Department for Environment
Food and Rural Affairs

Measuring the impacts on global biodiversity of goods and services imported into the UK

Methodology Report

February 2013

Simon Croft (Stockholm Environment Institute, University of York)
Elena Dawkins (Stockholm Environment Institute, University of York)
Chris West (Stockholm Environment Institute, University of York)
Cecile Brugere (Stockholm Environment Institute, University of York)
William Sheate (Collingwood Environmental Planning Ltd)
David Raffaelli (Environment Department, University of York)

SEI STOCKHOLM ENVIRONMENT INSTITUTE

THE UNIVERSITY of York

Collingwood Environmental Planning

Contents

1. Introduction

Continued growth in global demand for goods and services is a major contributor to on-going environmental degradation. However, increasing trade liberalization and the resulting increased complexity of supply chains has led to the geographic separation of production-driven environmental impacts from the point of consumption. Although able to acknowledge their potential importance as drivers of environmental and societal impact, consumers and policy makers therefore find it increasingly difficult to understand the real-world consequences of their consumption.

This study focuses on one potential impact of the pressure that growing demand and resource extraction and use place on the environment – the loss of biodiversity. The main drivers of biodiversity loss globally are: habitat change (including land use change and modification of rivers and water withdrawal); over-exploitation; invasive alien species; pollution; and anthropogenic climate change. As the project focus is UK consumption, we concentrate on identifying production activities overseas that generate export goods for the UK that could also lead to local biodiversity loss. The most likely direct and pertinent drivers in this case will be habitat change, overexploitation and pollution.

We will address this challenge by linking two distinct research streams: the analysis of trade pathways and supply chains (both direct and indirect) for both goods and services to identify important sources of production; and identification and analysis of production systems and products and their impacts based on geographically specific biodiversity-relevant information. Combining these research streams involves overcoming issues with data-availability (for trade, production, and biodiversity) and assessment of complex resource pathways and environmental interactions.

This document focuses on the methodology used to link financial trade data with physical production and environmental data. The result of the methodology is an allocation of production, in terms of tonnes of a good produced, to UK consumption. This production information can then in turn be linked to biodiversity drivers and impacts.

2. Background to the method development and project

The UK Department for Environment, Food and Rural Affairs (Defra) have commissioned this study to investigate the impacts that consumption in the UK has on biodiversity overseas. The main objective of this work is to provide a database-driven methodology for linking UK imports to geographically-defined impacts on biodiversity. Researchers at the Stockholm Environment Institute (SEI), University of York, with support from Collingwood Environmental Planning (CEP) have proposed a method to answer this question which utilises and applies recent developments in multi-regional input-output (MRIO) modelling. The method is outlined in this paper. The research project started in November 2011, with a final report to be published and database provided by March 2013.

An interim report was provided to Defra in March 2012 and included a review of existing methodologies, trade and consumption data availability and biodiversity indicators. Following this process the project team recommended that a multi-regional input-output approach is applied to ensure that all indirect impacts of UK consumption are captured (a key requirement of the project brief). From knowledge of MRIO approaches, and previous work conducted in conjunction with partners on the OPEN:EU project (http://www.oneplaneteconomynetwork.org/), we then recommended the development of a methodology, based on the latest footprinting analysis (as presented by Ewing et al. (2012)), to link financial multi-regional input-output (MRIO) tables with physical production and trade information. The main advantage of this approach, as highlighted by Ewing et al. (2012), is the maintenance of product-level detail traditionally found in physical flow accounting, combined with the full supply-chain detail that MRIO analysis provides. This approach has previously been used to create the 'Footprint Family' of indicators for tracking human pressure on the planet, as described in Galli et al. (2012).

Due to its capacity to retain product-level detail and full supply chain impacts, this MRIO-F approach is an appropriate technique to begin to link UK imports to geographically-defined impacts on biodiversity. However, whilst the MRIO-F method is particularly useful for tracking the impacts associated with consumption for any pollutant or resource use that has the same effect regardless of the origin of production (such as greenhouse gases), impacts that vary depending on the local conditions (such as water use or biodiversity) require more information to attempt to link consumption to the actual impacts on the ground. The methodology described in this paper will therefore be used to extend these latest footprint family developments and MRIO-F modelling advances to capture both a broad set of indicators of the drivers of biodiversity loss, along with possible methods to link these to data on local impacts and localised effects.

We start with an overview of the methodological approach proposed for this study which in theory can be applied to any production system given sufficient production and trade information. Details are given of the assumptions taken to link the physical and financial data sets. We explain how a GTAP8 (Global Trade Analysis Project: https://www.gtap.agecon.purdue.edu) MRIO table linked to physical agricultural production data from FAO is implemented based on the work of Ewing et al. (2012) and Peters, Andrew, and Lennox (2011). In the case where extensive trade information is not readily available we detail an alternative technique for linking the physical data to the MRIO table. This is briefly summarised following the explanation of the physical use table extension. We finish with an overview of how production data may be linked to drivers of biodiversity loss. We hope that this will contribute to the further development and progress of these types of indicators and further raise the profile of monitoring both direct and indirect impacts of activities through consumption and trade-based methods.

3. Methodology and model development

The following section describes the method used to link physical data to a financial MRIO table.

3.1. The model structure overview

The basic data structure of the model is based on environmentally-extended input-output analysis, with the inclusion of the latest developments for incorporating physical data described by Ewing et al. (2012) and is shown in Figure 1.

Ewing et al. (2012) describe various different options for linking the monetary and physical datasets; the simplest approach for integrating physical data with a financial MRIO is a direct allocation of all primary products to their producing MRIO sectors, and the most complex would be to use physical data directly within the MRIO table. The second option, whilst theoretically more powerful, would require disaggregation of the MRIO sectors which is a resource intensive task and requires data that is not consistently available across all products. The first option is useful and applicable when physical data are limited and is described in more detail in Section 4.2.

As a compromise between the most simple option and a full sectoral disaggregation, Ewing et al. (2012) describe a method which involves the construction of a 'physical use' matrix. This has the advantage of maintaining product level detail in physical units, and can be linked to the MRIO table using datasets that contain highly aggregated information about product supply and use. In this case rows in the physical use table represent physical products by country, columns represent sectors of the MRIO model and the country where they are located. The advantage of this method is that detailed mass flow accounts (and therefore product-level information) can be maintained alongside the original monetary structure.

