life Cycle Assessment Scenarios under Large-Scale Technological Change. *Environmental Science & Technology*, **49**(18), 11218–11226. doi:10.1021/acs.est.5b01558.

Heath, Garvin A., & Mann, Margaret K. 2012. Background and Reflections on the Life Cycle Assessment Harmonization Project. *Journal of Industrial Ecology*, **16**, S8–S11. doi:10.1111/j.1530-9290.2012.00478.x.

Heijungs, Reinout, & Suh, Sangwon. 2002. *The computational structure of life cycle assessment*. Vol. 11. Springer Science & Business Media.

Henriksson, Patrik John Gustav, Guinée, Jeroen Bartholomeus, Heijungs, Reinout, Koning, Arjan, & Green, Darren Michael. 2014. A protocol for horizontal averaging of unit process data-including estimates for uncertainty. *The International Journal of Life Cycle Assessment*, **19**(2), 429–436. doi:10.1007/s11367-013-0647-4.

Ingwersen, Wesley W., Hawkins, Troy R., Transue, Thomas R., Meyer, David E., Moore, Gary, Kahn, Ezra, Arbuckle, Peter, Paulsen, Heidi, & Norris, Gregory A. 2015. A new data architecture for advancing life cycle assessment. *Int. J. Life Cycle Assess.*, **20**(4), 520–526. doi:10.1007/s11367-015-0850-6.

ISO 14044. 2006. *Environmental management — Life cycle assessment — Requirements and guidelines*. ISO, Geneva, Switzerland.

ISO 14071. 2014. *Environmental management — Life cycle assessment — Critical review processes and reviewer competencies: Additional requirements and guidelines to ISO 14044:2006*. ISO, Geneva, Switzerland.

Khan, Huda, Caruso, Brian, Corson-Rikert, Jon, Dietrich, Dianne, Lowe, Brian, & Steinhart, Gail. 2011. DataStaR: Using the Semantic Web approach for Data Curation. *International Journal of Digital Curation*, **6**(2), 209–221. doi:10.2218/ijdc.v6i2.197.

Kuczenski, Brandon. 2015. Partial ordering of life cycle inventory databases. *The International Journal of Life Cycle Assessment*, **20**(12), 1673–1683. doi:10.1007/s11367-015-0972-x.

Kuczenski, Brandon, Sahin, Cetin, & Abbadi, Amr El. 2016. Privacy-preserving aggregation in life cycle assessment. *Environment Systems and Decisions*, **37**(1), 13–21. doi:10.1007/s10669-016-9620-7.

Laurent, Alexis, Clavreul, Julie, Bernstad, Anna, Bakas, Ioannis, Niero, Monia, Gentil, Emmanuel, Christensen, Thomas H., & Hauschild, Michael Z. 2014. Review of LCA studies of solid waste management systems – Part II: Methodological guidance for a better practice. *Waste Management*, **34**(3), 589–606. doi:10.1016/j.wasman.2013.12.004.

Lloyd, Shannon M., & Ries, Robert. 2007. Characterizing, Propagating, and Analyzing Uncertainty in Life-Cycle Assessment: A Survey of Quantitative Approaches. *Journal of Industrial Ecology*, **11**(1), 161–179. doi:10.1162/jiec.2007.1136.

Majeau-Bettez, Guillaume, Wood, Richard, & Strømman, Anders Hammer. 2014. Unified Theory of Allocations and Constructs in Life Cycle Assessment and Input-Output Analysis. *Journal of Industrial Ecology*, **18**(5), 747–770. doi:10.1111/jiec.12142.

McManus, Marcelle C., Taylor, Caroline M., Mohr, Alison, Whittaker, Carly, Scown, Corinne D., Borrion, Aiduan Li, Glithero, Neryssa J., & Yin, Yao. 2015. Challenge clusters facing LCA in environmental decision-making—what we can learn from biofuels. *The International Journal of Life Cycle Assessment*, **20**(10), 1399–1414. doi:10.1007/s11367-015-0930-7.

Mesirov, J. P. 2010. Accessible Reproducible Research. *Science*, **327**(5964), 415–416. doi:10.1126/science.1179653.

Mila i Canals, L., Asselin, A.-C., Cruypenninck, H., Liu, S., Ingwersen, W., & Macdonald, T. 2015. Towards a Global Network of Interoperable LCA Databases. *In: Proceedings of SETAC Europe 25th Annual Meeting (SETAC Europe 2015)*.

Modahl, Ingunn Saur, Askham, Cecilia, Lyng, Kari-Anne, Skjerve-Nielssen, Camilla, & Nereng, Guro. 2012. Comparison of two versions of an EPD, using generic and specific data for the foreground system, and some methodological implications. *The International Journal of Life Cycle Assessment*, **18**(1), 241–251. doi:10.1007/s11367-012-0449-0.

Pauliuk, Stefan, Majeau-Bettez, Guillaume, & Müller, Daniel B. 2015. A General System Structure and Accounting Framework for Socioeconomic Metabolism. *Journal of Industrial Ecology*, **19**(5), 728–741. doi:10.1111/jiec.12306.

Pelletier, Nathan, Ardente, Fulvio, Brandão, Miguel, Camillis, Camillo De, & Pennington, David. 2014. Rationales for and limitations of preferred solutions for multi-functionality problems in LCA: is increased consistency possible? *The International Journal of Life Cycle Assessment*, **20**(1), 74–86. doi:10.1007/s11367-014-0812-4.

Peters, Glen P. 2007. Efficient algorithms for Life Cycle Assessment, Input-Output Analysis, and Monte-Carlo Analysis. *Int. J. Life Cycle Assess.*, **12**(6), 373–380. doi:10.1065/lca2006.06.254.

Plevin, Richard J., Delucchi, Mark A., & Creutzig, Felix. 2013. Using Attributional Life Cycle Assessment to Estimate Climate-Change Mitigation Benefits Misleads Policy Makers. *Journal of Industrial Ecology*, **18**(1), 73–83. doi:10.1111/jiec.12074.

