



Virtual water trade and water footprint of agricultural goods: the 1961-2016 CWASI database

Stefania Tamea¹, Marta Tuninetti¹, Irene Soligno¹, Francesco Laio¹

¹Department of Environment, Land and Infrastructure Engineering, Politecnico di Torino, Torino, 10129, Italy

5 *Correspondence to:* Stefania Tamea (stefania.tamea@polito.it)

Abstract. To support national and global assessments of water use in agriculture, we build a comprehensive and harmonized database of water footprint and virtual water trade (VWT) data for hundreds of agricultural goods. The water footprint, indicating the water needed for the production of a good, including rainwater and water withdrawals, is expressed as a volume per unit weight of commodity (or unit water footprint, uWF), which is here estimated at the country scale for every year in the period 1961-2016. The uWF is also differentiated, where possible, between production and supply, referring to local production only and to a mixed role of local production and import, respectively. The VWT data, representing the amount of water needed for the production of a good and virtually exchanged with the international trade, are provided for each commodity as bilateral trade matrices, between origin and destination countries, for every year in the period 1986-2016). The database, developed within the EU-funded CWASI project, improves upon earlier datasets because it takes into account the annual variability of the uWF of crops, it accounts for both produced and imported goods in the definition of the uWF and it traces goods across the international trade up to the origin of goods' production. The CWASI database is available on the Zenodo repository at doi.org/10.5281/zenodo.3987468 (Tamea et al., 2020) and welcomes contributions and improvements from the research community to enable analyses specifically accounting for the temporal evolution of the uWF.

20 **1 Introduction**

There has been a booming interest in the concept of Water Footprint (WF) since its introduction about 10 years ago (Hoekstra & Chapagain, 2007, 2008). The water footprint offers a common approach, language and method to a wide range of analyses and multidisciplinary studies, as well as for its capability to convey environmental messages to the public. Also the concept of Virtual Water, originally proposed by J. A. Allan (1998) and from which the WF originated, is growing in popularity among both the scientific community and the general public. Virtual water is the volume of water needed to produce a certain good that is virtually traded when the good is exchanged among countries. Such virtual flow defines the international virtual water trade (VWT) and represents a metric that is suitable to analyse



environmental and multidisciplinary aspects related to global trade of agricultural goods, to water management and to agricultural policy. Similarly, the WF identifies the freshwater needed for the production of goods along the full supply chain, eventually separating between water from precipitation (green) or from surface/ground-water bodies (blue). The WF provides a quantitative framework to analyse the volume of water embedded in agricultural goods and the efficiency of water use, when the metric is computed per unit weight of each good (hereafter referred to as the unit water footprint, or uWF). The term *unit water footprint* is here introduced to unify the current terminology which includes “water footprint”, used indifferently for volumes and for volumes per unit weight, “crop water footprint” which excludes livestock products, or “virtual water content” mainly used within the context of trade (see for e.g., Hoekstra et al., 2011; Konar et al., 2011; Dalin et al., 2012; Tuninetti et al., 2015).

Assessment of WF and VWT requires a relatively large amount of data combining production and trade data (in tonnes) and unit water footprint data (in cubic meters per tonne). As for uWF data, the first remarkable effort about data sharing has been made by the Water Footprint Network, which published a large dataset of uWF for several primary and processed agricultural goods having crop and animal origin (Mekonnen & Hoekstra, 2010a, 2010b). This database, named WaterStat, includes average values over the period 1996-2005 and has been the basis of the water footprint assessment as proposed, e.g., in Hoekstra et al. (2011). Other uWF datasets exist, which are based on spatially distributed models coupling the soil water balance with vegetation growth (see, e.g., Tuninetti et al., 2015, and references therein); such databases mostly refer to a single year or period or to long-term simulations. Other datasets, referring to blue water or to scarcity-weighted indicators, are also available within the literature related to Life Cycle Assessment (e.g., Pfister et al., 2011; Pfister et al., 2016). The temporal variability of uWF has been seldom considered. Few examples include water scarcity indexes (e.g., Pfister & Bayer, 2014), or annual time series of uWF in the EORA database, based on assumptions about the economic growth of different production sectors (Lenzen et al., 2013). Recently, Tuninetti et al. (2017) proposed a Fast-Track methodology to adapt crop-based uWF data from WaterStat through agricultural yield data.

Focusing on international trade, physical trade flow databases are maintained for example by the Food and Agriculture Organization (FAOSTAT) or directly by the United Nations (UN-COMTRADE). Early publications by the Water Footprint Network (e.g., Hoekstra & Chapagain, 2011) are based on the combination of such trade databases and WaterStat to produce WF assessments. Another relevant framework for trade data are Input-Output tables, tracing supply chains across sectors and countries, whose worldwide dimension is captured by global multi-regional input-output tables (MRIO) (see Tukker & Dietzenbacher, 2013, for a review). In such a framework, some MRIO databases offer specific water-related extensions, quantifying water volumes associated to international trade (e.g., Geschke &



60 Hadjikakou, 2017). Two relevant examples are the EORA database (Lenzen et al., 2013) and EXIOBASE (Stadler et al, 2108), both including a water assessment distinguishing between green and blue water and including the temporal variability, although product categories and geographical regions are more aggregated than in the present study. Supply chains and trade of specific products, with their impact on the local environment and the water resources are also the objectives of the TRASE project developed by the Stockholm Environment Institute and the Global Canopy Programme
65 (SEI, 2019). Such project focuses on a limited set of products, although accurately investigating their supply chain and environmental effects.

In an attempt to classify different methods for VWT and WF assessment, two approaches can be identified: the bottom-up and the top-down approach. The *bottom-up* approach refers to a process-based analysis, with a detailed description
70 of production processes and associated water volumes. Within such approach, the uWF of each good is multiplied by the (produced or traded) quantity of such good and resulting water volumes are then summed across goods. WaterStat is the main example of a bottom-up approach The *top-down* approach aims at tracing full supply chains throughout economic sectors and different countries. Input–output analyses, frequently used in Economics for environmental assessments, belong to this approach (Duarte & Yang, 2011). Bottom-up approaches do not consider the entire supply
75 chain of goods and can be affected by truncation errors when used to assess the water footprint of final consumption (Feng et al., 2011). At the same time, bottom-up techniques can offer high resolution enabling, for example, to differentiate the water associated to the production of a large variety of single (agricultural) products. A major problem affecting bottom-up approaches is the identification of the geographic origin of produced goods (Hubacek & Feng, 2016). In many cases, product re-export disconnects producing and consuming countries, now allowing a correct
80 identification of dependencies and externalities. In the present work, we improve the traditional bottom-up approach by identifying the origin of produced goods and reconstructing the supply chain of agricultural goods, implementing the method proposed in Kastner et al (2011). With such improvement, the VWT quantified in the present database aims both at best estimating the water embodied in bilateral trade and at providing accurate estimates of the total virtual water embedded in final consumption (Feng et al., 2011; Lenzen et al., 2013).

85

In this publication, we present an open database of virtual water trade, including the annual trade matrices (years 1986-2016) and the annual virtual water export (years 1961-2016) associated to a large number of agricultural products, as well as their unit water footprint in all countries (years 1961-2016), referring to the sum of green water (originated from rainfall) and blue water (originated from surface- and ground-water bodies). Starting from the uWF dataset in
90 Mekonnen & Hoekstra (2010a, 2010b), we extend it to consider the temporal variability of uWF. Improvements also include the differentiation between the production-side and supply-side of uWF. The new time-varying uWF are applied



to the FAO datasets of agricultural production, of country export, and to the reconstructed detailed trade matrices, generating the CWASI database.

