Economic Systems Research

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

INVESTIGATING ALTERNATIVE APPROACHES TO HARMONISE MULTI-REGIONAL INPUT–OUTPUT DATA

Arne Geschke, Richard Wood, Keiichiro Kanemoto, Manfred Lenzen & Daniel Moran

To cite this article: Arne Geschke, Richard Wood, Keiichiro Kanemoto, Manfred Lenzen & Daniel Moran (2014) INVESTIGATING ALTERNATIVE APPROACHES TO HARMONISE MULTI-REGIONAL INPUT–OUTPUT DATA, Economic Systems Research, 26:3, 354-385, DOI: 10.1080/09535314.2014.937069

To link to this article: https://doi.org/10.1080/09535314.2014.937069

Published online: 22 Jul 2014.

Economic Systems Research, 2014
Vol. 26, No. 3, 354–385, http://dx.doi.org/10.1080/09535314.2014.937069

Routledge
Taylor & Francis Group

INVESTIGATING ALTERNATIVE APPROACHES TO HARMONISE MULTI-REGIONAL INPUT–OUTPUT DATA

ARNE GESCHKE[a*], RICHARD WOOD[b], KEIICHIRO KANEMOTO[c], MANFRED LENZEN[a] and DANIEL MORAN[b]

[a]*ISA, School of Physics A28, The University of Sydney, NSW, Australia;* [b]*Industrial Ecology Program, NTNU, Trondheim, Norway;* [c]*Institute of Decision Science for a Sustainable Society, Kyushu University, Fukuoka, Japan*

(*Received 12 November 2013; accepted 16 June 2014*)

Over recent years a small number of global multi-regional input–output (MRIO) databases were developed to describe the entire global economy at high sector detail. We investigate the differences that arise out of applying different construction procedures for two global MRIO databases: The EXIOBASE database, developed as part of the EU FP6 & 7 programs and the Eora database developed at the University of Sydney. The procedures used in EXIOBASE involve a high degree of interrogation and adjustment throughout the construction of the data set, whilst the Eora MRIO relies on single-step mathematical programming techniques and high-performance computing. We unravel the effect of the different approaches taken to develop the databases by undertaking a number of combinatorial experiments in which we exchange parts of the construction process between the EXIOBASE and Eora build pipelines. We conclude that Eora's highly automated data reconciliation approach produces MRIO databases that are of comparable quality to those constructed with EXIOBASE's multi-step approach. However, the reliability and robustness of the resulting MRIO database largely depend on the level of detail and reliability of the underlying raw data.

Keywords: Multi-regional input–output analysis; Matrix balancing; Constrained optimisation; Automation; Matrix distance

1. INTRODUCTION

Multi-regional input–output tables (MRIOs) are becoming increasingly policy-relevant as climate policy negotiators are looking at consumption-based accounting (The Brookings Institution, 2013). Considerable work is now being undertaken on developing MRIO databases (Daniels et al., 2011; Feng et al., 2011; Skelton et al., 2011; Su and Ang, 2011; Ewing et al., 2012; Lenzen et al., 2012; Liu et al., 2012; Wilting, 2012; Andrew and Peters, 2013; Lenzen et al., 2013; Tukker et al., 2013b; Wiedmann and Barrett, 2013; Galli et al., 2013). However, comparisons (see editorial of this special issue) have shown that these MRIO databases yield diverging results, for example for carbon footprints of nations (Peters et al., 2012). Some of these divergences are due to different classifications and levels of aggregation (see Steen-Olsen et al., 2014), some due to different initial assembly and reconciliation techniques, and others are due to different source data being used (Owen and Barrett, 2013). In general, all of the existing MRIO databases are affected by shortcomings

*Corresponding author. E-mail: arne.geschke@sydney.edu.au

in at least one of these areas of divergence; there exists currently no MRIO database which is better than any other.

The availability of several MRIO models is beneficial in testing the impacts of the different construction techniques on final results. As a consequence the question arises of how we can explore the different construction techniques across the models. Can we apply a construction technique from one database, and apply it to the sector classifications and/or source data of another database? And: How would the resulting MRIO database differ from those original databases? To answer these questions, we undertake a number of numerical experiments where we mix the initial estimate, constraints (i.e. data sources), and reconciliation methods of three MRIO frameworks: EXIOBASE v1.0,[1] EXOBASE v2.0 and Eora.[2] In particular, we take combinations of initial estimates (IEs) and constraints from all three frameworks, and apply the Eora reconciliation method. The outcomes of these experiments are then compared with the 'pure' EXIOBASE v1.0, EXIOBASE v2.0, and Eora MRIO databases.

One of the purposes of this work is to compare the performance of single-step highly automated matrix reconciliation techniques with multi step-wise approaches. To this end we employ a derivative of the KRAS-type (Lenzen et al., 2010a) constrained optimisation technique to resolve all differences in data and balances in one-step to construct global, multiregional input–output tables for the year 2007.

In particular, we investigate the following MRIO components:

(1) Initial input data
   (a) 2000 EXIOBASE v1.0 final data set,
   (b) EXIOBASE v2.0 initial data set,[3]
   (c) Eora data set for 2006. According to the workflow used for Eora, this serves as the IE for 2007.
(2) Constraints
   (a) Eora constraints data set,
   (b) EXIOBASE v2.0 constraints data set.
(3) Reconciliation approaches
   (a) EXIOBASE v1.0 (only for the original EXIOBASE v1.0 data set),
   (b) EXIOBASE v2.0 (only for the original EXIOBASE v2.0 data set),
   (c) Eora.

Using these components we come up with eight combinations of initial inputs, constraints and reconciliation methods. For each of these combinations, we construct an MRIO

---

[1] The EXIOBASE database is currently available in two different versions: EXIOBASE v1.0 (developed from the EXIOPOL project) is an MRIO framework for the year 2000, EXIOBASE v2.0 (developed from the CREEA project) is an MRIO framework for the year 2007. EXIOBASE v2.0 is the result of EXIOPOL's follow-up project. The two databases (EXIOBASE v1.0 and EXIOBASE v2.0) differ in size and detail, but mainly cover the same countries and regions. The different structures (particularly the sector classifications) of EXIOBASE v1.0 and EXIOBASE v2.0 are discussed in Section 2.2. EXIOBASE is available under http://www.exiobase.eu/. The database if briefly described in Appendix 1. For more detailed information, see Tukker et al. (2013a) and Tukker (2013)

[2] For more information on Eora, see Appendix 4, Lenzen et al. (2012, 2013) and Moran (2013).

[3] A fully assembled IE data set for EXIOBASE v2.0 was not available at the time of this project. The construction of the IE data set that was used during this project is described in Section 2.

database in EXIOBASE v2.0 output format. We compare the MRIO databases resulting from the various combinations using a set of matrix distance measures obtained from the literature. These matrix distances provide an estimate of the 'closeness' of any of our MRIO databases with their 'pure' counterparts.

All calculations that are performed during this project are controlled by a software package called AISHA, developed at the University of Sydney during the construction phase of Eora (Geschke et al., 2011). AISHA is capable of constructing large-scale MRIOs of arbitrary structure. AISHA is explained in Section 2.

This article unfolds as follows. In Section 2 we describe our methods and data, as well as the MRIO databases which were constructed by combining the various combinations and components. Section 3 presents the results obtained from the comparison of the different databases. In Section 4 we draw conclusions and give an outlook.

2. METHODS

2.1. Numerical Experiments

We evaluate combinations of build pipeline components of three MRIO databases: EXIOBASE v1.0 (in the following abbreviated by 'X'), EXIOBASE v2.0 (C) and Eora (E). We distinguish three such components: (1) the IE data set, (2) the set of constraints imposed during the reconciliation process and (3) the reconciliation method. Note that the EXIOBASE v2.0 IE used to construct databases in this study was constructed by expanding export and import vectors according to trade shares. We used the approach developed by Andrew and Peters for the GTAP database (Andrew and Peters, 2013). Therefore, we abbreviate this IE as 'G', or refer to it as the 'A&P-approach'. The original EXIOBASE v2.0 IEs are not fully assembled. Instead, individual country IEs are constructed, which are individually reconciled. The fully reconciled individual data sets are then used as the domestic data blocks within the multi-regional framework. Finally, this multi-regional framework is completed by constructing the trade blocks using a method similar to the approach presented by Andrew and Peters (2013).

Each database can be uniquely represented by combining these shortcuts into a three-letter abbreviation. The first letter refers to the IE, the second letter to the constraints data set and the final to the construction method. The abbreviation ECE, for example, refers to the database featuring Eora's IE, EXIOBASE v2.0's constraints data set and Eora's construction method (Table 1).

The MRIOs 1–3 in Table 1 were constructed according to the individual methodology used for each database. These databases were not constructed during this project, only taken as a reference point for the matrix distances reported in Section 3. The databases 4–8 were constructed according to the Eora construction method during the project. The mathematical problem for reconciling MRIO databases as it is used in Eora's construction methodology is given by

$$\min_{\boldsymbol{\tau}} f(\boldsymbol{\tau}, \boldsymbol{\tau}^0, \boldsymbol{\sigma}_{\boldsymbol{\tau}^0}, \boldsymbol{\sigma}_{\mathbf{c}}) \quad \text{s.t. } \mathbf{M}\boldsymbol{\tau} = \mathbf{c}, \quad \mathbf{l} \leq \boldsymbol{\tau} \leq \mathbf{u}. \tag{1}$$

Within this equation, the databases are treated in a vectorised form. The IE is given by $\boldsymbol{\tau}^{\mathbf{0}}$, the constraints set is given by $\mathbf{M}\boldsymbol{\tau} = \mathbf{c}$. The final, fully reconciled data set is given by the variable $\boldsymbol{\tau}$.

TABLE 1. Databases used for this study.

No	IE data set	Const. data set	Construction	Short	Notes
1	EXIOBASE v1.0	EXIOBASE v1.0	EXIOBASE v1.0	XXX	Database not constructed during this project, but used as a reference
2	EXIOBASE v2.0	EXIOBASE v2.0	EXIOBASE v2.0	CCC	Database not constructed during this project, but used as a reference
3	Eora	Eora	Eora	EEE	Database not constructed during this project, but used as a reference
4	EXIOBASE v1.0	EXIOBASE v2.0	Eora	XCE	
5	EXIOBASE v2.0	EXIOBASE v2.0	Eora	GCE	see Section 2.4.1
6	Eora	EXIOBASE v2.0	Eora	ECE	
7	EXIOBASE v1.0	Eora	Eora	XEE	
8	EXIOBASE v2.0	Eora	Eora	GEE	see Section 2.4.1

Notes: The databases numbered 1–3 were used as reference tables for this study. The databases 4–6 were reconciled according to the EXIOBASE v2.0 constraints set, the remaining databases were reconciled according to the Eora constraints set.

For example, if we were to identify each of the variables by putting the name of the database into the subscript of that variable, then database no. 4 would be given by

$$\begin{aligned} \min_{\boldsymbol{\tau}_{\mathrm{XCE}}} \quad & f\left(\boldsymbol{\tau}_{\mathrm{XCE}}, \boldsymbol{\tau}^{0}_{\mathrm{EXIOBASE\ v1.0}}, \boldsymbol{\sigma}_{\boldsymbol{\tau}^{0}_{\mathrm{EXIOBASE\ v1.0}}}, \boldsymbol{\sigma}_{\mathbf{c}_{\mathrm{EXIOBASE\ v2.0}}}\right) \\ \text{s.t.} \quad & \mathbf{M}_{\mathrm{EXIOBASE\ v2.0}}\boldsymbol{\tau}_{\mathrm{XCE}} = \mathbf{c}_{\mathrm{EXIOBASE\ v2.0}}, \quad \mathbf{l} \le \boldsymbol{\tau}_{\mathrm{XCE}} \le \mathbf{u}. \end{aligned}$$

The boundary vectors **u** and **l** are invariant under the data sets as they depend on the final classification (which is is equal for all databases).

The entire work flow of models 4–8 is controlled by a software package called AISHA, developed at the University of Sydney during the construction phase of Eora (Geschke et al., 2011). AISHA is capable of constructing large-scale MRIOs of arbitrary structure. The key steps in the AISHA workflow are

(1) Construction of a fully populated IE.
(2) Construction of a constraints data set according to the user's specifications.
(3) Reconciliation of MRIO.
(4) Recovery of the result of the reconciliation process, and data export in various formats.

