About

resourcetrade.earth/about

Methodological overview

resourcetrade.earth has been developed by Chatham House to enable users to explore the fast-evolving dynamics of international trade in natural resources, the sustainability implications of such trade, and the related interdependencies that emerge between importing and exporting countries and regions.

The trade data on this site are from the Chatham House Resource Trade Database (CHRTD). The CHRTD is a repository of bilateral trade in natural resources between more than 200 countries and territories. The database includes the monetary values and masses of trade in over 1,350 different types of natural resources and resource products, including agricultural, fishery and forestry products, fossil fuels, metals and other minerals, and pearls and gemstones. It contains raw materials, intermediate products, and by-products.

Dealing with complexity

Bilateral statistics are critical to understanding global resource trade, but existing data are often difficult to access and use. The original data source for the CHRTD is International Merchandise Trade Statistics (IMTS). IMTS data are collected by national customs authorities and compiled into the United Nations Commodity Trade Statistics Database (UN Comtrade) by the United Nations Statistics Division. UN Comtrade utilizes three distinct trade classification systems: the Harmonized Commodity Description and Coding System (HS), the Standard International Trade Classification (SITC), and Broad Economic Categories (BEC). Of these, the CHRTD employs the HS taxonomy (1996 revision) which assigns HS codes to all forms of traded goods in a hierarchical structure (2, 4, and 6 digit codes respectively represent commodity chapters, headings, and subheadings).

Across the resource landscape there are alternative repositories of trade data available, but they do not offer the breadth and depth of analysis that UN Comtrade permits. For example:

- UNCTAD data have the greatest temporal availability, with some aggregate categories dating back to 1948, but even more recent series lack commodity-level (6-digit HS code) detail.

- UN FAO provides comprehensive agriculture, forestry, fisheries and aquaculture bilateral trade data, but not other resource domains. Much of the available data shares the same origins as UN Comtrade and is not necessarily any more accurate.

- USDA Global Agricultural Trade System (GATS) is similarly domain constrained and focuses on US trading partners. (Similarly the European Commission's EUROIND database covers only trade with and between EU countries.)

- <u>IEA</u> provides comprehensive energy balance and energy flow statistics, but trade statistics are limited to gas flows within, and to, Europe.

- <u>EIA data</u> include national energy consumption, production, import and export statistics, but lack bilateral detail.

- The world oil and gas databases developed by the Joint Organisations Data Initiative (<u>JODI-Oil</u> and <u>JODI-Gas</u>) were developed as transparency tools rather than trade databases. Unlike other databases, JODI does not adjust reported figures or substitute missing figures, so coverage is incomplete.

- Commercial sources such as the <u>BP Statistical Review of World Energy</u> provide comprehensive energy trade data, but no bilateral dimension.

UN Comtrade is therefore arguably the most comprehensive source of merchandise trade statistics available; volumetric and monetary value data are catalogued under more than 5,000 HS codes, and the monetary values of trades are available as far back as 1962. However, it does present several challenges for users focusing on resource trade, which the CHRTD and the resourcetrade.earth site address:

- The HS system is not easy to use: the HS nomenclature has evolved historically as a pragmatic and comprehensive industrial taxonomy for the broad range of internationally traded goods.

- The scale hinders simple queries: with over <u>3 billion</u> trade records since 1962, finding the right data is not always easy and the size of data queries can be difficult to manage.

- The IMTS data are of variable quality: missing data, trade mispricing, unreported and illegal trade, and general mistakes and inconsistencies all cast doubt over the reliability of certain reported trade flows.

- The presence of between one and four data points for every trade flow complicates use: both exporters and importers are expected to report trade values and trade masses. If these records are incomplete or do not correspond with one-another, there may be uncertainties about the reliability of data and/or about which reporter is the most authoritative.

