October 2020

# Working Paper

015.2020

# Seasonality Fingerprint on Global Trading of Food-commodities. A Data-mining Approach

S. Quaini, S. Saccani, S. Vergalli, L. Assom, M. Beria, A. Codello, M. Monaco, R. Sabatini

2030 Agenda
Series Editor: Sergio Vergalli

# Seasonality Fingerprint on Global Trading of Food-commodities A Data-mining Approach

By S. Quaini[1], S. Saccani[1], S. Vergalli[1,2], L. Assom, M. Beria, A. Codello[3], M. Monaco, R. Sabatini

1 Fondazione Eni Enrico Mattei, Milano
2 University of Brescia, Department of Economics and Management
3 Universidad de la Republica, Montevideo

**Summary**

We analyze the United Nations commodities trade database (UN comtrade), comprised of international commodities exchanges in volume and price with monthly resolution. We introduce a trade impact index to quantify the impact, in terms of distance travelled, of importing a specific food raw commodity in a specific period of the year and in a specific country of the world. This index captures the seasonal exchange of raw commodities in an insightful and concise manner.

**Keywords**: Trade, Food Commodities, Resource Access and Availability, Index, Fingerprint, Data Mining

**JEL Classification**: F18, Q17, C8, C43

*Address for correspondence:*
Sergio Vergalli
Fondazione Eni Enrico Mattei
C.so Magenta, 63
20123 Milano
Italy
E-mail: sergio.vergalli@feem.it

# Seasonality Fingerprint on Global Trading of Food-commodities. A Data-mining Approach

Luigi Assom, Marco Beria, Alessandro Codello*, Maurizio Monaco, Stefania Quaini†, Riccardo Sabatini, Sebastiano Saccani‡ and Sergio Vergalli§

**Abstract**. We analyze the United Nations commodities trade database (UN comtrade), comprised of international commodities exchanges in volume and price with monthly resolution. We introduce a *trade impact index* to quantify the impact, in terms of distance travelled, of importing a specific food raw commodity in a specific period of the year and in a specific country of the world. This index captures the seasonal exchange of raw commodities in an insightful and concise manner.

**Keywords**: Trade, Food Commodities, Resource Access and Availability, Index, Fingerprint, Data Mining
**JEL Classification**: F18, Q17, C8, C43

PACS numbers: 43

## INTRODUCTION

International trade of food commodities has increased rapidly in volume [1] and complexity [2] over the last few decades. Just to name a few examples, trade of crops and livestock products from China increased from about 0.6 Bn$ in 2000 to about 3 Bn$ in 2013, while in 2016 Brazil exported 53% of its soybean production [3]. In the European Union, the internal export of agricultural goods among member states accounted for an estimated 11% of all exports in 2018. Exports between EU members and the rest of the world accounted for another 7% [4]. This trend is fueled also by the changes in dietary habits, especially in developing countries that are experiencing an increase in the consumption of meat [5], and general population growth. In fact, total meat production in the developing world tripled between 1980 and 2002, from 45 to 134 million tons [6] and production of livestock feed is expected to more than double as global population reaches 10 billion [7].

The scientific community pointed out various effects of this trend. For instance, many countries rely more and more on imported goods for their food consumption and supply [8, 9]. The land area and resources devoted to exports rather than internal consumption is increasing, and it is currently estimated that one fifth of all cropland globally is devoted to exports and increasing, while the amount of land devoted to local consumption is approximately constant [1, 10, 11]. This re-organization of the agriculture and food sector has the effect of displacing ecological and social impacts. Effects such as greenhouse gas emissions and land and water use get in fact moved from consuming to producing countries [12]. Such impacts are typically not prioritized in policy-making [13–15]. Furthermore, with currently ongoing international trade tension between USA/China and the possible consequences of Brexit, easily quantified insight into international food trade become increasingly appealing. Given the volume of most international trade data-sets, it is impossible to obtain such insights without creating representative metrics, valid across different food commodities and time of the year.

Motivated by this need, in this work we create an index, that we call *Trade Impact Index* (TII), that is indicative of the distance a given imported product must travel in order to reach the consumer, in a particular time of the year. This index is closely related to the concept of food miles [16, 17]. Though food miles are not a comprehensive environmental impact metric [17–20], and in fact displacing production to countries with a better yield may actually reduce such impact [21–24], food miles are still a valuable metric as the greenhouse gases emission due to transport is still significant, even though not always major, component of total emission related to production and consumption. We believe TII is also a valuable tool in raising awareness about global production practices, interconnections and resilience of countries, especially in terms of food security. In fact food trading contributes to countries inter-dependencies that are susceptible of adverse cascading effects in case of extreme environmental, geo-political or health related events. The COVID-19 pandemic has in fact triggered intense discussions about the vulnerability of the world's food systems and food supply chains and about the roles of different types of supply chains, e.g. local vs. global, in providing food security.

After introducing the TII, it is applied in a data mining context. This allows for a novel analysis of the global food market. We go beyond previous work [25] as we analyze data on a monthly timescale, therefore capturing sea-

*Instituto de Fisica, Facultad de Ingenieria, Universidad de la Republica, Montevideo, Uruguay
†Fondazione Eni Enrico Mattei
‡Fondazione Eni Enrico Mattei
§Department of Economics and Management, University of Brescia and Fondazione Eni Enrico Mattei, Corso Magenta 63, Milan, Italy

| **Field** | **Data Type** |
|---|---|
| **Reporting month** | Numeric |
| **HS commodity Code** | Numeric |
| **Reporting country Code** | Numeric |
| **Partner country Code** | Numeric |
| **Trade flow type** | Categorical |
| Weight (in Kilograms) | Numeric |
| Value (in US-Dollars) | Numeric |

TABLE I: Fields available in the UN comtrade main table. Fields in bold font being are the table key.

sonal pattern in food commodities trading. To the best knowledge of the authors this pattern is mostly ignored in macro environmental impact studies, though seasonality of fresh food products necessarily has an impact on international trade. Also, we do not complement the data with Multi Regional Input Output (MRIO) models, which are found at times to strongly disagree with physical data [26] and do not offer a time resolution short enough for our purposes, but we aim at making the best possible use of physical data, with the least amount of assumptions possible.

## THE DATA-SET

We analyze the United Nations commodities trade database (comtrade) [27], containing data collected by the UN International Merchandise Trade Statistics (IMTS) [28]. In turn, these data are collected by national customs authorities. The UN comtrade arguably contains the most detailed bilateral monthly information about international trades. The database is in essence a list of records, having the fields specified in Table I, the fields in bold font being the table key. Though various commodity code systems are available in the UN comtrade database, we use the most commonly used and recommended Harmonized Commodity Description and Coding System (HS) classification [29]. HS-codes are identified by three couples of digits, descriptive of the properties of the product. The first two digit represent the highest level of aggregation (01 - “Animals; live”, 02 - “Meat and edible meat offal”, 03 - “Fish and crustaceans”, etc.). Every additional couple of digits increases the classification detail. In this paper, only categories pertaining to food commodities are studied (the first two digits of the HS-codes fall between 01 and 24). We will be using HS-codes having four or six digits.

