Economic Systems Research

ISSN: 0953-5314 (Print) 1469-5758 (Online) Journal homepage: https://www.tandfonline.com/loi/cesr20

A network approach for assembling and linking input–output models

João Rodrigues, Alexandra Marques, Richard Wood & Arnold Tukker

To cite this article: João Rodrigues, Alexandra Marques, Richard Wood & Arnold Tukker (2016) A network approach for assembling and linking input–output models, Economic Systems Research, 28:4, 518-538, DOI: 10.1080/09535314.2016.1238817

To link to this article: https://doi.org/10.1080/09535314.2016.1238817

ECONOMIC SYSTEMS RESEARCH, 2016
VOL. 28, NO. 4, 518–538
http://dx.doi.org/10.1080/09535314.2016.1238817

Routledge
Taylor & Francis Group

OPEN ACCESS

A network approach for assembling and linking input–output models

João Rodrigues[a], Alexandra Marques[a], Richard Wood[b] and Arnold Tukker[a]

[a]Institute of Environmental Sciences (CML), Leiden University, Leiden, The Netherlands; [b]Programme for Industrial Ecology, Energy and Process Technology Department, NTNU, Trondheim, Norway

ABSTRACT

Input–output (IO) models, describing trade between different sectors and regions, are widely used to study the environmental repercussions of human activities. A frequent challenge in assembling an IO model or linking several such models is the absence of flow data with the same level of detail for all components. Such problems can be addressed using proportional allocation, which is a form of algebraic transformations. In this paper, we propose a novel approach whereby the IO system is viewed as a network, the topology of which is transformed with the addition of virtual nodes so that available empirical flow data can be mapped directly to existing links, with no additional estimation required, and no impact on results. As IO systems become increasingly disaggregated, and coupled to adjacent databases and models, the adaptability of IO frameworks becomes increasingly important. We show that topological transformations also offer large advantages in terms of transparency, modularity and increasingly importantly for global IO models, efficiency. We illustrate the results in the context of trade linking, multi-scale integration and other applications.

ARTICLE HISTORY
Received 3 December 2015
In final form 16 September 2016

KEYWORDS
Input–output (IO) model; assembling and linking; network approach; topological transformation; virtual nodes

1. Introduction

Input–output (IO) models have been increasingly used to study the environmental, social and economics impacts of human activities in an ever more interconnected world (Miller and Blair, 2009; Liu et al., 2015). This type of analysis requires IO tables that describe trade between all sectors in the economy of interest in a given year, and if the economy describes several regions they are called multi-regional IO (MRIO) tables. In a globalized world where production and consumption processes are spatially disconnected, the integration of different data streams is critical to allow tracking environmental pressures through global supply chains (Wood et al., 2015).

In practice it often happens that no data is available at the level of detail we desire but instead we only possess aggregate information (Miyazawa, 1968; Batten and Martellato, 1985; Joshi, 1999; UN, 1999; Su et al., 2010; Lenzen, 2011; Bachmann et al., 2014;

CONTACT João Rodrigues j.rodrigues@cml.leidenuniv.nl

Supplemental data for this article can be accessed here. http://dx.doi.org/10.1080/09535314.2016.1238817

Streicher and Stehrer, 2014; Edens et al., 2015). Although different from an economic or environmental point of view, all these problems are essentially identical if seen through the lens of inference: we would like to know something but because of lack of data or time we only know sums thereof. The most commonly applied solution is to split the known aggregates using some form of proxy data to perform proportional allocation.

To the best of our knowledge the literature reviewed above takes a strictly heuristic approach, not providing a reason as to why proportional allocation should be used. However, there is a multitude of other possible solutions to the problem of missing disaggregate data subject to aggregate information when assembling or combining IO models (Lindner et al., 2012). In this paper we take a step back to look at this problem from first principles. We base ourselves on the work of Jaynes on inductive inference (2003) and invariance transformations (1973), which we operationalize in a novel approach for assembling and combining IO models. This new approach, which we christened topological transformation, involves thinking about the IO system as a network, and re-form the problem of estimating missing arcs by that of defining virtual nodes. This subject is explored in Section 2. Our results benefit researchers that encounter a situation of missing disaggregate data during the assembling or linking of IO models.

A natural follow-up question is how the results obtained from topological transformations differ from those obtained from the standard proportionality assumption. This is explored in Sections 3 and 4. For concreteness, Section 3 offers an in-depth application of the network approach to the particular case of trade linking, presenting a theoretical comparison of both approaches and a simple numerical illustration. Section 4 then shows how this approach can help in other circumstances and, in particular, in the linking of multi-scale IO models, a subject which we believe to be of great and increasing importance.

A different question is the implication of the choice of approach (topological transformation vs. proportionality assumption) in terms of the practical workflow of IO analysis. This is addressed in Section 5, which discusses the performance of both approaches under the headings of transparency, modularity and data storage and computation. We now briefly explain why we believe these topics to be important.

It is common for empirical data to be inconsistent, a problem addressed by applying a data balancing procedure (Lindner et al., 2012; Wood et al., 2014) such that detailed technological data are reconciled to aggregate industry data, while respecting material balances. If this procedure is applied to algebraically transformed data, then it becomes impossible to trace back the inconsistency to individual data sources, no matter how good the documentation is. Using a topological transformation, instead, the data balancing procedure adjusts directly the source information, so it is always possible to step back and to find out exactly which input datum was wrong and needs to be refined. We believe the coming years will witness an expansion in the collaborative compilation of IO models (Lenzen et al., 2012b). And we also believe that a clear link between inputs and outputs is a necessity to ensure data quality in such a collaborative environment.

If MRIO work is going to keep advancing and maximizing opportunities to couple to other models/databases in order to provide necessary resolution in following source impacts to local consumers we see a strong need for optimizing the process of MRIO construction. This paper makes a contribution in that direction by showing that it is possible to combine different datasets in a unified model without the need to perform algebraic transformations on that data. This preservation of data integrity is crucial as we move to increasingly decentralized procedures of data assembly, in which different users might want to combine and reshuffle overlapping datasets (Liu et al., 2015).

Currently, developers and users of some MRIO databases are beginning to suffer because of the sheer seize of these databases (millions of variables covering industries in every region of the globe). For example, EXIOBASE (Tukker et al., 2013) currently covers only 49 regions because of the computational limitations of having both high sectoral detail and country coverage (Tukker et al., 2011). The producers of Eora (Lenzen et al., 2012a) had to develop their own custom balancing algorithms (Lenzen et al., 2009) instead of using commercially available solvers in order to streamline the memory requirements in balancing their global MRIO table as much as possible. Efforts to link agricultural databases to MRIO databases are producing databases of some 2.5 billion variables under conventional approaches (Bruckner, 2015). It turns out that a lot of these millions and billions of data points do not correspond to any true source data but are instead artifacts obtained by applying proportionality assumption to a far smaller number of empirical data points. Topological transformations drastically reduce the computational needs in MRIO modelling and therefore open up a whole new research frontier.

The paper concludes with Section 6, which offers some final remarks.

2. The network approach

2.1. Problem formulation

This paper addresses the problem of assembling an IO model or linking several IO models when some flow values of interest are missing, but aggregates thereof are available.

This situation is often addressed by performing imputation of the missing flow data (Lenzen, 2011; Rodrigues, 2016) and then balancing/reconciling the system (Lahr and de Mesnard, 2004; Rodrigues, 2014) until the aggregate constraints are satisfied. In some circumstances proxy data can be used to obtain a more informed imputation, but when no additional information is available practitioners can resort to proportional allocation of the aggregates among the several disaggregate items.

In this section we address this problem not by proposing a particular ad hoc method but by asking what (if any) general solution is suggested from first principles. The following sections will apply the general solution we found to this problem in different situations.