Rainville, Anne, Hawkins, Richard, & Bergerson, Joule. 2015. Building consensus in life cycle assessment: the potential for a Canadian product category rules standard to enhance credibility in greenhouse gas emissions estimates for Alberta's oil sands. *Journal of Cleaner Production*, **103**(sep), 525–533. doi:10.1016/j.jclepro.2014.06.067.

Reap, John, Roman, Felipe, Duncan, Scott, & Bras, Bert. 2008. A survey of unresolved problems in life cycle assessment Part 1: goal and scope and inventory analysis. *Int. J. Life Cycle Assess.*, **13**(4), 290–300.

Recchioni, Marco, Mathieux, Fabrice, Goralczyk, Malgorzata, & Schau, Erwin M. 2013. *ILCD Data Network and ELCD Database: current use and further needs for supporting Environmental Footprint and Life Cycle Indicator Projects*. Tech. rept. EUR 25744 EN. European Commission, Joint Research Centre, Luxembourg.

Subramanian, Vairavan, Ingwersen, Wesley, Hensler, Connie, & Collie, Heather. 2012. Comparing product category rules from different programs: learned outcomes towards global alignment. *The International Journal of Life Cycle Assessment*, **17**(7), 892–903. doi:10.1007/s11367-012-0419-6.

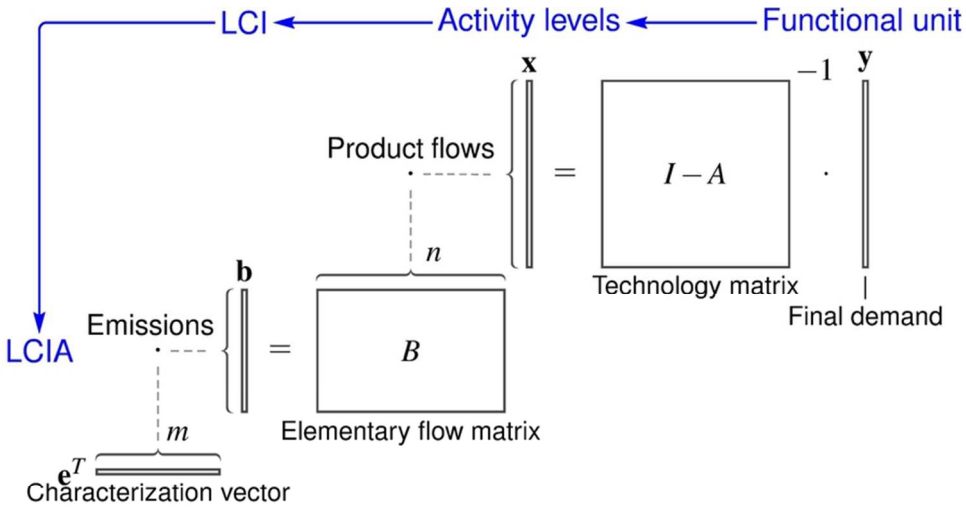
Suh, Sangwon, & Huppes, Gjalt. 2005. Methods for Life Cycle Inventory of a product. *Journal of Cleaner Production*, **13**(7), 687 – 697. doi:http://dx.doi.org/10.1016/j.jclepro.2003.04.001.

Suh, Sangwon, Weidema, Bo, Schmidt, Jannick Hoejrup, & Heijungs, Reinout. 2010. Generalized Make and Use Framework for Allocation in Life Cycle Assessment. *J. Ind. Ecol.*, **14**(2), 335–353. doi:10.1111/j.1530-9290.2010.00235.x.

Turconi, Roberto, Boldrin, Alessio, & Astrup, Thomas. 2013. Life cycle assessment (LCA) of electricity generation technologies: Overview, comparability and limitations. *Renewable and Sustainable Energy Reviews*, **28**(dec), 555–565. doi:10.1016/j.rser.2013.08.013.

UNEP/SETAC. 2011. *Global Guidance Principles for Life Cycle Assessment Databases*. Tech. rept. United Nations Environment Programme.

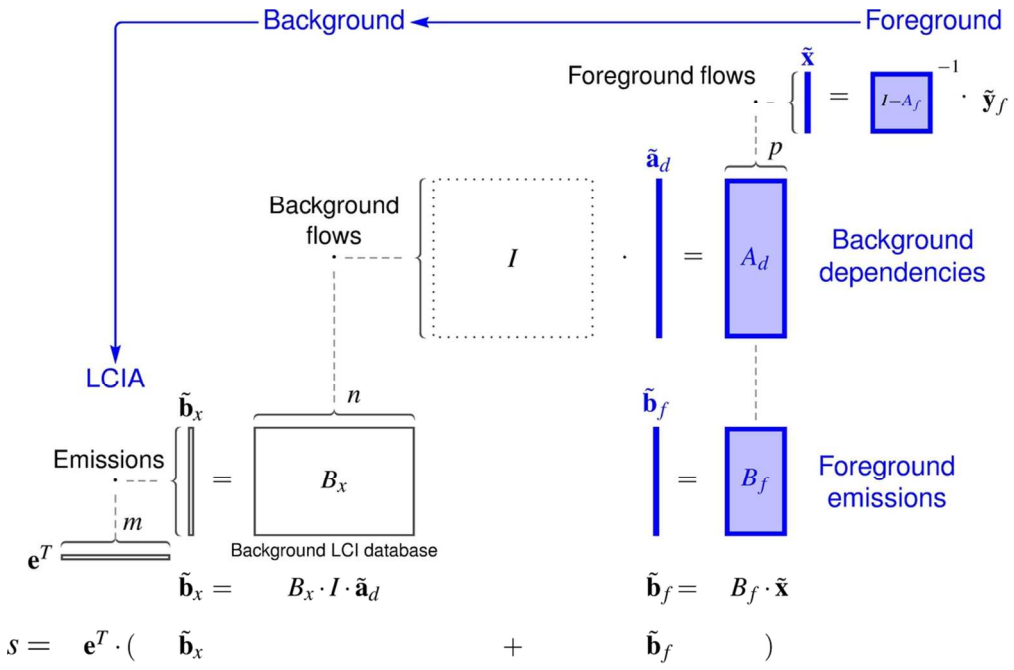
van der Harst, Eugenie, & Potting, José. 2013. A critical comparison of ten disposable cup LCAs. *Environmental Impact Assessment Review*, **43**(nov), 86–96. doi:10.1016/j.eiar.2013.06.006.