Constructing matrices of the use of physical commodities ('physical use tables') and linking these to the financial MRIO table is the most complex part of the model development and requires a number of assumptions that are detailed below. The research team have also developed a number of algorithms for overcoming some of the data restrictions. Whilst the methods used to develop this model have followed the mathematical descriptions outlined in Ewing et al. (2012), some assumptions and development approaches may differ in order to utilise the implemented datasets, chosen for their ability to accommodate potential biodiversity-specific impacts.

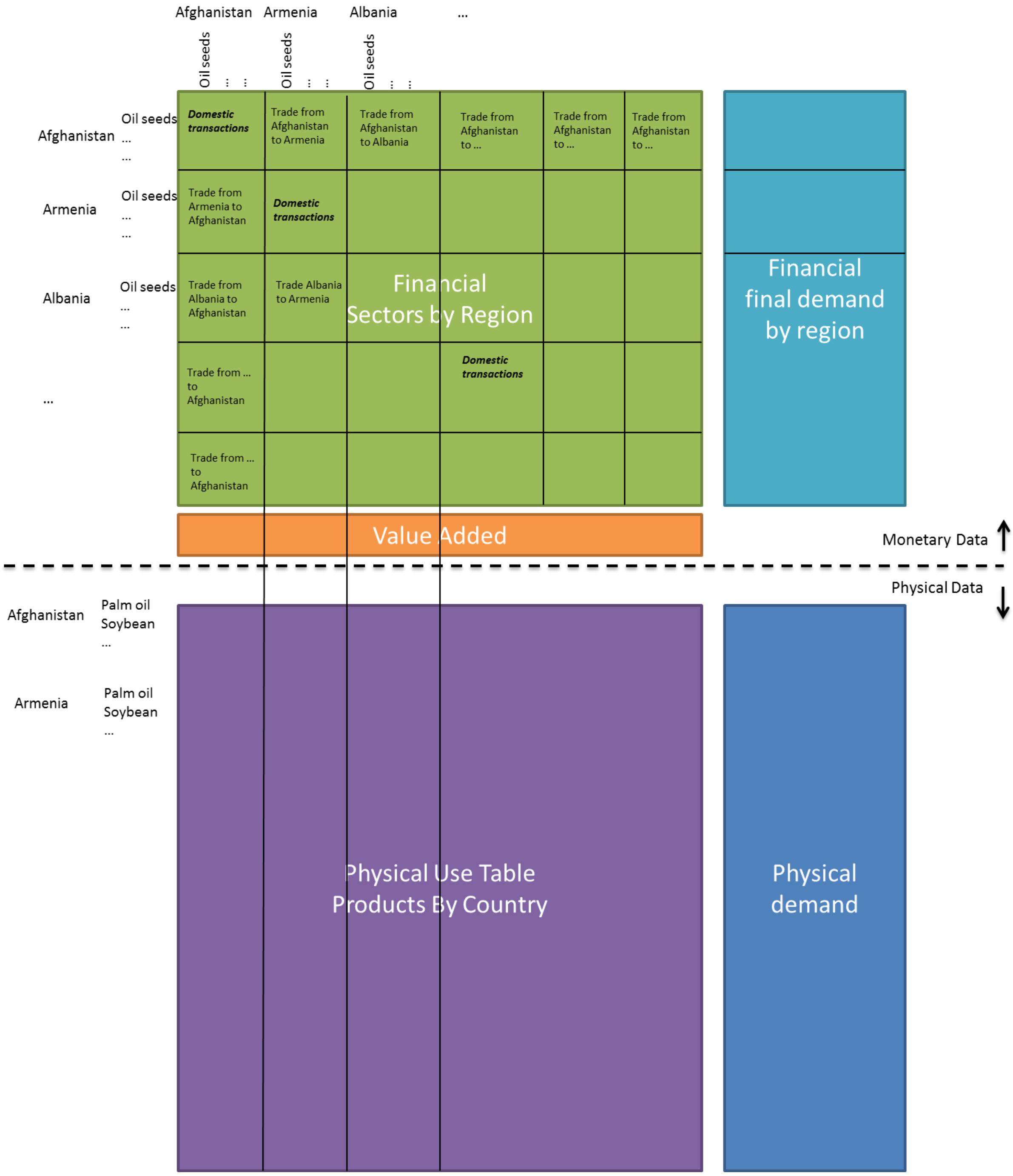

Figure 1: Diagram of MRIO physical model adapted from: Brad R. Ewing, Troy R. Hawkins, Thomas O. Wiedmann, Alessandro Galli, A. Ertug Ercin, Jan Weinzettel, Kjartan Steen-Olsen, Integrating ecological and water footprint accounting in a multi-regional input–output framework, Ecological Indicators, Volume 23, December 2012, http://www.sciencedirect.com/science/article/pii/S1470160X12000714

3.2. Data preparation for linking physical and MRIO tables

3.2.1. MRIO table

In order to track both direct and indirect flows of goods and services through the world economy (and therefore also for UK consumption) it is first necessary to construct a multi-regional input-output (MRIO) table, which represents the supply and use of goods and services, in monetary values, between numerous world regions and their industrial sectors. We have chosen to use GTAP data to construct our MRIO, following the method of Peters, Andrew, and Lennox (2011). Further details of the construction of the GTAP data are given in Sections 4.1.1. and 4.1.2.

3.2.2. Physical data collection

To create a product-level physical use table, a dataset that provides production and trade data for many products across multiple countries (in physical units) is required. If this trade information extends to physical trade between sectors then allocation based on the detailed physical data can be completed, however, in the majority of cases trade information is limited to inter-region trade, in which case other data must be used to inform use across sectors.

3.2.3. Concordances between datasets

In order to link the two datasets, products, sectors, regions and countries must be matched so that the physical data can be allocated to the most appropriate monetary sector in the MRIO table. As there is no international harmonised system for physical and monetary accounting, the MRIO and physical product data sources are likely to contain different region and industry categorisations. Therefore a concordance table to link the datasets must be developed. This is often based on manual matching between classification systems and various assumptions taken according to product descriptions and any existing published concordances.

