A method for cleaning international food trade data for regional analysis: The Pacific Food Trade Database

Version: 10 April 2022 based on Version 2.1 of the database.

Tom D. Brewer[1*], Neil L. Andrew[1], Michael K. Sharp[2], Anne Marie Thow[3], Helani Kottage[1], Stuart Jones[4]

[1] Australian National Centre for Ocean Resources and Security, University of Wollongong, Australia

[2] The Pacific Community, Noumea, New Caledonia

[3] Menzies Centre for Health Policy, The University of Sydney, Sydney, Australia

[4] Stats NZ, Wellington, New Zealand

* Corresponding author

tbrewer@uow.edu.au

Citation: Brewer, T.D., Andrew, N.L., Sharp, M.K., Thow, A.M., Kottage, H., Jones, S. 2022. *A method for cleaning international food trade data for regional analysis: The Pacific Food Trade Database.* Version 2.1. Pacific Community working paper.

CONTENTS

ABSTRACT

International trade in food, and the agreements that govern that trade, play important roles in the development of nations and the wellbeing of people. Evidence-based food policy requires analysis and interpretation of trade flows among countries. The authoritative source of global trade data, United Nations' Comtrade database and its many derivatives, are appropriate for coarse analysis but are limited by the availability and accuracy of national reporting. Exploratory analysis of trade flows with and among 18 countries and territories in the Pacific region revealed significant and consequential errors in UN Comtrade and the derived CEPII BACI databases. We describe a stepwise process to identify and correct errors to develop a reliable food trade database for the region. The method includes both expert review of the plausibility of trade flows and rule-based identification of unit price outliers and subsequent imputation associated quantities. For the period 1995–2018, a total of 9,711 (3%) trade flows contained categorical errors in some combination of exporter * importer * commodity and were re-categorized or deleted. A further 13,177 (4.2%) trade flows had implausibly large or small unit prices. Just 0.01% of outlier trade flows (particularly for rice) accounted for 98% of error in terms of quantity. The imputation process reduced the total dataset from 314.67 million t to 80.31 million t. Final detailed review of each commodity (at sub-heading level) within each country reduced the total to quantity 73.87 million t, mostly resulting from the removal of commodities such as baitfish and salt which were not primarily for human consumption. We conclude that Comtrade and derived databases do not, in their native state, provide a reliable source of information about food trade in the Pacific region. While details are unique to this dataset, the step-wise mixed method approach developed is of broad applicability to regional and national analysis of international food trade data for improved food trade policy development.

Keywords: Outliers; Comtrade; BACI; Pacific; food security; trade data cleaning

INTRODUCTION

In an ever more interconnected world, the flow of food among countries, and the agreements that govern that trade, play important roles in the development of nations and the wellbeing of people. Increasing trade, in part resulting from extensive liberalization of economic policy since the 1950s and increasingly efficient transportation, has had both positive and negative implications for food security and nutrition (Baker, Friel, Schram, & Labonte, 2016; Geyik, Hadjikakou, Karapinar, & Bryan, 2021; Thow, 2009). Increased access to a broader diversity of foods such as fruits and vegetables and reduced volatility in food availability (Brooks & Matthews, 2015; Gillson & Fouad, 2014) can support improved dietary quality and reductions in undernutrition (García-Dorado, Cornselsen, Smith, & Walls, 2019). These trends can lead to reduced food insecurity at an aggregate level (Kerr, 2011; Pyakuryal, Roy, & Thapa, 2010). In contrast, there is a clear link between the importation of ultra-processed foods and beverages and increased incidence of non-communicable disease (e.g. Estimé, Lutz, & Strobel, 2014; Thow, 2009).

The analysis of reliable food trade data should make important contributions to understanding the complex relationships between trade and food availability and so improve trade policy (Thow et al., 2011). Their potential contribution is, however, compromised by omissions and inaccuracies. Sources of error can be attributed to a range of factors including incorrect attribution by trade partners, the use of different data sources to compile datasets (e.g. customs records or mirror data), incorrect commodity attribution, measurement error associated with failure to adhere to protocols and non-reporting (Ortiz-Ospina, Beltekian, & Roser, 2018; United Nations, 2008). As acknowledged on the World Bank online trade data platform, '… *despite all efforts made by national and international agencies, data quality may vary among countries*…' (World Bank, 2021). In 2001, for example, Pakistan reported US$236 million t of exports to China but China reported US$557 million t for the same year (World Bank, 2021). Analyses of food trade dynamics that do not address the potentially significant errors in global datasets will have consequences for research outputs derived from international trade data, policy development and, ultimately, food system outcomes.

Regional studies offer a unique opportunity to examine trade flows within and among economic and trade communities, to understand how trade dynamics impact on regional food systems (Thow et al., 2015). The Pacific region is an exemplar of unreliable food trade data. In many analyses, Pacific Island Countries and Territories (PICTs; see Appendix 1) are subsumed into 'Asia-Pacific', included with Australia and New Zealand as 'Oceania', or simply missing altogether, and thereby marginalized in global discourses around food trade and security. As a point of context on the paucity of reliable data for food policy in the region, no PICTs are included in the global food security index (https://foodsecurityindex.eiu.com/), which tracks national food security of 113 nations through time. Reliable time series of high resolution food commodity data are essential to attempts to better describe and interpret these changes in food acquisition and consumption, and to formulate evidence-based policy for improved economic and health outcomes for the region.

To support broader analyses of food security and nutrition in the Pacific region (e.g. Andrew et al., 2022), we describe a method for cleaning international trade data from the region. The Pacific region is comprised of small island-state economies with relatively diverse imports and limited exports. The region is increasingly dependent on globalized trade in food commodities (Gewertz & Errington, 2010; Hawkes, 2010) and vulnerable to volatility in those markets and a range of other external drivers (Farrell et al., 2020). Declining per capita production combined with growing imports of unhealthy foods and increased incidence of non-communicable diseases demands change in understanding the regional food system (Andrew et al., 2022), and this requires reliable food trade data. Although the specifics of the analysis and its conclusions are unique to the region, the method described herein offers a transparent process for cleaning international food trade data, and a cautionary tale concerning the uncritical use of global datasets.

— Sources of international trade data and rationale for this analysis

National food trade data held by PICTs were explored for their suitability, however we were unable to access sufficiently long and consistent time-series across all PICTs and concerns remained about the granularity of commodity classification and variation in reporting methods. Alternative data sources included a range of international organizations, notably United Nations Comtrade (https://COMTRADE.un.org), International Trade Centre (http://www.intracen.org/default.aspx), World Bank (https://wits.worldbank.org/default.aspx), Food and Agriculture Organization's FAOSTAT (http://www.fao.org/faostat/en/#data/TM), World Bank (https://data.worldbank.org/topic/trade) and World Trade Organization (https://www.wto.org/english/res_e/statis_e/statis_e.htm). These sources provide standardised accounts at the national level at varying resolutions of commodity detail and availability of country data. Importantly, none of these sources systematically reconciles trade flows reported by exporters and importers, which are often vastly different.

The CEPII-BACI international trade database (Gaulier & Zignago, 2010) is an international trade database derived from the UN Comtrade database. The BACI database adds significant value to UN Comtrade by reconciling reporting differences among countries and filling gaps created by non-reporting of trade flows. As Gaulier and Zignago (2010) note, some countries, including some PICTs, do not report trade statistics to the United Nations. BACI utilizes mirror data (trade flows described by the trade partner) to provide a more complete and coherent set of trade flows. Further, in BACI, quantities have been converted from non-standard units into metric tonnes and values to Free On Board (FOB) equivalent expressed in US$.

CEPII-BACI (herein BACI) fills many gaps in reporting in the UN Comtrade database and so provides a more complete platform for analyses of food trade. Both datasets are reliant on information provided to UN Comtrade by countries. BACI may also introduce error because trade flows from mirror data are equally open to miss-reporting. The impact of error generated from the use of mirror data could be disproportionate for small PICT economies (e.g. Tuvalu) when errors are- introduced from much larger trading partners (e.g. Australia). The stated purpose of the BACI database is to provide an international trade database covering the largest number of countries at the highest degree of product-detail, for the longest period (Gaulier & Zignago, 2010). It does not provide detailed country – or commodity – specific cleaning, but offers the best foundation from which to build a more reliable regional database.

Explorations of trade flows with PICTs in BACI suggested there were many errors, including numerous records of implausible quantities, and incorrect country attribution that could only be corrected by a systematic and partially expert-based and non-statistical approach to recognition and treatment of errors. Our guiding principle was to limit changes to the primary data, adjusting or deleting only implausible trade flows and retaining those that were merely improbable. Examples of implausible trade flows are provided in each stage outlined below, and include, for example, export of commodities not produced by the PICT (e.g. temperate zone fruit such as apples or cashew nuts).

— Stages in database development

The described method comprised step-wise process in five stages (Figure 1):

1. Database development, in which the original download was reduced to only data of interest
2. Removal or re-categorization of implausible **exports** from PICTs based on empirical exploration of the data and expert elicitation
3. Removal or re-categorization of implausible **imports** by PICTs based on empirical exploration of the data and expert elicitation
4. Identification and imputation of outliers in unit price from plausible trade flows using a rule-based imputation method
5. Systematic cleaning at country * HS6 commodity resolution for both imports and exports across all 18 PICTs

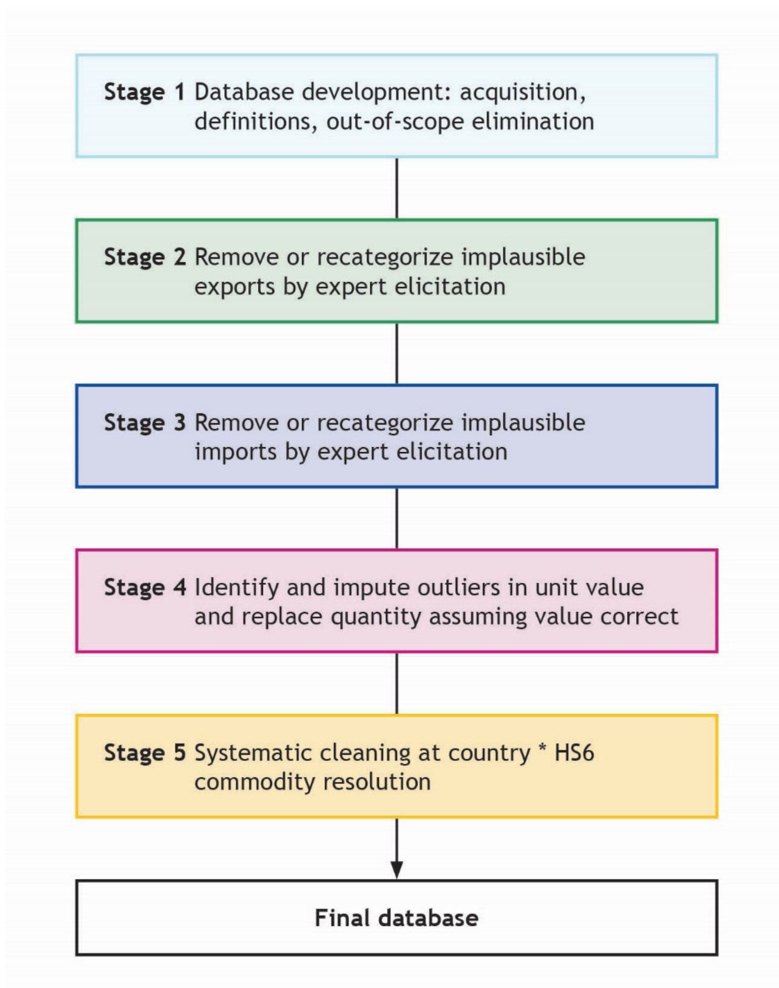


Figure 1. Sequence of stages in database development and cleaning. See text for additional description of sub-stages.

Separately or in combination, the elements of a trade flow ('exporter', 'importer', 'commodity', 'quantity', 'value', 'year', and the derived variable 'unit price') offer different lens through which to recognize errors. All elements are subject to error, but for the purposes of our analysis 'year' and ‘value’ were assumed to be reported without error (see Appendix 8 for exploration of this assumption). These assumptions leave combinations of ‘exporter’, ‘importer’, ‘commodity’, ‘quantity’ and ‘unit price’ as clues in the identification of error. Combinations of ‘exporter’, ‘importer’, ‘year’ and ‘commodity’ are explored on a categorical basis in Stages 2 and 3, and ‘quantity’ is dealt with by identification and imputation of outliers in unit price in Stage 4. In Stages 2 and 3 ‘quantity’ may be used as a lens to recognize implausible trade flows, but deletions were not made on the basis of ‘quantity’ *per se*. **As examples, consider these implausible trade flows from the raw BACI download (cleaning stage where the trade flow was highlighted in parentheses):**

- 1,134 t of poppy seeds exported from Tuvalu to Sweden in 2001 (Stage 2)
- 190 t of undenatured ethyl alcohol from Nauru to the Republic of Moldova in 1996 (Stage 2)
- 15.6 million t of brown rice exported from Australia to Papua New Guinea in 2000 (Stage 4)

Further examples of deleted and adjusted trade flows associated with each stage are provided in Appendices 3 to 7 and in the following text.

STAGE 1: DATABASE DEVELOPMENT

— Stage 1.1: Data acquisition

Data were downloaded from the CEPII BACI database (http://www.cepii.fr/CEPII/en/bdd_modele/presentation.asp?id=1) on 25 March, 2020. The download contained all trade flows for all countries from 1995 to 2018 at the sub-heading (6-digit) level under the United Nations Harmonized Commodity Description and Coding Systems (HS) https://unstats.un.org/unsd/tradekb/Knowledgebase/50018/Harmonized-Commodity-Description-and-Coding-Systems-HS. The HS classification undergoes a major review every five years, resulting in code changes. Almost always, codes from earlier classifications are split, creating more six digit codes with each revision. In this application we used the HS92 classification to maximize the number of years in the time series comparable to other databases (see http://www.cepii.fr/DATA_DOWNLOAD/baci/doc/DescriptionBACI.html for a version comparison). The data included year, exporter country code, importer country code, sub-heading commodity code, quantity (t), and value (US$000).

— Stage 1.2: Data definitions

BACI provides commodity descriptions at the sub-heading (6-digit) level separately. Descriptions were mapped to commodity codes. BACI uses Comtrade country codes to denote importer and exporter within the primary data, and provides country names and ISO codes separately. Country codes were mapped to country names and ISO codes provided by BACI.

