Resources, Conservation & Recycling 186 (2022) 106526

Contents lists available at ScienceDirect

Resources, Conservation & Recycling

journal homepage: www.elsevier.com/locate/resconrec

Full length article

Advancing UN Comtrade for Physical Trade Flow Analysis: Review of Data Quality Issues and Solutions

Chuke Chen [a,b,c], Zhihan Jiang [a,b], Nan Li [a,b,c,*], Heming Wang [d], Peng Wang [a,b], Zhihe Zhang [d], Chao Zhang [e,f], Fengmei Ma [a,b], Yuanyi Huang [a,g], Xiaohui Lu [a,h], Jianlimin Wei [a,b,c], Jianchuan Qi [i], Wei-Qiang Chen [a,b,c,*]

[a] Key Lab of Urban Environment and Health, Institute of Urban Environment, Chinese Academy of Sciences, Xiamen 361021, P. R. China
[b] Xiamen Key Lab of Urban Metabolism, Xiamen 361021, P. R. China
[c] University of Chinese Academy of Sciences, Beijing 100049, P. R. China
[d] State Environmental Protection Key Laboratory of Eco-Industry, Northeastern University, Shenyang 110819, P. R. China
[e] School of Economics and Management, Tongji University, Shanghai 200092, P. R. China
[f] UN Environment-Tongji Institute of Environment for Sustainable Development, Shanghai 200092, P. R. China
[g] College of Civil and Transportation Engineering, Shenzhen University, Shenzhen 518060, P. R. China
[h] Nottingham University Business School (NUBS), University of Nottingham Ningbo China, Ningbo, 315100, P.R. China
[i] Ganjiang Innovation Academy, Chinese Academy of Sciences, Ganzhou 341119, P. R. China

ARTICLE INFO

Keywords:
UN Comtrade
Physical trade
Outlier
Missing value
Bilateral asymmetry
Material flow analysis database

ABSTRACT

International trade has been considered a critical driving force of material flows and their environmental pressures, which has been a global research hotspot. The United Nations Commodity Trade Statistics Database (UN Comtrade) is the original and probably the most widely-used data source to support the physical trade analysis. However, data discrepancies have been discovered in UN Comtrade, which may lead to diametrically conflicted conclusions if not properly addressed. To promote applications of UN Comtrade, this article reviews data statistics criteria and preprocessing procedures, discusses three main data quality issues (outliers, missing values, and bilateral asymmetries), and reviews methods to explore adequate options. It is revealed that data quality issues existed in data of almost all the commodities, reporters, and periods, but existing methods are subject to certain limitations. Furthermore, this article presents a brief introduction of our following work on addressing these issues.

1. Introduction

Globalization has revolutionized the commodity trade by promoting the geospatial separation and redistribution of production and consumption activities worldwide. Along with boosting global economic development, trade has also been driving significant cross-border transfers of pressures on natural resources and the environment. Scholars have been devoted to evaluating the effects of commodity trade on environmental issues, such as water use (Yang et al., 2006), climate change (Lewis and Witham, 2012), pollutant emissions (Li and Hewitt, 2008), land use (Taherzadeh and Caro, 2019), raw material extractions (Schaffartzik et al., 2014), biodiversity loss (Lenzen et al., 2012), etc. The analysis and discussion over these results have provided substantial support for global environmental governance.

In this research hotspot, the accessibility, completeness, and reliability of commodity trade data are crucial. The United Nations Commodity Trade Statistics Database (UN Comtrade, https://comtrade.un.org) is an original and one of the most widely used commodity trade data sources in this field. UN Comtrade plays an important role in many aspects, such as (1) establishing trade networks, (2) building trade-related databases, and (3) conducting material flow analysis (MFA). For trade networks, Gephart and Pace (2015) established a global seafood trade network with UN Comtrade to describe the structure and evolution of the global seafood trade. Wang et al.(2020) and Shi et al. (2021) established global plastic waste trade networks using UN Comtrade to interpret global plastic waste trade and investigate the influence of the Chinese import ban on plastic. Moreover, UN Comtrade has been a reliable data source for constructing trade-related databases. For

* Corresponding authors.
E-mail addresses: nanli@iue.ac.cn (N. Li), wqchen@iue.ac.cn (W.-Q. Chen).

https://doi.org/10.1016/j.resconrec.2022.106526
Received 15 March 2022; Received in revised form 1 June 2022; Accepted 3 July 2022
Available online 15 July 2022

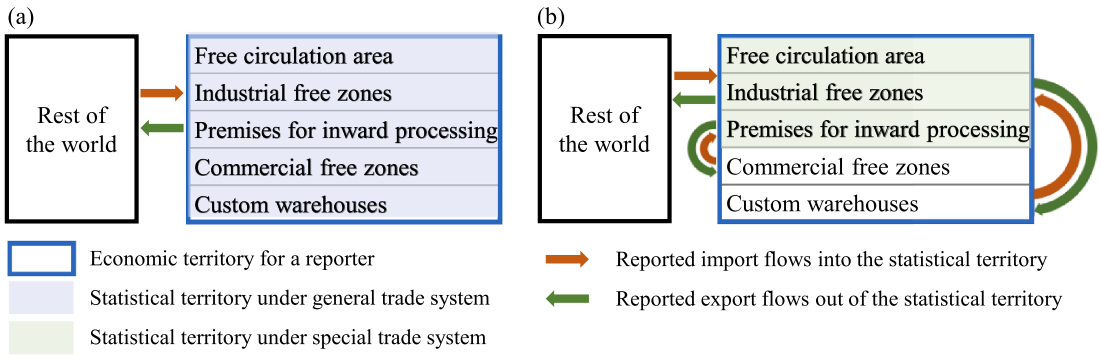

Fig. 1. Territorial elements and potential imports and exports under (a) the general trade system and (b) the special trade system (United Nations, 2010).

example, the Global Trade Analysis Project (GTAP) database established by Aguiar et al. (2019) and the World Input-Output Table (WIOT) built by Dietzenbacher et al. (2013), which are widely used as multi-regional input-output (MRIO) tables, acknowledge UN Comtrade as their critical data sources. UN Comtrade has also been used to uncover cross-border material flows. Qu et al. (2019) interpreted UN Comtrade data of waste plastics, unsorted scrap papers, discarded textile materials, and vanadium slags to understand their global flows, and thus revealed the profound effects of policy changes on the global circular economy of solid waste. Mohammadi et al. (2021) conducted MFA based on UN Comtrade, estimating the flows and stocks for 206 product types related to e-waste. The MFA knowledge of, such as types of uses and international trade, could contribute to the assessment of material criticality, and the evaluation of resources sustainability (Chen et al., 2019; Chen and Graedel, 2012).

UN Comtrade is well-acknowledged for the broad coverage, which is attributed to its establishing mechanism. It relies on the original data provided by all the involved countries or areas (i.e., the reporters) rather than mathematical approaches, which would ensure the consistency of the data from each reporter. However, this also causes discrepancies. The most critical issue is that reporters have different statistic criteria, which are the standards and processes of gathering observations or measurements. The mismatch of statistical criteria may result in different definitions or understandings of the data among reporters. For instance, data provided by a reporter using the general trade system may include the transaction that occurred in the free zone, while data provided by one applying the special system may not. To solve these problems, UN Comtrade has set up a preprocessing mechanism and, in most cases, the published data have already been preprocessed.

However, there still exist concerns in the practical application of UN Comtrade data. Data discrepancies can still be identified, making scholars have to further process data for their studies. For example, Dittrich and Bringezu (2010) found that the UN Comtrade data is insufficient for determining the shipping volume of international trade because the net weight data were incomplete. It results in difficulties in calculating physical metrics (e.g., domestic material consumption, resource productivity) when measuring and monitoring material flows. Besides, (Gaulier et al., 2008) reported unrealistic observations in UN Comtrade, which have large impacts on variations of the average and the median of unit value. It might particularly result in substantial biases in representing price change (e.g., price indices, unit value indices) and thus a misleading economic analysis (Silver, 2007). Moreover, some studies found huge differences in both monetary and physical values between mirror trade statistics. These deviations would create enormous uncertainties or even lead to diametrically conflicted conclusions in characterizing trade patterns of countries/areas (Ferrantino et al., 2012; Hamanaka, 2011), understanding the stance of macroeconomic policies and structural conditions (Robert C. Feenstra et al., 1998), etc.

The data quality issues in UN Comtrade identified above can be categorized into three types: outliers, missing values, and bilateral asymmetries. (1) Outliers refer to the sampling points that lie outside the main body of the data group, which may cause significant deviations in the results (Narins, 2018; Yeats, 2012). (2) Missing values refer to the incomplete data values, which are mostly in the physical commodity trade dataset. The incompleteness of UN Comtrade may impede its applications in certain contexts, and the calculation with missing values may be unreliable. (3) Bilateral asymmetries mean that importers' records do not match the exporters' records in the same transaction. In some cases, the results, calculated from either import or export data, might be considerably different or even opposite (Bahmani-Oskooee et al., 2013).