2.2. Bridging Different Regional and Sectoral Classifications

The three MRIO databases that were used in this publication have different classification systems. These are:

- *EXIOBASE v1.0*. EXIOBASE v1.0's classification system is a homogeneous supply-use tables (SUTs) classification. EXIOBASE v1.0 has 44 regions, each of these represented by an SUT structure. EXIOBASE v1.0 focuses on European countries and major economies outside Europe. The first 43 regions of EXIOBASE v1.0 are dedicated to individual countries, the last region summarises the remaining countries into a Rest-of-the-Word (RoW) region. EXIOBASE v1.0 features 129 industries and 129 products for each of its regions. For this project, only the basic price sheet was used.
- *EXIOBASE v2.0* (EXIOBASE 2.0). EXIOBASE v2.0's classification system is a homogeneous classification SUT-classification. EXIOBASE v2.0 has 48 regions in total, each of these represented by a supply-use structure. The first 43 regions are identical to those in EXIOBASE v1.0, but the RoW-region is divided into five different regions. Within each of these five RoW-regions, all RoW-countries belonging to a particular geographic region are summarised. EXIOBASE v2.0 offers a higher sectoral detail than EXIOBASE v1.0. It features 163 industries and 200 products for each of its 48 regions. EXIOBASE v2.0 also offers a number of valuations on top of the basic price sheet. However, for this project only the basic price sheet will be used. EXIOBASE v2.0 is available for the year 2007.
- *Eora*. The Eora database has a heterogenous classification structure. It features 187 regions, which all correspond to individual countries. The sector classifications for each country range from 25 to 500 sectors. Where possible, the native classifications are preserved for individual countries. Hence, Eora features industry input–output tables (IIOT), commodity input–output tables (CIOT), as well as SUT structures. Eora features five valuation sheets (basic price, trade margin, transport margin, taxes and subsidies), and it is available as a time series for each year from 1990 until 2011.

All MRIO databases in this project were constructed

(1) for the year 2007 (the same year that EXIOBASE v2.0 was constructed for),
(2) in EXIOBASE v2.0's regional and sectoral classification and
(3) using Eora's five valuation sheets.

The classification defined by points 2 and 3 will be referred to as the *final classification*.

The final classification has the following regional and sectors structure

- 48 regions,
- each region to be represented in the SUT structure,
- 163 industry sectors,
- 200 product sectors,
- 7 final demand sectors,
- 6 value-added sectors,

and the valuations

- basic price,
- trade and,
- transport margin,
- taxes and
- subsidies.

For the data preparation, it is therefore generally sufficient to develop concordances that map Eora or EXIOBASE v1.0 to the EXIOBASE v2.0 classification.

For all databases, the unit for the data is million Euros. This is the generic unit for EXIOBASE v2.0 and EXIOBASE v1.0. Eora, however, is published in thousand US-dollars (USD). The official IMF-based exchange rate for the year 2007 is given by 1 Euro = 1.3705 USD.[4]

All databases (nos. 4–8 in Table 1) were reconciled in the final classification. Hence, the EXIOBASE v1.0 data set, the Eora IE data set and the Eora constraints data set had to be converted, in order to align with the final classification.

2.3. Overall Work Flow

The process to obtain an MRIO database can be summarised in the following four steps.

(1) One of the available IE is selected and vectorised as a $N \times 1$ vector $\boldsymbol{\tau}^0$ (assuming the IE contains N values).
(2) All available constraints' data (assume M points) are collated into a vector $\mathbf{c}$.
(3) An $M \times N$ matrix $\mathbf{M}$ is set up that contains constraint coefficients describing the relationship between M constraints' data points and N MRIO table elements. In addition, vectors $\mathbf{l}$ and $\mathbf{u}$ of the dimension $N \times 1$ are constructed that contain lower and upper boundaries on all MRIO elements in $\boldsymbol{\tau}$. These lower and upper boundaries result from definitions of accounting variables. For example, the boundaries for changes in inventories are $[-\infty, +\infty]$, those for subsidies are $[-\infty, 0]$, and those for remaining MRIO elements are $[0, +\infty]$. Finally, two vectors $\boldsymbol{\sigma}_{\boldsymbol{\tau}^0}$ and $\boldsymbol{\sigma}_{\mathbf{c}}$ holding information about the reliability (in the form of standard deviations) for each element of $\boldsymbol{\tau}^0$ and $\mathbf{c}$ are constructed.
(4) A constrained optimisation algorithm is invoked for finding a solution for $\boldsymbol{\tau}$ that best fulfils the constraints $\mathbf{M}\boldsymbol{\tau} = \mathbf{c}$ and $\mathbf{l} \leq \boldsymbol{\tau} \leq \mathbf{u}$, whilst minimising the departure of $\boldsymbol{\tau}$ from its IE $\boldsymbol{\tau}^0$. The complete data reconciliation problem then becomes (see Equation 1).

$$\min_{\boldsymbol{\tau}} f(\boldsymbol{\tau}, \boldsymbol{\tau}^0, \boldsymbol{\sigma}_{\boldsymbol{\tau}^0}, \boldsymbol{\sigma}_{\mathbf{c}}) \quad \text{s.t. } \mathbf{M}\boldsymbol{\tau} = \mathbf{c}, \quad \mathbf{l} \leq \boldsymbol{\tau} \leq \mathbf{u}.$$

The optimisation step is necessary because the number of MRIO elements by far exceeds the number of constraints, and there is not enough information to analytically solve the system for $\boldsymbol{\tau}$. The objectives 'best fulfils' and 'minimises departure' can be specified mathematically. Whilst there exists a plethora of optimisation approaches, the literature on input–output table estimation favours variants of the RAS iterative scaling method (Bacharach, 1970), and Quadratic Programming algorithms (van der Ploeg, 1988). These methods differ by the quantitative specification for penalties that are imposed for any departure of $\boldsymbol{\tau}$ from the initial data $\boldsymbol{\tau}^0$. This departure becomes necessary in order to fulfil the constraints $\mathbf{M}\boldsymbol{\tau} = \mathbf{c}$ and $\mathbf{l} \leq \boldsymbol{\tau} \leq \mathbf{u}$.

A detailed description of the data reconciliation process based on Geschke et al. (2011) is given in Appendix 1.

---

[4] A full list of exchange rates is available under http://unstats.un.org/unsd/snaama/dnlList.asp.

Assuming that constraint data points $\mathbf{c}_i$ are the means of normally distributed sets of observations, we estimate standard deviations of these constraints' data points. These estimations are based on published data or expert interviews, but mostly set according to certain world views on the uncertainty of various sets of constraints data. Generally, it can be found that smaller constraint data values are associated with higher relative standard deviations, and vice versa.

Second, a modified RAS optimisation algorithm is employed in order to fit standard deviations $\boldsymbol{\sigma}_{\boldsymbol{\tau}_j}$ to an error propagation formula $\boldsymbol{\sigma}_{c_i} = \sqrt{\sum_j (\mathbf{M}_{ij}\boldsymbol{\sigma}_{\boldsymbol{\tau}_j})^2}$. This procedure is consistent with the estimation of the (vectorised) MRIO $\boldsymbol{\tau}$, based on constraints data $\mathbf{c}$. In fact, the error propagation formula can be derived from the optimisation condition $\mathbf{M}\boldsymbol{\tau} = \mathbf{c}$. The $\boldsymbol{\sigma}_{\boldsymbol{\tau}}$ are influenced by two factors. The first is an uncertainty characteristic: the smaller the uncertainty $\boldsymbol{\sigma}_{\mathbf{c}_i}$ of a constraints data point $\mathbf{c}_i$, the smaller the uncertainty $\boldsymbol{\sigma}_{\boldsymbol{\tau}}$ of MRIO elements addressed by this constraints data item. The second is a data conflict characteristic: the pre-modified-RAS IE $\boldsymbol{\sigma}_{\boldsymbol{\tau}^0}$ of the $\boldsymbol{\sigma}_{\boldsymbol{\tau}}$ is set to the difference between the MRIO IE $\boldsymbol{\tau}^0$ and the MRIO final solution $\boldsymbol{\tau}$. This difference is influenced by the conflict in the constraint data, because conflicting constraint data lead to movements in elements during optimiser runs. For further details, see Lenzen et al. (2010b).

2.4. Data Preparation

2.4.1. Concordances and Preparation of the IE Data sets

In order to convert the Eora IE or EXIOBASE v1.0 IE given in Eora or EXIOBASE v1.0 classification into IE data sets given in EXIOBASE v2.0 classification, concordances $\mathbf{C}^{\text{Eora}}_{\text{pre}}$ and $\mathbf{C}^{\text{Eora}}_{\text{post}}$ must be be generated. Assume $\mathbf{T}^{\text{Eora}}_{\text{Eora Class}}$ is Eora's intermediate demand matrix in Eora classification, and $\mathbf{T}^{\text{Eora}}_{\text{EXIOBASE v2.0 Class}}$ is the matrix holding the same data following the conversion into EXIOBASE v2.0 classification. Then $\mathbf{C}^{\text{Eora}}_{\text{pre}}$ and $\mathbf{C}^{\text{Eora}}_{\text{post}}$ should operate as follows (assuming that the normalisation described in this section has been carried out):

$$\mathbf{T}^{\text{Eora}}_{\text{EXIOBASE v2.0 Class}} = \mathbf{C}^{\text{Eora}}_{\text{pre}}\mathbf{T}^{\text{Eora}}_{\text{Eora Class}}\mathbf{C}^{\text{Eora}}_{\text{post}}.$$

The concordances are constructed in two steps:

(1) For each country in Eora, construct the pre- and post-concordances between the Eora classification for the particular country and EXIOBASE v2.0's SUT structure. This is achieved using the harmonised system (HS) classification as a bridging classification. Concordances between Eora and HS as well as EXIOBASE v2.0 and HS are used to construct country-wise Eora-to-EXIOBASE v2.0 concordances.

(2) The country-specific concordances are then combined to form the full pre- and post concordances $\mathbf{C}^{\text{Eora}}_{\text{pre}}$ and $\mathbf{C}^{\text{Eora}}_{\text{post}}$. The placing of each individual country-specific concordance into $\mathbf{C}^{\text{Eora}}_{\text{pre}}$ and $\mathbf{C}^{\text{Eora}}_{\text{post}}$ then reflects which of Eora's countries are represented as individual regions in the EXIOBASE v2.0 classification (such as the EU-27 countries, China, or the USA), and which ones are summarised into broader regions (such as the majority of the African countries which are summarised in one of EXIOBASE v2.0's RoW-regions).

In a similar way, concordances are constructed to convert the final demand and value-added matrices.

Eora's IE for the year 2007 is, according to the Eora workflow, the final Eora model for the year 2006. Converting each sheet of this MRIO from Eora classification into the final classification can be achieved by applying pre- and post-maps $\mathbf{M}_{\text{pre}}^{\text{Eora}}$ and $\mathbf{M}_{\text{post}}^{\text{Eora}}$ such that

$$\mathbf{T}_{\text{EXIOBASE v2.0 Class}}^{\text{Eora}} = \mathbf{M}_{\text{pre}}^{\text{Eora}} \mathbf{T}_{\text{Eora Class}}^{\text{Eora}} \mathbf{M}_{\text{post}}^{\text{Eora}}.$$

The pre- and post-maps $\mathbf{M}_{\text{pre}}^{\text{Eora}}$ and $\mathbf{M}_{\text{post}}^{\text{Eora}}$ are the normalised concordances $\mathbf{C}_{\text{pre}}^{\text{Eora}}$ and $\mathbf{C}_{\text{post}}^{\text{Eora}}$ which were described at the beginning at this section. The normalisation was achieved by prorating the concordances according to a feasible proxy. The normalisation yields a map. The difference between a map and a concordance is that a concordance only contains 0 and 1 as values. If a sector in the source classification is disaggregated by a concordance, its values would be fully assigned to multiple sectors in the target classification, which in return leads to double counting. Hence, if a sector is to be disaggregated, then each affected transaction value must be distributed across all target sectors. A proxy vector determines the ratios between the different sectors of the target classification that are to be used for a disaggregation. A map is a normalised concordance which considers these ratios. Hence, a map may feature any given value between 0 and 1. By using a map for the conversion of Eora data into EXIOBASE v2.0 classification, the total sector outputs of Eora are preserved in the EXIOBASE v2.0 classification.

The proxy used for the normalisation of $\mathbf{C}_{\text{pre}}^{\text{Eora}}$ and $\mathbf{C}_{\text{post}}^{\text{Eora}}$ to obtain $\mathbf{M}_{\text{pre}}^{\text{Eora}}$ and $\mathbf{M}_{\text{post}}^{\text{Eora}}$ was the gross output vector of the final EXIOBASE v2.0 model (database CCC in Table 1), which was the only available proxy in the EXIOBASE v2.0 classification.

