Relaxing the import proportionality assumption in multi-regional input-output modelling

Simon Schulte*,†, Arthur Jakobs and Stefan Pauliuk

Abstract

In the absence of data on the destination industry of international trade flows most multi-regional input-output (MRIO) tables are based on the import proportionality assumption. Under this assumption imported commodities are proportionally distributed over the target sectors (individual industries and final demand categories) of an importing region.

Here, we quantify the uncertainty arising from the import proportionality assumption on the four major environmental footprints of the different regions and industries represented in the MRIO database EXIOBASE. We randomise the global import flows by applying an algorithm that randomly assigns imported commodities block-wise to the target sectors of an importing region, while maintaining the trade balance.

We find the variability of the national footprints in general below a coefficient of variation (CV) of 4%, except for the material, water and land footprints of highly trade-dependent and small economies. At the industry level the variability is higher with 25% of the footprints having a CV above 10% (carbon footprint), and above 30% (land, material and water footprint), respectively, with maximum CVs up to 394%.

We provide a list of the variability of the national and industry environmental footprints in the online SI so that MRIO scholars can check if a industry/region that is important in their study ranks high, so that either the database can be improved through adding more details on bilateral trade, or the uncertainty can be calculated and reported.

Keywords: import proportionality assumption; uncertainty; environmentally-extended multi-regional input-output; carbon footprint; land footprint; water footprint; material footprint; footprint analysis

Introduction

Multi-regional input-output modelling (MRIO) is widely applied to study the relationship between economic activities and their upstream environmental, social and economic impacts [1, 2]. To do so, MRIO links national inter-industry accounts with international trade data [3]. Although extensive inter-regional trade information is available, they often lack the required level of detail needed to compile MRIO tables without having to rely on strong assumptions [4, 5]. The compilation of MRIO tables requires information on the bilateral trade between industry $i$ in region $a$ and target sector $j$ in region $b$ for all regions and industries covered by the MRIO. The target sector can either be an individual industry (intermediate consumer) or a final consumer (typically distinguishing households, government and capital formation). However, data on imported inputs are not available at the level of individual in-
dustries and final demand categories but by “Broad End-use Category” (BEC) [6]. As the name implies the BEC classification system only broadly distinguishes between intermediate consumption, household consumption, capital goods and mixed end-use (products where the end-use is unclear e.g. cars can be purchased both for household consumption and as capital goods).

In effort to overcome this lack of information, MRIO compilers proportionally distribute the imported commodities over the target sectors in the importing region, which is often referred to as the “proportionality assumption” [4, 7]. The proportionality assumption underlies all current MRIO tables, however the level at which the proportionality assumption is applied varies. EXIOBASE [8], Eora [9] and GTAP [7] for instance allocate imports to the target sectors without the differentiation in intermediate use, consumption, and capital (i.e. not using BEC data) thus using the same proportions for all target sectors be it industries or final consumers. WIOD, by contrast, take the BEC data to distribute imports to aggregate end-use sectors but then also use the same proportions for all industries and final demand categories, respectively [5]. One reason why only WIOD uses the BEC data is that the BEC data by no means covers all industries, countries and years at the resolution required to produce such high resolution MRIOs as EXIOBASE or Eora.

Although the assumption that imports are distributed proportionally among individual industries and end-consumers might provide a practical solution to the lack of more detailed data, the assumption might be flawed for several reasons, possibly biasing the outcome of MRIO studies. In the following, we first address some aspects that could lead to ‘real’ import shares differing significantly from the proportionality assumption. We then work out the conditions that must be fulfilled for the proportionality assumption to lead to a bias in the MRIO results.

Why the proportionality assumption might bias MRIO-based footprints

One reason why the proportionality assumption might be flawed is the aggregation bias: Due to the aggregation of firms to broader industry sectors, one such a sector might include rather heterogeneous products (in terms of physical properties and/or prices) [10]. In EXIOBASE for example, firms producing raw unfabricated leather and those producing luxury leather handbags are aggregated into one ”Leather and leather products” sector [8]. Imagine three countries: Country a and b both export leather and leather products to country c. However, a is specialised in raw leather, while b exports mainly luxury leather handbags. In country c two industries i and j use imported leather, however, sector i buys raw leather from country a (to further process it), while sector j imports luxury leather handbags from country b (to retail them). However, since this knowledge on the exact use of imports is not available for MRIO compilers, they assume that for both industries i and j the shares of imported leather products coming from country a and b, respectively are the same.

Depending on the application case this assumption of proportional import shares could lead to biases in the outcome of MRIO studies if further conditions are met: If a MRIO is used to study the environmental implications at the level of industry sectors (e.g. study the footprint of sector i [11, 12], or of a consumption basket [13, 14]), the proportionality assumption biases the outcome when there is a large variation in the impact intensities between the industries of the exporting countries.
Presumably, the "Leather and leather products" sector in country \(a\) exporting cheap raw leather has a higher carbon emission intensity (i.e. emissions/Euro) than the same sector in country \(b\) producing expensive luxury leather handbags. Hence sector \(i\)'s carbon footprint would be underestimated because the upstream emissions related to \(i\)'s use of imported leather products are averaged among both importing countries, instead of taken from the carbon intense leather sector of country \(a\).