These issues limit the applications of UN Comtrade, making it necessary to comprehensively review the status quo, causes, and existing solutions to these issues. Presenting this critical information, this review (1) gives a whole picture of this widely-used database to first-time users and (2) gives insights into uncertainties of related studies to intermediate or advanced users. Furthermore, we can seek prospective solutions based on previous efforts to improve the data quality of UN Comtrade, thus promoting its applications in physical trade analyses considering environmental impacts.

In this review article, we first introduce critical details of UN Comtrade data that may cause data discrepancies, including the reporters' different statistic criteria and the preprocessing procedures by the UN Statistics Division (UNSD). Then we investigate the three major data quality issues (i.e., outliers, missing values, and bilateral asymmetries) by revealing the situations, causes, and potential consequences of these issues. Finally, we review and evaluate the methods used in previous studies to solve these issues and discuss feasible solutions to promote the data quality of UN Comtrade. This article is the first in a three-part series presenting our work on addressing data quality issues of UN Comtrade. In the following articles, we will present our efforts in processing outliers and missing values, respectively. With these issues addressed, the improved dataset can be accessed online via https://www.macycle.org/improved-un-comtrade-data/.

2. Critical Details of UN Comtrade Data

UN Comtrade contains a repository of global trade data, including the year, value, quantity, and net weight of imported or exported commodities. The data are reported by 209 reporters' statistics authorities annually since 1962 and monthly since 2000. As reporters adopt different statistic criteria, the UNSD unifies and standardizes the reported trading data before publishing. This section reveals the critical details of UN Comtrade data, which may cause statistical discrepancies and thus jeopardize the data quality. We first describe different statistic criteria and uncover the potential consequences of the criteria selection. Then we introduce the preprocessing methods by the UNSD.

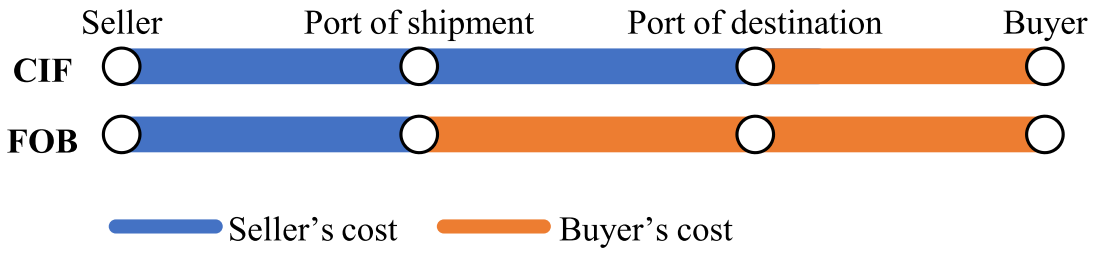

Fig. 2. The simplified profile of tasks and costs associated with the transportation and delivery of goods for FOB and CIF proposed by the International Chamber of Commerce (Montezuma, 2021).

2.1. Statistic Criteria of Reporters

Reporters use different statistic criteria for reporting data to the UNSD. That is, they gather observations or measurements with different standards and systematic processes. Reporters adopt different statistic criteria in the trade system, monetary valuation, partner attribution, and commodity classification system.

2.1.1. Trade System

Reporters report commodity transaction records, and whether a transaction is reported depends on the trade system that a reporter adopts. There are two types of trade systems, the general and the special trade systems. The difference between them lies in their statistical territory where the trade happens. As shown in Fig. 1, the statistical territory under the general trade system coincides with the economic territory, while that under the special trade system comprises only a particular part. For example, the trade that occurred in commercial free zones and custom warehouses may not be included in the records by reporters applying the special trade system but is included in reporters adopting the general trade system. The difference might result in an imbalance between import and export values.

As shown in the metadata given by UN Comtrade, 79 reporters (e.g., Australia, Canada, Hongkong of China, and the United States) adopt the general trade system, while 38 (e.g., Austria, Brazil, Colombia, France, and Germany) use the special trade system. 87 reporters use different trade systems in different years, and the selected trade systems of 5 reporters are not given by the UNSD (Jordan, Nepal, Anguilla, Togo, and Bosnia Herzegovina). Moreover, it is noteworthy that Estonia and Indonesia adopt different trade systems simultaneously. Estonia uses the special trade system for trade flow among European Union (EU) countries and the general trade system for extra-EU trade flow. Indonesia uses the general trade system for exports and the special trade system for imports.

2.1.2. Monetary Valuation

The difference in monetary valuation of commodity trade is caused by different international shipping agreements. In most cases, the import trade value is evaluated based on the Cost, Insurance, and Freight (CIF) agreement, while the export trade value with the Free/Freight on Board (FOB). The difference between CIF and FOB is presented in Fig. 2. For exporters (or sellers), the CIF trade value involves the costs during the process from the exporters' locations to the ports of destination (where goods are discharged in another country), while that for importers (or buyers) involves the costs from the ports of destination to the importers' locations. In contrast, for the FOB agreement, the boundary of costs between the exporters and importers is at the port of shipment (where the goods are loaded onboard the ship). Considering the involved process, it is evident that there may exist a difference in trade values between importer's and exporter's statistics. Usually, the CIF-type values of imports tend to be higher than the FOB-type values of exports (UN Comtrade, 2009).

Noting that CIF and FOB apply only when goods are transported by sea or inland waterway, which means they are not applicable for other means of transport (e.g., by railroad). According to the United Nations (2010), there are substitutive valuation methods for non-waterborne transportation. Carriage and Insurance Paid to (CIP) can substitute CIF for delivery by any carrier except ships. To substitute FOB, Free Carrier (FCA) and Delivered at Frontier (DAF) can be used, which prices goods exported, for example, by car or airplane, and by railroad or pipeline, respectively. Since the cost accounting mechanisms of CIF and CIP are almost the same, they are both referred to as CIF-type for estimating import trade values. For the same reason, FOB, FCA, and DAF are referred to as FOB-type for exports.

However, there are exceptions. Australia, Canada, and South Africa always use FOB-type when valuating import goods. Moreover, 28 reporters use different agreements in all these years. 6 reporters (e.g., Algeria and Peru) have used CIF-type to report export trade value since 2000. Cambodia even adopted a different method named "Other," which remains unknown since the UNSD has not given an explanation. Besides, 23 reporters (e.g., Brazil, Mexico, Poland) have used FOB-type to report import trade value. These special cases make the bias even higher between the values of imports and exports. The complete list of special cases is given in Fig. A1.

2.1.3. Partner Attribution

One of the critical attributes in the reported commodity trade data is the partner. However, reporters define partners differently. For a reporter's import trade partners, the most often used attribution is the "origin," representing the place where the goods originate. For a reporter's export trade partners, the most frequently used attribution is the "last known destination," which is the final destination country that the exporters know at the time of exportation. Ideally, for a single transaction, if the importer defines the exporter as the origin and the exporter defines the importer as the last known destination, the record of this transaction will match on both sides. For example, we assume that the definitions of partners adopted by Chile and China follow the situation above, and Chile exports 10 tons of salmon fish to China via Japan. In this case, China will report an import of 10 tons of salmon fish from Chile, while Chile will report an export of 10 tons of salmon fish to China.

However, some reporters define partners with different attributions. Other definitions for a reporter's partners of imports include "country of purchase," "country of consignment," and "origin/consignment for intra-EU." The country of purchase is the country where the seller of the goods resides, and the country of consignment is the country where the goods pass through. The origin/consignment for intra-EU means that the reporters define partners as the origin for extra-EU trade and the consignment for Intra-EU trade (EUROSTAT, 2007). Meanwhile, other definitions for a reporter's partners of exports include "country of sale," "country consignment," and "country of consumption." The country of sale is the country where the purchaser of the goods resides, and the country of consignment is also the passing-through country. The country of consumption is the country in which the goods are expected to be used for private/public consumption or as intermediate inputs of the

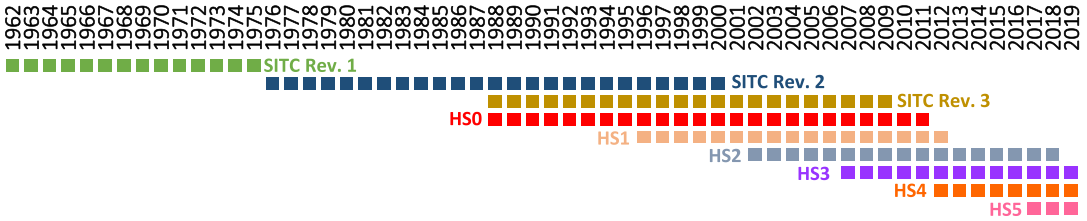

Fig. 3. Classification systems adopted by reporters over the years.

production process (United Nations, 2010). This discordance can cause significant discrepancies. For example, we suppose Chile exports 10 tons of salmon fish to China, passing through Japan. China uses origin to define exporters, while Chile defines consignment as importers. In this case, Chile would record an export of 10 tons of salmon fish to Japan, while China would report 10 tons of salmon fish imported from Chile. This would cause a difference in the statistic between China and Chile. From 1962 to 2019, 79 and 32 reporters did not always use "origin" for their imports and "last known destination" for their exports. For example, Poland used 'country of purchase' before 1991 and 'origin' after to attribute import partners, while for their export partners, Poland used 'country of sale' before 1991 and 'last known destination' after. Complete details are provided in Fig. A2and Fig. A3. For the trade that these reporters are involved in, the statistic they report may not be compatible with their partners' reports, which may cause significant statistical discrepancies.