For the EXIOBASE v1.0 data set, the concept for the construction of the concordances is identical to the one for the Eora-to-EXIOBASE v2.0 concordances. Required are two (normalised) concordances such that

$$\mathbf{T}_{\text{EXIOBASE v2.0 Class}}^{\text{EXIOBASE v1.0}} = \mathbf{C}_{\text{pre}}^{\text{EXIOBASE v1.0}} \mathbf{T}_{\text{EXIOBASE v1.0 Class}}^{\text{EXIOBASE v1.0}} \mathbf{C}_{\text{post}}^{\text{EXIOBASE v1.0}}.$$

The construction of concordances to convert the EXIOBASE v1.0 data set from EXIOBASE v1.0 classification into an EXIOBASE v1.0 data set in EXIOBASE v2.0 classification was less complex than for the Eora data set.

EXIOBASE v1.0's first 43 regions are identical to those of EXIOBASE v2.0, and both EXIOBASE v1.0 and EXIOBASE v2.0 follow a strictly homogenous SUT structure. Hence, it is only required to construct one country-wise concordance in order to account for the different sector classifications of EXIOBASE v1.0 and EXIOBASE v2.0. This country-wise concordance is then be used for each of the first 43 countries in EXIOBASE v1.0.

The construction of the final part of the concordances presents a situation which has not been encountered during the construction of the Eora-to-EXIOBASE v2.0 concordances: regional disaggregations. EXIOBASE v1.0 only features one RoW-region, EXIOBASE v2.0 features five. Hence, EXIOBASE v1.0's RoW-region must be disaggregated to EXIOBASE v2.0's five RoW-regions.

The task of converting the EXIOBASE v1.0 IE in EXIOBASE v1.0 classification into an EXIOBASE v1.0 IE in EXIOBASE v2.0 classification is essentially identical to the conversion of the Eora IE in Eora classification into an Eora IE in EXIOBASE v2.0 classification. Following the preparation of the concordances $\mathbf{C}_{\text{pre}}^{\text{EXIOBASE v1.0}}$ and $\mathbf{C}_{\text{post}}^{\text{EXIOBASE v1.0}}$, these concordances are normalised using the same EXIOBASE v2.0 proxy (EXIOBASE v2.0's gross output vector) to obtain pre- and post maps. The conversion is then given by

$$\mathbf{T}^{\text{EXIOBASE v1.0}}_{\text{EXIOBASE v2.0 Class}} = \mathbf{M}^{\text{EXIOBASE v1.0}}_{\text{pre}} \mathbf{T}^{\text{EXIOBASE v1.0}}_{\text{EXIOBASE v1.0 Class}} \mathbf{M}^{\text{EXIOBASE v1.0}}_{\text{post}}.$$

The use of EXIOBASE v2.0's gross output vector as a proxy introduces a certain element of bias or interdependency into the pre- and post-maps. At this point, processed EXIOBASE v2.0 data are used to convert Eora data from Eora classification into EXIOBASE v2.0 classification. This is clearly not desirable, as it, at least theoretically, mixes Eora data and processed EXIOBASE v2.0 data. Implicitly, the converted Eora data set contains information obtained through EXIOBASE v2.0's data processing method. In order to avoid this problem, a different proxy vector in EXIOBASE v2.0 classification must be used. For example, the gross output of a fully assembled yet unprocessed EXIOBASE v2.0 IE in EXIOBASE v2.0 classification. Such an IE is not available, as the EXIOBASE v2.0 construction method does not include the construction of a fully populated IE. Hence, the EXIOBASE v2.0 gross output vector, which is used as a proxy in this case, is the only available proxy vector in EXIOBASE v2.0 classification. However, the converted Eora table still does not contain any processed EXIOBASE v2.0 data. The proxy vector is only used to split Eora data points across a number of sectors in the case of sectoral disaggregation. Hence, the bias caused by the use of this proxy vector is expected to be negligible.

The methodology used to construct the EXIOBASE v2.0 database did not follow the closed optimisation approach that is given in Equation 1. EXIOBASE v2.0 was constructed using a multi-stage process. A full single set of IE data (which is required for the reconciliation method used in this study) was not generated during the EXIOBASE v2.0 project.

In order to obtain an IE data set for the EXIOBASE v2.0 database that can be used for the reconciliation approach of Equation 1, we use available EXIOBASE v2.0 raw data for the domestic data, and initial trade data based on UN Comtrade data and UN main aggregates data. There was no initial estimate trade data for EXIOBASE v2.0 available. Both these data sets were originally misaligned, and therefore had to be converted into the final classification. The strategy that was used to construct the IE data is described in Andrew and Peters (2013). We refer to this strategy as the *A&P approach*.

Since this IE is not the same data set as the underlying data set of the EXIOBASE v2.0 model, it is not referred to as **C** in the list of different databases in Table 1. Instead, this IE data set is referred to as **G**. It was used for the databases 5 and 8 (short form GCE, GEE) in Table 1.

2.4.2. *Preparation of the Constraints' Data sets*

The EXIOBASE v2.0 constraints data set was constructed from various sets of source data and directly imported into AISHA.

This constraints data set can be broadly divided into three different types of constraints.

(1) *Balancing and boundary constraints.* These are the standard constraints to balance an MRIO table, and the definition of the boundaries for each element of the MRIO.

These constraints are independent of the actual source data for the EXIOBASE v2.0 constraints set.

(2) *SUT constraints data.* The SUT constraints data set is the most important of all source data sets for the EXIOBASE v2.0 constraints. The SUT data contain aggregated SUT tables for every region of the EXIOBASE v2.0 classification. Additionally, constraints data for final demand, value added, as well as import- and export totals are given. The source data are given in a 59-industry/59-product classification. This is the standard Eurostat table classification for EU countries, based on NACE.[5] All data of the SUT data set are used as so-called point constraints (as opposed to ratio constraints), meaning that the actual values are used as constraints. See Appendix 1, Equation A4 for an example on how data are used as point constraints.

(3) *Numerous source data sets on trade, product groups, and domestic SUT tables.* All remaining source data are implemented as ratio constraints. The difference between ratio constraints and point constraints is that ratio constraints do not constrain elements of the MRIO to equal a certain value, but merely to constrain a (sub-)set of values within the MRIO to be in the same ratio to one another as the source data points are. The easiest example would be a two-points ratio constraint defining that MRIO element $\boldsymbol{\tau}_i$ is for example twice a big as MRIO element $\boldsymbol{\tau}_j$, without defining the actual absolute values of neither $\boldsymbol{\tau}_i$ nor $\boldsymbol{\tau}_j$. See Appendix 1, Equation A7 for an example on how data are used as ratio constraints.

Generally speaking, using data as ratio constraints generates more non-zero coefficients in the constraints matrix **M** in Equation 1 compared to using data as points constraints. Since for the EXIOBASE v2.0 constraints data set, large amounts of source data were used as ratio constraints, the number of non-zero coefficients in **M** added up to approx. 20 billion. The implications of this large amount of data on the computational implementation of the reconciliation routine will be discussed in Section 2.5.

The Eora constraints data set was available in Eora's classification from the Eora database. Since Eora's sheets are identical to the sheets of the final classification, the conversion constraints data set in Eora classification into a constraints data set in EXIOBASE v2.0 classification requires the conversion of the regional and sectoral classification only, without altering the structure of the valuations. There are generally two ways to reconcile an IE in EXIOBASE v2.0 classification subject to the Eora constraints set:

(1) Convert the IE in EXIOBASE v2.0 classification into and IE in Eora classification, execute the reconciliation using a constraints data set in Eora classification, and then convert the result back into EXIOBASE v2.0 classification (two conversions required).

(2) Convert the constraints data set in Eora classification into a constraints data set in EXIOBASE v2.0 classification, and complete the reconciliation task in EXIOBASE v2.0 classification (one conversion required).

The first option is mathematically and computationally easier. The conversion of an IE in EXIOBASE v2.0 classification into an IE in Eora classification is similar to the conversions

[5] NACE is the *Statistical classification of economic activities in the European Community*. For more information on NACE, see http://epp.eurostat.ec.europa.eu/portal/page/portal/nace_rev2/introduction.

described in Section 2.4.1. But the first option requires two conversion: first the conversion of the data set from EXIOBASE v2.0 classification into Eora classification, and secondly the conversion of the result of the reconciliation process from Eora classification back into EXIOBASE v2.0 classification. In each conversion process, the normalisation of the concordance matrices introduces certain assumptions to the converted data set. This results in the loss of data accuracy. While the conversion of data sets was necessary in this study, the authors tried to minimise the number of conversions for each data set. Hence, option two was chosen for the reconciliation of data according to Eora constraints: the conversion of Eora's constraints data set in Eora classification into Eora's constraints data set in the final classification.

This conversion presents the most complex part of this study. The complexity arises from a combination of different factors. Firstly, the sheer amount of data that have to be processed in order to convert Eora's constraints matrix from Eora classification into the final classification. Secondly, the mathematical complexity of task. And finally, the implementation of an algorithm which carries out the conversion task in an acceptable time frame using the given computing resources. The basics of the mathematical concepts of this conversion will be discussed in this section and the computational challenges will be briefly discussed in Section 2.5.

Eora's constraints data set for the year 2007 is available from the Eora database. The data sources considered for this constraints data set are discussed in detail in Lenzen et al. (2012, 2013).

For this section, let $\mathbf{M}_{\text{Eora}}$ be Eora's constraint matrix in Eora classification, and let $\mathbf{M}_{\text{final}}$ be Eora's constraint matrix in the final classification.

According to Equation 1, the constraints matrix $\mathbf{M}_{\text{Eora}}$ (not to be confused with the pre- and post maps $\mathbf{M}_{\text{pre}}$ and $\mathbf{M}_{\text{post}}$ that were used earlier) operates on the vectorised representation $\boldsymbol{\tau}$ of an MRIO $\mathbf{T}$. The rows of the constraints matrix refer to the different constraints, the columns refer to the elements in the vectorised MRIO $\boldsymbol{\tau}$. Hence, a conversion of $\mathbf{M}_{\text{Eora}}$ to $\mathbf{M}_{\text{final}}$ only calls for a suitable post-map for $\mathbf{M}_{\text{Eora}}$. The number of rows in $\mathbf{M}_{\text{Eora}}$ and $\mathbf{M}_{\text{final}}$ is identical, since the number of constraints is invariant under a classification change.

Assume that $\mathbf{G}_{\text{M}}$ is the required post-map, then

$$\mathbf{M}_{\text{final}} = \mathbf{M}_{\text{Eora}}\mathbf{G}_{\text{M}} \tag{2}$$

holds.

The pre- and post concordances for the conversion of any data set in Eora classification into the final classification operate on the table format $\mathbf{T}$ of the MRIO. They do not operate on the vectorised format $\boldsymbol{\tau}$ that the constraints matrix $\mathbf{M}_{\text{Eora}}$ operates on. Hence, $\mathbf{G}_{\text{M}}$ must be constructed from these pre- and post concordances.

The issue of data disaggregation was already discussed in Section 2.4.1: if a data point within the Eora classification is disaggregated when converted into the final classification, the concordances have to be normalised in order to avoid double counting of this data point. For the constraints matrix, the issue of (dis-)aggregation and normalisation is the contrary: disaggregation does not require any normalisation, but aggregation does.

Consider two points $\boldsymbol{\tau}_1^{\text{Eora}}$ and $\boldsymbol{\tau}_2^{\text{Eora}}$ in the vectorised MRIO in Eora classification. Consider further the conversion in the final classification maps $\boldsymbol{\tau}_1^{\text{Eora}}$ onto two points, say $\boldsymbol{\tau}_1^{\text{final}}$ and $\boldsymbol{\tau}_2^{\text{final}}$, and the point $\boldsymbol{\tau}_2^{\text{Eora}}$ is mapped onto a single point, say $\boldsymbol{\tau}_3^{\text{final}}$. If this conversion is normalised, then $\boldsymbol{\tau}_1^{\text{Eora}} = \boldsymbol{\tau}_1^{\text{final}} + \boldsymbol{\tau}_2^{\text{final}}$ holds. For a constraint equation in Eora classification

$$m_1\boldsymbol{\tau}_1^{\text{Eora}} + m_2\boldsymbol{\tau}_2^{\text{Eora}} = c$$

the equivalent equation in the final classification is

$$m_1(\boldsymbol{\tau}_1^{\text{final}} + \boldsymbol{\tau}_2^{\text{final}}) + m_2\boldsymbol{\tau}_3^{\text{final}} = c.$$

This is simply

$$m_1\boldsymbol{\tau}_1^{\text{final}} + m_1\boldsymbol{\tau}_2^{\text{final}} + m_2\boldsymbol{\tau}_3^{\text{final}} = c.$$

Hence, in the case of disaggregation, the matrix coefficient for a particular element in the Eora classification is simply used for each of the corresponding elements in the final classification. No special treatment of the coefficients m_1 and m_2 is necessary to yield an equivalent constraints equation in the final classification.