$$s = \mathbf{e}^T \cdot \mathbf{b} = \mathbf{e}^T \cdot B \cdot \mathbf{x} = \mathbf{e}^T \cdot B \cdot (I-A)^{-1} \cdot \mathbf{y}$$

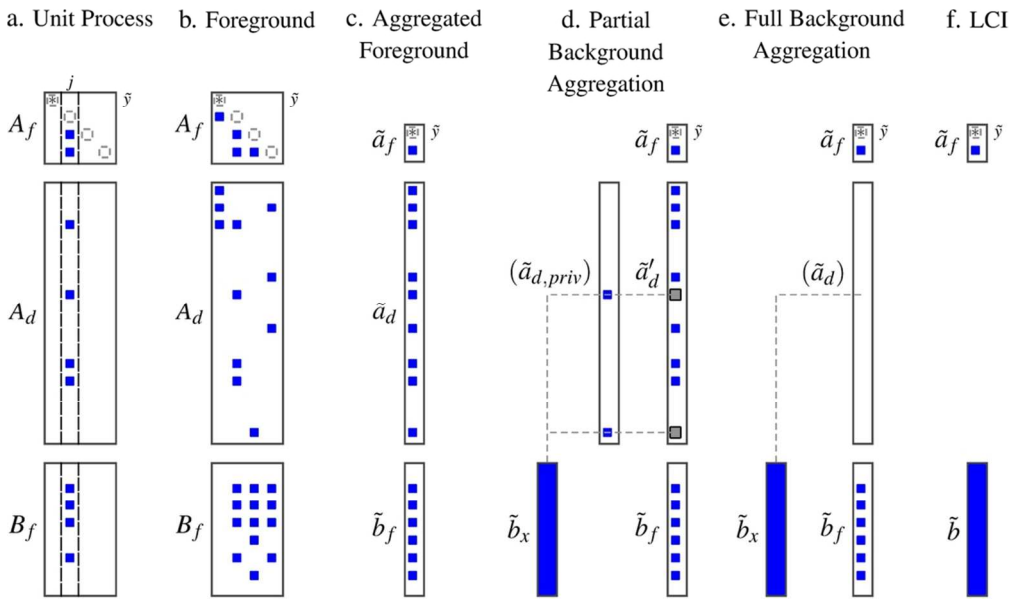
Traditional matrix structure of an LCA computation.

78x44mm (300 x 300 DPI)



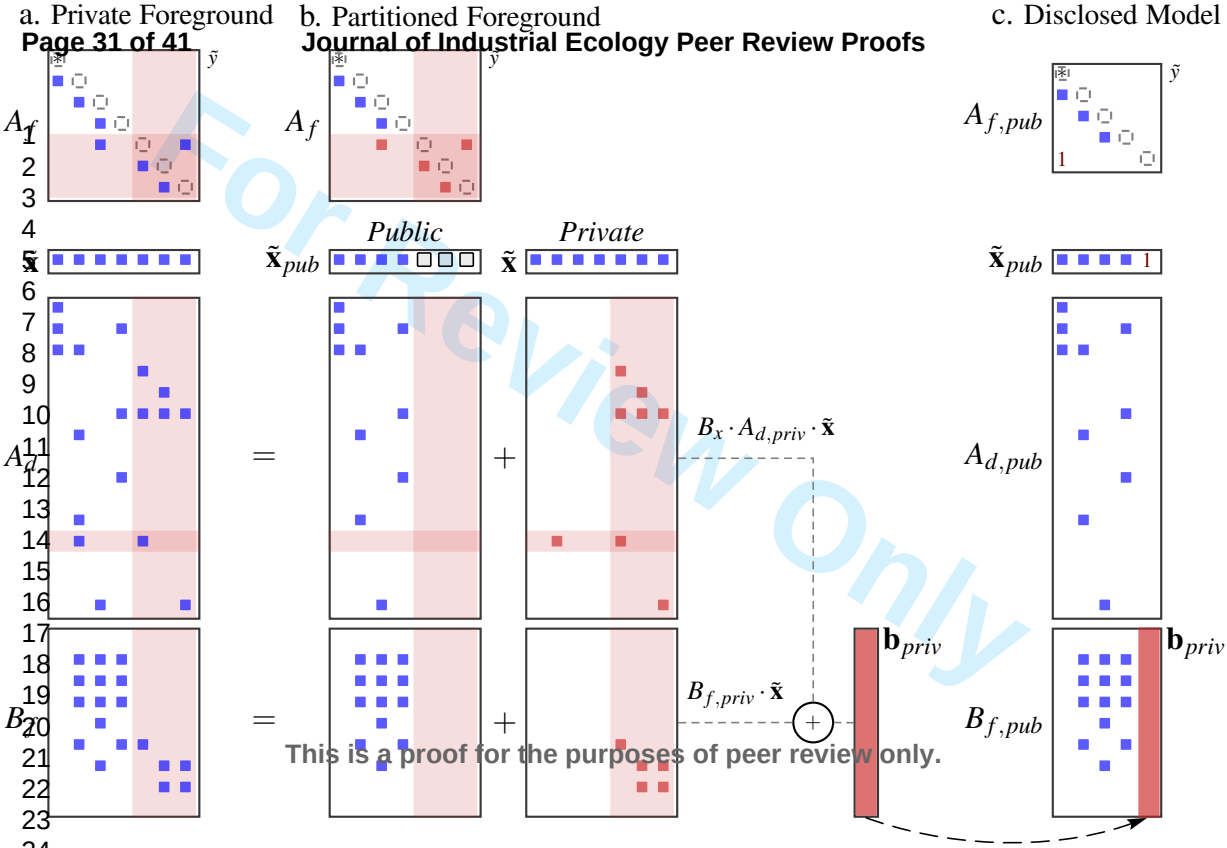
Matrix structure of an LCA foreground study. The background technology matrix has been incorporated into a background LCI database Bx, and replaced in the figure with a placeholder identity matrix I. The foreground and aggregated foreground elements are highlighted.