Trade flow type can be either import or export. In case the trade flow is an import (export), the reporting country is the importing (exporting) country, while the partner is an exporting (importing) country. There exist two other trade flow types (re-import and re-exports), however these are seldom used for food commodities, see

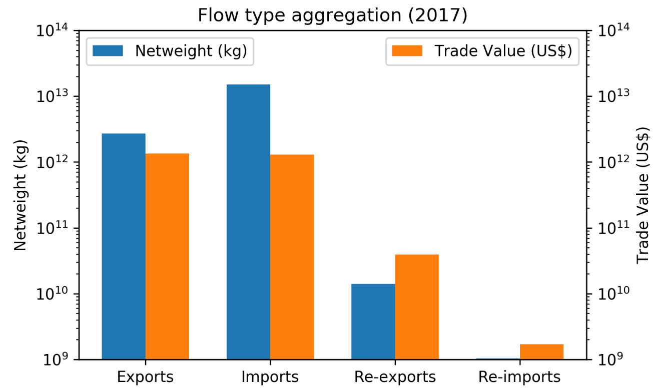



FIG. 1: (Color online) Aggregation of traded net weight and value for the four trade flow types available in the UN comtrade database, 2017 data

FIG. 1.

## DATA PREPROCESSING

In this section, we deal with the problem of translating a set of records of trade flows to an actual representation of the flow of goods, while exploiting maximally the information available within UN Comtrade. A number of well-known issues arise when addressing this problem, for a comprehensive discussion we refer to [30]. We will focus on the following ones:

- record with inconsistent traded volumes (in Kilograms) or/and value (in US-Dollars);
- inconsistent reporting: dealing with double reporting, single reporting, missing reporting;
- re-imports and re-export flows: intermediate steps in trading may hinder the actual travel of a particular good.

We will denote with $v^i_{ab,t}$ the reported value of a certain good transiting from country $a$ to $b$, being of trade type $t$, either import or export, for the commodity code $i$. The index $t$ is there to denote the reporting country, $a(b)$ if $t$ =export(import). We will use $z^i_{ab,t}$ as reported weight of the same trade. All quantities will be calculated on a monthly basis, unless specified otherwise. We do not insert the reporting month index as the procedure reported in the following will be the same for each month and will not depend on the preceding of following months’ values. We list in the following the approach followed to mitigate each of the issues.

### Records with inconsistent traded volumes (in Kilograms) or/and value (in US-Dollars)

Of the total $8.4 \cdot 10^6$ documented trades from 2017, the traded net weights of roughly $6 \cdot 10^5$ trades were not reported. Traded values were missing for only about $1.4 \cdot 10^3$ of the records. We believe this is because the traded value, being linked to insurance and shipping liabilities, is a more important datum and therefore more reliably reported. To deal with these lacking data entries, the relation between weight and value of traded goods is studied. Missing data is then inferred from this relation.

Figure 2 depicts the overall distribution of the reported weights and values of traded commodities (zero and missing values are not included in the histogram). Inspection of Figure 2 reveals that the weights and values approximately follow a lognormal distribution. Trade records for which the traded mass was below 10Kg were noticeably frequent and inexpensive. Such records are therefore considered anomalies and are not included in the estimation of the relation between mass and value. The weight and value distributions allow us to define a proxy price from the global average price per kg, as in equation (1).

$$V^i = \frac{\sum\limits_{ab,t} v^i_{ab,t}}{\sum\limits_{ab,t} z^i_{ab,t}} \tag{1}$$

where the summation is only taken over records for which both the weight $z^i_{ab,t}$ and the value $v^i_{ab,t}$ are available. Records missing both a weight and a value are excluded from further investigation. For each record $i$ such that $v^i_{ab,t}$ is missing but $z^i_{ab,t}$ is provided, we then interpolate the missing value from equation (1) by taking

$$v^i_{ab,t} = z^i_{ab,t} V^i \tag{2}$$

And likewise, missing values $z^i_{ab,t}$ are interpolated as

$$z^i_{ab,t} = v^i_{ab,t} / V^i \tag{3}$$

**Inconsistent reporting**

Because of the raw nature of the IMTS data, trades may be reported twice (once by both the exporter and the importer) or only once (either by the exporter or the importer). In the worst case, trades may not be reported at all. This is a well known problem for bilateral trade data, and it can also depend on less obvious factors, including different categorization of the same good by importer and exporter country [30]. Also, in case of bilateral reporting, acceptable agreement between two records is not guaranteed. 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 [30].

In the literature, these issues of inconsistent reporting are often addressed by assuming that different countries have different reporting qualities [31]. Only data from the

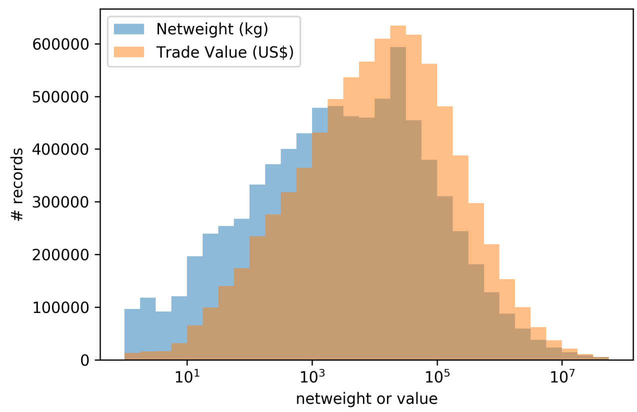



FIG. 2: (Color online) Distribution of valid traded weight and values

country considered most reliable is then considered in every trade. The difference in reporting qualities is clearly noticeable in Figure 3, depicting the number of traded products reported by each country in a given time interval.

In this paper, the total exported mass $z^i_{a,b}$ of a commodity $i$ from a country $a$ to a country $b$ are taken to be the maxima of the mass and value as documented by exporter $a$ and importer $b$. This is expressed more concisely in equation (4).

$$z^i_{ab} = max\left(z^i_{ab,import},\ z^i_{ab,export}\right) \tag{4}$$

Where we denote the presumed actual trade flows from country $a$ to $b$ by $z^i_{ab}$ ($t$ index dropped). This is because:

- when only one record is present in the grouping operation, we are using all the information available;
- when more than one record is present, we argue that the one reporting higher quantity for the trade aggregates a larger number of IMTS data, therefore it's probably a more accurate value.

In this way, we also recover partial trade flows involving non reporting countries.