In the case of aggregation, the situation is different. For this example, assume an MRIO

$$\mathbf{T} = \begin{pmatrix} t_{11} & t_{12} \\ t_{21} & t_{22} \\ t_{31} & t_{32} \end{pmatrix},$$

a row-wise vectorisation, i.e.

$$\boldsymbol{\tau} = \begin{pmatrix} t_{11} \\ t_{12} \\ t_{21} \\ t_{22} \\ t_{31} \\ t_{32} \end{pmatrix},$$

and the following constraints:

$$\begin{aligned} t_{21} + t_{22} &= 4, \\ t_{31} + t_{32} &= 5, \\ t_{11} + t_{12} &= 3. \end{aligned} \tag{3}$$

Hence, the constraints equation for this MRIO is

$$\underbrace{\begin{pmatrix} 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 & 0 & 0 \end{pmatrix}}_{=:\mathbf{M}_{\boldsymbol{\tau}}} \boldsymbol{\tau} = \underbrace{\begin{pmatrix} 4 \\ 5 \\ 3 \end{pmatrix}}_{=:\mathbf{c}}.$$

Further, consider the following pre-concordance for **T**

$$\mathbf{C}_{\text{pre}} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \end{pmatrix}.$$

Applying this pre-concordance to **T** aggregates the second and the third row as follows:

$$\mathbf{C}_{\text{pre}}\mathbf{T} = \underbrace{\begin{pmatrix} t_{11} & t_{12} \\ t_{21} + t_{31} & t_{22} + t_{32} \end{pmatrix}}_{=:\mathbf{F}}.$$

Using the same concept for the vectorisation of **F** that was used for the vectorisation of **T**, the vectorisation φ is

$$\varphi = \begin{pmatrix} t_{11} \\ t_{12} \\ t_{21} + t_{31} \\ t_{22} + t_{32} \end{pmatrix}.$$

Since the second and third rows of **T** are aggregated in the final MRIO **F**, the first and second equations of the constraints set in Equation 3 cannot be explicitly formulated for **F**. At this point, another assumption must be made: 'normalisations' of the sums $t_{21} + t_{31}$ and $t_{22} + t_{32}$ such that

$$\begin{aligned} p_1 \cdot (t_{21} + t_{31}) &= t_{21,} \\ (1 - p_1) \cdot (t_{21} + t_{31}) &= t_{31} \end{aligned} \tag{4}$$

for a $0 < p_1 < 1$ and

$$\begin{aligned} p_2 \cdot (t_{22} + t_{32}) &= t_{22,} \\ (1 - p_2) \cdot (t_{22} + t_{32}) &= t_{32} \end{aligned} \tag{5}$$

for a $0 < p_2 < 1$. Hence, a certain ratio between the elements in a sum must be assumed in order to be able to formulate the original constraints given in the first two equation of the constraints set in Equation 3. This strategy is similar to the disaggregation of MRIOs where a certain ratio between the different target sectors of the disaggregation is assumed. For this study, the authors chose the gross outputs of rows of **T** (in our case: Eora) as ratios for p_1 and p_2. This implies that $p_1 = p_2$. For this example, assume that $p_1 = p_2 = 0.3$. Then the constraint equations 3 can be reformulated for **F** as

$$\begin{aligned} 0.3 \cdot (t_{21} + t_{31}) + 0.3 \cdot (t_{22} + t_{32}) &= 4, \\ 0.7 \cdot (t_{21} + t_{31}) + 0.7 \cdot (t_{22} + t_{32}) &= 5, \\ t_{11} + t_{12} &= 3. \end{aligned} \tag{6}$$

These constraints can be reformulated as

$$\underbrace{\begin{pmatrix} 0 & 0 & 0.3 & 0.3 \\ 0 & 0 & 0.7 & 0.7 \\ 1 & 1 & 0 & 0 \end{pmatrix}}_{=:\mathbf{M}_\varphi} \varphi = \begin{pmatrix} 4 \\ 5 \\ 3 \end{pmatrix}.$$

Now that $\mathbf{M}_\tau$ and $\mathbf{M}_\varphi$ are both available (for the big picture: $\mathbf{M}_\tau$ refers to Eora, $\mathbf{M}_\varphi$ refers to EXIOBASE v2.0), $\mathbf{G_M}$ can be constructed. It is given by

$$\mathbf{G_M} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0.3 & 0 \\ 0 & 0 & 0 & 0.3 \\ 0 & 0 & 0.7 & 0 \\ 0 & 0 & 0 & 0.7 \end{pmatrix}. \tag{7}$$

And hence

$$\mathbf{M}_{\varphi} = \mathbf{M}_{\boldsymbol{\tau}} \mathbf{G}_{\mathbf{M}}$$

A similar procedure yields the matrix $\mathbf{G}_{\mathbf{M}}$ belonging to a post concordance, which is applied to the initial MRIO $\mathbf{T}$. Once both $\mathbf{G}_{\mathbf{M}}$ matrices have been constructed, they must be multiplied to form the final $\mathbf{G}_{\mathbf{M}}$ matrix.

As mentioned before, the shape of $\mathbf{G}_{\mathbf{M}}$ depends on the vectorisation of the MRIOs. A different vectorisation will yield a different row- and column order in $\mathbf{G}_{\mathbf{M}}$. To understand the structure of $\mathbf{G}_{\mathbf{M}}$ for this example, consider the 'normalised', transposed version of $\mathbf{C}_{\mathrm{pre}}$ (called $\mathbf{M}_{\mathrm{pre}}^{\mathrm{T}}$).

$$\mathbf{M}_{\mathrm{pre}}^{\mathrm{T}} = \begin{pmatrix} 1 & 0 \\ 0 & 0.3 \\ 0 & 0.7 \end{pmatrix}.$$

For this particular vectorisation of $\mathbf{T}$ and $\mathbf{F}$ vectorisation, $\mathbf{G}_{\mathbf{M}}$ of Equation 7 resembles a 'tiled' version of $\mathbf{M}_{\mathrm{pre}}^{\mathrm{T}}$ in a sense that the three different non-zero values of $\mathbf{M}_{\mathrm{pre}}^{\mathrm{T}}$ are represented in diagonal matrices of dimensions 2×2. These 2×2-tiles are arranged in the same structure within $\mathbf{G}_{\mathbf{M}}$ as the corresponding values in $\mathbf{M}_{\mathrm{pre}}^{\mathrm{T}}$. Further investigations of the structure of $\mathbf{G}_{\mathbf{M}}$ reveal that the tiles have the same number of rows as the pre concordance $\mathbf{C}_{\mathrm{pre}}$.

The structure of $\mathbf{G}_{\mathbf{M}}$ changes if $\mathbf{G}_{\mathbf{M}}$ is developed for a post map. In this case, no tiles appear in $\mathbf{G}_{\mathbf{M}}$. Instead, the matrix $\mathbf{M}_{\mathrm{pre}}^{\mathrm{T}}$ is replicated in a certain structure.

The authors considered constructing $\mathbf{G}_{\mathbf{M}}$ as a complete matrix, but the following two reasons prevented this.

(1) *Size of* $\boldsymbol{G_M}$. The vectorised version $\boldsymbol{\tau}$ of Eora is a vector containing 10^9 values. The vectorised version of an MRIO in the final classification is even bigger, containing approx. 1.2×10^9 values. Therefore, the matrix $\mathbf{M}_{\mathrm{Eora}}$ has approx. 10^9 columns, and a complete $\mathbf{G}_{\mathbf{M}}$ would have the dimensions $10^9 \times (1.2 \times 10^9)$. If $\mathbf{G}_{\mathbf{M}}$ was to be constructed, it would be saved in sparse format. But a priori estimations of the number of non-zero elements revealed that $\mathbf{G}_{\mathbf{M}}$ would contain approx. 5×10^{10} elements. This would require too much RAM for storing the matrix, let alone executing the operation given in Equation 2.

(2) *Structure of* $\boldsymbol{G_M}$. A general formula to construct $\mathbf{G}_{\mathbf{M}}$ from the pre- and post concordances $\mathbf{C}_{\mathrm{pre}}^{\mathrm{Eora}}$ and $\mathbf{C}_{\mathrm{post}}^{\mathrm{Eora}}$ is impossible since the structure of $\mathbf{G}_{\mathbf{M}}$ depends on the chosen vectorisation. AISHA uses different ways of vectorising the MRIO depending on what mode AISHA is running in. Hence, a unique formula for constructing $\mathbf{G}_{\mathbf{M}}$ does not exist.

Therefore, the concepts described in this section to convert Eora's constraint matrix in Eora classification into an Eora constraints matrix in the final classification were applied individually to each element of the constraints matrix $\mathbf{M}$.

2.4.3. Standard Deviations for the IE and Constraints' Sets

AISHA is able to consider the reliability of the input data during the reconciliation. Whereas the standard deviation data for the Eora IE and constraints data set are already estimated in the Eora database, the standard deviations for all other data have to be estimated.

The authors follow the general assumption that small values in a data set are less reliable than large values. Additionally, the constraints' data are always considered more reliable than the initial estimate data. For this project, the constraints' data receive standard deviation values that are three times smaller than those of the IE data sets, making the constraints' data three times more reliable in the reconciliation process. This is in general the same strategy that was chosen during the construction of the Eora database.

For more information on the estimation of data reliability for MRIO input data, refer to Lenzen et al. (2010b) and Wiedmann et al. (2010).

2.5. Data Preparation and Reconciliation

The calculations were carried out on two computers. The main computer featured 12 3.4 GHz Intel cores, 296 GB of fully shared RAM and 8TB of hard disc space. The second computer featured 8 2.4 GHz Intel cores with a total of 192 GB of fully shared RAM and 8TB of hard disc space. Both machines were fully utilised during this project. A total of approx. 5 TB of data were produced.

2.5.1. Converting Eora's Constraints Set into Final Classification

This was computationally the most resource- and time-demanding step in this study. As mentioned in Section 2.4.2, the matrix of $\mathbf{G}_{\mathbf{M}}$ was not explicitly calculated. Instead, the process of converting the constraints matrix into the final classification was carried out individually for each non-zero coefficient in the constraints matrix.

The algorithm performing this element-wise conversion was programmed in parallel and executed on the smaller of the two computers. Using seven of the eight available cores and approx. 160 GB of RAM, the total runtime for the conversion of $\mathbf{M}_{\mathrm{Eora}}$ into EXIOBASE v2.0 classification was approx. 80 h.

2.5.2. Database Reconciliation

Since all reconciliation tasks were carried out in the final classification, most of the individual parts of the input data for the reconciliation were identical in size (such as the IE, the boundary vectors or the standard deviation data sets for the IE), regardless of which database was to be reconciled. Only the constraints matrix $\mathbf{M}$ and the right-hand side vector $\mathbf{c}$ differ in size for the Eora constraints data set and the EXIOBASE v2.0 constraints data set, respectively.

The Eora constraints data set featured approx. 2,462,000 constraints and 5×10^9 non-zero elements in $\mathbf{M}$. The total amount of input data for the reconciliation algorithm when using the Eora constraints data set was approx. 100 GB. The EXIOBASE v2.0 constraints matrix had less constraints – approx. 907,000 – but more non-zero elements in $\mathbf{M}$: approx. 2×10^{10}. The reason for the high number of non-zero coefficients was the extensive use of ratio constraints, which need far more coefficients per constraint than point constraints. The total amount of input data for the reconciliation algorithm when using the EXIOBASE v2.0 constraints data set was more than 400 GB.

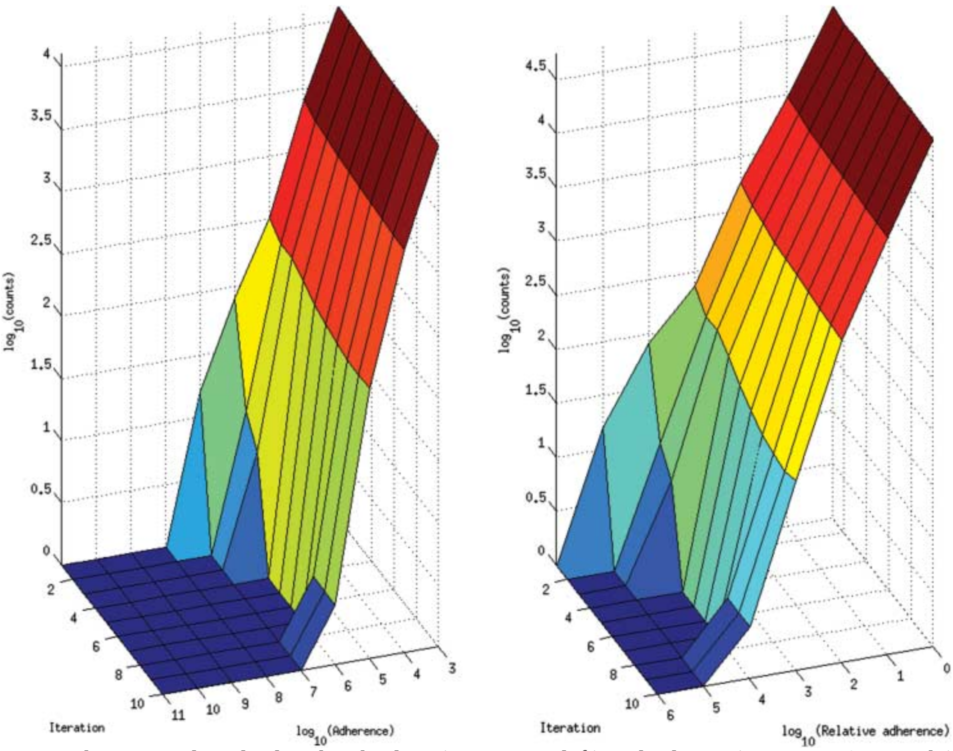
All reconciliation tasks were executed using a parallelised version of the KRAS algorithm (see Lenzen et al., 2010a for more information on KRAS). The run-time for each reconciliation on the larger computer with full utilisation of all available cores was approx. 24 h.