If a MRIO practitioner is interested in the environmental implications at the national level (e.g. [15]), a large variation in the impact intensities between the industries of the exporting countries, however, is not yet sufficient. Then additionally the products of industries \(i\) and \(j\) have to differ in terms of what proportion is consumed domestically and what proportion is exported. If industry \(i\) (importing from \(a\)) mainly produces for export, while \(j\)'s luxury leather handbags are mainly for the domestic market, the proportionality assumption would lead to an overestimation of \(c\)'s national carbon footprints because the upstream emissions related to industry \(j\)'s use of imported leather products would partly incorporate emissions which are actually linked to \(i\)'s leather imports (which should not show up in the footprint of country \(c\) since \(i\) produces for export).

There exist also other reasons why imports might not be distributed proportionally among target sectors. These are, for example, geographical reasons, e.g. the location of firms from one sector closer to large harbours or towards the border to a neighbour country, or historical grown trade relations between firms in different countries.

**Literature review, research gap & research question**

Even though several MRIO compilers mention the potential problems related to the proportionality assumption [8, 7, 9, 5], so far only few studies approached the problem quantitatively. Puzzello (2012) investigate the effect of the proportionality assumption on the factor content (capital, labour, services, ...) of trade [16]. The author compares the results of two different methods to compile the Asian MRIO table, one which applied the proportionality assumption to trade flows, and another survey-based including bilateral details on trade. She finds the net bias introduced by the proportionality assumption to be small "only because the biases on exports and imports of factor services cancel each other out".

In their working paper Milberg and Winkler (2010) studied the error from the proportionality assumption on the estimate of the effect of offshoring on the German labour demand for 36 sectors [17]. They estimated the effect of offshoring applying the same econometric model but with two distinct data sets: one based on the proportionality assumption and one with additional details on the use of German imports. They found a large difference in the coefficient estimates between the two versions, in many cases even with reversed signs.

Feenstra and Jensen (2012) did a similar comparison for the estimates of material offshoring from US manufacturing [18]. They calculated the shares of imported intermediate inputs of individual manufacturing sectors both with and without applying the proportionality assumption. They found a high correlation between the offshoring shares calculated with the distinct methods.

Recently, Jiang et al. (2020) compared the material footprints of China and Chinese provinces based on two approaches: one with the assumption of proportional
provincial import shares, and one with the inclusion of detailed data on Chinese
inter-provincial trade [19]. They found that the Chinese national material footprint
are not significantly influenced by the choice of methods. Across provinces, however,
the authors found variations in the material footprints between the two methods in
the range from -9% to +14%. For disaggregated materials the differences between
both methods were even ranging between -48% and +34%.

From the literature review we identify two major research gaps. First, all studies
so far are regional specific, i.e. they only investigate how the calculated impacts
change when including bilateral trade details for one country/region. Second, all
but one study investigate economic effects (offshoring, factor content), with only
Jiang et al. (2020) considering an environmental impact indicator. The effect of
the proportionality assumption on other environmental indicators, such as carbon,
land or water footprints, however, have not be investigated so far. Which is surpris-
ing given that environmental footprinting has been a major field of application of
MRIOs in the last decades [2, 20, 21, 22].

Against this background, we test how sensitive environmental footprint estimates
are towards changes in the allocation of import flows not only for individual regions
but for the entire world trade. Since we do not know the true import shares we ran-
domise the allocation of import flows to different target sectors while keeping fixed
the available information, namely (i) the total imports per country and product,
and (ii) the total use of product import per target sector. The aim of our study is
to quantify the maximum uncertainty introduced by the proportionality assump-
tion. We focus on the four most commonly used types of environmental footprints
(carbon, material, land, and water) on two levels (country and industry).

Material and Methods

We use EXIOBASE (Version 3.4), a MRIO database that is based on the propor-
tionality assumption and is widely used for environmental footprint analysis [8]. We
use the industry-by-industry table for the year 2011 in current prices in its original
resolution (163 industries, 44 countries + 5 RoW regions). To quantify the uncer-
tainty of environmental footprints introduced by the proportionality assumption
we undertake the following steps: We generate 4897\(^{[1]}\) MRIO tables matrices with
globally randomised import allocations. With each of these new MRIO tables we
then calculate national and industry footprints and investigate the variability of
these footprints.

Generating a MRIO table with globally randomised import allocations

We first show our approach to randomise the allocation of the imports of one product
to the target sectors in one country. To generate a entire new MRIO table with
globally randomised import allocations this procedure has to be repeated for all
industries and countries covered by the MRIO.

