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

$$\begin{aligned}
\tilde{B} \cdot (I - \tilde{A})^{-1} \cdot \tilde{\mathbf{y}} &= \left[\begin{array}{cc} \mathbf{0} & B \end{array} \right] \cdot \left[\begin{array}{c} 1 \\ \mathbf{q} \end{array} \right] && (5) \\
&= B \cdot \mathbf{q} && (6) \\
&= B \cdot (I - A)^{-1} \cdot \mathbf{y} && (7)
\end{aligned}$$

thus completing the proof.

1.2 "Flattening" the background database

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

$$\tilde{A} = \left[\begin{array}{cc} A_f & 0 \\ A_d & A \end{array} \right]; \quad \tilde{B} = \left[\begin{array}{cc} B_f & B \end{array} \right] \tag{8}$$

$$\tilde{A}_{flat} = \left[\begin{array}{cc} A_f & 0 \\ A_d & 0 \end{array} \right]; \quad \tilde{B}_{flat} = \left[\begin{array}{cc} B_f & B_x \end{array} \right] \tag{9}$$

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:

$$I* = \left[\begin{array}{cc} I & 0 \\ 0 & Z \end{array} \right]^{-1} \cdot \left[\begin{array}{cc} I & 0 \\ 0 & Z \end{array} \right] \tag{10}$$

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

$$\tilde{B} \cdot (I - \tilde{A})^{-1} = \left[\begin{array}{cc} B_f & B \end{array} \right] \cdot I* \cdot \left[\begin{array}{cc} I - A_f & 0 \\ -A_d & I - A \end{array} \right]^{-1} \tag{11}$$

$$= \left[\begin{array}{cc} B_f & B \end{array} \right] \cdot \left[\begin{array}{cc} I & 0 \\ 0 & Z \end{array} \right]^{-1} \cdot \left[\begin{array}{cc} I & 0 \\ 0 & Z \end{array} \right] \cdot \left[\begin{array}{cc} I - A_f & 0 \\ -A_d & Z \end{array} \right]^{-1} \tag{12}$$

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:

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

But these terms are equivalent to the flattened LCIDB:

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

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

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

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:

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

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

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

.

Substituting this into the LCA system equation:

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

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

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

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

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

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

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

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 $\mathbf{e}$).

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 $\tilde{\mathbf{b}}_f$

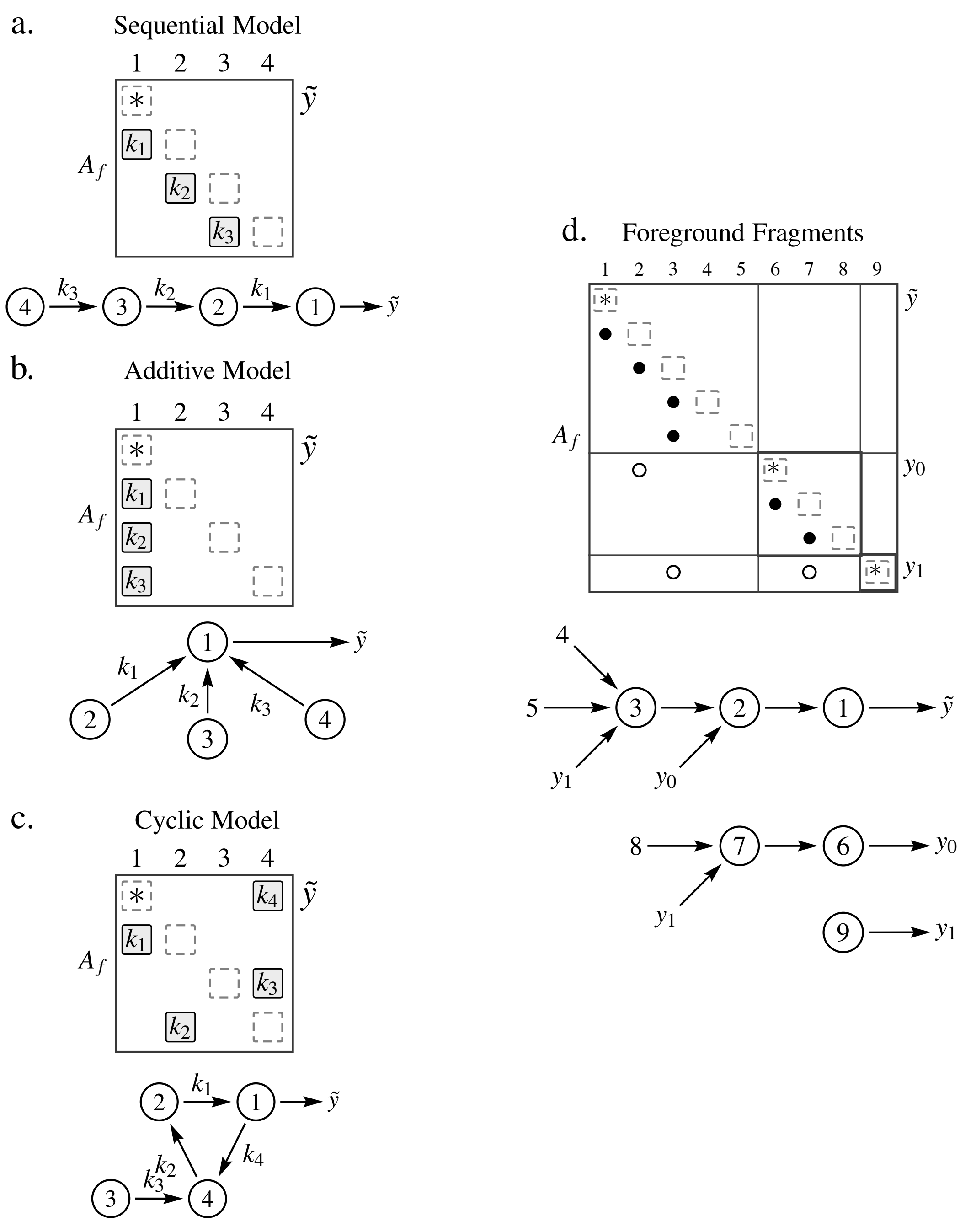

Figure S1: Equivalent matrix representations and graphs for different foregrounds.

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.