



Comparing imports based on microdata with FIGARO

Adam N. Walker

Nieke Aerts

Yoram Kunkels

Oscar Lemmers

Eline van Leeuwen

Khee Fung Wong

The project was made possible by Eurostat Grant SMP-ESS-2022-EGD-IBA, project 2022-NL-EGD, project number 101112968.

CBS Den Haag
Henri Faasdreef 312
2492 JP The Hague
P.O. Box 24500
2490 HA The Hague
+31 70 337 38 00
www.cbs.nl

project number

PR002518
National Accounts
22 December 2023

Index

1. Abstract	4
2. Introduction	5
3. Data	7
3.1 FIGARO	7
3.2 Statistics Netherlands data	7
4. Methods	9
4.1 Preparing the international trade data	9
4.2 Adding the sector dimension of the producer of imports	13
4.3 Analysing the international trade data	13
5. Results – a comparison with FIGARO	16
5.1 Comparison of raw data	16
5.2 Comparison of the foreign production requirement	22
6. Conclusion	25
7. References	27
8. Description of the input , intermediary and output data files	28
8.1 Input files	28
8.2 Intermediate files	31
8.3 Output files	32
9. Description of the R-scripts	34

1. Abstract

High quality international trade data is necessary for analysis of global supply chains and the calculation of environmental footprint indicators. This study employs a procedure using microdata to create national accounts consistent import data. It maximally uses the national data in the European Statistical System (ESS), such as data about trade in goods and services, trade by enterprise characteristics (TEC and STEC), the general business register and national accounts. This allows for more detail in products, industries and countries.

An analysis of the newly compiled data is carried out with a comparison to the import data in the FIGARO multi-region input-output table. In addition, the “foreign production requirement” of the Netherlands derived using FIGARO is compared to that derived with the use of microdata. This provides more insights into the effects of microdata use for footprint calculations. The findings show that the microdata results conform closely to FIGARO at the aggregate level. Several important differences do however exist at the detailed level. The imports of retail and wholesale services are substantially less when microdata is used. Trade with certain countries, particularly the United States, the United Kingdom and Belgium, is affected. There are also important differences regarding the imports from foreign mining and quarrying sectors and from foreign manufacturers of coke and petroleum, which are likely to have a noticeable effect on footprint results. The scripts for creating the international trade data, including an example with dummy data, are provided with the intention that other National Statistical Institutes can employ these techniques.

2. Introduction

Environmental footprints are becoming an increasingly popular and necessary indicator across multiple policy realms. This is because stakeholders are becoming increasingly aware of nature of international supply chains and the disconnect between the location of economic activities and the location at which the negative impacts of these activities occur. It is for example increasingly expected that businesses and nations work to reduce their footprints instead of simply working towards reducing their internal/domestic impacts¹. Multi-Regional Input-Output Tables (MRIOTs) provide the necessary data to calculate footprints that fully account for global supply chains.

A problem with MRIOTs is that they are generally not consistent with the national accounts and international trade data of individual countries. This is particularly a problem for the Netherlands because of the prevalence of re-exports, which are not properly accounted for in MRIOTs due to lack of publically available data. Footprints from MRIOTs substantially misestimate the footprint of the Netherlands for this reason, both in size and in allocation to countries. Therefore, Statistics Netherlands developed the SNAC (Single-country National Accounts Consistent) method (Edens et al., 2015). SNAC replaces Dutch data in the MRIOT and rebalances the system. Unfortunately, SNAC has proved too costly and time consuming to be used for the production of regular statistics. Statistics Netherlands is therefore developing a simplified SNAC method. This method is a hybrid method, referred to by Tukker et al. (2018) and Palm et al. (2019) as “simplified SNAC”. The idea is to increase national accounts consistency without creating a single national accounts consistent MRIOT. In the future, this will be achieved by calculating and summing three separate footprints, illustrated with the example of the Netherlands.

- 1) Emissions within the Netherlands resulting from final consumption of Dutch products as calculated with the SCIO (Single-Country Input-Output table) of the Netherlands
- 2) Emissions outside of the Netherlands directly and indirectly resulting from the foreign production required to produce products in the Netherlands for final consumption within the Netherlands. This requires the SCIO, the MRIOT and data on the imports.
- 3) Emissions in and outside the Netherlands resulting from final consumption of non-Dutch products within the Netherlands as calculated using the MRIOT.

Because there is no need to merge the SCIO and the MRIOT and no need to rebalance the system, simplified SNAC is more amenable to use in automation in an IT-system for the regular production of statistics, thus facilitating lower costs. Statistics Netherlands has extensively tested the simplified SNAC method and found that it can be used to produce high quality footprint estimates (Walker et al., 2023).

The most complicated part of the simplified-SNAC procedure relates to the second of the three separate footprints described above. Within this part of the calculation, the international trade data plays a key role by allowing the value chains in the Netherlands, as described by the SCIO, with the value chains in the rest of the world, as described by the MRIOT. It is also a key area in which national statistical agencies can employ their own data to improve the quality of the resulting footprints. This can be achieved by the integration of microdata at trader level. In this way, we can obtain a much higher quality than in existing MRIOTs. In addition, we obtain more detail with respect to products, industries and countries. MRIOT compilers do not have access to the unique micro data that an NSI has. It is known (Lemmers, 2013; Lemmers and Wong, 2019) that using microdata leads to different estimates of the share of imports from a country that is used for re-exports. This immediately affects the level of footprints, since imports from country A will have more embodied emissions than imports of the same product from country B. Therefore, we use the microdata to improve estimates about imports of a country for use in the Netherlands. Another quality improvement is obtained by using the microdata at the trader level to assign imports of a product from a country to the correct industries. Aerts et al. (2022) show that there is considerable heterogeneity here as well. For example, imports of palm oil from Malaysia are used by different

¹ See <https://ghgprotocol.org/sites/default/files/standards/ghg-protocol-revised.pdf>, the Greenhouse Gas Protocol for the corporate emission calculations

industries than imports from Indonesia. Using the microdata in this part will not change the total footprint of the Netherlands, but it will improve the footprint estimates at industry level. Policy makers can then target the industries that they find important, and industries get further insights in their supply chain.

The challenge in employing the micro-data is to ensure that the resulting dataset is consistent with the data in the SCIO. At Statistics Netherlands, the micro-data is not consistent with the macro totals in the national accounts. The principal of simplified SNAC is to employ wherever possible national accounts data of the Netherlands and to only employ data which deviates (which deviates as little as possible) where necessary. Therefore, the micro-data on international trade needs to be adjusted to ensure conformity with the national accounts

The aim of this study is to create and test a dataset pertaining to imports into the Netherlands that can be used in the simplified SNAC method. This dataset will employ all available microdata and be consistent with the SCIO of the Netherlands. Scripts (in the R language), documentation and dummy data will be provided as part of this project to researchers from other countries to produce the data. In order to demonstrate the effects of using microdata, two tests are carried out in this study. Firstly, the resulting import data for the Netherlands is compared to the import of the Netherlands according to the Eurostat's MRIOT known as FIGARO. The second test involves calculating the "foreign production requirement" (FPR) of the Netherlands using the data created and comparing this to the FPR created with FIGARO. The FPR is defined as the amount of output in sectors outside of the Netherlands that is directly required by producers in the Netherlands to fulfil the final demand for Dutch products. The FPR functions as the "glue" which connects the supply chain as described by the SCIO to the supply chain as described by the MRIOT, and therefore has an important role in the simplified SNAC process. In order to see how much of a difference the use of micro-data makes in the simplified SNAC process, a comparison is made between the FPR calculated only with FIGARO and the FPR calculated with the import data created in this project. These comparisons are made with data pertaining to the year 2020.

Section 3 of this report describes the data that is used to achieve the research aims. Section 4 explains the method used to compile the international trade data and to calculate the FPRs. Section 5 contains the comparisons to FIGARO and the analysis of the FPR. Section 6 concludes. Sections 7, 8 and 9 contain references, description of the datasets and R-scripts, respectively.

This report is principally of interest to the international statistical community. A basic understand of international classification systems and national statistics is therefore assumed.

3. Data

The data for this study is taken either from the FIGARO MRIO or from Statistics Netherlands. Statistics Netherlands data pertains to the input-output table and the international trade data.

3.1 FIGARO

FIGARO stands for Full International and Global Accounts for Research in input-Output Analysis. This dataset contains multi-regional input output tables in both the industry and the product dimensions, as well as inter-country supply and use tables. As of December 2023, the available time series runs from 2010 to 2021. We used 2020 in this study. FIGARO is considered by Statistics Netherlands to be a reliable and high quality dataset, meaning it is suitable material for comparison. Also, Statistics Netherlands plans to use FIGARO as an input for simplified SNAC. As such, it is useful to gain experience with using the dataset.

FIGARO is constructed by firstly creating balanced bilateral trade data. This involves reconciling various data sources and dealing with issues surrounding valuation layers. Secondly, a full set of national supply and use tables are collated including extra information on the use of imported products versus the use domestic products. From these data, inter-country supply and use tables are compiled and balanced. From these tables, the MRIOT is derived using a standard transformation. A full description of the method can be found in Remond-Tiedrez and Rueda-Cantucho (2019).

3.2 Statistics Netherlands data

The data from Statistics Netherlands is used to connect the use within the Netherlands to each imported good or service. In other words, for each product X imported from country Y it is labelled according to its use which can be re-exports, final consumption or intermediate use and additionally by which industry it is used (if applicable). In order to do this several different data sources are used, as listed below. We indicate the availability of each data source in other EU-countries. It is not necessary (and not expected) that all EU countries have the same data as Statistics Netherlands. Below we indicate, where possible, how to proceed when less data is available.

1. From national accounts, **detailed input-output** data is used. Each row in this data set contains a commodity (in our case 7-digit CPA extended), supply (e.g., imports or an industry), use (e.g., an industry or exports), value, valuation (e.g., basic prices, taxes). The valuation is important for the estimation of imports needed for re-exports: here all that is not basic prices is estimated to be production in the Netherlands for Dutch re-exports.

Availability: it is expected that of all data sources, this is the most problematic one for other countries. Not every EU-country compiles yearly an input-output table. Yet, it is expected that when they do so, they start from the supply and use tables, make the detailed information similar to above, then aggregate to industry-by-industry tables. For countries where that is the case, they can use that data. For countries where this is not the case, or where tables are not compiled yearly, it is possible to use the FIGARO input-output table and the underlying supply and use tables. The only information that then would be lacking is the estimate of imports for re-exports at commodity level.

2. **Trade in goods** contains the total trade in goods disaggregated to country, commodity (CN 8-digit), trade indicator (*import, re-export or export not re-export*).

Availability: except for re-exports, this is compulsory in the European Statistical System. There are various ways to estimate re-exports at the required level, see for example Roos (2006).

3. **Trade in goods on trader level.** Each row in this data set contains a commodity (CN 8-digit), country (in our case this is previous owner and not country of origin), trade indicator (*import, re-export or export not re-export*) and trader within the Netherlands (identified for example by VAT registration number. Note that even if the trader level data is not available for all trade in goods, it can still be used to improve the result.

Availability: except for re-exports, this is compulsory in the European Statistical System. There are various ways to estimate re-exports at the required level, see for example Roos (2006).

4. **Bridge table between traders and enterprises.** A bridge table between the trader (in the Netherlands: VAT number) and the enterprises in the General Business Register. If the link VAT number to enterprise is not a one-to-one relation some dividing key is necessary to distribute the amounts of one VAT number over the different enterprise numbers.

Availability: it is expected that each EU-country has such a bridge table, since trade in goods data has to be collected at VAT number and there are mandatory publications about trade in goods at enterprise characteristics (TEC). Even without this bridge table, the trade data at VAT level will improve country detail.