— Stage 1.3: Out-of-scope data elimination

Trade flows outside the scope of the database, or with incomplete records, were excluded in the following order:

1. Commodities not for human consumption were excluded based on HS sub-heading definitions and cross referenced with the Central Product Classification v1.1 (United Nations, 2002) as required. HS Chapters 01-04, 07-12, 15-22, 24 and 25 (salt only, HS250100) were retained with exceptions. Although not usually considered a food or beverage commodity, tobacco (HS Chapter 24) was retained because it was of interest for future analyses of linkages between consumption and health.
2. Trade flows that did not include a PICT as either importer or exporter were removed. Pitcairn and Norfolk Island were treated as non-PICTs; the former because of its extremely small size (ca. 50 people live there), and the latter because it is administratively part of Australia.
3. The Pacific territories of the United States of America (American Samoa, Commonwealth of Northern Mariana Islands and Guam) were treated as non-PICTs because there were no trade flows between these territories and the USA; nor were there trade flows recorded for American Samoa or Guam for the years 1995–1999.

At completion of Stage 1.3, the dataset included 314,509 trade flows and 581 unique commodities at HS6 level for 18 PICTs (Table 1) spanning the years 1995–2018. The total value of the trade flows was US$79.6 billion and the total quantity was 315,337,498 t.

Table 1. Total records for included PICTs at end of Stage 1. Note, records, quantity and value exceed the totals provided in text above due to double counting of between-PICT trade records. FSM = Federated States of Micronesia.

PICT	Export records (n)	Import records (n)	Export quantity (t)	Import quantity (t)	Export value (US$000)	Import value (US$000)
Cook Islands	803	8,527	72,615	266,950	251,331	470,472
FSM	994	13,640	890,045	881,671	1,343,534	910,449
Fiji	26,672	33,964	13,241,079	8,218,330	9,516,102	6,309,732
French Polynesia	3,285	45,942	331,402	3,689,421	675,662	6,946,327
Kiribati	874	10,737	1,662,121	779,127	1,131,581	655,772
Marshall Islands	950	5,823	1,170,993	884,050	1,655,213	411,636
Nauru	401	4,327	23,247	125,917	29,653	193,867
New Caledonia	3,548	41,852	162,153	3,430,365	582,607	5,993,793
Niue	291	3,111	11,349	59,395	17,366	261,936
Palau	386	12,352	89,182	333,430	452,642	632,261
Papua New Guinea	7,911	23,724	31,705,134	237,460,620	21,158,439	9,850,637
Samoa	3,403	17,623	489,163	2,513,195	441,382	1,621,273
Solomon Islands	1,885	11,841	1,798,794	1,929,556	2,242,276	1,362,341
Tokelau	782	1,085	89,750	143,203	38,887	46,416
Tonga	1,812	13,484	296,681	800,265	276,318	900,714
Tuvalu	410	4,376	97,987	135,593	173,305	94,923
Vanuatu	2,273	14,993	2,333,190	710,976	3,021,106	848,238
Wallis and Futuna Islands	47	7,967	717	141,526	1,071	246,883

STAGE 2: REMOVING OR REVISING IMPLAUSIBLE PICT EXPORT RECORDS

—Stage 2.1: Implausible exports (exporter * commodity combination)

The data were reviewed to recognize implausible combinations of exporter and commodity for each PICT. Trade experts from The Pacific Community (SPC), relevant national agencies and National Statistics Offices completed a survey in which they were presented with the complete global list of 82 food and beverage commodities at HS4. **Respondents answered the following question for each commodity:**

'*In your opinion, in the last 25 years, has [their PICT] exported this food type?'*

In instances where more than one expert responded and answers differed, the commodity was included as plausible. Several PICTs, notably Fiji, New Caledonia and Samoa, act as trading hubs for small PICTs, importing commodities and re-exporting them to their final destination. Respondents categorized re-exports as plausible exports. The resulting list of implausible exports were assumed to have been incorrectly coded by the importing country and introduced as part of the BACI reconciliation process.

This stage removed 4.7% of all PICT export records in the database or 2,692 implausible trade flows comprising 344,583 t or 0.63% of total export quantity in the database at this point (Figure 2; Appendix 2). The total value removed was US$192,174,000 or 0.45% of all PICT export value in the database at this point. Removed trade flows include commodities not produced in low elevation tropical climates and from PICTs that do not export food (e.g. Tokelau); **examples include:**

- Olive oil exported from Tuvalu
- Live horses for food exported from Wallis and Futuna
- Chocolate exported from Tokelau
- Fresh apples, pears and quinces exported from Nauru

— Stage 2.2: Plausible exports (PICT exporter – PICT importer combination)

Of the 2,692 candidate implausible trade flows isolated in Stage 2.1, 305 were between PICTs. These inter-PICT trade flows were inspected by regional trade experts to ensure none were plausible re-exports. In instances where there was discrepancy between expert opinions, weight of evidence, including expert commentary or third-party verification, was used to assign a judgement. 148 records were considered plausible as re-exports and reinstated. In making this judgement we assumed PICTs did not re-export foods to countries outside the region. The remaining 157 records comprised commodities likely to be imported by the importing PICT but unlikely to be produced by the exporting PICT. These records were retained within the database with exporter name changed to 'Unknown'. This stage reinstated all 305 inter-PICT trade flows corresponding to 27,460 t (Figure 2; see Appendix 3 for illustrative examples of inter-PICT re-trades). **Examples of inter-PICT trade flows included:**

- 21,726 t of rice from Solomon Islands to Papua New Guinea (in this instance the exporter was changed to 'Unknown')
- 250 kg of tobacco from Kiribati to Tuvalu, a plausible retrade.

— Stage 2.3: Implausible exports (PICT exporter – importer combination) isolated by quantity or frequency

Although most implausible exports were identified in Stage 2.1, as a further check the combination of PICT exporter and non-PICT importer was used to isolate implausible flows. This stage was used to isolate instances where exporter-commodity combinations were plausible but the non-PICT importing country for that trade flow was not. We created a matrix of PICT exporter by non-PICT importer for: (i) frequency of trade flows at HS6, and (ii) quantity (t) of trade for all commodities. This level of disaggregation was required because stage 2.1 was conducted at HS4 and some trade flows plausible at HS4 might not be at HS6. These matrices were inspected for unusually large quantities and frequencies of trade flows, and isolated instances where there were limited trade flows between countries in the whole data set. A set of 900 trade flows were identified for detailed review. If the combination of exporter-importer-commodity was plausible but the quantity was not then the trade flow was retained, to be further reviewed in Stage 4.

Two trade experts reviewed and scored the 900 records and provided supporting text in some instances. A score of 0 was given if it was implausible that the PICT exports the commodity to any country (including other PICTs). A score of 1 was given where it was implausible that the PICT exports the commodity to the import country recorded, but it is plausible that the PICT exporter exports the commodity. A score of 2 was given in instances where the PICT exporter – non-PICT importer is plausible given commodity and year. Additional score assignment was given to incorrect commodity description, incorrect country name and in instances where records were retained unchanged.

Experts also provided commentary in some instances. For example, both experts noted that export of 'Crustaceans: frozen, n.e.s. in item no. 0306.1 (whether in shell or not, whether or not cooked by steaming or by boiling in water)' (HS030619) from Cook Islands was likely to be pearl shell (HS0508). In 139/900 instances experts differed – a judgement whether to exclude the record was made on the weight of evidence (e.g. supplementary verification by other experts), with a bias toward inclusion. **Other anomalies identified in this stage required a change either in the name of the importing countries or the commodity. Records were returned to the database as:**

i. 39 without alteration.
ii. 98 with importer changed to 'Unknown'.
iii. 80 with importer name changed to 'PICT unknown' (where the trade flows was assumed to be between PICTs).
iv. 15 with importer changed to 'Unknown', and the commodity changed from a palm oil commodity [HS151110 (n = 8), HS151190 (n = 7)] to comparable copra oil commodities [HS151311 (n = 8), HS151319 (n = 7)]. This occurred in instances where PICTs do not produce palm oil and expert reviewers judged the export to be copra oil. These records were for Fiji, Marshall Islands and Samoa. Other examples of trade flows where importer was changed to 'Unknown' include 5,918 t of Copra from Papua New Guinea to Pitcairn, and 1,375 t of skipjack or stripe-bellied bonito (tuna) from FSM to Mauritius.
v. 24 exports from Fiji with importing country changed from Christmas Island (the Australian external territory in the Indian Ocean) to Kiribati - reviewers concluded the Christmas Island referred to was the island of Kiritimati (Christmas Island) in Kiribati. An additional 83 exports from Fiji to Christmas Island not reviewed by the experts were also changed as above.

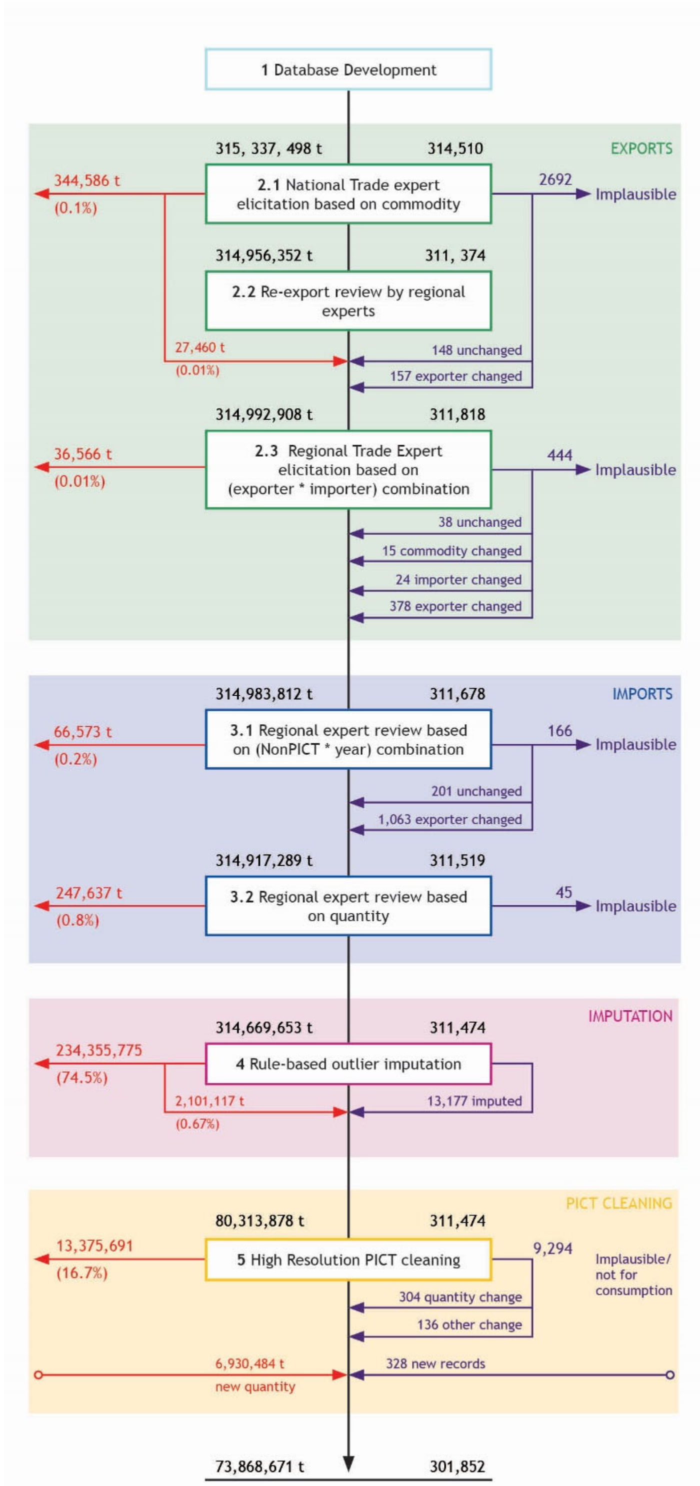


Figure 2. Outcomes at each stage of database development and cleaning. See text for additional description of sub-stages. The numbers (purple text) and quantities (red text) of trade flows reviewed and either returned or deleted are indicated. Quantities and numbers of trade flows in black refer to trade flows present prior to cleaning in each stage.

At the completion of Stage 2, cleaning of PICT exports including inter-PICT trades, a total of 2,832 records (5%) were removed from the database as implausible. A total of 574 trade flows were retained within the database with some adjustments, such as importer and exporter attribution and commodity (see Appendix 4 for illustrative examples). A total of 53,222 PICT export records were retained within the database in their original form.

STAGE 3: REMOVING AND ADJUSTING IMPLAUSIBLE PICT IMPORT RECORDS

Categorical identification of implausible imports was more difficult than exports because, not only were there far more trade flows (N = 257,782), there was also a much greater diversity of food and beverage commodities imported, particularly by PICTs with significant tourism sectors, and from diverse exporters. Because our focus was on the Pacific region, we were more concerned with the plausibility of the PICT importer-commodity combination than the identity of the exporter. **This focus has implications for the cleaning process; consider the following imports to Federated States of Micronesia (FSM) in the downloaded dataset:**

- 'Fish preparations: mackerel, prepared or preserved …' (HS160415) from Mali in 2005
- 'Meat preparations of swine …' (HS160241) from Saudi Arabia in 2005
- 'Fish preparations: sardines, sardinella …' (HS160413) from Mongolia in 2009

In these cases the exporters were implausible but the combination of PICT importer – commodity was plausible so it was not, *a priori*, defensible to delete the trade flow to FSM. We therefore created an exporter code 'Unknown' to retain trade flows that were plausible imports to the PICTs.

In a smaller number of instances, the PICT importer-commodity combination was considered implausible irrespective of the exporter, and the trade flow was deleted from the database. **Examples of such flows include:**

- 2,569 t of palm nuts and kernels from Nigeria to Marshall Islands in 2002
- 2,362 t of tobacco from Zimbabwe to Tokelau in 2004
- 16,010 t of cashew nuts from Burkina Faso to Vanuatu in 2013

— Stage 3.1: Implausible imports (exporter – PICT importer combination) isolated by quantity

Given the large number of imports, emphasis was placed on identifying those with large quantities that could be influential in national and regional analyses. For each PICT, we created a matrix of quantity of trade flows for all commodities by non-PICT exporter by year. This approach enabled detailed examination of all trade pairings through time. The matrices were inspected for unusual patterns in trade flows including very large quantities, single trade flows for country pairings and sequences of annual trade flows of improbable commodities in one or several blocks only. If a single year of trade flow between exporter and PICT importer appeared anomalous, all trade flows within the year were inspected. If the combination of exporter–PICT importer-commodity was plausible but the quantity was not then the trade flow was retained, to be further reviewed in Stage 4.