We refer to the sets of exporting regions as \(R\), exporting industries as \(I\), importing
regions as \(S\) and importing target sectors (comprising industries and final demand

\(^{[1]}\) The number results from the maximum run time of 40 hours allowed by the
BWunicluster which was used for the analysis.
categories) as $J$. Matrices (capital letters) and vectors are represented as bold characters. For demonstration in this paper, we use a simple MRIO system with four regions, four industries/products and two final demand categories to exemplify how the imports of product $i \in I$ (say "leather products") from the exporting countries $R$ are randomly allocated to the six target sectors $J$ of region $s \in S$ (say Germany) (see Figure 1). Since the proportionality assumption concerns only the inter-industry matrix $Z$ and the final demand matrix $Y$ at that stage we omit the other elements of typical EE-MRIO tables (primary inputs, total output, environmental extensions).

The problem we are facing is the allocation of the import flows of a given good (here: $i = \text{leather products}$) from different countries $R$ to different target sectors $J$ in a given country (here: $s = \text{Germany}$), where both (i) the total amount of imports by each exporting country $r \in R$, and (ii) the total amount of imported inputs for each target sector $j \in J$ are known. This problem can be represented in the form of a matrix (the "import matrix" $T_{R,s,i,J}$), where both the row sums $s_{Rsi}$ ( = imports of product $i$ to region $s$ by region of origin $R$) and column sums $u_{siJ}$ ( = inputs of product $i$ by target sector $J$ in region $s$) are known, but cell entries are not. Formally expressed we know thus:

$$\sum_{j \in J} t_{Rsij} = s_{Rsi} \quad (1)$$
$$\sum_{r \in R} t_{rsiJ} = u_{siJ} \quad (2)$$

Figure 1 shows how we extract the import matrix $T_{R,s,i,J}$ from the inter-industry matrix $Z$ and the final demand matrix $Y$. Using our example from before, the import matrix $T_{R,s,i,J}$ depicts the flows of imported leather products (i) from different regions $R$ to different target sectors $J$ in Germany ($s$). For the sake of readability, we omit the subscripts in the following. Summing $T$ row-wise we get the vector of import flows $s$ depicting the imports (supply) of leather products from different exporting regions to Germany (Equation 1). Summing $T$ column-wise we get the vector $u$ depicting the use of the imported leather products in different industries and final consumption categories in Germany (Equation 2).

Now, the aim is to randomly allocate the region-specific supply $s$ to the industry-specific use $u$. Thus, we want a ‘new’ randomised import matrix $T'$. We follow the compilers of the most prominent global MRIOs EXIOBASE [8], Eora [9] and GTAP [7] and do not - unlike WIOD [5] - include information on the BEC.

We apply an algorithm to randomly allocate $s$ to $u$ block-wise which works as follows (see Figure 2 A-C, a pseudocode version of the algorithm can be found in Additional file 2):

**Step 1:** We start by taking the supply of region 1 ($s_1 = 1$st element of $s$) and the use of a randomly chosen target sector $i$ ($u_i = i$th element of $u$).

**Step 2:** Now we differentiate three cases: If the supply from country 1 equals or is smaller than the use of industry $i$ (case 1 or 2, respectively) we allocate the entire supply of country 1 to industry $i$. If, however, the supply of country 1 is larger than
the use of industry \(i\) \((\text{case 3})\), the fraction of country 1’s supply which equals the entire use of \(i\) is allocated to \(i\). In Figure 2 A-C these three cases are illustrated.

**Step 3** is depending on which case has occurred in the previous step:

- In case 1, both the entire supply of country 1 and the entire need of industry \(i\) have been accounted for. Thus, we can go over to the next country 2 and compare its supply with the next randomly chosen industry following the procedure described under step 2 and 3.
- In case 2, the entire supply of country 1 has been accounted for but not the need of industry \(i\). Thus, we go over to the next country 2 and compare its supply with the remainder of industry \(i\) following the procedure described under step 2 and 3.
- In case 3, the entire need of industry \(i\) is met, but country 1 still has imports left. Therefore, we continue with the next randomly chosen industry (in our example \(p_4\)) and compare its need with the remainder of country 1’s supply following the procedure described under step 2 and 3.

We run this algorithm until the supplies of all countries have been accounted for and the needs of all industries have been met. This condition will certainly be reached since all trade flows in MRIO tables are balanced, i.e. the total imports of product \(i\) into region \(s\) equals the total use of imported inputs \((\sum_r s_{rsi} = \sum_j u_{sij})\).

Carrying out the above outlined procedure for the imports of each product into each region results in ‘new’ matrices \(Z_{new}\) and \(Y_{new}\) where all imported products into all countries are randomly allocated to the target sectors while keeping fixed both, (i) the total imports per country and product, and (ii) the total use of product import per industry sector.

Our approach is strictly speaking not a randomisation, since we do not consider all possible versions of the import matrices. In the lack of knowledge on bilateral trade details we should have to assume that each of the theoretically infinite possible versions of this import matrix is evenly likely. However, with our algorithm we only consider the extreme versions of the import matrix. With ”extreme” we mean that our algorithm produces import matrices where imports are bundled and allocated block-wise to the target sectors (Figure 2 A-C). Thus, we miss all versions of the import matrix \(T\) where the import flows from different regions are split and distributed over a large number of target sectors (i.e. all target sectors import a bit from country \(a\), a bit from \(b\) and so on, Figure 2 E).