Introducing clarity

The Chatham House Resource Trade Database reorganizes data around natural resources. As the IMTS and HS systems contain all types of traded goods - including manufactured goods - analysing natural resource trade flows in UN Comtrade typically requires amalgamating a variety of HS codes. The difficulty of this varies: products that have a long history of being traded extensively are captured in greater detail than products that are traded less frequently. For example, there is a single HS code associated with rare earth elements, but several hundred codes assigned to steel and steel products. The CHRTD overcomes this problem by selecting over 1,350 HS codes that are identifiable as raw materials or relatively undifferentiated intermediate products, and grouping them by resource type. For example, copper ores and concentrates, intermediate copper products

such as mattes, bars or wires and copper scrap are classified into a single 'copper' category, enabling global copper trade to be tracked at different stages of the value chain. The CHTRD employs a five-tier resource taxonomy permitting queries to be as atomised or aggregated as required.

The Chatham House Resource Trade Database employs a systematic approach to identify and manage data gaps and errors. The CHRTD is subject to the same data gaps and weaknesses as are apparent in other sources of international merchandise trade data. (For a full discussion of reporting asymmetries see Markhonko, 2014.) However, it exploits the maximum information available within UN Comtrade to assess the reliability of individual trade records, and to present as complete and as reliable a picture as possible. The approach taken relies on two assumptions. First, for each trade flow the values (US$) and masses (kg) reported by the exporter and importer should approximate to one-another. The reported monetary values are unlikely to be exactly the same since exports are typically reported Free On Board (FOB), whereas imports are typically reported on a Cost for Insurance and Freight (CIF) basis. Second, we expect the reported prices per tonne to relate to world market prices. Unlike some alternative approaches to reconciling importer and exporter reports, no assumptions are made about the general reliability of country reporting across multiple commodities or years; each individual report is assessed on its own merit.
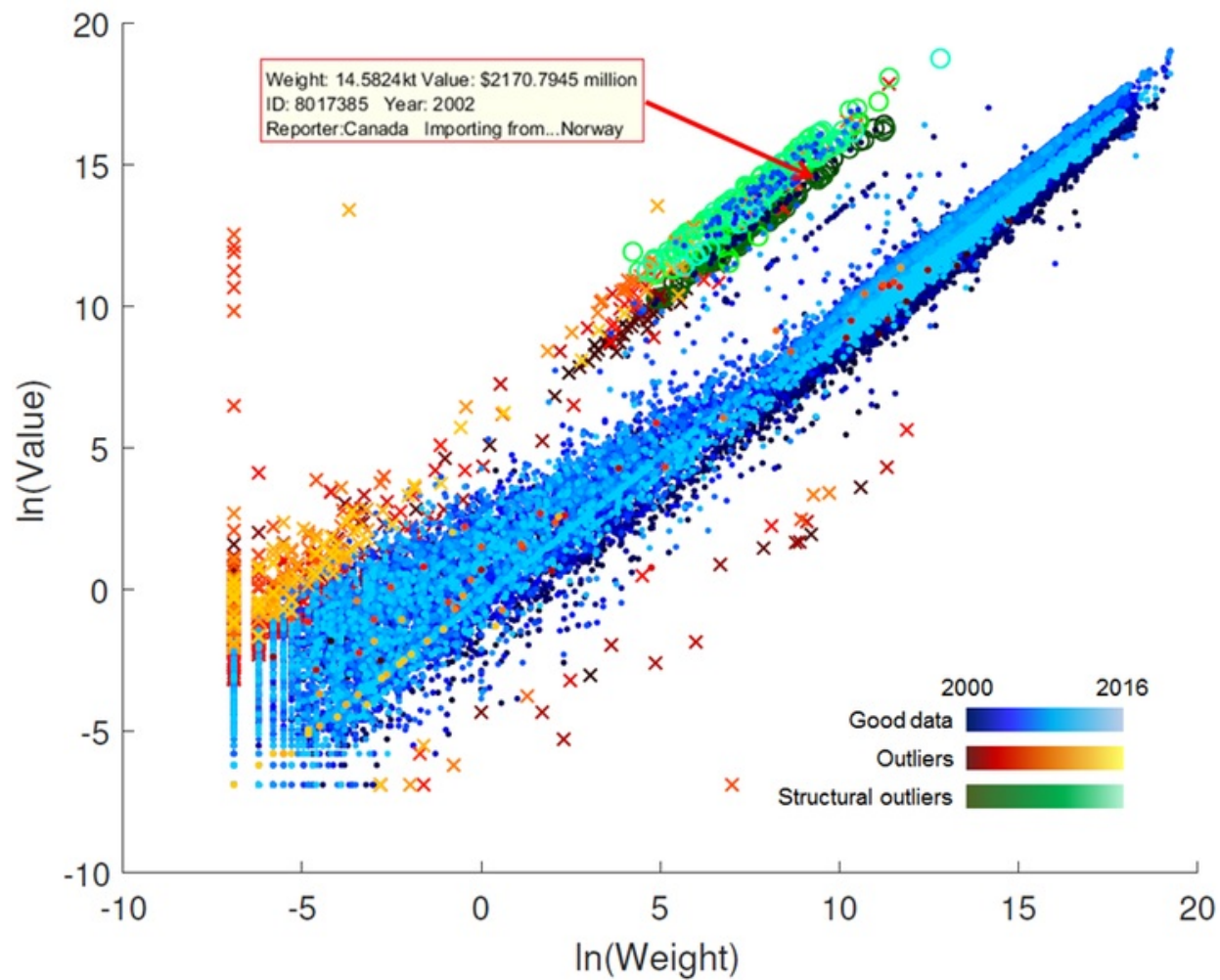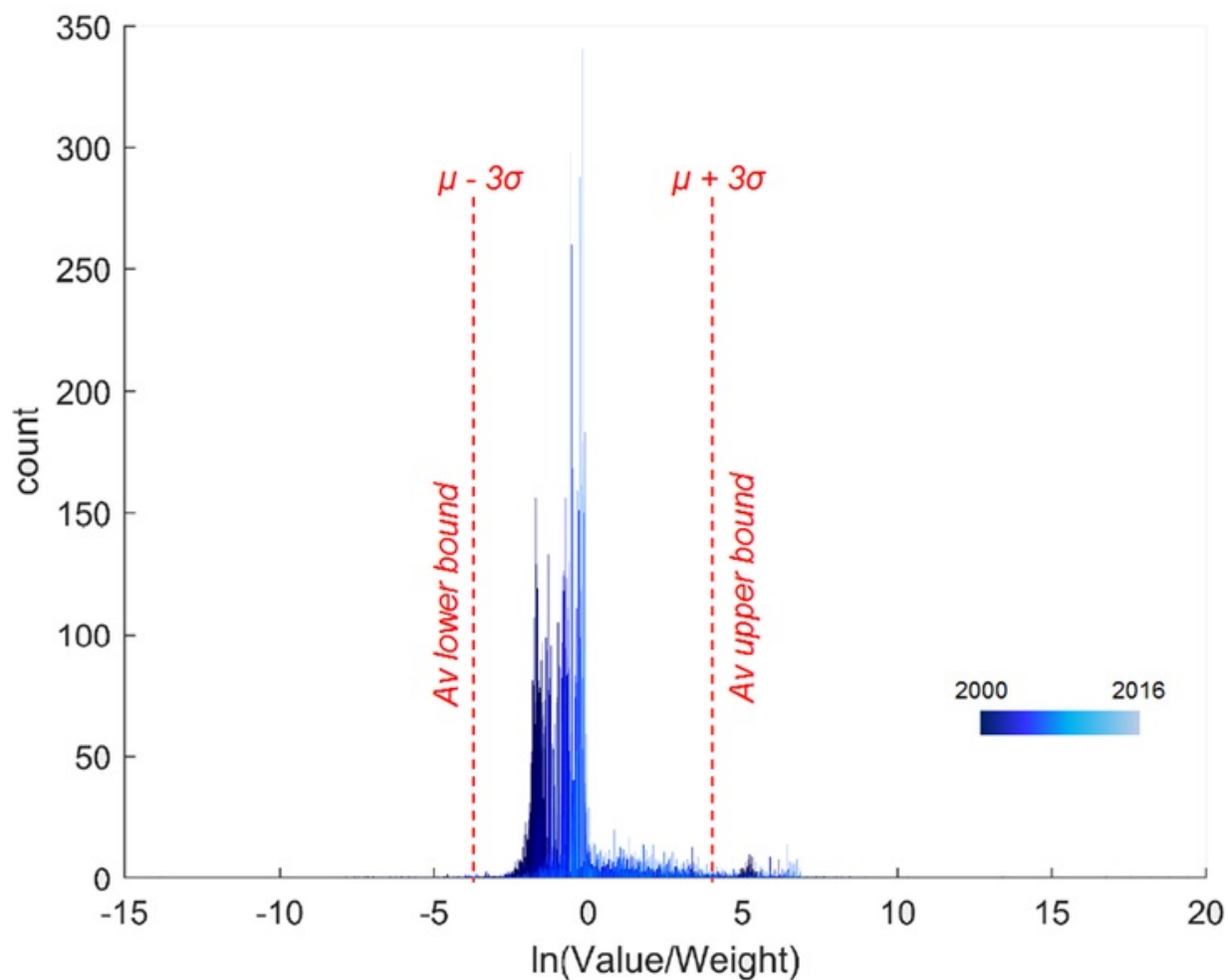