2.1.4. Commodity Classification System

Three commodity classifications are used in UN Comtrade, which are the Harmonized System (HS), Standard International Trade Classification (SITC), and Broad Economic Categories (BEC). The HS has six versions, HS0, HS1, HS2, HS3, HS4, and HS5, published in 1992, 1996, 2002, 2007, 2012, and 2017, respectively. The SITC, maintained by the UN, has four versions (SITC Rev.1-4) which were respectively updated in 1951, 1974, 1988, and 2004 (United Nations, 2006). Take SITC Rev.4 as an example. There are ten sections, 67 divisions, 262 groups, 1023 subgroups, and 2970 headings or items. The commodity code is a 4- or 5-digit number, such as 001.5 and 012.11. For 001.5, "001" represents the group with section 0 and division 01, while "5" is the sub-group with no basic headings (United Nations, 2006). The BEC is a 3-digit classification, which groups transportable goods according to the end-uses and main outputs of corresponding industries.

The remarkable thing is that in one year, each reporter report data with only one commodity classification, but different reporters do not necessarily adopt the same commodity classification. In all these years, reporters have used nine classifications mentioned above, excluding SITC Rev.4 and BEC. The reason for not using SITC Rev.4 could be that this version was intended as an analytical tool to analyze trade data in HS. The use of HS was officially endorsed in compilating and disseminating international trade statistics in 1993, and since then reporters using SITC were less. Notably, after 2010, SITC was no longer adopted by any reporters, as shown in Fig. 3. However, the trade data in SITC dating back to the year 1951 could enable a long-term analysis, so to ensure strict period-to-period comparability, in 2004, the SITC Rev.4 was prepared to harmonize the SITC and HS as much as possible (United Nations, 2006). As for the BEC, this classification was an aggregated classification of the SITC to facilitate broad economic analyses of trade and production (United Nations, 2018). This indicates that its classification is too rough for customs to supervise each type of merchandise and goods (e.g., checking the customs duty) and may not be suggested by the statistical authorities. Hence, SITC Rev.4 and BEC were not adopted by any countries/areas to report commodities. More importantly, as Fig. 3 shows, there often exist multiple classifications in one year, e.g., SITC Rev.2, SITC Rev.3, HS0, and HS1 were adopted in 1998. Besides, these classifications vary greatly. For example, SITC Rev.2 code 00141 is equal to HS0 code 010511 plus code 010519; HS0 codes 382471, 382479, and 382490 are merged into HS2 code 382490. More examples can be found in the official report (United Nations, 2017). It is difficult to precisely decompose SITC Rev.2 code 00141 into HS0 code 010511 and 010519, so different classifications might result in data discrepancies in statistics (Chen et al., 2019).

2.2. Preprocessing by the UNSD

To deal with these issues above, the UNSD processed the original data from reporters before publishing. As reported by the UNSD, the preprocessing procedures include the conversion of classifications and the conversion/estimation of missing values and units. This section will discuss the general preprocessing procedures by the UNSD.

2.2.1. Classification Conversion

For commodity trade analysis, commodity classification is the most likely source of uncertainties. Thus, the UNSD has made efforts to harmonize the classifications of UN Comtrade. The data reported in the latest version are converted into earlier versions with the conversion system designed by UNSD. Data by 6-digit HS classification codes are converted first. Then data by 4-digit and 2-digit codes are aggregated with the converted data with a 6-digit code. For example, all the data reported by the 6-digit classification codes of HS 17 is converted into those of HS12, HS07, etc., respectively. Subsequently, the data by 4-digit codes in earlier versions is the sum of data with all the corresponding 6-digit codes, and the data by 2-digit codes is the sum of data with all the corresponding 4-digit codes.

Reportedly, this conversion procedure may induce several problems (UNSD_KH, 2021). First, it may break the temporal continuity of some commodities, i.e., statistics of some codes are available earlier but are unavailable in recent years, which is attributed to the changes in commodity classification systems. For example, to convert HS3 code 010594 to HS2, the statistics of 010594 would be directly and solely correlated to HS2 code 010592. However, according to the commodity classification system, 010594 in HS3 is the aggregate of 010592 and 010593 in HS2. This conversion will result in the absence of the statistics of 010593 in HS2 during the years using HS3 to report, and the statistics of 010592 in those years would be larger than the actual. Second, it would cause the loss of confidential information. Some reporters only report data at the 2-digit classification level due to confidentiality. Nevertheless, since the conversion is conducted only at the 6-digit level and the converted 2-digit data was the sum of converted 6-digit data, the converted 2-digit data would not contain confidential information. Third, it would cause the loss of net weight data at higher classification levels. When data by the 6-digit classification codes are summed up to higher classification levels, the net weight data cannot be aggregated due to unit differences.

To reduce the effects of all these issues, UN Comtrade also provides the original data from each reporter. Users can retrieve "HS as reported" or "SITC as reported" data, which are the raw data provided by each reporter. The UNSD also offers additional metadata, including the version of the original classification, original currency, etc. Combining both, users could design their own preprocessing methods for particular

Table 1
World Customs Organization (WCO) Recommended Units and Examples

Index	WCO Abbreviation	Description	Examples (in HS0 6-digit code)
1	-	No quantity	Ivory (050710)
2	m^2	Area in square meters	Carpets and other textile floor coverings (570210)
3	1000 kWh	Thousands of kilowatt-hours	Electrical energy (271600)
4	m	Length in meters	Photographic film (370252)
5	U	Number of items	Horses; live, pure-bred breeding animals (010111)
6	2u	Number of pairs	Sports footwear; tennis shoes, etc. (640411)
7	l	Volume in liters	Wine; sparkling (220410)
8	kg	Weight in kilograms	Slate (251400)
9	1000u	Thousands of items	Ceramic building bricks (690410)
10	U(jeu/pack)	Number of packages	Playing cards (950440)
11	12u	Dozens of items	Eggs (040700)
12	m^3	Volume in cubic meters	Wood (440320)
13	carat	Weight in carats	Diamonds (710210)

purposes. Besides, the raw data can also be used as officially reported data to explore the trading behaviors of a certain country.

2.2.2. Conversion/Estimation of Missing Values and Units

In practice, reporters may not submit complete data as required by UNSD. In particular, reporters pay more attention to trade value data, while data of quantity and net weight are sometimes neglected. Moreover, reporters may submit data of the same commodity in different quantity units. For the missing values of quantity and net weight, the UNSD has a standardized estimation procedure to fill the blanks. At the same time, for the unit reconciliation, UNSD would convert all the original reported data into the 13 types of units (Table 1) recommended by the World Customs Organization. This conversion/estimation was conducted using reporter- or commodity-specific mathematical or empirical conversion factors provided by the UN (United Nations, 2019, 2013). To inform the users of this estimation process, a "flag" is set with 0, 2, 4, and 6 representing "No estimation," "Quantity estimation only," "Net weight estimation only," and "Both," respectively.

The major concern of this conversion or estimation procedure is that it only uses reporters' quantity values to estimate net weight values, or uses net weight values to estimate quantity values. The trade values, however, are not considered. Given that some reporters may provide biased quantity values due to misreporting, the estimation of net weight values may have significant uncertainties, which may, for example, make the unit price of goods unrealistically large. In reality, the unit price of a certain good (i.e., trade value divided by quantity value) varies much less among transactions than that of the solely quantity value. For this concern, it is better to take trade value into account to enhance the reliability of the conversion/estimation process.

3. Data Quality Issues of UN Comtrade

Due to the statistical criteria selection and preprocessing procedures discussed above, three major data quality issues in UN Comtrade have been reported: outliers, missing values, and bilateral asymmetries. We uncover these data quality issues with quantitative analyses on 580,063,082 records retrieved from UN Comtrade. These records cover 5037 commodities based on HS0 (the first and most used classification) and cover the period of 1988-2019. The definitions, causes, and consequences of these issues are presented as follows.