2.2. The IO network

In a symmetric IO system (Leontief, 1941; Miller and Blair, 2009) the economy is described by a set of industries or products, final demand and primary factors and the transactions between them (UN, 2010). We use $\mathbf{Z}$ to denote the matrix of inter-industry transactions, $\mathbf{Y}$ the matrix of final demand, $\mathbf{V}$ a matrix of primary inputs and $\mathbf{x}$ total output. Additionally, $\mathbf{r}$ is a vector of environmental interventions.

For ease of notation consider that the MRIO model is harmonized, with every of the n_R regions having n_F sectors each, each producing a homogeneous commodity, and n_Y final demand categories. We use lowercase italics to denote an element of the matrix and vector objects. The formulas explaining how environmental impacts are calculated are summarized in Supplementary Information A.1.

An IO system describes a network (Duchin and Levine, 2010; Kagawa et al., 2013; Heijungs, 2015), a mathematical object defined by a set of nodes or vertices and a set of arcs or edges connecting them. The set of arcs can be summarized in what is known as an IO table, but can also be called an adjacency matrix whose entry (i,j) represents the arc (or flow) from node i to node j. In an IO system the nodes are industries, final demand and primary input categories nested within countries, and the arcs are flows between them.

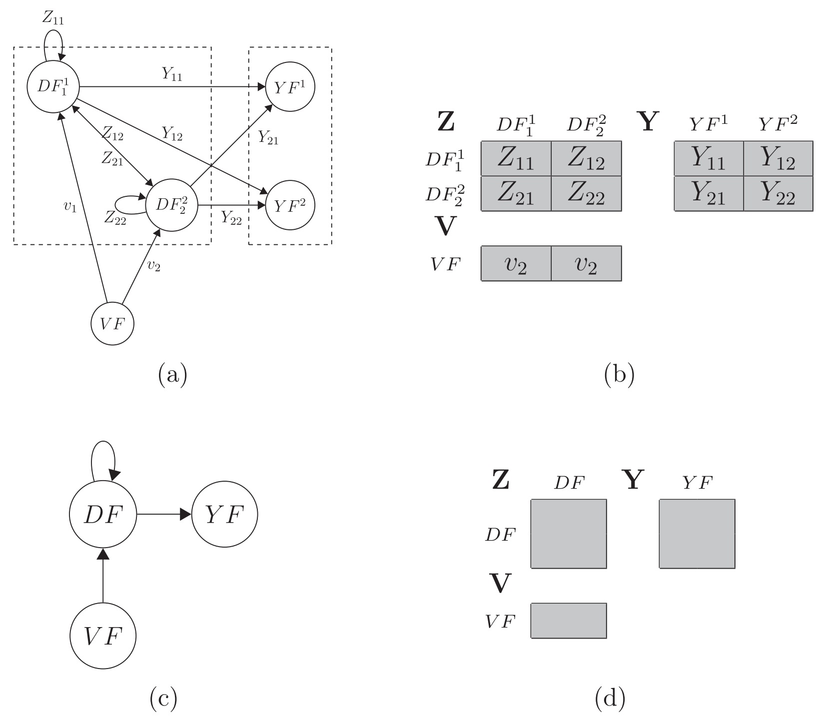
Let DF_i^a denote the domestic sector i of region a, and $DF_i^a \rightarrow DF_j^b$ denote the arc between DF_i^a and DF_j^b, whose flow $Z(a,i:b,j)$ is stored in row k and column m of $\mathbf{Z}$, where $k = i + (a-1)n_F$ and $m = j + (b-1)n_F$.

$\mathbf{Y}$ represents flows $DF_i^a \rightarrow YF_j^b$, where YF_j^b is the node representing the final demand category j in region b. $\mathbf{V}$ represents flows $VF_i \rightarrow DF_j^b$, where VF_i is the node representing the primary input category i (in this model primary input payments are strictly domestic). The corresponding arc flows are $Y(a,i:b,j)$ and $V(i:b,j)$.

An IO vector describes an object associated with a single node. For example, $x(a,i)$ is the total flow through DF_i^a.

Figure 1 illustrates the IO economic network using block diagrams and adjacency matrices in an economy with two countries and a single primary input, final output and industry

Figure 1. Representation of the IO network model. (a) Diagram of elements, (b) matrix version, (c) block diagram and (d) matrix version.



Arrow = arc, circle = node. DF = domestic firm, YF = final demand, VF = primary inputs. Z = internal flows, Y = final outputs, V = primary inputs.

categories. For simplicity these figure and the subsequent ones omit total output and environmental interventions. Figure 1(a) is a diagram with every node in the system (*DF* for domestic firm, *YF* for final demand and *VF* for primary input) and the arcs connecting them. Figure 1(b) is a matrix representation, identifying the rows and columns as nodes and entries as arcs. In the remainder of this paper we will deal with more complex topologies, and for clarity will not depict every node and arc element but rather blocks thereof, as illustrated in Figure 1(c,d), labelling only the nodes with the adjacency matrix indicating non-empty arc elements.

2.3. Invariance transformations

The assembly and linking of IO models subject to missing disaggregate data is a problem of inductive inference (Jaynes, 2003), i.e. a situation in which there is no unique answer but multiple options satisfy the desired criteria, which in this case are the different ways in which the aggregate constraint can be split over the various disaggregate items. Jaynes (1973) suggests that, in some circumstances, it is possible to reduce a problem of inductive inference to a problem of deductive inference, i.e. a situation in which there is a single solution, by means of invariance transformations.

An invariance transformation is an operation which leaves some property of the system identical to its initial state and thus uncovers a hitherto hidden symmetry in the original problem formulation. A geometrical example is the rotation of a circle, used for time keeping in analog clocks: when a hand turns 360° (the transformation) it returns to the initial position (invariance).

In the context of inference, by reducing the number of degrees of freedom in the original problem formulation, the identification of an invariance transformation may allow rephrasing an ill-posed (inductive) problem as a well-posed (deductive) one. Supplementary Information A.2 presents examples of invariance transformations.

To apply an invariance transformation to IO model assembly and linking it is necessary to recast the problem in the context of network theory. Instead of focusing our attention on determining the arcs that connect existing nodes (flows in a fixed IO classification), it is possible to conceptualize the empirical aggregate information itself as a set of adjacency matrices connecting existing nodes to virtual nodes not originally present in the problem formulation.

The method of topological transformation thus consists of, instead of estimating detailed (but unknown) arcs connecting existing nodes, adding virtual nodes in the network such that existing empirical data can be mapped to available arcs. An example of virtual nodes are importing agents, which collect imports from different countries and deliver outputs to domestic sectors, thus mediating between international trade in products and use of imported products by different domestic industries. More generally, nodes may also be real or virtual domestic industries – such as electricity distributors that provide different mixes of electricity.

The expansion of the original set of nodes and replacement of a direct and unknown arc between two real nodes by a sequence of arcs starting and ending in real nodes but passing by a virtual node is a topological transformation because the structure of interactions in the network (its topology) has changed. This changes IO model assembly from an ill- to a well-posed problem because all flows in the original formulation are now accounted for and there is no need to perform any additional imputation. This transformation constitutes an invariance because it does not require any information which was not already present in the original problem formulation.

3. Trade linking

3.1. Proportional allocation

We now address the trade linking problem, which deals with the by far most common situation of missing disaggregate data in IO analysis (IOA).