This process identified a total of 1,430 trade flows for further inspection (Figure 2) by regional trade experts. Following review, 201 (0.1%) trade flows were returned to the database unaltered, 4 were returned with exporter changed from Christmas Island to Kiribati, 1,059 were returned to the database with exporter name changed to 'Unknown', and 166 trade flows were deleted (Figure 2, see Appendix 5 for illustrative examples of adjusted records). The set of eliminated records included two anomalous clusters of trade flows, between Nigeria and Niue and between Sweden and Tuvalu as well as other implausible trade flows. **Examples of eliminated records include:**

- 7,451 t of cocoa beans exported from Nigeria to Niue in 2011
- 351 t of alcohol exported from Sweden to Tuvalu in 2002

- 40 t of pepper exported from Vietnam to Tokelau in 2002

— Stage 3.2: Implausible imports isolated by quantity for each commodity

In this stage we reviewed all imports by PICTs, with data sorted by PICT * HS2 (Chapter) and quantity. This analysis allowed focused review of the larger quantities traded with each PICT, within each commodity chapter. The purpose of this analysis was to re-check imports for errors that might have been overlooked in stage 3.1 in instances where the exporter consistently exported large volumes with a PICT through time; such a pattern would not have justified review in stage 3.1.

Reviewing all import records, focusing on large quantities, identified 45 records (0.1% of the database at this stage) that were deemed implausible in terms of six elements in the database (exporter, importer, commodity, quantity, value, and year), and eliminated from the database. In most instances it is likely the commodity was exported by the exporter, but not imported by the PICT (see Appendix 6 for illustrative examples of implausible records). **Examples of excluded trade flows include:**

- 1,250 t of bovine, sheep and goat fat from Australia to Palau in 2012
- 124,710 t of cigarettes from Indonesia to Solomon Islands in 2017
- 32 t of live animals for food from Tanzania to Kiribati in 2017

At the end of this stage, 3,036 trade flows were eliminated (0.96% of total trade flows) corresponding to 314,669,653 t. Two ISO codes were amended for ‘unknown’ and ‘PICT unknown’, changing the nomenclature to ‘UNK’ and ‘PICTUNK’ respectively. Additionally, we changed ‘Other Asia, not elsewhere specified’ to ‘Taiwan and other Asia, nes.’ and provided it with the unique ISO of ‘TOA’ because Taiwan was not included separately and is likely to comprise the majority of trade flows categorised as ‘Other Asia, not elsewhere specified’ (https://unstats.un.org/unsd/tradekb/Knowledgebase/Taiwan-Province-of-China-Trade-data).

STAGE 4: CLEANING UNIT PRICE OUTLIERS

The preceding stages modified or removed implausible records based on the identity of exporters, importers and traded commodity. More difficult errors to treat were those with plausible pairings of trading countries and commodities, but implausible quantities (t) or values ($). Further, it is possible that errors in both could remain undetected if the unit price appeared plausible. **Examples of such implausible trade flows included:**

- 48.5 million t of brown husked rice exported from Australia to PNG in 2001. The net value of this trade was recorded as US$35,469,139 - roughly 70 cents per t. This quantity of rice far exceeded Australia's total net annual rice production of around 800,000 t in 2001, and would equate to roughly 10 t per capita for the PNG population.
- 689,000 t of copra exported from Kiribati to Philippines in 2009. The net value of this trade was recorded as US$322,296 – roughly 50 cents per t.
- 21,650 t of raw cane sugar from Papua New Guinea to New Zealand in 2000. The value of this trade was recorded as US$1,598 – roughly 13 cents per t.
- 1 kg of miscellaneous food preparations (HS210690) from Fiji to Solomon Islands in 2015. The value of this trade was US$54,143.

There was no *a priori* justification to remove such trade flows completely, but such were the quantities, they would be problematic in interpretation of trends in trade flows even at regional aggregated scales. Detailed investigation of all trade flows for potential errors in quantity or price was considered too *ad hoc* and interventionist in the absence of prohibitive investments in expert review.

Below we detail the method used to identify outliers in unit price at HS6 and replace them with imputed values based on median unit price. We assume value is reported correctly and use the imputed unit price to correct quantity. Value data is more likely to be correct, primarily because it is reported in a standard unit and is used for calculating import and export taxes and duties, whereas quantity units are highly variable and less consistently reported (Food and Agriculture Organisation, United Nations Statistics Division, International Trade Centre 2019). Below we describe the two-stage process for recognizing and treating outliers. The sensitivity of results to assumptions and methods in identifying and replacing outliers is described in detail in Appendix 7.

— Stage 4.1: Identifying unit price outliers

We used Tukey's (1977) interquartile range method to recognize outliers for unit price ($/t) transformed into natural log space following convention in identifying quantity outliers in trade data (Food and Agriculture Organisation, United Nations Statistics Division, International Trade Centre, 2019). All available trade data (1995-2018) for each HS6 code were pooled across all PICTs and to maximise the diversity and number of observations in each sample. The largest sample size was food preparations n.e.s. (HS210690, N = 4,575) and the smallest sample sizes were for live carp (HS030193, N = 2), worked barley grain (HS110421, N = 2) and castor oil seeds (HS120730, N = 2). Only 14 commodities at HS6 had N < 10.

Upper and lower fences were set at multiples of the interquartile range where the lower fence = $Q_1 - k(Q_3 - Q_1)$ and upper fence = $Q_3 + k(Q_3 - Q_1)$, where Q = quartile and k = multiplier. By convention, k is set at 1.5 (Tukey, 1977) which, if the data were perfectly normally distributed, would result in ca. 0.7% of observations being recognized as outliers (Jones, 2018; Tukey, 1977). Following sensitivity analysis (Appendix 7), we retained k at 1.5. Example distributions are shown in Figure 3.

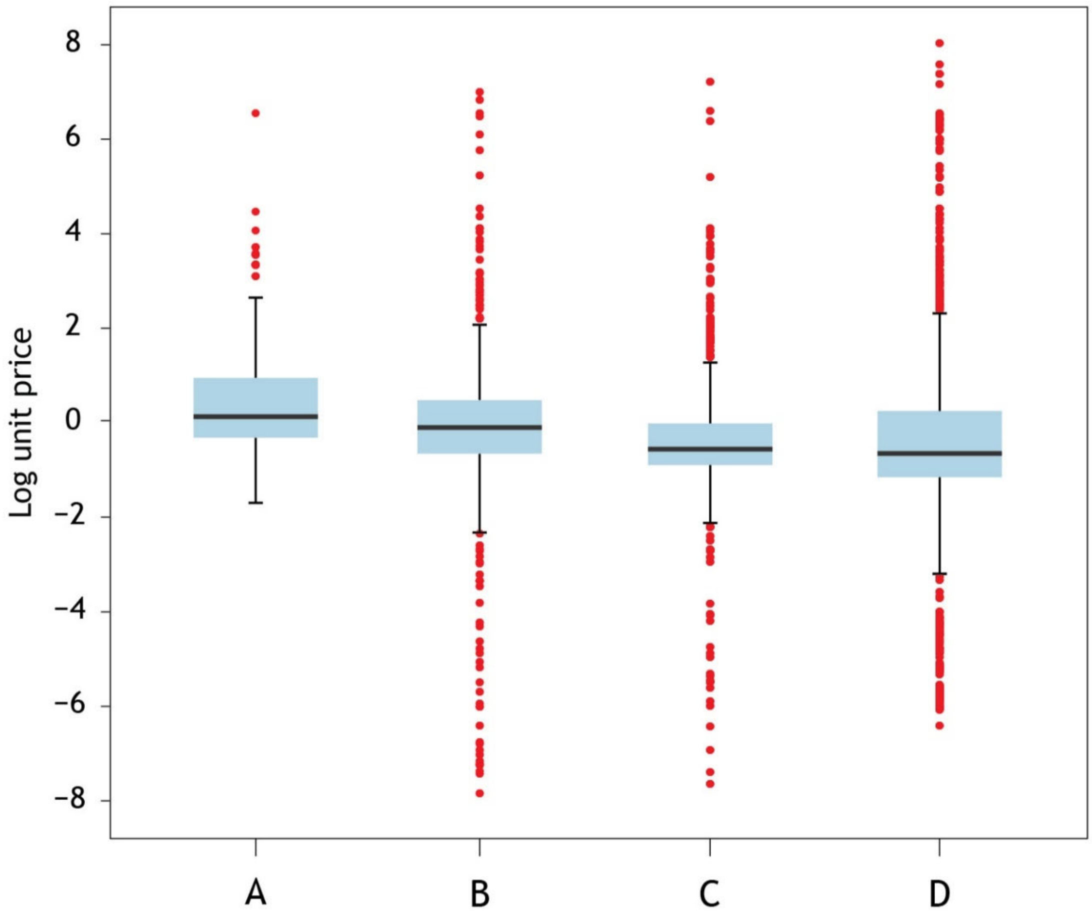


Figure 3. Illustrative box and whisker plots of log unit price of commodities showing trade flows identified as outliers at k = 1.5. Outliers shown as red dots. A = Vegetable roots and tubers: sweet potatoes, with high starch or inulin content (HS071420); B = Cereals: rice, semi-milled or wholly milled (HS100630); C = Sugars: cane sugar, raw, in solid form (HS170111); D = Water other than mineral and aerated not containing added sugar (HS220190).

In instances where an outlier unit price was for a trade flow between two PICTs, the record was associated with the exporting PICT. A total of 1,210 between-PICT trade flows were recognized as outliers, the majority of which were from Fiji, a re-export hub for the region.

A total of 13,177 (4.2%) trade flows were identified as unit price outliers, accounting for 236,456,892 t or 75% of the total quantity in the dataset at this stage of cleaning (Figure 2). The vast majority (93%) of outliers were in trade flows of less than 100 t, but just 20 trade flows accounted for 96% of outlier quantity (Table 2). Of these 20 outliers, 18 were exports of rice (HS1006) from Australia to Papua New Guinea (PNG). Of the total number of outliers, 26% (8,209,833 t) were exports from PICTs and 84% (228,984,458 t) were imports to PICTs (numbers exceed 100% due to between-PICT trades).

Of the 11,967 trade flows between PICTs and non-PICTs identified as outliers, 19% of the number of outliers (232,099,299 t or 98% of total outlier quantity) involved exports or imports to/from PNG. The remaining outlier trade flows between PICTs and non-PICTs were spread among PICTs from the rest of Melanesia (35%), Micronesia (13%), and Polynesia (33%). For the great majority of outlier trade flows, the bulk of the outlier quantity was found in a small number of trade flows within each commodity sub-heading (HS6), typically fewer than 10, with the remainder contributing a much smaller quantity (Appendix 8).

Table 2. Twenty largest trade flows (by quantity) recognized as unit price outliers.

Year	Exporter	Importer	HS6 code	HS4 Name	Value (US$000)	Quantity (t)	Unit Price ($/kg)
2001	Australia	PNG	100630	Rice	35,469	48,501,785	0.0007
2001	Australia	PNG	100620	Rice	11,487	43,229,962	0.0003
2000	Australia	PNG	100630	Rice	34,305	31,499,092	0.0011
2002	Australia	PNG	100620	Rice	13,164	24,967,188	0.0005
2002	Australia	PNG	100630	Rice	15,128	18,246,080	0.0008
2000	Australia	PNG	100620	Rice	5,799	15,667,328	0.0004
2002	Australia	PNG	100610	Rice	11,636	8,992,090	0.0013
2002	Australia	PNG	100640	Rice	4,610	7,813,048	0.0006
1998	Australia	PNG	100610	Rice	17,530	6,630,408	0.0026
1998	Australia	PNG	100630	Rice	43,121	5,451,703	0.0079
2003	China	PNG	100610	Rice	22,630	4,412,524	0.0051
2002	PNG	Bangladesh	180100	Cocoa beans	2,122	3,810,940	0.0006
2000	Australia	PNG	100610	Rice	918	1,622,893	0.0006
2003	Australia	PNG	100640	Rice	5,029	1,426,630	0.0035
2003	Australia	PNG	100610	Rice	6,985	1,138,825	0.0061
1998	Australia	PNG	100620	Rice	6,041	1,016,821	0.0059
2002	Ukraine	PNG	100620	Rice	284	955,200	0.0003
2002	USA	PNG	100610	Rice	1,839	867,650	0.0021
2016	PNG	Philippines	030343	Frozen fish	68,597	725,632	0.0945
2002	USA	PNG	100630	Rice	1,652	716,862	0.0023

— Stage 4.2: Unit price outlier imputation

Outlier unit prices were imputed at the HS6 level with the median of non-outlier unit prices from a sample of similar trade flows. Imputed unit prices were then used with value ($) to revise quantities. Various imputation methods were compared, including the use of standard unit prices within HS6 (Food and Agriculture Organisation, United Nations Statistics Division, International Trade Centre, 2019), and the most suitable method used here (see Appendix 7 for methods description and sensitivity analysis, Appendix 8 & 9 for a full breakdown of imputations across the tested methods). Medians at HS6 were estimated from combinations of PICT * YEAR with a minimum sample size of N ≥ 20 (see Method 8 in the sensitivity analyses in Appendix 7). If the N ≥ 20 sample size rule was not satisfied for a single year then years were added in increments of two to a maximum of 21 years. The outlier year was then taken as the mid-year (e.g. if three years was required to achieve n ≥ 20 for outliers in 1997 then the sample was drawn from 1996, 1997 and 1998). If the sample size remained <20 at this point then the median was estimated from all years for that PICT, irrespective of sample size. The chosen method, while complex, accounted for the most variation in unit prices given available categorical variables for isolating unique median unit prices.

Imputation of outliers towards either end of the time series, notably the first or last years, occasionally 'required' sample years outside the dataset (before 1995 or after 2018). In these instances the median was estimated with the remaining 'in scope' years (see Figure 4 for illustration). This rule was designed to use years closest to the outlier year to estimate the median and therefore minimize the impact of systemic change in unit price through time. In 1,964 (15%) instances there was fewer than 20 non-outlier data points within the PICT across all years. In these instances the process was repeated using non-outlier data from all PICTs where there was a sample size of 20 or greater. Finally, in 21 of the 1,964 instances the minimum sample size was not reached from all PICTs and the median unit price of the sample within HS6 across all PICTs, irrespective of sample size was used.

