The PIOLab: Building global physical input-output tables in a virtual laboratory
The PIOLab - Building global physical input-output tables in a virtual laboratory

Wieland, Hanspeter¹, Lenzen, Manfred², Geschke, Arne², Fry, Jacob², Wiedenhofer, Dominik³, Eisenmenger, Nina³, Schenk, Johannes⁴ & Giljum, Stefan¹

¹ Institute for Ecological Economics, Vienna University of Economics and Business, Vienna, Austria
² ISA, School of Physics A28, The University of Sydney, Sydney, NSW 2006, Australia
³ Institute of Social Ecology, University of Natural Resources and Life Sciences, Vienna, Austria
⁴ Montanuniversität Leoben, Chair of Ferrous Metallurgy, Leoben, Austria

Abstract:

Informed environmental-economic policy decisions require a solid understanding of the economy’s biophysical basis. Global physical input-output tables (gPIOTs) collate a vast array of information on the world economy’s physical structure and its interdependence with the environment. However, building gPIOTs requires dealing with mismatched and incomplete primary data with high uncertainties, which makes it a time-consuming and labor-intensive endeavor. We address this challenge by introducing the PIOLab: A virtual laboratory for building gPIOTs. It represents the newest branch of the Industrial Ecology virtual laboratory (IELab) concept, a cloud-computing platform and collaborative research environment through which participants can use each other’s resources to assemble individual input-output tables targeting specific research questions. To overcome the lack of primary data, the PIOLab builds extensively upon secondary data derived from a variety of models commonly used in Industrial Ecology. We use the case of global iron-steel supply chains to describe the architecture of the PIOLab and highlight its analytical capabilities. A major strength of the gPIOT is its ability to provide mass-balanced indicators on both apparent/direct and embodied/indirect flows, for regions and disaggregated economic sectors. We present the first gPIOTs for 10 years (2008-2017), covering 32 regions, 30 processes and 39 types of iron/steel flows. Diagnostic tests of the data reconciliation show a good level of adherence between raw data and the values realized in the gPIOT. We conclude with elaborating on how the PIOLab will be extended to cover other materials and energy flows.

Keywords:

Industrial ecology, material flow analysis, physical input-output tables, virtual laboratories, environmental input-output analysis, footprint, consumption-based indicators, iron, steel
1. INTRODUCTION

Progress towards sustainability necessitates comprehensive, quantitative research analyzing the relation between socioeconomic activities and biophysical processes across different spatiotemporal scales (Haberl et al., 2019). Socioeconomic activities are directly and indirectly enabled by ‘free gifts of nature’ (Daly, 1968; Duchin, 2009; Leontief, 1970), i.e. the extraction of resources from, and dilution of pollutants to the natural environment. These environmental pressures are a dominant driving force of global environmental change – e.g. global warming, biodiversity loss and environmental degradation (Steffen et al., 2015). Due to globalization, production and consumption activities take place at an increasing spatial distance (Fischer-Kowalski et al., 2015; Schaffartzik, Mayer, Eisenmenger, & Krausmann, 2016), giving rise to environmental burden shifting between regions through international trade (Oita et al., 2016; Peters, Minx, Weber, & Edenhofer, 2011). Global models that reflect such biophysical interdependencies can play a pivotal role in informing environmental-economic policy decisions.

A wide spectrum of models and accounting frameworks are available to industrial ecologists, each describing certain aspects of the socioeconomic metabolism (SEM), i.e. the biophysical flows exchanged between societies and their natural environment as well as the flows within and between social systems (Haberl et al., 2019; Pauliuk, Majeau-Bettez, & Müller, 2015). This includes process-based life-cycle assessment (LCA; EU JRC, 2010), environmentally extended monetary input-output analysis (EEIOA; Miller & Blair, 2009), material or substance flow analysis (MFA/SFA; Baccini & Brunner, 2012) and economy-wide material flow accounting (ew-MFA; Krausmann, Schandl, Eisenmenger, Giljum, & Jackson, 2017). To overcome the limitations of individual methods, various combinations of approaches have been presented. For example, the waste input-output (WIO) model was developed to explicitly take into account the interdependence between flows of goods and wastes in input-output (IO) analysis (Nakamura & Kondo, 2002); hybrid LCAs were proposed to relax the system boundary selection problem (Suh et al., 2004) and more recently, dynamic MFA techniques (Müller, Hilty, Widmer, Schluep, & Faulstich, 2014) are increasingly applied to incorporate stock-flow relations in ew-MFA (Krausmann, Wiedenhofer et al., 2017; Wiedenhofer, Fishman, Lauk, Haas, & Krausmann, 2019). Such cross-fertilizations between modelling schools have the potential to be more comprehensive and informative than either approach by itself.

Physical IO tables (PIOTs), sometimes referred to as IO-based MFA (Courtonne, Alapetite, Longaretti, Dupré, & Prados, 2015; Nakamura, 2011), are conceived as a way for combining the strengths of IO modeling and (ew-)MFA (Wachs & Singh, 2018). When constructed following an economy-wide perspective using a single physical unit (e.g. metric tonnes), PIOTs represent a mass balanced map of biophysical flows between socio-metabolic processes including society-nature interactions. These principles bring a number of advantages. First, PIOTs can circumvent the homogeneous price assumption of EEIOA, which implies proportionality between monetary...
and physical flows (Weisz & Duchin, 2006). In fact, physical quantities are not sold at the same price to all consumers, an issue already discussed in early energy IO analysis (Bullard & Herenden, 1975). Second, PIOTs capture all physical flows whether they have a monetary value or not, including wastes and secondary materials. This is an advantage over monetary IO tables (MIOT), whenever research questions focus more strongly on the biophysical dimension of economic activities. For example, as EEIOA is increasingly used for analyzing circular economy strategies (Aguilar-Hernandez et al., 2019; Çetinay, Donati, Heijungs, & Sprecher, 2020; Donati et al., 2020; Tisserant et al., 2017; Towa, Zeller, & Achten, 2020; Wiebe, Harsdorff, Montt, Simas, & Wood, 2019), the reliance on MIOTs has been identified as a major limitation because circularity policies are usually defined in physical units (Aguilar-Hernandez, Sigüenza-Sanchez, Donati, Rodrigues, & Tukker, 2018). Third, by tracing material flows through economic sectors in a mass-balanced manner, PIOTs open up the black box of ex-MFA and thereby widen its applicability for policy-oriented analysis that focuses on specific economic activities, such as manufacturing, construction, household consumption or public procurement (Altimiras-Martin, 2014; Giljum & Hubacek, 2009). Finally, PIOTs contribute to widening the empirical scope of MFA studies, whose main interest is the quantification of material cycles. So far, most MFA/SFA studies are geographically constrained to single (world) regions and many times trade flows of final goods are ignored (Chen & Graedel, 2012). Due to globalized production-consumption systems, however, closing material cycles must be conceived at the global level (Graedel, Reck, Ciacci, & Passarini, 2019), an undertaking for which a global multi-regional PIOT (gPIOT) provides a practical framework.

gPIOTs can function as an integration framework for different data sources. The System of Environmental-Economic Accounting (SEEA; United Nations et al., 2014), which follows the accounting conventions from the System of National Accounts, arranges physical flows and stocks in a series of accounts. Global physical supply-use tables (PSUTs), the building blocks of gPIOTs, play a key role in this regard. Data for gPIOTs must be obtained from a variety of sources, such as energy and waste accounts and statistics on production, recycling, emissions and international trade. The application of the mass balance principle means that a coherent picture emerges from the integration of the different data sources.