3.1. Outliers

Outliers are the observations that differ significantly from others. For example, one of UN Comtrade records shows that 10,145,704,000,204,100 kilograms of hakes were imported by Eswatini in 2016, greater than the sum of all the other countries in that year. This number is too large to be normal, which might be caused by misreports by the reporters or incorrect estimation by the UNSD. It should be noticed that there is no strict boundary between outliers and normal data, so whether a data point is considered abnormal is closely associated with the detection criteria. For example, 2% of UN Comtrade data were considered outliers by a modified 3-sigma rule method (Benkovskis and Wörz, 2012), while 30% of UN Comtrade data were detected outliers according to percentile curves. Identified and reported in many previous studies, this issue is prevalent and could have a disproportionate effect on statistical results (e.g., the mean), resulting in misleading interpretations or analysis (Andrey A. Gnidchenko, 2018; Silver, 2007; Silver and Heravi, 2007).

3.2. Missing Values

Missing values in UN Comtrade, especially for the trade value, quantity, and net weight, may lead to significant inconvenience in data processing, calculation, and analysis. It is caused by either the reporters

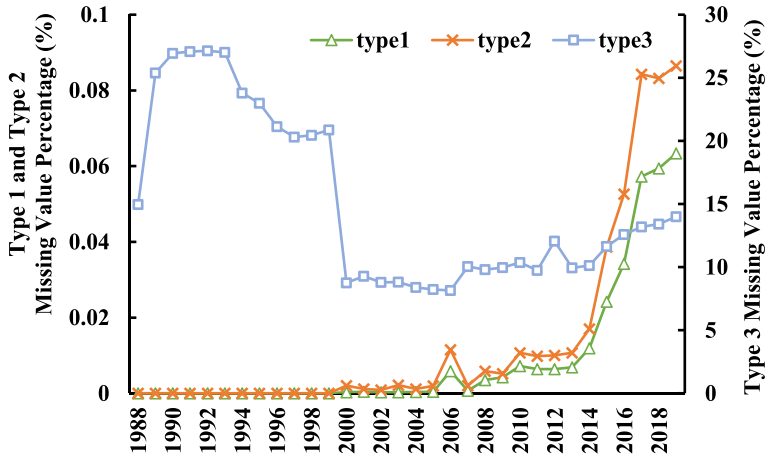

Fig. 4. The proportions of records with missing values by types in the total amount of UN Comtrade data from 1988-2019. Type 1 refers to records in which both trade value and net weight are missing. Type 2 represents records with only trade value missing. Type 3 indicates records with only net weight missing.

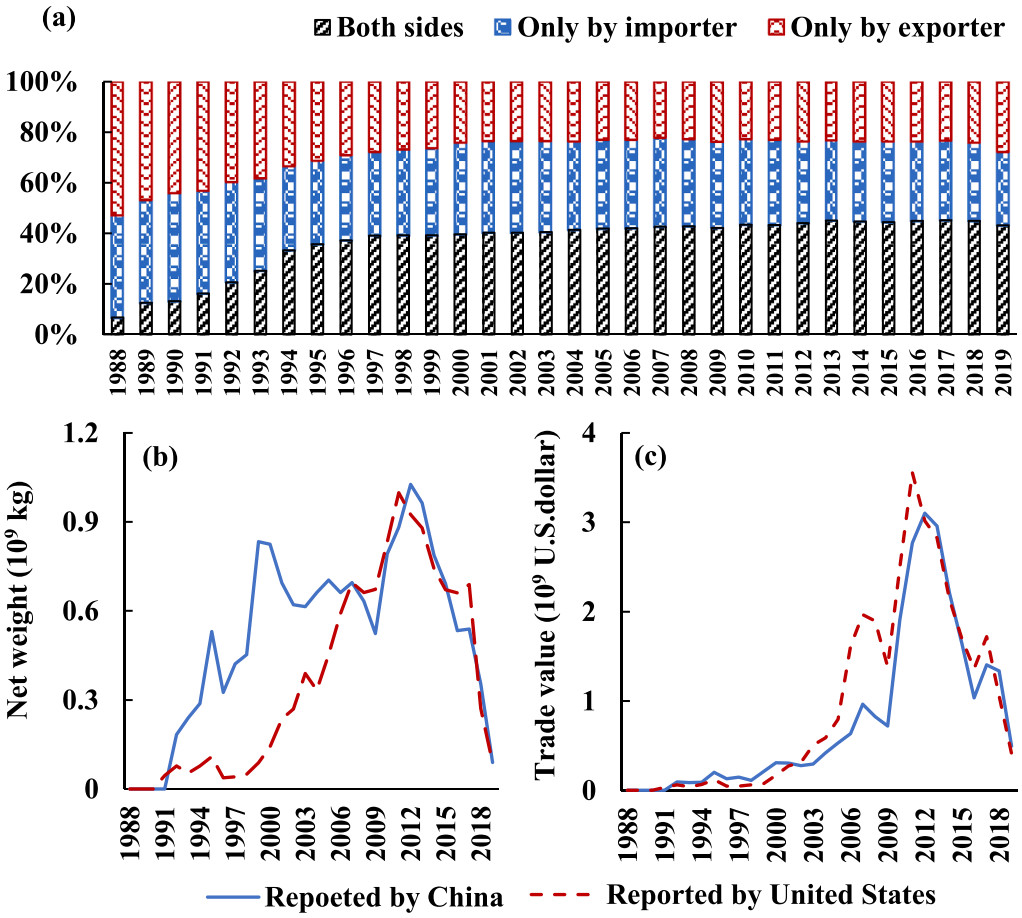

Fig. 5. Origins of bilateral trade data. (a) The trade data of the same commodity transaction reported by both sides, only by the importer, and only by the exporter. The net weight (b) and trade value (c) of the transaction of commodity 740400 (copper; waste and scrap) in HS0 between China and the United States.

or the UNSD. Reporters may fail to report data for the following reasons: (1) non-compliance of the units between reporters and the WCO; (2) confidentiality issues; (3) delays in data processing; (4) erroneous reporting (Farhan, 2015). Meanwhile, the UNSD may leave some data blank, or as missing in the preprocessing procedures, if these two conditions happened: (1) the unit given by reporters is challenging to convert to a WCO recommended unit; (2) the net weight is difficult to estimate with the trade value.

According to the statistical results, we categorized the records with missing values into three types and counted the number of records for each category: (1) both trade value and net weight are missing (45938 records, 0.013% of the total); (2) only trade value is missing (24937, 0.007%); (3) only net weight is missing (22709945, 6.5%). Noted that, for records of type 1, the transactions did happen, but the details (i.e., trade value or net weight) were not reported in the records ('0' or 'null'). There will be no record in UN Comtrade if the transaction never happened. As Fig. 4 shows, the proportions of records for type 1 and 2 in the total amount have increased during 1988-2019, while that of type 3 has decreased since 2000, indicating a better data coverage for the net weight. However, from the perspective of reporters and commodities, the quality of trade value data is relatively better than that of net weight in this period. During 1988-2019, 53.6% of the reporters and 88.2% of the commodities have records of type 2, while records of type 3 exist in all the reporters and commodities. The proportions of reporters and commodities with type 1 records are 53.1% and 83.2%, respectively.

3.3. Bilateral Asymmetries

In general, a commodity transaction is a completed agreement between a buyer and a seller. Thus, ideally, a transaction should be reported by both the importer and the exporter. However, only less than 50% of the commodity transactions (less than including net weight (Fig. 5b) and trade value (Fig. 5c), varied widely between both sides, which might be attributed to different statistical criteria mentioned in Section 2.1. Both incomplete data acquisition and inconsistent values result in bilateral asymmetries.

To quantify the bilateral asymmetry level, we introduced the bilateral asymmetric rate, which measures the degree of bilateral asymmetries of the commodity trade reported by both the importer and the exporter (equation 1). X represents the net weight value or the trade value of a transaction. In the transaction of commodity c between reporter i and reporter j in the year t, X_{ijtc} is calculated with data provided by reporter i while X_{jitc} is with data provided by reporter j. The rate ranges from 0 to 1. The closer the value to 1, the higher the asymmetric degree.

$$a_{ijtc} = \frac{\left|X_{ijtc} - X_{jitc}\right|}{X_{ijtc} + X_{jitc}} \quad (1)$$

Furthermore, we established the asymmetric index to measure the asymmetric degree of a specific commodity, which is computed by the average bilateral asymmetric rate of all the commodity transactions (equation 2).