So instead of randomly sampling out of all possible versions of the import matrix, we only sample out of all extreme ones. Given the number of repetitions to be limited by computational issues - in our case to 4897 repetitions - with our approach we increase the probability to capture the extreme ends of the ”real” distribution of the uncertainty of the respective footprint. Thus, we come closer to an estimate of the maximum possible uncertainty of the respective footprint which is the aim of our study.

**Calculating environmental footprints**

We calculate the four most used environmental footprints: carbon, land, material and water [23]. Following Steinmann et al. (2018) [23] we define these footprints as the consumption-based ...
... emissions of the GHGs CO\textsubscript{2}, CH\textsubscript{4}, N\textsubscript{2}O, SF\textsubscript{6}, hydrofluorocarbons (HFC), and perfluorocarbons (PFC) weighted by their global warming potential based on a time horizon of 100 years [24] (carbon footprint)
... area of land required by forestry, agriculture, infrastructure, etc. (land footprint)
... mass of all used extractions including metal ores, others minerals, wood, fish, and crops (material footprint)
... volume of the total blue water consumption (water footprint).

To calculate the environmental footprints at the national and industry level we use \(Z_{\text{new}}\) and \(Y_{\text{new}}\), along with the stressor matrix \(S\) containing the relative environmental impacts per unit of output, the characterisation matrix \(C\) that weights the environmental impacts according to the four footprints, and the output per sector \(x\), all provided by EXIOBASE [8, 1].

We first calculate the Leontief inverse matrix \(L\) as
\[
L = (I - Z_{\text{new}} \hat{X}^{-1})
\]
where \(I\) is the identity matrix and \(\hat{X}^{-1}\) is a square matrix with \(1/x_i\) on the main diagonal and zeros elsewhere.

We calculate industry footprints in the form of industry footprint intensities \(E_{ifi}\) defined as a sector’s footprint for producing one monetary unit of its product
\[
E_{ifi} = CSL
\]
and total industry footprints \(E_{ift}\) defined as a sector’s footprint for the entire production that is finally demanded:
\[
E_{ift} = E_{ifi} \hat{y}
\]
where \(\hat{y} = \sum_j y_{ij}\) is the total final demand by product.

While for the analysis of the footprint variability we use \(E_{ifi}\) (thus avoiding one matrix multiplication; even though results would be the same as the total final demand \(y\) is a constant), we take \(E_{ift}\) as a measure of a industry footprint’s relevance to put the footprint’s variability into perspective.

National footprints \(E_{nf}\) we calculate as
\[
E_{nf} = E_{ift} Y_{\text{new}}
\]

Footprints calculated with the standard MRIO table, thus based on the proportionality assumption (by replacing \(Z_{\text{new}}\) with \(Z\) and \(Y_{\text{new}}\) with \(Y\) in Equation 3, 5 and 6) are noted as \(E_{ift}^{\text{ImpProp}}\) and \(E_{nf}^{\text{ImpProp}}\), respectively. Note that we do not include direct (Scope 1) emissions since they are invariant of the proportionality assumption.

We carry out 4897 MC runs, thus resulting in samples of 4897 different carbon, water, land and material footprints. We use two measures to quantify the variability within these samples:
The coefficient of variation (CV) is used as a measure of the relative variation:

\[ CV = \frac{\sigma}{\mu}, \tag{7} \]

where \( \sigma \) is the standard deviation of the sample and \( \mu \) the mean.

Since the CV does not give any information about the exact appearance of a distribution (e.g. its skewness, number of modes, etc.), we take a closer at the sample distributions for some interesting industries/nations by looking at their probability density function and describing their variability in terms of their 2.5th and 97.5th percentiles, i.e. below which 2.5% and 97.5% of the 4897 simulation results fall.

**Results**

We present the sensitivity of different environmental footprints when relaxing the proportionality assumption on two different levels: First, at the level of nations, and second, at the level of industries.

**National Footprints**

Figure 3 A shows the relative variability (CV) of the national footprints compared to the size of the footprints assuming proportional import shares. The points and country labels are coloured by the percentage of the footprint which is sourced from imports.

Overall, the variability of the carbon footprints is lower on average compared to the other three footprint types (boxplots). Footprints with a high import share generally show a higher variability. Since regions with a large footprints (e.g. China, USA or India) mostly have a low import share they also show a lower variability as compared to most of the regions with smaller footprints.