Logical operations are used to produce a transparent decision on the relative reliability of each data point and to reconcile the importer and exporter reports into a single record. Each record incorporates the value and mass of the given commodity trade between the two countries in the given year. In each case we consider the degree of similarity between the importer and exporter reports. In cases where either trade partner reports the monetary value and the mass of the trade (some reports contain only the value), we perform a distributional analysis of the value-to-mass ratio for all trades of the given commodity in the given year, i.e. the reported price per tonne is assessed relative to the global distribution of unit prices for the same commodity in the same year. The vast majority of these unit price commodity-year distributions are lognormal in nature, so we take the natural logarithm of the unit prices to derive normal distributions (ln($/kg)). Based on this transformation we identify outliers that are greater than three standard deviations from the mean value ($\mu \pm 3\sigma$). This results in approximately 0.3 per cent of the unit prices being flagged as outliers. This is illustrated in the below figure, which plots the 17 distributions for live sheep unit prices for 2000-2016.
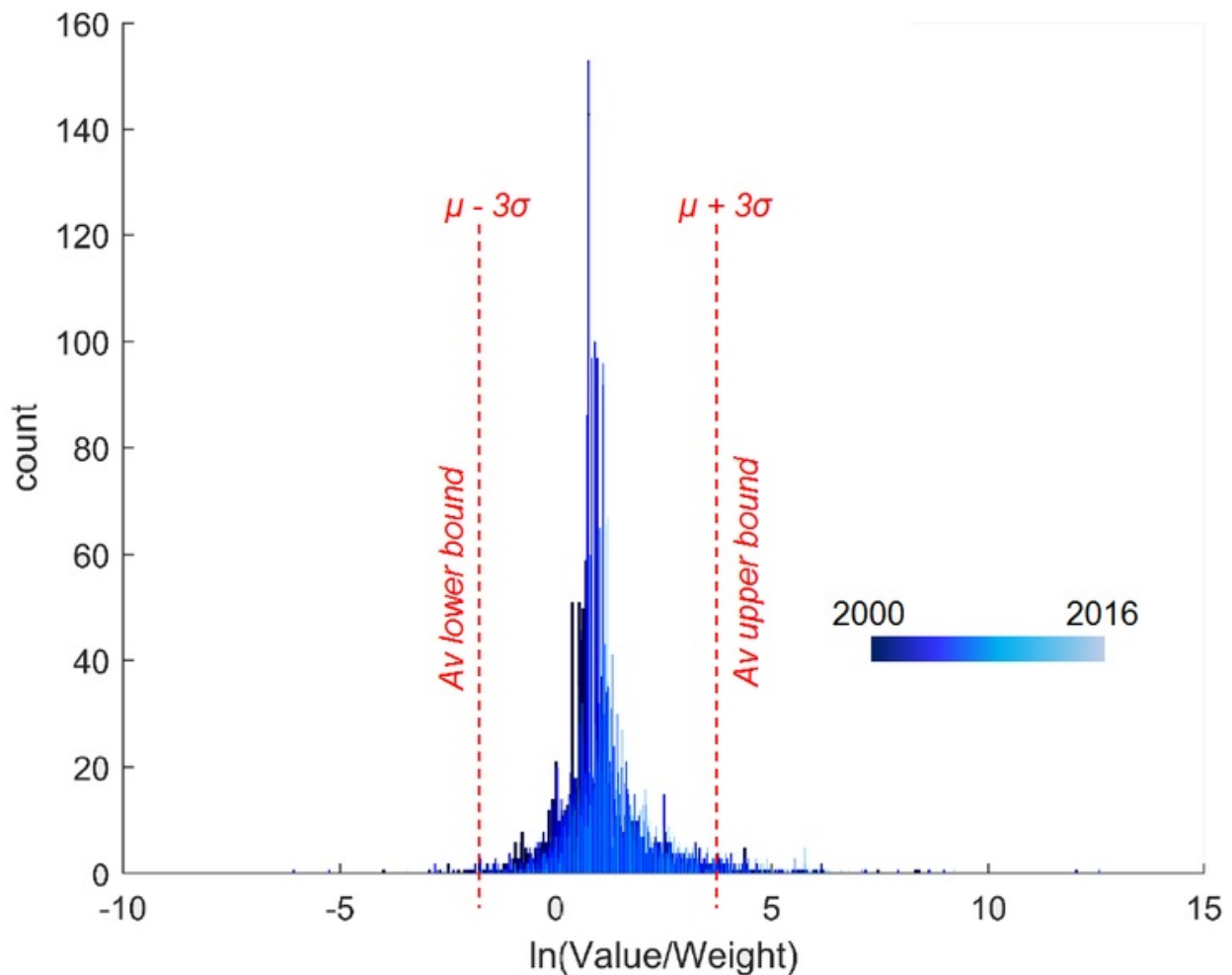
Not all commodity-year distributions have a lognormal structure however; some exhibit other structural properties that need to be accounted for. For example, where there are two distinct markets or where a commodity code amalgamates multiple distinct commodities, distributions may be bimodal. We therefore algorithmically identify whether secondary structures are present in the distribution and whether outlying data points require deflagging as outliers. The number of bins is defined by the Freedman-Diaconis rule. Secondary structures are investigated by defining windows equal to one-twentieth of the total number of bins, in one bin increments from the upper and lower bounds ($\mu \pm 3\sigma$). If any of these windows contain twice or more the expected number of data points (where the expectation is defined by the fitted normal distribution), the dataset within that window is defined as a secondary structure. However, if more than one of these structures is identified in either tail, we select that window which has the greatest difference between actual and expected number of data points as the secondary structure. Within a secondary structure, individual trade reports are only de-flagged as outliers if the natural logarithm of the reported value (ln$) is greater than the median natural logarithm value (med ln$) for the given commodity-year distribution, as at low values there is greater natural spread in the reported data.