5. **Link between enterprises and enterprise groups.** An enterprise might belong to an enterprise group that consists of more than one enterprise, for example a multinational with many activities.

Availability: it is expected that each EU-country has such data in the General Business Register.

6. **Trade in services** contains the total trade in services disaggregated to country, commodity (in our case a code that links to 7-digit CPA codes), trade indicator (*import, export or SPE flow*).
- Availability:* it is expected that each EU-country has such data, possibly without the distinction SPE/non-SPE, since it will not be relevant for each country.

7. **Trade in services on trader level.** Each row in this data set contains a commodity (in our case a code that links to 7-digit CPA codes), country, trade indicator (only imports are used), trader within the Netherlands (CBS internal enterprise number). Again, even if there is no trader level data available for all trade in services, the data that is there can still be used to improve the result.
- Availability:* the Services Trade Enterprise Characteristics (STEC) statistics will be compulsory at EU-level in a few years. Some countries already have such data. Others might be able to link (part of) their survey data to traders.

8. **General Business Register (GBR)** provides the industry and the enterprise group an enterprise belongs to.

Availability: it is expected that each EU country has a General Business Register.

9. **Coupling information** between national accounts and trade data. First, bridging tables between the commodities and types of services in the international trade data and the products in national accounts. Second, knowledge about which adjustments are being made in order to merge the trade data with the national accounts data is important. We illustrate this with four examples. First, we removed the flows of processing trade and returned goods. These are not considered flows of trade in goods in national accounts. Second, imports for production abroad. These imports are not included in trade data and they are included in national accounts imports. Therefore, estimations of these amounts as included in national accounts imports are removed from the national accounts imports and these imports are handled separately. Third, some products, such as water imports are not included in trade data of the Netherlands. For such commodities, the origin can be estimated based on other data sources. We chose to handle this commodity by commodity and set a country distribution for imports not included in trade data. For example, for water imports the country distribution as found in UN Comtrade data is used. Fourth, sometimes, for large enterprises there is a specific adaptation of the original trade data to adhere to the principles of national accounts. Then this adaptation is put on our trade input data as well.
- Availability:* it is expected that all EU-countries have bridging tables between trade in goods and services. It is expected that they all have information about the adaptations because of conceptual differences such as processing trade and returned goods. Information about production abroad or specific information about adapted data at large enterprises will be scarcer. Yet countries with a Large Case Unit might have some information.

4. Methods

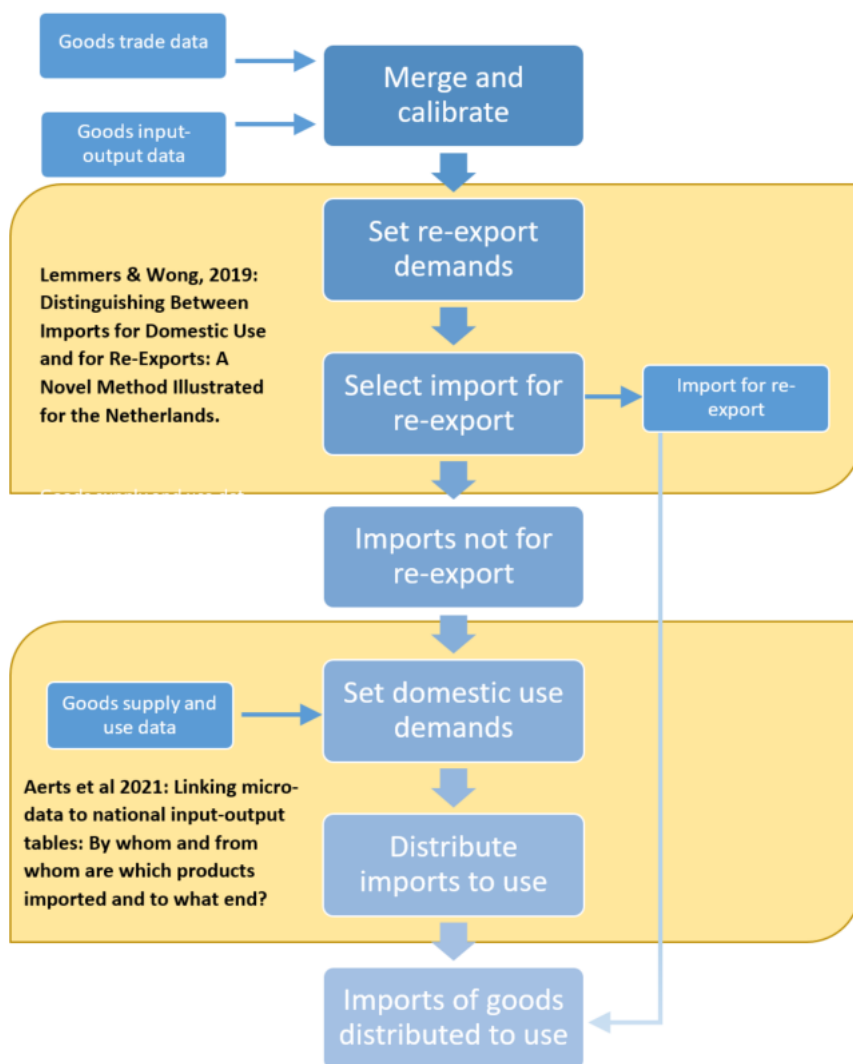
This section describes how the international trade data is prepared using microdata such that the results are consistent with the SCIO of the Netherlands. It then explains how the product dimension (what is imported) is replaced with the sector dimensions (which foreign sector produces that which is imported). The method for comparing the resulting data with FIGARO is as explained earlier, in terms of both the raw data, and the calculation and comparison of the FPR.

4.1 Preparing the international trade data

For the Netherlands, there is a big difference between the trade in goods and trade in services data. Therefore, allocating imports to their use is done separately for goods and services.

4.1.1 Trade in Goods

When allocating the imports to use we want to use as much detail as possible. Within Statistics Netherlands it is so that for allocating imports for re-exports a lot more information is available than for imports for other uses. Therefore, allocating the imports of goods to use is done in two steps: first imports for re-exports are allocated, and then imports for other uses are allocated. As shown in the figure below, allocating imports for re-exports follows the methodology from Lemmers and Wong (2019). Allocating imports of goods for other uses follows the methodology from Aerts et al. (2022).



The process is started by merging and calibrating the trade in goods data with the detailed input-output data. There are two special cases handled separately.

1. Imports for production abroad are included in the input-output data but not in the international trade data. Therefore, the estimated amounts of imports for production abroad are removed from the input-output data. These are imports for which the use is known but the country from which they are imported is not. These are assigned to a country based on the distribution of the imports within this commodity. For example, if 100 euro of commodity X is imported for production abroad and from the trade data we know that 80 percent of this commodity is imported from Belgium and 20 percent from Germany we estimate that 80 euros worth of imports for production abroad of commodity X is imported from Belgium and 20 euros from Germany.
2. Merchanting trade is another such type: this is included in the detailed input-output data but not in the trade in goods data. This is included in the trade in services data and will be taken care of later.

After the special imports are removed from the dataset, we continue with merging and calibrating the two datasets. Basically, the resulting data should adhere to the totals as given by national accounts. National accounts trade data in the detailed IOT uses 7-digit commodity codes (CPA 2008 extended codes). On the other side, the international trade data uses 8-digit CN codes. Firstly, the international trade data is transferred to the 7-digit commodity codes of national accounts, from here on denoted by *commodityIOT*. Then the totals per commodityIOT and per type (import, export, re-export) are compared between trade as in the detailed IOT and in the international trade data.

For each commodityIOT and type, the international trade data is altered as follows:

- a. If the two totals are equal, nothing is changed.
- b. If the total in national accounts is lower, each trade flow from international trade data is decreased by the same percentage so that the totals are equal.
- c. If the total in national accounts is higher, the international trade data is not just increased by a percentage because we should not assume that the missing imports are connected to the companies listed. Only if the difference is small we feel that this is a valid assumption. Therefore, we set two bounds: an absolute bound δ and a relative bound ϵ^2 . If the absolute difference between the totals is larger than δ or when the relative difference is larger than ϵ times the international trade total, international trade data is not increased. Instead additional rows are added where the partner country is set to be the average of the partner country of the imports of this commodity and the enterprise information is left empty. If the increase does not cross the set bounds, each trade flow from international trade data is increased by the same percentage so that the totals are equal.

The remaining process consists of two steps. The first step compiles the data about imports for re-exports. The second step compiles the data about imports for domestic use. That can be for production in the Netherlands (possibly for exports) and final use in the Netherlands.

Compiling data about imports for re-exports

When the two datasets are calibrated, the import for re-export demands are set. At Statistics Netherlands, there is also detailed (enterprise based) data available on re-exports. However, in this data it is not given how much of this value consists of imports. We follow Lemmers & Wong (2019) to obtain an estimate. For each commodity we get an average imports per re-export share. In national accounts data for each re-exported commodity we get the total value of re-exports and the value of imports for re-exports at basic prices level. We assume that on average the imports needed for one euro of re-exports of a commodity is equal to the share of basic prices of imports for re-exports of the total of re-exports in one euro. Using these shares, we can create a dataset that has imports needed for re-exports on an enterprise x commodity level. Now these demands are going to be filled using the scaled detailed imports dataset following the algorithm of Lemmers & Wong (2019).

² In the current project, we used $\delta = 500$ million and $\epsilon = 0.25$.

This algorithm consist of six rounds. In each round as much of the demands as possible are satisfied, each following round will have weaker conditions on assigning imports to demands. In the last round, it is assured that all demands are satisfied.

The conditions on which imports data has to match to the re-exports data for each round are:

1. VAT number and commodity code (CN8).
2. VAT number and the first 6 digits of the commodity code (CN6)
3. enterprise number and the first 6 digits of the commodity code (CN6)
4. enterprise group and the first 6 digits of the commodity code (CN6)
5. enterprise group and the national accounts commodity code (CPA7)
6. on product level only, subsequently on: CN8; CN6+CPA7; CN4+CPA7; CN6; CN4; CPA7; CPA2

It may occur that not all demands are satisfied after these steps. For example for the Netherlands water import is not included in international trade data. For such a commodity the country of origin is set using information from different data sets, e.g., UN Comtrade.

Compiling data about imports for domestic use

After the imports for re-exports are assigned, the imports within the trade data that are not intended for re-exports are left to assign to a use. This follows the method from Aerts et al. (2022). For this, we use the industry classification of the trader when available. It is known that most of these imports for which no trader is available are destined for re-exports.

Conversely, from the National Accounts data, we have information about the imports by product per and industry. There is one category that cannot be used immediately: an imported good can be placed in stock in one year and taken out of stock at a later time. In such a cases, we observe negative imports due to the inventory mutation in the data. Although this negative value does obviously not represent actual demand, it is still an imported good that is available for use by this respective industry. In order to meet the demand, it may be necessary to make the goods held in stock available. We add these goods to the traded data with an unknown origin. In terms of footprints, these imports have already been paid for when they were imported and placed in inventory. Therefore, they are not included in the footprint calculation.

Per industry (i) and commodity (c) we are given a demand (from National Accounts data), denoted by $D[i, c]$. Per industry and commodity, we also have a total amount of imports, denoted by $TI[i, c]$ (from trade data). Additionally, we have a separate category that represents the total imports for each commodity that could not be matched to an industry: $TI[NA, c]$, where NA stands for Not Allocated.