3.3. Creating and applying the MRIO-physical use model

3.3.1 Creating the physical use table

Assume physical production data covers R countries and S products, and the MRIO table provides information for R* regions and S* sectors:

Physical production data is in form of an R x S matrix, with each element a quantity of production for one product in one country.

The physical trade data is represented by an R x R x S three-dimensional matrix, with each "layer" detailing the trade for a specific product, with each value in that layer a quantity exported from one region to another. Diagonal entries within a layer incorporate the production information and trade information to represent domestic use of domestic production. For a chosen product within the physical dataset, it follows that production/trade is therefore represented by an R x R matrix.

Applying the concordance relationship between the physical production/trade data and the financial MRIO table, each country in R is related to its respective region classifications within R*. The data can then be transformed into an R x R* matrix, with all exports (and domestic use) from regions in R allocated to regions in R*.

This newly allocated trade is then distributed amongst all sectors (S*) and final demand (y within MRIO framework) within each region in R*, weighted according to industry and final demand for imported sectors from the MRIO data. The data is therefore in the form of an R x (R* x S*) two-dimensional matrix, with the exports from each region in R (rows) distributed and assigned to each sector in S* within each region in R* (columns).

Each row in this matrix is then divided element-wise by the x vector (total output) calculated from the MRIO dataset to ascertain the physical quantity demand per monetary unit of output for each of the regions and sectors within the MRIO dataset.

3.4.3 Applying Leontief

Once the physical use and MRIO table have been constructed the model must be run to calculate the use of physical products associated with a given demand.

The Leontief inverse is calculated from the MRIO data and multiplied by the physical quantity demanded per unit financial output. Multiplying this by the final demand data (y) from the MRIO model produces an (R* x S*) x (R x R*) matrix, with a row for each GTAP sector within each GTAP region, and a column for production in each country in R due to demand in each region in R*.

Production due to final demand from a specific region therefore forms a subset of this result. To obtain production due to demand in a single region, in the above step it is only required to multiply by the final demand vector for the desired region.

4. Implementation of the method with real data

The preceding section includes an overview of the methods. This section explains how we have used available trade and production data to construct the MRIO and physical use matrices, along with the assumptions required to do this.

4.1. Data acquisition and preparation

4.1.1. GTAP8 (2007) Data

GTAP8 is the latest dataset to be released by GTAP, and includes both 2004 and 2007 datasets for 57 sectors within 129 global regions. To work with the GTAP8 database, it is necessary to download the relevant dataset and accompanying programs from the GTAP website and then extract the required information. This data extraction was done by first outputting a disaggregated set of the most recent (2007) data (any desired aggregation can later be performed) in the GTAPAgg program which comes with the download. The outputted ZIP file contains a file ‘basedata.har’ which includes all of the required data. An SEI-created ‘.bat’ file then uses the ‘har2csv’ executable (obtained by downloading the GEMPACK software suite) to extract a number of different arrays of data in nine separate ‘.txt’ files.

4.1.2. Creation of an MRIO table from GTAP8 database

We followed the method for constructing a MRIO table (of 57 sectors and 129 world regions) from the GTAP database given in Peters, Andrew, and Lennox (2011). A custom Matlab script, created by SEI, was used to read and import the data from the GTAP ‘.txt’ files and generate Z, y, X and L matrices (for the MRIO analysis), which are output and saved in both .csv and .mat formats. These are then in a usable form for direct examination or use in Matlab for further calculations, respectively.

For each region within the GTAP database the geographic origin of commodities is known, however information about the distribution of these commodities across sectors is resolved only in terms of 'domestic' or 'imported' commodity use. An implicit assumption in the MRIO table, therefore, is that imports of a given commodity to a specific region are treated identically in terms of their distribution across different sectors, regardless of their country of origin. Similarly, due to the concordances between the two datasets, different commodities end up being treated in the same way. For example, "Asparagus" and "Strawberries" will both come under the "Vegetables, fruits, nuts" sector in the GTAP dataset and consequently be distributed across sectors according to the same weighting.

4.1.3. Physical production: FAO (2007) data

We have chosen to focus primarily on products of primary agricultural production (but see Section 4.2) as FAO statistics on these products are detailed in comparison with other physical commodity datasets. FAO datasets can be downloaded directly from the faostat.fao.org website. Choosing 2007 data to match the latest available GTAP data, different datasets need to be downloaded for different commodity types, and for each commodity type there are datasets for production and trade (as well as additional information). The data is downloaded in '.csv' format which can then be imported into Matlab where SEI's custom scripts reorganise the data into a usable 'physical use' format. This reorganisation consists of a number of methodological steps and involves tackling a number of data issues:

4.1.3.1. Region irregularities

There are some issues/irregularities within the FAO data. For example, Chinese production and trade data within the FAO dataset is described under a number of different country names and codes. There are four sub-regions ([41] `China: Mainland'; [96] `China: Hong Kong SAR'; [128] `China: Macau SAR' and [214] `China: Taiwan province of'), and two listings simply under the name `China' (country codes [351] and [357]).

All production data is assigned to country listing [351] `China', with no listings reported for the other entries. However, the trade data is much less consistent; exports from Chinese regions are reported by entries [96], [128], [351] and [357], and entries [41], [96], [128], [351] and [357] are also reported as exporting partners by non-Chinese trade partners.

After consulting with FAO on the subject, clarification was received that the four sub-regions should sum to the value of the [351] `China' listing, and that [357] `China' is an aggregation of [41] `China: Mainland' and [214] `China: Taiwan province of'. It was further clarified that whilst production was originally reported in disaggregated form for the four sub-regions, production is now only reported in aggregated form for all of China (i.e. entry [351]) due to changes in data acquisition in recent years.

Given the importance of production data (and specifically production origin) in the methodology for assigning impacts, it is necessary to consider Chinese trade in the same aggregated form for consistency. It would therefore be preferable to just take the aggregated values for trade to and from entry [351] `China' as an overall value for China's trade. Unfortunately, not all regions report this value, and so in these cases an aggregation has to be made manually from the data. The easiest and most efficient way to do this is to either sum the values for the four sub-regions or, equivalently, sum the reported value for the semi-aggregated [357] `China' region and the remaining two sub-regions. However, the inconsistent data reporting does not make such an approach viable. Some countries only report an aggregated trade value, some report for the four sub-regions and others report a combination of semi-aggregated and sub-region. To complicate things further, some report a mixture of the above methods, often with “duplicated” entries.