Since Leontief’s seminal work on the economic IO model (Leontief, 1936), various environmental-economic frameworks have evolved for integrating non-monetarized ‘ecological commodities’ in IO analysis (Ayres & Kneese, 1969; Daly, 1968; Forssell & Polenske, 1998; Isard et al., 1968; Leontief & Ford, 1972) but it was only from the 1990s onwards that IO models using

---

1 It should be noted, as discussed by Weisz and Duchin (2006), that monetary and physical IO tables would produce identical results, e.g. for indirect resource requirements of final consumption i.e. footprint indicators, when both tables are related to each other by a price matrix.
a single physical unit were increasingly developed (Hoekstra & van den Bergh, 2006). First PIOTs were constructed for Germany (Stahmer, Kuhn, & Braun, 1997), Netherlands (Konijn, Boer, & van Dalen J., 1995), Denmark (Gravgård Pedersen, 1999) and Finland (Mäenpää, 2002). According to Hoekstra (2010), these models have been one-off, large-scale efforts that were not able to provide convincing solutions to the primary problem i.e. the large investment costs associated with data collection and processing. To cope with the lack of physical data, hybrid IO techniques were developed, such as the ‘waste IO approach to material flows’ (WIO-MF; Nakamura, Nakajima, Kondo, & Nagasaka, 2007) or UPIOM (Nakamura, Kondo, Matsubae, Nakajima, & Nagasaka, 2011), which use proxy information to synthesize PIOTs from detailed MIOTs. However, these approaches are only applicable for regions where detailed MIOTs are available, e.g. Japan (Nakatani, Maruyama, & Moriguchi, 2020; Ohno, Sato, & Fukushima, 2018) or the United States (Chen, Graedel, Nuss, & Ohno, 2016; Nuss, Chen, Ohno, & Graedel, 2016). Other studies employed process engineering models to build PIOTs for the sub-national level (Singh, Compton, Hawkins, Sobota, & Cooter, 2017; Wachs & Singh, 2018). Lenzen and Lundie (2012) constructed a PIOT for the dairy sector of New Zealand, utilizing engineering knowledge and RAS. Kovanda (2019) compiled the first SEEA-conform PIOT for the Czech Republic. Bösch and colleagues (2015) constructed a PIOT for wood and paper flows in Germany. Besides these mostly single-region models, global mixed-unit IO models have been constructed as well, e.g. hybrid EXIOBASE (Merciai & Schmidt, 2018) or models of global biomass flows (Bruckner et al., 2019; Croft, West, & Green, 2018).

Like their monetary counterparts, PIOTs are generally underdetermined systems, i.e. not all data elements of the table are explicitly informed by primary data. A recurring challenge of all PIOT construction efforts is dealing with incomplete and mismatched data sources with high uncertainties, which makes the construction process a time-consuming and labor-intensive undertaking. However, in order to be relevant for policy making, IO models need to be created and updated in a timely, continuous and cost-efficient way (Wiedmann, Wilting, Lenzen, Lutter, & Palm, 2011).

The present work addresses this challenge by introducing the PIOLab, a virtual laboratory for building gPIOTs. It represents the newest branch of the Industrial Ecology virtual laboratory (IELab) concept, which is a collaborative research environment with a cloud-based high-performance computing platform and a single-step reconciliation engine. This enables modelers to use each other’s resources to assemble their own tailor-made large-scale multi-regional input-output model (MRIO) that is fit for purpose in a cost-efficient and timely way (Lenzen, 2014). IELabs are an evolving niche of IO research that hitherto were mainly used for compiling environmentally extended monetary IO databases (Geschke & Hadjikakou, 2017). Since the construction of the forerunner IELab for Australia (Lenzen, Geschke, Malik et al., 2017; Wiedmann, 2017), various branches have been developed including a global MRIO laboratory.
(Lenzen, Geschke, Abd Rahman et al., 2017) and several regional i.e. subnational MRIO modeling suits for various countries including for example Japan (Wakiyama, Lenzen, Geschke, Bamba, & Nansai, 2020), China (Wang, 2017), Indonesia (Faturay, Lenzen, & Nugraha, 2017), Taiwan (Faturay, Sun et al., 2020) or the United States (Faturay, Vunnava, Lenzen, & Singh, 2020).

The PIOLab is the first IELab designed for building purely physical IO tables. To overcome the lack of primary data in the physical domain, the PIOLab philosophy is to build extensively upon secondary data derived from accounting frameworks and models each describing certain aspects of SEM, such as MFA, ew-MFA, WIO-MF and LCA. To describe the architecture of the PIOLab and to highlight its analytical capabilities, we show-case the first gPIOT of global iron-steel supply chains which covers 32 regions, 30 processes and 39 types of iron/steel flows for the years 2008-2017.

Steel was selected as a first case study for several reasons. It is a key building block of manufactured capital i.e. in-use stocks (Graedel, 2010; Song et al., 2020; Weisz, Suh, & Graedel, 2015) and a strategic material for sustainable development (Müller, Wang, & Duval, 2011; Pauliuk & Müller, 2014), but also a major concern for local pollution (e.g. particulate matter) and CO₂ emissions (Ryberg, Wang, Kara, & Hauschild, 2018; van der Voet, 2013). The iron and steel industry is the largest industrial source of CO₂ emissions from fossil fuels (5% in 2014) and due to the high energy intensity of steel production particularly challenging to decarbonize (Carpenter, 2012; Davis et al., 2018). With steel being a highly recyclable material, one of the most important technological options for the steel industry, besides energy and material efficiency measures (Allwood, 2013; Allwood, Cullen, & Milford, 2010), is to transition towards a circular economy through increasing the use of steel scrap as feedstock (Nechifor et al., 2020). Considering all this in the context that metal supply chains are increasingly organized on the global level (Schaffartzik et al., 2016), iron and steel provides an excellent case study to present the challenges faced in the compilation of gPIOTs and the utility of the PIOLab.

The paper is structured as follows. Section 2 describes the methodological features of the PIOLab and its application to the iron-steel case. Section 3 presents an analysis of the alignment of the final gPIOT with its raw data sources and process descriptions, which includes realizations of input and output ratios of the blast furnace process. Furthermore, this section presents a comparison of ew-MFA headline indicators and an overview over international trade in iron-steel supply chains. Section 4 (discussion and outlook) elaborates on the advantages, shortcomings and policy relevance of the PIOLab approach and elaborates on upcoming developments.

2. METHODOLOGY

The method section is structured in three parts. It starts by briefly summarizing the features that are common to all IELabs (2.1). General characteristics and unique features of the PIOLab
approach are presented in 2.2, followed by the particularities of building iron-steel gPIOTs (2.3), i.e. data sources and important modelling assumptions.

2.1. The IELab concept

An overarching goal of the IELab concept is to maximize flexibility so that users can compile their own tailor-made MRIO database that fits individual needs and research goals while at the same time giving access to data and code repositories that others have contributed. The IELab architecture rests on three pillars.