95 The database addresses several needs: (i) the need for a comprehensive and harmonized database of uWF, WF and VWT, (ii) the need to adopt unit water footprints that vary in time, as recently pointed out by D’Odorico et al. (2019), (iii) the need to disentangle the production-side and the supply-side uWF to coherently assess the WF of production and consumption, (iv) the need for ready-to-use detailed trade matrices, accurately tracing goods’ trade and origin, suitable for network analyses. The database has been developed within the EU-funded CWASI project “Coping with water
100 scarcity in a globalized world” and it is shared through an online open-access repository (Tamea et al., 2020). In a relatively recent overview of the field, the research lines that originated from the concept of WF were identified (Hoekstra, 2017). These are (i) the role of trade and globalization in goods production and consumption and how they affect local water issues, (ii) the comparison of water requirements with water availability and renewability, (iii) the supply-chain approach applied to water management, and (iv) the combination of green, blue and grey indicators in a
105 single framework. With the CWASI database we aim at contributing to these research lines and provide all researchers with an up-to-date and ready-to-use starting point for their research. The database will welcome additions and external contributions that may possibly become available in the future and will represent an open and shared source of data on water footprint and virtual water trade.

2 Data and preliminary arrangements

110 From the database of the Food and Agricultural Organization of the United Nations (FAOSTAT, 2018), we collected 31 years (1986-2016) of trade data of agricultural goods, including crop-based and animal-based commodities. Data originate from national accountings and are available with the following information:

- reporting country (with FAO code),
- type of trade (import or export),
- 115 - partner country reported within the trade record (with FAO code),
- year,
- commodity (with FAO code), unit of measure, quantity.

From FAOSTAT (2018), we also collected 56 years of data on country export (1961–2016), regardless of the destination, and of agricultural production data, having this information:

- 120 - producing/exporting country (with FAO code),
- year,



- commodity (with FAO code), unit of measure, quantity.
From FAOSTAT (2018), data of agricultural yield and harvested area have also been collected per each considered crop, country and year in the period 1961-2016. The reference unit water footprint for every commodity and country, averaged over the period 1996-2005 are taken from WaterStat (Mekonnen & Hoekstra, 2010a, 2010b), as well as the product fraction and the value fraction needed for the computation of the uWF of processed crops.

2.1 Commodities

Production and trade data collected from FAOSTAT (2018) include crops, processed crops, livestock primary, livestock processed and live animals. The commodities currently included in the CWASI database are 357 and are those whose FAO code or item name or description could be associated to a WaterStat database entry (commodities are listed in Table SI.1) The number of commodities includes all products in the “Crop” production statistics of FAO, many processed crops with the exception of feed products (such as bran and cake); animals and animal-based products are included for most relevant species although not for others (such as camels, rodents, ecc.). Among all commodities, some are considered in both trade and production data, some appear only in trade and some other appear only in production. The primary cause in the difference of total commodities considered in production and trade is that production data are only available for primary goods and for few processed goods, while trade includes primary and a larger set of processed goods. For example, the flour of wheat or the bread are only available as trade data because production data only show the primary commodity (wheat). Conversely, yams or sugar cane are only available as production data because their trade is not recorded in the FAO statistics, possibly because they are not internationally exchanged as raw product. The number of commodities, differentiated between production and trade have been subdivided into 9 categories, and it is specified in Figure 1. The FAOSTAT database provides for each commodity and year the amounts produced (or traded) in any given country (or pair of countries) expressed in tons or heads, depending on the type of product.

2.2 Countries

The analysis includes all geographical/political/economical entities reporting (or reported for) at least one product and one year, either in the trade or the production data. From 1961 to 2016, agricultural goods were produced and traded among 255 entities having a temporary or permanent activity (the full list is reported in Table SI.2). Not all the 255 countries were active along the whole considered period, as they underwent political-administrative changes. Examples include is the collapse of the USSR, the separation of Eritrea from Ethiopia, or the splitting of Belgium and Luxembourg, which were considered a single entity until year 2000. Despite being inactive, a country may be reported by partners as importing or exporting goods. Values reported for a country outside its range of active years are



associated to the corresponding larger active country (e.g., a trade reported towards USSR in 1992 is associated to the Russian Federation). China has been considered as the aggregation of the following non-overlapping FAO entries: “China, Mainland”, “China, Hong Kong SAR”, “China, Macao SAR”, “China, Taiwan Province of”. There are also few entries of unclear allocation (Neutral Zone, Unspecified): they are listed but values are not considered. The set of associations is detailed in Table SI.3 The period of activity of each country is listed in the Supplementary Information, together with the associations adopted during country inactivity.

2.3 Trade matrices

The detailed trade data provided by FAOSTAT include the international trade records reported by each country. Reporting countries across the years are 184, whereas the remaining ones up to 255 are only reported by others. There is a total of 9 million records (i.e., trade flows per country pairs, per item and per year, for the commodities included in the CWASI dataset) and the number of records reported by each country is detailed in Figure 2. These records are used to reconstruct the trade matrix M for each item and year, having dimensions 255 x 255 and carrying the exporting countries on the rows and the importing countries on the columns. The matrix element $M(i, j)$ thus identifies the trade flow from country i to country j , which is different than the flow from country j to country i , i.e. $M(i, j) \neq M(j, i)$. Sub-national trade is not considered in these matrices and the terms on the diagonals are zeros.

A major problem arising in the construction of the trade matrices is that the same trade flow can be reported twice in the FAOSTAT database, once by the exporting country and once by the importing country. When a trade flow is reported by only one of the two countries, the reported flow is considered in the matrix (single record); this is the case for 40% records in the database. All other records are “double” and require a comparison between the declarations of the exporting and the importing countries, which are generally quite different and show a mean (absolute) relative difference, across all items, countries and years, of 61%. The choice of a value from two double records is called “reconciliation” and it is required for 42% of the non-null flows in the final trade matrices. The adopted method is based on the identification of the most reliable reporting country, among the two involved in each flow, and the use of the flow it reports. The reliability of countries is measured per item and per year with a data-based approach detailed in the following and adapted from Gehlhar (1996).

For each product, p , and year, t , two trade matrices are built, one matrix collecting all “Importer-Reported” flows and the other matrix collecting the “Exporter-Reported” flows. The matrices have the same structure and dimensions, with the exporter countries on the rows and the importing countries on the columns. Then a reliability index is calculated for each country, c , differentiating between import and export.

First, an accuracy measure (A) is defined for every flux, from country i to country j , as



$$A(i, j) = \frac{|IR(i, j) - ER(i, j)|}{\max\{IR(i, j), ER(i, j)\}}, \quad (1)$$

with $IR(i, j)$ being the importer-reported trade flux and $ER(i, j)$ being the exporter-reported flux. The measure is modified from Gehlhar (1996) to maintain the conceptual symmetry between import and export. The smaller is the measure and the more similar is the information reported by the importing and exporting country.

Then, the reliability of each country is measured, separately for import and export, based on the comparison between the flows reported by the country and by its trade partners. For every country, c , the reliability index for imports, $RI_{imp}(c)$, and for exports, $RI_{exp}(c)$, are defined as follows:

$$RI_{imp}(c) = \frac{\sum_j^{acc} IR(j, c)}{\sum_j^{all} IR(j, c) - IR(w, c)}, \quad RI_{exp}(c) = \frac{\sum_i^{acc} ER(c, i)}{\sum_i^{all} ER(c, i) - ER(c, w)}. \quad (2)$$

where $IR(j, c)$ is the flux from country j to c , as reported by c (importer-reported), and $ER(c, i)$ is the flux from c to any country i , as reported by c (exporter-reported) respectively. Σ^{all} is the sum of all import or export fluxes reported by c and Σ^{acc} is the sum of acceptable fluxes only, defined as the fluxes whose accuracy A is smaller than an acceptance threshold, set to 20% as in Gehlhar (1996). $IR(w, c)$ and $ER(c, w)$ are, respectively, the import from, and the export to, the worse partner w , defined as the ones having the maximum (worse) flow-weighted accuracy measure (WA) defined, for import and export fluxes, as

$$WA_{imp}^c(j) = A(j, c) \cdot \frac{IR(j, c)}{\sum_j^{all} IR(j, c)}, \quad WA_{exp}^c(i) = A(c, i) \cdot \frac{ER(c, i)}{\sum_i^{all} ER(c, i)}. \quad (3)$$

$A(j, c)$ is the accuracy level of flux $IR(j, c)$ and $A(c, i)$ is the accuracy level of flux $ER(c, i)$; the denominators are, respectively, the sum of all imports and all exports reported by country c .