102x67mm (300 x 300 DPI)



Matrix structures for different forms of foreground aggregation.

95x56mm (300 x 300 DPI)



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

2017-11-16

Supporting Information

1 Mathematical Relations

1.1 Final Demand Augmentation

This is a proof of Equation 2 in the manuscript.

Given a life cycle inventory database (LCIDB) containing input-output matrix A and emission matrix B, and a final demand vector $\mathbf{y}$, construct an augmented inventory database $\tilde{A} = \left[\begin{smallmatrix} 0 & \mathbf{0} \\ \mathbf{y} & A \end{smallmatrix}\right]$, $\tilde{B} = [\mathbf{0}, B]$, and canonical functional unit $\tilde{\mathbf{y}} = [1, 0, 0, \ldots, 0]^T$. Show that

$$B \cdot (I - A)^{-1} \cdot \mathbf{y} = \tilde{B} \cdot \left(I - \tilde{A}\right)^{-1} \cdot \tilde{\mathbf{y}} \tag{1}$$

We begin by constructing $(I - \tilde{A})^{-1}$. Because the matrix is block triangular, we only need to determine the value of $\mathbf{q}$ in the matrix below:

$$I = (I - \tilde{A})^{-1} \cdot (I - \tilde{A}) \tag{2}$$

$$\begin{bmatrix} 1 & \mathbf{0} \\ \mathbf{0} & I \end{bmatrix} = \begin{bmatrix} 1 & \mathbf{0} \\ \mathbf{q} & (I - A)^{-1} \end{bmatrix} \cdot \begin{bmatrix} 1 & \mathbf{0} \\ -\mathbf{y} & (I - A) \end{bmatrix} \tag{3}$$

It can be seen that $\mathbf{q} - (I - A)^{-1} \cdot \mathbf{y} = 0$ is a necessary condition to satisfy the equality, and thus $\mathbf{q} = (I - A)^{-1} \cdot \mathbf{y}$.

We know that $\tilde{\mathbf{y}}$ selects the first column of its argument, so:

$$(I - \tilde{A})^{-1} \cdot \tilde{\mathbf{y}} = \begin{bmatrix} 1 \\ \mathbf{q} \end{bmatrix} \tag{4}$$

And therefore

$$\begin{aligned} \tilde{B}\cdot(I-\tilde{A})^{-1}\cdot\tilde{\mathbf{y}} &= \begin{bmatrix} \mathbf{0} & B \end{bmatrix}\cdot\begin{bmatrix} 1 \\ \mathbf{q} \end{bmatrix} && (5)\\ &= B\cdot\mathbf{q} && (6)\\ &= B\cdot(I-A)^{-1}\cdot\mathbf{y} && (7) \end{aligned}$$

thus completing the proof.

1.2 "Flattening" the background database

Given a canonical LCA foreground study, show that the following formulations are equivalent:

$$\tilde{A} = \begin{bmatrix} A_f & 0 \\ A_d & A \end{bmatrix}; \quad \tilde{B} = \begin{bmatrix} B_f & B \end{bmatrix} \tag{8}$$

$$\tilde{A}_{flat} = \begin{bmatrix} A_f & 0 \\ A_d & 0 \end{bmatrix}; \quad \tilde{B}_{flat} = \begin{bmatrix} B_f & B_x \end{bmatrix} \tag{9}$$

where $B_x = B\cdot(I-A)^{-1}$.

It is sufficient to show that $\tilde{B}\cdot(I-\tilde{A})^{-1} = \tilde{B}_{flat}\cdot(I-\tilde{A}_{flat})^{-1}$.

Using the shorthand $Z = I-A$, we define the following identity:

$$I* = \begin{bmatrix} I & 0 \\ 0 & Z \end{bmatrix}^{-1}\cdot\begin{bmatrix} I & 0 \\ 0 & Z \end{bmatrix} \tag{10}$$

We introduce the identity in between the two terms of the LCIDB:

$$\tilde{B}\cdot(I-\tilde{A})^{-1} = \begin{bmatrix} B_f & B \end{bmatrix}\cdot I*\cdot\begin{bmatrix} I-A_f & 0 \\ -A_d & I-A \end{bmatrix}^{-1} \tag{11}$$

$$= \begin{bmatrix} B_f & B \end{bmatrix}\cdot\begin{bmatrix} I & 0 \\ 0 & Z \end{bmatrix}^{-1}\cdot\begin{bmatrix} I & 0 \\ 0 & Z \end{bmatrix}\cdot\begin{bmatrix} I-A_f & 0 \\ -A_d & Z \end{bmatrix}^{-1} \tag{12}$$