For each tuple industry (possibly including NA), commodity and country (l) there is an import value $I[i, c, l]$ in the international trade data. We create a dataset with a row for each such tuple, followed by $I[i, c, l]$. In each row, we add a demand $D[i, c, l]$ using the following definition:

$$D[i, c, l] = (D[i, c]) / (TI[i, c]) * I[i, c, l]$$

Note that, for tuples which industry is NA, this demand is zero and for others the demand will be higher than the available imports. This way, the imports for which the industry is set will be matched first. When there are no more unmatched imports satisfying both industry and commodity, the imports where industry is not set are used. This is particularly the case for the imports by wholesalers and transporters in the trade data, which are typically known as intermediaries facilitating the imports of the using industry.

When there are no available imports for commodity c and industry i , we need to use a different equation, as the denominator $TI[i, c]$ in the previous equation is 0. In this case, the demand is divided evenly over the rows:

$$D[i, c, l] = (D[i, c]) / (\#(TI[i, c] = 0 \& D[i, c] \neq 0))$$

Here $\#(TI[i, c] = 0)$ stands for the number of times that $TI[i, c]$ is equal to zero. Unfortunately, even at the highly aggregated commodity level of National Accounts not all demands can be satisfied. Therefore, we stepwise match and aggregate in order to satisfy all the demands. In total, we go through the following eight distribution rounds:

1. CPA 7-digit and industry
2. CPA 6-digit and industry
3. CPA 4-digit and industry
4. CPA 7-digit
5. CPA 6-digit
6. CPA 4-digit
7. CPA 2-digit and industry
8. CPA 2-digit

In rounds 1, 2, 3 and 7 only imports for which the industry is set are distributed. In the other rounds, the industry is not taken into account and the available imports are just simply proportionately distributed over the demand. With respect to imports going to domestic consumption, we only have information on the commodity; hence, these will be assigned in rounds 4 to 6 and round 8.

4.1.2 Trade in services

For trade in services we follow the same strategy: try to use the detailed information available first and distribute the rest equally.

Within trade in services, there are three special categories: government services, travel services (consumption abroad) and services from special purpose entities (SPEs). For these three services types there is only limited information available at Statistics Netherlands. For such services, there is trade partner information available in the trade in services dataset but no industry information. Therefore, the imports of services of these types are distributed proportionally over the partner countries.

For other service types, there is additional information available for some but not all of the services imports, as the trade in services on trader level is a sample. The additional information contains industry of the importing enterprise and this is used to match imports to intermediate use. Secondly, the service types in the service trade data does not have a one-to-one correspondence to the commodities (CPA) used in the national accounts data. Each service type can be mapped to one or more different commodities. For example, one service type is "industrial services" which can be mapped to many different commodities. The detailed import data is used in two rounds:

1. Distribution on full industry and commodities to which the service type can be mapped
2. Distribution on partial industry (2-digit) and commodities to which the service type can be mapped.

These two steps use do not yet assign all trade in services to countries. The rest of imports of national accounts is assigned to countries proportionally. Again, the trade in service dataset is used to assign a country to the imports that was not matched in the previous steps. We use the average origin of the specific service type, taking into account the earlier assigned trade to prevent bias. That is, if some part is already assigned to country X in the steps above, we remove such a share from the trade in services dataset so that the distribution in the final result equals the distribution in the trade in services dataset.

4.2 Adding the sector dimension of the producer of imports

The result of the international trade data preparation is a data file that contains a product dimension. In other words, the data shows *what* has been traded. This data then needs to be used in the simplified-SNAC process. This process uses data with a sector dimension. In other words, simplified-SNAC requires data regarding not what is being traded, but with *who* produced the traded products. It is therefore necessary to add the sector dimension.

The problem with adding a sector dimension is that a given product can be made by many different sectors. Therefore, extra data is required to inform which sectors make specific products. For example, in the UK, IT services are provided by a much wider array of sectors than only the IT services sector. Additionally, the IT services sector in the UK can also provide research services. The information what is produced by who is fortunately all available in the FIGARO supply table; even better sources might exist as well. Import of IT services and import of research services can therefore be allocated proportionally to all the sectors in the UK that produce these services. The product dimension can then be removed from the data. The remaining information pertains only to how much product in total each sector has exported to Netherlands, or returning to our example, how much IT services and research services in total has been exported to the Netherlands from the UK IT services sector.

4.3 Analysing the international trade data

The international trade data is analysed in two ways. Firstly, the raw data is compared to the data in the FIGARO input-output table. Secondly, FPRs (Foreign Production Requirements) are calculated using both FIGARO and Statistics Netherlands data. The methods required to make these comparisons are explained in this section.

4.3.1 Raw data comparison

The comparison of the raw data is fairly self-explanatory. As mentioned in section 3.2, after replacing the product dimension with the sector dimension, the international trade data has the following dimensions: producing sector, producing country, importing Dutch sector. Data with these exact dimensions can be extracted from the FIGARO input-output table. This facilitates comparison between the two data sources.

It is important to note that this study only compares the *imports for intermediate use*. Imports for final consumption are not analysed here, although they are obviously important for environmental footprint calculations.

4.3.2 Foreign production requirement (FPR) comparison

As mentioned in the introduction, the FPR has an important role in the simplified SNAC method, being the “glue” that links supply chains within the Netherlands to the foreign part of the supply chain. We define the FPR formally. To begin, let us define the following terms:

m \equiv number of sectors in FIGARO in total

n \equiv number of sectors in FIGARO in a single country

L \equiv Leontief matrix for the Netherlands ($n \times n$)

F \equiv A vector of final consumption in the Netherlands (n)

Z \equiv A vector of the output per sector in the Netherlands (n)

Two identities are required to calculate the Leontief matrix: the intermediate block of an input-output table and the output per sector. For this report, the Leontief matrix is calculated using the intermediate block for the Netherlands taken from FIGARO and with the output per sector in the Netherlands taken also from FIGARO³.

³ When the SAMCA method is operationalized for the calculation of footprints, the intermediate block for The Netherlands is taken from internally available national accounts data, which employs a more detailed sectoral classification.

This allows for the most precise comparison between the different results. The Statistics Netherlands import data dimensions have already been converted to sectors (section 3.2). It can therefore be organized in a matrix defined as follows.

$Y \equiv$ matrix of imports according to Statistics Netherlands data by domestic sector and foreign sector ($n \times m$)

The matrix Y therefore shows how much is imported by each sector (row) in the Netherlands from each sector in FIGARO (columns). The columns referring to the Netherlands are populated with zeros (the Netherlands does not import from itself). The cells of each column show how much is imported by each sector in the Netherlands from that foreign sector. The matrix Y is used to calculate the following identity,

$J_{SN} \equiv$ import intensities according to Statistics Netherlands ($n \times m$)

which is calculated as

$$J_{SN} = Y * \text{diag}(Z^{-1})$$

The FPR is calculated as

$$J_{SN} * L * \text{diag}(F) \quad (1)$$

which results in a matrix with dimensions ($m \times n$). Because the matrix of intensities J_{SN} is populated with zeros in columns relating to the Netherlands, the cells of the FPR matrix referring to the Netherlands are zero also. The FPR matrix contains information on the amount directly produced by each foreign sector to satisfy the final demand in the Netherlands for the products produced by sectors in the Netherlands. The FPR is a direct requirement in the sense that intermediate producers in the Netherlands require it directly. It is indirect in the sense that it is not directly demanded by Dutch consumers: it is indirectly demand via supply chains in the Netherlands. It does not include the foreign production that is required indirectly to satisfy the direct foreign production requirements. For example, if a Dutch consumer buys a German car from a Dutch retailer who imported the car from Germany, then the FPR is for one car. It does not include the production required to produce the car, i.e. the agricultural output required to produce leather seats is not included.

The FPRs as described above is calculated using the import data created by Statistics Netherlands as given by matrix Y . In order to provide material with which to compare and to gain insights into the effects of using Statistics Netherlands data, we also calculate the FPR using X : an alternative to Y which is defined as

$X \equiv$ matrix of imports according to FIGARO ($n \times m$)

The matrix X is taken directly from the intermediate block of the FIGARO input-output table by selecting the rows that relate to the Netherlands and setting all values in the columns relating to the Netherlands to zero. Now set

$$J_{FIGARO} = X * \text{diag}(Z^{-1})$$

Accordingly, we can calculate the FPR anew as

$$J_{FIGARO} * L * \text{diag}(F) \quad (2)$$

The FPR according to formula (2) has the same dimensions as the FPR calculated according to formula (1). This facilitates comparison in detail. Formulas (1) and (2) differ only because the import intensities are different, and

these differ only because the import data is different. Differences between the two FPRs are therefore solely attributable to differences between the import data according to Statistics Netherlands and the import data according to FIGARO.

5. Results – a comparison with FIGARO

Section 4 explained how comparisons could be made in terms of both the raw data and the FPR. This section makes these comparisons and analyses them. All data pertains to the year 2020, the most recent year in FIGARO when this study was carried out.

5.1 Comparison of raw data

While this study focuses on the imports for intermediate use, it is pertinent to begin by analysing the total import and its allocation to either final consumption or intermediate use. By construction, the total import (including re-exports) is the same according to Statistics Netherlands and FIGARO. Differences do however exist when re-exports are excluded. Total imports excluding re-exports according to FIGARO is €328 bn. According to Statistics Netherlands, this value is 3% higher at €339 bn.

We can now proceed to analysing the allocation of imports to either final consumption, intermediate use or re-exports. FIGARO does not include data on re-exports. However, we can infer the value of re-exports in FIGARO by taking the difference between total imports according to Statistics Netherlands data and the sum of imports for final consumption and intermediate use. These results are shown in figure 1.

Figure 1: Import of the Netherlands per use type according to FIGARO and Statistics Netherlands.

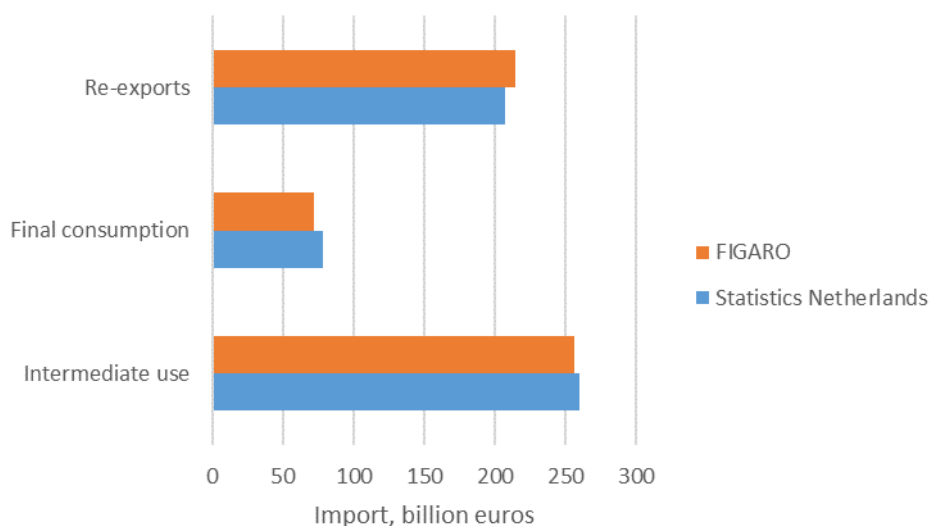


Figure 1 shows that re-exports are smaller according to Statistics Netherlands data. Secondly, the figure shows that the exclusion of re-exports has had a proportionally smaller effect on imports for intermediate use. In other words, Statistics Netherlands data allocates less import to re-exports and relatively fewer imports for re-exports to final consumption than to intermediate use.