The outcome of this process is illustrated in the below figure: 17 (2000-2016) independent commodity-year distributions are shown for crude oil, with earlier years represented by darker shading. In the scatter plot, data points identified as outliers are shaded red-yellow, and data points that are within the bounds are shaded blue. Green-shaded data points are those that are initially identified as outliers but which are then deflagged and treated as

reliable on the basis of the secondary distributional structure identification process. This process is repeated for every commodity-year distribution. A sample of distributions with secondary structures identified by the algorithm are manually checked by researchers to ensure there is a valid rationale for their identification (e.g. the Norway to Canada trade highlighted below will be compared with other trade reports in the secondary structure to understand whether the data points represent plausible market dynamics).

μ - 3σ   Av lower bound

μ + 3σ   Av upper bound

2000   2016

count

ln(Value/Weight)

Weight: 14.5824kt Value: $2170.7945 million
ID: 8017385 Year: 2002
Reporter:Canada Importing from...Norway

2000   2016
Good data
Outliers
Structural outliers

ln(Value)

ln(Weight)

Following the outlier identification process, mirror reports of the same trade by both trade partners are reconciled into one report. If both partners report a non-outlying unit price, then a weighted average of the two reports is recorded; the weighting factor is calculated according to the relative distances from world average unit prices. Data points that are deemed unreliable and irreconcilable are labelled as such and quarantined.

A full specification of this methodology will be made available at a later date.

Environmental impacts

Estimates of embodied carbon dioxide, land, and water, where available, are indicative of the environmental impacts of resource trade.

Carbon dioxide

Embodied carbon dioxide volumes are calculated by multiplying trade volumes by product-level carbon intensity factors. The emission factors employed are from Sato (2014). They are world-average, cradle-to-gate factors, defined in physical terms (kg CO_2/kg product).

Sato (2014) presents a detailed discussion, and sensitivity test, of the advantages and disadvantages of using world average factors relative to country-specific factors. The cradle-to-gate system boundary accounts for emissions generated throughout the production phase of a product's lifecycle, including the production of inputs, up until the factory gate, i.e. before the product is transported to the consumer. This contrasts with alternative system boundaries such as gate-to-gate, cradle-to-grave (including the use and disposal phases of the product) and cradle-to-cradle (including recycling).

The cradle-to-gate carbon intensity factors assume that all production inputs are sourced domestically, i.e. they consider only domestic supply chains and exogenously include trade in intermediate and final products. This system boundary results in issues of double-counting emissions at aggregated levels, since emissions associated with relatively unprocessed materials will also be recorded against products that have the original materials as inputs. For aggregate national emissions inventories, double-counting is more of an issue for countries with significant trade volumes relative to the size of their economy, and for countries engaged in significant processing, with large import contents in their exports. This cradle-to-gate approach, however, is well-suited for comparing trade-adjusted emission inventories at a more detailed product-level. As such, to avoid double-counting and over-representing emissions, we report embodied CO_2 emissions only at individual product level rather than for categories of aggregation.

The carbon intensity factors provided by Sato (2014) are recorded against Standard International Trade Classification (SITC) revision 3, 4 digit resolution product codes. These were converted to the HS codes used by the CHRTD using UNSD correspondence tables. Carbon intensity factors are available for 93 per cent of the HS product codes included within the CHRTD.

Land and water

Embodied land area and water volumes are calculated for a sub-set of agricultural products

by the Global Landscapes Initiative of the Institute on the Environment at the University of Minnesota, building on analysis they previously published (MacDonald *et al.* 2015). This is a refinement of the approach developed by Kastner *et al.* (2011), which uses caloric equivalence to relate processed goods to their root crop. As such it is only possible to employ this methodology for commodities with a caloric value. CHRTD trade data for 2000-15 replace FAO trade data as the input trade data, other input values, for example caloric equivalence factors, production volumes, area harvested, and water productivity are derived from the same sources specified in MacDonald *et al.* (2015). Currently this site displays only embodied blue water values; we also have values for embodied green water, which we hope to add to the site in a future revision. Please contact us for further details.