As such, a careful and methodical approach is taken to construct the data for China. As stated above, production data is simply extracted for the [351] `China’ entry. For trade, the processing of the data is split into two parts to handle exports to and exports from China, though the methodology is effectively identical.

For reported exports from China, reported trade from any of the six Chinese regional entries is isolated for each product and to each of the other FAO countries. If it exists, trade reported from the four sub-regions is summed, as is trade reported from [357] `China’ and the [96] and [128] sub-regions. These two summed values are then compared to any reported values for [351] `China’, and the largest of these three values is assigned as the total Chinese export of the given commodity to the given region. Exports to China are handled in the much the same way. For a given product and exporting country, all reported exports to any of the six Chinese regional listings are isolated, and the same aggregations as above are performed. As with outgoing trade, the largest of the three attained values is chosen as the export value from the specified region to China.

The choice of the largest of these values is an attempt to account for gaps in the data. When a value is reported for [351] `China’, as an aggregation of the four sub-regions, it should theoretically be larger than or equal to (depending on data reporting) any of the summed values. Checking the data reveals that this holds for all reported trade data. In effect, the algorithm takes the value for trade to/from [351] `China’ where it exists, and when it doesn’t takes the largest value from the sum of the four sub-regions or the sum of the semi-aggregated [357] `China’ and the [96] and [128] sub-regions.

4.1.3.2. Incomplete data

The FAO data is downloaded in the form of ‘.csv’ files containing matrices comprised of columns denoting codes for the reporting country, partner country (where applicable), and commodity, and a column for the relevant data value (e.g. quantity). The data consists only of reported values for production and trade, and so needs to be expanded into a sparser dataset (i.e. have gaps in the data filled with zero-valued entries where no production or trade occurs) in order to create a full trade list which can be formed into a complete inter-regional trade matrix.

4.1.3.3. Data irregularities

The FAO export data includes an “unspecified” region entry for exports reported with no detail of the recipient country. There are different options available for dealing with this data. One option is to make assumptions based on the trade patterns of the countries that have listed exports to an unspecified region, and extrapolate this pattern to redistribute the unspecified exports to other countries. Such assumptions would be difficult to justify, however. An alternative option is to remove the unspecified region from the data. This will affect the total exports, and in turn the net balances, of all countries reporting exports to an unspecified region since they will no longer register in the data. A third approach, that is currently being adopted, is to treat the “unspecified” regions as a country in itself. Since there are no reported imports from unspecified regions (and obviously no reported production data), this new “country” exists within the data purely as an importer (and in our model has no concordant region within the GTAP MRIO and is therefore excluded from the MRIO analysis). Whilst this approach does not in any way attempt to reallocate the exports to their true recipients, as mentioned above the assumptions necessary to do so are problematic to defend. Unlike the second option mentioned above, this approach does leave the total export quantities intact. As a result, whilst the data is not precise in terms of destination for these exports, it is as complete as possible in terms of accounting for absolute export quantities (which is important for calculating re-exports; see Section 4.1.5 and Annex 2) without the need for contentious assumptions.

4.1.3.4. Import/Export mismatches and data reporting errors

Available crop data from FAO consists of production, export and import quantities. Export data and import data often do not match up (i.e. what Region *i* says it exports to Region *j* does not match what Region *j* says it imports from Region *i*), and net balances for regions (i.e. production + imports - exports) can also be negative.

The mismatch of import and export data can be caused by a number of factors, such as time lag due to transport (e.g. product leaves exporter in December and arrives with importer in January), differing reporting periods (e.g. one region reports annually over the period January-December, another region reports annually over the period April-March) and misclassification of products (e.g. exporting region classifies product as 'Wheat', importing region reports same product as 'Miscellaneous grains') to name a few. Human error and mistakes may occur in addition.

A negative net balance for a region can occur as a result of the above discrepancies in the reporting of production, import and export data, but can additionally also occur when all reporting is theoretically correct. For example, a commodity such as wheat grain can be stored for significant periods of time and in significant quantities. Consequently, if one year the wheat grain harvest is particularly poor in a given region, production quantities will be lower than usual yet export quantities may remain unaffected due to the use of stored wheat grain to meet demand. As such, a negative net balance should not be assumed to be incorrect.

As the import and export datasets do not match up, it is necessary to pick one or the other (or some combination of the two) to describe trade. It has been decided within this project to utilise the export dataset to construct the trade tables since exploration of the data reveals it to contain more information than the import dataset.

4.1.4. GTAP-FAO concordance

The FAO and GTAP datasets differ significantly in their country/region and commodity/sector categorisations, with many regions and sectors within the GTAP dataset being aggregations of the countries and commodities appearing within the FAO data. Consequently, it is necessary to construct concordance matrices to link countries and commodities within the FAO dataset to the regions and sectors within the GTAP dataset, respectively (see Equation 1 and Equation 2). For example, the FAO dataset contains production and trade information for "Asparagus", whilst in the wider GTAP sectors this would be aggregated with other items under the much broader "Vegetables, fruits, nuts" sector. Similarly, the FAO dataset contains explicit data for "Papua New Guinea", but within the GTAP dataset this is grouped within the "Rest of Oceania" region. The Country-Region concordance between the FAO and GTAP datasets is relatively simple to construct since a large number of countries in the FAO dataset have direct counterparts in the GTAP listings and the database documentation on the GTAP website is comprehensive in its description and breakdown of the regional aggregations. Comparatively, composing the Product-Sector concordance table is a much larger undertaking; very few products have direct equivalents, and whilst documentation also exists regarding the GTAP sectoral groupings it is less comprehensive. Although some products are very easy to assign to the relevant sector, others require more consideration and we have used our judgement when assigning concordance.

$$R(r_1\, r_2 \dots r_{236}) \quad \rightarrow \quad R^*(r_1^*\, r_2^* \dots r_{129}^*)$$

Equation 1: Regional concordances from the FAO to the GTAP datasets with 236 counties reassigned to 129 regions. Set R denotes the set of countries in the FAO dataset, and set $R^*$the set of regions in the GTAP dataset.

$$S(s_1\, s_2 \vdots s_{183}) \quad \rightarrow \quad S^*(s_1^*\, s_2^* \vdots s_{57}^*)$$

Equation 2: Commodity/sector concordances from FAO to GTAP datasets see an even greater degree of aggregation. When considering FAO data for primary agriculture products, 183 commodities have been assigned to the 57 sectors within the GTAP dataset. Set S denotes the set of commodities in the FAO dataset, and set $S^*$the set of sectors in the GTAP dataset.