First, a generalized raw data management system that streamlines the extraction of information from data sources and organizes it for further processing. The functions that perform this task are called ‘data feeds’. Processed raw data are matched to a root classification, which is the maximum sectoral and regional detail that is theoretically achievable\(^2\), assuming universal availability of highly disaggregated data (Lenzen et al., 2014). To be used as constraints in the reconciliation process, raw data need to be concorded to the root classification.

The second pillar is a single-step reconciliation engine that compiles the MRIO table in a fully automated way. It requires an initial estimate of the table as input for reconciling the constraints provided by the data feeds. The resultant table is termed a ‘base table’, where the so-called base classification is any unique aggregation of the root classification. The reconciliation engine of the IELab is AISHA\(^3\) (Geschke, Lenzen, Kanemoto, & Moran, 2011). It provides different reconciliation algorithms including RAS-type methods like KRAS (Lenzen, Gallego, & Wood, 2009) or least-square constrained optimization approaches (Ploeg, 1988). Besides merely aligning the row and column sums of tables, such algorithms can incorporate constraints on arbitrarily sized and shaped subsets of matrix elements, including for example known sums of elements and ratios between flows. These algorithms consider information on the uncertainty i.e. standard errors (SE) of constraints and the initial estimate, thereby enabling the handling of conflicting constraints and finding compromise solutions.

The third pillar of the IELab is a cloud-based, high-performance computer (HPC) architecture with parallel computing capabilities, in order to undertake the large-scale constrained-

\(^2\) Aggregation of more detailed raw data may lead to an undesirable loss of valuable information. See for example arguments in Lenzen (2011). The overarching goal of the data management system of the IELab is to maximize the information that can be extracted from a raw data source and avoid information to be buried in the unfortunate aggregation of the source data.

\(^3\) Originally, AISHA has been developed in the course of the construction of the Eora MRIO. The acronym stands for “An Automated Integration System for Harmonized Accounts”.

---
optimization task in a reasonable amount of time. Each IELab builds on the same general architecture. A decisive feature of the IELab concept is the large degree of flexibility afforded by avoiding the need to lock-in any particular MRIO structure, that is, a regional and sectoral classification for the base table, at the time the laboratory is constructed (Lenzen et al., 2014). For further details on the IELab concept, we refer to the editorial of a special issue on virtual laboratories by Geschke and Hadjikakou (2017). The IELab architecture and workflow⁴ as realized in the PIOLab is shown in Figure 1.

Figure 1: A generalized representation of the PIOLab architecture and the workflow. Yellow squares stand for datasets i.e. arrays that contain information. Blue circles depict functions or general processing steps. The main purpose of feed functions is to transform the data and its classification in a way that it can be used as an input in the single-step reconciliation process (AISHA). Two general types of data feeds exist i.e. “data feeds for data constraints” which are based on primary or secondary data as well as “data feeds for fundamental constraints” which includes for example balancing constraints.

⁴ When a root classification is constructed and data feeds are given, the workflow in the IELab consists of the following steps: First, users define a structure for the base table by selecting regional and sectoral classifications, which is realized by constructing root-to-base concordances. Second, users provide AISHA an initial estimate (for each element) of the envisioned base table and choose the constraints that this base table should adhere to by selecting specific data feeds. In the end, users initiate an AISHA run to balance the table and reconcile it with the constraints Lenzen, Geschke, Abd Rahman et al. (2017).
2.2. The PIOLab approach

Initial estimate and secondary data sources

AISHA requires initial estimates for all elements of the base table, but primary data for all physical flows is usually not readily available (Kovanda, 2019). The PIOLab approach is to combine secondary data derived from available socio-metabolic accounting frameworks and models whenever primary data is not available. This can include life-cycle inventory databases (Wernet et al., 2016) for designing more detailed process descriptions; results from dynamic material flow models (Pauliuk, Wang, & Müller, 2013) for incorporating addition to and withdrawal from societal material stocks; or hybrid IO models, such as the waste IO approach to material flows (WIO-MF; Nakamura et al., 2007) for producing initial estimates of physical inter-industry flows.

The WIO-MF framework occupies a special place in the list of secondary data providers due to its ability to estimate large numbers of unknown material flows. WIO-MF allows inferring a physical flow matrix from monetary IO tables (Chen et al., 2016; Nakamura et al., 2007; Nakatani et al., 2020; Nuss et al., 2016; Nuss, Ohno, Chen, & Graedel, 2019; Ohno, Nuss, Chen, & Graedel, 2016). A filter matrix $\mathbf{M}$ containing zeros and ones is used to remove monetary flows i.e. inputs from the technology matrix $\mathbf{A}$ of the conventional monetary MRIO model, which do not correspond with apparent physical quantities ($\mathbf{A}_{\text{new}} = \mathbf{A} \circ \mathbf{M}$). Here, $\circ$ indicates an element-wise multiplication of two matrices. Subsequently, a new Leontief inverse can be calculated ($\mathbf{L}_{\text{new}} = (\mathbf{I} - \mathbf{A}_{\text{new}})^{-1}$) to estimate a modified gross production vector $\mathbf{x}_{\text{new}}$ by post-multiplying the inverse with the total final demand vector ($\mathbf{x}_{\text{new}} = \mathbf{L}_{\text{new}} \times \mathbf{y}$). The modified variables are used to allocate the material use $\mathbf{p}$ of industries to final demand $\mathbf{y}$ via $\mathbf{D} = \mathbf{q} \mathbf{L}_{\text{new}} \hat{\mathbf{y}}$, where $\mathbf{q}$ stands for the diagonalized direct material use intensity vector $\mathbf{q} = \mathbf{f} \mathbf{p} \mathbf{x}_{\text{new}}^{-1}$. Note that the multiplication of sectoral material use $\mathbf{p}$ with a yield vector $\mathbf{f}$ results in a flow matrix $\mathbf{D}$ where only useful output is present$^6$. When constructed from a MRIO database, $\mathbf{D}$ reflects a multi-region physical flow table and the material composition of final products, covering for example flows from higher manufacturing to end-use.

The decision for or against using a WIO-MF approach for making initial estimates is up to the PIOLab user and it can be adapted - for specific regions, processes and flows - when deemed necessary for a particular research question. The construction of the WIO-MF model obliges modelers to make simplifying assumptions that introduce uncertainties. This includes assumptions on the physical or monetary nature of flows, yields of processes and price homogeneity. However, because AISHA requires uncertainty-estimates, i.e. standard errors for all elements of the initial

---

$^5$ Viewed from the perspective of a structural path analysis, $\mathbf{L}_{\text{new}}$ and $\mathbf{x}_{\text{new}}$ reflect only such paths that are assumed to be physical inter-industry flows.

$^6$ For clarifications on the distinction between useful output and gross output (i.e. throughput) in (physical) IO models, we refer to section three in the supplementary information of Pauliuk et al. (2015).
estimate, the reconciliation process addresses these assumptions. Given the unfavorable situation of general data scarcity in the physical domain, we believe the WIO-MF approach represents a generalizable estimation method for filling data gaps in the initial estimate of gPIOTs.

**Root classification**

Modelling the full scale and composition of all biophysical flows in one highly detailed gPIOT is very ambitious and by no means always necessary, since useful analysis can be conducted for any of the individual materials or substances of the full set of flows. Importantly however, conceptual and physical consistency is a key feature of the PIOLab, which is operationalized by defining a specific root classification.