The staple of the so-called non-survey approaches[1] is the proportionality or trade share assumption whereby bilateral exports are allocated to importing industries and final demand categories according to the trade shares of the importing region, yielding a Chenery–Moses type of MRIO model (Moses, 1955; Batten and Martellato, 1985; OECD, 2006; Miller and Blair, 2009). Recent examples of the application of this assumption to the assembly of MRIOs are Peters et al. (2011), Lenzen et al. (2012a), Tukker et al. (2013) and Wood et al. (2015), where detailed bilateral trade data[2] are linked to national IO tables (which describe how much of a given commodity is imported, but not its country of origin).[3] Imputation can occur at different levels of detail, e.g. the WIOD database uses a non-survey method that relies on also including aggregated bilateral trade data by broad economic category (usually capital goods, consumption goods and intermediate inputs) (Dietzenbacher et al., 2013b).

An MRIO requires the specification of the trade that takes place between every two sectors from every two countries. Such information is not available as source data and must therefore be estimated from other sources (Oosterhaven et al., 2008). In the case of trade, this is often a table of internationally traded commodities (i.e. the exports of given commodity from a given region to all industries of another region), and a set of national tables indicating the imports of every sector by commodity type but not by source country (i.e. the imports a given commodity type imported from all countries to a given industry from a given country).

For concreteness, we consider the following information is available:

- Intermediate use of domestic inputs, $vdf(a, i, j)$, for every region a, product i and industry j;
- Final use of domestic inputs, $vdy(a, i, j)$, for every region a, product i and final demand category j;
- Intermediate use of imported inputs, $vif(a, i, j)$, for every region a, product i and industry j;

[1] If detailed information about the trade between different regions is available then we are in the realm of an inter-regional IO model (Isard, 1951). The construction of an IRIO requires a survey-based approach to collect large volumes of microdata. This subject falls outside the scope of the present paper, which specifically addresses the problem of missing detailed information.

[2] E.g. http://comtrade.un.org/, with bilateral trade in harmonized commodity description and coding systems classification reported by commodity, exporter and importer, but not how these commodities are allocated to end use sectors within a country.

[3] Earlier MRIO projects (UN: unstats.un.org/unsd/servicetrade/, Eurostat: ec.europa.eu/eurostat/statistics-explained/index.php/International_trade_in_services and OECD: www.oecd.org/tad/services-trade/) faced related problems when addressing the estimation of bilateral trade flows in services.

- Final use of imported inputs, $viy(a, i, j)$, for every region a, product i and final demand category j;
- International trade in products, $vxw(a, b, i)$, for every export region a, import region $b \neq a$ and product i.

More detailed information might be available, for example, a breakdown of intermediate vs. final exports (Lenzen et al., 2012a; Dietzenbacher et al., 2013b). For brevity we will not consider these additional situations although they can be easily incorporated in the general framework presented here.

The first two data items can be introduced directly in the intermediate and final consumption matrices, while the latter three are conventionally combined with the trade share or proportional allocation method (Moses, 1955; IDE, 2006; OECD, 2006; Peters et al., 2011; Wood et al., 2014) yielding:

$$Z(a, i : a, j) = vdf(a, i, j); \tag{1}$$

$$Y(a, i : a, j) = vdy(a, i, j); \tag{2}$$

$$Z(a, i : b, j) = \frac{vxw(a, b, i)}{vxw(*, b, i)} vif(b, i, j) \quad \text{where } b \neq a; \tag{3}$$

$$Y(a, i : b, j) = \frac{vxw(a, b, i)}{vxw(*, b, i)} viy(b, i, j) \quad \text{where } b \neq a. \tag{4}$$

In the previous expression $*$ denotes summation over all elements:

$$vxw(*, b, i) = \sum_a vxw(a, b, i).$$

Notice that in this case the export data were used to obtain shares which are applied to the import data. Naturally, if all data are consistent, the reverse strategy can be followed too.

In practice an MRIO is constructed in several stages that alternate aggregation, disaggregation and balancing of conflicting data, but in some form or another Equations 2 and 3 always form part of the toolbox of MRIO compilation (Wood et al., 2015).

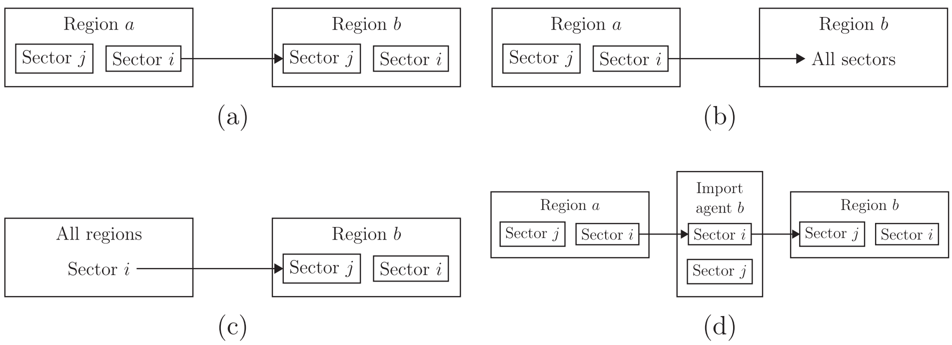
Figure 2 illustrates the trade linking of inter-industry transactions (the case of international trade for final consumption is similar). Figure 2(a) identifies the missing disaggregate flow, which under proportional allocation is estimated directly from available aggregate information on exports (Figure 2(b)) and imports (Figure 2(c)).

3.2. Topological transformation

To apply an invariance transformation to the trade linking problem it is necessary to identify the existing nodes and arcs in the original system and to examine whether the empirical aggregate information (the export of goods, $vxw(a, b, i)$, and the use of imported goods, $vdf(a, i, j)$ and $vdy(a, i, j)$) can be conceptualized as a set of adjacency matrices connecting existing nodes to virtual nodes not originally present in the problem formulation.

Figure 2 illustrates trade linking with topological transformation. Figure 2(a) represents the unknown arc between known nodes, $DF_i^a \rightarrow DF_j^b$. Figure 2(b) represents an existing

Figure 2. Representation of the trade link problem. (a) Missing disaggregate data, (b) aggregate export data, (c) aggregate import data and (d) topological transformation.



arc between a known and a virtual node, $DF_i^a \rightarrow IF_i^b$, where IF_i^b is a virtual node representing a hypothetical sector of region b which imports product i, and this flow is calibrated using international trade in products, $vxw(a, b, i)$. Figure 2(c) represents an existing arc between a virtual and a known node, $IF_i^b \rightarrow DF_j^b$ or $IF_i^b \rightarrow YF_j^b$, that is, arcs connecting node IF_i^b and, respectively, DF_i^b and YF_i^b, and are calibrated with the use of imported goods, $vif(b, i, j)$ and $viy(b, i, j)$. Finally, Figure 2(d) shows a transformed topology in which the known nodes are connected to one another through existing arcs and a virtual node, $DF_i^a \rightarrow IF_i^b \rightarrow DF_j^b$.