The highest variability in carbon, material and water footprints can be seen in Luxembourg (LUX) with CVs of approximately 0.02 (carbon) and 0.06 (material and water). Taiwan (TWN) shows the highest variability for the land footprint with a CV of close to 0.07. The percentage of these footprints sourced from imports is 80% (LUX, carbon), 96% (TWN, land), 99% (LUX, material) and close to 100% (LUX, water). Other regions of interest with a relatively high CV and a relevant footprint size are - in the case of carbon footprints - Switzerland (CHE) and Belgium (BEL) with CVs of 0.016 and 0.011, respectively. In the case of land footprints, regions worth mentioning are the Netherlands (NLD, CV of 0.038) and Belgium (BEL, CV of 0.029), which both have a high population density and are highly dependent on imports of land-intensive food products and materials. RoW Asia and Pacific (WWA) stands out when looking at its material footprints having the third largest absolute material footprints with more than 5000 Mt and the second highest variability with a CV of 0.047. In the case of water footprints, interesting regions are again the Netherlands (CV of 0.06) and, less pronounced, Belgium (CV of 0.031).

Figure 3 B shows the distribution of the 4897 simulated national footprints exemplary for some selected regions (the one earlier mentioned). For distributions of all national footprints please refer to Figure S1 in the Additional file 1. The footprints where normalised by dividing each sample by the mean of all 4897 samples. The
violin plots show the probability density and the range of all samples. The boxplots show the inter-quartile range (IQR) where 50% of all samples are situated (boxes), the sample’s median (horizontal line) and the range from the 2.5th to the 97.5th percentile (whiskers). The colours indicate the actual size of the footprints on a log scale according to the colour bar given on the right. The red points show the footprints calculated with the standard version of EXIOBASE (which is based on the proportionality assumption), also normalised by the mean of all samples.

The 95% confidence interval (CI) of Luxembourg’s carbon footprint ranges between -4% around the mean. Both, Luxembourg’s material and water footprint distribution are skewed towards higher values with a range between -9% and +14%, and -10% and +12%, respectively. Taiwan’s land footprint ranges between -11% and +10% around the mean (95% CI). Taiwan’s land footprint and WWA’s material footprint both show multi-modal distributions, i.e. they have several modes (local maxima in their density function).

The difference between the footprint using the proportionality assumption and the sample mean, is particularly pronounced for the material footprints of Luxembourg (+10%), WWA (-8%) and Belgium (+6%). The three countries exemplary listed with their water footprint show a difference around +3 to +6%, while the countries listed with their carbon and land footprints show only a comparably smaller deviance.

Industry footprints
For the industry footprints (Figure 4) we see a similar pattern as for the national footprints. The carbon footprints are less variable on average than land, material and water footprints (boxplots). Industries with a higher variability (CV) also have a higher import share. However, a high import share does not necessarily go hand in hand with a higher CV (compare Figure S3 in the Additional file 1). With maximum CVs between 1.87 (land footprint) and 3.94 (material footprint) the variability at the level of industry sectors is significantly higher than at the regional/national level.

Industries of interest, i.e. with a relatively high CV and a relevant footprint size are - in the case of the carbon footprint - the 'Electricity by biomass and waste' (i40.11.g) sectors in Portugal (PRT) and India (IDN) with CVs of 2.19 and 1.66 respectively. In the case of land footprints, the sector ‘Other non-ferrous metal ores and concentrates’ (i13.20.16) in Germany (DEU, CV of 1.87), Italy (ITA, CV of 1.70) and Greece (GRC, CV of 1.62) stand out. Industries of interest in their material footprint are the Indian ‘Electricity by petroleum and other oil derivatives’ sector (IDN: i40.11.f) with a CV of 3.94, the Brazilian ‘Electricity by coal’ sector (BRA: i40.11.a) with a CV of 3.60 and the Turkish ‘Electricity by wind’ sector (TUR: i40.11.e) with a CV of 3.27. In the case of water footprints the sectors with a prominent role are the ‘Wool, silk-worm cocoons’ sectors (i01.o) in Ireland (IRL) and - less pronounced - Finland (FIN) with CVs of 2.05 and 1.77, respectively, the Chinese ‘Sale, maintenance, repair of motor vehicles, motor vehicles parts, motorcycles, motor cycles parts and accessories’ sector (CHN: p50.a) with a CV of 1.92, the Slovenian ‘Processing vegetable oils and fats’ sector (SVN: i15.e) with a CV of 1.98, and the German ‘Retail trade services, except of motor vehicles and motorcycles;
repair services of personal and household goods’ sector (DEU: i52) with a CV of 1.69.

When taking a closer look at the distributions of the 4897 simulated industry footprints for selected industries (Figure 5), we see that most distributions are multi-modal, i.e. they have two or more local maxima in their density function. Additionally, most distributions have a positive skew.

The most extreme upward deviation can be seen for the Portuguese ‘Electricity by biomass and waste’ (PRT: i40.11.g) sector’s carbon footprint with a 97.5th percentile of 1000%. Since the footprint calculated with the default version of EXIOBASE is even a bit below the sample mean, this finding implies that this sector’s carbon footprint can be more than 10 times higher than the current EXIOBASE version’s result. The largest downward deviation can be seen for the water footprint of the Italian ‘Mining of other non-ferrous metal ores and concentrates’ sector with a 2.5th percentile of 5%, implying a possible underestimation of this footprint by a factor of 20.