**Re-imports and re-export flows**

Use of re-import and re-export flows in the UN comtrade database is rather rare, see Figure 1. Even though IMTS criteria recommend recording as partner the country of origin of the product in case of imports, and the ultimate country of destination in case of exports [28], there are cases in the databases of countries exporting and importing in the same month large and comparable amounts of the same product code. For example, Germany reports a large amount of both imports and exports of product code 151321 - "Vegetable oils; palm kernel or babassu oil . . . ". Because according to FAOstat data [3] the country does not have any production of palm kernels, we

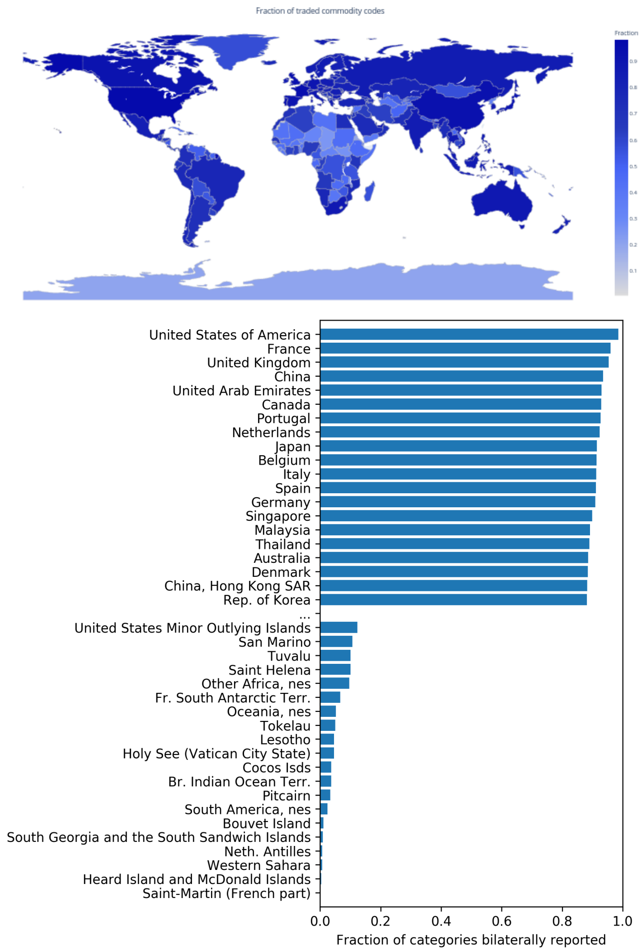



FIG. 3: Fraction of total commodity codes traded by each country, as bilaterally reported in 2017 (highest and lowest 20 countries in bar plot)

can conclude that the reported exports of commodity 151321 are actually a form of re-export. We thus treat re-import/re-export as if being reported as import/export flows, and instead seek a different strategy to retrieve the ultimate origin/destination of products. This strategy is based on the consideration that if a particular country registers a large amount of import and export of the same good, this country is probably is just a country of transit for that good. We will make the following assumptions:

- All goods within a certain category are equally valuable and indistinguishable from each other.
- Each country consumes its local production first, then exports the excess production or imports a quantity to satisfy the excess demand.
- All goods transiting a country are redistributed in the next step in the network evenly, without regard to their original provenience.

Also, because we are not focusing on a small subset of products such as other authors, we ignore conversions between food commodities. Though this a recognized limitation of the present analysis, we thus avoid introducing very large and parametric transformation matrices, that would hinder our goal of making the least number of assumptions possible. The procedure we use is the one employed in [25], but we substitute the assumption of internal consumption being proportional to imports and local production with local production as being consumed first. This allows us to proceed in the analysis without using data for local production, focusing only on exported production, that we assume being the difference between exports and imports.

As the procedure is repeated for each commodity in the same manner, in this section we will temporarily drop the commodity index $i$. We will denote with $p_a$ the exported production quantity of country $a$. Using our hypothesis, it can be calculated as:

$$p_a = \max\left(\sum_b z_{ba} - \sum_b z_{ab},\ \ 0\right). \tag{5}$$

We denote with $\hat{\boldsymbol{p}}$ the diagonal matrix having $p_a$ values along its diagonal (similar notation is used for other diagonal matrices). With $x_a$ we denote the values of the domestic production plus imports (hereafter DMI). Because we are ignoring production destined for internal consumption, it is set as the sum of imports plus the exported production

$$x_a = \sum_a z_{ab} + p_a \tag{6}$$

We seek at calculating $\bar{\boldsymbol{R}}$, the matrix having elements $\bar{r}_{ab}$, the amount of the exported production flow of good $i$ from country $a$ being consumed in $b$. Using the procedure outlined in [25], it can be calculated as

$$\bar{\boldsymbol{R}} = \hat{\boldsymbol{c}} \cdot \boldsymbol{R} \tag{7}$$

$$\boldsymbol{R} = (\boldsymbol{I} - \boldsymbol{A})^{-1} \cdot \hat{\boldsymbol{p}} \tag{8}$$

Where $\hat{\boldsymbol{c}}$ is a diagonal matrix having along its diagonal the values of the internal consumption shares of DMI

$$c_a = \frac{1}{x_a}\left(x_a - \sum_a z_{ab}\right) + \epsilon \tag{9}$$

We add a small constant $\epsilon$ because, even though following the assumptions above countries with $p_a > 0$ do not have a significant consumption of imports, it will allows us in the following to calculate the TII also for those countries. $\boldsymbol{A}$ is the matrix of export shares, defined as:

$$\boldsymbol{A} = \boldsymbol{Z} \cdot \hat{\boldsymbol{x}}^{-1} \tag{10}$$

where $\boldsymbol{Z}$ is the matrix having elements $z_{ab}$.

To illustrate the effect that the procedure has on estimating the original source of products, we report results for a few different commodities. We choose some having very different trading patterns, in order to observe the results in different scenarios. In particular we considered the Inverse Participation Ratio ($IPR$) of a commodity production, defined as:

$$IPR = \sum_a P_a^2 / \left( \sum_a P_a \right)^2 . \tag{11}$$

Where $P_a$ is the annual production of country $a$. The $IPR$ is a measure of how many terms in a sum of positive quantities effectively contribute to the total value (in this case the global production). If a single country accounts for the entire global production, then $IPR = 1$, while if $N$ elements all contribute equally, $IPR = 1/N$. Using the $IPR$ calculated on 2017 production data from FAOstat [3] as guiding metric, we selected the following products:

- commodity code 151321 - "Vegetable oils; palm kernel or babassu oil ...", a commodity having large $IPR$, representing a very concentrated production in a small number of countries (Indonesia and Malaysia mainly),
- commodity code 1003 - "Barley", a commodity having a very low $IPR$, indicative of a distributed production,
- commodity code 0805 - "Citrus fruit; fresh or dried", a seasonal product.

We compare in Table II the annual aggregations of $z_{ab}^i$ with aggregations of $\hat{r}_{ab}^i$ for the selected products, regarding imports of Austria, a small landlocked country. The table also reports production data, to help us assess the validity of the procedure. It is worth stressing that production data are only reported for reference, the calculation of $\hat{r}_{ab}$ being based solely on trade data.

The effect of the procedure is particularly noticeable in the first case. Original data in the UN comtrade (contradicting IMTS guidelines) make it look like the biggest source of palm kernels is Germany, that has no production of such good, while aggregations based on the retraced values agree much better with known large producers. Notice also how this procedure shows that this product ultimately comes from much further away than what is originally reported by the UN comtrade database. For this case, comparison data between $z_{ab}^i$ and $\hat{r}_{ab}^i$ are also reported in Figure 4. Notice like exports from Ecuador, Germany and the Netherlands disappear in $\hat{r}_{ab}^i$, as they are re-exporters of this commodity.