Let T denote the topologically transformed system (and not transpose, which is denoted by $'$) and let domestic industries, DF_i^a, be indexed before import industries, IF_i^a. The total number of intermediate nodes (i.e. neither primary nor final) is now $2n_R n_F$, and the non-empty elements of $\mathbf{Z^T}$ and $\mathbf{Y^T}$ are

$$Z^{\mathrm{T}}(a, i; 1 : a, j; 1) = vdf(a, i, j); \tag{5}$$

$$Y^{\mathrm{T}}(a, i; 1 : a, j) = vdy(a, i, j); \tag{6}$$

$$Z^{\mathrm{T}}(a, i; 1 : b, i; 2) = vxw(a, b, i); \quad \text{where } b \neq a. \tag{7}$$

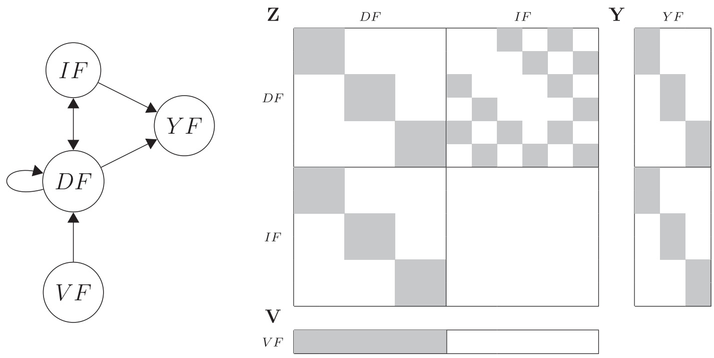
$$Z^{\mathrm{T}}(b, i; 2 : b, j; 1) = vif(b, i, j);$$

$$Y^{\mathrm{T}}(b, i; 2 : b, j) = viy(b, i, j),$$

A new element was added to the notation. Now there are two blocks of industries (domestic and imports), and the number after the semicolon indicates to which block the element belongs to. Because domestic industries are indexed before import industries, the former are in position 1 and the latter in position 2. When there is a single block (as in the case of final demand), the block index is omitted.

Figure 3(a) shows the block diagram of the topological transformation approach to trade linking. Figure 3(b) shows the matrix objects, identifying the non-empty entries in the case of three countries, two industries and a single primary input and final demand category. The third region is not a country but a composite thereof and thus exhibits self-exports, i.e. international trade of a commodity from and to itself.

Figure 3. Topological transformation of trade linking. (a) Block diagram and (b) matrix representation.



(a) (b)

We have followed through with two different approaches to the trade linking problem. It is natural to ask now whether the results are the same or different. Supplementary Information A.3 shows that multipliers or footprints calculated using algebraic or topological transformations yield exactly the same result.

3.3. Illustrative example

We now present a simple numerical example to illustrate the trade share and the topological transformation approaches to trade linking: we show how the two systems are constructed, how their results compare and conclude by examining how it is possible to query them about data not originally available.

Consider a system with three regions and two sectors per region, whose source data are presented in Table 1(a–e), where all monetary data are in million euros and GHG emissions are in ktCO_2eq. Using the notation of Section 3.1, domestic intermediate use is *vdf*, domestic final use is *vdy*, imported intermediate use is *vif*, imported final use is *viy* and exports are *vxw*.

Using this source data our goal is to build an MRIO that will tell us what is the impact in terms of GHG emissions in every sector in the world if domestic consumption of products from sector 1 in region 1 decrease by 10% and there is a corresponding increase in the consumption of imported goods from sector 1.

Note that as global production networks emerge, indirect trade flows among rest of the world economy become equally important as direct exports and imports. A global supply chains analysis using both global intermediate and final demand input coefficients with global Leontief inverse such as this one includes feedback effects via third countries. That is, when we look at the impact of a stimulus in sector 1 of region 1 in emissions in sector 2 of region 2, it does not mean that there is a sale of sector 2 in region 2 to the final demand of region 1. For example, the supply chain might have involved any sector of region 3 as an intermediate.

To answer the research question using the algebraic transformation approach, it is necessary to calculate the trade shares and estimate international inter-industry transactions using Equations 2 and 3, using the information from tables *vif*, *viw* and *vtw*. All other values

Table 1. Source data in the illustrative example.

Region 1	Sector 1	Sector 2	Final
Sector 1	1299	1604	907
Sector 2	1982	2171	475

Region 2	Sector 1	Sector 2	Final
Sector 1	1773	2355	774
Sector 2	2291	1096	774

Region 3	Sector 1	Sector 2	Final
Sector 1	2340	1425	801
Sector 2	781	2439	1288

(a) Intermediate and final use of domestic products.

Region 1	Sector 1	Sector 2	Final
Sector 1	4263	4410	2294
Sector 2	3722	3877	1934

Region 2	Sector 1	Sector 2	Final
Sector 1	3207	5014	1929
Sector 2	4356	4457	3042

Region 3	Sector 1	Sector 2	Final
Sector 1	3990	4223	1852
Sector 2	4511	3325	2597

(b) Intermediate and final use of imported products.

Sector 1	Region 1	Region 2	Region 3
Region 1	0	5331	4416
Region 2	6824	0	5649
Region 3	4142	4818	0

Sector 2	Region 1	Region 2	Region 3
Region 1	0	5540	5218
Region 2	4418	0	5214
Region 3	5115	6315	0

(c) Export tables.

	Sector 1	Sector 2
Region 1	2291	3324
Region 2	5749	872
Region 3	1905	4527

(d) Primary inputs.

	Sector 1	Sector 2	Final
Region 1	620	680	30
Region 2	720	350	50
Region 3	630	900	10

(e) Carbon emissions.

Table 2. MRIO obtained using the trade share method.

Z		R1		R2		R3		**Y** R1	R2	R3
		S1	S2	S1	S2	S1	S2			
R1	S1	1299	1604	1684	2633	1750	1853	907	1013	813
	S2	1982	2171	2036	2083	2256	1663	475	1421	1299
R2	S1	2653	2744	1773	2355	2239	2370	1427	774	1039
	S2	1725	1797	2291	1096	2254	1662	896	774	1298
R3	S1	1610	1666	1522	2380	2340	1424	866	916	801
	S2	1997	2080	2320	2374	781	2439	1038	1620	1288
$\mathbf{v}'$		2291	3324	5749	872	1905	4527			
$\mathbf{r}'$		620	680	720	350	630	900	$\mathbf{h}'$: 30	50	10

can be mapped directly from the source data, leading to the MRIO described in Table 2. All matrix objects follow the notation of Section 2.2, with **h** being the direct emissions of final demand and **v** being the vector of primary inputs. Also, for brevity 'region' and 'sector' are hereafter abbreviated to R and S, respectively.

We need to identify how much is 10% of the domestic final demand of S1 by R1, in this case 90.7. This decrease in consumption needs to be allocated among the imported final demand of S1 by R1. In the absence of additional information and keeping in the spirit of this approach, we assume proportional allocation. The shares of imports for this product category from R2 and R3 are $1427/(1427 + 866) = 0.63$ and $866/(1427 + 866) = 0.37$,

Table 3. MRIO obtained using topological transformation.

Z			DF						IF						**Y** (YF)		
			R1		R2		R3		R1		R2		R3				
			S1	S2	S1	S2	S1	S2	S1	S2	S1	S2	S1	S2	R1	R2	R3
DF	R1	S1	1299	1604							5331		4416		907		
		S2	1982	2171								5540		5218	475		
	R2	S1			1773	2355			6824				5649			774	
		S2			2291	1096				4418				5214		774	
	R3	S1					2340	1425	4142		4818						801
		S2					781	2439		5115		6315					1288
IF	R1	S1	4263	4410											2294		
		S2	3722	3877											1934		
	R2	S1			3207	5014										1929	
		S2			4356	3877										3042	
	R3	S1					3990	4223									1852
		S2					4511	3325									2597
v′		VF	2291	3324	5749	872	1905	4527									
r′		RF	620	680	720	350	630	900									
h′															30	50	10

hence the stimulus vector is (in thousand euros):

$$[-90.7 \quad 0 \quad 57.1 \quad 0 \quad 33.6 \quad 0]'.$$

The parameters of the IO model are calculated applying Equations A.4–A.6 from Supplementary Information using the data of Table 2, and the result is that impact of this stimulus in GHG emissions on the different sectors is (in ktCO_2eq)

$$[-4.27 \quad -0.23 \quad 1.90 \quad -0.06 \quad 1.40 \quad -0.47]'.$$

The net impact is -1.73 ktCO_2eq, meaning that the net effect is a reduction in global GHG emissions, in which the reduction in R1 compensates the increases in other regions.