The difference between imports for intermediate use according to FIGARO and Statistics Netherlands is 2.3%. We proceed by analysing this difference firstly in terms of the countries from which the Netherlands imports. The results are shown in figure 2. We will then analyse the imports in terms of the foreign sectors which produce them (figure 3) and in terms of the sectors in the Netherlands which import them (figure 4).

Figure 2. Import of the Netherlands for intermediate use per exporting country according to FIGARO and Statistics Netherlands, sorted according to the magnitude of the difference between the value, with a selection of the 12 countries for which the difference is most positive, and the 12 for which the difference is most negative.

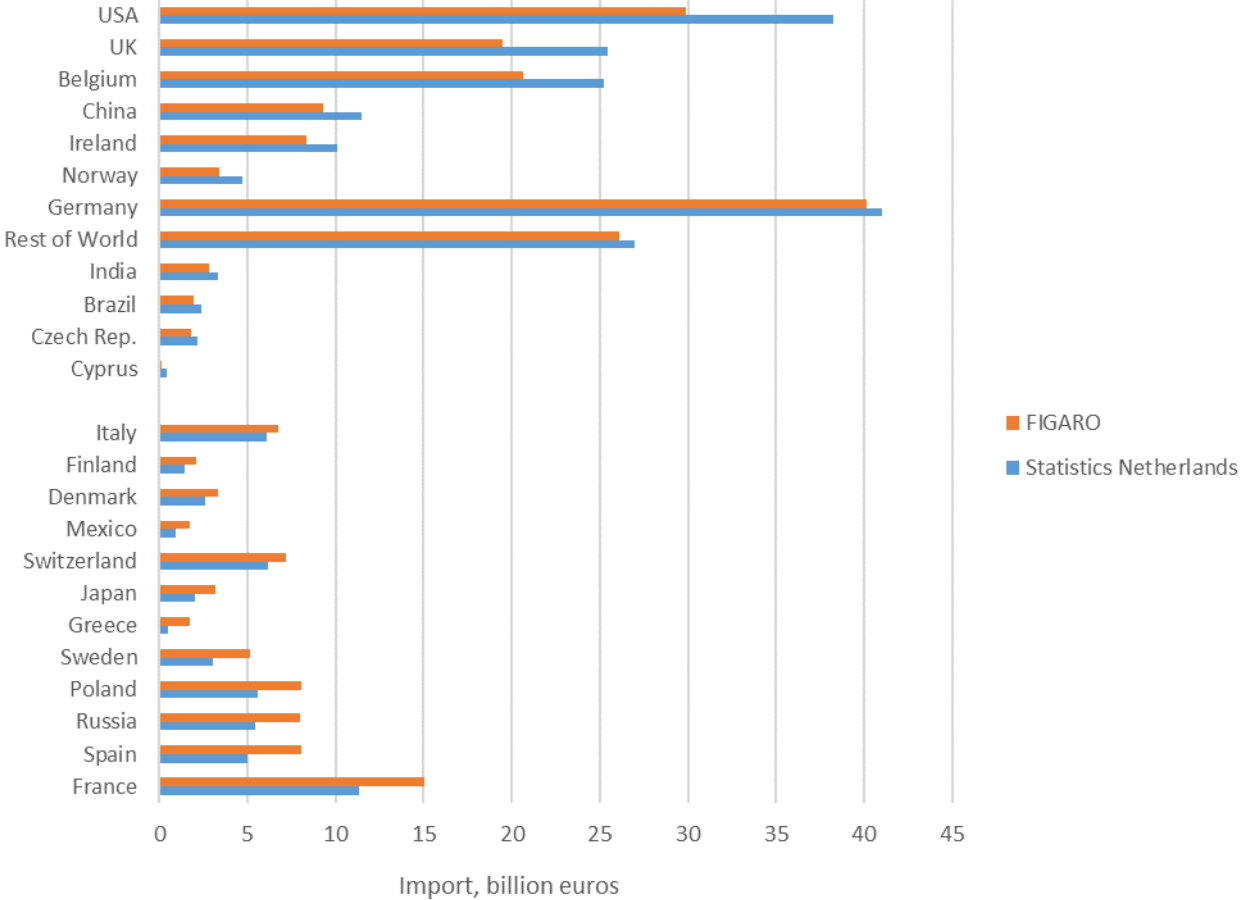


Figure 2 shows that the differences in imports for intermediate use can be larger at the level of country than at the total level shown in figure 1. In particular, we see large differences for important trading partners such as the USA, the UK and Belgium, for which import is greater according to Statistics Netherlands data. We also see many cases for which import is greater according to FIGARO, most notably France, Spain, Russia and Poland.

To explain these differences, we need to know whether they results from shifts between import for final consumption and intermediate use, or because of differences in the total, which will predominantly be determined by differences in re-exports at country level. Unfortunately, information on re-exports at country level is not available in FIGARO. However, we can analyse the extent to which imports shift between imports for intermediate use and imports for final consumption in the Statistics Netherlands data. Whatever difference is not explained by this shift, must be explained by another factor, such as differences in re-exports. Let us take the USA as an example. Not shown in figure 2 (because it pertains only to imports for intermediate use), is that the total value of imports for intermediate use plus final consumption is 15% greater in the Statistics Netherlands data relative to FIGARO. Further, according to FIGARO, 82% of imports from the USA are used for intermediate consumption, whereas according to Statistics Netherlands, this value is 88%. There are therefore two effects at play. Firstly, Statistics Netherlands estimates greater imports in total from the USA (this is predominantly explained by the treatment of re-exports). Secondly, there is a shift from imports for final consumption towards imports for intermediate use in Statistics Netherlands data relative to FIGARO.

The example of the USA is however not representative in general. Generally, we actually see that the shift occurs in the other direction, namely from imports for intermediate consumption towards imports for final

consumption. Imports only shift from imports for final consumption towards intermediate use for 17 of the 45 countries and notably, this does not occur for the other countries at the top of figure 2, namely the UK, Belgium and China. In these cases, the difference in total imports “drives” the increase in imports for intermediate consumption in spite of a shift towards imports for final consumption. In the case of the countries towards the bottom of figure 2, imports generally follow the pattern of shifting from intermediate consumption towards imports for final consumption. This shift compounds a reduction in total imports and leads to lower imports for intermediate consumption according to Statistics Netherlands data.

Instead of considering the countries from which the Netherlands imports, we can also consider the sectors abroad from which the Netherlands imports. These results are shown in figure 3.

Figure 3. Import of the Netherlands per exporting sector according to FIGARO and Statistics Netherlands, sorted according to the magnitude of the difference between the value, with a selection of the 12 countries for which the difference is most positive, and the 12 for which the difference is most negative.

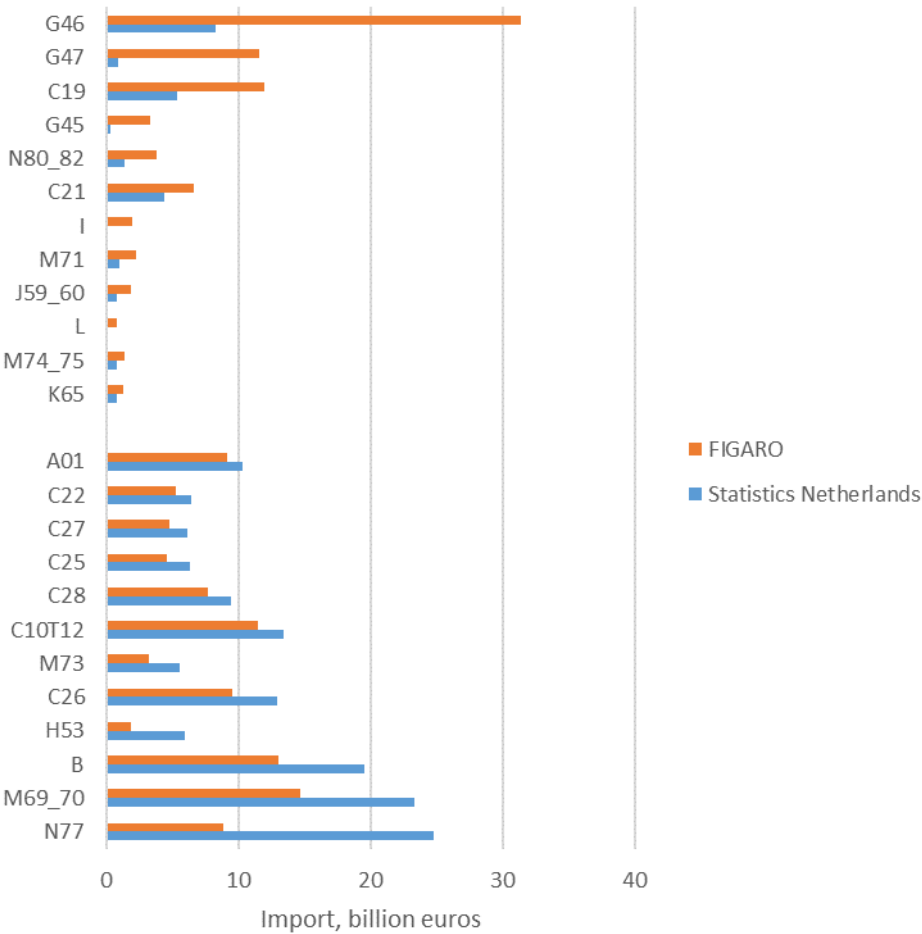


Figure 3 shows more substantial differences than figure 2. There are several cases where Statistics Netherlands imports (for intermediate use) are lower than that in FIGARO. This is particularly the case for wholesale (G46), retail (G47), coke oven products and oil refinery (C19) and trade in and repair of automobiles (G45). There are also several cases where Statistics Netherlands import data are higher than those in FIGARO. This is particularly the case for renting and leasing services (N77), financial, legal and administrative services (M69 in M69_70) and activities of head offices and management consultancy (M70 in M69_M70) and finally, mining and quarrying (B).

The differences shown in figure 3 provides more information on many of the differences shown in figure 2. Let us first consider a case where import according to Statistics Netherlands is greater than the import according to

FIGARO. This is the case for imports from the USA (see figure 2). Analysis of the imports from the USA at sector level reveals that this difference is predominantly due to differences in N77 conforming to those figure 3. The same is true for the UK, although in this case M69_70 also plays a large role. Let us now consider a case where import according to Statistics Netherlands is less than the import according to FIGARO. This is the case for imports from France (see figure 2). Analysis of the imports from France at sector level reveals that this difference is predominantly due to differences in G46 conforming to those figure 3. The sector C19 also plays a large role. These results also hold for Russia and Spain.

The differences in imports per exporting sector are, in the first instance, a function of differences in data regarding which products are imported from which countries. An analysis of the FIGARO use table shows that a similar value of CPA G46 products are imported into the Netherlands to the value of products of NACE G46 in the FIGARO input-output table. In both cases, this value is considerably larger than the value of both imports of CPA G46 and imports from foreign NACE G46 in the Statistics Netherlands data. While there are some differences caused by ancillary production (i.e., products other than CPA G46 produced by NACE G46 and the production of CPA G46 outside of NACE G46), these do not have a substantial effect on the results.

We now summarise the differences between imports per Dutch importing sector in figure 4.

Figure 4. Import per Dutch importing sector according to FIGARO and Statistics Netherlands.