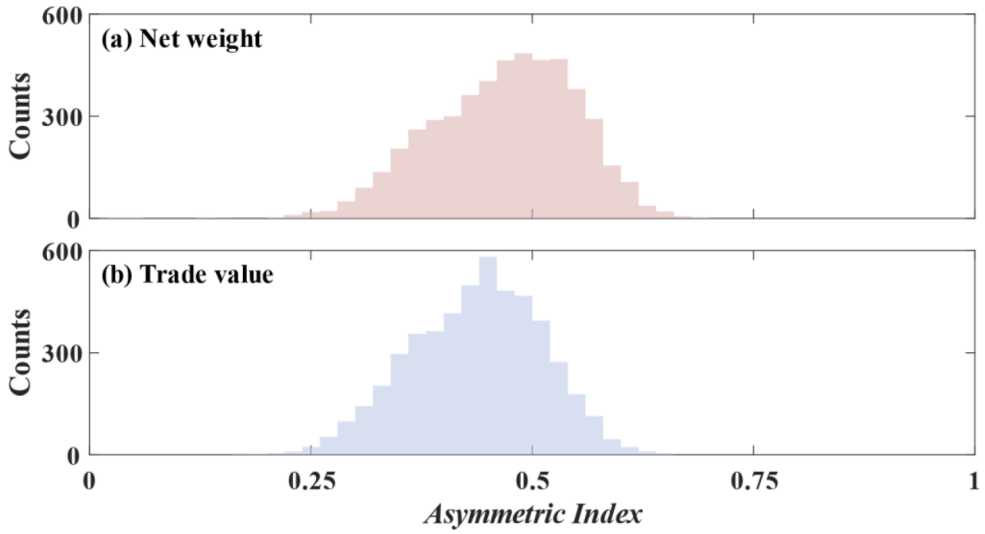

Fig. 6. The histogram of the *asymmetric index* of (a) net weight and (b) trade value.

$$asymmetric\ index_c = \frac{1}{n_c}\sum_{ijt} a_{ijtc} \quad (2)$$

Fig. 6 shows the histogram of the asymmetric index of net weight and trade value of all the commodities. The x-axis represents the *asymmetric index*, and the y-axis represents the number count of commodities' occurrences in each interval of the *asymmetric index*, graphically representing the distribution of the commodities' *asymmetric index*. As shown in Fig. 6, the *asymmetric index* is centered around 0.5 and 0.4 for net weight and trade value, respectively, indicating that of a certain transaction, net weight or trade value provided by one country would be 3 and 2.3 times higher or lower than that provided by the other country on average. This large discrepancy would cause serious results. For example, Bahmani-Oskooee et al. (2013) illustrated that drastically different results for the impact of exchange rate on trade could arise by choosing data from different transaction sides. Gehlhar (1996) demonstrated that this discrepancy would reduce the credibility of simulation exercises based on trade structure.

4. Existing Methods of Addressing Data Issues

For better UN Comtrade data applications, several attempts to address these data quality issues of outliers, missing values, and bilateral asymmetries have been made. This section intends to comprehensively review and evaluate the existing methods for each issue.

4.1. Outliers

Existing studies have identified outliers in trade value or net weight based on different outliers' definitions. Some scholars regarded values that are significantly different from adjacent years as outliers (Damerval, 2012; Saki et al., 2019), while others defined outliers as those whose unit prices are either too high or too low (Benkovskis and Wörz, 2013; T. Brewer et al., 2020; Giljum et al., 2014). Correspondingly, various methods for detecting and processing outliers have been developed.

For the time-series outliers, methods of time-series analysis have been adopted for detection. Saki et al. (2019) used the autoregressive integrated moving average models (ARIMA) to detect outliers to forecast the comparative advantage of the United States textile. This method is well-performed in forecasting by removing non-stationarity caused by outliers. Damerval (2012) tested three methods to detect outliers, including wavelets, Kalman filter, and forward search methods, to analyze the European Union's policies. However, these detecting methods are not suitable for UN Comtrade data because they might define values that contain valuable information as abnormalities. For example, detected time-series outliers might be caused by infrequent but real events, such as the drought (Collins, 1998), money laundering (Damerval, 2012), etc. In this way, outlier detection based on time-series analysis will introduce biases in subsequent analysis by directly removing these influential outliers from further analysis.

To detect the outliers with abnormal unit prices, methods of testing statistical properties have been applied. Benkovskis and Wörz (2012) conducted 3-sigma tests to detect outliers and remove them before calculating the real effective exchange rate. Brewer et al. (2020) used the boxplot method to identify outliers to build a food trade database. Nevertheless, these statistical methods have limitations in assumptions or parameter settings. For example, the 3-sigma rule is under the normality assumption, while the parameters used in the boxplot method are too rigid and uniform for most cases. As a result, some usual data might be identified as outliers. We tried different widely-used methods to detect outliers for all records in UN Comtrade. Detailed results (e.g., outlier distributions and deviations, quantitative comparative analyses) will be presented and thoroughly discussed in our next article of this series.

4.2. Missing Values

Three basic approaches have been used to estimate missing values, which are: (1) average global unit price; (2) bilateral data; (3) export price index (XPI), and import price index (MPI).

The first approach was firstly proposed by Dittrich and Bringezu (2010), who used the average global unit price (for a specific commodity in a certain year) to estimate missing net weight values. The average global unit price is calculated by the sum of all this commodity's existing net weight values that year divided by the sum of all this commodity's existing monetary data that year. This method assumes all the reporters make deals for this commodity at the same price, which is disputable for the following reasons. First, this method is sensitive to outliers, which would lead to a biased estimation of the average global unit price. Second, the unit price varies among reporters. For some commodities, like HS 711100 (Metals, clad with platinum, semi-manufactured), this unit price variance would be significant. In this case, this method will over/underestimate the missing values in net weight for the reporters that are trading at more expensive/lower prices within this commodity group.

The second method was used by Gaulier and Zignago (2010) for building the international trade database. They used the bilateral trade

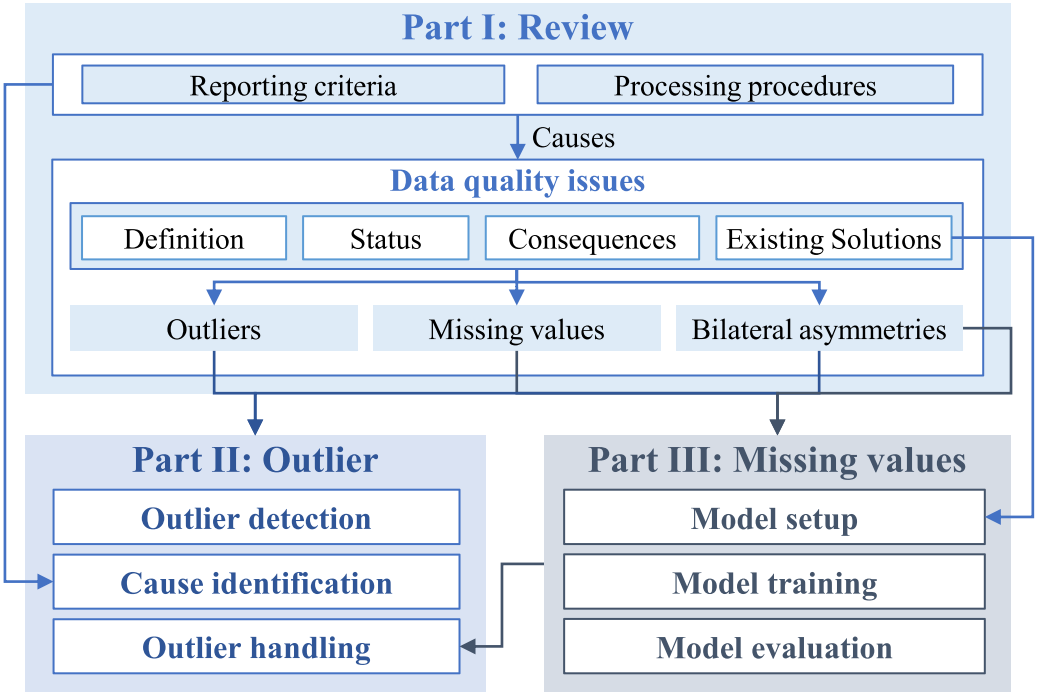

Fig. 7. The diagram of our work.

data to impute missing values. However, as not all bilateral data are reported by both the exporter and the importer, this method is not applicable for those reported only by one side. Besides, as mentioned in section 3.3, the partner's data is not a reasonable estimation of missing values as the bilateral trade data are often mismatched.

The third method was used by Ong et al. (2013) and Farhan (2015). In these two studies, they used stochastic regression and multiple imputation models to estimate missing net weight values, respectively, concerning XPI and MPI data. This approach's issue is that both MPI and XPI are only available for a small number of reporters. Its application would be limited to reporters whose MPI and XPI cannot be found.