The MRIO which is obtained by implementing topological transformations has the same structure as Figure 3(b). The non-empty arc blocks of **Z** and **Y** are zoomed in along the edges of Table 3.

To answer the research question, we again find that 10% of the domestic final demand of S1 by R1 is 90.7 and then simply subtract this value to the corresponding final demand for imports, leading to the stimulus vector (in thousand euros):

$$[-90.7 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0 \quad 90.7 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0]'.$$

The parameters of the IO model are calculated in the same way as before, by applying Equations A.4–A.6 of Supplementary Information, but now the data are given by objects that have the structure and content described earlier in this section. The impact of this stimulus in GHG emissions on the different sectors is (in ktCO_2eq)

$$[-4.27 \quad -0.23 \quad 1.90 \quad -0.06 \quad 1.40 \quad -0.47 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0]'.$$

That is, the first six-entry block of the impact – pertaining to domestic industries – is identical to the solution obtained using proportional allocation, while the second six-entry block – pertaining to import industries – yields zero emissions. This is exactly what we expected, given the results of the preceding section: the impacts on real nodes is the same using both approaches, while the impact on virtual nodes is zero.

It is important to note that since a topological transformation preserves information, it is always possible to recover the equivalent algebraically transformed disaggregate data – with the caveat that the practitioner must invoke the proportionality assumption explicitly. We consider this transparency in the clear divide between data and assumptions a major advantage. As an example, consider we wish to determine the imports of S1 produced by R2 consumed by the final demand of R3. Using the algebraically transformed system, Table 2, it is trivial to find an estimate of $Y(2, 1:3) = 1039$. No equivalent figure can be found in the topologically transformed data of Table 3 – because such data were not originally available, after all the number 1039 is not a source datum but was instead obtained using an assumption. To recover it under the topologically transformed system the user is forced to make an explicit hypothesis as to how imports of S1 from R2 are allocated among the different consumers within R3. We find total imports of S1 by R3 as $Y(3, 1; 2:3, 1) = 1852.2$, imports of S1 from R1 to R3 and from R2 to R3, respectively, as $Z(1, 1; 1:3, 1; 2) = 4416$ and $Z(2, 1; 1:3, 1; 2) = 5649$. Using proportionality we recover $1852.2 \times 5649/(4416 + 5649) = 1039$, as expected.

4. Model linking and other applications

4.1. The general approach

Thus far we considered topological transformations in the context of trade linking a harmonized symmetric MRIO table in basic prices.

The general approach to perform a topological transformation can be summarized as follows:

- Given absence of information about a disaggregate flow of interest (i.e. an arc connecting two nodes);
- And availability of aggregate flow data whose source or destination nodes are the same as those of the missing disaggregate flow of interest;
- Reformulate the model introducing virtual nodes as necessary so that available flow data can be mapped to arcs connecting pairs of nodes in the transformed model.

Topological transformations have been or can be used in many other contexts besides trade linking, although they have not necessarily been recognized as such. A few of these circumstances are listed in Table 4: trade linking was addressed earlier, multi-scale model integration is addressed now and the remainder are deferred to Supplementary Information B. This inventory is by no means exhaustive, but tries to cover various problems which exhibit different economic and/or environmental interpretations but which are fundamentally the same from the point of view of inference.

Due to space limitations, in the remainder of this section we use the network approach in one application only: the integration of a foreground country into a background MRIO. We select this application because we believe it is the one of most immediate interest to practitioners, besides trade linking. We close this section by showing how a topologically transformed system is updated when more detailed information becomes available.

Table 4. List of situations in the assembling and linking of IO models in which there is missing disaggregate data but aggregate information is available.

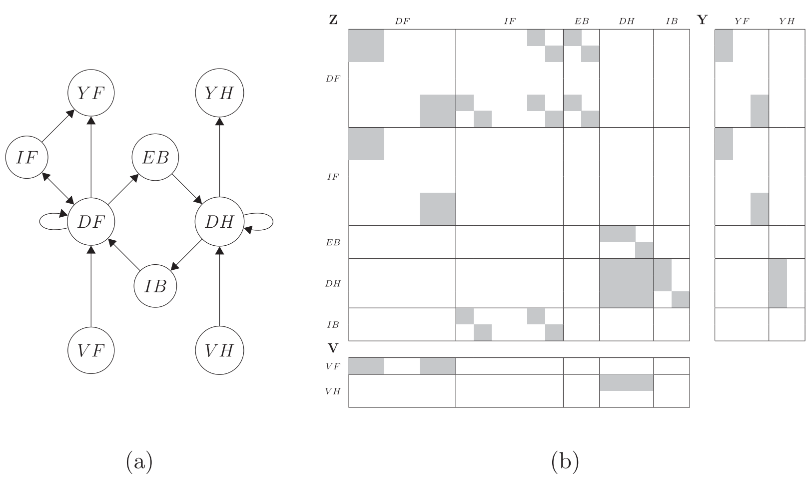
Problem	Short description
Trade linking	Detailed information between the flow of goods and services between industries in different countries is not known but exports of the bulk of a given product category between two countries is, as are the imports of a given product type by every sector in given country (Batten and Martellato, 1985)
Multi-scale integration	Combining a country foreground superior data (possibly a subnational MRIO) in a global MRIO, when the sector classification in both systems differ (Bachmann et al., 2014; Edens et al., 2015)
Supply-use transformations	In supply-use systems it may happen that a given product category is generated by multiple industries (UN, 1999)
Trade and transport margins	Information about the total volume of domestic or international trade and transport margins paid by a particular industry but lack of discrimination as to which margin provider actually delivered the service (Streicher and Stehrer, 2014)
Closed and hybrid models	A monetary description of the production side of the economy is combined with either the consumption side (Miyazawa, 1968) or a non-monetary description of production, such as process-based life-cycle inventories (Joshi, 1999) or a physical description of the energy (Guevara and Rodrigues, 2016) or waste sector (Nakamura and Kondo, 2002)
Extensions classification mismatch	In an extended IO model the level of detail of available industry information in an IO table may differ from that of the environmental extensions (Su et al., 2010; Lenzen, 2011)

4.2. *Multi-scale integration*

A global MRIO combines information from multiple sources, which had to be reconciled. Hence, its figures represent a compromise between different constraints and may therefore not respect data of a particular country (Edens et al., 2015). More generally, if the MRIO is harmonized, there simply might be more detailed national data available, possibly even a subnational MRIO (Dietzenbacher et al., 2013a; Bachmann et al., 2014; Wenz et al., 2014).

Figure 4 illustrates a global MRIO of which one particular region was replaced by superior national-level data. The background global MRIO, identical to the one described in the previous section, is assumed to have three world regions (with the last one being

Figure 4. Topological transformation of multi-scale integration. (a) Block diagram and (b) matrix version.



composite), each with two sectors, one primary input and one final demand category per region. The foreground country corresponds to the middle region of the MRIO, has three industries and two primary input and final demand categories, one of which are imports and exports, respectively.

The background model has four node blocks: primary inputs, *VF*; domestic firms, *DF*; import firms, *IF*; and final demand, *YF*. The foreground model has three node blocks: primary inputs, *HF*; domestic firms, *DH*; and final demand, *YH*. Note that imports and exports are categories of *VH* and *YH* which now must be emancipated to virtual node blocks *EB* and *IB*, respectively. The former will mediate international trade from the background to the foreground model and the latter the reverse flow.

The linking of the two models is performed in several steps. First, to avoid double-counting, all flows related to the foreground country are removed from the background model, hence the empty rows and columns corresponding to the second region of *DF* and *IF* in Figure 4(b). Second, the export and import flows to the foreground country and replaced by flows to and from virtual node blocks *EB* and *IB*, respectively. These bridge nodes have the same sector classification as the background model and this operation only involves moving the source data from one place to another.