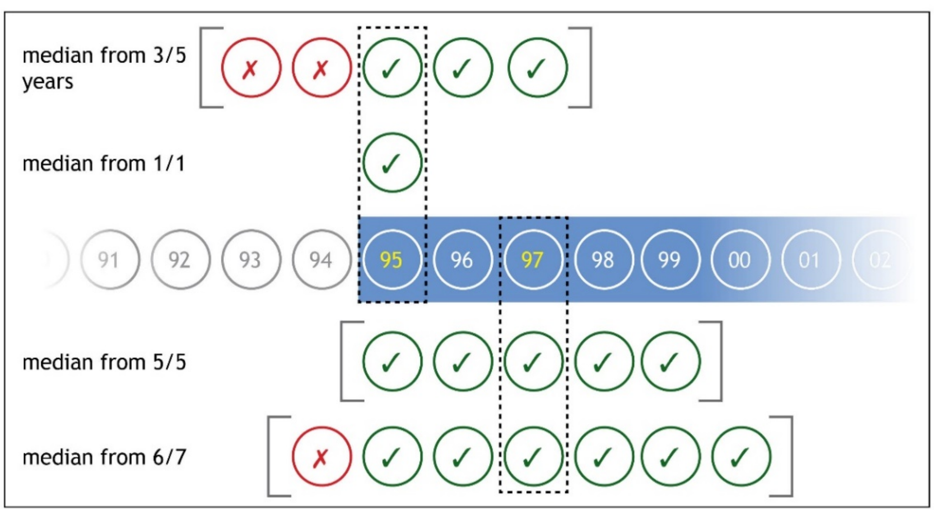


Figure 4. Method used to calculate medians of unit price when the required years were out of scope of the dataset. Illustrative examples are shown for 1995 (above the timeline) and 1997 (below). The middle year in the sample contains the outlier (in dashed rectangle). Years before 1995 are outside the scope of the dataset. Sample sizes were calculated from in-scope years and more added as required to satisfy the minimum sample size rule for each method.

STAGE 5: SYSTEMATIC CLEANING AT COUNTRY * HS6 COMMODITY RESOLUTION

To this point data cleaning focused on deleting and adjusting records at a coarse level (stages 1-3) and imputation of quantities (stage 4). The approach taken, and review of changes in the data, suggests that the data at the end of stage 4 was significantly less variable at the regional and sub-regional scales, and at aggregations of commodity (HS) groupings such as at the Chapter level. However, concerns remained regarding data quality and interpretation at higher resolutions, such as individual countries and commodities. Country-level variation caused by error that must be addressed by country-level cleaning is likely to be more acute for countries with fewer trade flows and smaller total quantities. Given the broad objective of the database was to inform policy related to food trade for individual countries, country level cleaning, as described below, was considered necessary. **As examples of PICT-level concerns of data quality, consider the following:**

- Solomon Islands has implausible records for tonnage of rice imported from 1995-2008 (~100 t in total per year for the first 4 years). Not addressing this concern would skew any analysis of food security for Solomon Islands. Similar concerns relating to rice data were observed for Papua New Guinea, Vanuatu, Fiji, and Samoa.
- Kiribati imported 7,864 t of salt in 2008, which equates to roughly 64 kg per capita per year. Significant volumes of salt are imported to many Pacific nations for tuna preservation, rather than for consumption. Overlooking the use of salt in tuna preservation would have drastic consequences for dietary analysis.
- Numerous records of sugar beet molasses exported from Fiji, including 45,329 t in 2007. Fiji is not a commercial producer of sugar beet, and retrade of the commodity is implausible given the context.
- Numerous implausible trades, when the context of individual PICTs (remoteness, population, dietary preferences etc.) are considered: 110 t of kidney beans (HS071333) from China to Kiribati in 2009; 246 t of vegetable fats and oils (HS151620) from Ecuador to Cook Islands in 2012. While most such errors would not be influential at regional and sub-regional scales of analyses, they could dramatically skew food and beverage trade analysis at the national scale, so could not be overlooked.

In addition to anomalous records, some commodity sets contained significant residual error not addressed by preceding stages. This error was particularly egregious for both rice and wheat products which are both central to food security and nutrition for the region. These commodities experience significant miss- and under-reporting for the region, partly caused by data suppression by exporting countries. To provide insight into these concerns, and how they were addressed, we elaborate on wheat and wheat milling products.

There are seven HS6 codes for wheat and wheat products, but trades are dominated by just two: HS100190 (meslin and wheat other than durum) and HS110100 (wheat or meslin flour). Together these two codes account for 94% of wheat and wheat product trade flows by volume (end of stage 4). The largest exporter, by tonnage, of wheat and wheat products to the region is Australia. The largest importer of wheat grain is PNG. Fiji, PNG, New Caledonia and Solomon Islands have flour mills and the majority of imports are wheat grain from Australia; the remainder of PICTs import mostly wheat flour. A number of PICTs import the bulk of their wheat products as milled flour from Fiji which was imported as grain from Australia. In addressing concerns with wheat data in stage 5 no adjustments are made to convert flour into grain equivalents, and 23tempts are made to control for any wheat refuse products that are likely to be used as animal feed.

Within the broad understanding of high regional dependency on wheat and derived products, and import of bulk grain to PICTs with commercial milling facilities, major concerns were observed within the data. For example, the database shows that import of wheat (HS100190) from Australia to Fiji was virtually non-existent until 2002, and then climbed annually from around 68,000 t in 2002 to 190,000 t in 2018. This was corroborated by Comtrade data. Australian Bureau of Statistics export data showed significant exports of HS100190 to Fiji from 1995 forward; roughly 69,000 t in 1995 up to 111,000 t in 2002. Substituting data on Australian exports of HS100190 with Australian Bureau of Statistics data provided a more plausible trend and had a dramatic effect on wheat import data for Fiji. Similar issues and substitutions were conducted for the other major importers of HS100190: New Caledonia, Papua New Guinea, and Solomon Islands. The updated PNG estimates were comparable to other analysis (M. R. Bourke & Harwood, 2009; citing Gibson, 2001), differing by an annual average of only 1.5% between 1997 and 2007. This similarity adds weight of evidence that these major adjustments were warranted and essential to improving the accuracy of the database. The major effects on these changes do also highlight, again, the concerns relating to global food trade datasets when conducting academic and policy analysis. Similar concerns were identified in rice data, which is detailed in Appendix 10.

Cleaning to the end of stage 4 broadly addressed error relating to incorrect commodity*PICT attribution (stages 2 and 3), and the majority of quantity error through the unit price imputation method employed in stage 4. One dimension of cleaning, not yet conducted, yet necessary for reliable country level analysis was to review HS6 quantities (t) through time, and by trade partnership. Specifically, experts reviewed matrices of HS6*year and HS6*trade partner country, for both imports and exports for each PICT (i.e. 4 matrices per PICT), across all commodity chapters, equating to roughly 1200 individual matrices (Figure 5).

Generally, Stage 5 methods were applied to all HS Chapters in the database, one chapter at a time, to enable comparison of quantities between similar commodities. However, in some instances it made more sense to review specific HS6 groups within, or across, chapters because they were more similar. For example different HS6 were aggregated to a) rice and rice flour, b) wheat and wheat flour, and c) other cereals and flours across Chapter 10 and 11. Additionally, some chapters such as HS02 (Meat) contained numerous HS6 codes. In such instances the chapter was disaggregated to smaller sets of HS6-level groupings with more similar characteristics. In the case of HS02, ‘bovine’, ‘sheep’, ‘poultry’, ‘swine’, and ‘other’ were reviewed separately (4 matrices per PICT for each). This detailed, high resolution review of data enabled observation of temporal changes in quantity imported and exported across all PICTs at HS6, and associations between PICTs and their trade partners, in terms of quantity at HS6. While requiring a significant time investment, it was necessary to make the data stable at high resolution commodity definition across PICTS to ensure the database useful for national food policy analysis.

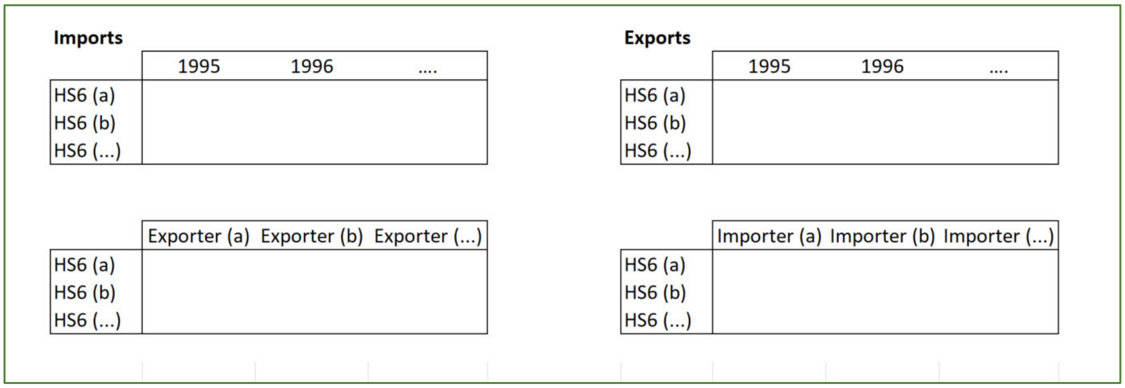


Figure 5. Matrix configuration constructed and reviewed for each PICT within each of the aggregated commodity sets. Each set of matrices were populated with quantities. Note, matrices included all HS6 within each commodity set, all years, and all trade partners, for both imports and exports.

Once all four matrices were constructed for each PICT within each commodity grouping, the quantities within the matrices were visually inspected. Commodity*year and commodity*trade partner quantities that appeared implausible, including as major anomalies in the time series, were subsequently flagged. Each of these flagged quantities were then tabulated for review (example in Table 3). Numerous attributes related to each data point were acquired to enable better informed decision making relating to the plausibility of each trade record. In particular, Comtrade data were reviewed to identify whether one or both countries reported the trade. This was essential in determining data plausibility due to the use of mirror data, by BACI, to generate trade quantities and values. Other sources such as national statistics records were also used, where possible, to verify records of concern. Grey literature estimates and other sources were also used in instances where official data was not deemed adequate to make an assessment. The sum of all sources of evidence were then reviewed to make a final determination on each trade flow of concern. Determinations varied including deletion from the database or changing the quantity to the quantity reported to Comtrade by one of the trade partners. In some instances, substitution data was used because neither the PFTD (stage 4) data or Comtrade data were considered robust. These commodity-specific changes are described in detail in Appendix 10.

Table 3. Example of a specific trade record (meat export from Austria to Cook Islands in 1999) that was reviewed in stage 5, including the record details and the final determination. [(a)] HS6 commodity code; [(b)] HS6 commodity definition; [(c)] Unique record identifier; [(d)] Whether, or not, the record was imputed during stage 4.

20120 [(a)]	Meat; bovine, cuts with bone in (excluding carcasses) [(b)]			PFTD stage 4 quantity		Comtrade HS92 quantity						
PFTD ID [(c)]	Year	Exporter	Auxillary dat	Australia	Other	Exporter reported	PICT Importer reported	Imputed? [(d)]	PFTD stage 5 change	PFTD stage 5 quantity	Change date	Notes
942	1999	Austria	n/a	n/a	72.5	72.5	no	no	delete	0	3/09/2021	

Additionally, some commodities were eliminated from the database because, with the acquisition of new commodity-specific information, they were determined to not be dominantly for human consumption. For example, all of HS01 (live animals) were removed because their quantities are small and they are primarily traded as breeding stock. Some commodities with volatile, and sometimes confusing, trends were reviewed by commodity-specific experts. Others were similarly reviewed because the commodity may have been traded for reasons other than human consumption. For example, salt, as outlined above, was excluded due to its role in tuna preservation. Instances of commodity deletion, and reasons for deletion, are outlined in Appendix 10.

Guiding the stage 5 review process were a number of principles that were applied to making changes to the database. First, and following earlier stages was the principle of plausibility – if any particular record was considered unlikely, but plausible, it was retained without adjustment. In terms of procedural structure we followed the principle of reviewing each set of commodities (e.g. rice, wheat, pork, bovine) as an isolated set of data to be cleaned to maximise the reliability of those data for future analyses focused on specific commodity sets.

— Differences in temporal trends among data sources and cleaning stages

To reveal the importance of cleaning food and beverage trade data for the Pacific we compared the quantity (t) of trade flows for commodities cleaned here with the same commodities contained within the BACI and Comtrade datasets (Figure 6). Importantly, differences between Comtrade and BACI estimates are likely to be dominated by the inclusion of mirror data in the BACI database. HS chapters 10 and 12 are shown independently of other data due to the dramatic differences between data sources. Both Comtrade and BACI data contained the suite of implausible rice trades between Australia and Papua New Guinea, which can be seen in the Cereals trend (Figure 6A). The cleaning process described here adjusted these implausible trades and shows a less volatile trend at the coarse resolution shown here. The dramatic volatility in Comtrade data for oil seeds and oleagic fruits (HS12, Figure 6B) is mostly driven by records of copra and palm nut and kernel. BACI methods ameliorated the observed volatility of Comtrade data. Similarly, Comtrade quantity estimates exhibited dramatic volatility across the remaining commodities (Figure 6C). This volatility is partly addressed by BACI, including increased mean estimates through the incorporation of mirror data. The outputs of the analysis presented here further smooths the data to reveal a stable trend, reflective of gradual increase in trade occurring for the region. At the resolution shown in Figure 6, it would not be possible to reliably produce food policy for the region using either Comtrade or BACI data.

Samoa and Tuvalu illustrate the different purposes of the cleaning process (Figure 7). In Samoa, the categorical cleaning in Stages 2 and 3 had little impact, but the imputation of outliers in plausible trade flows reduces the overall variability observed in the original data due to outstanding volumes of trade reported for the years 2001 to 2005 and 2009 and 2017. Without this correction, false conclusions would be formed on the overall quantity of trade in Samoa with potential implications on food trade policy. In Tuvalu, in contrast, the cleaning process that occurred during stage 2 and 3 removed a cluster of implausible exports to Sweden in 2001 and 2002. A diverse range of food types appeared in these trade flows, none of which were exported by Tuvalu in the period of the dataset. We note that the ISO code for Tuvalu (TUV) is similar to Turkey (TUR), a nation that does export a diverse range of dried fruits, nuts among other commodities. These errors were removed in Stage 2 of cleaning. The variability in the overall volume of trade was further adjusted through the correction of the outstanding quantities of beverages and tobacco, among other commodities reported after 2005. The notable difference between stage 4 and stage 5 for Samoa is primarily explained by the exclusion of baitfish and salt (see Appendix 10 for reasoning).