We tried to estimate missing values of UN Comtrade by developing a model employing various statistical methods, which takes the difference in commodities, reporters, and years into account. Details (e.g., data preprocessing, model framework development) about this improved model and comparative analyses of the model performance will be presented and thoroughly discussed in the third article of our series.

4.3. *Bilateral Asymmetries*

Previous studies have tried to harmonize the bilateral asymmetries based on the "data quality" of the reporters or the causes of asymmetries (e.g., CIF-FOB ratio, etc.).

Gehlhar et al. (2010), de Saint Vaulry (2008), and Shaar (2019) introduced an indicator (the quality index) to assess the reliability of each reporter and assumed the reporters with a higher quality index would provide more reliable data than those with a lower quality index. Then, for the bilateral asymmetric trade data, they put higher weights on the reporters with higher data quality. For example, if the United States has a data quality index of 3 while Japan's is 2, the harmonized trade data of both imports and exports will be the sum of 60% of the United States' value and 40% of Japan's. However, the indicator is subjective and straightforward, reflecting only a partial picture of reality. Our calculation shows that some developed countries have much lower quality indexes than developing countries, which is beyond expectation. For example, the data of Israel is far worse than that of Moldova according to the quality index, but developed countries are more likely to have better data quality. Moreover, the data quality of Australia is almost one of the lowest among all reporters. To adjust the asymmetries, indicator-based methods focus on reconciling with the weighted average of bilateral data. These methods aim to establish a balance between data of both sides and ignore the difference in the causes of bilateral asymmetries, such as the existence of indirect trade and various partner attributions (Shaar, 2019).

Beyond these, previous studies also tried to deal with this issue from the causes of asymmetries. For example, the UNSD (United Nations, 2013) harmonizes the data by setting up a model to address the issues that could lead to bilateral asymmetries, such as the trade system, monetary valuation, partner attribution, etc. However, an important factor, the indirect trade, is neglected in this case. Indirect trade means the goods are transported through a third reporter. For example, the United States exports goods to mainland China via Hong Kong, China. If mainland China does not know the original exporter of these goods, the reported partner would be Hong Kong of China, while the partner reported by the United States would be mainland China. Many studies regard this as the most crucial factor for bilateral asymmetries (Dong Guo, 2010; Marko Javorsek, 2016). However, the identification of indirect flows is challenging, which requires information at a lower level instead of nation-level data (Adrian Chesson, 2018). For example, with obtained information from monthly enterprise surveys, the intermediary role of Hong Kong SAR, China was identified in bilateral relationships between mainland China and the United States (United Nations, 2013). Furthermore, the adjustment of trade statistics between two countries should consider the specific role of each intermediary party (e.g., outward processing, transiting). For example, in 2009, China and the United States governments adjusted the imbalance by calculating the added value of indirect shipments of processed goods, given Hong Kong's involvement in outward processing (U.S.-China JCCT, 2009). These studies were conducted with the collaborative efforts of national governments, which signed off on the exchange of trade data at the firm level. However, the data are not published due to confidentiality. Due to the unavailability of data, it is impossible to develop a general approach to analyzing all bilateral asymmetries. For now, the appropriate solution is to take a deep-dive analysis for each case under the collaborative

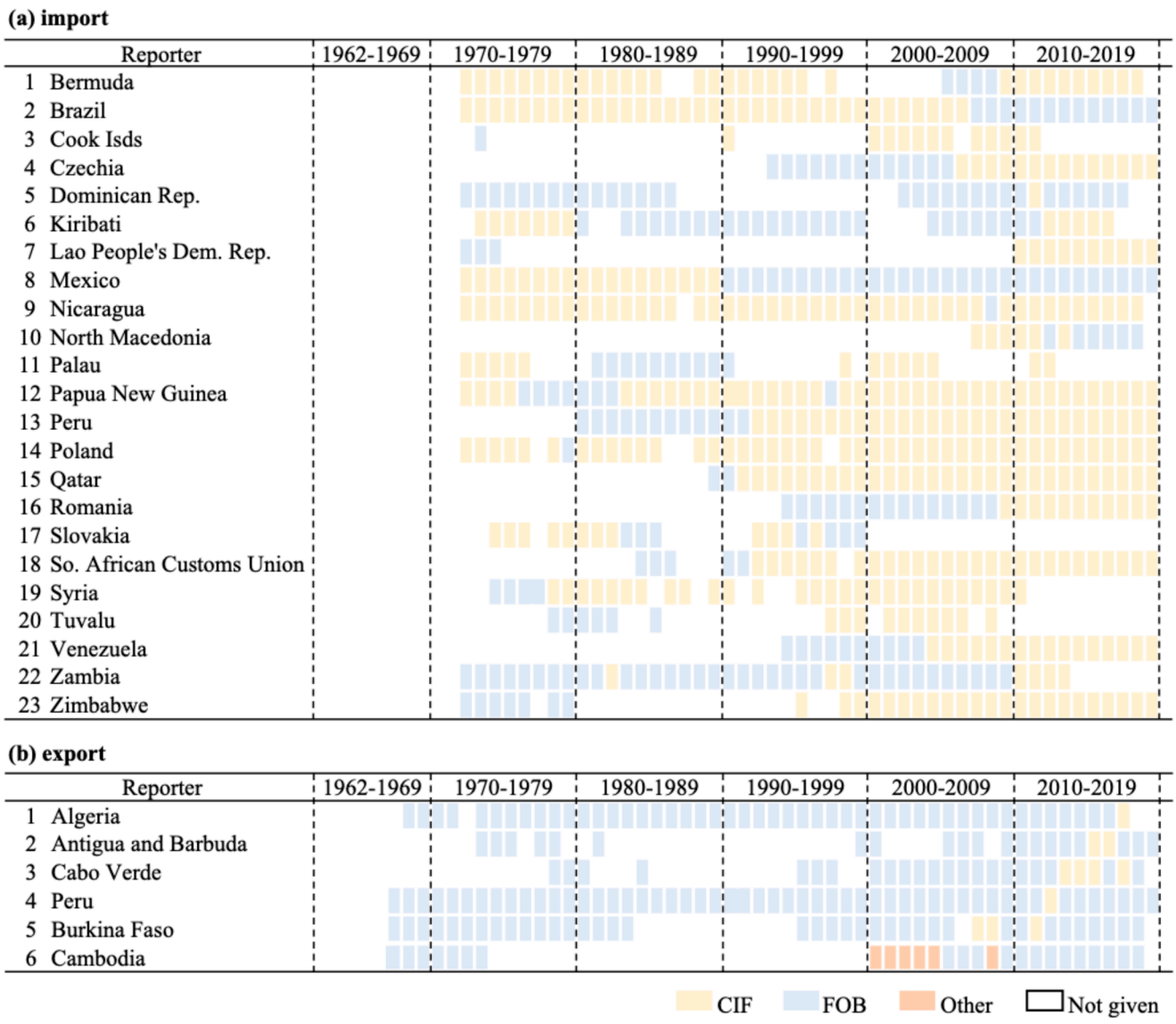

Fig. A1. Reporters using different shipping agreements to report trade value. Six reporters have used CIF to value export trade, while 23 reporters have used FOB for import value. The reporters are listed in alphabetical order.

efforts of two countries, as suggested by the Office for National Statistics in the United Kingdom (Adrian Chesson, 2018). In analyses concerning transactions between two reporters, uncertainties caused by the bilateral asymmetry issue should be fully taken into account.

5. Concluding remarks and future work

This review conveyed to the users of UN Comtrade the most critical details of standards or systematic processes to gather observations or measurements (i.e., statistic criteria) by reporters, and correspondingly preprocessing procedures by the UNSD. This review also intends to inform the users of the existence of data quality issues in UN Comtrade as well as their negative impacts on practical data applications. Furthermore, this review provides an overview of existing methods aiming at addressing each issue. We found that: (1) Different reporters would use different criteria in a particular year, and one reporter would use different criteria in different years; (2) Together, inevitable limitations of data gathering and preprocessing methods caused three data quality issues of outliers, missing values, and bilateral asymmetries. (3) These data quality issues would severely jeopardize the quality of further analysis. For example, outliers, existing in data of almost all commodities and reporters, would cause biased results (e.g., unrealistically high or low unit price); approximately 6.5% of records lack net weight values, which may lead to underestimation of flows in material flow analysis; etc. More than 2 times difference in data between different transaction sides (i.e., bilateral asymmetries), selection of one of which may lead to different results. (4) Most importantly, previously proposed methods are subject to certain limitations.