3. RESULTS

3.1. Convergence of the Reconciliation Problems

The mixing of the three different construction components results in feasible reconciliation problems in every case. A feasible problem is characterised by the fact that the reconciliation method can find a solution that considers all given constraints. If the

FIGURE 1. Convergence plot for the database XCE.



Notes: These two plots display the absolute ($\|\mathbf{M}\boldsymbol{\tau} - \mathbf{c}\|$, left) and relative ($\|\mathbf{M}\boldsymbol{\tau} - \mathbf{c}\|/\|\mathbf{c}\|$, right) constraints adherence for the first 10 iterations of the reconciliation algorithm. Small adherences mean that the constraints data **c** are well realised by the MRIO table $\boldsymbol{\tau}$. The *x*-axis (horizontal) shows the order of magnitude of the constraints adherence, the *z*-axis (vertical) shows the order of magnitude of the number of constraints that have the violation indicated on the *x*-axis, and finally, the *y*-axis shows the iteration number. The better the constraints are adhered to, the steeper the curve is (for each iteration), and the further to the right-hand side it is located . Both plots clearly show convergence, since the surface is steeper and further to the right as the iteration numbers increase. The slight increase in the number of constraints with a constraint adherence between 0 and 0.5 in the log-scale for iterations 9 and 10 is misleading. While the number of constraints showing small adherences increases, the residual (not shown in this plot) continues to decrease, and the algorithm keeps converging.

constraints data set contains a large number of conflicting constraints, this is not possible. In this case the reconciliation problem would be infeasible and the final database would violate some of the given constraints significantly. KRAS was designed to achieve convergence during the reconciliation process if even conflicting constraints are present in the constraints set. In order to achieve this, KRAS considers reliability information given for the right-hand side values of the constraints' equations. Based on the reliability of each constraint, KRAS finds the 'best compromise' for conflicting constraints. See Lenzen et al. (2010a) for more information on KRAS.

Using KRAS for the reconciliation, convergence was achieved for all databases that were constructed in this project (Figure 1).

3.2. Analysis and Comparison of the Different Databases

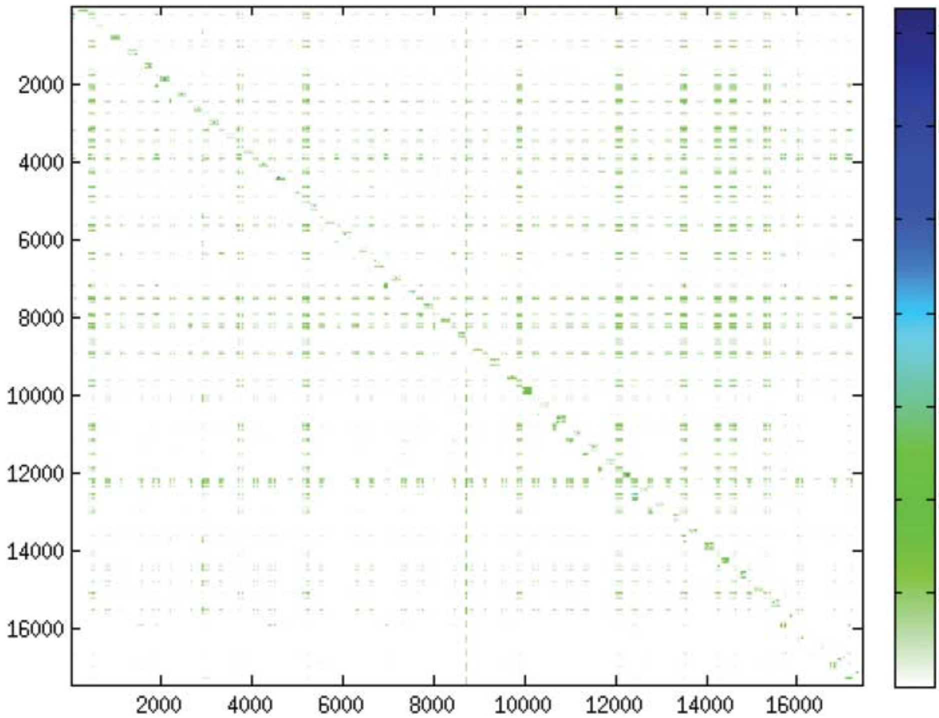
The results were analysed in two different ways:

(1) Each database is visualised in a heat map plot to verify that the overall structure does not contain any unexpected values (such as extremely large values in trade blocks or negative values in the intermediate demand block of the basic price sheet). Figure 2 shows the heat map of the basic price sheet of the intermediate transaction block for the database GCE. The heat map shows that the basic price sheet has a normal structure: The blocks on the main diagonal refer to the supply tables, which show, in general, higher values than other blocks in the sheet. Some regions show a higher import/export activity, indicated by strips of more prominent blocks in vertical and horizontal directions. There are no red dots in the plot, indicating that all values in the basic price sheet are non-negative (as expected).

(2) All possible pairings of the models listed in Table 1 are compared. Different measures for the comparison of input–output matrices have been discussed in the literature for well over 20 years (Harrigan et al., 1980; Knudsen and Fotheringham, 1986; Günlük-Şenesen and Bates, 1988; Gallego and Lenzen, 2006). Following these recommendations and analyses, we chose four different measures to evaluate actual metric distances between matrices, the 'goodness of fit' or correspondence between the entries of two matrices, and the correlation between the compared matrices. Suppose **A** and **B** are two matrices of equal size $n \times m$. The norms used for the comparison are:

Name	Abbreviation	Formula
Mean absolute difference	MAD	$\dfrac{\sum_{ij}(\lvert aij - b_{ij}\rvert)}{nm}$
Euclidean metric distance	EMD	$\dfrac{\sum_{ij}\sqrt{(aij - b_{ij})^2}}{nm}$
Regression coefficients	$1 - R^2$	$\dfrac{\text{SSE}}{\text{SST}}$ (see footnote [6])
Pearson's correlation coefficient	$1 - \text{CORR}$	$1 - \dfrac{\text{cov}(\mathbf{A}, \mathbf{B})}{\sigma_{\mathbf{A}}\sigma_{\mathbf{B}}}$

FIGURE 2. Heatmap of the intermediate demand block of database GCE.



Notes: The intermediate demand block has the dimensions $17,424 \times 17,424$, each value within this block is displayed as a pixel. The colour of the pixel depicts the order of magnitude of the absolute value of the corresponding value in the database. This plot displays the basic price intermediate demand block of the database GCE.

The matrix distance measures are applied to three different blocks of the basic price sheet: the intermediate demand block, the final demand block and the value-added block.

The original EXIOBASE v1.0 was included in these analyses. Note that, unlike all other databases considered in this project, EXIOBASE v1.0 was compiled for the year 2000 and not for the 2007.

Note that the regression coefficient and correlation coefficient are subtracted from 1 to match the concepts of MAD and EMD of displaying smaller values the more the investigated matrices are alike.

The results of the different metrics are displayed in Tables 2–5.

MAD and EMD provide measures of the distance between two different databases. As MAD is based on the 1-norm and EMD is based on the Euclidean norm, values for MAD are always larger than or equal to those for EMD.

[6] The term SSE denotes the *summed square of residuals*. The residual is the difference between two corresponding values in each of the compared databases. The term *SST* denotes the *total sum of squares*.

TABLE 2. Euclidean metric distance for the intermediate demand table in million Euros.

	GEE	XEE	ECE	GCE	XCE	EEE	CCC
XXX	0.19	0.009	0.01	2.25	0.049	0.008	0.007
CCC	0.197	0.011	0.009	2.29	0.05	0.007	
EEE	0.198	0.011	0.006	2.29	0.05		
XCE	0.203	0.049	0.046	2.24			
GCE	2.28	2.29	2.29				
ECE	0.199	0.010					
XEE	0.199						

Notes: The database pairs XXX/CCC, XXX/EEE, and CCC/EEE feature the lowest EMDs. For all pairings that do not include the databases which were based on the A&P-approach IE (GCE and GEE), the corresponding pairing shows a low value for EMD. Database pairings featuring the GCE or GEE databases have significantly larger values.

TABLE 3. Mean absolute difference for the intermediate demand table in million Euros.

	GEE	XEE	ECE	GCE	XCE	EEE	CCC
XXX	6.93	0.232	0.322	10.55	0.190	0.352	0.309
CCC	6.92	0.429	0.331	10.53	0.434	0.320	
EEE	6.96	0.446	0.119	10.55	0.454		
XCE	6.94	0.267	0.384	10.47			
GCE	9.76	10.61	10.57				
ECE	6.96	0.387					
XEE	6.91						

Notes: Similar to the results in Table 2, the MAD values are also the lowest for database pairings that do not include the GCE or GEE databases, and the highest for database pairings featuring the GCE or GEE databases.

But neither of these two measures can provide information about the goodness of fit or correlation between the two matrices.

The regression coefficient $1 - R^2$ provides a measure for the goodness of fit or correspondence between the values of two matrices. R^2 can only take values between 0 and 1. Generally speaking, the closer R^2 is to 1 (and the closer $1 - R^2$ is to 0, respectively), the better the entries of one matrix fit the entries of the other one. However, Knudsen and Fotheringham (1986) discuss that in some cases R^2 can yield relatively high values despite the fact that two matrices may differ substantially. Hence, the regression coefficient must always be interpreted together with the EMD/MAD results to assess the differences between two matrices accurately. Pearson's correlation coefficient gives information on the degree to which two different matrices are linearly correlated. In our case, the smaller the value $1 - \text{CORR}$ is, the higher the degree of correlation. Theoretically, the value of CORR can also become negative, resulting in values larger than 1 for $1 - \text{CORR}$. This would indicate that the corresponding databases are negatively linearly correlated. This behaviour has not been observed in any combination of databases and would indicate that at least one of the databases contains significant errors. Similar to EMD/MAD and $1 - R^2$, the correlation coefficient cannot stand by itself as an accurate measure for the comparison of two matrices. Consider the vectors $\mathbf{a} = [1, 2, 3, 4, 5]^{\mathrm{T}}$

TABLE 4. Regression coefficients $1 - R^2$ for the intermediate demand table.

	GEE	XEE	ECE	GCE	XCE	EEE	CCC
XXX	0.064	0.597	0.525	0.002	0.048	0.676	0.779
CCC	0.103	0.443	0.618	0.004	0.030	0.786	
EEE	0.074	0.369	0.800	0.003	0.026		
XCE	0.010	0.061	0.262	0.921			
GCE	0.013	0.003	0.186				
ECE	0.067	0.285					
XEE	0.098						

Notes: The largest values for the regression coefficient belong to those database pairings that were amongst those with the smallest values for EMD and MAD: XXX/CCC, XXX/EEE, and EEE/CCC. All of these model pairings were constructed according to different constraints' sets. For other pairings, databases with identical constraints' sets often yield small values for $1 - R^2$ (such as CCC/XCE, GCE/ECE, or EEE/GEE).

TABLE 5. Correlation coefficients 1 − CORR for the intermediate demand table.

	GEE	XEE	ECE	GCE	XCE	EEE	CCC
XXX	0.746	0.227	0.274	0.949	0.779	0.177	0.116
CCC	0.679	0.334	0.213	0.935	0.826	0.113	
EEE	0.727	0.397	0.105	0.941	0.838		
XCE	0.899	0.751	0.487	0.039			
GCE	0.883	0.943	0.568				
ECE	0.740	0.456					
XEE	0.686						

Notes: The three pairings of those databases that were not constructed in this project (XXX/EEE, CCC/EEE, and XXX/CCC) have much smaller values (i.e. higher correlation) than most other pairings.

and $\mathbf{b} = [1,000,\ 1,001,\ 1,002,\ 1,004,\ 1,005]^{\mathrm{T}}$. Then the correlation between these two vectors equals 1 (i.e. total correlation), despite the fact that **a** and **b** differ substantially.