The analysis using the CHRTD is innovative as it produces estimates of embodied resource volumes on each bilateral product flow, as opposed to previous estimates that consider the resources embodied on the overarching transfer of root crops between country of origin and the target country of final consumption. This is also considered in the current analysis but as it is not compatible with displaying direct bilateral trade flows, it is not currently included on this site. Overviews of the key assumptions and decisions taken in developing this approach, and of the post-processing procedure used to derive values for bilateral product flows from the existing MacDonald *et al.* (2015) analytical approach are outlined below.

The embodied land and water volumes reported are those associated with producing the root crop from which the traded commodity is derived, not with further processing stages. Because calculations are based on the root crop, unlike the embodied carbon dioxide calculations, which relate to the traded product, it is possible to aggregate embodied land areas and water volumes across different products derived from the same root crop.

Key assumptions

1. Export commodities are sourced evenly throughout the exporting country

2. For calculation of re-exported crops there is no differentiation between imported and domestically-produced crop products

3. Trades that have more than one intermediary between source and target can be neglected. (This assumption is consistent with the MacDonald *et al.* (2015) and Kastner *et al.* (2011) methods.)

4. All commodities from a given root crop are fungible. For example, if the US exports soy cake to the UK and the UK only exports soybeans to France, the soy cake is considered part of the total soy volume that can be exported from the UK. In other words, there is no tracking of which commodities can be converted into which other commodities.

5. Commodities for which FAOSTAT only reports a single value for wet and dried production are excluded

Post-processing procedure

The post-processing procedure is a method for estimating the properties underlying the re-

export corrected trades. Each re-export corrected trade as reported using the procedure outlined in MacDonald *et al.* (2015) describes a trade of commodities based on a root crop produced in country A, then traded to, and consumed in, country Z. These commodities may have been traded directly from A to Z, or they may have been traded from A to {B,C,D …} and then from {B,C,D …} to country Z. Commodities which flow from A to B, B to Y, and then Y to Z are not considered.

The post-processing procedure works as follows for each root-crop trade in the re-export corrected matrix:

[For notation, denote the source country as "A" and denote the target country as "Z", denote the tonnes of root crop traded in the re-export corrected analysis as T_{REC}, let {B,C,D … Y} denote all possible intermediate trading countries, and limit the calculation to a single year]

- Suppose that T_{REC} =1,000 tonnes of root crop "Maize" traded.
- First, consider direct trades from A to Z. Identify the set of all trades {T_{AZ}} in the CHRTD that list a maize-based commodity traded from A to Z.
- Sum the total tonnes of root commodity in the set of trades {T_{AZ}} to get the total direct trade in tonnes (T_{direct}).
- Suppose this sum is T_{direct} = 1,500 tonnes (i.e. $T_{REC} < T_{direct}$), then this is a case where all trade from A to Z is direct trade. Write out to an intermediate file for each of the trades {T_{AZ}} with the traded tonnes decreased by a factor of (1,000/1,500). As such the direct trades written out to the intermediate table correspond to the re-export corrected trade from A to Z.
- Suppose that T_{direct} is 600 tonnes (i.e. $T_{REC} > T_{direct}$), then this is a case where indirect trade is necessary to explain the trade flows. Here 400 tonnes (i.e. T_{REC} - T_{direct}) are calculated as re-exported in intermediate trade flows from A to Z.
- To allocate these 400 tonnes among the trades, identify all possible trades from A to {B,C,D… Y} and then from {B,C,D,…Y} to Z. For each intermediate country {B,C,D, … Y} calculate the total possible trade in each maize-based commodity. (For example, suppose that 10 tonnes of maize are traded from A to B, and then 50 tonnes of maize are traded from B to Z. Consider only 10 tonnes as a possible intermediate trade flow in this case.)
- Calculate T_{int} as a sum of all of these total possible trades through intermediate countries {B … Y}. In almost all cases, this intermediate trade T_{int} will be greater than (T_{REC} - T_{direct}). Then write out to an intermediate file each of the trades {T_{AB}, T_{AC}, T_{AD} …} and {T_{BZ}, T_{CZ}…} and multiply the traded tonnes decreased by a factor (T_{REC} - T_{direct})/(T_{int}).
- Differentiate the first-export from the second-export as follows: the embodied land and water coefficients are associated with the first-export only.
- Usually this factor (T_{REC} - T_{direct})/(T_{int}) will be fairly small. However, as we go through the lines of the re-export corrected trade matrix, we will be rewriting these trades multiple times.
- As a final step, sum embodied land area and water volumes from like lines in the intermediate file. Here, "like lines" refers to lines which have a given commodity from country C to country D. (Note: there will be many "like lines" because there may be one direct trade from C to D, and there can be many trades where C to D is the first