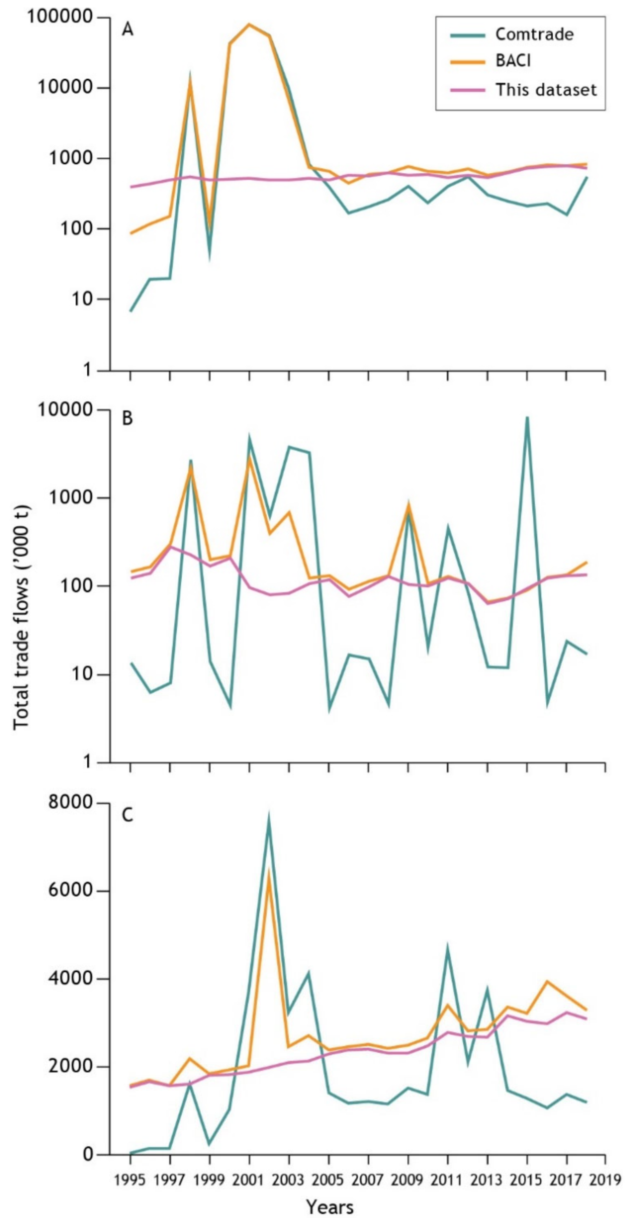


Figure 6. Selected examples of differences among databases in the quantity (t) of food traded (sum of imports and exports). UN Comtrade and CEPII-BACI are global public databases, 'This dataset' is the research database developed in this paper by cleaning the BACI database. **A)** HS Chapter 10 (Cereals), **B)** HS Chapter 12 (Oil seeds, oleagic fruits etc.), and **C)** All other HS Chapters included within this database. Comtrade data includes exports and imports reported by PICTs with the 'World'. Note log scale in Figures 6A and B.

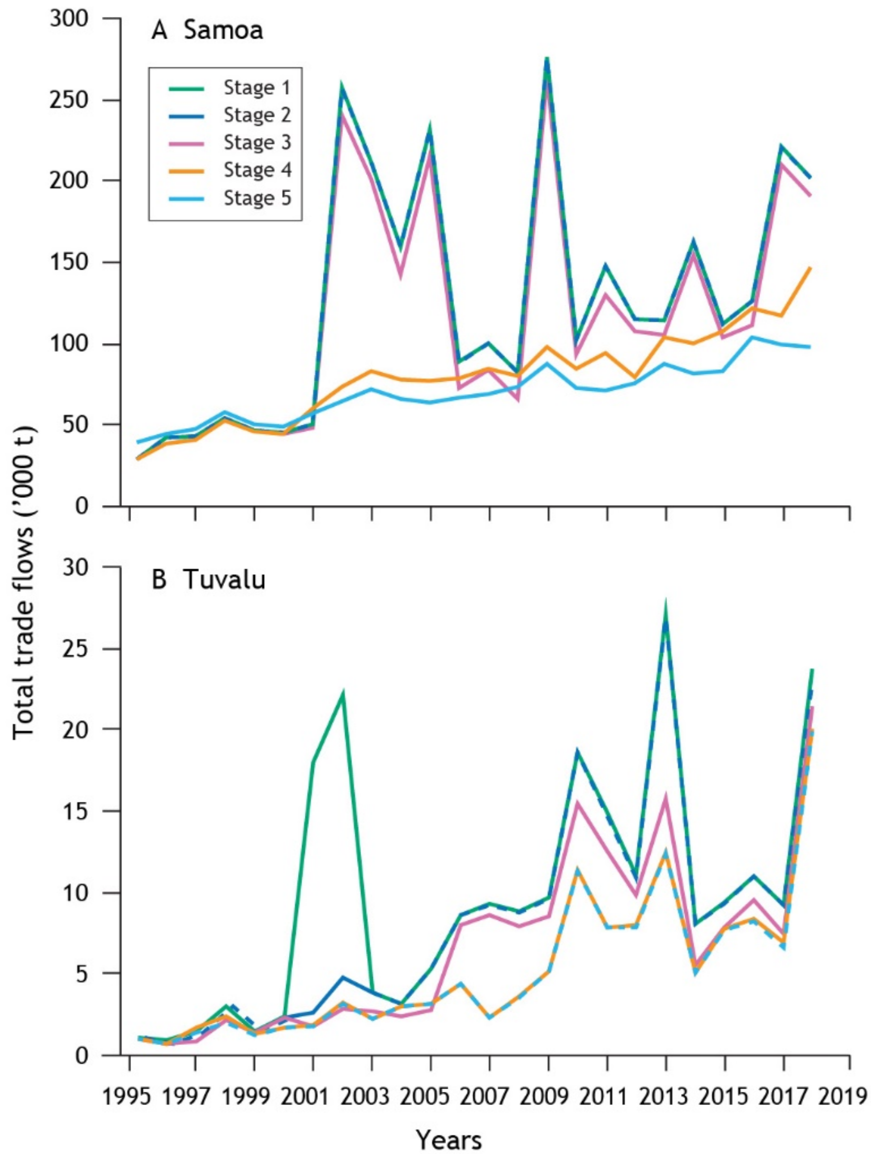


Figure 7. Total quantity (t) imported to, and exported from, Samoa and Tuvalu through time. Each line shows quantity trends at the end of each stage of either data preparation or cleaning. Stage 1 is BACI data, reduced to the relevant data set, as described in Stage 1. Note, the difference between stage 4 and stage 5, for Samoa, is primarily explained by commodities excluded during stage 5 cleaning.

— Effect of Stages 2 & 3 cleaning on PICT and commodity chapter data

The effect of data cleaning in stages 2 and 3 was not homogenous across PICTs (Table 4). The expert review of trade flows had the most dramatic effect on smaller PICTs, notably Niue and Tokelau, where national and regional experts could confidently judge the plausibility of exports. We assume implausible exports arose in the BACI dataset from incorrectly coded 'exporting' countries. Such trade flows were removed (Stage 2) because they were not relevant to the Pacific database. Samoa, Papua New Guinea, and particularly Fiji all re-export food commodities to smaller PICTs and so there was less confidence in judging exports to be implausible.

Table 4. Percentage of data [quantity (t), count of trade flows, and value] removed in cleaning stages 2 & 3. PICTs are ranked in decreasing order of percentage change in quantity.

PICT	Quantity (t)	Count	Value ($)
Niue	64.70	5.67	80.65
Tokelau	44.64	44.24	68.40
Tuvalu	17.97	7.08	22.35
Nauru	15.05	5.73	11.38
Solomon Islands	4.56	1.15	0.72
Vanuatu	3.96	1.47	1.02
Cook Islands	3.07	2.04	1.25
Palau	1.47	0.57	0.36
Tonga	1.72	0.44	0.66
FSM	0.89	0.62	0.23
Wallis and Futuna Islands	0.46	0.47	0.28
Marshall Islands	0.79	1.46	0.53
New Caledonia	0.37	0.21	0.39
Kiribati	0.13	1.18	0.28
French Polynesia	0.06	0.18	0.04
Samoa	0.06	0.24	0.13
Papua New Guinea	0.03	0.15	0.38
Fiji	0.00	0.06	0.00

There was significant variation in the effect of stage 2 and 3 cleaning on commodity groups (Table 5). For example, more than half of the quantity of tobacco was removed as being implausible (3.14% of total trade flows of tobacco). The great majority of these were recorded as exports from PICTs to other countries identified in stage 2. For other commodity Chapters, less than 10% of the quantity traded was removed across other commodity chapters. While this percentage appears trivial, it could be highly influential for analysis of commodity Sub-headings within single PICTs.

Table 5. Percentage of data [quantity (t), count of trade flows, and value ($)] removed in cleaning stages 2 & 3 by Commodity Chapter. Chapters are ranked in decreasing order of percentage change in quantity.

Commodity Chapter (HS2)	Quantity (t)	Count	Value ($)
Tobacco and manufactured tobacco substitutes (HS24)	59.48	3.14	11.90
Edible fruit, nuts, peel of citrus fruit, melons (HS08)	6.27	2.56	5.57
Live animals (HS01)	5.51	3.39	4.05
Milling products, malt, starches, inulin, wheat glute (HS11)	1.76	1.18	1.24
Vegetable, fruit, nut, etc food preparations (HS20)	1.41	0.62	1.40
Dairy products, eggs, honey, edible animal products nes (HS04)	1.11	1.23	0.94
Coffee, tea, mate and spices (HS09)	0.96	0.81	1.76
Edible vegetables and certain roots and tubers (HS07)	0.94	1.17	0.68
Salt, sulphur, earth, stone, plaster, lime and cement (HS25)	0.88	2.32	0.94
Sugars and sugar confectionery (HS17)	0.71	0.73	0.74
Oil seed, oleagic fruits, grain, seed, fruit, etc, nes (HS12)	0.68	3.12	2.15
Cereal, flour, starch, milk preparations and products (HS19)	0.66	0.65	0.43
Animal, vegetable fats and oils, cleavage products, et (HS15)	0.47	1.30	0.17
Meat and edible meat offal (HS02)	0.29	0.63	0.14
Miscellaneous edible preparations (HS21)	0.22	0.84	0.15
Cocoa and cocoa preparations (HS18)	0.20	0.87	2.69
Beverages, spirits and vinegar (HS22)	0.19	0.92	0.29
Meat, fish and seafood food preparations nes (HS16)	0.12	0.51	0.22
Fish, crustaceans, molluscs, aquatic invertebrates ne (HS03)	0.10	0.29	0.11
Cereals (HS10)	0.04	1.42	0.61

— Effect of Stage 4 cleaning on quantity data

In contrast to the categorical cleaning based on expert elicitation in stages 2 and 3, the imputation process was most impactful in larger PICTs, notably Papua New Guinea and Samoa (Table 6). The most effected cells were rice in Papua New Guinea, beverages in Tuvalu, and sugar in Marshall Islands. In almost all cases of large (>20%) change the imputation reduced the quantity in the trade flow; the exception was a 29% increase in miscellaneous food preparation in Papua New Guinea.

Table 6. Percentage change (+/-) in quantity (t) by PICT and HS2 resulting from revised quantities based on imputation of median unit price. PICTs are ranked from left to right and HS chapters from top to bottom in decreasing order of total absolute value of change. Cells with >20% change are highlighted in yellow. n.d. indicates no data. See Table 5 for chapter inclusions and Appendix 11 for complete list of commodity exclusions.

Chapter	Papua New Guinea	Samoa	Kiribati	Tuvalu	Marshall Islands	Wallis and Futuna Islands	Tokelau	Fiji	Solomon Islands	Vanuatu	New Caledonia	Palau	Nauru	French Polynesia	Tonga	Niue	Cook Islands	FSM
HS22	-5.6	-5.7	-18.7	-91.2	-82.4	0.2	-47.3	-0.2	-11.6	-8.5	-3.3	-9.9	-0.9	-1.7	-24.6	6.0	0.9	0.1
HS24	-1.8	-2.4	-19.3	-68.8	-1.3	-46.9	0.0	-48.9	-37.0	-1.9	-6.9	-0.7	-5.9	0.0	-0.1	0.0	0.1	-1.0
HS01	-27.3	15.2	0.0	0.0	0.0	-92.5	0.0	0.0	-0.1	-1.3	-15.3	0.0	0.0	-58.8	0.3	0.0	0.0	0.0
HS10	-96.9	-80.9	0.1	0.2	0.1	0.0	0.0	-5.4	-4.5	0.1	-0.1	0.2	5.7	0.1	2.0	0.0	10.1	0.0
HS17	-39.5	-58.1	-1.4	-2.1	-86.0	0.2	9.8	1.3	-1.7	0.2	-0.4	-1.0	0.1	0.0	-0.1	0.5	0.7	0.0
HS18	-78.8	-11.0	-28.8	0.0	0.7	-0.8	0.4	-0.4	0.9	-47.9	0.0	-22.2	-0.9	0.2	-4.9	2.8	0.9	0.8
HS11	-2.6	-29.1	-48.3	-14.6	0.0	-0.9	5.8	-16.8	2.1	-9.8	0.1	0.2	0.1	0.0	2.2	0.4	-1.5	0.1
HS12	-1.1	-0.2	-87.8	0.0	0.0	0.0	0.0	-4.2	1.4	-0.1	-0.2	6.2	-18.7	0.0	0.1	0.0	8.8	0.0
HS21	29.2	-5.3	-0.5	-5.5	0.0	-0.2	13.2	-16.4	-1.2	0.1	-20.3	-5.8	0.4	0.1	0.2	2.0	-0.7	0.4
HS04	-6.9	3.9	2.5	-34.9	1.7	13.8	5.8	0.8	2.5	0.6	-5.1	1.3	-4.8	-2.3	4.3	5.1	1.9	2.4
HS08	2.3	-26.3	2.0	0.3	-8.8	-1.7	-27.6	-1.3	-0.3	2.5	-3.0	-1.1	-3.3	0.0	-3.5	4.3	-1.5	2.0
HS07	-23.3	-7.9	-25.7	-0.6	0.1	1.4	3.4	-1.3	5.6	2.1	-2.2	-1.2	1.4	0.5	-5.4	-0.5	2.5	0.0
HS16	0.1	13.4	-1.5	1.7	-0.1	4.1	2.4	14.3	-25.2	0.0	0.0	0.8	-0.1	-1.3	10.1	1.6	-0.1	2.6
HS09	-0.2	-9.5	-0.4	0.0	-10.5	0.2	-7.4	-0.9	-0.2	-2.8	1.8	-8.8	17.0	2.2	-5.9	3.5	4.7	1.5
HS03	-25.1	-5.0	-9.5	-0.2	-0.2	0.0	0.0	-3.1	-0.3	-0.1	-23.2	-2.3	-0.3	0.8	-0.2	2.5	-0.2	-0.3
HS15	0.8	0.9	-14.2	0.1	-1.0	0.2	15.4	-9.7	1.7	0.2	0.0	2.9	-0.1	0.5	-0.4	9.6	1.6	1.6
HS19	-4.0	0.1	-8.1	-10.2	-2.1	-0.6	4.8	-13.8	0.6	-3.8	-0.4	-3.8	-2.4	0.0	-1.1	2.2	1.4	-1.0
HS25	0.4	-0.9	2.0	0.0	16.6	0.0	n.d.	1.1	0.2	-6.9	0.5	0.5	7.4	0.1	0.3	0.0	0.4	3.1
HS20	-0.1	-0.1	0.6	1.8	-7.4	0.3	-4.4	1.0	-4.9	-0.4	-9.2	-2.3	0.1	0.7	0.2	-1.0	-1.9	3.3
HS02	-0.2	-10.6	3.8	-0.2	0.0	0.0	0.0	-0.2	2.2	-6.4	0.0	0.1	0.0	0.0	0.2	-0.5	0.9	0.0

— Effect of Stage 5 cleaning on commodity, chapter and year data across PICTs

Stage 5 cleaning reduced the total quantity by 8% across the database. This decrease was primarily explained by the complete removal of some commodities including baitfish, salt, and palm nut and kernel (see Appendix 10 for rationale for exclusion and Appendix 11 for comprehensive list of commodities excluded from initial data download to final end of stage 5 database). Not including the commodities that were completely removed in stage 5 the total quantity increased by 1.7%. This small net increase in total quantity masks the significant increases and decreases in quantity that have occurred for individual PICTs across Chapters and through time (Tables 7, 8), reinforcing the importance of this stage of cleaning for country-level food and beverage analysis. While most changes in stage 5 resulted in deletion of records, this was counteracted by the addition of a significant quantity of rice data across Papua New Guinea, Solomon Islands, Vanuatu, Fiji and Western Samoa (Tables 7, 8; Appendix 10). A large proportion of HS03 (seafood) was removed from the database because three HS Sub-headings (HS030371, HS030374, HS030749) were deemed to be dominantly imported for bait in the purse seine tuna fishery (Appendix 10). Smaller PICTs including Tokelau, Nauru, Kiribati, Niue, and Wallis and Futuna were more influenced across commodity Chapters by this stage of cleaning (Table 8), further highlighting the importance of this stage.