The tables for the final demand and value-added blocks are given in Appendix 1.

3.3. Discussion

Tables 2 and 3 show that the smallest values for MAD and EMD are achieved if both databases are based on similar IEs. The database pairing CCC/XXX, where both databases were constructed according to the same methodology, displays among the smallest values for MAD and EMD. For the models XXX, CCC and EEE, a lot of effort was put into the construction of the IE. The result of these efforts is that all three IEs already provide a good approximation of the final MRIO. Hence, the final MRIO is probably relatively close to the IEs. This is reflected by small values for MAD and EMD between all database combinations for which both databases are based on the EXIOBASE v1.0, EXIOBASE v2.0 or Eora IE. This indicates that neither of these IEs had to be adjusted substantially during the reconciliation process, regardless of which constraints data set or reconciliation method was used.

The A&P-approach IE (G) that was constructed as part of this study did not undergo such a detailed data mining and construction process. Tables 2 and 3 clearly show that those databases that were based on the A&P-approach IE are not comparable to the other databases. Depending on the distance measure that is applied, the differences in the values for MAD and EMD are around two orders of magnitude larger when compared with databases that use the EXIOBASE v1.0, EXIOBASE v2.0, or Eora IE. Additionally, the comparison of the databases GCE and GEE shows that although both of these databases are based on the same IE, the values for EMD and MAD for this pairing are still much larger than other for database combinations. This indicates that during the reconciliation process, the imposed constraints cause the reconciliation engine to calculate a result which has departed significantly from the initial data set. Both the Eora and the EXIOBASE v2.0 constraints' data sets were constructed with great effort, and it is obvious that the constraints' sets contain information that conflict significantly with the low-effort IE 'G'. These conflicts are much smaller when the Eora-, EXIOBASE v2.0- or EXIOBASE v1.0 IE is used. Hence, the construction of a meaningful IE has a significant effect on the final model.

The performance of the Eora-style reconciliation process compared with the EXIOBASE v2.0- and/or EXIOBASE v1.0 reconciliation process cannot be directly assessed. Due to the nature of the EXIOBASE v2.0 construction process (which does not require a fully populated IE), it was impossible to compare the Eora construction method and the EXIOBASE v2.0 construction method for two identical IE and subject to the same constraints data set. A 'CCE'-model could not be constructed during this study because EXIOBASE v2.0 was constructed in a multi-stage process that is not based on a fully constructed IE. Hence, a EXIOBASE v2.0 IE data set was not available. The A&P-approach, which was used to construct a fully populated IE based on EXIOBASE v2.0 raw data, is similar to the approach that is used to construct the IE in EXIOBASE's original multi-stage. Analysis of the databases that were constructed based on this IE (GCE and GEE) shows that reconciling this IE according to the subject to the Eora or EXIOBASE v2.0 data set yields significant differences in the resulting databases. Both the EMD and the MAD values for the pairing GCE/GEE are amongst the highest. In fact, every pairing that features the GCE or GEE databases has significantly higher values for EMD and MAD than every other pairing. Initially, it is unclear if these discrepancies are due to the IE or constraints' data sets, or the construction methods of the individual databases. An indirect comparison of the different construction methods of Eora and EXIOBASE v2.0 shows the following: the database pairing XXX/CCC shows similar values for EMD and MAD as the combinations EEE/CCC and EEE/XXX do for these metrics. Both Eora's IE and EXIOBASE v1.0 data sets (used as an IE in this study) are reconciled with AISHA according to the Eora constraints set (databases EEE and XEE) and the EXIOBASE v2.0 constraints set (databases ECE and XCE). Comparing these databases with the original EXIOBASE v2.0 model (combinations XCE/CCC and ECE/CCC, as well as XEE/CCC and EEE/CCC) reveals that the reconciliation behaves in similar ways on both initial data sets, independent of the constraints data set that was chosen. Hence, AISHA's performance is comparable when applied to different data sets.

Table 4 reveals that if a small Euclidean norm is calculated for a pairing of two databases, this does not necessarily imply that the corresponding regression coefficient $1 - R^2$ is small as well. The database pairings CCC/XXX, EEE/CCC and XXX/EEE have amongst the highest values for $1 - R^2$, despite having low values for MAD and EMD. Two of the databases that were reconciled using the EXIOBASE v2.0 constraints data set (GCE, and XCE) show a small value for $1 - R^2$ when compared with the original EXIOBASE v2.0 database CCC. The same holds for the databases that were reconciled using the Eora constraints data set (XEE and GEE) when compared with the original Eora data set EEE. This indicates that if two databases are reconciled subject to the same constraints data set, the pairing of these two databases tends to have a lower value for $1 - R^2$. Also, this observation is independent of the reconciliation method that was used. However, the impact of the constraints data set on $1 - R^2$ is not as obvious as the impact of the IE on MAD and EMD.

Finally, the analysis of the correlation coefficients (Table 5) shows that the three original databases (XXX, EEE and CCC) are highly correlated to one another. For the databases that were calculated during this study, the higher the correlation, the more the three construction components are alike. For example, the discussion of Tables 2 and 3 already showed that the comparison of databases based on the 'G' IE with databases that are based on different IEs (E, C or X) yields very large values for both MAD and EMD. The same pair-wise combinations of databases show poor correlation in Table 5. An example of a high correlation is the database combination XEE/EEE. This combination shows small matrix distances for both norms, and relatively good goodness of fit in Table 4. The database pairing XEE/EEE also shows a relatively high correlation.

4. CONCLUSIONS AND OUTLOOK

In this study we compared different approaches for harmonising global MRIO databases, notably the multi-stage process of the EXIOBASE v1.0 and EXIOBASE v2.0 MRIO databases, and the single automated reconciliation step of the Eora database. In a number of numerical experiments we investigated the suitability of the software suite AISHA, initially developed for the construction of Eora, for processing the EXIOBASE v2.0 and EXIOBASE v1.0 data sets. These experiments allowed us to examine the effects of the different construction methods and input data sets on the final MRIO databases. Tables 2 and 3 give a clear indication that a well-constructed IE has a significant impact on the quality of the final MRIO database. Since AISHA requires a fully constructed IE for only one year of a time series (the so-called base year), the researcher must carefully choose a year for the IE in which enough meaningful data are available to support the construction of an IE. During the construction of the Eora database, the year 2000 was identified as the most suitable year for the IE, and data from a large number of sources were used to construct Eora IE for the year 2000 (see Lenzen et al., 2012 for details).

The impact of the constraints data is less obvious at first. This is due to a number of reasons.

Firstly, any impact that the constraints data set has on MAD or EMD would have been overshadowed by the far more dominant impact of the IE. The other metrics that were used – the regression coefficient and Pearson's correlation coefficient – can provide misleading results. Knudsen and Fotheringham (1986) describe cases for which the regression coefficient does not accurately reflect the 'goodness of fit' of two compared data sets. Numerical tests during this study showed that Pearson's correlation coefficient can react quite sensitively to outliers. Hence, the regression coefficient and the correlation coefficient reflect phenomena in the MRIOs that are not only caused by data given in the constraints' data sets. However, the results indicate that the use of a certain constraints data set is reflected in the regression coefficient. This supports previous research results. For example, Lenzen et al. (2006) found that using a large number of accurate constraints addressing the majority of MRIO elements ensures good results during the reconciliation process. Finally, the comparison of the correlation between the databases shows that higher the correlation achieved between two databases, the more both underlying data sets – the IE and the constraints data set – are alike.

We conclude that if the IE and constraints data set are both well constructed, AISHA can construct MRIO databases that match the quality of databases that were constructed according to different strategies, such as EXIOBASE v1.0 and EXIOBASE v2.0.

The good convergence behaviour together with the analyses of the final MRIO databases means that it is viable to produce a future time series of the EXIOBASE v2.0 MRIO database using Eora's automated reconciliation method. It is however necessary to construct a fully populated IE for the first year of the time series. The construction of the IE is not automated within the AISHA tool. Therefore, the benefits of AISHA's highly automated construction workflow can only be utilised once a (potentially labour-intensive) construction of a suitable IE has been completed.

Large-scale MRIO tables are essential for on-going assessment of environmental impacts on a global scale. In order to fulfil the requirements of such applications in the future, global MRIO tables must ideally become less expensive to compile, more frequently published with more timely data, and more accurate and comprehensive. Existing databases often face the obstacle of high production costs, infrequent publication intervals and/or they produce inaccurate results. Ideally, one would borrow the best attributes from different MRIO initiatives. In this study we showed that mixing construction components of different MRIO databases results in MRIO databases of comparable quality to the original ones. More importantly, we have shown that a highly automated workflow in the production process is able to approximate MRIO databases that were constructed in more labour-intensive processes. Further research has to be undertaken in order to assess to what degree an automated construction process can replicate the processes used for more labour-intensive MRIO databases, and how elements of other construction processes can be integrated into AISHA in order to assist this.

References

Andrew, R.M. and G.P. Peters (2013) A Multi-Regional Input-Output Table Based on the Global Trade Analysis Project Database (GTAP-MRIO). *Economic Systems Research*, 25, 99–121.

Andrew, R., J. Lennox and G. Peters (2010) Greenhouse Gas Emissions Embodied in New Zealand's Trade: Methodology (Paper prepared for the 18th International Input-Output Conference).

Bacharach, M. (1965) Estimating Nonnegative Matrixes From Marginal Data. *International Economic Review*, 6, 294–310.

Bacharach, M. (1970) *Biproportional Matrices & Input-Output Change*. Cambridge, Cambridge University Press.

Bouwmeester, M. and J. Oosterhaven (2008) Methodology for the Construction of an International Supply-Use Table (Paper prepared for the Intermediate Input–Output Meeting, Sevilla, Spain).

Daniels, P.L., M. Lenzen and S.J. Kenway (2011) The Ins and Outs of Water Use – A Review of Multi-Region Input–Output Analysis and Water Footprints for Regional Sustainability Analysis and Policy. *Economic Systems Research*, 23, 353–370.

Deming, W. and F. Stephan (1940) On Least-Squares Adjustments of a Sampled Frequency Table when the Expected Marginal Total are Known. *Annals of Mathematical Statistics*, 11, 427–444.

Ewing, B.R., T.R. Hawkins, T.O. Wiedmann, A. Galli, A. Ertug Ercin, J. Weinzettel and K. Steen-Olsen (2012) Integrating Ecological and Water Footprint Accounting in a Multi-Regional Input-Output Framework. *Ecological Indicators*, 23, 1–8.

Feng, K.A., A. Chapagain, S. Suh, S. Pfister and K. Hubacek (2011) Comparison of Bottom-Up and Top-Down Approaches to Calculating the Water Footprints of Nations. *Economic Systems Research*, 23, 371–385.

Friedlander, D. (1961) A Technique for Estimating a Contingency Table, Given the Marginal Totals and Some Supplementary Data. *Journal of the Royal Statistical Society, Series A*, 124, 412–420.

Gallego, B. and M. Lenzen (2006) Estimating Generalised Regional Input-Output Systems: A Case Study of Australia. In: M. Ruth and B. Davíðsdóttir (eds.) *The Dynamics of Regions and Networks in Industrial Ecosystems*. chapter 5. Cheltenham, UK, Edward Elgar Publishing, 55–77.

Galli, A., J. Weinzettel, G. Cranston and E. Ercin (2013) A Footprint Family Extended MRIO Model to Support Europe's Transition to a One Planet Economy. *Science of the Total Environment* 461–462, 813–818.

Geschke, A., M. Lenzen, K. Kanemoto and D. Moran (2011) AISHA: A Tool to Construct a Series of Contingency Tables (Paper prepared for the 19th International Input–Output Conference, Alexandria, VA, USA, June 2011).

Glen, P., R. Andrew and J. Lennox (2011) Constructing an Environmentally-Extended Multi-Region Input-Output Table Using the GTAP Database. *Economic Systems Research*, 23, 131–152.

Günlük-Şenesen, G. and J.M. Bates (1988) Some Experiments with Methods of Adjusting Unbalanced Data Matrices. *Journal of the Royal Statistical Society*, 151, 473–490.

Harrigan, F.J., J.W. Mcgilvray and I.H. Mcnicoll (1980) Simulating the Structure of a Regional Economy. *Environment and Planning A*, 12, 927–936.

Knudsen, D.C. and A.S. Fotheringham (1986) Matrix Comparison, Goodness-of-Fit, and Spatial Interaction Modeling. *International Regional Science Review*, 10, 127–147.