For the second case, there is a smaller effect in terms of the order of the top exporters, though the quantities involved decrease significantly. This is a consequence of Austria being an exporter of barley as well, therefore within our assumptions not all the imported barley is destined to internal consumption but much of it get passed on into the trade network.

For the third case, we notice the procedure helps in reducing inconsistencies between production and exports (again Germany resulting a large exporter of citrus fruit having no reported production). Also, the order of the bigger exporters is changed, though not as dramatically as for the palm kernel case.

## THE TRADE IMPACT INDEX

In this Section, an index is derived to quantify the trading impact, in term of travelled distance, of a certain commodity in a specific period of the year to or from a specific country of the world. The index is based on the traded quantity (in kilograms) and the length of the trade-route (in kilometers).

Consider a commodity $i$ traded between countries $a$ and $b$. We denote with $l_{ab}$ the shortest distance on the globe between the population centers of the countries $a$ and $b$, calculated through the Haversine formula (see [32]), as stated in equation (12).

$$l_{ab} = 2R_{\text{Earth}} \arcsin \sqrt{\sin^2\left(\frac{\phi_a - \phi_b}{2}\right) + \cos\phi_a \cos\phi_b \sin^2\left(\frac{\lambda_a - \lambda_b}{2}\right)} \tag{12}$$

where $R_{\text{Earth}}$ is the Earth's radius and $\phi_l$ ($\lambda_l$) is the latitude (longitude) of the country $l$'s center of population. This is calculated as the population centroid of cities above 15000 inhabitants, data from [33]. We again denote by $\hat{r}_{ab}^i$ the mass of commodity $i$ exported from country $a$ to country $b$ during the time period of interest. For each commodity $i$ and each country $a$, we can now compute the average distance of imports of $i$ to $a$, weighted by the weight shipped. This value is denoted by $c_a^i$ as defined in equation (13).

$$c_a^i = \frac{\sum_{b \neq a} l_{ba} \bar{r}_{ba}}{\sum_{b \neq a} \bar{r}_{ba}} . \tag{13}$$

For each commodity $i$ and country $a$, $c_a^i$ can be interpreted as the mean distance traveled by commodity $i$

| 151321 – Vegetable oils; palm kernel or babassu oil ... | | | 1003 – Barley | | | 0805 – Citrus fruit; fresh or dried | | |
|---|---|---|---|---|---|---|---|---|
| **Exporter** | **Imports from trade data ($z$)** | **Production (FAOstat)** | **Exporter** | **Imports from trade data ($z$)** | **Production (FAOstat)** | **Exporter** | **Imports from trade data ($z$)** | **Production (FAOstat)** |
| Germany | 11.20 | 0 | Hungary | 51 | 1404 | Spain | 63 | 125 |
| Nigeria | 0.81 | 2281 | Czechia | 46 | 1712 | Germany | 26 | 0 |
| Iran | 0.30 | 0 | Slovakia | 23 | 545 | Italy | 23 | 20 |
| Denmark | 0.05 | 0 | Germany | 19 | 10853 | Greece | 8 | 1 |
| Slovenia | 0.03 | 0 | Serbia | 5 | 305 | Turkey | 8 | 10 |
| **Exporter** | **Imports after procedure ($\hat{r}$)** | **Production (FAOstat)** | **Exporter** | **Imports after procedure ($\hat{r}$)** | **Production (FAOstat)** | **Exporter** | **Imports after procedure ($\hat{r}$)** | **Production (FAOstat)** |
| Indonesia | 6.55 | 3830 | Hungary | 21 | 1404 | Spain | 79 | 125 |
| Malaysia | 3.13 | 2281 | Czechia | 7 | 1712 | Greece | 7 | 1 |
| Nigeria | 1.15 | 120 | Slovakia | 2 | 545 | South Africa | 7 | 30 |
| Thailand | 0.82 | 236 | Romania | 1 | 1907 | Turkey | 6 | 10 |
| Iran | 0.30 | 0 | Denmark | 1 | 3992 | Italy | 5 | 20 |

TABLE II: Aggregated value of imported net weight (in milions of Kg) in 2017 in Austria for a few commodities (see text). In the top rows data for $z^i_{ab}$ while in the lower rows data from $\hat{r}^i_{ab}$. It can be noticed how the procedure helps in removing inconsistencies with production data from FAOstat [3].

to reach country $a$. Thus, a large value of $c^i_a$ in physical terms means that most of the mass of commodity $i$ that is imported to country $a$ is transported from very distant locations. The Trade impact Index (TII) of commodity $i$ and country $a$, denoted by $T^i_a$ is now obtained by normalizing $c^i_a$ using half of the Earth's circumference, as in equation (14)

$$T^i_a = \frac{c^i_a}{\pi R_{\text{Earth}}} \tag{14}$$

For given $i$ and $a$, if $T^i_a = 1$, this means that all of commodity $i$ imported to country $a$ is imported from the diametrically opposed side of the planet. Likewise, a value of $T^i_a$ close to zero means that most of the imported mass of $i$ to $a$ is imported from very nearby countries.

Now that we have defined the TII, we wish to briefly illustrate the effect that the retracing procedure described in the previous section has on the calculation. The value of the TII for product 151321 imported in Austria is 0.47 . If the retracing procedure was not used (using elements of $z_b a$ instead of $\bar{r}_{ba}$ in equation 13), it would result in a TII value of 0.04, undermining the utility of the index.

Note that is possible to define an analog quantity to the $T^i_a$ focusing on export rather then imports, by changing the summations in 13 on $b$ instead of $a$. As we are mainly interested in the $TII$ as an awareness instrument for consumers rather than producers, we do not pursue this analysis here.

## ANALYSIS OF THE TII

In this section, we will study the characteristics of the index and its ability to meaningfully describe food trading patterns. We will focus on the year 2017, on countries that are beyond the 5% quantile of total traded value for that year. This is done in order to remove from the analysis countries either very small or with very poor reporting.

In Fig 5 we report the TIIs for the countries Italy, Netherlands and Brazil for seasonal commodities such as citrus-fruits, grapes, apricots and peaches, and processed commodities related to these such as fruit-juices. Observe that, as opposed to fresh fruits, the TII of which oscillate with the seasonality, fruit-juices can be stocked and traded on demand. This results in an almost flat TII for the latter commodity category. We also show the behavior of the TII for bread, pastry and other bakers' products, as well as for food preparations/sauces, all processed commodities, described by almost flat TIIs. We observe that Brazil has a high value of the TII due to the long distance from its trading partners. Observe also the reversed seasonality of seasonal fruits such as grapes or apricots between the countries in the northern and southern hemispheres.