Using the associative property of matrix multiplication, it can be shown that $M\cdot N^{-1} = (N\cdot M^{-1})^{-1}$ for any invertible M and N of the same rank, so:

$$\begin{aligned} \tilde{B}\cdot(I-\tilde{A})^{-1} &= \begin{bmatrix} B_f & B \end{bmatrix}\cdot\begin{bmatrix} I & 0 \\ 0 & Z \end{bmatrix}^{-1}\cdot\left(\begin{bmatrix} I-A_f & 0 \\ -A_d & Z \end{bmatrix}\cdot\begin{bmatrix} I & 0 \\ 0 & Z \end{bmatrix}^{-1}\right)^{-1} && (13)\\ &= \begin{bmatrix} B_f & B \end{bmatrix}\cdot\begin{bmatrix} I & 0 \\ 0 & Z^{-1} \end{bmatrix}\cdot\left(\begin{bmatrix} I-A_f & 0 \\ -A_d & Z \end{bmatrix}\cdot\begin{bmatrix} I & 0 \\ 0 & Z^{-1} \end{bmatrix}\right)^{-1} && (14)\\ &= \begin{bmatrix} B_f & B\cdot Z^{-1} \end{bmatrix}\cdot\left(\begin{bmatrix} I-A_f & 0 \\ -A_d & I \end{bmatrix}\right)^{-1} && (15) \end{aligned}$$

But these terms are equivalent to the flattened LCIDB:

$$\tilde{B}\cdot(I-\tilde{A})^{-1} = \left[\begin{array}{cc} B_f & B\cdot(I-A)^{-1}\end{array}\right]\cdot\left(\left[\begin{array}{cc} I-A_f & 0 \\ -A_d & I\end{array}\right]\right)^{-1} \quad (16)$$

$$= \left[\begin{array}{cc} B_f & B_x\end{array}\right]\cdot\left(I-\left[\begin{array}{cc} A_f & 0 \\ A_d & 0\end{array}\right]\right)^{-1} \quad (17)$$

$$= \tilde{B}_{flat}\cdot(I-\tilde{A}_{flat})^{-1} \quad (18)$$

thus completing the proof.

1.3 The Study Foreground Equation

Equation 5 in the manuscript is derived from Equation 4.

We begin by computing $(I-\tilde{A}_{flat})^{-1}$, again taking advantage of the matrix's block triangularity. Find Q such that:

$$\left[\begin{array}{cc} (I-A_f) & 0 \\ -A_d & I\end{array}\right]\cdot\left[\begin{array}{cc} (I-A_f)^{-1} & 0 \\ Q & I\end{array}\right] = \left[\begin{array}{cc} I & 0 \\ 0 & I\end{array}\right] \quad (19)$$

It can be seen that $Q = A_d\cdot(I-A_f)^{-1}$ is a necessary condition to satisfy the equality, and thus:

$$(I-\tilde{A})^{-1} = \left[\begin{array}{cc} (I-A_f)^{-1} & 0 \\ A_d\cdot(I-A_f)^{-1} & I\end{array}\right] \quad (20)$$

.

Substituting this into the LCA system equation:

$$s = \mathbf{e}^T\cdot\tilde{B}_{flat}\cdot(I-\tilde{A}_{flat})^{-1}\tilde{\mathbf{y}} \quad (21)$$

$$= \mathbf{e}^T\cdot\left[\begin{array}{cc} B_f & B_x\end{array}\right]\cdot\left[\begin{array}{cc} (I-A_f)^{-1} & 0 \\ A_d\cdot(I-A_f)^{-1} & I\end{array}\right]\cdot\left[\begin{array}{c} \tilde{\mathbf{y}}_f \\ \mathbf{0}\end{array}\right] \quad (22)$$

where $\tilde{\mathbf{y}}_f$ is a canonical functional unit having the same dimension as A_f. Thus:

$$s = \mathbf{e}^T\cdot\left[\begin{array}{cc} B_f & B_x\end{array}\right]\cdot\left[\begin{array}{c} (I-A_f)^{-1}\cdot\tilde{\mathbf{y}}_f \\ A_d\cdot(I-A_f)^{-1}\cdot\tilde{\mathbf{y}}_f\end{array}\right] \quad (23)$$

$$= \mathbf{e}^T\cdot\left(B_f\cdot(I-A_f)^{-1}\cdot\tilde{\mathbf{y}}_f + B_x\cdot A_d\cdot(I-A_f)^{-1}\cdot\tilde{\mathbf{y}}_f\right) \quad (24)$$

$$= \mathbf{e}^T\cdot\left(B_f + B_x\cdot A_d\right)\cdot(I-A_f)^{-1}\cdot\tilde{\mathbf{y}}_f \quad (25)$$

$$= \mathbf{e}^T\cdot\left(B_f + B_x\cdot A_d\right)\cdot\tilde{\mathbf{x}} \quad (26)$$

thus completing the derivation.

1.4 Meeting the Disclosure Objectives

1.4.1 Transparency

The objective for transparency was identified as requiring computability, completeness, and reproducibility. These requirements can be met by a disclosure that clearly describes the identities of each row or column of A_f, A_d, and B_f containing non-zero entries, and the locations and values of those entries. In principle, a reader with this information would have the capability to construct the augmented LCIDB and perform the computation in Eq. 5. In actuality, while it is easy to reproduce a set of sparse matrices, there is considerable potential for ambiguity in stating the identities of the rows and columns of those matrices.

The foreground nodes, which make up the columns of A_f, can be chosen freely by the study author according to the objectives of the study. Large studies may contain hundreds of foreground nodes, and the nodes can correspond to physical activities, logical operations, unit conversions, accumulation or distribution points, or any other aspect of model construction that can be reflected in a process-flow diagram. The only requirement on their disclosure is that the identity of each node's reference flow, including its unit of measure, is clearly stated. In formalizing the study as a normalized direct requirements matrix, the activity level of each foreground node necessarily equals the magnitude of the total reference flow emanating from the node.

For each row in A_d containing a non-zero entry, the author must unambiguously identify the exact dataset used, including the version of the database, as well as the exact process and reference flow selected; the dimension (reference quantity or unit) of each reference flow must be specified; and the sign of the numeric entry in A_d must be consistent with the implementation of the process in the background database. Similarly, for each row in B_f containing a non-zero entry, the author must unambiguously identify the substance being exchanged with the environment, the compartment or context into which it is being exchanged, and the reference quantity or unit associated with the flow. Sign consistency must also be assured. If LCIA indicator results are included, the author must also unambiguously identify the method computed (identity of **e**).