4.1.5. Calculating re-exports

The reported import and export data provided by FAO does not necessarily provide information regarding the true origin of the imported/exported product; i.e. no distinction is made between the export of a domestically produced product and the re-export of a previously imported product. In order to correctly assign the impact of production to the final consumer, re-exports need to be considered.

Whilst the data in FAOStat for crop products contains no explicit information regarding re-exports, by making certain assumptions and developing and applying an algorithm it is possible to use implicit information within the data to form an approximation of re-exports. At the broad scale, three possible options are:

- (a) Available natively produced products are assumed to be exported before any imported goods are re-exported.
- (b) Available imported goods are assumed to be re-exported before any natively produced products are exported.
- (c) Exported products consist of both natively produced and imported products, which are exported in proportion to available quantity (or some weighted variation thereof).

A simple example of these options is provided in Annex A. In each case the region exports the same quantity of goods and is left with the same net quantity. What changes is the composition of the export and, consequently, the remaining products which the region retains.

In general, in Case (a), if production is greater than total exports, then no re-exporting will occur. Conversely, in Case (b), if total imports are greater than total exports, then no exporting of the domestically produced commodity will occur, and all outgoing trade will be in the form of re-exports. Case (c) ensures a combination of exports of natively produced products and re-exports (wherever production and trade of the same commodity occurs within a given region), but requires rules to be specified to govern this.

As well as these differences, each of these options depends on common broad assumptions about trade behaviour, as well as production, in order to be implemented. The FAO trade data account for production and trade information aggregated over an annual period, and therefore provide no detail regarding when a given commodity was produced or exported/imported within this period. Certain commodities will see higher levels of temporally correlated production (for example seasonal products) and this will possibly manifest itself as periodic patterns in trade. This is a common issue when trying to assign re-exports to periodically reported trade data.

Another issue arising when addressing re-exports is the previously mentioned negative net balances of production, imports and exports. If one is assuming that exports consist of a combination of natively produced products and re-exported imports, when a negative net balance occurs (that is, production + imports – exports < 0) it must be decided from where the commodity to meet this extra demand is originating.

We have selected case (c), weighted according to available quantities of domestic and imported commodities, for use within this project. Full details of the algorithms developed to account for these re-exports are presented in Annex B.

4.1.6. Creating the physical use table

This step is achieved by using the FAO physical dataset to define production and inter-country trade quantities, and the GTAP MRIO monetary table to disaggregate the inter-region trading down to the inter-sector trading level.

Combining the two datasets in this way creates a new dataset which is a combination of both the FAO countries and commodities and the GTAP regions and sectors. For a given country and commodity within the FAO dataset, the final use and demand is presented in the form of a list of the GTAP regions and sectors.

4.2 Dealing with other physical datasets

When dealing with other production datasets, the information present is often not as complete as that which can be derived from the FAO datasets for agriculture. In these instances, the methods described in Section 3.3.1 for implementing the data and creating the physical use table needs to be suitably adjusted.

The FAO datasets for fisheries production and trade provide a good example. Whereas the agriculture data provides detailed accounts of production and inter-country trade for each product (where applicable), the fisheries data provides values for total production, import, export and re-export quantities. Whilst the addition of re-export data is useful as it avoids the need for estimation (see Section 4.1.5), for each product in each region only four values are available: production, total imports, total exports and total re-exports, with no information as to the origin or destination of imports, exports and re-exports.

In such cases it is not possible to make a full physical use table from the available data and the simplest alternative workaround, as highlighted in Ewing et al. (2012), is to simply allocate the production information to the concordant producing region and sector within the GTAP data. Whilst this method is not as powerful as the physical use table approach, as the use of products is governed by the monetary sector-aggregated information within the MRIO, it allows retention of product-level production information. In cases where data is not sufficient to construct a full physical use table, we will adopt this alternative methodology.

5. Environmental extension overview

The hybrid MRIO system translates units of monetary final demand into the corresponding units of physical production (in tonnes for agricultural products) occurring in different countries that are required to fulfill demand. An environmental extension can be linked to this data if further information is known about the environmental impact associated with a unit of production. For the purposes of assessing potential impact to biodiversity, where impacts are generally locally specific, it is important that this impact data is resolved to country level (i.e. assuming adequate data, regional or world-average impacts should be avoided). Additionally, where possible, given the retention of product-level production detail it is also preferable to have product-level impact data.

Along with the production and trade values mentioned above, FAOStat also contains relatively detailed information about impacts associated with primary production. For example, yield and land use data for most crop types in most regions is available, which when scaled to the production figures coming from the model can estimate the land required for production. FAO also collates information on fertiliser application per unit production, although this dataset is relatively incomplete.

For the impacts of water use, the Water Footprint Network (WFN) (http://www.waterfootprint.org) provide information per unit of production for different crops in different countries for green, blue and grey water. Green water measures water stored naturally in the soil, blue water measures fresh surface and groundwater (i.e. that used for irrigation), and grey water measures water pollution. The WFN and FAO also provide information on water scarcity for different catchments/regions that can be used in conjunction with the water footprint data to assess the likely impact of water extraction vs. availability (see Lenzen et al (2012a) for an example).

The extensions above attempt to quantify potential country-level drivers of biodiversity loss. Biodiversity datasets are also available that can inform us of the likely biodiversity impacts associated with production. For example, WWF have compiled a list of Ecoregions which may be overlaid with country-level production. The IUCN prepare a global Red List of threatened species and BirdLife International identify Important Bird Areas. These datasets include details of the anthropogenic drivers of potential loss which can be linked to the associated production information from the model. A similar approach has been adopted in a recent study by Lenzen et al. (2012b).