In Figure 6, the mean TIIs of each reporting country are represented graphically. It is calculated as $T_a = \sum_{i,b} T^i_a \bar{r}_{ba} / \sum_{i,b} \bar{r}_{ba}$. Notice how gradations of color are related to the geographical positions of the countries. Notice also that among countries with high TII there are relatively isolated islands. This can be easily understood considering the distance of trading partners for such places. Interestingly, one of the countries with the highest TII is the Falkland islands, that is an isolated island with limited trade with neighbouring Argentina because of territorial disputes, and therefore is forced to trade

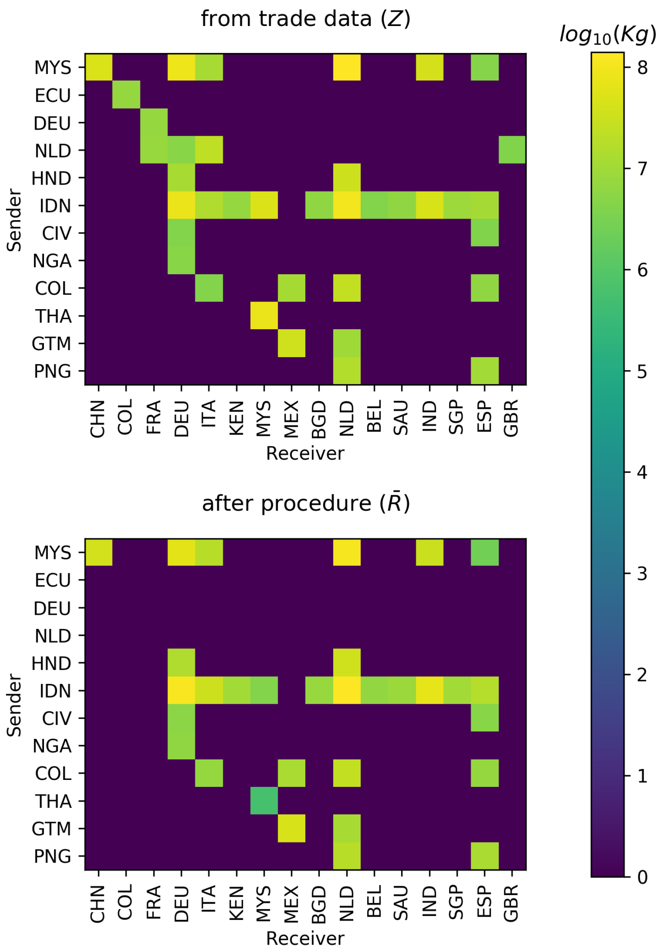
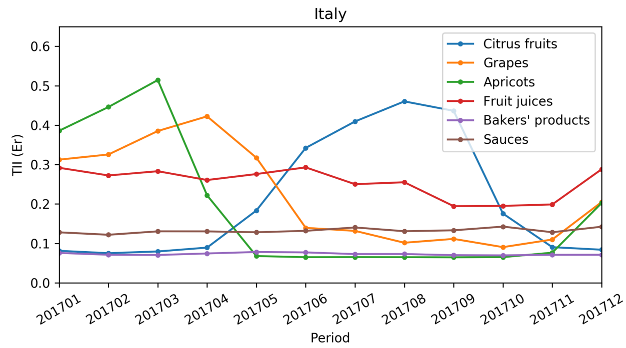
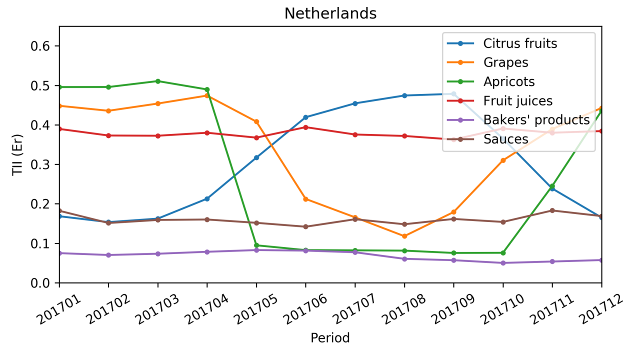


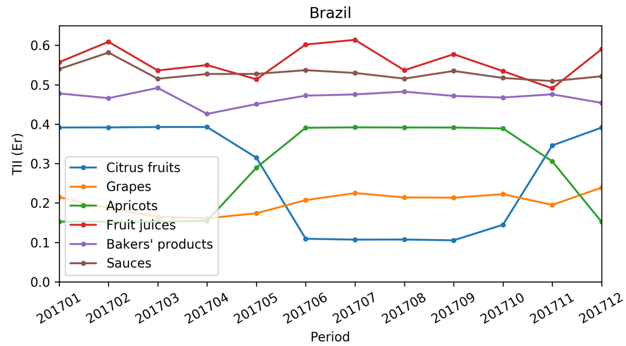

FIG. 4: Elements of the trade matrix for commodity code 151321 - Vegetable oils. . . . In the top rows data for $z^i_{ab}$ while in the lower rows data from $\hat{r}^i_{ab}$. Countries reported using their alpha-3 code.




(a)




(b)




(c)

FIG. 5: For Italy , Netherlands 5(b), and Brazil 5(c) in the year 2017 we show the values of the TII at different months for the commodities 0805 (blue), 0806 (green), 0809 (red), 2009 (cyan), 1905 (purple), and 2103 (yellow), describing respectively the citrus fruits, grapes, apricots, fruit-juice, sauces and bakers' products, and food preparation. The index captures remarkably well the difference between seasonal commodities, characterized by long fluctuations in phase with their seasonality, and processed products, marked by almost flat indexes

with faraway partners. China is also a country with a high TII, due to its ever increasing use of imports from South America and Africa [34]. Countries with a low TII are instead smaller, mostly land-facing countries that have neighbouring trading partners. We think that this has two main reasons. On the one hand, a smaller country tends to have lower distances in formula 12, and also smaller countries imply that we are in practice looking at a finer spatial scale, resulting in a lower TII. On the other hand, small land facing countries have an easy and short trade route to their neighbours, implying the the low TII also results from a physically tighter trade relationships with neighbouring countries. In the opinion of the authors, this is the case for European countries, that tend to have a low TII on average. This also can be related within the food culture of each country, and eventually it may be also a fingerprint about how "local" the local nutritional attitudes are. It can also be argued that the procedure used to trace original sources does not completely remove re-import and re-export biases, because we are ignoring products transformations, thus resulting in an underestimated TII for nearest neighbours countries having strong trade interconnections.