Reliability indexes are calculated by country, commodity, year and flow direction (import and export). This because the reliability of a country in reporting import and export may be different, the attitude of a country to over-report or under-report may differ by products, e.g. depending on taxation, and the reliability of a country may change in time, e.g. according to socio-political factors. The direction- and commodity- averaged RI of reporting countries are shown in Figure 2 with the darker (lighter) line corresponding to the newest (oldest) values. Countries more involved in trade and reporting more information (to the left) are characterized, on average, by a larger reliability, while countries less involved in trade have lower average reliability, which used to be very low in the past. Current RI values, instead, are more uniform across countries.

Having computed all reliability indexes, the “reconciled” trade matrix for each item and year is built, combining importer-reported and exporter-reported data. In fact, each matrix element $M(i, j)$ is taken from the IR or ER matrix if the importing country j or the exporting i has a greater reliability index. Where the reliability indexes are equal, the country having larger acceptable fluxes is chosen.



3 Unit water footprint

The unit water footprint measures the amount of water required to produce a unit amount of product and it can be expressed as m^3/ton or, equivalently, as l/kg . Depending on the type of commodity, different approaches are applied for the computation of the unit water footprint. In the present work we propose a differentiation between the uWF of production (uWF_p) and the uWF of supply (uWF_s) of primary crops. The uWF_p refers to locally-produced crops whose water footprint depends on the crop actual evapotranspiration and crop yield, with values which are here estimated annually starting from 1961. The uWF_s refers to the domestic supply of primary crops, which relies both on local production and on international trade. Domestic supply is used as human consumption, food manufacturing, feed for livestock and as export towards other countries. The impossibility to track local production and imports into consumption and exports, within each country, makes the uWF_s the best indicator to be used in conjunction with consumption and export data. The uWF_s of primary crops is computed averaging local production and imports, after identifying the countries of origin of the goods thanks to an appropriate procedure applicable from 1986.

Processed crops are produced from a root product which may or may not originate from local production. The absence of systematic FAO data about the production of processed crops prevents the differentiation between the unit water footprint of production and of supply. Therefore, processed crops considered in this study will have a single unit water footprint, depending on country and year, computed applying some conversion factors to the uWF_s of the root product. Finally, animal-based products are here considered only with the WaterStat values, without temporal variability.

3.1 Unit water footprint of locally-produced primary crops in time

When considering the production of primary crops, the unit water footprint of production, uWF_p, is a function of the actual evapotranspiration along the growing period of the crop and the crop actual yield. The present work considers the sum of green water (originated from rainfall) and blue water (originated from surface- and ground-water bodies). Due to precipitation, evapotranspiration, and yield fluctuations, the uWF_p exhibits significant spatio-temporal variability.

We computed the uWF_p in a given year by means of the Fast-Track (FT) method, introduced by Tuninetti et al. (2017). This method is based on the use of the WaterStat dataset (Mekonnen & Hoekstra, 2010a, 2010b) and assumes that the temporal variability of the unit water footprint, not detailed in WaterStat, is mainly expressed by a ratio of agricultural yields. Yield implicitly expresses many factors, including environmental and climatic conditions, harvested areas and agricultural practices and its temporal variations dominate over the variability of the water volumes used (evapotranspired) by crops. According to the Fast-Track method, the unit water footprint of an agricultural product p produced in country c in year t , i.e. $uWF_{p,c,p,t}$, reads:



$$uWFp_{c,p,t} = \overline{uWFp_{c,p,T}} \cdot \frac{Y_{c,p,t}}{\overline{Y_{c,p,T}}}, \quad (4)$$

where $\overline{uWFp_{c,p,T}}$ is the reference unit water footprint provided by WaterStat (Mekonnen & Hoekstra, 2010a, 2010b) corresponding to an average in the period $T=1996-2005$, $\overline{Y_{c,p,T}}$ is the average crop yield over the same period T , and
 245 $Y_{c,p,t}$ is the annual crop yield in a generic year t in the range 1961-2016. The average crop yield is obtained as an average of the annual yields in the years 1996-2005, weighted by the harvested areas across the years in country c . Equation (4) is best referred to the total (green plus blue) unit water footprint, but it may also be applied separately to the two components of green and blue water.

250 The Fast-Track method keeps implicitly constant the actual evapotranspiration of crops, equal to the average in the period T . This has been shown to introduce an uncertainty around $\pm 10\%$ of the uWF estimates for wheat, maize, rice, soybeans, which is lower than the model uncertainty in the water footprint assessment (see Tuninetti et al., 2017). The Fast-Track method, initially applied to 4 crops, has been here extended to a large set of primary products, including cereals, fruits, vegetables, seeds, luxury food and non-edibles. The extension is justified by the fact that water stress
 255 affects the evapotranspiration of different crops in a similar way, the only difference being the phases of the growing periods affected by water stress and the crop coefficients describing the water requirements along the growing periods. Water stress is assumed not to affect irrigated crops, implying that actual evapotranspiration matches maximum potential evapotranspiration. Uncertainty associated to the Fast-Track method has been sparsely checked on other crops and the range of errors found in Tuninetti et al. (2017) has been confirmed.

260 3.2 Primary-equivalent trade matrix

For the correct identification of countries of origin of the crops traded internationally, the reconstruction of a primary-equivalent trade matrix, \mathbf{M}_{eq} , is necessary (Kastner et al., 2011). This is defined as

$$\mathbf{M}_{eq} = \mathbf{M}_p + \sum_{dp} \left(\mathbf{M}_{dp} \cdot \frac{f_v}{f_p} \right), \quad (5)$$

where \mathbf{M}_p is the trade matrix of any root-product, \mathbf{M}_{dp} is the trade matrix of the derived products (dp) and f_p, f_v are the
 265 product fraction and value fraction which convert the derived products into a root-product equivalent quantity. The summation is extended to all derived products which originate from the same root product and, in the case of a multi-step supply chain, Eq. (5) is applied iteratively until reaching a root product that is also a primary crop. The product fraction, f_p , is defined as the weight of a derived product obtained from a ton of input product. For example, a ton of nuts with shells leads to $f_p (<1)$ tons of shelled nuts. The value fraction, f_v , is the market value of the derived product



270 divided by the aggregated market value of all derived products resulting from a ton of input product. For example, in a
production process of wheat flour there are other economically valuable by-products (e.g., wheat germs to feed
animals); hence, the value of wheat flour constitutes only a portion (i.e., the value fraction) of the total value generated
by the process. Product fractions and value fractions used in the CWASI database are time- and space- invariant and are
taken from Mekonnen & Hoekstra (2010a, 2010b), as well as the root products and the full supply chains of the
275 considered commodities.