Discussion

Most published MRIO-based footprints come with no uncertainty by default. Given the high potential uncertainty due to several assumptions made in the process of compiling MRIO tables [25] this is problematic if these results are used for or influence decision making, as the robustness of the decision in relation to the footprint information used cannot be assessed due to the lacking uncertainty of the latter [26]. This piece quantifies the effect of one assumption underlying all global MRIOs - the proportionality assumption - on the four major environmental footprints of the different EXIOBASE regions and industries.

At the country level, the variability is in general below a coefficient of variation (CV) of 4%, except for the material, water and land footprints of highly trade-dependent and globally very small (Luxembourg, Malta, Latvia, Lithuania) and small economies (Belgium, Netherlands), and the RoW-region ‘Asia and Pacific’. More serious problems, however, are to be expected when interpreting the geographic locations of national footprints (such as e.g. [15, 27]).

At the industry/product level we find the bias introduced by the proportionality assumption to be substantially higher as compared to national footprints. 25% of the industry footprints covered by EXIOBASE show a CV above 10% (carbon footprint), above 30% (land and material footprint), and above 34% (water footprint), respectively. Some industry footprints, show a possible bias of 1000% or more indicating that assuming proportional import shares might lead to over- or under-estimation of these footprints by a factor of 10 or more. Our findings thus confirms that MRIO-based footprints at the industry/product level need to be treated with great care [25].

The greater variability at the industry level compared to the national level can be explained by the fact that national footprints are the sum of a multitude of industry footprints. Hence, at the more aggregated level of nations variability at the industry level will partly cancel out each other. Moreover, as elaborated in the introduction, for the industry footprints to be biased by the proportionality assumption only one condition has to be met (a large variation in the impact intensities between
the industries of the exporting countries), while in the case of national footprints a second condition has to be satisfied (a large variation in the proportions of the importing industries that is consumed domestically).

The sample distributions of most industry footprints and some national footprints (Taiwan’s land footprint, WWA’s material footprint) are multi-modal, i.e. they have two or more local maxima in their density functions. One explanation could be that the variability of these footprints depends very much on the allocation of imports of only one (or few) products. Changing the allocation of only this (these few) product imports leads to a large change in the footprint (local maxima) while the allocation of all other global import flows has only a minor influence (little dispersion around the local maxima). Further research could zoom into these footprints, e.g. via a structural path analysis/decomposition, to determine the origin of the uncertainty introduced by the proportionality assumption [28].

In general, we find that country and industry footprints with a high import share show a higher variability (compare also Figure S2 and S3 in the Additional file 1). This finding is in line with Jiang et al. 2020 who found a high correlation between the percentage of the China’s material footprint sourced from imports and the error of the footprint introduced by the proportionality assumption [19]. This relationship also seems to explain the overall lower variability of the carbon footprints (with relatively low import shares), compared to material, land, and water footprints (with import shares up to 100%). The finding that a high import share does not automatically lead to a high variability in the footprints suggests that a high import share is a necessary but not sufficient condition for a high uncertainty in national/industry footprints with regard to the proportionality assumption.

With the algorithm we applied in this study we provided an upper boundary estimate of the uncertainty that might arise from the proportionality assumption in MRIO analysis. By only considering ‘extreme’ versions of how to allocate imports to target sectors, the sample distributions of the footprints (and all numbers derived from them) cannot be seen as the ‘true’ uncertainty distributions under the assumption that each possible allocation is evenly likely. However, we consider our study as a relevant contribution to the quantification of the maximum possible uncertainty that may arise from relaxing the proportionality assumption in MRIO.

One possible reason why the actual maximum uncertainty could be even larger than stated in this study, however, is that the uncertainty estimate is based on only 4897 iterations. Theoretically, with our allocation algorithm there exist almost infinite possibilities to construct ‘new’ \(Z\) and \(Y\) matrices. The computational limit of 4897 iterations is almost entirely owed to the calculation of the Leontief inverse which is - given the size of EXIOBASE - computational expensive even when solved as a system of linear equations (see also [29]).

In a next step, the question whether the inclusion of trade data from the BEC substantially decreases the uncertainty of the footprints could be answered qualitatively. Given the patchy and aggregate nature of the BEC data, however, compilers of highly disaggregated MRIOs such as EXIOBASE would still have to rely on strong assumptions. Another interesting research question would be the effect of the sectoral/regional resolution of a MRIO on the uncertainty introduced by the
proportionality assumption. This could either be approached by conducting a similar analysis for other MRIOs having different sectoral/regional resolutions or using the same database but aggregating sectors/regions step-wise.

To contextualise our results we can compare the footprint variability we found to the variability found in other studies. Most studies to date dealing with the uncertainty in MRIO-based footprint analysis compared the results across databases [30, 31, 32, 33, 34, 35]. An exception is the work of Lenzen et al. (2010) that uses (inferred) standard deviations of the raw data to analyse how these uncertainties propagate to UK’s carbon footprint estimate [36]. Almost all studies examine carbon footprints, only Giljum et al. (2019) focuses on material footprints [37] while for MRIO-based land and water footprint we could not find any study on uncertainty with a scope comparable to our work which would make it possible to compare our results with.