As long as there is no universal root classification for all material and energy flows, the approach of the PIOLab is to conceptually separate the socio-economic metabolism (SEM) into two sections, the ‘modelled core system’ and ‘other SEM processes’. The modelled core system is the specific subset of socio-metabolic processes and biophysical flows that the researcher is primarily interested in and wants to model in detail, such as iron and steel supply chains in this paper. This core system is represented by physical Supply-Use tables $S$ and $U$ and the final use matrix $Y$, shown in Figure 2 in yellow. For stock-building materials (such as steel), the final use matrix represents gross additions to material stocks. Scope and highest possible level of detail of the modelled core system is set by the root classification. Physical flows between the modelled core system and other SEM processes - i.e. all inputs to and outputs from the core system that are processed, treated or used by other societal activities - are represented by additional rows/inputs $O$ and columns/outputs $P$, shown in blue in Figure 2. Hereafter, these flows are termed SEM-inputs and SEM-outputs.

Finally, introduction of rows and columns for inputs from nature $E$ and the discard of residuals $W$ represent the interaction of the modelled core system with the natural environment, shown in green in Figure 2. Inputs from nature together with SEM-inputs constitute the ‘boundary inputs’ to the core system. SEM-outputs plus discard of residuals represents ‘boundary outputs’ (see Figure 2).

This conceptualization yields a full system description of SEM with varying degrees of detail, following the mass balances principle as well as the SEEA guidelines. Moreover, it allows modelers and PIOLab users to invest their efforts on the subset of biophysical flows and socio-

---

7 The separation into modeled core system and other SEM processes is similar to the foreground/background system distinction in LCA.

8 This terminology follows the recommendations of SEEA. In ew-MFA, ‘discard of residuals’ corresponds to ‘domestic processed outputs’ (DPO) and ‘inputs from nature’ represents ‘used extraction of materials’ including air and oxygen taken up during combustion.
metabolic processes that are most relevant for their research, while retaining conceptual consistency.

Figure 2: A two-region example of the generalized physical supply-use table (PSUT) framework as realized in the PIOLab. The tables of the core system are shown in yellow, flows to and from other SEM-processes in blue and society-nature interactions in green. Use tables (U) are matrices that record the intermediate inputs of processes with the dimension flow-by-process. The Supply tables (S) record useful outputs (i.e. flows that are used by other processes of the core system which is different from its throughput) of processes with the dimension process-by-flow. Final use matrices (Y) show final consumption products, which also includes gross additions to societal material stocks for stock building materials such as steel. \( U_{1,1}, U_{1,2} \) and \( U_{2,1}, U_{2,2} \) as well as \( Y_{1,1}, Y_{1,2} \) and \( Y_{2,1}, Y_{2,2} \) stand for trade flows of intermediates and final products respectively. In general, columns depict the inputs and rows the outputs of processes. Summation of all process inputs yields \( \mathbf{x}' = i \mathbf{U} + k \mathbf{O} + m \mathbf{E} \), where ' stands for transposition of vectors and \( i, k, m \) for appropriate summation row vectors containing ones. Summation of all process outputs gives \( \mathbf{x} = S \mathbf{j} + P \mathbf{l} + W \mathbf{n} \), where \( j, l \) and \( n \) stand for appropriate summation column vectors containing ones. Vector \( \mathbf{x} \) stands for total supply-use of processes (i.e. throughput) and \( \mathbf{q} \) for total supply-use of flows of useful output which is why \( \sum \mathbf{q} < \sum \mathbf{x} \).

The PIOLab differs from other IELabs by including non-market flows in its root classification, i.e. physical flows that are not distributed through markets and therefore have no monetary value. The sector root classifications of the global MRIO Lab (Lenzen, Geschke, Abd Rahman et al., 2017) is based on the Harmonized System/Central Product Classification (HSCPC) for traded commodities and in its current version differentiates 6357 products. The root classification of the global MRIO laboratory does not, for example, include such flows as molten/hot metal (e.g. liquid steel) or scrap from processing (e.g. the rolling and forming of steel) since they are situated within steel plants and therefore not distributed via markets. However, these flows are important when
constructing gPIOTs, as they constitute inputs and outputs of crucial transformation processes and therefore are decisive for maintaining mass balances and the integrity of the overall system description.

### 2.3. Building iron-steel gPIOTs

The root classification of the forerunner PIOLab for the global iron-steel metabolism differentiates 76 transformation processes, 266 types of ferrous material flows and 221 regions. The regional root classification is identical to the one of the global MRIO Lab. All processes that use ferrous materials are included, i.e. mining, iron- and steelmaking, casting, rolling, forming, waste management, steel-using manufacturing sectors as well as the final use-phase i.e. fixed capital formation and final consumption of households or governments. This also includes flows of recycled steel scrap from processing (i.e. rolling and forming) and manufacturing (also termed ‘new scrap’). Extraction of crude ore and the combustion of oxygen by furnaces represent inputs from nature. Wastes discarded to the natural environment, such as greenhouse gas emissions, represent outputs to nature. Coke, flux and EoL-scrap (‘end of life scrap’ which is also termed ‘old scrap’) are treated as SEM-inputs. By-products of the iron- and steelmaking process, such as blast furnace gas or slag, are classified as SEM-outputs.

Table 1 lists all primary and secondary data for setting up the initial estimate of the base table and writing data feed constraints. The statistical yearbooks of the World Steel Association report 27 outputs of the iron and steel industry (intermediate and finished steel products) in 135 countries, accounting for approximately 85% of global steel production (World Steel Association, 2018). Iron ore grades and extraction volumes are taken from the UNEP-IRP global material flow database (UNEP-IRP, 2017). Information on physical trade flows is sourced from BACI (Gaulier & Zignago, 2010), which differentiates a large number of finished steel products (187). This is why the root classification for flows (266) is more detailed than for processes (76). Data on the regional supply of EoL-Scrap is sourced from the dynamic MFA model of Pauliuk and colleagues (2013). Coke consumption and production of blast furnace gas is taken from IEA’s energy balances (IEA, 2012). We use the MRIO database EXIOBASE (Wood et al., 2015) to build a WIO-MF model for providing initial estimates of the steel flows in manufacturing and to link 10 steel-using manufacturing sectors with final consumption. Information on manufacturing yields, required for the WIO-MF model, are sourced from the global steel flow model of Cullen and colleagues (2012). Cullen and colleagues’ model also provides estimates for the end-use split, i.e. the share in finished steel products used by different manufacturing sectors.
Table 1: Summary of data sources used for constraints and in the setup of the initial estimate table of the iron-steel gPIOT. Data sources that are not used for formulating data feeds and setting constraints (see column on the far right) are only used for constructing the initial estimate.

In general, most data sources report the useful output of processes i.e. the Supply-table such as the amount of liquid steel produced in basic oxygen furnaces (BOF) in China. There is little information on the input-side of these processes i.e. the Use-table (e.g. amount of steel scrap and pig iron used in BOF in China). The PIOLab uses process descriptions sourced from MFAs to derive initial estimates for the input-side (i.e. the Use-table) and at the same time formulate ratio constraints for the reconciliation algorithm.