The final and important part is now the introduction of bridge matrices $EB \rightarrow DH$ and $DH \rightarrow IB$, connecting the bridge nodes to the foreground model. These are the import and export vectors, which are removed from foreground primary inputs and final demand. These vectors may need to be rearranged to perform aggregation or disaggregation, since the number of elements in the source and destination node blocks do not necessarily match. A more thorough discussion of bridge matrices can be found in Supplementary Information A.4.

In this case the number of sectors in the foreground model is larger, so $EB \rightarrow DH$ performs a disaggregation and $DH \rightarrow IB$ an aggregation. This formulation of multi-scale data integration allows not only to represent both systems using native classifications, but also using native units, since model linking will automatically perform the unit conversion.

In this particular example we considered the situation of import and export vectors for clarity, but naturally more information can be used, if available (e.g. exports within and outside the EU).

4.3. Handling new detailed information

The problem described in Section 4.2 can be viewed as a reformulation of a topologically transformed trade linked global harmonized MRIO in which a country data block in the background model is replaced by superior data. But it is natural to ask whether it is possible to replace smaller datasets. The practical procedure will depend on the nature of this newly available information so for concreteness in this section we consider two specific cases: a single datum and a sub-global inter-regional data block.

Recall the trade link framework laid out in Section 3, whereby there is only export information by product and region but not destination sector and import information by sector and product, but not origin region. Consider that this is the only information which is available for all regions.

Suppose now that there is additional source information about a single inter-industry international transaction, $Z(a, i; 1 : b, j; 1)$. Is it possible to integrate this data in the topologically transformed system? Naturally yes. The only matter to take into account is to avoid double-counting and to ensure that this quantity is removed from the aggregate flows $Z(a,i;1:b,i;2)$ and $Z(b,i;2:b,j;1)$.

There is a smooth transition between a topologically transformed system (to model missing disaggregate data) and a system where all disaggregate data are specified from source data. The more of the latter is available the more hollowing out of the former takes place.

We now consider the case in which inter-regional intersectoral transactions become available from survey data, revisiting the numerical example of Section 3.3. In situations in which MRIO construction routines assume proportionality, new code has to be written to take the new information into consideration, whereas in the topological transformation procedure this new information can be inserted almost directly – with the caveat of checking for double-counting.

Consider that the three-region two-sector system of Table 1(a–e) still holds but additional detailed source-based information about the inter-regional intersectoral flows of R1 and R2 is now available, described in Table 5(a,b).

This new information is integrated in the topologically transformed system, Table 6 as follows. The inter-regional information concerning R1 and R2 (Table 5(a,b)) is introduced directly in the upper left quadrant of **Z** and **Y** while the information pertaining to R3 remains the same. The trade between R1 and R2 in the upper right quadrant of **Z** is simply removed.

Finally, the use of undifferentiated imported products needs to be reduced by the amount that is now available from survey data. By subtracting Table 5(a,b) to the first two tables of Table 1(b) we obtain Table 5(c,d). These reduced flows are edited in the lower right quadrant of **Z** and the bottom half of **Y** and the system is now consistent.

5. Discussion

5.1. Transparency in data processing

A topological transformation is an alternative to proportional allocation for the handling of missing disaggregate flow data in the construction of an IO model when aggregate flow data

Table 5. Additional source data concerning inter-regional flows between regions 1 and 2.

	Sector 1	Sector 2	Final
Sector 1	2701	2646	1477
Sector 2	1369	2084	965

(a) Intermediate and final use of products generated in region 2 and consumed in region 1.

	Sector 1	Sector 2	Final
Sector 1	1107	2629	1596
Sector 2	1970	2165	1405

(b) Intermediate and final use of products generated in region 1 and consumed in region 2.

	Sector 1	Sector 2	Final
Sector 1	1562	1764	816
Sector 2	2353	1793	969

(c) Intermediate and final use of imported products of region 1 not originating in region 2.

	Sector 1	Sector 2	Final
Sector 1	2100	2385	334
Sector 2	2386	2292	1636

(d) Intermediate and final use of imported products of region 2 not originating in region 1.

Table 6. Topologically transformed MRIO incorporating inter-regional source data concerning regions 1 and 2.

Z			*DF*						*IF*						**Y** *YF*		
			R1		R2		R3		R1		R2		R3				
			S1	S2	S1	S2	S1	S2	S1	S2	S1	S2	S1	S2	R1	R2	R3
DF	R1	S1	1299	1604	1107	2629							4416		907	1596	
		S2	1982	2171	1970	2165								5218	475	1405	
	R2	S1	2701	2646	1773	2355							5649		1477	774	
		S2	1369	2084	2291	1096								5214	965	774	
	R3	S1					2340	1425	4142		4818						801
		S2					781	2439		5115		6315					1288
IF	R1	S1	1562	1764											816		
		S2	2353	1793											969		
	R2	S1			2100	2385										334	
		S2			2386	2292										1636	
	R3	S1					3990	4223									1852
		S2					4511	3325									2597
v′	*VF*		2291	3324	5749	872	1905	4527									
r′	*RF*		620	680	720	350	630	900							**h′**		
															30	50	10

are available. Although it does not alter the result of the computations when performing the analysis, a topological transformation improves the transparency of the data processing procedure itself by using the least possible data manipulation.

To visualize the perils of mathematical artifacts consider the illustrative example reported in Section 3.3. The algebraically transformed data of Table 2 reports blocks of data of widely discrepant quality without any obvious disclaimer. Value $Z(1,1:1,1) = 1299$ is a source datum, which (hopefully) embodies expert knowledge (Manski, 2014). Neighbouring value $Z(1,1:2,1) = 1684$, however, is just a crude estimate that in fact can take any value in the range from 0 to the minimum of $vxw(1,2,1) = 5331$ and $vif(2,1,1) = 3207$.

The reporting of official statistical data and estimated data side by side without additional information raises unnecessary concerns about the quality of the database as a whole. Clearly, there are benefits in avoiding, where possible, the processing and storing of mathematical manipulations that do not improve the overall accuracy of the model (see Jensen, 1980 for a discussion on these points).

A practical situation in which the conflation of true data and artifacts resulting from algebraic transformations matters is in data balancing (Lahr and de Mesnard, 2004). Consider the problem of trade linking when the values of the import matrix do not match the corresponding sum of bilateral trade data, a common real-world situation. For example, van der Linden and Oosterhaven (1995) use RAS balancing after the MRIO initial estimate construction and EXIOBASE (Wood et al., 2015), WIOD (Dietzenbacher et al., 2013b) and Eora (Lenzen et al., 2014) also use modern balancing algorithms.

If there is an inconsistency between two source data values that should match and do not it is because one is wrong and the other correct (or at least less wrong). If this inconsistency is resolved by balancing an MRIO after proportional allocation has been performed then each item in the reconciled system will be a product of multiple original data points so it is now impossible to clearly identify where the inconsistency originated. This process is therefore opaque, leading to information degradation.

Alternative, if balancing is performed on the topologically transformed model, there is a one-to-one correspondence between an element in the balanced system and an original data point. It becomes possible to identify the data items which exhibit larger deviations and prioritize data refining using expert knowledge to improve specifically those items. Even if it is impractical to refine the source data it is still possible to perform data balancing anew but this time using different reliability or data quality indices (Rodrigues, 2014) according to the degree of inconsistency observed for each datum.

5.2. Modularity in analysis

A different aspect with regard to which topological transformations may be of use is for the seamless integration of different models. The use of algebraic transformations requires manipulating datasets in a way which prevents their subsequent reuse, a situation which does not occur with topological transformations. For clarity, consider an example in which we wish to link a monetary domestic supply-use table to a sectorally disaggregated energy model (Guevara and Rodrigues, 2016).