The most up-to-date study of the uncertainty of MRIO-based carbon footprints we are aware of is from Rodrigues et al. (2018) [30]. The authors analysed the uncertainty of national and industry carbon footprints using the variability in the footprints between five different global MRIO tables. At the level of national footprints they found CVs of 6% (USA), 9% (China) up to 16% for the Netherlands, and at the level of industry footprints CVs between 10 and 213%. It should be noted, however, that they used a much lower sectoral resolution only distinguishing 17 industry sectors. A similar analysis with regard to material footprints was conducted by Giljum et al. (2020) [37]. They reported the variability of country material footprints as % difference in footprints between three different MRIO databases. To allow comparison to our results, we took their raw results and calculated the CV of selected country footprints. Taiwan, Slovakia and the Netherlands showed the highest variability in their material footprints with CVs of 40%, 37% and 25%, respectively. At the lower end range the US and the German material footprints with CVs of 6% and 2%, respectively.

When comparing these numbers with our results, we find that for national carbon footprints the variability caused by the proportionality assumption is only 1 to 4% of the inter-database variability found by Rodrigues et al. Although the maximum variability of industry footprints we found (219%) is in the same range compared to Rodrigues et al., due to the much higher sectoral resolution of our data, we can assume that when aggregating our results to an equivalent industry resolution the variability will be considerably reduced. Similarly, for national material footprints the variability caused by the proportionality assumption is only 0.2 to 2% of the inter-database variability found by Giljum et al. Hence, we conclude that the import proportionality assumption only leads to a small variability of environmental footprints as compared to those assumptions made in the process of compiling environmentally-extended MRIOs that differ between individual databases, such as the source data, the use of the territorial versus residential principle, the chosen balancing algorithm or the breakdown/allocation of extensions (see also [26]).

However, as the variability of some industry and a few country footprints is too great to be ignored, and to help researchers that use MRIO to study environmental footprints at the national or sectoral level, we provide our main results in the supplementary (Additional file 3) in the form of .xlsx tables containing measures of
the variability (CV, 2.5th and 97.5th percentiles) of the environmental footprints of all regions and industries covered by EXIOBASE. With the help of these tables researchers can check if a industry/region that is important in their study ranks high, so that either the database can be improved through adding more details on bilateral trade, or the uncertainty can be calculated and reported.

Competing interests
The authors declare that they have no competing interests.

Author’s contributions
SS, SP and AJ conceived and designed the research. SS performed the computations, analysed the results and wrote the paper with inputs from all authors.

Funding
SS and SP were supported by a grant by the Eva-Mayr-Stihl Foundation. AJ’s work was funded by the Swiss National Science Foundation grant number 407340 172445 as part of the National Research Program “Sustainable Economy: resource-friendly, future-oriented, innovative” (NRP 73).

Acknowledgements
We thank Richard Wood for his helpful comments on an earlier draft. Further, we thank the BwUniCluster for providing a high performance computing cluster which gave us the opportunity to conduct such a data and computing intensive study.