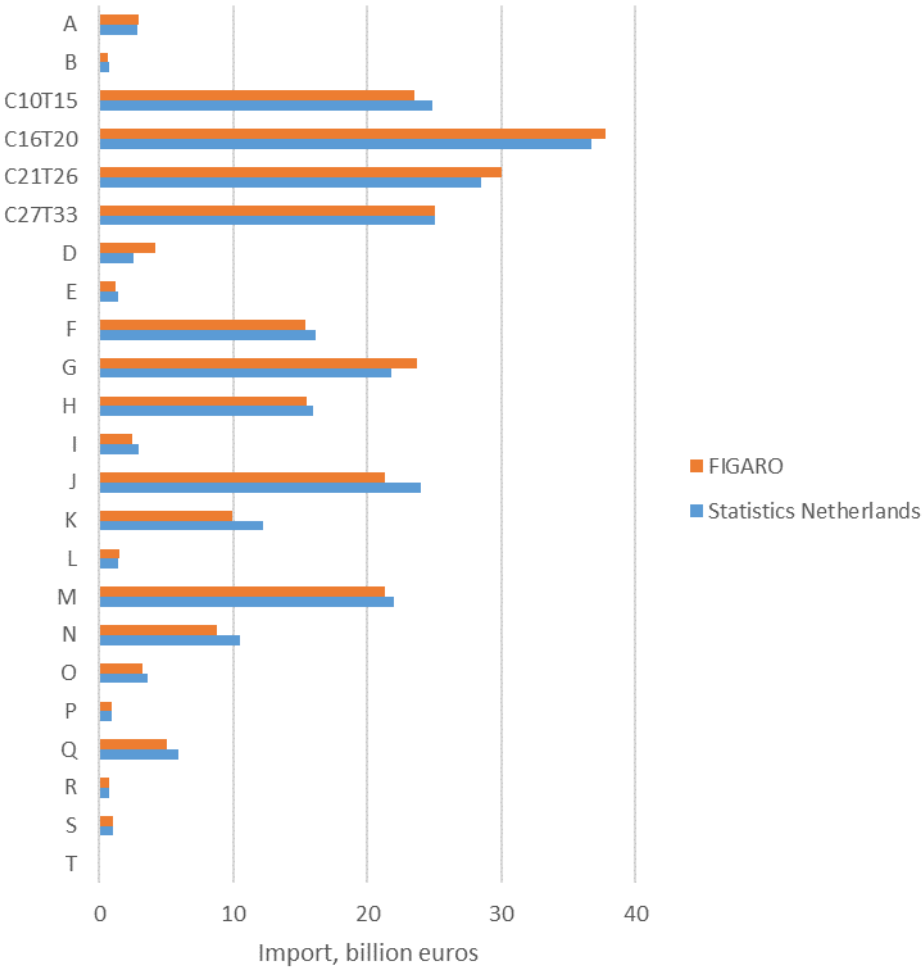


Figure 4 shows that the import per sector in the Statistics Netherlands data is approximately equal to that in FIGARO. The largest exceptions is the energy sector (NACE D) and wholesale and retail (NACE G) for which imports according to Statistics Netherlands data are lower, and information and communication (NACE J), for which

imports according to Statistics Netherlands data are higher. Specifically, within NACE C the exceptions are C19 (coke oven products and oil refinery, 3% difference), C20 (chemicals industry, 4% difference) and C26 (electronic and optical devices, 14% difference). Because the energy sectors D, C19 and C20 are important sectors for environmental footprints, it is pertinent to consider these sectors in more detail. All these sectors show a shift between NACE B and NACE C19. These two sectors are closely related because NACE B supplies raw materials to NACE C. Countries that import oil products can either import crude products from foreign NACE B sectors or import refined products from foreign C19 sectors. The Netherlands has a large oil refinery industry and thus both options are feasible. FIGARO shows that sectors D, C19 and C20 in the Netherlands import more from foreign NACE C19 sectors and less from foreign NACE B sectors. Statistics Netherlands data shows relatively more imports from foreign NACE B sectors and relatively less from foreign NACE C19 sectors. This is shown in figure 5. In addition, figure 6 shows the imports of from NACE B and NACE C19 sectors by all Dutch intermediate sectors together.

Figure 6 shows that, according to both datasets, the Dutch economy imports approximately the same value of products from foreign NACE C19 and NACE B sectors. It also shows that the distribution between B and C19 is different. Specifically, Statistics Netherlands data shows a larger import for foreign NACE B sectors and a lower import from foreign NACE C sectors.

When considering figures 5 and 6, it is important to note that although the monetary value of the imports is less according to Statistics Netherlands; this does not necessarily imply that the tonnage is less. This is particularly the case because the value per ton of refined products is greater than the value of crude products. The key point is that refining crude products requires energy and energy production is emission intensive. If the Netherlands imports less refined products and more raw products then it is expected that the environmental footprint is lower. The reason is that generally, electricity production is less emission intensive in the Netherlands compared to foreign countries.

Figure 5: Imports by the energy sector (D) in the Netherlands from foreign coke oven products and oil refinery sectors (C19) and foreign mining and quarrying sectors (B).

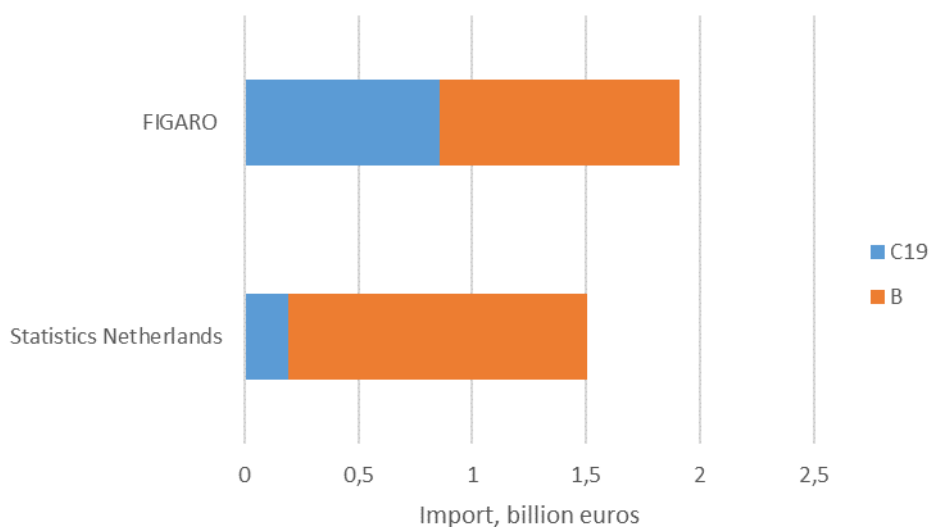
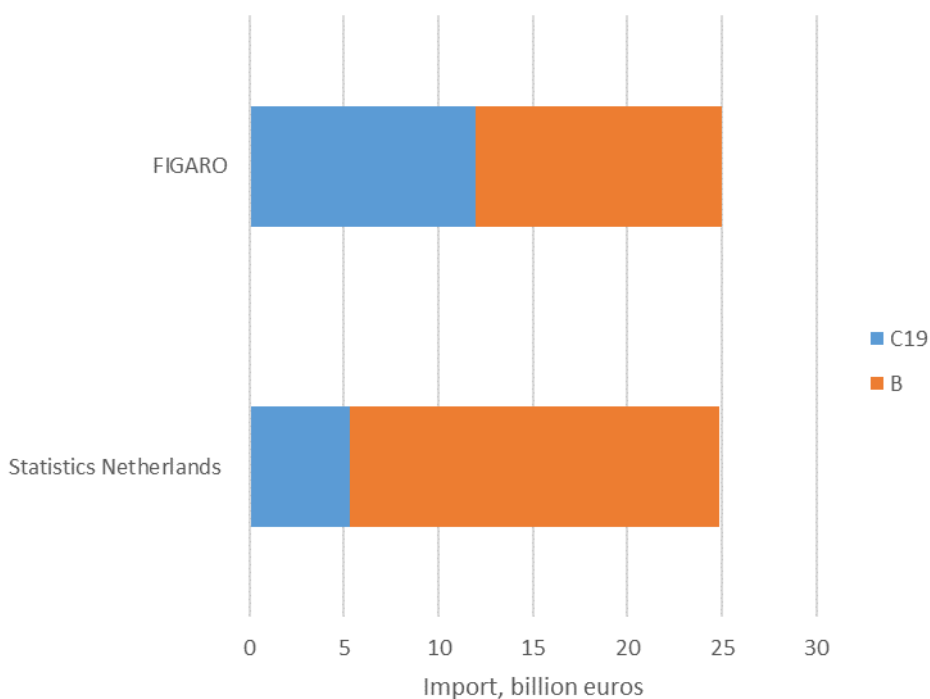


Figure 6: Imports by all intermediate sectors in the Netherlands from foreign coke oven products and oil refinery sectors (C19) and foreign mining and quarrying sectors (B).



From the analysis of the raw data, the following conclusions can be drawn:

- At the level of the country from which the Netherlands imports, the data of Statistics Netherlands conforms closely to that from FIGARO at total level. Differences are largest for the UK, USA and Belgium. For these countries, import is higher according to Statistics Netherlands. The differences are also large for France, Spain, Russia and Poland. In those cases the import of the Netherlands is greater according to FIGARO
- At the level of the foreign sectors from which the Netherlands imports, the data of Statistics Netherlands conforms closely to that FIGARO for some sectors. The most substantial differences are found or retail and wholesale trade, for which the import of the Netherlands is less according to Statistics Netherlands. Large differences are also found for certain service sectors (N77, M69_70). In those cases, the import of the Netherlands is less according to FIGARO.

- The imports per sector in the Netherlands according to FIGARO conform reasonably closely to Statistics Netherlands data, although there are some differences that will have an effect on environmental footprint results.

5.2 Comparison of the foreign production requirement

The FPR shows how much production in foreign countries can be allocated to Dutch final consumption via Dutch production chains. These are the imports used by Dutch enterprises to produce for Dutch final demand. The FPR of the Netherlands according to FIGARO is €101 billion, which is equal to 0.069% of all foreign production. The FPR of the Netherlands according to Statistics Netherlands is €105 billion, which is equal to 0.072% of all foreign production. This shows that the amount of foreign production required is greater according to Statistics Netherlands. This conforms to the first result of the previous chapter, namely that imports are greater according to Statistics Netherlands than FIGARO.

More detail on this result is provided in the following sections by considering the FPR per foreign country and the FPR per Dutch sector. In order to aid in interpreting the results, it is useful to quickly remind ourselves what FPR is and how it relates to the import. Firstly, it is important to remember that we are only considering the imports indirectly related to intermediate consumption. Secondly, it is important to remember that the FPR is a function of final demand in the Netherlands. The greater the final demand, the greater the FPR. Thirdly, the FPR is a function of Dutch supply chains. The FPR therefore will be greatest when final demand focusses on goods and services produced in the Netherlands that, directly or indirectly, require many imports in their production. Fourthly and finally, FPR is the *direct* requirement at the Dutch border (figuratively speaking). The foreign production required to produce the goods to be imported is *not* included.