of an intermediate trade flow from C to Z, and many trades where C to D is the second trade of an intermediate trade from A to D.)

National indicators

National environmental, socio-economic, governance, and resource-dependence indicators contextualize the significance of resource trade. Selecting a country name on the trade map gives you the option to display a country profile. These profiles provide a suite of indicators of the country's absolute and relative standing across a range of environmental, socio-economic, governance, and resource-dependence domains. For each country and for any given indicator, we present values in the indicator series' native denomination and the country's percentile rank, which is a measure of the percentage of countries - with data available for the given indicator - that have an equal or worse score where normative judgements can be applied (100 is best, 0 is worst). Taken together, this suite of indicators helps to contextualize the importance of natural resources to the nation's development.

Data sources

Trade data

Chatham House Resource Trade Database, UN Comtrade

Embodied environmental data

- Embodied carbon dioxide: calculated from Chatham House Resource Trade Database and carbon dioxide coefficients from Sato, M. (2014). Product level embodied carbon flows in bilateral trade. *Ecological Economics*, 105, 106–117. Article. Dataset.

- Embodied land: calculated from Chatham House Resource Trade Database and other sources by the Global Landscapes Initiative of the Institute on the Environment at the University of Minnesota following the approach adopted in MacDonald, G., Brauman, K., Sun, S., Carlson, K., Cassidy, E., Gerber, J. and West, P. (2015). Rethinking Agricultural Trade Relationships in an Era of Globalization. *BioScience*, 65(3), 275-289.

- Embodied green and blue water: calculated from the Chatham House Resource Trade Database and other sources by the Global Landscapes Initiative of the Institute on the Environment at the University of Minnesota following the approach adopted in MacDonald, G., Brauman, K., Sun, S., Carlson, K., Cassidy, E., Gerber, J. and West, P. (2015). Rethinking Agricultural Trade Relationships in an Era of Globalization. *BioScience*, 65(3), 275-289.

National indicators

- Environmental indicators

- Socio-economic indicators

- Governance indicators

- Resource dependence indicators

Country groupings

- Composition of geographic regions, Least Developed Countries, Landlocked Developing Countries, and Small Island Developing States are defined by the UN Statistics Division.
- Composition of other country groupings are obtained from the respective sources:

Terms of use

The Terms and Conditions of use for www.chathamhouse.org apply in full to resourcetrade.earth ('our site'), with the following exceptions and limitations:

- resourcetrade.earth/data and data accessed or downloaded thereon are licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International Licence

- Attributions must clearly provide a link back to resourcetrade.earth. A preferred citation is given below.

- Data from, and based on, UN Comtrade are used and re-disseminated under licence.

- Users are kindly requested to report any errors or deficiencies in our site by emailing

Citation

When citing or reproducing any data or material from resourcetrade.earth, please use the following reference:

Chatham House (2018), *'resourcetrade.earth'*, http://resourcetrade.earth/

Contact us

Acknowledgements

- The Chatham House Resource Trade Database builds on an earlier version developed by Jaakko Kooroshy.

- Trade data are calculated from source data provided under licence by UN Comtrade.

- Embodied land and water data are calculated in partnership with Global Landscapes Initiative of the Institute on the Environment at the University of Minnesota.

- Site by Applied Works

- Generously funded by MAVA and UK aid from the UK government