3.3 Supply-side unit water footprint of primary crops

The country supply of a primary crop results from the sum of local production and imports, where imports may occur
from producing or non-producing countries, the latter case testifying a re-export of goods produced elsewhere.
Therefore, the unit water footprint of supply, uWFs, is proportionally contributed by local production and by trade,
280 specifying the relative contribution of every country from which the goods originated from, considering re-exports and
the goods processing, if necessary. For each primary-equivalent crop and each year, we can define a column vector, S ,
containing the supply of all countries as rows. This vector is calculated as the sum of the production vector, P , and of
the imports obtained from the bilateral trade matrix M_{eq} , where $M_{eq}(i, j)$ identifies the trade flow from i to j as

$$S = P + M'_{eq} \cdot I, \quad (6)$$

285 where I is a column vector of ones (i.e., a summation vector) and M'_{eq} is the trade matrix transposed. Hence, the uWFs
of a country depends both on the domestic uWFp (through P) and on the uWFp of the origin countries, where the
product is produced.

In order to trace the actual origin of the country's supply, namely tracing its origin back to the country where it was
produced, we adopt the approach proposed by Kastner et al. (2011). First, we define a matrix R , where each element
290 $R(i, j)$ is the quantity of supply in country i that is produced in country j . A first approximation of R can be based on
reported flows only, and being equal to the sum of a diagonal matrix with elements of the P vector on the diagonal, i.e.
 $diag(P)$, and the transposed trade matrix, M'_{eq} .

However, this approximation misses the fact that exporting countries may obtain the exported products not only from
local production, but also from import. To account for this fact, a matrix of export shares, A , can be defined as

$$295 \quad A = M'_{eq} \cdot diag(S^{-1}), \quad (7)$$

where $A(i, j)$ is the share of country j 's supply that is exported to country i . The term $diag(S^{-1})$ denotes a diagonal
matrix made up by the reciprocal elements of S . In turn, the imported and re-exported products may partly originate



from local production and import, and so on, recursively. It has been shown by Miller and Blair (2009), that such procedure converges to

$$300 \quad \mathbf{R} = (\mathbf{I} - \mathbf{A})^{-1} \cdot \text{diag}(P), \quad (8)$$

where \mathbf{I} is the identity matrix. For further details and exemplification, see Kastner et al. (2011).

The \mathbf{R} matrix identifies where the supply of each country originates from. Therefore, knowing the uWfP of the primary crop in such countries, we can now define the unit water footprint of supply in country c and year t of the primary product p , i.e. $uWFS_{c,p,t}$, as

$$305 \quad uWFS_{c,p,t} = \sum_{j=1}^{255} uWFP_{j,p,t} \cdot \frac{R_{p,t}(j,c)}{S_{c,p,t}}. \quad (9)$$

The evaluation of uWFs corresponds to a weighted average of the uWfP values, where the weights are the actual fractions of supply, S , traced back to their origins. Eq. (9) is valid for every primary crop p and year t , considering that trade matrices, production vectors and uWfP values change from year to year. It is worth noticing that because the trade matrices are available from 1986 only, the uWFs can be built from that year only.

310 **3.4 Unit water footprint of processed crops in time**

Processed crops are based on the processing of root products, which are available as country's supply. The time-varying unit water footprint thus depends on that of the root product and on the conversion factors, i.e. (Hoekstra, et al., 2011),

$$uWF_{c,dp,t} = uWFS_{c,p,t} \cdot \frac{f_v}{f_p}, \quad (10)$$

315 where $uWF_{c,dp,t}$ is the unit water footprint of the processed crop (or derived product, dp), $uWFS_{c,p,t}$ is the unit water footprint of supply of the root product, p , from which it derives, c and t are the country and year considered, respectively, f_p is the product fraction and f_v the value fraction of the processed crop. The method takes into account the temporal variability associated to both the crop production, through the Fast-Track method applied to the primary crop, and the evolution of trade, through the Kastner's method applied to the crop supply; the method does not include

320 water inputs for processing of goods. When supply chains are formed by multiple steps, for example in the case of bread, made with flour, made in turn with wheat, Eq. (10) is applied routinely at each step. Within the CWASI dataset, the longest supply chain is made of 4 steps, leading to final products such as refined sugar or chocolate.

Equation (10) describes the unit water footprint without differentiating between production and supply. This is because the absence of FAO data of production of most processed crops hinders the application of the Kastner's method (Sect. 325 3.3), thus an explicit accounting of countries of origin of trade, as in the case of primary crops. However, the trade of



processed crops is implicitly taken into account in the procedure, thanks to the use of the primary-equivalent trade matrix (Eq. (5)) which serves to compute the uWFs of the primary crop.

For the very few derived products without indication of the root product (Mekonnen & Hoekstra, 2010a), an association is made which is based on logical considerations (such as “Figs dried” deriving from “Figs”) or on similitudes of products. The “Sugar” products (Raw Sugar, Refined Sugar,...) were also missing the root-product, in that it can be either “Sugar Beet” or “Sugar Cane”. For these products, we have traced back the root product to the main product (beet or cane) available as country supply.

3.5 Unit water footprint of animal-based commodities

Animal-based commodities are available from FAOSTAT as grouped in three categories: “Live animals”, “Livestock primary”, and “Livestock processed”. Products of the first category are given in heads unit, which have been converted in tons according to FAO conversion factors (FAO, 2013). The missing conversion values for some countries (or animal products) have been assigned with an average value by category or considering similar animals. Due to the lack of reliable data about country-specific animal diets and its temporal variability as well as the lack of detailed trade matrices of feed crops, we do not currently provide a time-dependent unit water footprint for the animal-based commodities.

Nevertheless, we include these products in the present database adopting the country-specific values provided by WaterStat (Mekonnen & Hoekstra, 2010b) and without differentiating between production and supply (i.e., $uWFp = uWFs$). These values take into account the feed-animal-commodity global supply chain, considering locally-produced and imported feed (Mekonnen & Hoekstra, 2012) but are only available as time-averaged values over the period 1996-2005. Here data are generically referred to year 2000 and are arranged consistently with the rest of the CWASI database.

4 Virtual water trade and water footprint indicators

4.1 Water footprint and VWT data

The water footprint of agricultural production in a country and year is obtained by multiplying the production data, expressed in metric tons, by the corresponding (commodity, country, year) unit water footprint, considering the unit water footprint of production, $uWFp_{c,p,t}$, in the case of primary crops. A problem arises when a country was not a producer in the 1996-2005 decade, thus it does not have an associated value in the WaterStat database. In such case, the uWF in the closest producing country within a certain distance (10°) is taken; if no producing countries are found (e.g., in the case of remote islands, adverse climatic conditions, or small producing areas), then the global average weighted



355 by production, is used for that country. In the case of countries having experienced political discontinuity, for example
belonging to a larger country before years 1996-2005 (e.g., USSR), the reference value of uWF required in Eq. (4) is
computed as a production-weighted average of the values of countries belonging to the union and available in
WaterStat. This average value is then used to reconstruct the annual uWF from 1961 up to the year of the
disaggregation. After converting the agricultural production into water volumes, the overall water footprint of
production is obtained summing across all commodities. Care must be used to avoid double-accounting of water
360 footprints of primary and derived goods. For this reason, only primary products must be considered in aggregated
production data. In particular, when dealing with animal-based commodities, one should avoid the inclusion of
livestock and their products as well as the inclusion of crops used to feed the livestock. Primary, or single-accounting,
products to be included in the sum are indicated in the Supplementary Material, Table SI.1.