5.2.1 Foreign production requirement per foreign country

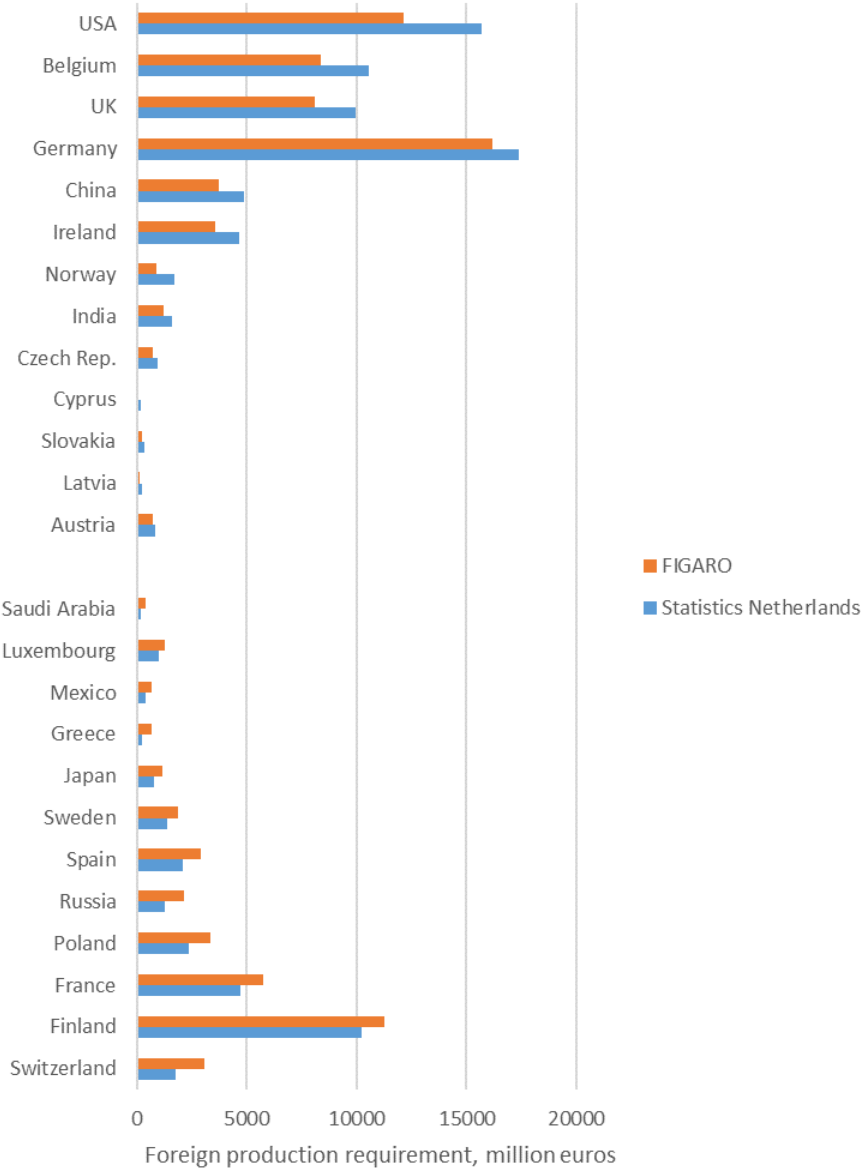
Figure 7 shows the differences between FPRs per country. There are several similarities between figure 7 and figure 2. This is to be expected because the FPR in a country is determined partly by the amount that the Netherlands imports from that country. For example, figure 2 shows that the imports from Belgium are 18% higher according to Statistics Netherlands. This partly explains the 21% differences in the FPRs. The difference in the differences can be explained by the products that the Netherlands imports from Belgium. The Statistics Netherlands data shows more imports of retail trade services. Retail trade services is an important sector for final consumption in the Netherlands, meaning that the modest difference in import translates to a sizeable difference in FPR.

We now discuss some differences between figure 2 and figure 7. Figure 2 shows that imports from Switzerland according to Statistics Netherlands are 17% lower than according to FIGARO. This translates into a 77% difference in FPR. This is especially interesting because Statistics Netherlands and FIGARO data both agree that the majority of imports from Switzerland consist of services. The differences are:

- the import of rental and leasing services of tangible goods (N77) and research, consultancy and specialised business services (M69_M70) is considerably higher according to Statistics Netherlands data
- the import of wholesale (G46) and retail (G47) and financial (K64) services are lower according to Statistics Netherlands data.

These differences have the largest effects on the FPRs of the Dutch retail and construction sectors. The retail sector has a greater FPR according to Statistics Netherlands data principally because of the greater FPR in Swiss sectors M69_M70 and N77. The FPR of the Dutch construction sector has a lower FPR according to Statistics Netherlands data because of the lower FPR in Swiss sectors G46, G47 and K64. As we shall see in the next section, the Dutch construction industry is very important in determining FPR.

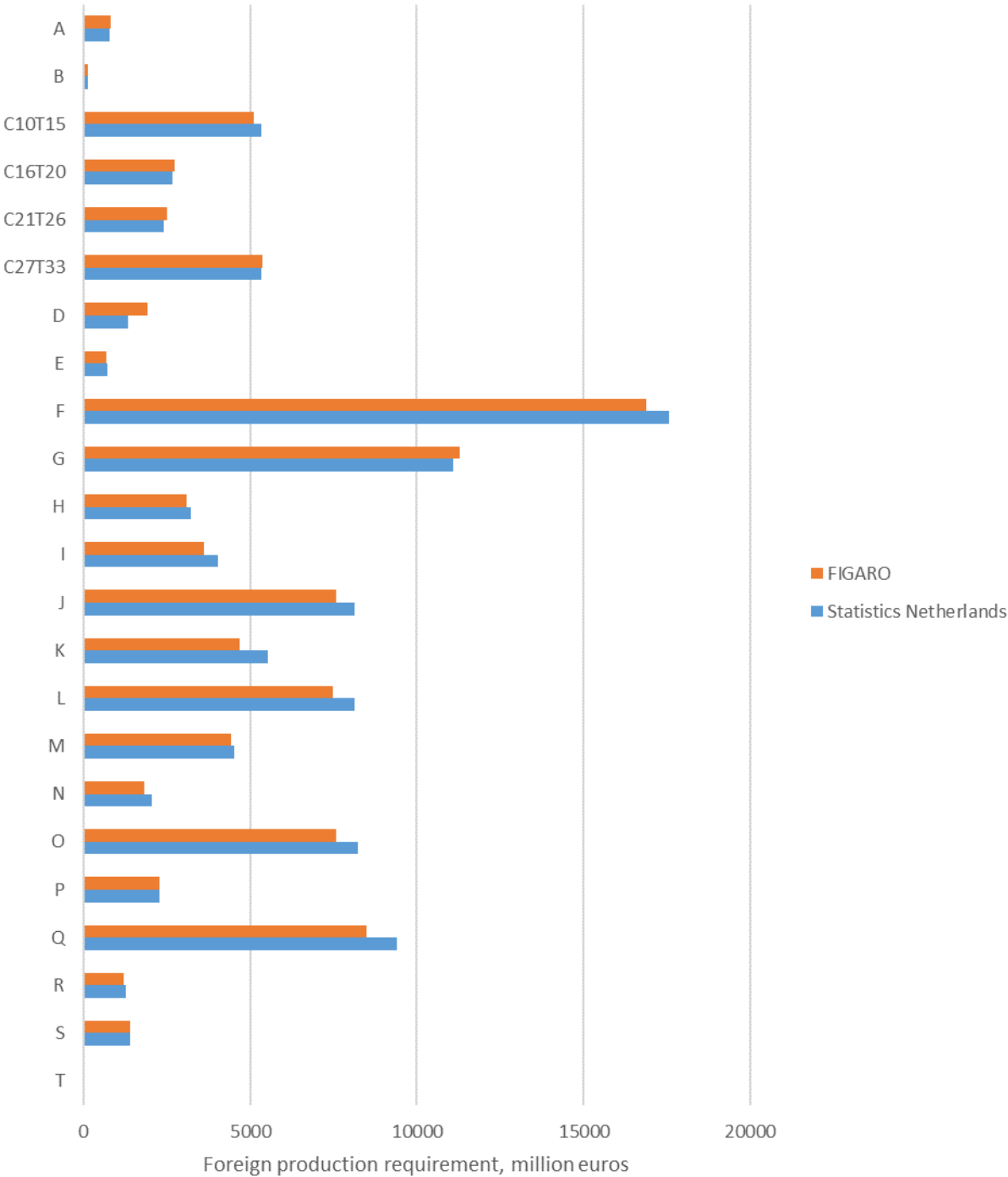
Figure 7. FPR per foreign country according to FIGARO and Statistics Netherlands, sorted according to the magnitude of the difference between the value, with a selection of the 12 countries for which the difference is most positive, and the 12 for which the difference is most negative.



5.2.2 Foreign production requirement per Dutch sector

We now present figure 8, which shows the FPRs of each Dutch sector. This figure corresponds to figure 4, which shows that import of each Dutch sector.

Figure 8. FPR per sector of the economy of the Netherlands according to FIGARO and Statistics Netherlands.



As mentioned in the previous section, we see in figure 8 that the Dutch construction industry leads to the most FPR. In this case, Statistics Netherlands and FIGARO results conform closely with each other. Because of the importance of this sector, differences in import country detail between Statistics Netherlands and FIGARO data can have a substantial effect on the results. An example is Switzerland, shown in the previous section.

Returning to figure 8, we see that the largest difference between the FPRs occurs for NACE F, for which the FPR is lower according to Statistics Netherlands data. This conforms to the results in figure 4 that shows lower imports by NACE F in the Netherlands. The second largest difference occurs for healthcare (NACE Q). Healthcare is an important sector for final consumption. This means that the modest differences in import translate to a larger difference in FPR. However, the difference is more pronounced in terms of the FPR than the raw import data.

6. Conclusion

In conclusion, let us begin by summarising the results. The comparison of the raw data has been made in terms of the countries from which the Netherlands imports, the foreign sectors from which the Netherlands imports and the imports per sector of the Dutch economy. The import values from Statistics Netherlands and from FIGARO concur reasonably well when the values per importing sector are analysed. There are however some larger differences where exporting sector and country are concerned. For example, Statistics Netherlands data estimates greater import from the USA, UK and Belgium and lower imports from France, Spain and Russia when compared to FIGARO. The differences are however most pronounced when considering the sectors from which the Netherlands imports. Large differences are seen for retail, wholesale and the manufacture of coke and petroleum (less import according to Statistics Netherlands data) and for business services, rental and leasing services and mining and quarrying (more import according to Statistics Netherlands data).

In the simplified SNAC procedure, it is however not the raw data which matters as such, but the FPR (Foreign Production Requirement) which is derived from the raw data. This is the foreign production that is directly exported to the Netherlands to be used for production for Dutch final demand. The differences in FPR can be different to those in the raw data. For example, the percentage difference in the FPR in Belgium is greater than the percentage difference in the raw data. This is because the raw data according to Statistics Netherlands contains greater import of retail services, which is an important sector for final demand. Similarly, differences in import of wholesale services by the Dutch wholesales sector lead to disproportionately large differences in the FPR of the Dutch wholesale sector because wholesale is also an important sector from a final demand perspective.

Another interesting difference is the shift in imports from C19 to B when Statistics Netherlands data is employed. This shift does not translate to substantial differences in total FPR. However, if greenhouse gas footprints were calculated, this shift would likely have an effect on the results.

The differences in the imports of wholesale and retail services are the most significant effect of employing Statistics Netherlands data as opposed to FIGARO. This results in substantially lower FPR and can be expected to result in lower environmental footprints. However, in general, at the aggregate levels, FIGARO and Statistics Netherlands data conform reasonably well. That the differences are most pronounced at the detailed levels demonstrate that the micro-data approach employed by Statistics Netherlands is beneficial.

An important possible shortcoming of this research is that it is carried out with data for only one year, namely 2020. COVID-19 caused considerable economic turmoil and some challenges for official statistics. For this reason it is possible that some of the results in this report have been affected the choice of year for the analysis.

Part of the research output of this project is a set of R-scripts (with dummy data) which National Statistical Institutes can employ using their own microdata. Hopefully, the insights in this report, and potentially the output of other projects using the same techniques, can be used to evaluate FIGARO and potentially lead to improvements in the way that FIGARO is constructed. Sharing micro-data directly with Eurostat is also an interesting possibility, as this would allow the data to be directly incorporated. Of course, confidentiality issues would have to be addressed.