The information available to link the two models is the use of energy carriers by industries in the economy, and the use of products from the rest of the economy by energy technologies. In a topologically transformed model both industries and products of the rest of the economy, as well as energy carriers and technologies are represented, and the linking can proceed without additional data transformations. In the algebraically transformed model (e.g. industry-by-industry and energy technology-by-technology) we would have to convert the use of energy carriers to use of energy technology outputs using an allocation transformation as well as converting the use of products by energy technologies to use of industry outputs.

Thus, topological transformations dispense with the explicit allocations (implicitly considering the industry-technology assumption) and allow the use of data-as-is in the hybrid model, eventually using bridge matrices (whose source information can always be recovered). This is very convenient for data reuse, since it allows the plug-and-play of different datasets without the need to retrace back previous data transformations to recover the original information.

We believe that in the future we will witness a higher level of data integration across spatial and sectoral scales. Topological transformation thus offer the possibility for the systematic exploration of different model configurations (Fox and Krishna, 1965; Hamilton et al., 1991; Robinson and Duffy-Deno, 1996) to obtain the optimal combination that maximizes resolution while minimizing noise resulting from data estimation.

5.3. Computation and storage

A topologically transformed system is sparser, often much more so, than its algebraically transformed counterpart. This opens new possibilities for the expansion in the detail and scope of the model, but also poses challenges, since conventional computational methods and data structures may be no longer suitable.

For concreteness, consider the trade link problem of Section 3. With n_R regions and n_S sectors, the number of elements in $\mathbf{Z}$ and $\mathbf{Y}$ in the algebraically transformed system is $n_A = n_R^2 n_S^2$. By contrast, in its topologically transformed counterpart it is $n_T = 2 \times n_R n_S^2 + n_R^2 n_S$. Hence the savings in storage requirements of using a topological transformation instead of proportional allocation are

$$1-\frac{n_T}{n_A}=1-\left(\frac{2}{n_R}+\frac{1}{n_S}\right).$$

To use examples of existing MRIOs to assess the magnitude of these storage gains consider that GTAP 7 has $n_R = 112$ and $n_S = 57$, yielding savings of 96.46% while Exiobase, $n_R = 48$, $n_S = 200$ and savings of 95.33%. Using the big O notation, if the number of sectors and regions are of the same order of magnitude $n_R \simeq n_A \simeq n$, the storage requirements of proportional allocation are $O(n^4)$, and the storage requirements of topological transformations are $O(n^3)$.

Naturally each of the mentioned MRIOs embodies additional information besides proportional allocation, but these numbers are nonetheless suggestive. If we wish to develop the field by integrating global and subnational MRIOs and sectorally detailed models, the storage savings become even larger. Notice also that data storage requirements have implications for computationally intensive operations such as structural path analysis or Monte Carlo integration.

A matrix with a large proportion of zero entries is said to be sparse and there are standard implementations to efficiently store such objects, such as compressed row/column storage (Golub and Van Loan, 1996). The implications of sparse matrix storage in the field of IOA and life cycle assessment have been earlier explored by Duchin and Szyld (1979) and Peters (2007). Note that a sparse data format requires saving the position of rows and column elements besides the value proper, so it involves an overhead, even though computing integers is lighter than computing floats.

Finally, the inverse of a sparse matrix is usually not sparse, and therefore it is convenient to implement the calculations of multipliers, footprints and other IO quantities of interest in a way which avoids the explicit presentation of the Leontief inverse and, ideally, only requires implementing matrix–vector or vector–vector operations. In Supplementary Information A.5 we provide an example of how this can be achieved.

6. Conclusions

The present paper proposes a novel approach to the problem of IO model assembly and linking in the presence of missing disaggregate data, based on the application of invariance transformations (Jaynes, 1973) to inductive inference (Jaynes, 2003). The conventional approach to this problem consists in the use of proportional allocation to estimate the missing data. Our approach, which we termed topological transformation, first recognizes that IO tables are adjacency matrices in a network and introduces additional nodes so that available data can be mapped to arcs present in the transformed system.

We described the problem of trade linking as a case-study to illustrate proportional allocation and topological transformation, and showed that they yield numerically identical results. We then proceeded to illustrate the application of topological transformation to different configurations which spanned a wide range of missing disaggregate data problems found in IO analysis. Finally, we discussed the implications of topological transformation in terms of data processing, data analysis and computational aspects.

Topological transformations are a tool to address missing disaggregate data problems and may contribute to improving data quality by increasing transparency, and offer new possibilities for the expansion of IO analysis, by cutting data storage requirements (Cheung, 2011) and by allowing for the seamless integration of datasets without additional algebraic manipulation.

We hope this combination of features will facilitate future collaborative efforts to address global environmental problems (Liu et al., 2015).

Acknowledgments

We thank editor Bart Los and two anonymous referees for their constructive criticism. Any errors the paper may contain are the sole responsibility of the authors.

Disclosure statement

No potential conflict of interest was reported by the authors.

ORCiD

J. Rodrigues http://orcid.org/0000-0002-1437-0059

References

Bachmann C., Roorda M.J. and Kennedy C. (2014) Developing a Multi-scale Multi-region Input–Output Model. *Economic Systems Research*, 27, 172–193.

Batten D. and Martellato D. (1985) Classical Versus Modern Approaches to Interregional Input–Output Analysis. *The Annals of Regional Science*, 19, 1–15.

Bruckner M. (2015) Hybrid Input–Output Approaches to Trade and Footprint Accounting. In: *The Impact of Consumption on Ecosystems and Biodiversity Beyond Borders – A TEEB Perspective*. Isle of Vilm. http://img.teebweb.org/wp-content/uploads/2015/12/Martin_Bruckner_hybrid_input-output_approaches.pdf

Cheung P. (2011) Input–Output Analysis and Contemporary Official Statistics (Communication at the 19th International Input–Output Conference, 13–17 June, Alexandria).

Dietzenbacher E., Guilhoto J.J.M. and Imori D. (2013a) The Role of Brazilian Regions in the Global Value Chain (21st International Input–Output Conference, Kitakyushu.

Dietzenbacher E., Los B., Stehrer R., Timmer M. and de Vries G. (2013b) The Construction of World Input–Output Tables in the WIOD Project. *Economic Systems Research*, 25, 71–98.

Duchin F. and Levine S.H. (2010) Embodied Resource Flows and Product Flows. *Journal of Industrial Ecology*, 14, 586–597.

Duchin F. and Szyld D.B. (1979) Application of Sparse Matrix Techniques to Inter-regional Input–Output Analysis. *Economics of Planning*, 15, 142–167.

Edens B., Hoekstra R., Zult D., Lemmers O., Wilting H. and Wu R. (2015) A Method to Create Carbon Footprint Estimates Consistent with National Accounts. *Economic Systems Research*, 27, 440–457.

Fox K.A. and Krishna Kumar K.T. (1965) The Functional Economic Area: Delineation and Implications for Economic Analysis and Policy. *Papers in Regional Science*, 15, 57–85.

Golub G.H. and Van Loan C.F. (1996) *Matrix Computations*. 3rd ed. Baltimore, MD, John Hopkins University Press.

Guevara Z. and Rodrigues J.F.D. (2016) Structural Transitions and Energy Use: A Decomposition Analysis of Portugal 1995–2010. *Economic Systems Research*, 28, 202–223. doi:10.1080/09535314.2016.1157456.

Hamilton J.R., Robinson M.H., Whittlesey N.K. and Ellis J. (1991) Economic Impacts, Value Added, and Benefits in Regional Project Analysis. *American Journal of Agricultural Economics*, 73, 334–344.