The ability of the index to capture seasonality for fresh products can be inspected by observing the correlation of

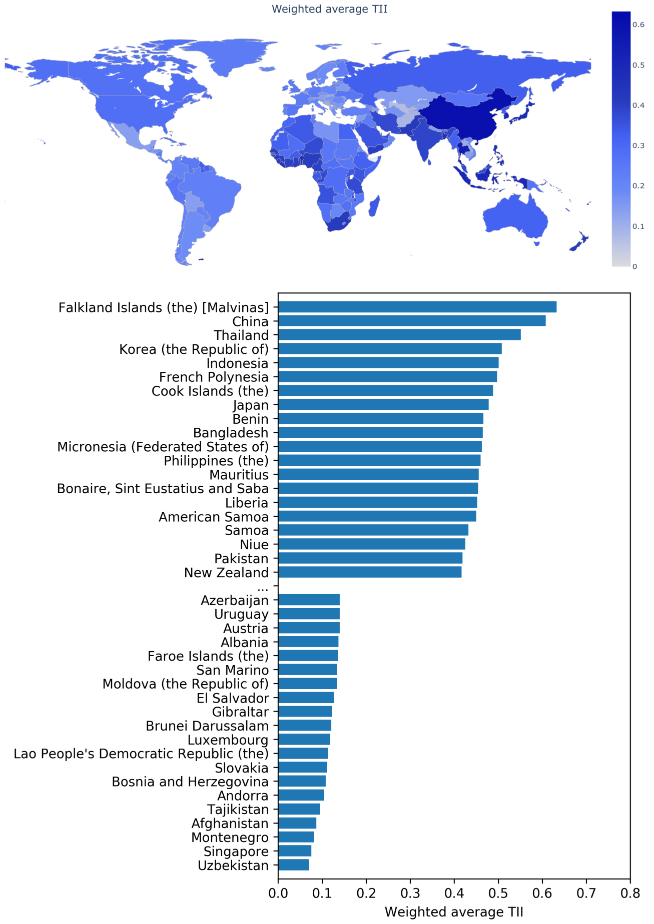



FIG. 6: (Color online) In the top panel the netweight weighted average TII for each reporting country is depicted with a gradation of blue. On the bottom one top and bottom 20 countries by the same measure. Notice how countries having large TII are mostly geographically isolated (islands).

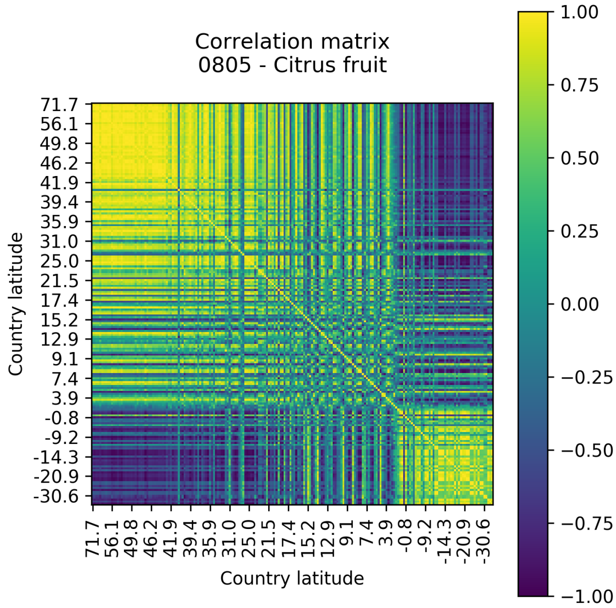



FIG. 7: Correlation matrix of the TII for product category citrus fruit between different countries. The country name is not reported, countries however are ordered by their center of population latitude.

the TII values among countries with different latitudes. In fact we observe that for fresh product the TII correlates positively with countries with similar latitudes, but more importantly it correlates negatively between countries in different hemispheres, see Figure. 7. This is an indication of the ability of the index to capture the difference in production at different latitudes in different times of the year.

In Figure 8, we show the statistics of the TIIs computed for all the four digits reported commodities in 2017. We observe that the shape of the histogram is characteristic of the specific country. Brazil has a peak for a value of the TII of about 0.45, while European countries have a much lower peak at about 0.08.

We furthermore implement a $k$-mean clustering algorithm to see how countries are clustered by the statistics of TIIs. We consider all the $T_a^{i,m}$ values ($m$ being the month index) for four digit commodity codes for each month of 2017 for each country $a$ (201 commodity codes each month). We consider in this section countries $a$ that have at least 400 valid $T_a^{i,m}$ values over $i, m$. We also consider only $i, m$ combinations that have a valid value in at least 180 countries $a$. Missing values of $T_a^{i,m}$ remaining after this selection are set to the average over $a$. We perform a principal component analysis (6 dimensions) on the obtained values, being $a$ the sample index. Finally, in this reduced dimensional space we perform a K-mean analysis where the number of clusters is fixed to be 10. See obtained results in (Fig. 9). Clusters are geographically contiguous, even though the TII values do not contain any explicit geographical information. This is a strong indication of the ability of the index to retain useful information about the trading habits of each particular country. This remains true even using a smaller number of clusters. Also, the cluster containing EU-28 countries is very stable to the number of clusters used, indicating a peculiar trading pattern within Europe.

It's useful to look at a test case to understand how is the index able to track global or local shift in food trading. We focus on the ban Russia enforced on imports of EU food products in August 2014 [35] as a retaliation for European economic sanctions related to the Donbass conflict in Ukraine. Thought the effect varies from one product to another, an overall increase of the TII is evident for most food products, as the need of importing from countries further away emerged. A marked effect is visible exactly at the enforcement of sanction, though the TII for some products such as frozen swine meat has a

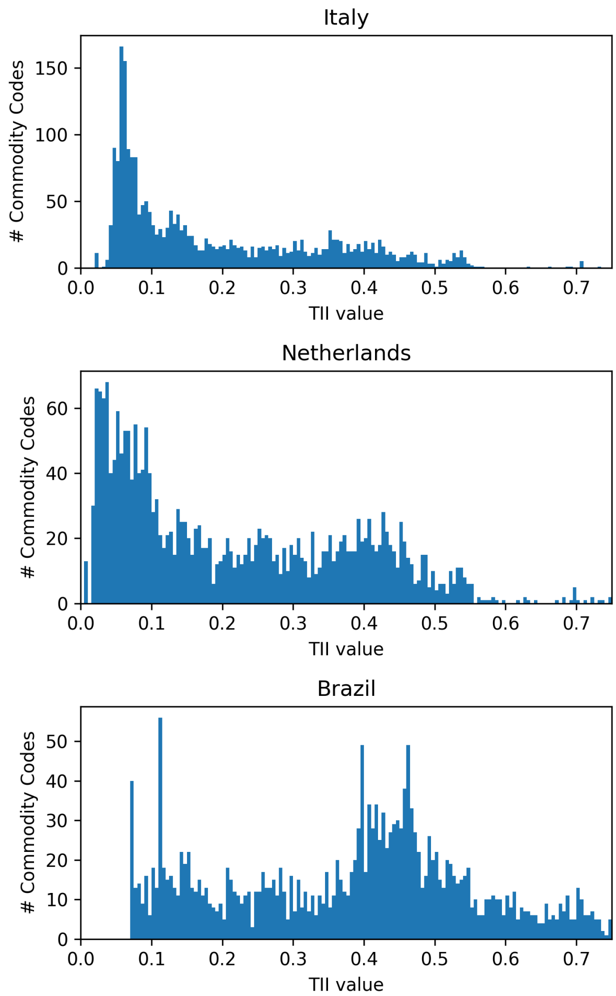



FIG. 8: For Italy, Netherlands, and Brazil in the year 2017, the histograms of the TIIs for all four digit food-commodities.