Work is ongoing at Statistics Netherlands to incorporate the data created in this project into the simplified SNAC procedure. The simplified SNAC procedure is intended as the basis for an array of footprint calculations: initially greenhouse gases and natural resource extraction, with the intention to expand the procedure to include land use and biodiversity footprints in the future. A time series of footprint results is desired. To achieve this, the

method for the import data will need to be regularly updated to account for revisions to the national accounts. Further, export data is also relevant for footprint calculations. Traditionally, footprints are calculated upstream in the supply chain using Leontief analysis. The alternative, which is increasingly relevant in Dutch policy circles, is to calculate footprints downstream in the supply chain using Ghosh analysis. Downstream footprints are considered necessary to properly account for all emissions for which a given sector could be considered 'responsible'. For example, in upstream analysis, refineries are considered 'responsible' for the emissions in the energy sector related to the electricity production required for refinery processes. In downstream analysis, refineries are 'responsible' for the emissions by cars running on the petrol created by the refineries. Downstream analysis therefore provides a useful additional perspective on the role of sectors within supply chains. To carry out downstream analysis, import data are no longer relevant but export data is relevant. Different procedures are required for export data because of differences in the available microdata. Procedures for downstream data therefore need to be developed and incorporated in the simplified SNAC procedure.

7. References

Aerts, N., Bohn, T., Lemmers, O., Wong, K.F. (2022). Linking micro-data to national input-output tables: By whom and from whom are which products imported and to what end? Presented at the 28th International Input-Output Conference, IIOA, Malaysia, Langkawi.

Edens, B., Hoekstra, R., Zult, D., Lemmers, O., Wilting, H., & Wu, R. (2015). A method to create carbon footprint estimates consistent with national accounts. *Economic Systems Research*, 27(4), 440-457.

Lemmers, O. (2013). Using enterprise data to improve trade in value added estimates. Presented at OECD Workshop "Measuring Trade in Value-Added".

Lemmers, O., & Wong, K. F. (2019). Distinguishing between imports for domestic use and for re-exports: a novel method illustrated for the Netherlands. *National Institute Economic Review*, 249(1), R59-R67.

Palm, V., Wood, R., Berglund, M., Dawkins, E., Finnveden, G., Schmidt, S., Steinbach, N. (2019). Environmental pressures from Swedish consumption – A hybrid multi-regional input-output approach. *Journal of Cleaner Production*, 228(10), 634-644.

Remond-Tiedres, I., Rueda-Cantucho, J. M. (2019). EU inter-country supply, use and input-output tables — Full international and global accounts for research in input-output analysis (FIGARO). Publications Office of the European Union.

Roos, J. (2006). Identifying and measuring re-exports and re-imports, paper presented at the 7th OECD International Trade Statistics Expert Meeting, Paris. STD/NAES/TASS/ITS(2006)18

Tukker, A., Giljum, S., & Wood, R. (2018). Recent progress in assessment of resource efficiency and environmental impacts embodied in trade: An introduction to this special issue. *Journal of Industrial Ecology*, 22(3), 489-501.

Walker, A.N., Lemmers, O., Horlings, E., Zult, D., Wijnen, V., Hooijmaaijers, S. (2023). Plan van aanpak voor een vervanger voor het SNAC-proces. Statistics Netherlands project PR0001422.

8. Description of the input , intermediary and output data files

This section describes the data files that are needed to run the scripts from the previous section, their intermediate and final results – the output. Depending on the data situation, NSIs can adapt the data and/or R-scripts to match their possibilities. It is not necessary to have the same data situation as in the Netherlands. The detailed data section (3.2) provides several examples and solutions when a certain data element is not available.

Statistics Netherlands created dummy data to illustrate the method and shared this data with Eurostat. This dummy data has almost no link with reality. For example, import values, VAT numbers and enterprise numbers were randomly generated. For each record in trade in goods and services, country and product were sampled separately from real data. Furthermore, the data is at less detail in industries and products than is internally available at Statistics Netherlands. However, the R-scripts are flexible, and will run properly with different industries and products as well.

The files are described by folder; this is the folder where the files have to be put.

8.1 Input files

8.1.1 Comtrade

Distribution999_2020 is an example of a file with a country distribution for a commodity in National Accounts where the trade data has no information about this commodity. In that case, the country distribution is taken from UN Comtrade. In the detailed Dutch data, such commodities exist. In the shared aggregated and dummy data, such commodities do not exist. The file contains the variables

- commodityIOT, internal, more aggregated level of products, used at National Accounts
- partnercountry, code for the partner country involved in trade
- share, the share of this country in imports

8.1.2 Goods

Traderproperties2020 is a file with information from the business register, to be linked to traders. It contains the variables

- enterprise_imaginary, internal enterprise identifier, here a random/imaginary number
- enterprisegroup_imaginary, internal enterprise group identifier, here a random/imaginary number
- industry, industry of the enterprise

Tradedata2020_a is a file with trade data matched – wherever possible – to the business register. It contains the variables

- commodityCN, most detailed level of commodities, here the Combined Nomenclature of the EU
- commodityIOT, internal, more aggregated level of products, used at National Accounts
- taxnr_imaginary, VAT number of the trader, here a random/imaginary number
- partnercountry, code for the partner country involved in trade
- enterprise_imaginary, internal enterprise identifier, here a random/imaginary number
- enterprisegroup_imaginary, internal enterprise group identifier, here a random/imaginary number
- industry, industry of the enterprise
- imports, value of imports
- exports, value of exports
- re_ex, value of re-exports

8.1.3 Input-output table (iot)

Ciffob_1_2020D has information about the ciffob corrections on goods. It contains the variables

- type, the type of ciffob corrections
- ciffob_cor, the value of the correction

Ciffob_2_2020D has information about the ciffob corrections on services. It contains the variables

- commodityIOT, a code for the ciffob correction
- SectorUse, the sector that “uses” this correction, equal to “ciffob_cor”
- country, the partner country, in the fictitious data “unknown”
- value, the value of the correction
- type, the type of ciffob corrections

Colnames_iot_2020D has all possible column names in the input-output table, industries and types of final demand. It is a value and does not contain variables.

dIOT_2020D has information from National Accounts about imports of services and their use. It contains the variables

- commodityIOT, internal, more aggregated level of products, used at National Accounts
- SectorUse, the sector (industry or type of final demand) that uses the services
- Lp, the value of imported services
- SectorSupply, the origin of the supply, different types of imports of services, in the fictitious data either “P721” or “P722”
- ValuationLayer, the valuation of the data. In the fictitious data always equal to “80”.

dIOT_basis2020D has information from National Accounts about trade in commodities. It contains the variables

- commodityIOT, internal, more aggregated level of products, used at National Accounts
- industry, the industry or final demand category that imports
- importsIOT, value of imports according to the input-output table
- reexportsIOT, value of re-exports according to the input-output table
- M_per_X, the amount of imports needed for 1 unit of re-exports of this commodity

domestic_demands_2020D is a file very similar to dIOTbasis_2020D, the main difference being that it does not contain the re-exports (P613 and P614). It contains the variables

- commodityIOT, internal, more aggregated level of products, used at National Accounts
- SectorUse, the industry or final demand category that uses the imports
- demand, value of imports according to the input-output table

import_reexport_demands_2020D describes the destination of imports for re-exports to different categories, in this data destinations P613 and P614. It contains the variables

- commodityIOT, internal, more aggregated level of products, used at National Accounts
- SectorUse, the destination of re-exports, either P613 or P614
- importsIOT, the value of imports for re-exports

8.1.4 Production abroad

Final_prod_abroad_2020D is a file that describes production abroad. This trade data is included in National Accounts, but it is not included in trade statistics. The file contains the variables

- commodityIOT, internal, more aggregated level of products, used at National Accounts
- industry, industry that trades
- partnercountry, code for the partner country involved in trade
- M, value of imports
- commodityCN, a concatenation of “pa” (production abroad) and the commodityIOT. There is no commodityCN from Combined Nomenclature known.

8.1.5 Services

lhd_response_2020 is a file with response data from trade in services. This has the advantage that an industry can easily be assigned, contrary to imputed data. The file contains the variables

- service, the type of service, using the Eurostat nomenclature
- country, the country from which the services were imported
- industry, the industry that imports
- value, the import value
- source, equal to 1 when there is country information and equal to 2 when there is no country information. This reflects the structure of the Dutch international trade in services sources.

serviceDistributionTable is a csv file that links a type of service in the trade data with a type of service in the National Accounts data. The file contains the variables

- service, the type of service in the trade data, using the Eurostat nomenclature
- value, the percentage of this type of service in trade data that is assigned to this type of service in National Accounts
- commodityIOT, internal, more aggregated level of products, used at National Accounts

Total_2020 is a file from trade in services statistics that describes total trade in services. It contains trade of Special Purpose Entities (SPEs⁴), xxx, and non-SPEs. The file contains the variables

- country, the country from which the services were imported
- service, the type of service in the trade data, using the Eurostat nomenclature
- value_exclSPE, the value imported by non-SPEs
- value_SPE, the value imported by SPEs

8.1.6 Stocks

Negative_rows_2020D is used to account for negative numbers in final demand. In our fictitious data, these appear in several types of individual consumption expenditure (P31Ax), changes in inventories (P52x) and acquisitions less disposals of valuables (P53). The file contains the variables

- commodityIOT, internal, more aggregated level of products, used at National Accounts
- P31Ax, the summed value of several types of individual consumption expenditure
- P52x, the summed value of changes in inventories
- P53, the value of acquisitions less disposals of valuables

⁴ <https://www.imf.org/External/NP/sta/bop/pdf/diteg9-11-bp.pdf> states: “Special Purpose Entities or, as they are called in the Netherlands, Special Financial Institutions (SFIs) are in the Netherlands established companies or institutions, regardless their legal form, whose shares are directly or indirectly held by non-residents and are mainly dealing with receiving funds from non-residents and channelling them to non-residents. SFIs may also be involved in managing participations outside the Netherlands on behalf of the parent company.”

Negative_stock_2020D is a file that contains the information about negative stocks; stocks that were imported in an earlier year to be used in the current year to fulfill demand. The file contains the variables

- commodityIOT, internal, more aggregated level of products, used at National Accounts
- partnercountry, always equal to "FromStock" since there is no partner country information
- commodityCN, a concatenation of "stock" and the commodityIOT. There is no commodityCN from Combined Nomenclature known.
- M, the value of imports

8.2 Intermediate files

These files are created while running the scripts. They are intermediate files; these are not meant to be used.