Table 7. Percentage change (+/-) in quantity (t) by PICT and HS2 between end of Stage 4 methods and end of Stage 5 methods. PICTs are ranked from left to right and HS chapters from top to bottom in decreasing order of total absolute value of change. Cells with >20% change are highlighted in yellow. 'Total % change' is the total % change in quantity for each HS Chapter resulting from Stage 5 cleaning. 'Total' is the overall change in quantity within the PFTD resulting from Stage 5 cleaning.

Chapter	Tokelau	Nauru	Samoa	Kiribati	FSM	Palau	Niue	Wallis and Futuna Islands	Cook Islands	Marshall Islands	Solomon Islands	Tonga	Tuvalu	Papua New Guinea	Fiji	French Polynesia	New Caledonia	Vanuatu	Total % change
HS03	-62.4	-61.7	-35.3	-61.2	-37.3	-27.0	-37.4	-6.9	-27.6	-2.0	-57.1	-10.8	-4.0	-14.4	-21.3	-0.2	-4.4	-29.7	-23.4
HS10	0.0	-2.6	-37.6	-10.0	-45.4	-1.4	0.0	3.1	-30.8	-63.0	58.8	-45.1	-4.4	17.4	18.7	-0.1	13.6	136.9	14.5
HS09	-89.7	-58.1	-16.0	-23.3	-24.9	0.0	-44.8	-57.5	-3.8	-12.6	-42.6	-15.4	0.0	-7.8	-0.3	-0.1	-4.0	-41.3	-8.3
HS24	-37.1	-24.7	-23.6	-24.8	-5.5	-14.9	-0.7	0.0	-15.3	0.0	-10.4	-1.9	-8.9	-2.0	-10.8	-0.8	-3.0	-2.9	-6.3
HS17	0.0	0.0	-30.7	-17.2	-18.8	-37.0	0.0	-26.1	0.0	-0.0	-5.6	-8.4	-7.0	-3.6	-1.1	-5.0	0.0	-11.0	-2.5
HS02	-89.2	-3.3	-0.5	-12.1	-12.9	-20.0	-4.9	-0.0	-1.2	0.0	-0.3	-1.7	-0.4	-0.5	-0.6	-0.2	-1.3	-1.8	-1.4
HS12	-80.4	-14.7	-0.4	0.0	-3.8	0.0	0.0	0.0	0.0	0.0	0.0	0.0	-15.8	-0.2	-0.0	0.0	0.0	-0.0	-0.1
HS20	-63.5	-19.5	-0.8	-0.8	-3.5	0.0	-0.5	-0.3	-1.4	-0.3	-1.9	0.8	-3.9	-8.5	-0.3	0.0	-0.2	-0.6	-1.6
HS18	-26.9	-21.3	-2.4	0.0	0.0	0.0	-21.9	0.0	-1.8	0.0	-0.2	-0.4	-17.5	-0.0	-0.4	-0.6	-0.7	-0.3	-0.1
HS07	-23.9	-6.7	-6.5	-3.6	-1.9	-3.5	0.0	-1.5	-0.1	-9.1	-18.5	0.0	0.0	-2.3	-0.5	0.0	0.0	-16.2	-1.1
HS04	-37.6	0.0	-2.1	-0.3	-1.0	-0.1	-1.1	-10.6	-1.0	-1.7	-2.5	-0.1	0.0	-0.2	-0.5	-1.6	-1.2	-0.4	-1.0
HS22	-28.6	0.0	-1.0	-5.2	-0.5	-3.8	0.0	0.0	-0.1	0.0	-0.7	-0.1	0.0	-1.5	-6.8	0.0	-0.5	-3.5	-1.7
HS08	-35.0	0.0	-2.0	0.0	0.0	-4.2	-2.2	-0.7	-1.8	0.0	0.0	-1.9	-0.3	-0.2	-0.0	-1.7	-0.9	-0.9	-0.9
HS11	0.0	0.0	-4.8	-2.0	0.0	0.0	-4.9	-10.9	-10.2	-3.8	-4.9	-0.5	0.0	-4.8	-2.3	-0.2	-0.1	-0.3	-2.5
HS16	-10.2	-3.8	-1.3	-1.5	-0.1	-11.3	0.0	-0.0	0.0	0.0	0.0	-0.4	0.0	-0.4	-1.2	-0.1	-0.1	-0.1	-0.7
HS15	-11.3	-1.5	0.0	0.0	-1.7	0.0	0.0	-3.2	-4.4	-3.5	0.0	0.0	0.0	-2.1	-0.4	-0.2	0.0	0.0	-1.9
HS19	-9.4	-2.0	0.0	0.0	-3.3	-3.4	0.0	-0.8	0.0	-0.5	-3.0	-0.0	0.0	-2.2	0.0	0.0	-0.4	-0.2	-0.9
HS21	-10.0	0.0	-0.0	0.0	-0.3	0.0	0.0	-0.5	-0.1	0.0	0.0	0.0	0.0	-0.9	-0.0	0.0	-0.0	-1.2	-0.3
Total																			1.7

Table 8. Percentage change (+/-) in quantity (t) by PICT and YEAR between end of Stage 4 methods and end of Stage 5 methods. PICTs are ranked from left to right in decreasing order of total absolute value of change. Cells with >20% change are highlighted in yellow. 'Total % change' is the total % change in quantity for each HS Chapter resulting from Stage 5 cleaning.

Year	Tokelau	Solomon Islands	Samoa	FSM	Vanuatu	Papua New Guinea	Marshall Islands	Nauru	Kiribati	New Caledonia	Cook Islands	Fiji	Palau	Wallis and Futuna Islands	Niue	Tuvalu	Tonga	French Polynesia	Total % change
1995	-100.0	58.9	39.3	0.0	23.1	36.5	0.0	-0.1	-6.4	25.6	-13.3	11.0	0.0	0.0	-6.6	0.0	-2.4	0.0	21.7
1996	0.0	51.2	21.6	0.0	24.8	43.0	-0.3	-0.6	-10.5	16.6	-0.3	9.4	-4.1	0.0	-19.9	0.0	0.0	-0.0	22.3
1997	-27.4	45.5	16.9	0.0	10.6	42.6	-0.2	-1.3	-3.1	23.9	0.0	11.2	-3.8	-0.4	-15.7	-17.6	-1.6	-0.8	23.2
1998	-72.9	47.2	13.3	0.0	14.1	8.7	0.0	-3.3	-3.3	21.9	-0.4	14.1	-3.4	-1.6	-8.7	-15.5	-9.0	-0.0	11.0
1999	-100.0	83.8	14.7	0.0	17.5	49.0	-68.7	-2.0	-2.0	5.3	-0.9	10.8	-1.7	0.0	0.0	-1.4	-3.4	-0.0	22.3
2000	-97.5	94.1	14.9	-0.2	23.4	8.0	0.0	-8.8	-0.2	1.0	-0.5	14.4	-7.0	-8.6	-0.1	0.0	-0.7	-0.8	9.9
2001	-55.2	25.1	-1.6	-0.8	32.8	16.8	-42.5	-6.2	-0.2	-1.0	-2.0	18.1	0.0	-7.0	-0.8	-1.1	-3.0	-0.0	13.2
2002	-37.3	119.7	-7.8	-5.0	37.0	2.3	-0.0	-5.3	1.4	0.9	-0.1	6.2	-3.6	-7.7	0.0	-1.2	-0.1	-0.0	4.2
2003	-58.4	26.2	-12.5	-29.9	17.7	-4.8	0.0	-7.0	2.1	-9.4	-4.4	6.0	-2.3	-6.8	0.0	0.0	-2.3	-0.7	-1.6
2004	-33.7	38.4	-13.4	-34.0	16.9	-9.9	0.0	-0.2	-2.5	4.2	-13.8	-2.2	0.0	-7.5	0.0	0.0	-0.7	-0.3	-4.9
2005	-48.3	49.0	-17.2	-34.3	-1.0	-5.6	-0.4	-1.9	-2.1	4.6	-12.9	0.1	0.0	-13.0	-3.6	0.0	-0.2	-0.0	-2.2
2006	-45.2	48.0	-13.0	-40.3	-3.3	15.4	0.0	-22.8	-0.3	4.5	-8.9	-0.7	-1.9	-9.8	0.0	0.0	-0.2	-0.0	5.8
2007	-62.0	42.5	-16.0	-43.2	-2.6	-1.6	0.0	-25.0	-0.7	2.6	-4.1	0.7	-4.1	0.0	-1.5	0.0	-0.2	0.0	-0.5
2008	-17.1	38.5	-7.0	-43.1	18.2	-1.0	-0.5	-33.0	-4.9	3.9	-5.7	1.7	-7.9	0.0	-1.2	0.0	-1.5	-0.0	0.7
2009	-57.1	2.6	-8.8	-18.3	-2.0	-0.5	-0.9	-30.2	-26.1	3.6	-4.0	-3.0	-10.2	0.0	-5.6	0.0	-1.2	-0.6	-1.9
2010	-60.4	22.4	-12.0	-15.1	-1.6	-4.9	-12.0	-3.4	-7.4	4.1	-0.9	-0.4	-7.5	-0.0	0.0	-1.7	-2.3	-0.8	-2.3
2011	-15.8	1.8	-22.1	-20.4	-2.5	-23.3	-0.0	-3.8	-11.2	3.6	-16.8	-5.3	-5.6	-1.3	0.0	-0.3	-2.0	-5.1	-14.9
2012	-49.9	0.5	1.4	-18.8	27.0	-13.3	-0.7	-1.3	-8.7	-0.5	-2.0	-1.1	-4.7	0.0	-5.7	-2.0	-2.2	-0.7	-7.8
2013	-57.9	2.7	-13.4	-25.3	4.1	-0.0	-0.5	-0.7	-9.1	-0.1	-6.0	-0.6	-2.3	0.0	0.0	-1.4	-2.7	0.0	-1.0
2014	-28.0	-0.9	-16.4	-30.5	18.4	-0.0	-41.4	-2.8	-15.2	-0.0	-7.9	-4.8	-31.3	0.0	0.0	-1.8	-2.7	0.0	-2.8
2015	0.0	-2.1	-21.9	-0.5	18.5	-0.1	-0.9	-1.6	-10.3	-0.1	-8.9	-2.0	-1.7	-0.1	-4.1	-1.5	-0.3	-0.1	-1.1
2016	-5.3	-11.1	-12.0	0.0	-8.1	-0.1	-1.1	-1.0	-37.0	-0.1	-10.8	-2.6	-1.2	0.0	-1.8	-1.2	-0.1	-0.5	-2.4
2017	-0.3	-4.4	-13.3	0.0	-0.8	-0.1	0.0	-0.6	-0.2	0.0	-9.9	0.1	-1.8	-8.4	0.0	-7.4	0.0	-0.7	-0.8
2018	-49.4	-10.9	-31.7	0.0	-0.5	-0.1	-9.4	-9.0	-3.4	0.0	-0.0	-0.8	-1.4	-8.7	-0.0	1.8	0.3	-1.0	-2.5

DISCUSSION

Comtrade is the most important and authoritative source of international trade statistics. Many third party databases use Comtrade data in some form, often through web interfaces that simplify access and queries. All these databases are, ultimately, reliant on data reported by countries. Although the United Nations Statistics Division (UNSD) employs a range of measures to standardize data and identify statistical outliers (e.g. Food and Agriculture Organisation, United Nations Statistics Division, International Trade Centre, 2019) many errors remain. UNSD, CEPII-BACI and others provide explicit caveats on the use of the data, and caution against over-interpretation of results. The results of our analysis confirm that great care should be taken in drawing conclusions from uncleaned Comtrade and derivative databases. For the 18 small countries and territories of the Pacific region we conclude that uncleaned Comtrade-derived data does not provide a sufficiently robust foundation for analysis of international food trade.

The process of expert elicitation to identify error at the coarse (stages 2 & 3) and granular (stage 5) levels were each essential to eliminating significant amounts of error. In particular, the investment in stage 5 makes the data useable for observing temporal trends in key sets of commodities within individual PICTs. Similarly, imputation (stage 4) significantly improved the data, particularly for some PICTs and some commodities (Figures 6, 7; Table 6). Despite the significant investment required to develop this database, the output data compared to the input data show that, particularly for small nations, this cleaning was essential prior to conducting any analysis of food and beverage trade for the Pacific.

The broader purpose of conducting this research was to identify the magnitude and distribution of error in trade data and provide a method that can be applied to improve data quality for analyses. Changes to data across PICTs and commodities were dramatic and consequential to estimates of trade flows and, by inference, estimates of the availability of food in national and regional food systems. In addition to a relatively small number of egregious errors, with either absurd quantities or implausible exporters or importers, there were thousands of more minor errors that, in sum would be influential for analyses at more granular scales.