The classification of the base table differentiates 30 processes, 39 flows and 32 regions. The number of regions in the base classification is primarily determined by the region classification of the WIO-MF model (EXIOBASE) and the number of processes mainly by the detail provided by the World Steel Association. 6 of the 32 regions represent country groups, such as Rest-of-the-World (RoW) America or RoW Europe. The 26 countries included in the base table account for 89% of global iron extraction and 88% of global steel production. Note that waste discarded to nature is not explicit in the base table because it is assumed that all by-products of iron and steelmaking (furnace slags, dusts, sludges or top gases i.e. off gases) are subject to some form of further treatment and hence are all classified as SEM-outputs. The root and base classifications are

---

9 The regional root-2-base concordance is included in the supporting information (SI-2) showing which countries are included in which RoW regions.
included in supporting information 2 (SI-2). Figure 3 visualizes the general layout and the different sections of the base table by showing heat-maps of the domestic PSUTs of China for 2008.

A: Supply table and boundary output block

B: Use table and boundary input block

Figure 3: Heat-maps of the physical Supply-Use tables (PSUTs) of China 2008. 3-A: The Supply table (S) depicting outputs of processes (in the rows), is shown at the top together with boundary flows of SEM-outputs (P). 3-B: The Use table (U), depicting inputs of processes (in the columns), together with boundary inputs (Q, E) and final use block (Y), is shown at the bottom. The PSUTs shown here include 30 processes, 39 flows, 5 boundary inputs and 8 final outputs of which 6 refer to final use and 2 to SEM-outputs i.e. by-products for further treatment. The cells are colored according to the quartiles they correspond to where dark colors represent larger values. Yellow signifies zero. The data used to for the heat-maps can be found in the supporting information (SI-1).
3. RESULTS

This section begins with an analysis of constraint realizations, followed by a comparison of ew-MFA headline indicators that are derived from the iron-steel gPIOT with results from established databases. The analysis of the constraint realization and the ew-MFA indicators is based on the gPIOT for the year 2008 since this is the base year for which the initial estimate has been constructed. In the end, we present an analysis of international trade flows in the iron-steel supply chain to highlight its analytical capabilities. This analysis is based on the more recent gPIOT for 2014. The gPIOTs are available as a time series for the years 2008-2017.

3.1. Analysis of constraint realizations

As a diagnostic test of the constructed base table, we present a rocket plot (Figure 4-A) to show how well the final gPIOT adheres to its raw data constraints. Each raw data set is accompanied with standard errors (SE) so that the reconciliation engine can find a compromise solution for conflicting data points, adhering more to any data that are tagged with relatively low SE. As indicated in the method section, besides raw data on trade, mostly process outputs i.e. production values (e.g. iron ore, EoL-Scrap and various steel products) were available for use as data point constraints. In the optimization run presented here, we have assigned higher SE to trade data than to production data. This can be seen in the rocket-plot shown in Figure 4-A (color of points accord to the different data sources where yellow stands for the trade data), comparing the raw data constraint values (x-axis) against the realized gPIOT values (y-axis) for the year 2008. The rocket-plot shows that larger constraint values, which are mostly production values, adhere more closely to the reported values. 17,092 data points were used as constraints in the reconciliation of which the large majority (16,172) refers to trade flows\textsuperscript{10}. The rocket-plot also shows the density distribution of data points (three circles in grey for 25, 50, 75%) to highlight areas of heavy overplotting. The outer circle, in the range of 10 – 500,000 tonnes, contains 75% of all data points.

Next, we present the realization of mass balances of the full system and individual processes. The balance of the full systems is visualized by the bar chart in Figure 4-C. The boundary input to the system comprises inputs from nature (2236 Mt of iron ore and 1192 Mt of air/oxygen) and inputs from SEM (429 Mt of coke/coal plus 251 Mt of EoL-Scrap plus ca. 190 Mt of flux). These inputs are eventually transformed into 1082 Mt of steel in products for final use, 1213 Mt solid/liquid (slag, dust and gangue) and 1994 Mt gaseous residuals, both for further treatment or use and hence classified as SEM-outputs. In total, boundary inputs amount to 4299 Mt and boundary outputs plus final use to 4290 Mt, which yields an imbalance of ca. 0.2 % of boundary inputs. In the

\textsuperscript{10} Note that the numbers of data points refer to the base classification, which is an aggregate of the root classification. The number of data points in the root classification is much higher.
optimization run presented here, mass balances of processes show larger imbalances (see scatter plot in Figure 4-B). Out of the 960 processes (32 regions with 30 processes), 32 show a relative imbalance larger than 15%, which are foremost section mill, rod mill and ingot casting with smaller throughputs.

Figure 4-A: The top left scatter plot shows a rocket plot which compares values as reported in data sources on the x-axis and the values as realized in the 2008 iron-steel gPIOT on the y-axis. Yellow indicates BACI trade data and dark blue other data sources. Due to over-plotting, the density distribution of data points (25, 50 and 75%) is shown with three circles (grey lines). The inner circle marks the area that comprises 25% of all data points and the outer circle 75%. Figure 4-B: The top right scatter plot compares total inputs (y-axis) and outputs (x-axis) of all processes in all regions of the base classification. Colors indicate the type of processes. The dashed lines mark +/- 15% deviation from the 45-degree equality line. Both scatter plots have a log10 scale in units of metric tonnes. Figure 4-C: The bar chart at the bottom shows the overall mass balance of the full system. It compares the boundary inputs (O + E) to boundary outputs (W + P) plus final use (Y). Boundary inputs reflect the sum of inputs from SEM and nature, which must equal the sum of final use i.e. final demand plus boundary outputs to SEM (by-products and wastes/emissions for further treatment) and nature (wastes/emissions without further treatment). Please note that, as mentioned in the method section, in the construction of the present gPIOT we assume that all wastes/emissions are subject to some form of treatment and hence boundary outputs to nature are zero. All results refer to the year 2008. All data can be found in the supporting information (SI-1).
As an example implementation of ratio constraints, we present the realization of technical input-output ratios for the blast furnace process\textsuperscript{11}. The Steel Manual (Stahlinstitut VDEh, 2015) is currently the PIOLab’s main technical process description for iron and steelmaking. The material balance of the blast furnace process is shown in Figure 5-A. It specifies that usually a tonne (t) of pig iron from the blast furnace requires inputs of approximately 1.55 t of iron ore, 0.45 t coke and coal, 1.25 t air/oxygen (for combustion) and 0.2 t of flux\textsuperscript{12} (mostly limestone). As a by-product, 2.1 t of blast furnace gas and 0.35 t of slag are produced, resulting in a total throughput of 3.45 t.

The realization of these ratio constraints is visualized in Figure 5-B. In general, iron ore input ratios, show the largest deviations, which indicates conflicts between different data sources (e.g. World Steel, BACI and input-output ratios). However, these variations are less than 0.17 t/t i.e. 9\% of the target value 1.55 t/t. The largest deviation for gas output ratios is 0.07 t/t i.e. 3\% of the target value 2.1 t/t. Other ratios, such as for the output of slag (with the target value 0.35 t/t) or input of coke (0.45 t/t), show even fewer variations compared to iron ore and gas. In terms of regions, the largest deviations from these ratio constraints are found for the Netherlands (NL), Great Britain (GB) and the United States (US). For all other regions, blast furnace ratios have been well realized.