References
1. Miller, R.E., Blair, P.D.: Input-output Analysis: Foundations and Extensions. Cambridge university press, Cambridge, UK (2009)
2. Wiedmann, T.: A review of recent multi-region input–output models used for consumption-based emission and resource accounting. Ecological Economics 69(2), 211–222 (2009). doi:10.1016/j.ecolecon.2009.08.026. Accessed 2019-08-15
3. Tukker, A., Koning, A.d., Owen, A., Lutter, S., Bruckner, M., Giljum, S., Studler, K., Wood, R., Hoekstra, R.: Towards Robust: Authoritative Assessments of Environmental Impacts Embodied in Trade: Current State and Recommendations. Journal of Industrial Ecology 22(3), 585–590 (2018). doi:10.1111/jiec.12716. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/jiec.12716. Accessed 2021-02-04
4. Rodrigues, J., Marques, A., Wood, R., Tukker, A.: A network approach for assembling and linking input–output models. Economic Systems Research 28(4), 518–538 (2016). doi:10.1080/09535314.2016.1326817. Accessed 2019-08-19
5. Dietzenbacher, E., Los, B., Stehrer, R., Timmer, M., Vries, G.d.: The Construction of World Input–Output Tables in the WIOD Project. Economic Systems Research 25(1), 71–98 (2013). doi:10.1080/09535314.2012.761180. Accessed 2020-07-09
6. OECD: BTDIXE Bilateral Trade in Goods by Industry and End-use, ISIC Rev.4 (2020). https://stats.oecd.org/Index.aspx?DataSetCode=BTDIXE_I4 Accessed 2020-07-21
7. Peters, G.P., Andrew, R., Lennox, J.: Constructing an Environmentally-Extended Multi-Regional Input–Output Table Using the Gap Database. Economic Systems Research 23(2), 131–152 (2011). doi:10.1080/09535314.2011.563234. Accessed 2020-07-09
8. Studler, K., Wood, R., Bulasavkaya, T., Södersten, C.-J., Simas, M., Schmidt, S., Usubiaga, A., Acosta-Fernández, J., Kuenen, J., Bruckner, M., Giljum, S., Lutter, S., Merciai, S., Schmidt, J.H., Theur, M.C., Plutzar, C., Kastner, T., Eisenmenger, N., Erb, K.-H., Koning, A.d., Tukker, A.: EXIOBASE 3: Developing a Time Series of Detailed Environmentally Extended Multi-Regional Input-Output Tables. Journal of Industrial Ecology 22(3), 502–515 (2018). doi:10.1111/jiec.12715. Accessed 2019-02-21
9. Lenzen, M., Moran, D., Kanemoto, K., Geschke, A.: Building eora: a global multi-region input-output database at high country and sector resolution. Economic Systems Research 25(1), 20–49 (2013). doi:10.1080/09535314.2013.769938. Accessed 2019-08-15
10. Majeau-Bettez, G., Pauliuk, S., Wood, R., Bouman, E.A., Strømman, A.H.: Balance issues in input–output analysis: A comment on physical homogeneity, aggregation bias, and coproduction. Ecological Economics 126, 188–197 (2016). doi:10.1016/j.ecolecon.2016.02.017. Accessed 2020-08-27
11. Wiedmann, T.O., Lenzen, M., Barrett, J.R.: Companies on the Scale. Journal of Industrial Ecology 13(3), 361–383 (2009). doi:10.1111/j.1530-9290.2009.00125.x. Accessed 2019-03-28
12. Huang, Y.A., Lenzen, M., Weber, C.L., Murray, J., Mattheus, H.S.: The Role of Input–Output Analysis for the Screening of Corporate Carbon Footprints. Economic Systems Research 21(3), 217–242 (2009). doi:10.1080/09535310903511348. Accessed 2019-04-10
13. Hardadi, G., Buchholz, A., Pauliuk, S.: Implications of the distribution of German household environmental footprints across income groups for integrating environmental and social policy design. Journal of Industrial Ecology 25(1) (2020). doi:10.1111/jiec.13045. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/jiec.13045. Accessed 2021-01-12
14. Ivanova, D., Studler, K., Steen-Olsen, K., Wood, R., Vita, G., Tukker, A., Hertwich, E.G.: Environmental Impact Assessment of Household Consumption. Journal of Industrial Ecology 20(3), 526–536 (2016). doi:10.1111/jiec.12371. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/jiec.12371. Accessed 2021-01-25
Figure 1  A simplified MRIO system with four regions, four industries/products and two final demand categories to exemplify how the import matrix $T_{RsiJ}$ is extracted from the inter-industry matrix $Z$ and the final demand matrix $Y$.

Figure 2  Sankey diagrams illustrating different possible ways to allocate $s$ to $u$. A - C: three possible outcomes of our algorithm. D: allocation based on proportional import shares. E: an example of an intermediate case which is not covered by our algorithm.

Additional Files
Additional file 1
Additional results

Additional file 2
A pseudo-code version of our algorithm we applied in this study.

Additional file 3
A zip folder containing the data behind Figure 3 A and Figure 4. The data of all individual model runs (needed to produce Figure 3 B and Figure 5) can be get upon request from the author

Additional file 4
Country Codes

Availability of data and materials
EXIOBASE is available at https://exiobase.eu/. All code needed to reproduce the results of this article is available in the github repository https://github.com/simschul/import_proportionality under the commit 990d4c66f227798472bd403e2105f9c6b9f4505
**Figure 3** A) The relative variability (CV) of the national footprints compared to their size assuming proportional import shares. Boxplots show the distributions of the CVs across the 49 regions. Country codes according to ISO 3166-1 alpha-3 except RoW regions (see Additional file 4). B) The sample distributions of the national footprints exemplary for some selected interesting regions. The footprints where normalised by dividing each sample by the mean of all 4897 samples. The violin plots show the probability density and the range of all samples. The boxplots show the inter-quartile range (IQR) where 50% of all samples are situated (boxes), the sample’s median (horizontal line) and the range from the 2.5th to the 97.5th percentile (whiskers). The red points show the footprints size assuming proportional import shares, also normalised by the mean of all samples.

**Figure 4** The relative variability (CV) of the industry footprints compared to their size assuming proportional import shares (Equation 5). The 10 industries with the smallest non-zero footprint and all industries with a zero footprint are not shown. Boxplots show the distributions of the CVs across all industries with a non-zero footprint.

**Figure 5** The sample distributions of the industry footprints exemplary for selected industries. The footprints where normalised by dividing each sample by the mean of all 4897 samples. The violin plots show the probability density and the range of all samples. The boxplots show the inter-quartile range (IQR) where 50% of all samples are situated (boxes), the sample’s median (vertical line) and the range from the 2.5th to the 97.5th percentile (whiskers). The red points show the footprints size assuming proportional import shares, also normalised by the mean of all samples.