The input of experts proved to be essential for the elimination of a vast number of export records (Stage 2.1) determined to be implausible, and crucial for identifying blocks of implausible trade flows such as observed between Tuvalu and Sweden. It was essential that this process (Stages 2 and 3) was conducted prior to the imputation or a significant number of records would have been used to estimate median unit prices, which would have compounded the existing error. Second, the structure used to generate the median unit prices associated with the identified outliers (stage 4.2, method 8) controls for natural variation in unit price that exists between commodities, countries, and years, thus delivering the most similar median unit prices to the identified outlier. This approach, compared to using only HS Sub-headings to estimate medians, as is common practise, would have the most dramatic effect on estimates where there is significant variation in unit prices between countries and years.

The method developed is applicable to any set of international trade data, including beyond food and beverages, but is particularly relevant to small countries with relatively simple export/import dynamics. For example, the described method could be applied to the 37 recognised Small Island Developing States (SIDS) outside the Pacific. SIDS tend to have growing populations, limited food resources, are prone to natural disasters and other external shocks, and are heavily reliant on food imports. Reliable trade data is essential to SIDS in ensuring food security and understanding the role of trade in diet related health outcomes, including non-communicable diseases. Similarly, the method could have application in global regions dominated by low-middle income countries such as Africa and Central America.

The data generated in this study represent an important new opportunity for food policy research in the Pacific region. In the Pacific, net food import dependence follows relatively recent and rapid trade liberalization, which has been associated with significant changes in diet (Thow & Snowdon, 2010). Diet-related non-communicable diseases now represent a significant social and economic burden in the region (Popkin, Corvalan, & Grummer-Strawn, 2020). Reliable data are critical to be able to monitor the impact of trade agreements on food environments and nutrition in the Pacific, and to develop effective, targeted policy responses (Ravuvu, Friel, Thow, Snowdon, & Wate, 2017). This includes analysis of specific trade agreements (e.g. PACER+), as well as more general trade and economic bi- and multi-lateral initiatives within the region (e.g. Pacific Step-up or PACHS17). Early analysis using the database show clear regional and sub-regional trends in the evolving role of trade in the Pacific food system (Andrew et al., 2022).

This study also has implications globally, in terms of supporting improved availability of high-quality data to inform policy priorities. First, from a trade policy perspective, improved data quality enables the assessment and monitoring of the impact of policy changes on trade flows. For example, assessing the impact of tariff changes on the balance of (food) trade for important economic sectors such as agriculture and industry. Recent research has demonstrated the value of this for both prospective and retrospective analyses of trade and investment agreements (Kawasaki, 2018; Sahu, 2019). Second, from a food policy perspective such data are useful across sectors governing food and food systems, to assess multiple aspects of food related trade, including the balance of trade in high (economic) value compared to low value foods, or the sustainability implications of food trade (Béné et al., 2019). Finally, from a nutrition policy perspective, such data enable more rigorous analysis of the potential impacts of trade on changing diets, which are linked to health concerns globally (Friel, Schram, & Townsend, 2020). Recent research from Central America provides further evidence for the long-term impacts of trade liberalization on diets, and access to healthy food, drawing on multiple sources of trade data. Development of reliable food trade databases can make a critical contribution to the design of effective food policy to mitigate potentially negative impacts from trade (Werner, Contreras, Mui, & Stokes-Ramos, 2019).

The database also contains minor structural limitations carried over from Comtrade and BACI. First, Comtrade and BACI do not report trade flows valued at less than US$1000. It is therefore likely that frequent, but small trades to PICTs and between PICTs are not reported. This non-reporting includes small volume 'shuttle trades', which occurs frequently throughout the Pacific, including on aircraft and smaller vessels. It is not possible to reliably estimate shuttle trade quantities though it is likely to be meaningful for some trade partnerships. Second, because some PICTs do not report trade flows to UNSD, the database will not include records of trade between any two non-reporting PICTs. The volume of these unreported trades is not likely to be large, however, because most non-reporting PICTs are geographically isolated and conduct most of their trade through larger reporting PICTs such as Samoa and Fiji. Third, Comtrade is updated as data is made available from reporting countries, and BACI acquires Comtrade data periodically. Consequently, it is likely that for the most recent years in the dataset, not all trade records were available at the time BACI acquired Comtrade data and any decline in trade value, volume or frequency in the most recent years should be treated with caution. Finally, caution must be used in interpreting patterns in cleaned data because some countries do not consistently provide trade records to UN Comtrade for particular commodities (e.g. rice and wheat exported from Australia) where there are commercial sensitivities. However, the application of mirror data by BACI is likely to largely ameliorate this potential error source.

ACKNOWLEDGEMENTS

We are indebted to in-country experts from PICTs for their contribution to stage 2.1 of the cleaning process. The following experts generously shared their knowledge of national trade flows: Enua Pakitoa, Pritesh Gounder, Kemueli Naiqama, Sharon L. Pelep, Aritita Tekaieti, Tiriara Ikam, Ipia Gadabu, Josie Tamate, Sition Suvulo, Louisa Baragamu, Reuben Sulu, Kele Lui, Anna Wells, Daniella Woiala, Mark Vurobaravu and Aliimuamua Malaefono Taua. We are thankful to commodity experts including, K. Hutton (rice), M. McCoy (baitfish, tuna, salt), J. Kinch (PNG fisheries), M. Amos (fish), E. Notere (Solomon Islands copra and palm oil), W. Sokimi (baitfish) for sharing their knowledge.

We are grateful to the team at CEPII BACI for fielding queries as the described method was developed. D. Abbott and N. Lal of SPC provided invaluable expertise in identifying the plausibility of trade flows and provided general expertise in this research. We are grateful to A. Romeo for discussion in the early stages of this work. N. Troubat and A. Dunn provided insightful and extensive feedback on the manuscript. We are grateful to Ethan Vu for his invaluable help in summarizing and checking data and Elle McNeill for graphics. This work was supported by The Australian Government through ACIAR projects FIS/2016/300 and FIS/2018/155.

— Disclaimer

The process described here and the resultant database were developed as a research tool and does not constitute an official record of trade flows with and among countries in the Pacific region. The views expressed are those of the authors and not their respective institutions.

REFERENCES

Aggarwal, C. C. (2017). An introduction to outlier analysis. In Outlier analysis (pp. 1-34): Springer.

Andrew, N. L., Allison, E. H., Brewer, T., Connell, J., Eriksson, H., Eurich, J. G., Farmery, A., Gephart, J., Golden, C., Herrero, M. (2022). Continuity and change in the contemporary Pacific food system. Global Food Security, 32, 100608.

Baker, P., Friel, S., Schram, A., Labonte, R. (2016). Trade and investment liberalization, food systems change and highly processed food consumption: a natural experiment contrasting the soft-drink markets of Peru and Bolivia. Globalization and Health, 12(1), 24.

Béné, C., Oosterveer, P., Lamotte, L., Brouwer, I. D., de Haan, S., Prager, S. D., Talsma, E.F., Khoury, C. K. (2019). When food systems meet sustainability–Current narratives and implications for actions. World Development, 113, 116-130.

Bourke, M. R., & Harwood, T. (Eds.). (2009). Food and Agriculture in Papua New Guinea. The Australian National University, Canberra: ANU E Press.

Bourke, R., McGregor, A., Allen, M., Evans, B., Mullen, B., Pollard, A., Wairiu, M., Zotalis, S. (2006). Solomon Islands Smallholder Agriculture Study. Volume 1. Main findings and recommendations.

Brooks, J., & Matthews, A. (2015). Trade dimensions of food security. OECD Food, Agriculture and Fisheries Papers, No. 77, OECD Publishing, Paris,. Retrieved from https://doi.org/10.1787/5js65xn790nv-en.

Bruffaerts, C., Verardi, V., Vermandele, C. (2014). A generalized boxplot for skewed and heavy-tailed distributions. Statistics & probability letters, 95, 110-117.

Estimé, M. S., Lutz, B., Strobel, F. (2014). Trade as a structural driver of dietary risk factors for noncommunicable diseases in the Pacific: an analysis of household income and expenditure survey data. Globalization and Health, 10(1), 48.

Food and Agriculture Organisation, United Nations Statistics Division, International Trade Centre (2019). Trade data processing and validation methodology. United Nations Department of Economic and Social Affairs, statistics division, economic statistics branch.

Farrell, P., Thow, A., Tutuo Wate, J., Nonga, N., Vatucawaqa, P., Brewer, T., Sharp, M., Farmery, A., Trevena, H., Andrew, N. (2020). COVID-19 and Pacific food system resilience: opportunities to build a robust response. Food Security, 12(4), 783-791.

Friel, S., Schram, A., Townsend, B. (2020). The nexus between international trade, food systems, malnutrition and climate change. Nature Food, 1(1), 51-58.

García-Dorado, S. C., Cornselsen, L., Smith, R., Walls, H. (2019). Economic globalization, nutrition and health: a review of quantitative evidence. Globalization and Health, 15(1), 15.

Gaulier, G., & Zignago, S. (2010). Baci: international trade database at the product-level (the 1994-2007 version).

Gewertz, D. B., Errington, F. K. (2010). Cheap meat: flap food nations in the Pacific Islands: University of California Press.

Geyik, O., Hadjikakou, M., Karapinar, B., Bryan, B. A. (2021). Does global food trade close the dietary nutrient gap for the world's poorest nations? Global Food Security, 28, 100490.

Gibson, J. (2001). Food security and food policy in Papua New Guinea. Institute of National Affairs, Port Moresby, INA Discussion Paper No. 83

Gillson, I., Fouad, A. (2014). Trade policy and food security: improving access to food in developing countries in the wake of high world prices. The World Bank.

Hawkes, C. (2010). The influence of trade liberalization and global dietary change: The case of vegetable oils, meat and highly processed foods. Trade, food, diet and health: Perspectives and policy options, 35, 59.

Hubert, M., Vandervieren, E. (2008). An adjusted boxplot for skewed distributions. Computational statistics & data analysis, 52(12), 5186-5201.

Jones, A. R. (2018). Probability, statistics and other frightening stuff (Vol. 2): Routledge.

Kawasaki, K. (2018). Economic Impact of Tariff Hikes-A CGE model analysis. Discussion paper No 18-05. National Graduate Institute for Policy Studies.

Kerr, W. A. (2011). The role of international trade in achieving food security. Estey Journal of International Law and Trade Policy, 12(2), 44-53.

Ortiz-Ospina, E., Beltekian, D., Roser, M. (2018). Trade and Globalization. Our World In Data. Retrieved from https://ourworldindata.org/trade-and-globalization

Popkin, B. M., Corvalan, C., Grummer-Strawn, L. M. (2020). Dynamics of the double burden of malnutrition and the changing nutrition reality. The Lancet, 395(10217), 65-74.

Pyakuryal, B., Roy, D., Thapa, Y. (2010). Trade liberalization and food security in Nepal. Food Policy, 35(1), 20-31.

Ravuvu, A., Friel, S., Thow, A. M., Snowdon, W., Wate, J. (2017). Monitoring the impact of trade agreements on national food environments: trade imports and population nutrition risks in Fiji. Globalization and Health, 13(1), 33.

Sahu, P. K. (2019). Tariff elimination under the trans-Pacific partnership and its impact on Indonesia's trade balance. Journal of Indonesian Economy and Business, 34(1), 16-33.

Schmidt, E., Fang, P. (2021). Papua New Guinea Agri-food Trade Trends. Dietary Change and Obesity. IFPRI Discussion Paper 02028. International Food Policy Research Institute

Seo, S. (2006). A review and comparison of methods for detecting outliers in univariate data sets. Masters Thesis, University of Pittsburgh.

Thow, A. M. (2009). Trade liberalisation and the nutrition transition: mapping the pathways for public health nutritionists. Public health nutrition, 12(11), 2150-2158.

Thow, A. M., Heywood, P., Schultz, J., Quested, C., Jan, S., Colagiuri, S. (2011). Trade and the Nutrition Transition: Strengthening Policy for Health in the Pacific. Ecology of Food and Nutrition, 50(1), 18-42.

Thow, A. M., Sanders, D., Drury, E., Puoane, T., Chowdhury, S. N., Tsolekile, L., Negin, J. (2015). Regional trade and the nutrition transition: opportunities to strengthen NCD prevention policy in the Southern African Development Community. Global health action, 8(1), 28338.

Thow, A. M., Snowdon, W. (2010). The effect of trade and trade policy on diet and health in the Pacific Islands. Trade, food, diet and health: Perspectives and policy options, 147, 168.

Tukey, J. W. (1977). Exploratory data analysis (Vol. 2): Reading, Mass.

United Nations (2002). Central Product Classification (CPC). Department of Economics and Social affairs, Statistics Division. Series M, NO. 77, Version 1.1.

United Nations (2008). International Merchandise Trade Statistics: Supplement to the Compilers Manual. Department of Economic and Social Affairs, Statistics Division. ST/ESA/STAT/SER.F/87/Add.1

Werner, M., Contreras, P. I., Mui, Y., Stokes-Ramos, H. (2019). International trade and the neoliberal diet in Central America and the Dominican Republic: Bringing social inequality to the center of analysis. Social science & medicine, 239, 112516.

World Bank (2021). World Integrated Trade Solution: 'Imports, Exports and Mirror Data with UN COMTRADE'. Retrieved from https://wits.worldbank.org/wits/wits/witshelp/content/data_retrieval/T/Intro/B2.Imports_Exports_and_Mirror.htm

APPENDICES

— Summary of Appendices

Appendix 1: Contextual attributes of PICTs included in the Pacific Food Trade Database

Appendix 2: Random examples of implausible exports from PICTs in Stage 2.1.

Appendix 3: Random examples of plausible trade flows between PICTs in Stage 2.2

Appendix 4: Random examples of trade flows changed or eliminated in Stage 2.3.

Appendix 5: Random examples of trade flows changed or eliminated in Stage 3.1.

Appendix 6: Random examples of trade flows eliminated in Stage 3.2.

Appendix 7: Sensitivity of results to methods and assumptions of Stage 4

Appendix 8: Summary of quantities post-imputation for the 10 tested methods in Stage 4.2.

Appendix 9: Summary of outliers treated by rules within each of the 10 tested methods in Stage 4.2.

Appendix 10: Specific changes for each commodity set reviewed in Stage 5, structured by chapter.

Appendix 11: Commodities excluded from HS92 Chapters HS01 – HS25 in the development of the PFTD

— Appendix 1: Contextual attributes of PICTs included in the Pacific Food Trade Database.

Population data from United Nations (https://population.un.org/wpp/Download/Standard/Population/). HDI (2018 global rank) data from Human Development Reports (http://hdr.undp.org/en/data#). Land Area and GNI per capita, PPP from World Development Indicators (https://databank.worldbank.org/source/world-development-indicators/).