365 Computation of the supply-side unit water footprint of goods enables the fast computation of the water footprint
associated to the consumption of commodities, under the hypothesis that consumption (and export) shares with the
country's supply the same mix of local and imported goods. The water footprint of consumption in a country and year
can then be obtained multiplying the consumed quantity of each good by the unit water footprint of supply, $uWF_{s_{c,p,t}}$
(per commodity, country and year). Finally, the virtual water trade is obtained by multiplying trade data, expressed in
370 metric tons, by the unit water footprint of supply, $uWF_{s_{c,p,t}}$ of the exporting country. Thanks to the new definition of
supply-side unit water footprint, this computation allows one to take into account the origin of goods, which are traced
back to their origin places across the supply chain, without being affected by re-export. In the few cases of goods
exported from countries not having an associated uWF (less than 1% of trade links over the whole period, mainly from
minor countries or remote islands), the global average uWF of supply is used, weighted by all countries' exports.

375 **4.2 The uWF index**

In the presentation of the results (Section 5), the (volumetric) water footprint and the virtual water trade are summed
across different commodities and the overall trends are assessed in time. However, the unit water footprint of different
commodities cannot be assessed as a whole, but only for one commodity at a time. To overcome such problem, an
appropriate index is constructed in analogy to some economic indices aggregating prices of different commodities,
380 calculated with the Laspeyres approach, such as the Agriculture Producer-Price Index (in FAOSTAT). The index is
built as the inverse ratio between the water footprint of production (in m^3) of all items (i) in all countries (c) in year
2000 and the WF obtained with the same quantities (year 2000) but with uWF in year t , i.e.



$$I(t) = \frac{\sum_{i,c} uWF(i,c,t) \cdot P(i,c,2000)}{\sum_{i,c} uWF(i,c,2000) \cdot P(i,c,2000)} \cdot 100. \quad (11)$$

385 In such way, $I(t)$ expresses the variation of uWF across all agricultural commodities, weighted by the productions in year 2000, $P(i, c, 2000)$. A similar index as in Eq. (11) can also be built for trade using, e.g., the exports of each country in year 2000 as weights. In addition, indexes (for production or trade) referring to single categories of goods can be built by aggregating only the goods belonging to a given category.

5 Results

390 The importance of considering a time-dependent unit water footprint is highlighted in Figure 3, which shows the temporal trends of the global average uWF of production of some commodities. The global average is computed by weighting the uWFp of each country by the country production of such crop. The relevance of the temporal change is evident, ranging from 4000 to 1500 m³/ton over the whole considered period. It is worth noticing that the uWF of production of other major crops are shown in Tuninetti et al. (2017). The values considered in WaterStat refer to the period $T=1996-2005$, highlighted by a grey shade in Figure 3. It is clear that the average value in such reference period is scarcely representative of the whole period considered in the present dataset. It is thus very important to consider the temporal variability of unit water footprint, especially in analyses spanning long periods or periods different than years 1996-2005.

400 The temporal variation of the uWF of production of crops is marked all over the world. If compared to the values averaged over the period 1996-2005 (as in the WaterStat database), the uWFp computed with the Fast-Track method at the beginning and at the end of the considered period are very different. Figure 4 shows the relative change of the uWFp of wheat in 1961 and 2016 with respect to the 1996-2005 average. The variation is quite uniform worldwide with improvements (decreases of uWF) in both periods which are consistent with the different duration of the two periods.

405 Extreme variations have occurred in China (largest improvement from 1961) and in African countries, showing large improvements in time but also occasional worsening due to unstable socio-economic conditions. It should be noticed that few countries worldwide do not produce wheat or miss FAOSTAT or WaterStat data: in such cases, uWFp values from nearby countries or worldwide averages are used instead. Having a uWF, for any good, in all countries and years is a need dictated by the chance that a non-producing country is an exporter (or re-exporter) of such good and thus

410 requiring a uWF for the conversion of the trade flow into virtual water.



415 A comparison between the uWF of production and supply of primary crops is very informative. Figure 5 highlights the absolute difference for wheat and soybean with red colours indicating countries where the uWFs is smaller than uWFp and green colours for the contrary. The more intense is the red colour, the more efficient is the crop import in saving global water resources because the imported crops are produced with lower uWF that the local uWF of production. This is the case for several African countries, some South-American ones and Thailand for wheat and several South-Asian countries for soybeans. On the contrary, the more intense is the green colour, and the more efficient the global production is, as compared to imports. The extreme case of non-producing (but importing) countries is highlighted by bold contours. This is observed in several Far-East countries for wheat and by most African countries for soybeans.

420

Considering all commodities together, the analysis of temporal evolution requires the use of a uWF index (Eq. (11)). The index built weighting by agricultural production decreases monotonically in time (Figure 6, left), being at +50% in 1961 and -7% in 2016. The trend is less marked than in Figure 3 because all goods, and not only wheat, are being considered in the index. The uWF index weighted with exports decreases even more starting with a +85% in 1961. The difference may be interpreted in two ways, the first being that in the past times, for a given good, the production in exporting countries was more efficient than in the other (producing but not exporting) countries. The second interpretation focuses on a given country and highlights changes in the relative composition of export, with increasing shares of water-efficient goods at the expenses of water-inefficient ones. It is worth noticing that the uWF index is built with all commodities, including those not having a uWF varying in time (e.g., animal-based products, as made explicit in Figure 6, right): considering such contribution, the index temporal variation would be even more marked. The uWF indexes by category, shown in Figure 6 (right), allows one to find similarities and differences. In time, traded fruits and vegetables have improved their uWF more than global production, with large discrepancies in the past between trade-weighted and production-weighted indexes. Traded cereals, especially in the past, where produced with high efficiency, while in more recent years the average uWF of traded cereals became larger than that of the global cereal production. This fact may be due to the increased participation in the international trade of developing countries (Carr et al., 2013), countries producing cereals with lower efficiency than major exporters (eg., the USA). Seeds/oils and luxury food had intermediate periods where trade was on average more efficient than global production, while non-edible goods are currently traded with low efficiency, i.e. the trade-weighted uWF index is larger than the production-weighted one.

440 The time-varying uWF in the CWASI database is used to assess the temporal evolution of virtual water trade across the years, considering the contribution of different categories of goods. Figure 7 updates previous versions published in the literature (e.g., Konar et al., 2011; Carr et al., 2013; Tuninetti et al., 2017) by either introducing the temporal variability of the uWF of crop-based goods, expanding the number of considered crops and/or extending the temporal range



considered. Total VWT has increased from 750 to 2400 km³/y in the considered period (about 1000 km³/y in 1986).
445 Major categories are cereals, luxury food, seeds/oils and vegetables, with the relative contribution of cereals, which was very large in the 60ies, being outperformed by the other categories in the most recent years. VW volumes associated to cereals has doubled in the considered period, while volumes associated to vegetables has grown 9-fold. The growth of animal-based products is remarkable, but it should be specified that it only reflects the increased trade quantity without considering the temporal variability of uWF.

450 **6 Conclusions**

The globalization of water resources through the international trade of food and agricultural goods is a remarkable global environmental change of our times, and the scientific community is devoting great effort to study it. The quantification of the volumes of water involved in the production and trade of agricultural goods is a key tool to investigate the water-food-trade nexus issues. This study presents an open-source database specifically developed for
455 this purpose. The main outcome of this study is the time-varying unit water footprint in the years 1961-2016 and the virtual water trade matrices for the years 1986-2016 of hundreds of commodities from the food and agricultural sector. The water footprint of production per commodity are available annually in the period 1961-2016. The current database includes a total of 26.8 million data, half of them being elements of the trade matrices. Figure 8 shows the number of shared data per variable and per year.

460 The introduction of a supply-side estimate of the unit water footprint brings much more detail in the water footprint accounting. This is a new concept and it is a key tool in the expedite accounting of the virtual water trade and also of the water footprint of consumption. It also enables the overcome of previous problems related to the non-consideration of re-export and it also enables a more accurate assessment of virtual water trade, with the correct identification of
465 countries of origin of traded goods.