\textsuperscript{11} Other process descriptions are incorporated as ration constraints for the two most important furnace types for steelmaking, i.e. basic oxygen furnace (BOF) and electric arc furnace (EAF), since their average scrap-input share differs significantly, with 0.2 tonnes and 0.8 tonnes per tonne of liquid steel respectively.

\textsuperscript{12} In metallurgy, flux refers to substances that are introduced in the smelting of ores to promote fluidity and to remove impurities in the form of slag. Limestone is commonly used for this purpose in smelting iron ores. Other materials used as fluxes are for example silica, dolomite or lime.
Figure 5-A shows the blast furnace material balance from the Steel Manual that is used for setting ratio constraints (unit: kg). Figure 5-B shows the realization of the input/output-ratios (per unit output of pig iron) in the final gPIOT for all regions where World Steel reports pig iron production. Country codes and underlying data can be found in the supporting information (SI-1 and SI-2). Results refer to the year 2008.

3.2. Comparison of gPIOT results with established databases

To show the extent with which PIOLab results differ from or agree with established databases, we sourced ew-MFA data for iron and steel flows from the UNEP-IRP global material flow database (UNEP-IRP, 2017) to compare these to the respective results from the iron-steel gPIOT. This comparison is shown in Figure 6 and includes Domestic Extraction (DE) and Domestic Material Consumption (DMC) as an indicator of national apparent consumption (Krausmann, Schandl et al., 2017). In addition, we calculated Raw Material Consumption (RMC) i.e. material footprint indicators (Wiedmann et al., 2015) with the gPIOT\textsuperscript{13} and compared them to results from the

\textsuperscript{13} Different types of physical IO tables and models can be constructed from the same set of PSUTs. The footprint-type indicators presented in this paper are derived from a gPIOT which is constructed using the approach proposed by Suh (2004), which is discussed in detail in the work of Altimiras-Martin (2014). IO models assume industry/process output to be homogeneous which is in conflict with a situation where process outputs are comprised of commodities \( Y \) and by-products \( P \) and \( W \). Suh proposed calculating footprint-type indicators via endogenization of by-products by treating them as negative inputs. In other words, matrices \( P \) and \( W \) (see Figure 2) are transposed and negative (\( -P' \) and \( -W' \)). This yields a physical IO model where \( P \) and \( W \) can be endogenously derived from final demand. In the present work, this is achieved by using a modified total process output vector \( x_{\text{new}} \) for building the IO table and subsequently total requirement matrix \( L \), which omits \( P \) and \( W \). Instead of \( x = Sj + Pi + Wn \) we estimate \( x_{\text{new}} = Sj \) which only contains ‘useful output’ and consequently \( \Sigma q = \Sigma x_{\text{new}} \). Using \( x_{\text{new}} \) and the multi-regional PSUTs, an IO model can be estimated following the commodity-by-industry framework as summarized in chapter five in Miller and Blair (2009). First, we calculate the industry input coefficients matrix \( B = U x_{\text{new}}^{-1} \) with the dimension product-by-
EXIOBASE MRIO (Wood et al., 2015). Indicators are presented on the level of single countries (scatter plots in Figure 6) and for five aggregated world regions (i.e. Asia/Pacific, America, Europe, Africa, Middle East) and China (bar charts in Figure 6). Hereafter, percentage deviations refer to the hypothetical mean value of the compared results.

The best alignment is observed for DE. On average, DE indicators for world regions vary by 0.8% and for countries by 3.4%. Compared to DMC and RMC, the deviations in DE results are rather small because the UNEP-IRP’s extraction account is also used for setting up data point constraints. Larger relative deviations are observed for DMC where results for world regions vary on average by 8.3% and by 22.9% for single countries. 12 out of the 32 regions show DMC deviations that are smaller than 15%, which includes China, India, Germany and Great Britain. The largest relative deviation in DMC is found for Sweden. Here the UNEP-IRP result (10.5 Mt) is around twice the size of the PIOLab result (5.1 Mt). The country with the second largest relative deviation is United States, where the gPIOT result (116.4 Mt) is around two thirds (66%) larger than the result from the UNEP-IRP database (69.9 Mt). Since DE results are very similar, differences in DMC must be rooted in the different estimates of physical trade flows and the physical trade balance (PTB; calculated via physical imports minus physical exports). Regarding the deviations in the DMC of Sweden, we see for example that UNEP-IRP and PIOLab data show very similar imports of ca. 6.8 and 6 Mt respectively. For exports, however, PIOLab results (29.8 Mt) are 4.6 Mt larger than the results reported in the UNEP-IRP data set (25.2 Mt). Deviations in the DMC of the United States are primarily a result of differences in imports, where the UNEP-IRP account reports 74.6 Mt and the gPIOT 111.2 Mt. Differences in the directionality of the PTB are only observed for one Rest-of-the-World (RoW) region i.e. RoW America (WL). Taking a closer look at the UNEP-IRP trade data, we found that global trade flows are not fully mass-balanced, meaning global imports and exports do not sum up to the same total, which might explain some of the differences observed in the comparison of UNEP-IRP and gPIOT results.

Of all the ew-MFA indicators compared, the largest relative deviations are observed for RMC. On average, RMC results for world regions vary by 29.4% and for single countries by 41.3%. Out of the 32 regions, 7 show variations that are smaller than 15%, which includes France, Great Britain, South Korea and Mexico. The largest relative deviations are observed for Brazil and Australia, two major iron ore extracting countries. For Brazil, the EXIOBASE result (133.9 Mt) is around 5 times larger than the PIOLab result (26.4 Mt). For Australia, the EXIOBASE result (60.3 Mt) exceeds the PIOLab result (17.8 Mt) by more than a factor of three. We find that the deviations in industry and the market share matrix \( \mathbf{D} = \mathbf{S} \mathbf{q}^{-1} \) with the dimension industry-by-product. This gives us the product-by-product technology matrix \( \mathbf{A} = \mathbf{BD} \) and subsequently the total requirement matrix \( \mathbf{L} = (\mathbf{I} - \mathbf{A})^{-1} \). All codes of the RMC calculus are included in the GitHub code repository. For further details on the communalities and differences of different physical IO models, we refer the interested reader to chapter three in the supplementary information of Pauliuk et al. (2015).
the RMC of Brazil and Australia are primarily a result of the differences in the RME of exports. For Brazil, the PIOLab result for the RME of exports (332 Mt) is 102.4 Mt larger than the result of EXIOBASE (229.6 Mt). For Australia, the PIOLab result (335 Mt) is 39.4 Mt larger than the EXIOBASE result (295.6 Mt). With the exception of China, we find that the directionality of raw material trade balances (RTB; calculated by subtracting RME of exports from the RME of imports), meaning whether a region is a net-importer or net-exporter of RMEs, is the same for EXIOBASE and the PIOLab. According to the PIOLab, China is a net-importer of RMEs (178.3 Mt), while EXIOBASE addresses China as a net-exporter (-98 Mt). Consequently, the PIOLab result for the RMC of China (1005.6 Mt) is 280.3 Mt larger than the EXIOBASE result (725.3 Mt). Data on aggregated trade flows can be found in the supporting information (SI-1).