When linking data quantitatively (for example assigning land use to production quantities), the data must be constructed to form a matrix with dimensions R x S (where R and S are the number of countries and products in the physical dataset, respectively), with each element in the matrix providing a unit impact per unit mass (e.g. for land use, ha/tonne) value for production of the corresponding product in the corresponding country.

Current output from the hybrid MRIO system is typically in the form of data for a single product (detailing the physical production in each country due to demand from each of the sectors in each region of the MRIO dataset). To convert this production information into environmental impact data, it is necessary to isolate the relevant data from the environmental extension matrix (the column corresponding to the product of interest), and multiply each element of the hybrid MRIO output by the appropriate entry (i.e. the corresponding country of production).

This environmentally-extended data will thus have the same dimensions as the hybrid MRIO output, but instead of describing the production in different countries due to demand across different regions and sectors, the data will now provide information as to the impact (e.g. ha of land used) associated with final demand.

6. References

Ewing, B. R., Hawkins, T. R., Wiedmann, T. O., Galli A., Ercin, A. E., Weinzettel, J. and Steen-Olsen, K. (2012) Integrating ecological and water footprint accounting in a multi-regional input–output framework. *Ecological Indicators*, 23, 1-8.

Galli, A., Wiedmann, T., Ercin, E., Knoblauch, D., Ewing, B. and Giljum, S. (2012). Integrating Ecological, Carbon and Water Footprint into a 'Footprint Family' of Indicators: Definition and Role in Tracking Human Pressure on the Planet. *Ecological Indicators,* 16, 100-112.

Lenzen, M., Bhaduri, A., Moran, D., Kanemoto, K., Bekchanov, M., Geschke, A. and Foran, B. (2012a) The role of scarcity in global virtual water flows. *ZEF-Discussion Papers on Development Policy,* 169.

Lenzen, M., Moran, D., Kanemoto, K., Foran, B., Lobefaro, L. and Geschke, A. (2012b) International Trade Drives Biodiversity Threats in Developing Nations. *Nature*, 486, 109-112.

Peters, G.P., Andrew, R. and Lennox, J. (2011) Constructing an environmentally extended multi-regional input-output table using the GTAP database. *Economic Systems Research*, 23, 131-152.

Annex A. Re-export options

As an example, assume a region has 1500 units of a domestically produced commodity, 1000 units of the commodity imported from other regions, and exports 2000 units.

In Case (a), as much of the demand as possible is met with domestically produced goods, with the remainder made up by re-exporting previously imported goods:

$$\begin{array}{ccc} Initial & Export & Remaining \\ \begin{pmatrix}1500\\1000\end{pmatrix}, & \begin{pmatrix}1500\\500\end{pmatrix}, & \begin{pmatrix}0\\500\end{pmatrix}. \end{array}$$

In Case (b), as much of the demand is met by re-exporting previously imported goods as possible, with domestically produced goods only exported as necessary to meet the demand:

$$\begin{array}{ccc} Initial & Export & Remaining \\ \begin{pmatrix}1500\\1000\end{pmatrix}, & \begin{pmatrix}1000\\1000\end{pmatrix}, & \begin{pmatrix}500\\0\end{pmatrix}. \end{array}$$

Finally, in Case (c), the trade is made up as a weighted proportion of exports of domestically produced goods (1500/(1500+1000) = 0.6) and re-exports of previously imported goods (1000/(1500+1000) = 0.4):

$$\begin{array}{ccc} Initial & Export & Remaining \\ \begin{pmatrix}1500\\1000\end{pmatrix}, & \begin{pmatrix}1200\\800\end{pmatrix}, & \begin{pmatrix}300\\200\end{pmatrix}. \end{array}$$

In each case the region exports the same quantity of goods (2000 units) and is left with the same net quantity (500 units). What changes is the composition of the export and, consequently, the remaining products which the region retains.

Annex B. Calculating re-exports

One algorithmic approach to calculating re-exports is outlined here with the example of a three-region model with one commodity. The three regions each produce a given quantity, x_i (i=1:3), annually. It follows that

$$Production = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}.$$

The trade of the commodity between the regions can be expressed by a number of trade terms, x_{ij} (i,j=1:3), where *i* denotes the exporting region and *j* the trade partner (importing country). It follows that in terms of a trade matrix, production can be considered as a region effectively trading with itself:

$$Production = \begin{pmatrix} x_{11} & 0 & 0 \\ 0 & x_{22} & 0 \\ 0 & 0 & x_{33} \end{pmatrix}.$$

Similarly, a number of trade terms x_{ij} ($i \neq j$, i,j=1:3), express the inter-region trading

$$Exports = \begin{pmatrix} 0 & x_{12} & x_{13} \\ x_{21} & 0 & x_{23} \\ x_{31} & x_{32} & 0 \end{pmatrix},$$

where the *j*th entry in the *i*th row is the export from Region *i* to Region *j*. Similarly, there is a corresponding trade matrix for imports (which if all data matched up, would be the transpose of the exports matrix). Ignoring re-exports, it follows that all exports of the commodity from a given region come from that region's production, and so it follows that the final distribution of production (*DOP*) of the commodity is described by

$$DOP = \begin{pmatrix} x_{11} - (x_{12} + x_{12}) & x_{13} & x_{13} \\ x_{21} & x_{22} - (x_{21} + x_{23}) & x_{23} \\ x_{31} & x_{32} & x_{33} - (x_{31} + x_{32}) \end{pmatrix}.$$

The *i*th row sums to the total production quantity of the commodity by Region *i*, and the different values denote its final distribution among all three regions. The sum of the *j*th column equals the total commodity possessed by Region *j*, and the different values their origin.

Any of the production or trade variables can be 0 without loss of generality, however if

$$x_{ii} < \sum_{j \neq i} x_{ij}$$

for any *i*=1:3, then a negative value will appear in the *i*th diagonal entry (i.e. exports are greater than production).

To account for re-exports, it is necessary to consider what has come into the region before deciding what goes out. Different methods for addressing this bring with them different sets of assumptions and effects.

One approach is to break the annual trade process down into a number of smaller time steps, and to calculate the traded commodities in the next time step by considering the commodity distribution at the current time. Within this framework, there are a number of different assumptions and choices that can be imposed regarding initial conditions and trade.