Given the vital role of UN Comtrade data, the serious impacts of data quality issues, and the limitations of existing methods, we have established an improved methodology to greatly increase UN Comtrade data quality, thus advancing analyses on physical trade flow. These efforts are presented in our three-part series of articles, and the relationships of these articles are depicted in Fig. 7. At first, this article comprehensively reviews data quality issues and solutions. The following two articles elaborate on the improved methodology to address these issues. It is worth noting that our previous attempts reveal that the global bilateral asymmetry issue caused by misreporting has been highly resolved based on trade data with outliers eliminated and missing values estimated. Asymmetries between two countries should be analyzed on a case-by-case basis. Therefore, in our future work, only two parts are designed to address the outlier and missing value issues, which are briefly

Fig. A2. Reporters using different partner attributions to report imports. Totally 79 countries/areas did not always use 'origin' to attribute their import partners.

presented as follows.

(1) Detecting and handling outliers. Existing methods for outlier detection focus on temporal comparison and statistical methods. However, most methods are only effective in specific groups of commodities for their contexts, and methods that are sensitive to subjective assumptions or parameter settings may also cause misjudgments. Besides, outliers may not be wrong. Ignoring the outliers or replacing them with estimated data may hide critical information behind the outliers. To avoid the potential loss of important information in dealing with outliers, we focus on the records with abnormal unit prices representing unreasonable commodity trade. In this regard, a mechanism of our work (Jiang et al., 2022) without strict assumptions would be developed to detect and handle outliers, which would have advantages in its suitability for trade data of various commodities and reporters.

(2) Estimating missing values. The major challenge for existing methods to process missing values is that they can hardly find

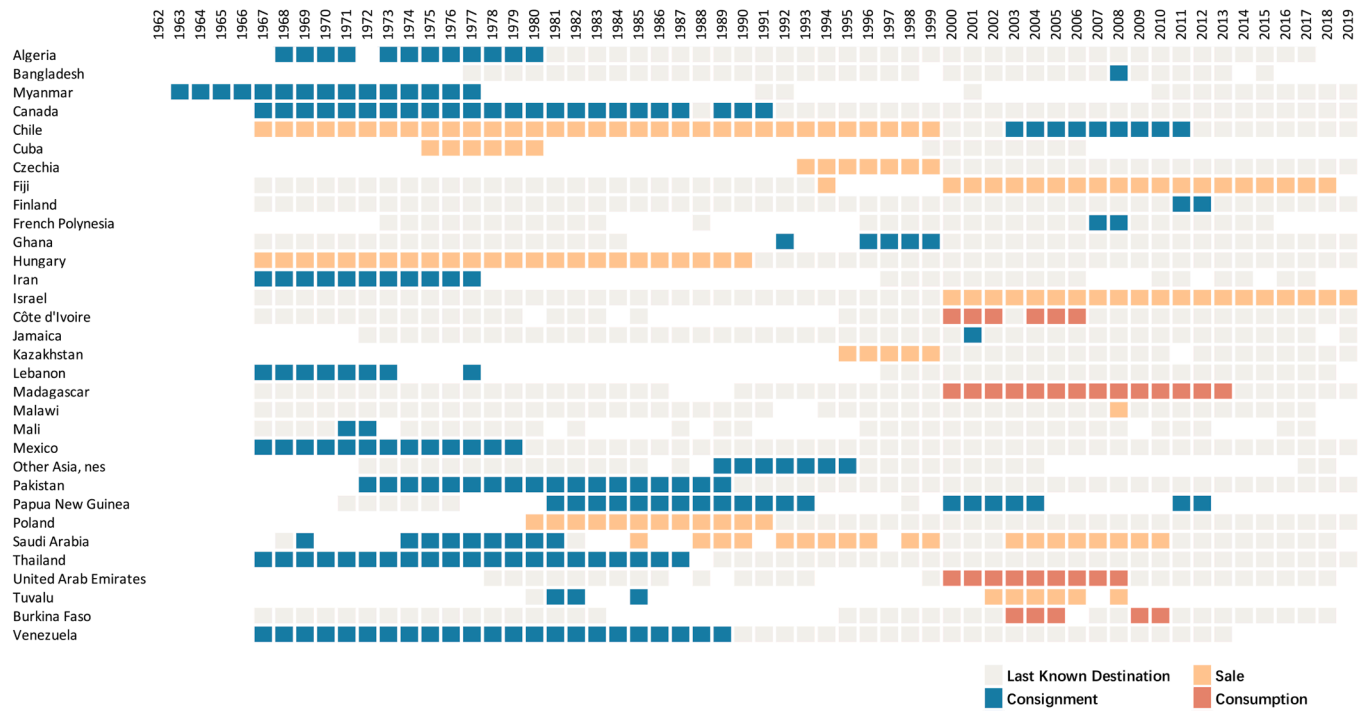

Fig. A3. Reporters using different partner attributions to report exports. Totally 32 countries/areas did not always use 'Last Known Destination' to attribute their export partners.

reliable data for estimation. The most widely-used method assumes the average prices per kilogram of one commodity are the same for all reporters in a specific year, which is not the case. To overcome this issue, instead of using the same average unit price globally, we try to capture the difference in prices among reporters. Besides, in our future work (Zhang et al., 2022) , a model employing statistical methods is established to fit missing values, considering differences in factors (i.e., commodities, reporters, and years) and thus promote estimation reliability.

CRediT authorship contribution statement

Chuke Chen: Resources, Formal analysis, Visualization, Writing – original draft. **Zhihan Jiang:** Resources, Investigation, Methodology, Validation, Formal analysis. **Nan Li:** Conceptualization, Supervision, Writing – review & editing, Funding acquisition. **Heming Wang:** Writing – review & editing, Funding acquisition. **Peng Wang:** Methodology, Writing – review & editing. **Zhihe Zhang:** Resources, Methodology. **Chao Zhang:** Writing – review & editing. **Fengmei Ma:** Writing – review & editing. **Yuanyi Huang:** Methodology, Validation, Data curation. **Xiaohui Lu:** Resources, Investigation. **Jianlimin Wei:** Resources, Investigation. **Jianchuan Qi:** Writing – review & editing. **Wei-Qiang Chen:** Conceptualization, Supervision, Writing – review & editing, Funding acquisition.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This research was supported by the National Natural Science Foundation of China (No. 52070178, No. 71961147003, No. 52170184, No.41871204, and No.52070034).

Appendix

References

Adrian, C., 2018. Asymmetries in trade data - diving deeper into UK bilateral trade data. The Office of National Statistics of the United Kingdom, London. https://www.ons.gov.uk/economy/nationalaccounts/balanceofpayments/articles/asymmetriesintradedatadivingdeeperintoukbilateraltradedata/2018-01-29#deep-dive-approach-sources-and-methods.

Aguiar, A., Chepeliev, M., Corong, E.L., McDougall, R., van der Mensbrugghe, D., 2019. The GTAP data base: version 10. J. Glob. Econ. Anal. 4, 1–27.

Zhang, Z., Jiang, Z., Chen, C., Zhang, X., Whang, H., Li, N., Wang, P., Zhang, C., Ma, F., Huang, Y., Qi, J., Chen, W.-Q., 2022. Advancing UN Comtrade for Physical Trade Flow Analysis: Addressing the Issue of Missing Values. Resour. Conserv. Recy. https://doi.org/10.1016/j.resconrec.106525.

Andrey A. Gnidchenko, 2018. Measuring product-level export quality: Evidence from Asian motor vehicles sector. FIW Working Paper, No. 185, FIW - Research Centre International Economics, Vienna.

Bahmani-Oskooee, M., Harvey, H., Hegerty, S.W., 2013. The effects of exchange-rate volatility on commodity trade between the US and Brazil. North. Am. J. Econ. Financ. 25, 70–93.

Benkovskis, K., Wörz, J., 2013. What drives the market share changes? Price versus non-price factors. European central bank working paper 1640, Berlin.

Benkovskis, K, Wörz, J., 2012. Non-price competitiveness of exports from emerging countries. Empir. Econ. 51, 707–735.

Brewer, T.D., Abbott, D., Lal, N., Sharp, M., Thow, A., Andrew, N., 2020. A method for cleaning trade data for regional analysis: The Pacific Food Trade Database Version 1 (1995-2016). Pacific Community working paper. New Caledonia.

Chen, W.-Q., Graedel, T.E., 2012. Anthropogenic cycles of the elements: A critical review. Environ. Sci. Technol. 46 (16), 8574–8586.

Chen, W.-Q., Ma, Z.J., Pauliuk, S., Wang, T, 2019. Physical and Monetary Methods for Estimating the Hidden Trade of. Materials. Resour. 8 (2), 89.

Collins, K.J., 1998. Outlook for US Agriculture. https://ageconsearch.umn.edu/record/33257.

Damerval, C., 2012. Identification of outliers in simultaneous time series using Wavelets and Forward Search. EUR - scientific and technical research reports. Luxembourg.

de Saint Vaulry, A., 2008. Base de données CHELEM-commerce international du CEPII. Working Papers 2008–09, CEPII Research Center. Paris.