Figure 6: Comparison of ew-MFA headline indicators from the PIOLab’s gPIOT with established databases, which includes Domestic Extraction (DE), Domestic Material Consumption (DMC) and Raw Material Consumption (RMC) i.e. the crude ore material footprint of nations. Here, DMC quantifies the national apparent consumption of ferrous materials, which is calculated via DE (of crude iron ore) plus physical imports minus physical exports of all ferrous materials, e.g. iron ore, pig iron and steel. RMC can be estimated by adding raw material equivalents (RME) of imports to and subtracting RME of exports from DE. Results for DE of crude iron ore and DMC of ferrous materials are sourced from the UNEP-IRP database and RMC is calculated using the EXIOBASE Mrio with DE from the UNEP-IRP database as the environmental extension. The global sums of DE, DMC and RMC add up to the same global total of 2.24 Giga tonnes. The bar charts at the top depict results for five aggregated world regions and China. The scatter plots at the bottom show country results. The data is included in the supporting information (SI-1). All results refer to the year 2008.
Variations in the RTB and the RME of exports can be interpreted in light of the different allocation logics i.e. value-based vs. physical allocation. When calculating RMC with EXIOBASE, raw material extraction is distributed to final demand following the monetary transactions between sectors and regions. Crude ore is allocated to all monetary supply chains, including the ones of for example service sectors that mostly serve domestic final demand. In contrast, the gPIOT-based RMC calculus follows a physical allocation based on the apparent physical flows between processes and regions. Consequently, the gPIOT calculus allocates crude iron ore exclusively to physical supply chains that distribute ferrous materials. The observed differences in the RME of exports suggest that with EXIOBASE a significant share of the extracted iron ore is allocated to domestic supply chains, which do not physically distribute iron and steel.

3.3. International trade in iron-steel supply chains

A major strength of the gPIOT framework is its ability to provide detailed sectoral information on both apparent/actual/direct physical flows and virtual/embodied/indirect flows (footprint-type indicators). To exemplify this, we provide a brief overview over global iron-steel supply chains and the scale of international trade flows. The following analysis is based on the gPIOT for the year 2014. To visualize the different processes and flows, a Sankey diagram is shown in Figure 7-A which is derived from an aggregated version of the gPIOT.

In 2014, the world consumed 1333 Mt of steel in final products, i.e. gross addition to material stocks. With ca. 345 Mt EoL-Scrap withdrawn from existing material stocks, this translates into a global net-addition to material stocks of 988 Mt. Construction is the most important sector, accounting for 59.5% of global steel consumption or 793 Mt, followed by machinery with 173 Mt. Together with motor vehicles (151 Mt), these three sectors account for 84% of global steel consumption.

With regard to international trade, we find that 13.5% of all the steel for final use is not manufactured within the country where the products are finally consumed. The construction sector stands out, as it sources only around 0.6% of its steel products for final use from international markets. On the other hand, manufactured products such as machinery and motor vehicles have trade shares of 32 and 31% respectively (see Figure 7-C). Products nec – i.e. office and electrical machinery, computers, communication equipment and other appliances - have trade shares of 39%. The largest trade share of all the flows in iron-steel supply chains is found for iron ore. 59.5% or

14 Please note that the exact numbers of these trade shares are dependent on the regional classification of the base table. A gPIOT with more disaggregated world regions would yield higher trade shares.
1209 Mt of iron ore is distributed via international markets (compare with the Sankey in Figure 7-A). Steel scrap (output of scrap preparation) has a trade share of 23%, followed by finished steel (output of rolling & forming) with 16%. Castings and products of reduced iron have smaller trade shares of 8 and 5% respectively.

The gPIOT can also be used to quantify upstream material inputs on the sector level. This is shown in Figure 7-B, which depicts the RMC and material multipliers reflecting the amount of reduced iron and steel scrap that are directly used in steelmaking (BOF and EAF) and indirectly required to deliver one tonne of steel for final use. The sector with the largest upstream scrap input per tonne is construction. To produce 1 tonne of manufactured steel for construction, approx. 0.6 tonnes of scrap and 0.93 tonnes of reduced iron are required. All other sectors show smaller scrap multipliers of less than 0.5 t/t, which points toward the fact that steel scrap is usually down-cycled (scrap from motor vehicles are used for construction) because of impurities (with other metals or materials).

In terms of RMC, we find again construction to be the most important sector with 1958 Mt of crude ore, followed by machinery (408 Mt), products nec (331 Mt) and motor vehicles (332 Mt). The regional distribution shows that ca. 53% of crude ore is embodied in international supply chains. Motor vehicles stand out with a trade share of embodied crude ore of 66%. Moreover, large upstream, i.e. indirect trade dependencies, are found for the construction sector (trade share of 50%), which stands in stark contrast to the trade share of 0.6% for manufactured construction materials (compare Figure 7-B with 7-C).
Figure 7-A: A Sankey diagram showing the various types of flows and processes that are included in the iron-steel gPIOT. The flows of the modelled core system are grouped into international trade flows (yellow color), where the two processes that are connected by the flow are situated in different regions, and domestic flows (dark blue color), where the two processes are in the same country. Figure 7-B shows for the different manufactured products indirect/upstream material inputs, which includes RMC (y-axis on the left) and material multipliers (y-axis on the right) reflecting the amount of reduced iron and steel scrap that are directly used in steelmaking (BOF and EAF) and indirectly required to deliver one tonne of steel for final use. RMC results are disaggregated into crude ore that is embodied in the final consumption of the country where it is extracted (domestic/in-country flow in dark blue color) and the ore that is embodied in international supply chains (yellow color). Figure 7-C shows steel for final use i.e. gross addition to material stocks by different manufactured products, which is also depicted on the far right side of the Sankey. Steel in final products that is internationally traded is shown in yellow color. Flows that are situated within regions are shown in dark blue color. The data is included in the supporting information (SI-1). The results refer to the year 2014.
4. DISCUSSION AND OUTLOOK

The iron-steel gPIOT presented here exemplifies how the PIOLab combines the strengths of different modelling schools and accounting frameworks commonly used in Industrial Ecology to synthesize a model that has the potential to be more comprehensive and informative than any individual approach.

Using an IO model in physical units makes footprint-type indicators insensitive to price inhomogeneity and fluctuations, including differences in taxation schemes and subsidies, which distort estimates of actual physical flows between sectors when a monetary allocation logic is applied. The comparison of results from the gPIOT and the monetary MRIO EXIOBASE revealed large differences in RMC indicators, especially for important iron ore and steel producing/consuming countries like Australia, Brazil and China. The present gPIOT yields more robust RMC indicators that are mass-balanced and thus better aligned with other ew-MFA headline indicators (such as DE and DMC) since they can be derived from the same integrated top-down database. The ability to generate a consistent set of indicators on different levels of aggregation is one of the main strengths of the gPIOT. The model can deliver highly aggregated information on the country level, but global material flows can also be disaggregated into different sectors and regions, which allows opening up the black box of ew-MFA and thereby widening its applicability for policy-oriented analysis that focuses on specific economic activities. For example, the analysis of upstream scrap inputs per tonne of final product revealed how scrap demand is strongly driven by the manufacturing of construction materials. Due to the explicit consideration of scrap supply and use, gPIOTs can provide a practical framework for analyses that aim at closing globalized material cycles.