The time-varying unit water footprint is still characterized by a shortcoming that can be improved in future assessments. This concerns with the uWF of animal-based commodities, which is kept constant in time. This limitation can be overcome when reliable data on the country-specific feed composition and diet of each animal type will be available along the considered time period. Despite this shortcoming, the overall open-source database presented in this work
470 aims to help the scientific community and policy makers to quantify and investigate the complex linkages between the global food system and water resource issues. Potential applications of the CWASI dataset range from supporting national-scale policies of water management as well as agricultural policies oriented to the optimization of water use or, ultimately, to provide indications for price formation or for trade agreements based on the efficient and sustainable use



of water resources worldwide. The CWASI database is shared through the Zenodo online open-access repository
475 (Tamea et al., 2020) and it is planned to be improved and updated in the future, capitalizing contributions from the
overall scientific community.

Data availability

Data are available on the Zenodo repository at doi.org/10.5281/zenodo.3987468 (Tamea et al., 2020).

Author contribution

480 ST, MT and IS analysed the data, developed the codes, and organized the dataset. All authors designed the methods and
FL supervised the work, ST prepared the manuscript with contributions from all co-authors.

Competing interests

The authors declare not to have competing interests.

485 Acknowledgments

The authors acknowledge funding support provided by the European Research Council (ERC) through the project
“Coping with water scarcity in a globalized world” (ERC-2014 CoG, project 647473). Information about project, data
and results may be found at the web site <https://watertofood.org/data>. Giuseppe Zaccaria is also acknowledged for early
contributions on the data collection and organization. The authors declare to have no conflict of interests.

490

References

- Allan, J.: Virtual water: A strategic resource - Global solutions to regional deficits, *Ground Water*, 36 (4), 545–546,
1998.
- 495 Carr, J. A., D’Odorico, P., Laio, F. and Ridolfi L.: Recent history and geography of virtual water trade, *PLoS ONE*, 8,
2, e55825, doi: 10.1371/journal.pone.0055825, 2013.
- Dalin, C., Konar, M., Hanasaki, N., Rinaldo, A., and Rodriguez-Iturbe, I.: Evolution of the global virtual water trade
network, *Proc. Natl. Acad. Sci. USA*, 109(16), 5989–5994, doi: 10.1073/pnas.1203176109, 2012.
- Dietzenbacher, E., Ignacio Cazarro, I., Arto, I.: Towards a more effective climate policy on international trade, *Nature
Comm.* 11, 1130, doi: 10.1038/s41467-020-14837-5, 2020.



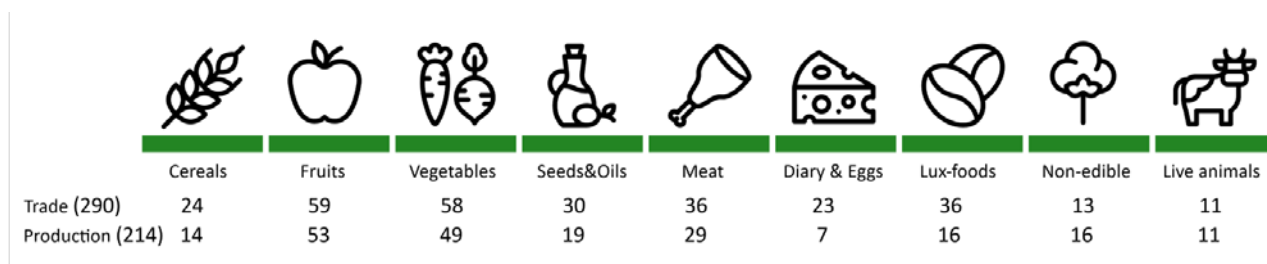
- 500 D’Odorico, P., Carr, J., Dalin, C., Dell’Angelo, J., Konar, M., Laio, F., Ridolfi, L., Rosa, L., Suweis, S., Tamea, S.,
Tuninetti, M.: Global virtual water trade and the hydrological cycle: patterns, drivers, and socio-environmental
impacts, *Environ. Res. Lett.*, 14, 5, 053001, doi: 10.1088/1748-9326/ab05f4, 2019.
- Duarte, R., Yang, H.: Input–Output and Water: Introduction to the Special Issue, *Econ. Syst. Res.*, 23, 341–351. doi:
10.1080/09535314.2011.638277, 2011
- 505 FAO: Technical Conversion Factors for Agricultural Commodities, Food and Agricultural Organization. Available
online at: www.fao.org/fileadmin/templates/ess/documents/methodology/tcf.pdf, 2013.
- FAOSTAT: Online database of the Food and Agriculture Organization. Available online at:
<http://www.fao.org/faostat/en/#home>. Accessed on 12-16/07/2018, 2018.
- Feng, K., Chapagain, A., Suh, S., Pfister, S., Hubacek, K.: comparison of bottom-up and top-down approaches to
510 calculating the water footprint of nations, *Econ Syst Res*, 23, 4, 371-385, doi: 10.1080/09535314.2011.638276,
2011.
- Gehlhar, M.: Reconciling bilateral trade data for use in GTAP, GTAP Technical paper, n.10, 1996.
- Geschke, A., Hadjikakou, M.: Virtual laboratories and MRIO analysis – an introduction, *Econ. Syst. Res.*, 29, 2, 143-
157, doi: 10.1080/09535314.2017.1318828, 2017.
- 515 Hoekstra, A.: Water footprint assessment: evolvement of a new research field, *Water Resour. Manage.*, 31, 3061-3081,
doi: 10.1007/s11269-017-1618-5, 2017.
- Hoekstra, A., Chapagain, A., Aldaya, M., and Mekonnen, M.: The Water Footprint Assessment Manual: Setting the
Global Standard, Earthscan, London, 2011.
- Hoekstra, A., Chapagain, A. K.: Globalization of water: Sharing the planet’s freshwater resources. Oxford, UK:
520 Blackwell Publishing, 2008.
- Hoekstra, A., Chapagain, A.: Water footprint of nations: Water use by people as a function of their consumption
pattern, *Water Resour. Manage.*, 21 (1), 35-48, doi: 10.1007/s11269-006-9039-x, 2007.
- Hubacek, K., Feng, K.: Comparing apples and oranges: Some confusion about using and interpreting physical trade
matrices versus multi-regional input–output analysis, *Land Use Policy*, 50, 194-201, doi:
525 10.1016/j.landusepol.2015.09.022,, 2016
- Kastner, T., Kastner, M., Nonhebel, S.: Tracing distant environmental impacts of agricultural products from a consumer
perspective, *Ecol. Econ.*, 70 (6), 1032-1040, doi: 10.1016/j.ecolecon.2011.01.012, 2011.
- Konar, M., Dalin, C., Suweis, S., Hanasaki, N., Rinaldo, A., and Rodriguez-Iturbe, I.: Water for food: The global virtual
water trade network, *Water Resour. Res.*, 47, W05520, doi: 10.1029/2010WR010307, 2011.
- 530 Lenzen, M., Moran, D., Kanemoto, K., Geschke, A.: Building EORA: a multi-region Input–Output database at high
country and sector resolution, *Econ Syst Res*, 25, doi: 10.1080/09535314.2013.769938, 2013.