Kronenberg, T. (2009) Construction of Regional Input-Output Tables Using Non-Survey Methods - The Role of Cross-Hauling. *International Regional Science Review*, 32, 40–64.

Lenzen, M., B. Gallego and R. Wood (2006) A Flexible Approach to Matrix Balancing Under Partial Information. *Journal of Applied Input-Output Analysis*, 11 & 12, 1–24.

Lenzen, M., B. Gallego and R. Wood (2010a) Matrix Balancing Under Conflicting Information. *Economic Systems Research*, 21(1), 23–44.

Lenzen, M., R. Wood and T. Wiedmann (2010b) Uncertainty Analysis for Multi-Region Input-Output models – A Case Study of the UK's Carbon Footprint. *Economic Systems Research*, 22, 43–63.

Lenzen, M., K. Kanemoto, D. Moran and A. Geschke (2012) Mapping the Structure of the World Economy. *Environmental Science and Technology*, 46, 8374–8381.

Lenzen, M., D. Moran, K. Kanemoto and A. Geschke (2013) Building Eora: A Global Multi-Region Input-Output Database at High Country and Sector Resolution. *Economic Systems Research*, 25, 20–49.

Liu, J., R. Liu and L. Yao (2012) Definition and Research Progress of Sustainable Consumption: From Industrial Ecology View. *Shengtai Xuebao/Acta Ecologica Sinica 32*, 20, 6553–6557.

Moran, D. (2013) The Eora MRIO. In: J. Murray and M. Lenzen (eds.) *The Sustainability Practitioner's Guide to Multi-Regional Input–Output Analysis*. Urbana-Champaign, IL, Common Ground Publishing, 65–70.

Oosterhaven, J., D. Stelder and S. Inomata (2008) Estimating International Interindustry Linkages: Non-Survey Simulations of the Asian-Pacific Economy. *Economic Systems Research*, 20, 395–414.

Owen, A. and J. Barrett (2013) Uncertainties in Environmentally Extended MRIO Tables Arising from Assumptions Made in their Construction and the Effect on their Usefulness in Climate Policy (Paper prepared for the 20th International Input-Output Conference).

Peters, G., S. Davis and R. Andrew (2012) A Synthesis of Carbon in International Trade. *Biogeosciences*, 9, 3247–3276.

van der Ploeg, F. (1988) Balancing Large Systems of National Accounts. *Computer Science in Economics and Management*, 1, 31–39.

Skelton, A., D. Guan, G.P. Peters and D. Crawford-Brown (2011) Mapping Flows of Embodied Emissions in the Global Production System. *Environmental Science and Technology*, 45, 10516–10523.

Steen-Olsen, K., A. Owen, E.G. Hertwich and M. Lenzen (2014). Effects of Sector Aggretation on CO_2 Multipliers in Multiregional Input–Output Analyses. *Economic Systems Research*, 1–19. doi:10.1080/09535314.2014.934325

Stephan, F. (1942) An Iterative Method of Adjusting Sample Frequency Tables when Expected Marginal Totals are Known. *Annals of Mathematical Statistics*, 13, 166–178.

Su, B. and B.W. Ang (2011) Multi-Region Input-Output Analysis of CO2 Emissions Embodied in Trade: The Feedback Effects. *Ecological Economics*, 71, 42–53.

The Bookings Institution (2013) A Fair Compromise to Break the Climate Impasse. Accessed April 26. http://www.brookings.edu/research/papers/2013/04/04-climate-emissions-grasso-roberts.

Tohmo, T. (2004) New Developments in the Use of Location Quotients to Estimate Regional Input–Output Coefficients and Multipliers. *Regional Studies*, 38, 43–54.

Tukker, A. (2013) EXIOBASE – A Detailed Multi-Regional Supply and Use Table with Environmental Extensions. In: J. Murray and M. Lenzen (eds.) *The Sustainability Practitioner's Guide to Multi-Regional Input-Output Analysis*. Urbana-Champaign, IL, Common Ground Publishing, 54–64.

Tukker, A., A. de Koning, R. Wood, T. Hawkins, S. Lutter, J. Acosta, J. Rueda Cantuche, M. Bouwmeester, J. Oosterhaven, T. Drosdowski and J. Kuenen (2013a) EXIOPOL – Development and Illustrative Analyses of a Detailed Global MR EE SUT/IOT. *Economic Systems Research*, 25, 50–70.

Tukker, A., A. de Koning, R. Wood, T. Hawkins, S. Lutter, J. Acosta, J. M.R. Cantuche, M. Bouwmeester, J. Oosterhaven, T. Drosdowski and J. Kuenen (2013b) EXIOPOL – Development and Illustrative Analyses for a Detailed Global MR EE SUT/EE SUT/IOT. *Economic Systems Research*, 25, 927–936.
Wiedmann, T. and J. Barrett (2013) Policy-Relevant Applications of Environmentally Extended MRIO Databases – Experiences from the UK. *Economic Systems Research*, 25, 143–156.
Wiedmann, T., R. Wood, M. Lenzen, J. Minx, D. Guan and J. Barrett (2010) The Carbon Footprint of the UK – Results from a Multi-Region Input-Output Model. *Economic Systems Research*, 22, 19–42.
Wilting, H.C. (2012) Sensitivity and Uncertainty Analysis in MRIO Modelling; Some Empirical Results with Regard to the Dutch Carbon Footprint. *Economic Systems Research*, 24, 141–171.

APPENDIX 1. THE EXIOBASE DATABASE IN SHORT

EXIOBASE is a Multi-Regional EE SUT based on the based on the input–output framework described in the well-established System of National Accounts, and includes environmental extensions already consistent with the SEEA. EXIOBASE comes in two versions:

- EXIOBASE 1.0. EXIOBASE v1.0 provides trade linked supply use tables for 43 countries and a rest of world, with a resolution of 129 sectors and products and a large number of environmental extensions. EXIOBASE v1.0 is available for the year 2000.
- EXIOBASE 2.0. EXIOBASE v2.0 provides trade linked supply use tables for 43 countries and five rest of world regions. EXIOBASE v2.0 is available for the year 2007.

EXIOBASE is constructed in the following steps:

(1) Creating harmonised EE SUT for individual countries. From available country SUT or IOT first SUT in basic prices and valuation layers are constructed, with a Use table split into a domestic and import part. This SUT is subsequently detailed with auxiliary data to the 129 sector and product resolution. Resource uses available from the SERI database (in turn based on sources such as FAO, USGS and others) and water and land use are allocated to the right sectors. Activity variables (IEA energy uses and other) are allocated to the right sectors, and using emission factors from various sources, consolidated in TNO’s TEAM model, are used to estimate emissions by sector.
(2) Transforming the MR EE SUT into an industry by industry and a product by product MR EE IOT.

The result is in essence the following data sets: a set of harmonised EE SUT for individual countries, a global MR EE SUT, and two types (industry by industry and product by product) of global MR EE IOT. For more information, refer to Tukker et al. (2013a) and Tukker (2013).

APPENDIX 2. THE EORA DATABASE IN SHORT

Eora is an MRIO framework featuring 187 individual countries. Eora provides fully trade-linked tables for the years 1990–2011, and features five valuations. Unlike the EXIOBASE data sets, Eora features a heterogeneous classification and features IIOTs, CIOT, and SUTs. During the construction of Eora, emphasis was placed on representing each country in its native classification, which led to its heterogeneous table structure. During the construction of Eora, a large variety of data sources were considered. These included data published by each national statistical agencies, UN official country data, UN Comtrade data, Eurostat data and data from other sources.

The construction of the Eora can be summarised in the following methodology.

(1) Choose a base-year (for Eora, this is the year 2000), assemble an IE from all available source. Use all remaining source data to construct the constraints matrix, then reconcile the database according the reconciliation problem given in Equation 1.
(2) Loop forward: For the following year, construct the constraints matrix from the available source data, use the final result of the previous year as IE, and reconcile the database using Equation 1.

This methodology is also used for the years prior to the base year 2000 by looping backwards. For more information, refer to Lenzen et al. (2012, 2013) and Moran (2013).

APPENDIX 3. METHODOLOGY FOR DATA RECONCILIATION

The harmonisation of the different MRIO data sets will be conducted using the AISHA tool (Geschke et al., 2011) which was developed by the University of Sydney. Within AISHA, the entire MRIO tables is reconciled according to previously defined constraints. The mathematical background of this concept is explained in this section. Let

$$\mathbf{T} = \begin{pmatrix} t_{11} & t_{12} & t_{13} \\ t_{21} & t_{22} & t_{23} \\ t_{31} & t_{32} & t_{33} \end{pmatrix}. \tag{A1}$$

Perhaps the most important constraints on MRIOs are the so-called *balancing constraints*. For the MRIO table $\mathbf{T}$ introduced in Equation A1 the balancing constraints are given (assuming that net taxes on products being included in value added) by a set of equations

$$\begin{aligned} t_{11} + t_{12} + t_{13} &= t_{11} + t_{21} + t_{31}, \\ t_{21} + t_{22} + t_{23} &= t_{12} + t_{22} + t_{32}, \\ t_{31} + t_{32} + t_{33} &= t_{13} + t_{23} + t_{33}, \end{aligned}$$

which is equivalent to

$$\begin{aligned} t_{11} + t_{12} + t_{13} - (t_{11} + t_{21} + t_{31}) &= 0 \\ t_{21} + t_{22} + t_{23} - (t_{12} + t_{22} + t_{32}) &= 0 \\ t_{31} + t_{32} + t_{33} - (t_{13} + t_{23} + t_{33}) &= 0. \end{aligned}$$

Since each of the diagonal elements t_{ii} for $i = 1, 2, 3$ appears twice with alternating signs in the corresponding equation, these equations simplify to

$$\begin{aligned} t_{12} + t_{13} - t_{21} - t_{31} &= 0, \\ t_{21} + t_{23} - t_{12} - t_{32} &= 0, \\ t_{31} + t_{32} - t_{13} - t_{23} &= 0. \end{aligned} \tag{A2}$$

Equation A2 are the mathematical representation of the condition that the sums of elements in the ith column must equal the sums over the elements of the ith row.

Similarly, other requirements can be expressed using the same notation. IO tables also have the condition that the sum of each column in the margins matrices (valuations 2 and 3 in our example) must equal 0. The resulting constraint equation is

$$t_{1i} + t_{2i} + t_{3i} = 0 \quad \forall\, i, \tag{A3}$$

where i denotes the column (destination) index in the margins tables.

In balancing constraints, the right-hand side values of the constraint equations (e.g. Equations A2 and A3) are 0. Other constraints might have different right-hand side values. In many examples of IO table estimation, information is available on the gross output of sectors (Deming and Stephan, 1940; Stephan, 1942; Friedlander, 1961; Bacharach, 1965; Tohmo, 2004; Gallego and Lenzen, 2006; Kronenberg, 2009). Suppose the gross output for sector *Primary, manufacturing, and utilities* is given by a monetary value c_1, then the corresponding constraint equation (sum over all elements of sector *Primary, manufacturing, and utilities* equals c_1) is

$$t_{11} + t_{12} + t_{13} = c_1. \tag{A4}$$

The vectorisation of $\mathbf{T}$ becomes necessary in order to represent the Equations A2 and A4 in a matrix-by-vector notation. Assume the following vectorisation $\boldsymbol{\tau}$ of the table $\mathbf{T}$

$$\boldsymbol{\tau} = \begin{pmatrix} t_{11} \\ t_{12} \\ t_{13} \\ t_{21} \\ t_{22} \\ t_{23} \\ t_{31} \\ t_{32} \\ t_{33} \end{pmatrix}. \tag{A5}$$

This vecorisation sorts the elements of $\mathbf{T}$ row-wise underneath one another. Using this vectorisation, the balancing constraints given in Equations A2 can be expressed in the following matrix-by-vector notation.

$$\underbrace{\begin{pmatrix} 0 & 1 & 1 & -1 & 0 & 0 & -1 & 0 & 0 \\ 0 & -1 & 0 & 1 & 0 & 1 & 0 & -1 & 0 \\ 0 & 0 & -1 & 0 & 0 & -1 & 1 & 1 & 0 \end{pmatrix}}_{=\mathbf{M}} \underbrace{\begin{pmatrix} t_{11} \\ t_{12} \\ t_{13} \\ t_{21} \\ t_{22} \\ t_{23} \\ t_{31} \\ t_{32} \\ t_{33} \end{pmatrix}}_{\boldsymbol{\tau}} = \underbrace{\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}}_{=\mathbf{c}}. \tag{A6}$$

This format already has the desired form of $\mathbf{M}\boldsymbol{\tau} = \mathbf{c}$. Additional constraints, for example those given in Equation A4, would simply be added to this set of equations by adding additional lines to the matrix $\mathbf{M}$ and the vector $\mathbf{c}$.