1.4.2 Authority and Primacy

It may be observed that many LCA studies make use of data sources that have been previously published but that have not been included in any LCI reference or background database. As long as these data sources can be integrated seamlessly into an LCA computation, it is not necessary to reproduce them in a disclosure. However, in most applications, previously published data must be re-implemented by the author in the LCA software context, and often this re-implementation requires a re-interpretation of the data source as a unit process inventory, where one exchange is recognized as a reference flow and other exchange values are reported in proportionality to the reference flow.

Because of these conditions, in the vast majority of cases the LCA study disclosure must include the author's reimplementations in order to achieve both the aims for authority and primacy. In so

doing, the disclosure enables a critical reviewer to evaluate whether the author's implementations are generally correct and complete.

In the future, the inclusion of external data automatically can be accomplished in the same way that reference LCI data could conceivably be included automatically: by the data providers making their information available using a stable semantic reference to a specialized Web-based application programming interface (API). This would have the benefits of enabling downstream users to access the information without having to re-implement it, thus reducing the size of the disclosure necessary to describe the PSM and simplifying the task of the modeler.

2 Foreground Configurations

Figure S1 illustrates the structure of some basic foreground configurations. Most product system models would include many of these elements mixed together. The first (a) is a sequential model, in which each node requires one foreground input and generates one output. This model is equivalent to a "gate to gate" model. Here the weights k_i indicate the amount of the preceding reference flow that is required by the subsequent node. Figure S1(b) shows an additive model, in which the outputs of several foreground nodes are added together, equivalent to a "mixer" or a horizontal average. In this arrangement the weights should add up to a unit output of the reference node. Finally, Figure S1(c) shows a foreground model with a cyclic dependency, where some of the reference output is consumed by another foreground node.

A typical PSM may contain multiple modules or fragments that are interconnected. An example of foreground composed of several fragments is illustrated in Figure S1(d). Here, the nodes labeled 1–5 represent one fragment, which generates the foreground's canonical reference flow $\tilde{y}$. This fragment requires two interior flows from separate fragments (y_0 and y_1), and has two unconnected flows (4 and 5). The reference y_0 is supplied by a second fragment, made up of nodes 6-8. The reference y_1 is supplied by another fragment made of only one node (9). The reference flow y_1 is consumed in two different places by the other fragments.

As noted, a reviewer with access to items *d-i* and *d-iv* in the disclosure would be able to construct A_f and automatically create a process-flow diagram, annotated with information about each node.

3 Examples

LCI databases often contain product systems that can be modeled as foreground studies because they describe products that are not required by the background. In this section, two product systems selected from LCI databases are used to illustrate the concept of structured publication. The systems selected because are complex enough to illustrate the premise but simple enough to review easily.

Each system is illustrated as a table that shows the foreground model, cutoff flows, background dependencies and foreground emissions included in the system. Aggregation results $\tilde{\mathbf{x}}$, $\tilde{\mathbf{a}}_d$, and $\tilde{\mathbf{b}}_f$