- Mekonnen, M., and Hoekstra, A.: The green, blue and grey water footprint of crops and derived crop products, Value of Water Research Report Series, n. 47, UNESCO-IHE Delft, the Netherlands, 2010a.
- Mekonnen, M., and Hoekstra, A.: The green, blue and grey water footprint of farm animals and animal products, Value of Water Research Report Series, n.48, UNESCO-IHE Delft, the Netherlands, 2010b.
- 535 of Water Research Report Series, n.48, UNESCO-IHE Delft, the Netherlands, 2010b.
- Mekonnen, M., and Hoekstra, A.: A global assessment of the water footprint of farm animal products, *Ecosystems*, 15, 401–415, doi: 10.1007/s10021-011-9517-8, 2012.
- Miller, R. E., and Blair, P. D.: *Input-Output Analysis: Foundations and Extensions*, Cambridge University Press, 2009.
- Pfister, S., Bayer, P., Koehler, A., and Hellweg, S.: Environmental impacts of water use in global crop production: hotspots and trade-offs with land use, *Environ. Sci. Technol.*, 45, 5761–5768, doi: 10.1021/es1041755, 2011.
- 540 hotspots and trade-offs with land use, *Environ. Sci. Technol.*, 45, 5761–5768, doi: 10.1021/es1041755, 2011.
- Pfister, S. and Bayer, P.: Monthly water stress: spatially and temporally explicit consumptive water footprint of global crop production, *J. Clean. Prod.*, 73, 52–62, doi: 10.1016/j.jclepro.2013.11.031, 2014.
- Pfister, S., Vionnet, S., Levova, T., Humbert, S.: Ecoinvent 3: assessing water use in LCA and facilitating water footprinting, *Int. J. Life Cycle Assess.*, 21, 1349–1360, doi: 10.1007/s11367-015-0937-0, 2016.
- 545 Stadler, K., Wood, R., Bulavskaya, T., Södersten, C. J., Simas, M., Schmidt, S., ... and Giljum, S.: EXIOBASE 3: Developing a time series of detailed environmentally extended multi-regional input-output tables, *Journal of Industrial Ecology*, 22(3), 502–515, doi: 10.1111/jiec.12715, 2018.
- SEI: A vision for Trase: 2016–2020, A report on TRASE (Transparency for Sustainable Economies), Stockholm Environment Institute and Global Canopy Programme. Available online at <https://trase.earth/>, 2019.
- 550 Tamea, S., Soligno, I., Tuninetti, M., and Laio, F.: CWASI database of virtual water trade and consumptive water footprint of agricultural products. Zenodo repository. Dataset accessed on 10/08/2020 at <http://dx.doi.org/10.5281/zenodo.3987468>, 2020.
- Tukker, A., Dietzenbacher, E.: Global Multiregional Input-Output frameworks: an introduction and outlook, *Economic Systems Research*, 25, 1, 1–19, doi: 10.1080/09535314.2012.761179, 2013.
- 555 Tuninetti, M., Tamea, S., D’Odorico, P., Laio, F., and Ridolfi, L.: Global sensitivity of high-resolution estimates of crop water footprint, *Water Resour. Res.*, 51 (10), 8257–8272, doi: 10.1002/2015WR017148, 2015.
- Tuninetti, M., Tamea, S., Laio, F., and Ridolfi, L.: A Fast Track approach to deal with the temporal dimension of crop water footprint, *Environ. Res. Lett.*, 12 (7), 074010, doi: 10.1088/1748-9326/aa6b09, 2017.



Figures



565 **Figure 1: Commodities considered in the analysis, split into 9 categories: number of items in the trade and production dataset. Icons from Flaticon.com.**

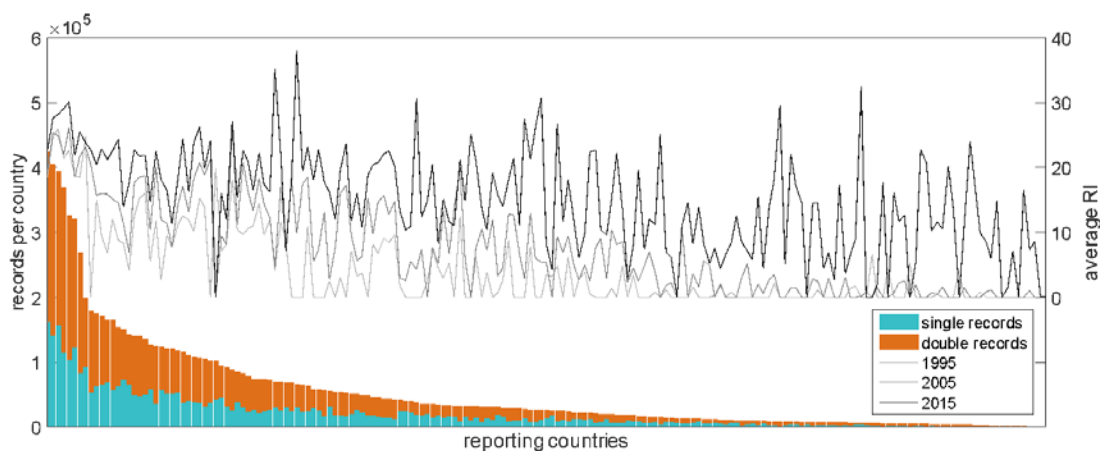
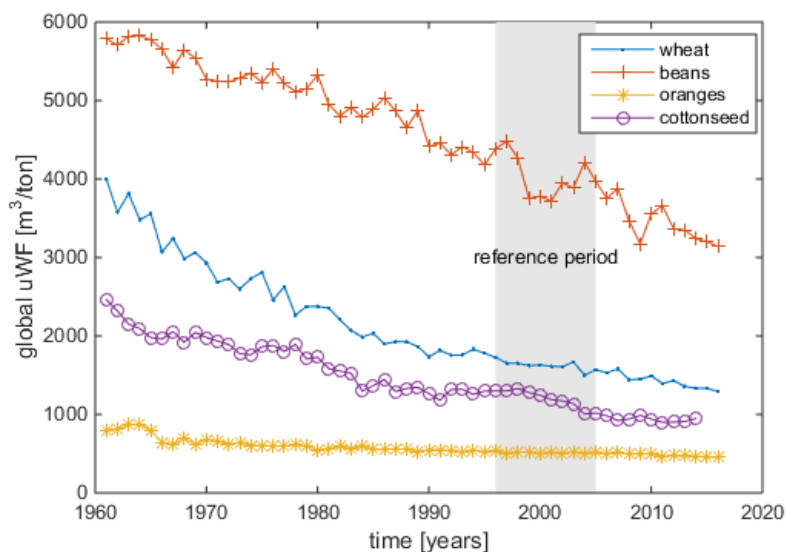
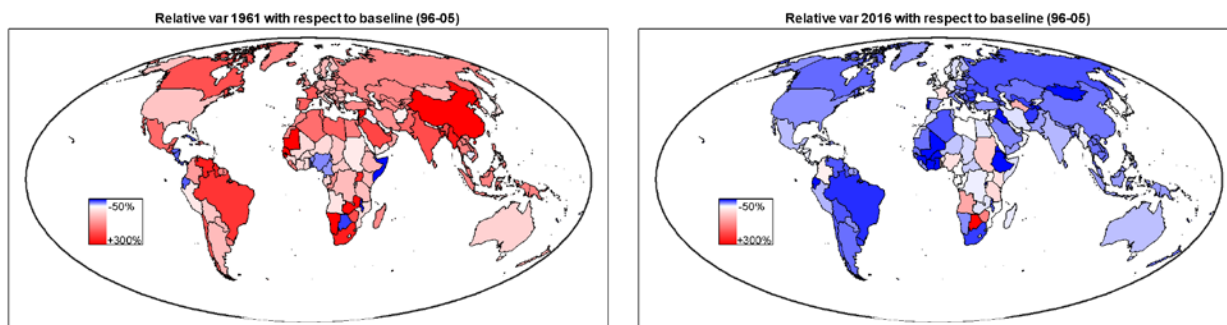


Figure 2: Number of single and double records per reporting country (including all partners, all items and all years). Right axis indicates the country-specific Reliability Index averaged over all items in 3 separate years.