Heijungs R. (2015) Topological Network Theory and Its Application to LCA and IOA and Related Industrial Ecology Tools: Principles and Promise. *Journal of Environmental Accounting and Management*, 3, 151–167.
IDE (2006) Asian International Input–Output Tables 2000 Volume 1. Explanatory Notes (Technical Report, Institute of Developing Economies, Japan External Trade Organization).
Isard W. (1951) Interregional and Regional Input–Output Analysis: A Model of a Space-economy. *Review of Economic and Statistics*, 33, 318–329.
Jaynes E.T. (1973) The Well-posed Problem. *Foundations of Physics*, 3, 477–492.
Jaynes E.T. (2003) *Probability Theory: The Logic of Science*. Cambridge, Cambridge University Press.
Jensen R.C. (1980) The Concept of Accuracy in Regional Input–Output Models. *International Regional Science Review*, 5, 139–154.
Joshi S. (1999) Product Environmental Life-cycle Assessment Using Input–Output Techniques. *Journal of Industrial Ecology*, 3, 95–120.
Kagawa S., Okamoto S., Suh S., Kondo Y. and Nansai K. (2013) Finding Environmentally Important Industry Clusters: Multiway Cut Approach Using Nonnegative Matrix Factorization. *Social Networks*, 35, 423–438.
Lahr M.L. and de Mesnard L. (2004) Biproportional Techniques in Input–Output Analysis: Table Updating and Structural Analysis. *Economic Systems Research*, 16, 115–134.
Lenzen M. (2011) Aggregation Versus Disaggregation in Input–Output Analysis of the Environment. *Economic Systems Research*, 23, 73–89.
Lenzen M., Gallego B. and Wood R. (2009) Matrix Balancing Under Conflicting Information. *Economic Systems Research*, 21, 23–44.
Lenzen M., Kanemoto K., Moran D. and Geschke A. (2012a) Mapping the Structure of the World Economy. *Environmental Science & Technology*, 46, 8374–8381.
Lenzen M., Moran D., Kanemoto K., Foran B., Lobefaro L. and Geschke A. (2012b) International Trade Drives Biodiversity Threats in Developing Nations. *Nature*, 486, 109–112.
Lenzen M., Geschke A., Wiedmann T., Lane J., Anderson N., Baynes T., Boland J., Daniels P., Dey C., Fry J., Hadjikakou M., Kenway S., Malik A., Moran D., Murray J., Nettleton S., Poruschi L., Reynolds C., Rowley H., Ugon J., Webb D. and West J. (2014) Compiling and Using Input–Output Frameworks Through Collaborative Virtual Laboratories. *Science of the Total Environment*, 485–486, 241–251.
Leontief W. (1941) *The Structure of the American Economy, 1919–1929*. Cambridge, MA, Harvard University Press.
Lindner S., Legault J. and Guan D. (2012) Disaggregating Input–Output Models with Incomplete Information. *Economic Systems Research*, 24, 329–347.
Liu J., Mooney H., Hull V., Davis S.J., Gaskell J., Hertel T., Lubchenco J., Seto K.C., Gleick P., Kremen C. and Li S. (2015) Systems Integration for Global Sustainability. *Science*, 347, 1258832.
Manski C.F. (2014) Communicating Uncertainty in Official Economic Statistics (Working Paper 20098, National Bureau of Economic Statistics, Cambridge, MA).
Miller R.E. and Blair P.D. (2009) *Input–Output Analysis: Foundations and Extensions*. 2nd ed. Cambridge, Cambridge University Press.
Miyazawa K. (1968) Input–Output Analysis and Interrelational Income Multiplier as a Matrix. *Hitotsubashi Journal of Economics*, 1, 38–55.
Moses L. (1955) The Stability of Interregional Trading Patterns and Input–Output Analysis. *The American Economic Review*, 45, 803–826.
Nakamura S. and Kondo Y. (2002) Input–Output Analysis of Waste Management. *Journal of Industrial Ecology*, 6, 39–63.
OECD (2006) The OECD Input–Output Database: 2006 Edition (Technical Report, Organisation for Economic Co-operation and Development, Paris).
Oosterhaven J., Stelder D. and Inomata S. (2008) Estimating International Interindustry Linkages: Non-survey Simulations of the Asian-Pacific Economy. *Economic Systems Research*, 20, 395–414.
Peters G.P. (2007) Efficient Algorithms for Life Cycle Assessment, Input–Output Analysis, and Monte-Carlo Analysis. *International Journal of Life Cycle Assessment*, 12, 373–380.

Peters G.P., Andrew R. and Lennox J. (2011) Constructing an Environmentally Extended Multi-regional Input–Output Table Using the GTAP Database. *Economic Systems Research*, 2, 131–152.
Robinson M.H. and Duffy-Deno K.T. (1996) The Role of the Oil and Gas Industry in Utah's Economy: An Economic Base/Input–Output Analysis. *Resource and Energy Economics*, 18, 201–218.
Rodrigues J.F.D. (2014) A Bayesian Approach to the Balancing of Statistical Economic Data. *Entropy*, 16, 1243–1271.
Rodrigues J.F.D. (2016) Maximum Entropy Prior Uncertainty and Correlation of Statistical Economic Data. *Journal of Business and Economic Statistics*, 34, 357–367.
Streicher G. and Stehrer R. (2014) Whither Panama? Constructing a Consistent and Balanced World SUT System Including International Trade and Transport Margins. *Economic Systems Research*, 27, 213–237.
Su B., Huang H.C., Ang B.W. and Zhou P. (2010) Input–Output Analysis of CO_2 Emissions Embodied in Trade: The Effects of Sector Aggregation. *Energy Economics*, 32, 166–175.
Tukker A., Edens B., Schmidt J., Masvar R., Pulles T. and Wood R. (2011) Compiling and Refining Environmental and Economic Accounts (Inception Report D2.1, Collaborative Project and Grant Agreement No: 265134, Delft. Funded by the EU's Seventh Framework Program – Theme ENV.2010.4.2.2-1).
Tukker A., de Koning A., Wood R., Hawkins T., Lutter S., Acosta J., Rueda Cantuche J.M., Bouwmeester M.C., Oosterhaven J., Drosdowski T. and Kuenen J. (2013) EXIOPOL – Development and Illustrative Analyses of a Detailed Global mr eeSUT/IOT. *Economic Systems Research*, 25, 50–70.
UN (1999) Handbook of Input–Output Table Compilation and Analysis (Studies in Methods (Series F) No. 74, United Nations, New York, NY).
UN (2010) System of National Accounts 2008 (Studies in Methods (Series F) No. 2, United Nations, New York, NY).
van der Linden J.A. and Oosterhaven J. (1995) European Community Intercountry Input–Output Relations: Construction Method and Main Results for 1965–85. *Economic Systems Research*, 7, 249–270.
Wenz L., Willner S.N., Radebach A., Bierkandt R., Steckel J.C. and Levermann A. (2014) Regional and Sectoral Disaggregation of Multi-regional Input–Output Tables – A Flexible Algorithm. *Economic Systems Research*, 27, 194–212.
Wood R., Hawkins T.R., Hertwich E.G. and Tukker A. (2014) Harmonising National Input–Output Tables for Consumption-Based Accounting – Experiences from Exiopol. *Economic Systems Research*, 26, 387–409.
Wood R., Stadler K., Bulavskaya T., Lutter S., Giljum S., Koning A.D., Kuenen J., Schutz H., Acosta-Fernandez J., Usubiaga A., Simas M., Ivanova O., Weinzettel J., Schmidt J.H., Merciai S. and Tukker A. (2015) Global Sustainability Accounting-developing Exiobase for Multi-regional Footprint Analysis. *Sustainability*, 7, 138–163.