For example, each region can initially be considered to be in possession of its total annual production of the commodity, such that the initial conditions are

$$DOP(0) = \begin{pmatrix} x_{11} & 0 & 0 \\ 0 & x_{22} & 0 \\ 0 & 0 & x_{33} \end{pmatrix}.$$

Initial trade can then only be of the domestically grown commodity (i.e. no re-exports). If the year is broken down into *n* segments of duration δt, after one time step the distribution will look like

$$DOP(\delta t) = \begin{pmatrix} x_{11} - \frac{1}{n}(x_{12} + x_{13}) & \frac{1}{n}x_{12} & \frac{1}{n}x_{13} \\ \frac{1}{n}x_{21} & x_{22} - \frac{1}{n}(x_{21} + x_{23}) & \frac{1}{n}x_{23} \\ \frac{1}{n}x_{31} & \frac{1}{n}x_{32} & x_{33} - \frac{1}{n}(x_{31} + x_{32}) \end{pmatrix}.$$

Now that initial trade has occurred, subsequent trade can include the re-exporting of the commodities along with exports. The composition of the subsequent trade then needs to be defined; one option is to have subsequent exports/re-exports in proportion to the quantity of the imported and domestic commodities possessed by each region.

For example, continuing the above example to time $2\delta t$, if we consider trade from Region 1 to Region 2, there will be the export of Region 1's domestically produced commodity as well as the re-export of Region 2 and Region 3's commodities previously imported into Region 1.

The first column shows the commodities possessed by Region 1. In the time δt to $2\delta t$, Region 1 needs to export a quantity of commodity to Region 2 equal to $\frac{1}{n}x_{12}$. To calculate the composition of this trade, it is simply required to normalise the total resource possessed by Region 1. It can be seen that

$$Region\ 1_{total} = x_{11} - \frac{1}{n}(x_{12} + x_{13}) + \frac{1}{n}x_{21} + \frac{1}{n}x_{31}$$

And so the export/re-export composition will be

$$\frac{\frac{1}{n}x_{12}}{x_{11} - \frac{1}{n}(x_{12} + x_{13}) + \frac{1}{n}x_{21} + \frac{1}{n}x_{31}} \begin{pmatrix} x_{11} - \frac{1}{n}(x_{12} + x_{13}) \\ \frac{1}{n}x_{21} \\ \frac{1}{n}x_{31} \end{pmatrix},$$

where the *i*th element of the resulting vector is the quantity of commodity produced in Region *i* exported/re-exported by Region 1 to Region 2.

In general, if at time *t* the *DOP* is

$$DOP(t) = \begin{pmatrix} x_{11}(t) & x_{12}(t) & x_{13}(t) \\ x_{21}(t) & x_{22}(t) & x_{23}(t) \\ x_{31}(t) & x_{32}(t) & x_{33}(t) \end{pmatrix},$$

trade from region *i* to region *j* in time *t* to time *t+* δt is given by

$$\frac{\frac{1}{n}x_{ij}}{x_{1i}(t) + x_{2i}(t) + x_{3i}(t)} \begin{pmatrix} x_{1i}(t) \\ x_{2i}(t) \\ x_{3i}(t) \end{pmatrix}.$$

Running this process iteratively for all n time steps (i.e. to time $t = n\delta t = 1\,year$) yields a final distribution of production matrix. In the case of calculating re-exports for the FAO agriculture data, *n* is taken to be 10,000 which provides sufficient levels of accuracy whilst keeping computational time manageable (this requires about 24 hours for the full dataset, but only needs to be calculated once).

The assumption that each region begins the annual period with its full annual production quantity of the commodity biases trade towards export rather than re-export. An alternative approach is to break the production up over the year across the different time steps. In this way, the system ceases to be closed (that is all commodities are placed in the system initially and simply redistributed over time), but rather a constant input is being made into the system in the form of production (production in region *i* acts like an import from a region external to the system).

If this approach is adopted, then initially all regions possess zero commodity. In each time step, the same behaviour occurs as above, but first an input is made due to the production. The process of the first time step looks like as follows:

$$DOP(\delta t)_1 = \begin{pmatrix} \frac{1}{n}x_{11} & 0 & 0 \\ 0 & \frac{1}{n}x_{22} & 0 \\ 0 & 0 & \frac{1}{n}x_{33} \end{pmatrix},$$

$$DOP(\delta t)_2 = \begin{pmatrix} \frac{1}{n}x_{11} - \frac{1}{n}(x_{21} + x_{13}) & \frac{1}{n}x_{12} & \frac{1}{n}x_{13} \\ \frac{1}{n}x_{21} & \frac{1}{n}x_{22} - \frac{1}{n}(x_{21} + x_{23}) & \frac{1}{n}x_{23} \\ \frac{1}{n}x_{31} & \frac{1}{n}x_{32} & \frac{1}{n}x_{33} - \frac{1}{n}(x_{31} + x_{32}) \end{pmatrix}$$

$$= \frac{1}{n}\begin{pmatrix} x_{11} - (x_{12} + x_{13}) & x_{12} & x_{13} \\ x_{21} & x_{22} - (x_{21} + x_{23}) & x_{23} \\ x_{31} & x_{32} & x_{33} - (x_{31} + x_{32}) \end{pmatrix}.$$

To expand to the general case like above, if at time *t* the distribution of production is

$$DOP(t) = \begin{pmatrix} x_{11}(t) & x_{12}(t) & x_{13}(t) \\ x_{21}(t) & x_{22}(t) & x_{23}(t) \\ x_{31}(t) & x_{32}(t) & x_{33}(t) \end{pmatrix},$$

It follows that trade from Region *i* to Region *j* in time *t* to time *t*+ δt is given by

$$\frac{\frac{1}{n}x_{ij}}{x_{1i}(t) + x_{2i}(t) + x_{3i}(t) + \frac{1}{n}x_{ii}} \begin{pmatrix} x_{1i}(t) \\ x_{2i}(t) + \frac{1}{n}x_{ii} \\ x_{3i}(t) \end{pmatrix},$$

where the "$\frac{1}{n}x_{ii}$" term in the vector is added to the *i*th row.