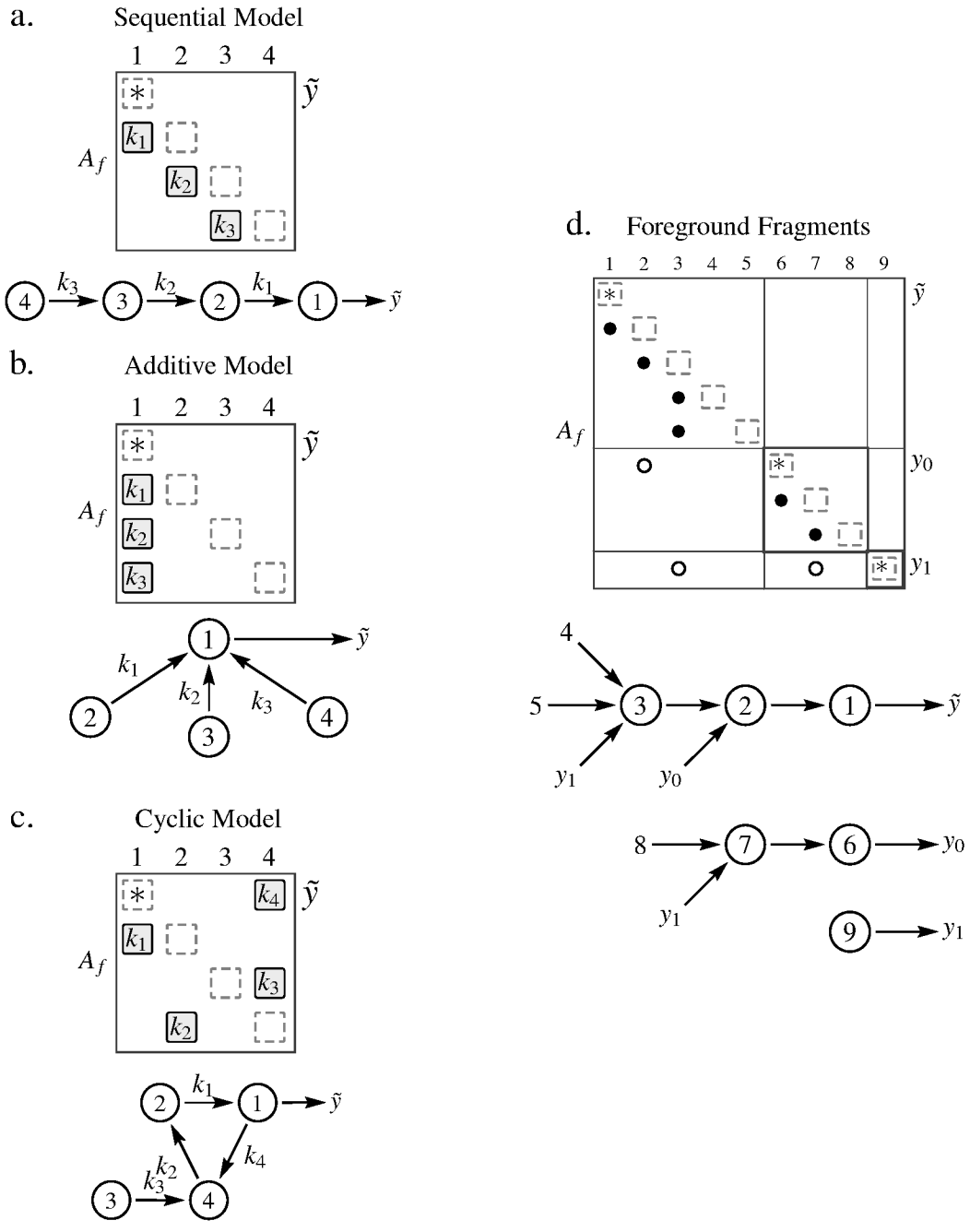


Figure S1: Equivalent matrix representations and graphs for different foregrounds.

are also reported. The table omits numeric data from most of the table for clarity. Instead, a black square indicates the presence of a nonzero value.

3.1 US LCI: Secondary Aluminum

The US LCI database contains a small background system of 39 processes, as well as 395 foreground product systems that range in size from one to 83 foreground flows.

The example product system reports the production of secondary aluminum from automotive scrap (Figure S2). The aluminum production process has direct requirements for two other foreground systems, including transportation services and quicklime production. The quicklime production itself depends on limestone extraction, which is also part of the foreground. The aluminum production requires two background systems, electricity and natural gas combustion, and reports 10 cutoff flows. The main input to the process, "Aluminum scrap, automotive," appears as a cutoff flow. Overall, the four foreground processes require the background for only combustion (five fuels), transport (three modes) and grid electricity. The same natural gas combustion and electricity models were used in all three non-transport foreground nodes.

The system also reports several direct emissions, mainly from the aluminum process. Only one resource consumption (input flow) is reported – the extraction of limestone. The direct emission modeling is limited in scope and includes a number of "unspecified" flows that may not be well characterized in many impact methods.

3.2 Ecoinvent: Organic Potatoes

Ecoinvent version 3 is provided in three different system models that reflect different linking strategies. The example is drawn from the "Allocation at the point of substitution" or APOS model, which includes 11,420 processes that produce 12,966 product flows. Of these, 10,282 are background flows and the rest are foreground flows.

The example system reports production of organic potatoes supplied to the global market (Table S3). The foreground includes nine nodes, of which six (nodes 3 through 8) form a cyclic dependency associated with the production of potato seeds. The Ecoinvent database's use of "markets" as mixer processes is evident in the table: node 0 is a mixer process that combines Swiss ("CH") production (node 1) weighted at 2.4% with rest-of-World ("RoW") production (node 2) weighted at 97.6%. That same market split (2.4 / 97.6) can also be seen in nodes 7 and 8.

Nodes 4 and 5, which make seeds ready "for setting," are mixed by node 8. Although nodes 4 and 5 are geographically distinguished (CH vs RoW), both can be seen to consume potato seed from the global market.

Looking at the dependency and emission lists, the different "signatures" of different kinds of processes can be seen: nodes 1, 2, 3, and 6 are clearly agricultural processes that require irrigation, tillage, manure and so on. Nodes 0 and 7 are visible as market processes, their only requirements being transport processes. Nodes 4 and 5 each consume electricity and require use of a

Aluminum, secondary, ingot, from automotive scrap, at plant [RNA]

(node) Foreground flows – A_f	0	1	2	3	
(0) Aluminum, secondary, ingot, from automotive scrap, at plant [RNA] (kg)	○				
(1) Aluminum recovery, transport, to plant [RNA] (kg)	1.03	○			
(2) Quicklime, at plant [RNA] (kg)	2.35e-05		○		
(3) Limestone, at mine [RNA] (kg)			1.87	○	
Foreground Node Weights $\tilde{x}$	1	1.03	2.35e-05	4.39e-05	
Input: CUTOFF Disposal, solid waste, unspecified, to sanitary landfill [CUTOFF Flows] (kg)	■		■		0.0842
Input: CUTOFF Filter media, at plant [CUTOFF Flows] (kg)	■				4.79e-05
Input: CUTOFF Lube oil, at plant [CUTOFF Flows] (kg)	■				8.75e-07
Input: CUTOFF Treatment gases, unspecified, at plant [CUTOFF Flows] (kg)	■				0.003
Input: CUTOFF Alloying additives, at plant [CUTOFF Flows] (kg)	■				0.021
Input: CUTOFF Chemicals, unspecified, used for wastewater treatment [CUTOFF Flows] (kg)	■				0.008
Input: CUTOFF Grain refiners, at plant [CUTOFF Flows] (kg)	■				0.0003
Input: CUTOFF Treatment salts, unspecified, at plant [CUTOFF Flows] (kg)	■				0.009
Input: CUTOFF Packaging, unspecified, at plant [CUTOFF Flows] (kg)	■				1.67e-05
Input: CUTOFF Aluminum scrap, automotive [CUTOFF Flows] (kg)	■				1.03

Background Dependencies – A_d	0	1	2	3	$\tilde{a_d}$
Transport, barge, average fuel mix [RNA] (t*km)			■		5.66e-07
Liquefied petroleum gas, combusted in industrial boiler [RNA] (l)			■		7.57e-10
Transport, combination truck, diesel powered [RNA] (t*km)		■	■		0.374
Diesel, combusted in industrial boiler [RNA] (l)			■	■	4.79e-08
Natural gas, combusted in industrial boiler [RNA] (m3)	■		■	■	0.223
Transport, train, diesel powered [RNA] (t*km)		■	■		0.0415
Electricity, at grid, US, 2000 [RNA] (kWh)	■		■	■	0.668
Bituminous coal, combusted in industrial boiler [RNA] (kg)			■	■	4.04e-06
Gasoline, combusted in equipment [RNA] (l)				■	2.25e-09