Dietzenbacher, E., Los, B., Stehrer, R., Timmer, M., De Vries, G., 2013. The construction of world input–output tables in the WIOD project. Econ. Syst. Res. 25, 71–98.

Dittrich, M., Bringezu, S., 2010. The physical dimension of international trade: Part 1: Direct global flows between 1962 and 2005. Ecol. Econ. 69, 1838–1847.

Dong Guo, 2010. Mirror Statistics of International Trade in Manufacturing Goods: The Case of China. Research and statistics branch working paper No. 19/2009. Vienna.

EUROSTAT, 2007. the intrastat system of the European Union. European Commission International trade statistics - methodology and classifications draft. Luxembourg. https://unstats.un.org/unsd/trade/WS%20Lima%202007/background%20documents/Eurostat%20-%20Meeting%20document%20-%20Item%207d.pdf.

Farhan, J., 2015. Overview of missing physical commodity trade data and its imputation using data augmentation. Transport. Res. C-Emer. 54, 1–14.

Ferrantino, M.J., Liu, X., Wang, Z., 2012. Evasion behaviors of exporters and importers: Evidence from the US–China trade data discrepancy. J. Int. Econ. 86, 141–157.

Gaulier, G., Martin, J., Méjean, I., Zignago, S., 2008. International trade price indices. CEPII Working Paper No. 2008/10. Paris.

Gaulier, G., Zignago, S., 2010. BACI: International Trade Database at the Product-Level (the 1994-2007 Version). CEPII Document De Travail No. 2010–23. Paris.

Gehlhar, M., 1996. Reconciling bilateral trade data for use in GTAP. Purdue University: West Lafayette.

Gehlhar, M., Wang, Z., Yao, S., 2010. GTAP 7 Data Base Documentation - Chapter 9.A: Reconciling Merchandise Trade Data. Center for Global Trade Analysis, Purdue University.

Gephart, J.A., Pace, M.L., 2015. Structure and evolution of the global seafood trade network. Environ. Res. Lett. 10, 125014.

Giljum, S., Dittrich, M., Lieber, M., Lutter, S., 2014. Global Patterns of Material Flows and their Socio-Economic and Environmental Implications: A MFA Study on All Countries World-Wide from 1980 to 2009. Resour. 3, 319–339.

Hamanaka, S., 2011. Utilizing the Multiple Mirror Technique to Assess the Quality of Cambodian Trade Statistics (ADB Working Paper Series on Regional Economic Integration No. No. 88). Asian Development Bank, Manila.

Jiang, Z., Chen, C., Li, N., Wang, H., Wang, P., Zhang, C., Ma, F., Zhang, Z., Huang, Y., Qi, J., Chen, W., 2022. Advancing Un Comtrade for Physical Trade Flow Analysis: Addressing the Issue of Outliers. Resour. Conserv. Recy. https://doi.org/10.1016/j.resconrec.2022.106524.

Lenzen, M., Moran, D., Kanemoto, K., Foran, B., Lobefaro, L., Geschke, A., 2012. International trade drives biodiversity threats in developing nations. Nature 486, 109–112.

Lewis, K., Witham, C., 2012. Agricultural commodities and climate change. Clim. Policy. 12, S53–S61.

Li, Y., Hewitt, C., 2008. The effect of trade between China and the UK on national and global carbon dioxide emissions. Energ. Policy. 36, 1907–1914.

Marko Javorsek, 2016. Asymmetries in international merchandise trade statistics: A case study of selected countries in Asia and the Pacific. ARTNeT Working Paper Series. Berlin.

Mohammadi, E., Singh, S.J., Habib, K., 2021. Electronic waste in the Caribbean: An impending environmental disaster or an opportunity for a circular economy? Resour. Conserv. Recy. 164, 105106.

Montezuma, J., 2021. International Commercial Terms Rules [WWW Document]. URL https://internationalcommercialterms.guru/ (accessed 5.11.21).

Narins, T.P., 2018. Chinese trade in Latin America compared to the European Union and the United States: The case of technology-intensive exports. Prof. Geogr. 70, 219–229.

Qu, S., Guo, Y., Ma, Z., Chen, W.-Q., Liu, J., Liu, G., Wang, Y., Xu, M., 2019. Implications of China's foreign waste ban on the global circular economy. Resour. Conserv. Recy. 144, 252–255.

Ong, G.P., Farhan, J., Chin, A.T.H., 2013. Stochastic Imputation of Missing Physical Commodity Trade Information Using Monetary Trade Data. Transp. Res. Record. 2354, 122–132.

Robert, C.Feenstra, Hai, Wen, Woo, Wing T., Yao, Shunli, 1998. The U.S. - China bilateral trade balance: its size and determinants (Working Paper No. NBER Working Papaer Series 6598). National Bureau of Economic Research, Cambridge, MA.

Saki, Z., Rothenberg, L., Moor, M., Kandilov, I., Godfrey, A.B., 2019. Forecasting US Textile Comparative Advantage Using Autoregressive Integrated Moving Average Models and Time Series Outlier Analysis. arXiv preprint arXiv:1908.04852.

Schaffartzik, A., Eisenmenger, N., Krausmann, F., Weisz, H., 2014. Consumption-based material flow accounting: Austrian trade and consumption in raw material equivalents 1995–2007. J. Ind. Ecol. 18, 102–112.

Shaar, K., 2019. Reconciling International Trade Data. ZBW – Leibniz Information Centre for Economics working paper. Hamburg.

Shi, J., Zhang, C., Chen, W.-Q., 2021. The expansion and shrinkage of the international trade network of plastic wastes affected by China's waste management policies. Sustainable Prod. Consumption. 25, 187–197.

Silver, M., 2007. Do Unit Value Export, Import, and Terms of Trade Indices Represent or Misrepresent Price Indices? (Working Paper No. Working Paper No. 07/121). International Monetary Fund, Washington D.C.

Silver, M., Heravi, S., 2007. Why elementary price index number formulas differ: Evidence on price dispersion. J. Econometrics. 140, 874–883.

Taherzadeh, O., Caro, D., 2019. Drivers of water and land use embodied in international soybean trade. J. Clean. Prod. 223, 83–93.

UN Comtrade, 2009. Differences between Imports and Exports, Reporters and Partners (partner, reporter, Trade valuation) [WWW Document]. URL https://unstats.un.org/unsd/tradekb/Knowledgebase/50136/Differences-between-Imports-and-Exports-Reporters-and-Partners (accessed 5.11.21).

United Nations, 2019. IMTS Bilateral asymmetries – how to measure, analyze, reduce and way forward. the United Nations Statistic Division Economic Statistics Branch technical report. New York. http://unstats.un.org/unsd/tradekb/Attachment441.aspx?AttachmentType=1.

United Nations, 2018. Classification by Broad Economic Categories, Rev.5: Defined in Terms of the Harmonized Commodity Description and Coding System (2012) and the Central Product Classification, 2.1, Statistical Papers (Ser. M). New York. https://doi.org/10.18356/7cff5416-en.

United Nations, 2013. International merchandise trade statistics: compilers manual, revision 1 (IMTS 2010-CM). United Nations, New York.

United Nations, 2010. International Merchandise Trade Statistics: Concepts and Definitions 2010. the United Nations Statistics Division Statistical Papers (Series M, No. 52). New York. https://unstats.un.org/unsd/publication/SeriesM/SeriesM_52rev3e.pdf.

United Nations, 2006. Standard international trade classification revision 4. the United Nations Statistics Division Statistical Papers (Series M, No. 34/Rev. 4). New York.

United Nations, 2017. Correlation and conversion tables used in UN Comtrade. United Nations, New York.

UNSD_KH, 2021. Reported and Converted Data /Conversion Issues - Comtrade - UN Statistics Wiki [WWW Document]. Reported and Converted Data / Conversion Issues. URL https://unstats.un.org/wiki/pages/viewpage.action?pageId=79009291 (accessed 12.18.21).

U.S.-China JCCT, 2009. Report on the Statistical Discrepancy of Merchandise Trade Between the United States and China. The 15th U.S.-China Joint Commission on Commerce and Trade (JCCT). Hangzhou, P.R.C https://www.commerce.gov/sites/default/files/migrated/reports/statisticaldiscrepancy.pdf.

Wang, C., Zhao, L., Lim, M.K., Chen, W.-Q., Sutherland, J.W., 2020. Structure of the global plastic waste trade network and the impact of China's import Ban. Resour. Conserv. Recy. 153, 104591.

Yang, H., Wang, L., Abbaspour, K.C., Zehnder, A.J., 2006. Virtual water trade: an assessment of water use efficiency in the international food trade. Hydrol. Earth. Syst. Sc. 10, 443–454.

Yeats, A.J., 2012. Implications of COMTRADE compilation practices for trade barrier analyses and negotiations. World. Bank. Econ. Rev. 26, 539–555.