As mentioned above, setting the system up with each region initially possessing their entire production of the commodity biases trade in favour of export over re-export. To demonstrate this, a simple two region system with one commodity can be used. If each region produces 1 unit of the commodity per year, and trades 0.5 units of the commodity per year, we have that

$$Production = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix},$$

and

$$Exports = \begin{pmatrix} 0 & 0.5 \\ 0.5 & 0 \end{pmatrix}.$$

If re-exporting is ignored, this leaves all commodities equally distributed between the two regions, with

$$DOP = \begin{pmatrix} 0.5 & 0.5 \\ 0.5 & 0.5 \end{pmatrix}.$$

If the first method for re-exporting is adopted with each region initially being allocated its entire annual production, the final production distribution matrix ends up as

$$DOP = \begin{pmatrix} 0.568 & 0.432 \\ 0.432 & 0.568 \end{pmatrix}.$$

With the second method, where production is introduced gradually with each time step, the final trade matrix generated is

$$DOP = \begin{pmatrix} 0.667 & 0.333 \\ 0.333 & 0.667 \end{pmatrix}.$$

Comparing results from the two re-export methods, it can be seen that the first method results in more of the domestically produced commodity being exported than the second method. The distribution of production from the second methods differs from when no re-exports are calculated by over 33%, and from the first method for re-exports by over 17%.

These methods provide simple and understandable means of generating re-exports from the available trade data. Implicit in the algorithm is the effect that the same item could be traded back and forth between two or more regions a large number of times (up to *n* times), which would economically make very little sense in reality and thus not occur. However, since the data does not explicitly deal with individual items/groups of items, but rather just a measure of mass or units produced and traded (i.e. there is no way to identify/track one tonne of wheat separately from another produced in the same region, for example), this real-world phenomenon does not manifest itself in the numerical results. It is the final answer provided for the distribution of production that is of interest, and not the intermediate steps.

Another effect of these algorithms is that they don't allow for the net balance of a commodity to be negative; the sums of the rows will always be equal to the production quantity of the respective regions. What this means in real terms is that if a country is reported as producing quantity *P,* importing quantity *I* and exporting *E* of a given commodity, if the exports exceed the sum of total production and imports (i.e. *E*>*P*+*I*), then the exports are implicitly scaled to *E*=*P*+*I*. Consequently, all domestically produced and imported quantity of the commodity will be exported, and no quantity will be left for domestic use. This can be changed to allow for negative net balance (i.e. allow a country to export a commodity it has not produced/imported in the current trade period), but as mentioned before, this requires further assumption about the origin of the commodity being traded.

Below is a real world example using the FAO data for Wheat production and trade. The data covers the top 5 producing regions by quantity (China, India, United States of America, Russian Federation and France), the UK, and all other regions aggregated into a "Rest of World" region.

The final distribution of production (*DOP*) of the commodity is calculated and shown for three methods of dealing with re-exports:

- (1) Re-exports are ignored and trade assumed to be as described by the export dataset.
- (2) Re-exports calculated assuming that all regions possess full production quantities at the beginning of the trade period.
- (3) Re-exports calculated assuming that all regions produce their commodities at a constant rate throughout the trade period (default method chosen for the project).

Production and export data for the top 5 Wheat producing countries, the UK and the Rest of the World from the FAO datasets are given as:

$$Production\ (\text{raw FAO data}) = \begin{array}{cc} \text{CHN} & 109298296 \\ \text{IND} & 75806700 \\ \text{USA} & 55820400 \\ \text{RUS} & 49368000 \\ \text{FRA} & 32763500 \\ \text{GBR} & 13221000 \\ \text{ROW} & 276333496 \end{array}$$

and

$Exports$ (raw FAO data) =

	CHN	IND	USA	RUS	FRA	GBR	ROW
CHN	0	0	3150	0	0	0	2333470
IND	12	0	40	0	0	2	195
USA	1312050	0	0	0	16461	107749	032822700
RUS	0	1242270	0	0	0	2566	013199300
FRA	50	0	112	142	0	167411	142188000
GBR	10	0	0	0	46485	0	1865020
ROW	63301	1078020	2386440	477123	495014	1261510	0

If Method 1 is applied and re-exports are ignored, all trade is comprised of exports from the producing region. The final distribution of production is given by

DOP=

	CHN	IND	USA	RUS	FRA	GBR	ROW
CHN	106962000	0	3150	0	0	0	23334700
IND	12	75806500	40	0	0	2	195
USA	1312050	0	21561400	0	16461	107749	32822700
RUS	0	1242270	0	34923900	0	2566	13199300
FRA	50	0	112	142	18377000	167411	142188000
GBR	10	0	0	0	46485	11309500	18650200
ROW	63301	1078020	2386440	477123	495014	1261510	270572000

Calculating re-exports according to Method 2 by initially assigning all produced commodities to the respective regions before implementing trade, the resulting distribution of production is calculated as:

DOP=

	CHN	IND	USA	RUS	FRA	GBR	ROW
CHN	106977000	3975.48	8069.92	1546.1	1473.77	4419.85	2302080
IND	12.2993	75806500	24.7579	0.13468	0.137229	2.26106	205.329
USA	1267740	55238.5	22564000	21473.1	32463.2	158827	31720600
RUS	1870.13	1258120	34763.7	35009900	8340.79	27363.8	13027700
FRA	2060.72	24179.80	37421.2	9519.98	18529800	180922	13979600
GBR	266.038	3082.68	4751.95	1197.85	33524.8	11421100	1757090
ROW	88411.5	975688	1302130	357533	329311	1056110	272224000

In contrast, when Method 3 is utilised and production is implemented as a progressive process with new production injected into the system as each time step, the resulting distribution of production is:

$DOP=$

	CHN	IND	USA	RUS	FRA	GBR	ROW
CHN	106992000	7416.23	7961.57	2305.81	1932.63	7518.49	2279600
IND	12.3257	75806500	16.5829	0.211581	0.18713	2.44134	209.176
USA	1242280	101366	23059200	31516	35563.4	192710	31157700
RUS	4421.38	1272410	37875.9	35075200	10957.1	44803.8	12922300
FRA	4785.87	45029.7	40611.2	14098.5	18629900	188910	13840200
GBR	579.803	5429.25	4891.46	1688.15	25075.4	11514500	1668830
ROW	93453.3	888629	800591	276284	231508	900275	273143000