The diagnostic tests of the base table built revealed a good level of adherence between reported data and the values realized in the gPIOT, comparable to constraint realizations in other IELabs (Wakiyama et al., 2020). Conflicts between different data sources are unavoidable when constructing multi-regional IO tables. In the optimization run presented here, we have assigned higher standard errors (SE) to trade data than to production data in order to give priority to a better realization of the latter due to two reasons. Firstly, trade data points are several orders of magnitude smaller than production data points and we assume, according to the law of large numbers, that the relative SE of for example crude steel production in China (512 Mega tonnes in the year 2008) tends to be smaller than that of welded tubes traded between Hungary and South Africa (1.1 t in 2008). Secondly, BACI reports only gross trade that includes re-exports, which are not present in an MRIO framework. We compared raw data of production and trade to reveal the scale of iron ore re-exports. According to the UNEP-IRP database, only 48 countries extract iron ore but BACI lists more than 100 exporting regions. The mismatch between trade and other data sources is relaxed by tagging trade data with relatively higher SE.
The flexibility of the PIOLab enables Industrial Ecologists to create large-scale models in a cost-efficient way that are tailored to specific needs. All data mapped to the root classification can be re-used when compiling new versions of the iron-steel gPIOTs in the future. New data is added with data feeds to formulate new constraints. For example, researchers might want to construct gPIOTs that reflect other data for iron ore extraction (Schaffartzik et al., 2014) or EoL-Scrap supply (Myers, Reck, & Graedel, 2019). Finished steel products can be broken down into different grades, such as stainless and (non-)alloy steel, using additional data sets or proxy information for disaggregation (Reck, Chambon, Hashimoto, & Graedel, 2010). Another possibility is to construct gPIOTs that align with reported official statistics of a particular country of interest. Similar to the SNAC (Single Country National Accounts Consistent) method which prepares a global MRIO that respects a given national monetary IO table (Edens et al., 2015), the PIOLab can incorporate more country-specific material flow data - e.g. steel yearbooks of China (CNKI, 2018) - to construct a gPIOT that is customized for policy applications for a certain region. In the same vein, modelers can impute further regional information on process descriptions (input-output ratios), yield factors (Wang, Jiang, Geng, Hao, & Zhang, 2014; Wang, Müller, & Hashimoto, 2015) or the supply and use of by-products like furnace slag (Rieger & Schenk, 2019) to compile global tables that better reflect regional differences.

Data infrastructures and related tools have developed impressively over the last decade in Industrial Ecology. Data collections such as the Yale Stocks and Flows database (Myers et al., 2019), the building material composition database (Heeren & Fishman, 2019), the UNEP-IRP ew-MFA accounts (Schandl et al., 2018) or the Industrial Ecology Data Commons (Pauliuk, Heeren, Hasan, & Müller, 2019) represent a cumulative body of knowledge that the PIOLab can build on in the future. The efficient reuse of such data collections for constructing gPIOTs for various material and energy flows requires the development of a universal root classification that covers all socio-metabolic processes and flows, which are described therein. Because the same flow could be measured in different units (mass, energy content or monetary value), the multi-layered SUT approach as proposed by Merciai (2019) would provide a suitable integration framework where relations between layers are imputed via ratio constraints which reflect calorific values (Joule/kg) and prices (Euro/kg). Such a multi-layered\textsuperscript{15} socio-metabolic root classification would constitute a model-independent reference system with clear and unambiguous definitions across modelling schools and accounting frameworks. This is very much in line with ongoing debates in the Industrial Ecology community regarding standardization and harmonization of data exchange and processing (Pauliuk, 2020; Petavratzi et al., 2018) as well as efforts to foster transparency and reproducibility of modelling results (Hertwich et al., 2018).

\textsuperscript{15} As a matter of fact, IELabs already employ such multi-layer capabilities by including different valuation sheets (basic prices, transport and trade margins).
An IELab capability that the PIOLab has not tapped so far\textsuperscript{16} is the possibility to create subnational and hierarchically nested multi-geography IO tables that disaggregate economy-wide material flows into different states and local regions. Because environmental and socioeconomic conditions can vary significantly within countries, IO models should aim to move from the aggregated national to a more detailed spatial level (Sun, Tukker, & Behrens, 2019). IELabs facilitate different regionalization methods including location quotient approaches, where national IO tables are disaggregated with proxy information to produce an initial estimate of the subnational table for the reconciliation routine (Flegg, Mastronardi, & Romero, 2016; Jahn, 2017; Jahn, Flegg, & Tohmo, 2020). Nesting such a subnational multi-regional table into a gPIOT would enable biophysical assessments of global supply chains that take into account the heterogeneity of environmental pressures and impacts on the subnational level (Moran, Giljum, Kanemoto, & Godar, 2020). This multitude of possibilities to refine, adjust and expand gPIOTs make the PIOLab a very promising tool for addressing a wide range of sustainability-related challenges in the future.

ACKNOWLEDGEMENTS

This project has received funding from the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation programme (FINEPRINT project, grant agreement No. 725525). This work was also financially supported by the Australian Research Council (ARC) through its Projects DP0985522, DP130101293, DP190102277, and LE160100066, as well as the National eResearch Collaboration Tools and Resources project (NeCTAR) through its Industrial Ecology Virtual Laboratory infrastructure VL 201. The authors thank Sebastian Juraszek for expertly managing the IELab’s advanced computation requirements.

\textsuperscript{16} Two other noteworthy IELab features exist that are relevant for material flow analysis. Firstly, the resultant IO tables reconciled by AISHA are accompanied with standard deviations that can be used for uncertainty assessments. In general, uncertainty analysis has received increasing attention in MFA Lupton and Allwood (2018); Laner, Rechberger, and Astrup (2014). The standard deviations from AISHA can be used for Monte-Carlo Simulations and to estimate confidence intervals of derived indicators see for example Moran and Wood (2014). Secondly, IELabs are increasingly used for disaster analysis. Faturay and colleagues (2020) quantified spillover effects resulting from different earthquakes and typhoons in Taiwan. Lenzen and colleagues (2019) quantified indirect economic damages from a tropical cyclone in Australia. Another recent study by Lenzen and colleagues (2020) estimated global economic impacts of the COVID-19 pandemic. Using the PIOLab for biophysical disaster analysis represents an interesting new research avenue.
The supporting information includes all data needed to reproduce the figures (SI-1), the domestic supply tables (SI-3) and the domestic use tables along with aggregated trade flows (SI-4) for the year 2008. Moreover, base and root classifications together with the root-2-base region aggregator can be found in SI-2. A GitHub repository with R scripts is openly available, allowing users to reproduce the figures and perform IO analysis with the gPIOT (www.github.com/fineprint-global/PIOLab). The time series of the gPIOTs (2008-2017) can be downloaded from Zenodo (https://zenodo.org/record/4385975).

REFERENCES


Proceedings of the National Academy of Sciences of the United States of America, 114(8), 1880–1885. https://doi.org/10.1073/pnas.1613773114


Myers, R. J., Reck, B. K., & Graedel, T. E. (2019). Ystadfdb, a unified database of material stocks and flows for sustainability science. *Scientific Data, 6*(1), 84. https://doi.org/10.1038/s41597-019-0085-7