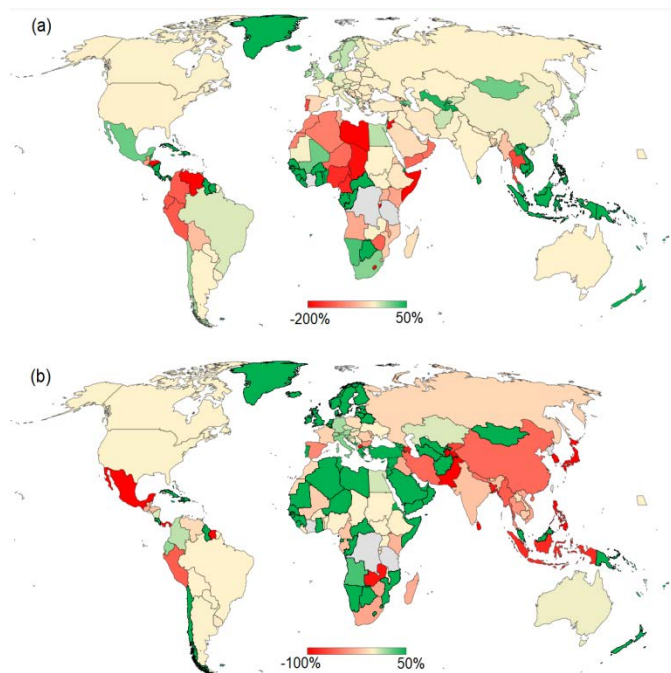
570

Figure 3: Production-weighted global uWFs along the period 1961-2016 for wheat, beans, oranges and cotton.

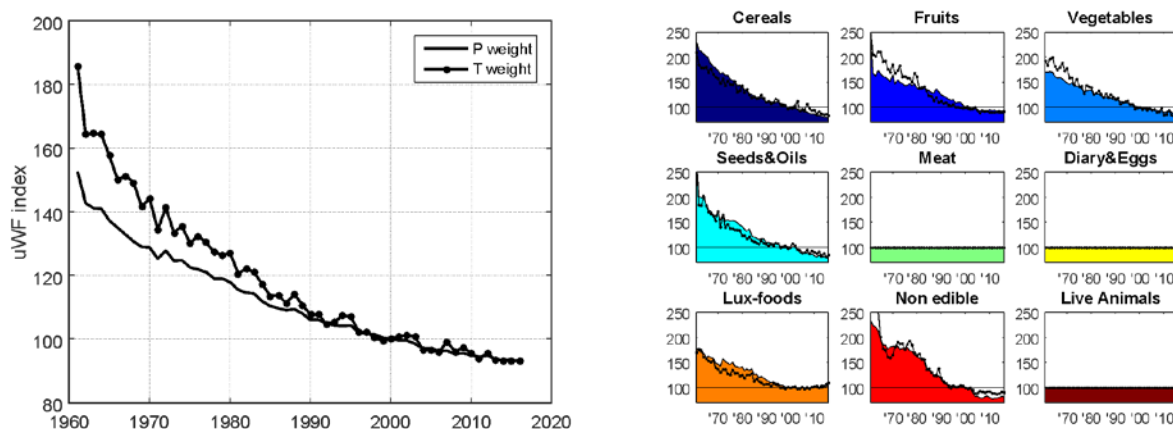


575

Figure 4: Relative change (in m³/ton) in the uWFp of wheat in 1961 (left) and 2016 (right) with respect to the average in 1996-2005, using identical color ranges: red/blue colors identify higher/lower values and color intensity scales with change values. (Maps created with Matlab® R14 software, Mapping Toolbox v.2.0.3).



580 **Figure 5: Percentage difference between the uWF of production and supply of wheat (a) and soybean (b) in year 2016, calculated as the difference between uWFs and uWFp, normalized by uWFs. Bold green countries do not produce the crop; hence they only have a supply-side uWF. (Maps created with Microsoft Power Map for Excel, © Microsoft).**



585 **Figure 6: Temporal variability of uWF indexes weighted with agricultural production (solid) and export (with dots) in year 2000, aggregated across all goods (left), and (right) split into the 9 categories of goods.**

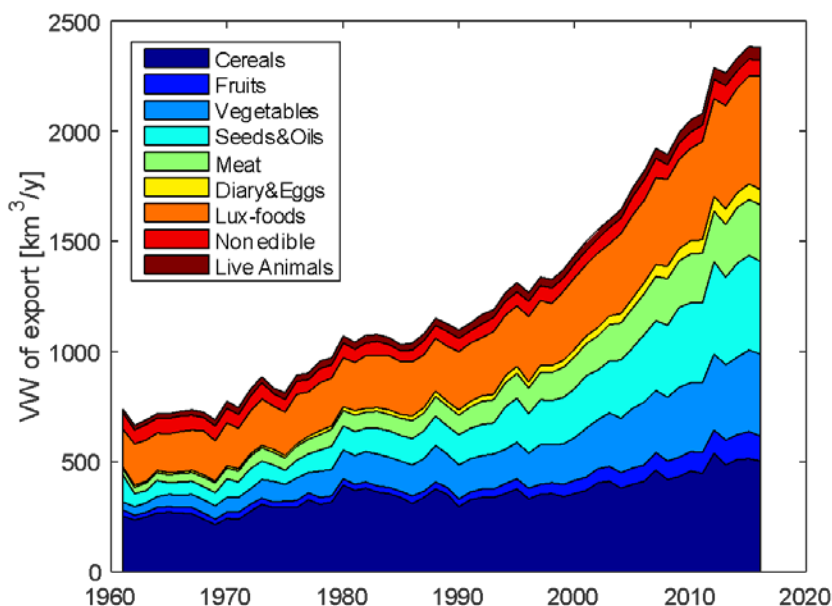
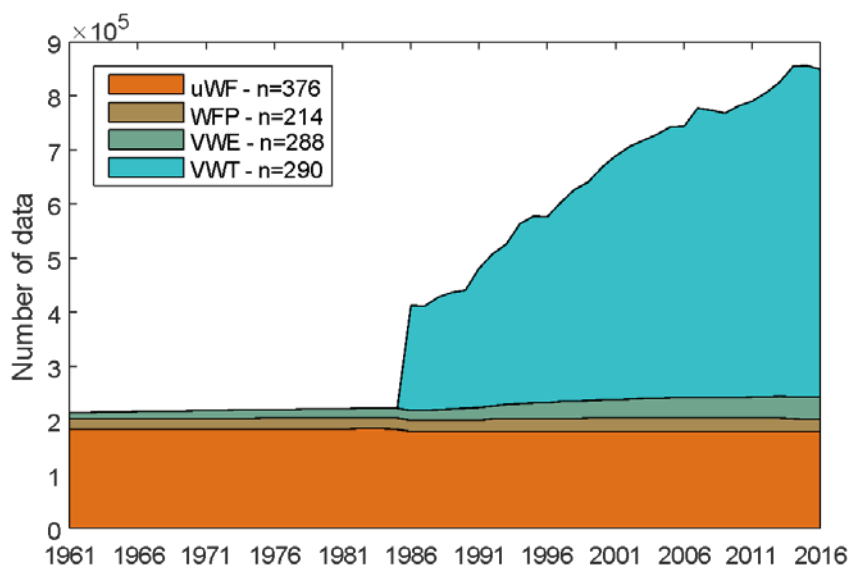


Figure 7: Global virtual water trade (as derived from export data) from 1961 to 2016 considering the 9 categories of goods from Figure 2.



590 Figure 8: Number of data in the CWASI database, per year and per variable (uWF: unit water footprint, WFP: water footprint of production, VWE: virtual water export, VWT: virtual water trade, n: number of commodities).