partial recovery after that. Some processed cheese products appear instead to be affected quite a few month lather, either as a result of a late enforcement of the ban or a late identification of a new trade partner for the product. Also, there is an increase in TII even before the ban was raised for some (frozen swine meat), suggesting either effects coming from the Donbass conflict directly or from some other unrelated effect. Some other categories (generic cheese and curd), display a marked effect because of the ban but also a full recovery within a few months. We suggest this might be due to increased production by neighbouring countries different from EU

## DISCUSSION AND OUTLOOK

We presented a Trade Impact Index, constructed on a monthly timescale, which is representative of the distance a given food product must travel to reach the destination

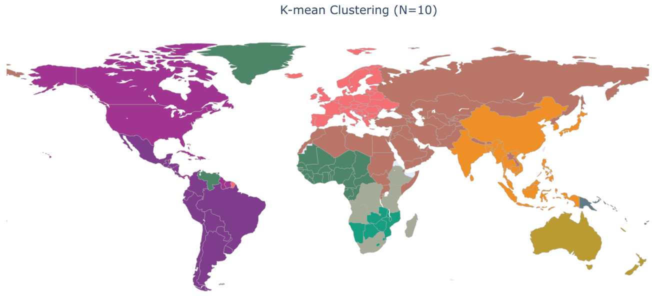



FIG. 9: Countries clustering based on their TII values (procedure in text). Clusters are highly geographically contiguous, even though the TII index does not contain explicit geographical information.

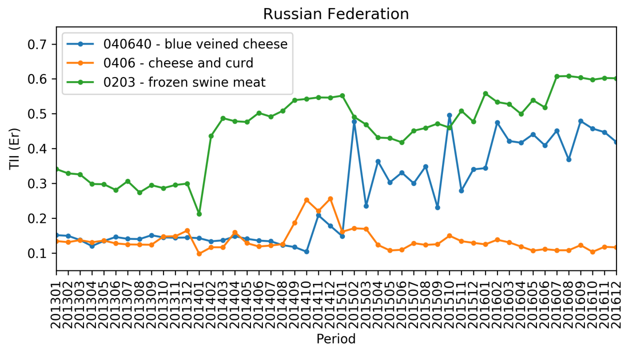



FIG. 10: Effect on the ban (August 2014) on import from European countries by the Russian federation on its TII.

country when subject to trading. The index is calculated using maximally the information contained in the UN comtrade database, while making the least amount of assumptions. It is built mixing together the kilograms of the traded commodity and the length over which this is traded, and is related to the concept of food miles. We employed a strategy to treat hidden re-import and re-export flows in IMTS data to retrace the final origin/destination of each commodity. We show the index is highly representative of trading habits of each country, by displaying its ability to cluster countries with similar geographical position. Also, TIIs of seasonal commodities is anti-correlated for countries pertaining to different hemispheres. This is an indication of the ability of the TII to capture seasonality of fresh food products. The index also captures recent shift in food trades, such as the one caused by the Russian Federation ban on imports of EU food commodities. We therefore argue that the TII is a valuable metric for raising awareness into consumers about international trading trend of food commodities.

We think further work needs to be done to address the problem or re-import and re-exports in a fully satisfactory manner, as this issue is aggravated by the transformation steps that are present in the food chain. These transformations steps cannot be easily treated without making a large number of assumptions in the form of comprehensive transformation matrices. On the other hand, the improvement of recording practices will mitigate this problem over time. Also, while the analysis presented here was dealing with international trade, it would be useful to complement to the present analysis local food production, as it would allows us to extend the scope of the analysis also to intra-national consumption patterns.


[1] T. Kastner, K.-H. Erb, and H. Haberl, “Rapid growth in agricultural trade: effects on global area efficiency and the role of management,” *Environmental Research Letters*, vol. 9, no. 3, p. 034015, 2014.

[2] G. K. MacDonald, K. A. Brauman, S. Sun, K. M. Carlson, E. S. Cassidy, J. S. Gerber, and P. C. West, “Rethinking agricultural trade relationships in an era of globalization,” *BioScience*, vol. 65, no. 3, pp. 275–289, 2015.

[3] FAO, “Faostat,” 2019.

[4] European Commission, *Agriculture in the European Union and the Member States - Statistical factsheets.* 2019.

[5] K. J., “Food consumption trends and drivers,” *Philosophical Transactions of The Royal Society B Biological Sciences*, vol. 365, no. 1554, pp. 2793–2087, 2010.

[6] World Bank, “Minding the stock : Bringing public policy to bear on livestock sector development. world bank,” 2009.

[7] D. Tilman, “Global environmental impacts of agricultural expansion: the need for sustainable and efficient practices,” *Proceedings of the National Academy of Sciences*, vol. 96, no. 11, pp. 5995–6000, 1999.

[8] H. Haberl, T. Kastner, A. Schaffartzik, and K.-H. Erb, “How far does the european union reach? analyzing embodied hanpp,” in *Social Ecology*, pp. 349–360, Springer, 2016.

[9] M. Porkka, M. Kummu, S. Siebert, and O. Varis, “From food insufficiency towards trade dependency: a historical analysis of global food availability,” *PloS one*, vol. 8, no. 12, p. e82714, 2013.

[10] A. Y. Hoekstra and M. M. Mekonnen, “The water footprint of humanity,” *Proceedings of the national academy of sciences*, vol. 109, no. 9, pp. 3232–3237, 2012.

[11] P. Meyfroidt, E. F. Lambin, K.-H. Erb, and T. W. Hertel, “Globalization of land use: distant drivers of land change and geographic displacement of land use,” *Current Opinion in Environmental Sustainability*, vol. 5, no. 5, pp. 438–444, 2013.

[12] K.-H. Erb, F. Krausmann, W. Lucht, and H. Haberl, “Embodied hanpp: Mapping the spatial disconnect between global biomass production and consumption,” *Ecological Economics*, vol. 69, no. 2, pp. 328–334, 2009.

[13] R. S. DeFries, T. Rudel, M. Uriarte, and M. Hansen, “Deforestation driven by urban population growth and agricultural trade in the twenty-first century,” *Nature Geoscience*, vol. 3, no. 3, p. 178, 2010.

[14] K. C. Seto, A. Reenberg, C. G. Boone, M. Fragkias, D. Haase, T. Langanke, P. Marcotullio, D. K. Munroe, B. Olah, and D. Simon, “Urban land teleconnections and sustainability,” *Proceedings of the National Academy of Sciences*, vol. 109, no. 20, pp. 7687–7692, 2012.

[15] D. R. Richards and D. A. Friess, “Rates and drivers of mangrove deforestation in southeast asia, 2000–2012,” *Proceedings of the National Academy of Sciences*, vol. 113, no. 2, pp. 344–349, 2016.

[16] A. Paxton, *The Food Miles Report: the Dangers of Long Distance Food Transport.* London: SAFE Alliance, 1994.

[17] A. Smith, P. Watkiss, G. Tweddle, A. McKinnon, M. Browne, A. Hunt, C. Treleven, C. Nash, and S. Cross, *The Validity of Food Miles as an Indicator of Sustainable Development.* DEFRA report, 2005.

[18] K. R. Chi, J. MacGregor, and R. King, “Fair food miles: Recharting the food miles map,” *Oxfam Policy and Practice: Agriculture, Food and Land*, vol. 9, no. 7, pp. 17–64, 2009.