Another important class of constraints is *ratio constraints*, which considers known ratios amongst table elements. For example, Andrew et al. (2010) applied ratio constraints during the estimation of an MRIO table centred on New Zealand (R. Andrew, personal communication, April 2011). More specifically, he imposed that the production structure of sectors in a table $\mathbf{T}$ to be constructed, should not deviate from the production structure in a known table $\mathbf{T}^*$. Assume the production structure of an economy is represented by a transaction table $\mathbf{T}$, the diagonalised gross output vector as $\hat{\mathbf{v}}$. Then the mathematical formulation of this concept is given by

$$\mathbf{A} = \mathbf{T}\hat{\mathbf{v}}^{-1}. \tag{A7}$$

The elements a_{ij} are interpreted as the amount of input that sector j requires from sector i per unit of j's gross output.

Virtually every constraint considered within this document is a combination of the basic constraints examples presented in this section. Combining these basic structures can result in constraints of almost arbitrary complexity, especially for large MRIO tables, and examples can be found throughout this document.

Finally, MRIO tables may also be subject to boundary constraints. Unlike the constraints discussed so far (which were all equality constraints), boundary constraints are inequality constraints that define a lower or upper boundary for elements of the MRIO tables. The most common boundary constraints in MRIO tables are restrictions on certain parts of the different valuation sheets. For example, basic-price intermediate demand should only have positive values. A subsidies sheets (our example's sheet 5) must only have non-positive values. Finally, changes in inventories are a component of final demand, that do not have any boundary restrictions because sectors in an economy can in subsequent years either add to, or draw from stocks, resulting in positive or negative changes in inventories.

Hence, for each element t_{ij} of $\mathbf{T}$ there may be numbers l_{ij} and u_{ij} that define boundaries for t_{ij}, i.e.

$$l_{ij} \leq t_{ij} \leq u_{ij}. \tag{A8}$$

All the concepts that were motivated for constraints yield constraint equations that are *linear* in the elements of the table. In order to express all constraints in a closed form, the MRIO tables must be vectorised. Assume an MRIO table $\mathbf{T}$ and its vectorisation $\boldsymbol{\tau}$, then *every* constraint can be expressed in the form

$$\mathbf{m}^{\mathrm{T}}\boldsymbol{\tau} = c \quad \text{with} \quad \mathbf{m}, \boldsymbol{\tau} \in \mathbb{R}^{I}, \quad c \in \mathbb{R}. \tag{A9}$$

The vector $\mathbf{m}$ contains the appropriate coefficients and the value c is the corresponding right-hand side value of the constraint. Assume that the vectorised MRIO MRIOtable $\boldsymbol{\tau}$ is subject to K constraints. Then each constraint can be represented in the form of Equation A9 with appropriate vectors $\mathbf{m}_k$ and right-hand side values c_k, $k = 1, \ldots, K$. More importantly, all K constraints can be summarised as

$$\underbrace{\begin{pmatrix} \mathbf{m}_1^{\mathrm{T}} \\ \mathbf{m}_2^{\mathrm{T}} \\ \vdots \\ \mathbf{m}_K^{\mathrm{T}} \end{pmatrix}}_{=:\mathbf{M}} \boldsymbol{\tau} = \underbrace{\begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_K \end{pmatrix}}_{=:\mathbf{c}}.$$

Hence all equality constraints can be elegantly written as

$$\mathbf{M}\boldsymbol{\tau} = \mathbf{c} \quad \text{with} \quad \mathbf{M} \in \mathbb{R}^{K\times I}, \boldsymbol{\tau} \in \mathbb{R}^{I}, \mathbf{c} \in \mathbb{R}^{K}. \tag{A10}$$

As mentioned before, boundary constraints may hold for each element of the vectorised MRIO $\boldsymbol{\tau}$ (as shown in Equation A8). Hence, for each element $\boldsymbol{\tau}_i$ of $\boldsymbol{\tau}$, there might be a lower boundary l_i and an upper boundary u_i. These boundary conditions can be expressed as

$$l_i \leq \boldsymbol{\tau}_i \leq u_i.$$

The values l_i and u_i can be summarised in vectors $\mathbf{l}$ and $\mathbf{u}$, so that the boundary conditions for the whole MRIO $\boldsymbol{\tau}$ are given by

$$\mathbf{l} \leq \boldsymbol{\tau} \leq \mathbf{u} \quad \boldsymbol{\tau}, \mathbf{l}, \mathbf{u} \in \mathbb{R}^{I}_{\infty} \quad \text{with} \quad \mathbb{R}_{\infty} = \mathbb{R} \cup \{\infty, -\infty\}. \tag{A11}$$

By using $\mathbb{R}_{\infty}$ it is possible to include those values $\boldsymbol{\tau}_i$ into Equation A11 that are subject to either only one boundary, or no boundaries at all.

Using this formulation of constraints and boundary conditions, the reconciliation task can be formulated as a constrained optimisation problem.

The initial MRIO table $\mathbf{T}^0$ is usually subject to constraints summarised in the matrix $\mathbf{M}$ and the right-hand side vector $\mathbf{c}$. hat is subject to more than one constraint. In order to reconcile the initial table $\mathbf{T}^0$, an optimisation algorithm has to be used to obtain an MRIO table $\mathbf{T}$ that adheres to all constraints given by $\mathbf{M}$ while violating the initial table $\mathbf{T}^0$ as little as possible. The violation of the initial table is measured by a so-called *objective function*.

Usually there are more elements in the MRIO table than there are constraints. Hence, the unknowns outnumber the constraints, resulting in the system being underdetermined. The system therefore exhibits too many degrees of freedom to yield a unique solution analytically.

In order to approach this problem, one has to make assumptions about the unknown elements in order to provide initial data for each value in the MRIO, yielding a generally imbalanced initial estimate $\mathbf{T}^0$ for the MRIO table (Oosterhaven et al., 2008; Glen et al., 2011; Bouwmeester and Oosterhaven, 2008).

Using $\mathbf{T}^0$, an optimisation problem to find an MRIO table $\mathbf{T}$ that fulfils all constraints (a so-called feasible MRIO table) can be formulated as

$$\min_{\boldsymbol{\tau}} f(\boldsymbol{\tau}, \boldsymbol{\tau}^0) \quad \text{subject to } \mathbf{M}\boldsymbol{\tau} = \mathbf{c},$$

where $\boldsymbol{\tau}$ and $\boldsymbol{\tau}^0$ are the vectorised representations of $\mathbf{T}$ and $\mathbf{T}^0$. Often the problem is augmented by introducing vectors of upper and lower boundaries $\mathbf{l}$ and $\mathbf{u}$ for $\boldsymbol{\tau}$.

$$\min_{\boldsymbol{\tau}} f(\boldsymbol{\tau}, \boldsymbol{\tau}^0) \quad \text{subject to } \mathbf{M}\boldsymbol{\tau} = \mathbf{c}, \quad \mathbf{l} \leq \boldsymbol{\tau} \leq \mathbf{u}. \tag{A12}$$

Assume further, that for value in the MRIO (in this case, this corresponds to each value in the vector $\boldsymbol{\tau}$) and for each of the values in the constraint vector $\mathbf{c}$, there exists a value for the standard deviation corresponding to the reliability to that particular value. Then the standard deviation values for $\boldsymbol{\tau}$ can be summarised in a vector $\boldsymbol{\sigma}_{\boldsymbol{\tau}}$, and the standard deviation values for $\mathbf{c}$ can be summarised in a vector $\boldsymbol{\sigma}_{\mathbf{c}}$. By including these vectors into the reconciliation problem given in Equation A12, the problem becomes

$$\min_{\boldsymbol{\tau}} f(\boldsymbol{\tau}, \boldsymbol{\tau}^0, \boldsymbol{\sigma}_{\boldsymbol{\tau}}, \boldsymbol{\sigma}_{\mathbf{c}}) \quad \text{subject to } \mathbf{M}\boldsymbol{\tau} = \mathbf{c}, \quad \mathbf{l} \leq \boldsymbol{\tau} \leq \mathbf{u}.$$

This equation was already presented as Equation 1. Note that Equations A12 and 1 do not specify the *objective function* f. In fact, the choice of a suitable and meaningful objective function *and* a powerful optimisation method that can solve Equation 1, has been a major topic in research over the last few decades. Bacharach (1970) presents the method which has been the most successful one so far: the RAS method.

AISHA offers a number of optimisation routines to solve the reconciliation problem. Generally, optimisation routines used for MRIO reconciliation are broadly divided into two classes: RAS-type algorithms and constrained optimisation algorithm. These categories are slightly misleading. RAS-type methods are in fact optimisation methods themselves. Bacharach (1970) motivated the objective function for RAS. AISHA offers reconciliation algorithms from both families. Hence, during this project both RAS-type and constrained optimisation routines will be used.

APPENDIX 4. RESULTS FOR FINAL DEMAND- AND VALUE-ADDED BLOCKS

TABLE A1. Euclidean metric distance for the final demand block in million Euros.

	GEE	XEE	ECE	GCE	XCE	EEE	CCC
XXX	6.81	0.859	1.06	6.85	0.891	1.34	1.19
CCC	6.77	0.913	0.564	6.57	0.956	0.663	
EEE	6.82	1.13	0.573	6.66	1.15		
XCE	6.76	0.632	0.767	6.67			
GCE	8.95	6.86	6.68				
ECE	6.76	0.679					
XEE	6.74						

TABLE A2. Mean absolute difference for the final demand block in million Euros.

	GEE	XEE	ECE	GCE	XCE	EEE	CCC
XXX	83.31	9.83	15.07	137.41	9.94	18.09	18.75
CCC	83.46	16.16	9.36	128.62	16.37	9.75	
EEE	84.38	16.03	4.82	131.63	16.15		
XCE	81.44	9.47	12.65	132.84			
GCE	190.32	135.69	130.75				
ECE	80.33	11.86					
XEE	76.35						

TABLE A3. Regression coefficients for the final demand block.

	GEE	XEE	ECE	GCE	XCE	EEE	CCC
XXX	0.000	0.218	0.020	0.009	0.191	0.025	0.022
CCC	0.001	0.003	0.583	0.169	0.049	0.637	
EEE	0.000	0.002	0.880	0.070	0.043		
XCE	0.000	0.073	0.054	0.133			
GCE	0.018	0.004	0.133				
ECE	0.000	0.002					
XEE	0.001						

TABLE A4. Correlation coefficients for the final demand block.

	GEE	XEE	ECE	GCE	XCE	EEE	CCC
XXX	0.996	0.533	0.858	0.903	0.562	0.839	0.851
CCC	0.966	0.938	0.236	0.588	0.779	0.201	
EEE	0.988	0.947	0.062	0.724	0.791		
XCE	0.988	0.728	0.766	0.634			
GCE	0.863	0.932	0.634				
ECE	0.985	0.949					
XEE	0.959						

TABLE A5. Euclidean metric distance for the value-added block in million Euros.

	GEE	XEE	ECE	GCE	XCE	EEE	CCC
XXX	14.75	3.89	8.27	7.101	3.58	12.45	11.40
CCC	18.30	11.99	6.40	13.19	11.78	7.52	
EEE	18.25	12.61	6.39	12.97	12.20		
XCE	13.56	3.31	8.19	3.85			
GCE	13.06	5.61	9.59				
ECE	16.02	8.48					
XEE	13.88						

TABLE A6. Mean absolute difference for the value-added block in million Euros.

	GEE	XEE	ECE	GCE	XCE	EEE	CCC
XXX	323.00	94.59	188.13	119.37	52.14	270.45	239.17
CCC	446.80	277.56	181.51	269.90	244.96	246.74	
EEE	459.54	287.37	141.89	268.26	253.32		
XCE	277.24	73.34	179.15	71.32			
GCE	222.07	96.48	200.72				
ECE	391.34	209.77					
XEE	268.47						

TABLE A7. Regression coefficients for the value-added block.

	GEE	XEE	ECE	GCE	XCE	EEE	CCC
XXX	0.000	0.702	0.296	0.003	0.919	0.121	0.261
CCC	0.001	0.183	0.795	0.035	0.259	0.698	
EEE	0.000	0.079	0.777	0.003	0.118		
XCE	0.001	0.670	0.293	0.005			
GCE	0.007	0.005	0.007				
ECE	0.001	0.231					
XEE	0.004						

TABLE A8. Correlation coefficients for the value-added block.

	GEE	XEE	ECE	GCE	XCE	EEE	CCC
XXX	0.978	0.162	0.455	0.938	0.041	0.652	0.488
CCC	0.966	0.572	0.109	0.812	0.490	0.164	
EEE	0.982	0.718	0.117	0.945	0.655		
XCE	0.978	0.181	0.459	0.929			
GCE	0.913	0.926	0.918				
ECE	0.974	0.518					
XEE	0.933						