Foreground Emissions – B_f	0	1	2	3	$\tilde{b_f}$
Output: Lead [air, unspecified] (kg)	■				2.16e-07
Output: Suspended solids, unspecified [water, unspecified] (kg)	■				2.56e-05
Output: Particulates, unspecified [air, unspecified] (kg)	■		■	■	1.78e-07
Output: NMVOC, non-methane volatile organic compounds [air, unspecified] (kg)	■				4.7e-05
Output: BOD5, Biological Oxygen Demand [water, unspecified] (kg)	■				1.35e-09
Output: COD, Chemical Oxygen Demand [water, unspecified] (kg)	■				8.3e-07
Output: Sulfur dioxide [air, unspecified] (kg)			■		3.52e-09
Output: Dissolved solids [water, unspecified] (kg)	■				2.38e-08
Output: carbon dioxide [air, unspecified] (kg)			■		1.8e-05
Input: Limestone [resource, ground-] (kg)				■	4.39e-05
Output: Heavy metals, unspecified [water, unspecified] (kg)	■				2.25e-05
Output: Organic substances, unspecified [water, unspecified] (kg)	■				1.29e-06
Output: Acids, unspecified [air, unspecified] (kg)	■				4.31e-05

Figure S2: A structured product system model for secondary aluminum, drawn from US LCI. Exchange values are replaced with black squares for clarity.

market for potato, organic [GLO]

(node) Foreground flows – A_f	0	1	2	3	4	5	6	7	8	
(0) potato, organic [GLO] (kg)	◌									
(1) potato, organic [CH] (kg)	0.024	◌								
(2) potato, organic [RoW] (kg)	0.976		◌							
(3) potato seed, organic, at farm [CH] (kg)				◌	0.024					
(4) potato seed, organic, at farm [GLO] (kg)					◌	1	1			
(5) potato seed, organic, for setting [RoW] (kg)						◌			0.976	
(6) potato seed, organic, for setting [CH] (kg)							◌		0.024	
(7) potato seed, organic, at farm [RoW] (kg)					0.976			◌		
(8) potato seed, organic, for setting [GLO] (kg)		0.11	0.11	0.16				0.16	◌	
Foreground Node Weights $\tilde{x}$	1	0.024	0.976	0.00315	0.131	0.128	0.00315	0.128	0.131	

Background Dependencies – A_d	0	1	2	3	4	5	6	7	8	$\tilde{a_d}$
building, multi-storey [GLO] (m3)						■	■			5.24e-06
electricity, low voltage [GLO] (kWh)						■				0.00933
potato haulm cutting [GLO] (ha)		■	■	■				■		5.04e-05
tillage, hoeing and earthing-up, potatoes [GLO] (ha)		■	■	■				■		0.000101
transport, tractor and trailer, agricultural [GLO] (metric ton*km)		■	■	■				■		0.00113
transport, freight, light commercial vehicle [GLO] (metric ton*km)	■				■					0.0258
tillage, ploughing [GLO] (ha)		■	■	■				■		5.04e-05
green manure, organic, until March [GLO] (ha)		■	■	■				■		5.04e-05
transport, freight, lorry, unspecified [GLO] (metric ton*km)	■				■					0.404
transport, freight train [GLO] (metric ton*km)	■				■					0.138
transport, freight, inland waterways, barge [GLO] (metric ton*km)	■				■					0.0857
copper oxide [GLO] (kg)		■	■	■				■		0.000107
tillage, harrowing, by spring tine harrow [GLO] (ha)		■	■	■				■		5.04e-05
potato planting [GLO] (ha)		■	■	■				■		5.04e-05
solid manure loading and spreading, by hydraulic loader and spreader [GLO] (kg)		■	■	■				■		0.717
potato grading [GLO] (kg)		■	■	■				■		1.13
tillage, harrowing, by rotary harrow [GLO] (ha)		■	■	■				■		5.04e-05
harvesting, by complete harvester, potatoes [GLO] (ha)		■	■	■				■		5.04e-05
transport, freight, sea, transoceanic ship [GLO] (metric ton*km)	■				■					0.525
liquid manure spreading, by vacuum tanker [GLO] (m3)		■	■	■				■		0.000565
electricity, low voltage [CH] (kWh)							■			0.00023
irrigation [GLO] (m3)			■					■		0.0173
tillage, currying, by weeder [GLO] (ha)		■	■	■				■		0.000101
application of plant protection product, by field sprayer [GLO] (ha)		■	■	■				■		0.000222
irrigation [CH] (m3)		■		■						0.000426

Foreground Emissions – B_f	0	1	2	3	4	5	6	7	8	$\tilde{b_f}$
Input: Occupation, construction site [natural resource, land] (m2*year)						■	■			2.1e-06
Input: Transformation, from unspecified [natural resource, land] (m2)						■	■			1.05e-06
Input: Transformation, to industrial area [natural resource, land] (m2)						■	■			1.05e-06
Input: Energy, gross calorific value, in biomass [natural resource, biotic] (MJ)		■	■	■				■		3.87
Output: Phosphate [water, ground-] (kg)		■	■	■				■		6.25e-06
Output: Cadmium, ion [water, ground-] (kg)		■	■	■				■		1.14e-09
Output: Zinc, ion [water, ground-] (kg)		■	■	■				■		1.35e-06
Output: Nitrate [water, ground-] (kg)		■	■	■				■		0.00696
… (31 rows omitted)										

Figure S3: A structured product system model for organic potato production, drawn from Ecoinvent v3.2 (APOS). Exchange values are replaced with black squares for clarity.

"multi-storey building." Taken together, the CH-locality processes appear to use CH irrigation and electricity supply but are otherwise similar to their RoW counterparts. The product model includes no cutoff flows.