[19] E. Schlich, “The ecology of scale: Assessment of regional energy turnover and comparison with global food (2 pp.),” *The International Journal of Life Cycle Assessment*, vol. 10, no. 3, pp. 171–172, 2005.

[20] E. Schlich, “Energy economics and the ecology of scale in the food business,” *New Research on Energy Economics*, 2008.

[21] D. Coley, M. Howard, and M. Winter, “Local food, food miles and carbon emissions: A comparison of farm shop and mass distribution approaches,” *Food policy*, vol. 34, no. 2, pp. 150–155, 2009.

[22] D. Coley, M. Howard, and M. Winter, “Food miles: time for a re-think?,” *British Food Journal*, vol. 113, no. 7, pp. 919–934, 2011.

[23] V. Sandström, H. Valin, T. Krisztin, P. Havlík, M. Herrero, and T. Kastner, “The role of trade in the greenhouse gas footprints of eu diets,” *Global food security*, vol. 19, pp. 48–55, 2018.

[24] F. Schierhorn, T. Kastner, T. Kuemmerle, P. Meyfroidt, I. Kurganova, A. V. Prishchepov, K.-H. Erb, R. A. Houghton, and D. Müller, “Large greenhouse gas savings due to changes in the post-soviet food systems,” *Environmental Research Letters*, vol. 14, no. 6, p. 065009, 2019.

[25] T. Kastner, M. Kastner, and S. Nonhebel, “Tracing distant environmental impacts of agricultural products from a consumer perspective,” *Ecological Economics*, vol. 70, no. 6, pp. 1032–1040, 2011.

[26] T. Kastner, A. Schaffartzik, N. Eisenmenger, K.-H. Erb, H. Haberl, and F. Krausmann, “Cropland area embodied in international trade: Contradictory results from different approaches,” *Ecological Economics*, vol. 104, pp. 140–144, 2014.

[27] United Nations, “Un comtrade: International trade statistics database,” 2019.

[28] United Nations Department of Economic and Social Affairs, “International merchandise trade statistics: Concepts and definitions,” 2010.

[29] United Nations International Trade Statistics, “Harmonized commodity description and coding systems (hs),” 2017.

[30] V. Markhonko, “Asymmetries in official international trade statistics and analysis of globalization,” *Aguascalientes, Mexico*, 2014.

[31] M. Gehlhar, “Reconciling bilateral trade data for use in gtap,” *GTAP Technical Papers*, p. 11, 1996.

[32] G. Korn and T. Korn, *Mathematical handbook for scientists and engineers: Definitions, theorems, and formulas for reference and review.* New York: Dover Publications,

Inc., 2000 [1922].

[33] http://download.geonames.org, 2019.

[34] J.-k. Huang, W. Wei, C. Qi, and X. Wei, "The prospects for china's food security and imports: Will china starve the world via imports?," *Journal of integrative agriculture*, vol. 16, no. 12, pp. 2933–2944, 2017.

[35] European Commission, "Russian import ban on eu products," 2014.

**NOTE DI LAVORO DELLA FONDAZIONE ENI ENRICO MATTEI**
**Fondazione Eni Enrico Mattei Working Paper Series**

Our Working Papers are available on the Internet at the following addresses:
http://www.feem.it/getpage.aspx?id=73&sez=Publications&padre=20&tab=1

**NOTE DI LAVORO PUBLISHED IN 2020**


1. 2020, FACTS Series, Alessandra Celani de Macedo, Nicola Cantore, Laura Barbier, Marco Matteini, Giorgia Pasqualetto, The Impact of Industrial Energy Efficiency on Economic and Social Indicators
2. 2020, 2030 Agenda Series, Nicola Comincioli, Sergio Vergalli, Effects of Carbon Tax on Electricity Price Volatility: Empirical Evidences from the Australian Market
3. 2020, 2030 Agenda Series, Marco Buso, Cesare Dosi, Michele Moretto, Do Exit Options Increase the Value-For-Money of Public-Private Partnerships?
4. 2020, FACTS Series, Ricardo Nieva, A Tragic Solution to the Collective Action Problem: Implications for Corruption, Conflict and Inequality
5. 2020, FACTS Series, Charles Fang Chin Cheng, Nicola Cantore, The Inclusive and Sustainable Development Index: a Data Envelopment Analysis Approach
6. 2020, 2030 Agenda Series, FACTS Series, Coker Eric, Cavalli Laura, Fabrizi Enrico, Guastella Gianni, Lippo Enrico, Parisi Maria Laura, Pontarollo Nicola, Rizzati Massimiliano, Varacca Alessandro, Vergalli Sergio, The Effects of Air Pollution on COVID-19 Related Mortality in Northern Italy
7. 2020, 2030 Agenda Series, Laura Cavalli, Giulia Lizzi, Port of the future : Addressing Efficiency and Sustainability at the Port of Livorno with 5G
8. 2020, FACTS Series, Federica Cappelli, Gianni Guastella, Stefano Pareglio, Institutional Fragmentation and Urbanisation in the EU Cities
9. 2020, FEP Series, Giacomo Falchetta, Nicolò Stevanato, Magda Moner-Girona, Davide Mazzoni, Emanuela Colombo and Manfred Hafner, M-LED: Multi-sectoral Latent Electricity Demand Assessment for Energy Access Planning
10. 2020, Local Projects Series, Marcella De Filippo, Annalisa Percoco, Angela Voce, Covid-19 e didattica a distanza.Il caso Basilicata, una regione a rischio digital divide
11. 2020, 2030 Agenda, Laura Cavalli, Sandro Sanna, Mia Alibegovic, Filippo Arras, Gianluca Cocco, Luca Farnia, Emanuela Manca, Luisa F. Mulas, Marco Onnis, Sandro Ortu, Ilenia G. Romani, Marta Testa, The Contribution of the European Cohesion Policy to the 2030 Agenda: an Application to the Autonomous Region of Sardinia
12. 2020, FACTS Series, Alexander Golub, Kristina Govorukha, Philip Mayer, Dirk Rübbelke, How does Climate Change Affect the Transition of Power Systems: the Case of Germany
13. 2020, FEP Series, Rossana Scita, Pier Paolo Raimondi and Michel Noussan, Green Hydrogen: the Holy Grail of Decarbonisation? An Analysis of the Technical and Geopolitical Implications of the Future Hydrogen Economy
14. 2020, FACTS Series, Marta Montinaro, Rupayan Pal, Marcella Scimitore, Per Unit and Ad Valorem Royalties in a Patent Licensing Game
15. 2020, 2030 Agenda, S. Quaini, S. Saccani, S. Vergalli, L. Assom, M. Beria, A. Codello, M. Monaco, R. Sabatini, Seasonality Fingerprint on Global Trading of Food-commodities. A Data-mining Approach

**Fondazione Eni Enrico Mattei**
Corso Magenta 63, Milano – Italia

Tel. +39 02.520.36934
Fax. +39.02.520.36946

E-mail: letter@feem.it
**www.feem.it**