8.2.1 Basisfiles

Basis_imports_distributed_2020D is the result of the process to assign imports (and partner countries) to the domestic use of commodities. The file contains the variables

- industry, importing industry
- commodityIOT, internal, more aggregated level of products, used at National Accounts
- commodityCN, most detailed level of commodities, here the Combined Nomenclature of the EU
- partnercountry, code for the partner country involved in trade
- imports, value of imports
- import_reexport, value of imports needed for re-exports
- M, value of imports not yet assigned
- M_needed, value of imports still needed for re-exports
- M_reexp, value of imports used for re-exports
- import_tot, total imports of this commodityIOT by this industry
- demand, total domestic demand of this commodityIOT by this industry
- M_domestic, value of imports used for domestic demand
- M_dom_demand, value of imports still needed for domestic demand
- com6, value distributed at commodityIOT (6 digit)
- com4, value distributed at commodityIOT (4 digit)
- com2, value distributed at commodityIOT (2 digit)
- com_industry, value distributed at commodityIOT by industry level
- com6_industry, value distributed at commodityIOT (6 digit) by industry level
- com4_industry, value distributed at commodityIOT (4 digit) by industry level
- commodity_only, value distributed at commodityCN level
- commodity6_only, value distributed at commodity level, using first 6 digits of the commodityCN
- commodity4_only, value distributed at commodity level, using first 4 digits of the commodityCN
- com2_industry, value distributed at commodityIOT (2 digit) by industry level
- commodity2_only, value distributed at commodity level, using first 2 digits of the commodityCN

Basis2020D is created by taking the import data from trade in goods statistics and merging it with and calibrating it to the import data from National Accounts. The file contains the variables

- commodityIOT, internal, more aggregated level of products, used at National Accounts
- commodityCN, most detailed level of commodities, here the Combined Nomenclature of the EU
- taxnr, VAT number of the trader, here a random/imaginary number
- partnercountry, code for the partner country involved in trade
- enterprise, internal enterprise identifier, here a random/imaginary number
- enterprisegroup, internal enterprise group identifier, here a random/imaginary number

- industry, the industry of the enterprise
- imports, the value of imports
- import_reexport, the value of imports needed for re-exports by this tax number. Note that there might be re-exports without corresponding imports at the tax number, e.g., when an affiliate imports and this tax number re-exports.

Import_reexport_2020D is the result of the process to assign imports (and partner countries) to the re-exports of commodities. It is very similar to the file basis_imports_distributed_2020D. The file contains the variables

- commodityIOT, internal, more aggregated level of products, used at National Accounts
- commodityCN, most detailed level of commodities, here the Combined Nomenclature of the EU
- taxnr, VAT number of the trader, here a random/imaginary number
- partnercountry, code for the partner country involved in trade
- enterprise, internal enterprise identifier, here a random/imaginary number
- enterprisegroup, internal enterprise group identifier, here a random/imaginary number
- industry, industry of the enterprise
- imports, value of imports
- import_reexport, value of imports needed for re-exports
- M, imports remaining after taking out imports for (own and of others) re-exports
- M_needed, imports still needed for re-exports
- d_tax0_CN8, value distributed by taxnr=0 (missing) and full commodityCN (8 digit)
- d_tax_CN8, value distributed by taxnr and full commodityCN (8-digit)
- d_tax0CN6, value distributed by taxnr=0 (missing) and commodityCN (6 digit)
- d_taxCN6, value distributed by taxnr and commodityCN (6 digit)
- d_enterprise_CN6, value distributed by enterprise and commodityCN (6 digit)
- d_enterprise_group_CN6, value distributed by enterprise group and commodityCN (6 digit)
- d_enterprise_comIOT, value distributed by enterprise and commodityIOT
- d_enterprise0_comIOT, value distributed by enterprise=0 (missing) and commodityIOT
- d_CN8, value distributed by full commodityCN (8 digit)
- d_CN6_comIOT, value distributed by commodityCN (6 digit) and commodityIOT
- d_CN4_comIOT, value distributed by commodityCN (4 digit) and commodityIOT
- d_CN6, value distributed by commodityCN (6 digit)
- d_CN4, value distributed by commodityCN (4 digit)
- d_comIOT, value distributed by commodityIOT
- d_comIOT2, value distributed by commodityIOT (2 digit)
- d_notresolved, the value of imports for re-exports that was not resolved
- M_reexp, imports remaining after taking out imports for (own and of others) re-exports

8.3 Output files

These files are created while running the scripts. They are output files containing information about the countries from where what was imported for whom.

8.3.1 Results

Imports_distributed_2020D contains the results for imports of goods by commodity, by country, by user (either an industry or a category of final demand). The file contains the variables

- commodityIOT, internal, more aggregated level of products, used at National Accounts
- partnercountry, the country from which the commodities were imported

- commodityCN, most detailed level of commodities, here the Combined Nomenclature of the EU
- type, provides information where in the process the link was made between partner country and use
- A01 etc., the value of imports by this industry/category of final demand

Imports_distributed_2020D_ciffob is very similar to the file imports_distributed_2020D; the only difference is that the ciffob correction has been added. The file contains the variables

- commodityIOT, internal, more aggregated level of products, used at National Accounts
- partnercountry, the country from which the commodities were imported
- commodityCN, most detailed level of commodities, here the Combined Nomenclature of the EU
- type, provides information where in the process the link was made between partner country and use
- A01 etc., the value of imports by this industry/category of final demand

Services_distributed_2020D contains the results for imports of services by product, by country, by user (either an industry or a category of final demand). The file contains the variables

- commodityIOT, internal, more aggregated level of products, used at National Accounts
- SectorUse, the industry or final demand category that uses the imports
- country, the country from which the services were imported
- value, the import value
- type, provides information where in the process the link was made between partner country and use

Total_imports_distributed_2020D contains the results for imports of goods and services by product, by country, by user (either an industry or a category of final demand). The file contains the variables

- commodityIOT, internal, more aggregated level of products, used at National Accounts
- partnercountry, the country from which the goods or services were imported
- commodityCN, most detailed level of commodities, here the Combined Nomenclature of the EU
- type, provides information where in the process the link was made between partner country and use
- SectorUse, the industry or final demand category that uses the imports
- value, the import value

9. Description of the R-scripts

This section briefly explains the R-scripts that were produced for this project. Since every country has a different data situation, it is not expected that another country can run them immediately. The R-scripts have to be adapted, but they will be valuable anyway.

In total, 14 scripts have been compiled to transform the input data to the output data. They are grouped into 7 stages:

- A. Create the basic data from trade in goods and National Accounts
- B. Merge the two data sets and calibrate trade in goods to National Accounts
- C. Estimate imports for re-exports and assign country detail
- D. Add country detail to imports for domestic use
- E. Final steps in trade in goods
- F. Assign country detail to imports of services
- G. Combine data of goods and services and create final output file

Additionally, there are the scripts **0_run_all** and **m_logmessage**. The first script sets several parameters that are used in the scripts and then runs them all. The second script ensures that the log is saved to a file.

We will now provide a brief description of each of the scripts.

A. Create the basic data from trade in goods and National Accounts

This part creates the basic data. It is here that the differences between National Statistical Institutes will be the largest. Therefore, we describe the scripts below and their output in the next section. Statistics Netherlands did not share these scripts with Eurostat.

A1_create_trader_properties_data_cbs collects all enterprises in the Netherlands in a given year, their corresponding enterprise group, international industry classification (SBI) of the enterprise and industry classification of the enterprise in the Dutch system of National Accounts (regkol).

A2_create_trader_data_cbs transforms the basic trade data and links it to enterprises. The basic trade data contains the tax number (VAT number), commodity (CN8), partner country, value of imported goods, value of exports of Dutch production, value of re-exports. Note that the trade data is adapted from the original trade in goods statistics; it is as close as possible to the concepts of National Accounts (BPM6/SNA 2008). The trade file is subsequently matched to a file that links traders and enterprises. It contains a percentage that shows how much of trade of this trader is assigned to that enterprise. In the next step, the information from the script A1 is matched to add industry classifications.

A3_create_national_accounts_data_cbs imports National Accounts data. First, the data about imports for domestic use. Second, the data about re-exports and all its inputs (among them are imports). Then a table **M_per_x** is created with the amount of imports needed per unit of re-exports at product level. Subsequently, all the data compiled in this step is linked resulting in 1 dataset.

B. Merge the two data sets and calibrate trade in goods to National Accounts

B_merge_and_calibrate ensures that the data from trade in goods statistics is calibrated to the trade data from National Accounts. However, in the case that import data from the first source differs too much from data from the second source, it is only scaled by a certain amount. The scaling is at the level of commodities of National

Accounts. All 8-digit commodity codes that are part of the same National Accounts commodity are scaled in the same way. In the last step, imports needed for re-exports by this trader are estimated with the earlier M_{per_x} data. This is at commodity level of National Accounts.

C. Estimate imports for re-exports and assign country detail

C_distribute_reexports is based on Lemmers & Wong (2019). It assigns imports for re-exports and the country of these imports. If a trader has sufficient imports for re-exports at commodity level, and 10% of his imports are from country A, then it is assumed that 10% of imports for re-exports are from country A as well. If there are not sufficient imports, the imports that are available are all use for re-exports. And in next steps the remaining imports are sought for. For example, they may not be at trader (VAT number) level, but at the enterprise, or at the enterprise group.

C1_aggregate_and_distribute_by is a script that is being used in the previous script. At a given level (trader – CN8, trader – CN6, enterprise – National Accounts commodity, enterprise group - National Accounts commodity, National Accounts commodity) it considers imports available, assigns them to re-exports, calculates how much is left/how much is still needed.

D. Add country detail to imports for domestic use

D_distribute_domestic_use continues with the imports that are not for re-exports. These are for domestic use and we apply the method described in Aerts et al. to add country detail. From National Accounts we have imports by industry and for final demand, by commodity, but not by country. From the previous step, we have imports by industry, commodity and country. These data are matched. Similar to the script C, the trade data is as much used as possible to assign country detail to the National Accounts data. Ideally, at product by industry (trade data), if not, less detailed product by industry (trade data) and so on. Note that the National Accounts data also contains production abroad. These imports are not in the trade in goods data. Therefore, these imports are assigned to “unknown origin”. Furthermore, there might be some products that are in the National Accounts data but not in the trade in goods data. These are added in this script as well.

D1_demands_for_domestic_use prepares an input file for the previous script. Basically, it has three parts, all in the realm of National Accounts. First, imports by industry at commodity level. Second, imports for stocks. This might be negative, namely when something is taken out of stocks and not from imports. Information of the year before for imports and country distribution is used. Third, imports for production abroad.

D2_aggregate_and_distribute_by_with_industry is a script that is similar to **C1_aggregate_and_distribute_by**. It links the trade data (commodity, industry, country) with National Accounts data (commodity, industry). Since this does not match 100%, this script solves that problem by first doing as much as possible at a detailed level (commodity by industry). Then at a less detailed level (commodity with less detail by industry) with the remainders, an even more less detailed level with the remainders and so on.

E. Final steps in trade in goods

E_check_sums checks whether the resulting files are in agreement with the original files of National Accounts. It checks totals of imports, imports for national use, imports directly used for exports, margins on imports, and imports directly used for consumption. It also checks these variables by industry. When the new data is not well-aligned with the original files, the script reports that at total level (or in which industry) something is not correct.

E1_ciffob_corrections adds the ciffob corrections from the IO-table in National Accounts to the dataset.

F. Assign country detail to imports of services

F_services_distribution breaks down the imports of services by type of service by industry (from National Accounts) by country. In the first step, government services, travel services (consumption abroad) and services from Special Purpose Entities (SPEs) are tackled. The trade data does not contain microdata and therefore no industry data. The added country information is the average distribution for imports of these services. In the second and third step, country detail is added using the survey information on trade by service type by country by enterprise. Ideally, everything is distributed at the full industry detail at National Accounts (step 2). Sometimes this is not possible and distribution is at less detailed 2 digit industries (step 3). These two steps use the script `F1_distribute_perserviceType`. They do not yet assign all trade in services to countries. In step 4, the rest of imports is assigned to countries. This is by the average origin of the specific service type, taking into account the earlier assigned trade to prevent bias.

F1_distribute_per_serviceType provides the ground work for step 2 and 3 of `F_services_distribution`. It assigns as much as possible from the survey in trade in services to industries in National Accounts. When that is not possible, it assigns trade to industries at a less detailed level.

G. Combine data of goods and services and create final output file

G_create_output_file combines the files for goods and services into one final output file. The data of National Accounts, import value by product by industry, is broken down by the country from which the Netherlands imports. For each commodity an indicator is added that shows the industry that commonly produces this commodity. This facilitates linking to a producing industry abroad.