PICT	Sub-region	Sovereignty	Population (2018)	HDI (2018 global rank)	GNI per capita, PPP (2018)	Land Area (km2)
Cook Islands	Polynesia	Sovereign State*	17,519	..	..	..
Fiji	Melanesia	Sovereign State	883,490	98	13,180	18,270
French Polynesia	Polynesia	Territory	277,673	..	..	4,000
Kiribati	Micronesia	Sovereign State	115,842	132	4,410	810
Marshall Islands	Micronesia	Sovereign State	58,412	117	5,090	180
Micronesia (FSM)	Micronesia	Territory	112,640	135	3,640	700
Nauru	Micronesia	Sovereign State	10,678	..	20,940	20
New Caledonia	Melanesia	Territory	279,986	..	..	18,580
Niue	Polynesia	Sovereign State *	1,610	..	..	..
Palau	Micronesia	Sovereign State	17,911	55	19,510	460
Papua New Guinea	Melanesia	Sovereign State	8,606,324	155	4,220	462,840
Samoa	Polynesia	Sovereign State	196,128	111	..	2,840
Solomon Islands	Melanesia	Sovereign State	652,856	153	2,320	28,900
Tokelau	Polynesia	Territory	1,318	..	..	..
Tonga	Polynesia	Sovereign State	103,199	105	6,520	750
Tuvalu	Polynesia	Sovereign State	11,505	..	6,100	30
Vanuatu	Melanesia	Sovereign State	292,675	141	3,250	12,190
Wallis & Futuna Is.	Polynesia	Territory	11,653	..	..	..

*: Self-governing in free association with New Zealand; ..: No data; HDI: Human Development Index

— Appendix 2: Random examples of implausible exports from PICTs removed from the database in Stage 2.1.

Year	Exporter	Importer	Value (US$000)	Quantity (t)	HS Chapter	Chapter definition	HS4 Heading
2017	Nauru	Singapore	69.2	18.0	09	Coffee, tea, mate and spices	0901
1996	Tokelau	Southern African Cust. Union	7.0	11.0	07	Edible vegetables and certain roots and tubers	0713
2015	Marshall Islands	Kiribati	4.8	0.0	09	Coffee, tea, mate and spices	0901
2008	Tokelau	Germany	10.1	18.8	20	Vegetable, fruit, nut, etc food preparations	2009
2003	Nauru	Nigeria	92.7	67.2	04	Dairy products, eggs, honey, edible animal product..	0405
2002	Tokelau	Former Sudan	95.2	107.1	20	Vegetable, fruit, nut, etc food preparations	2002
2005	Tokelau	Southern African Cust. Union	29.6	19.1	08	Edible fruit, nuts, peel of citrus fruit, melons	0812
2002	Tokelau	USA	4.7	2.5	21	Miscellaneous edible preparations	2102
1995	Kiribati	New Zealand	15.0	7.6	07	Edible vegetables and certain roots and tubers	0712
2010	Tokelau	United Republic of Tanzania	24.9	67.8	21	Miscellaneous edible preparations	2102
2001	Vanuatu	Australia	3.6	0.0	22	Beverages, spirits and vinegar	2204
2017	Tokelau	Seychelles	2.9	0.2	21	Miscellaneous edible preparations	2105
2009	Tokelau	France, Monaco	44.9	3.9	03	Fish, crustaceans, molluscs, aquatic invertebrates ne	0304
2001	Tuvalu	Sweden	7.1	10.0	07	Edible vegetables and certain roots and tubers	0705
2013	Tokelau	France, Monaco	3.4	0.5	03	Fish, crustaceans, molluscs, aquatic invertebrates ne	0302
1995	Wallis and Futuna Islands	Colombia	73.1	160.0	17	Sugars and sugar confectionery	1701
1998	Solomon Islands	USA	52.4	7.7	22	Beverages, spirits and vinegar	2208
1997	Nauru	Germany	2.1	1.0	07	Edible vegetables and certain roots and tubers	0708
2007	Tokelau	Mozambique	23.0	8.6	19	Cereal, flour, starch, milk preparations and products	1905
2007	Nauru	Slovenia	13.2	2.8	08	Edible fruit, nuts, peel of citrus fruit, melons	0805
2008	Vanuatu	Nigeria	43.3	12.8	04	Dairy products, eggs, honey, edible animal product…	0402
2001	Tuvalu	Sweden	15.9	5.0	09	Coffee, tea, mate and spices	0902
2018	Solomon Islands	Colombia	64.6	72.0	07	Edible vegetables and certain roots and tubers	0710
2017	Vanuatu	Suriname	6.9	3.8	16	Meat, fish and seafood food preparations nes	1602

— **Appendix 3:** Random examples of trade flows between PICTs established as plausible during stage 2.2 of cleaning process.

Year	Exporter	Importer	Value (US$000)	Quantity (t)	HS Chapter	Chapter definition	HS Heading
1997	Unknown	Kiribati	12.81	3.25	16	Meat, fish and seafood food preparations nes	1602
2004	Unknown	Fiji	4.30	0.20	21	Miscellaneous edible preparations	2103
2015	Marshall Islands	Kiribati	1.01	0.09	21	Miscellaneous edible preparations	2103
2003	Papua New Guinea	Fiji	1.59	0.58	09	Coffee, tea, mate and spices	0903
2013	Unknown	Fiji	69.68	175.43	10	Cereals	1006
1996	Unknown	Fiji	20.26	46.40	15	Animal, vegetable fats and oils, cleavage products..	1502
2006	French Polynesia	New Caledonia	40.07	5.38	11	Milling products, malt, starches, inulin, wheat glute	1105
2015	Kiribati	Tuvalu	9.03	0.45	22	Beverages, spirits and vinegar	2208
2008	New Caledonia	Vanuatu	1.20	0.19	15	Animal, vegetable fats and oils, cleavage products..	1507
2015	Marshall Islands	Kiribati	2.69	0.19	21	Miscellaneous edible preparations	2103
2015	New Caledonia	Vanuatu	2.26	0.14	04	Dairy products, eggs, honey, edible animal product..	0407
2008	Unknown	Kiribati	1.74	0.51	19	Cereal, flour, starch, milk preparations and products	1905
2018	Samoa	Tokelau	2.06	0.32	07	Edible vegetables and certain roots and tubers	0701
2015	Kiribati	Tuvalu	9.88	0.25	24	Tobacco and manufactured tobacco substitutes	2403
2012	Unknown	Nauru	1.48	0.30	11	Milling products, malt, starches, inulin, wheat glute	1102
2011	Unknown	Solomon Islands	5.39	1.06	02	Meat and edible meat offal	0210
2000	Vanuatu	PNG	16.08	6.10	16	Meat, fish and seafood food preparations nes	1602

— Appendix 4: Random examples of trade flows subject to expert elicitation process during Stage 2.3. 'Review scores': 0=Remove, 1=Importer 'unknown', 2= Importer 'PICT unknown', 3= Palm oil - Copra oil, 4 = Christmas Island - Kiribati, 5= retain unchanged.

Year	Exporter	Importer	Value (US$000)	Quantity (t)	HS Chapter	Chapter definition	HS Heading	Review Score
2004	FSM	Dominican Republic	1.90	0.69	17	Sugars and sugar confectionery	1704	0
2015	Kiribati	Sri Lanka	149.15	159.28	15	Animal, vegetable fats and oils, cleavage products, et	1513	5
2002	Nauru	Saint Vinc. and Gren.	7.63	2.00	03	Fish, crustaceans, molluscs, aquatic invertebrates ne	0305	0
2012	Samoa	Finland	87.06	18.98	09	Coffee, tea, mate and spices	0901	1
2004	New Caledonia	Gabon	72.78	11.67	22	Beverages, spirits and vinegar	2208	0
2017	Samoa	Poland	3.16	1.37	22	Beverages, spirits and vinegar	2204	0
2008	Fiji	Saudi Arabia	44.74	113.00	11	Milling products, malt, starches, inulin, wheat glute	1103	2
2008	Cook Islands	Croatia	11.62	1.01	20	Vegetable, fruit, nut, etc food preparations	2009	1
2002	New Caledonia	Nigeria	1,354.13	672.14	19	Cereal, flour, starch, milk preparations and products	1901	0
1995	Cook Islands	Slovakia	5.01	19.80	08	Edible fruit, nuts, peel of citrus fruit, melons	0803	1
2011	French Polynesia	Poland	2.92	0.36	04	Dairy products, eggs, honey, edible animal product nes	0406	0
2004	Samoa	Costa Rica	23.65	9.51	16	Meat, fish and seafood food preparations nes	1604	1
2009	Tonga	Indonesia	10.34	1.69	19	Cereal, flour, starch, milk preparations and products	1905	0
2011	Cook Islands	Ukraine	25.55	15.85	03	Fish, crustaceans, molluscs, aquatic invertebrates ne	0303	1
2017	Marshall Islands	Italy	621.77	993.81	15	Animal, vegetable fats and oils, cleavage products, et	1511	3

— Appendix 5: Random examples of trade flows changed or eliminated in Stage 3.1. Review Score: (0=Retain, 1=Christmas Island - Kiribati, 2= Exporter Unknown, 3 = Delete).

Year	Exporter	Importer	Value (US$000)	Quantity (t)	HS Chapter	Chapter definition	HS Heading	Review Score
2009	Senegal	FSM	2.03	1.77	20	Vegetable, fruit, nut, etc food preparations	2008	2
2014	Spain	Kiribati	834.22	859.42	03	Fish, crustaceans, molluscs, aquatic invertebrates ne	0303	0
2000	Sierra Leone	Papua New Guinea	27.86	12.80	11	Milling products, malt, starches, inulin, wheat glute	1106	2
2005	Former Sudan	Tonga	1.43	0.86	19	Cereal, flour, starch, milk preparations and products	1905	2
2010	Senegal	FSM	1.77	0.66	20	Vegetable, fruit, nut, etc food preparations	2008	2
2013	Saudi Arabia	FSM	8.84	14.00	04	Dairy products, eggs, honey, edible animal product nes	0401	2
2018	UAE	Palau	12.00	2.52	19	Cereal, flour, starch, milk preparations and products	1905	2
2011	Saudi Arabia	FSM	1.45	1.60	22	Beverages, spirits and vinegar	2202	2
2013	Senegal	FSM	4.02	1.29	19	Cereal, flour, starch, milk preparations and products	1905	2
2012	S. African Cust. Union	Solomon Islands	141.91	150.00	03	Fish, crustaceans, molluscs, aquatic invertebrates ne	0303	0
2008	Senegal	FSM	1.19	0.38	21	Miscellaneous edible preparations	2104	2
2003	Denmark	Cook Islands	105.84	288.60	03	Fish, crustaceans, molluscs, aquatic invertebrates ne	0303	0
2008	Senegal	FSM	9.53	14.84	04	Dairy products, eggs, honey, edible animal product nes	0401	2
2008	Armenia	Tonga	2.62	1.10	20	Vegetable, fruit, nut, etc food preparations	2005	2
1998	Sierra Leone	Papua New Guinea	25.09	4.31	11	Milling products, malt, starches, inulin, wheat glute	1106	2
2014	Brazil	Palau	100.95	79.51	02	Meat and edible meat offal	0206	0
2004	Zimbabwe	Tokelau	3454.24	2361.80	24	Tobacco and manufactured tobacco substitutes	2401	3
1999	Spain	Samoa	339.79	437.81	03	Fish, crustaceans, molluscs, aquatic invertebrates ne	0303	0
2010	Zambia	Tokelau	49.65	29.00	10	Cereals	1005	3
2010	S. African Cust. Union	Tokelau	64.39	69.94	08	Edible fruit, nuts, peel of citrus fruit, melons	0804	3
2010	Saudi Arabia	Tonga	7.46	2.71	03	Fish, crustaceans, molluscs, aquatic invertebrates ne	0303	2
2000	Serbia and Montenegro	Niue	3.00	1.00	11	Milling products, malt, starches, inulin, wheat glute	1106	2
2002	Sweden	Tuvalu	100.39	75.00	04	Dairy products, eggs, honey, edible animal product nes	0402	3

— Appendix 6: Random examples of trade flows identified as implausible in stage 3.2 and eliminated.

Year	Exporter	Importer	Value (US$000)	Quantity (t)	HS Chapter	Chapter definition	HS Heading
2017	Belgium-Luxembourg	Kiribati	16	1	01	Live animals	0105
2008	Ecuador	Tokelau	38	73	08	Edible fruit, nuts, peel of citrus fruit, melons	0803
1998	Indonesia	Palau	93	250	12	Oil seed, oleagic fruits, grain, seed, fruit, etc, ne	1205
1995	Hungary	Nauru	99	990	10	Cereals	1001
2012	Peru	Kiribati	321	101	09	Coffee, tea, mate and spices	0901
2010	Unknown	Tokelau	41	225	10	Cereals	1001
2011	USA	Niue	236	50	08	Edible fruit, nuts, peel of citrus fruit, melons	0802
2003	Germany	Tuvalu	827	662	04	Dairy products, eggs, honey, edible animal product nes	0402
2012	Australia	Palau	1,206	1,250	15	Animal, vegetable fats and oils, cleavage products, et	1502
2007	Ecuador	Tokelau	79	163	08	Edible fruit, nuts, peel of citrus fruit, melons	0803
1997	Canada	Tokelau	36	108	07	Edible vegetables and certain roots and tubers	0713
2006	India	Kiribati	43	100	08	Edible fruit, nuts, peel of citrus fruit, melons	0803
2017	Viet Nam	Papua New Guinea	61,044	12,283	09	Coffee, tea, mate and spices	0904
2009	USA	Niue	388	119	08	Edible fruit, nuts, peel of citrus fruit, melons	0802
2002	India	Niue	13	48	10	Cereals	1006
2010	Canada	Tonga	4,131	17,000	10	Cereals	1001
1998	Unknown	Niue	4,288	6,716	02	Meat and edible meat offal	0207
2011	Indonesia	Tuvalu	571	500	15	Animal, vegetable fats and oils, cleavage products, et	1511
2017	Un. Rep. Tanzania	Kiribati	39	32	01	Live animals	0102
1995	Colombia	Tuvalu	394	100	18	Cocoa and cocoa preparations	1804
2006	Indonesia	Kiribati	48	56	08	Edible fruit, nuts, peel of citrus fruit, melons	0804
2010	Uganda	Samoa	154	98	09	Coffee, tea, mate and spices	0901
1999	Thailand	Niue	345	4,500	07	Edible vegetables and certain roots and tubers	0714
1999	Indonesia	Niue	49	60	24	Tobacco and manufactured tobacco substitutes	2401
2008	Unknown	Niue	148	106	03	Fish, crustaceans, molluscs, aquatic invertebrates ne	0307

— Appendix 7: Sensitivity of results to methods and assumptions of Stage 4.

We tested the sensitivity of results to a range of assumptions and methods, including the impact of outlier identification, imputation method, sample size, and sample used to estimate medians. We detail these below.

Choice of Tukey's interquartile range method to detect outliers

The literature provides several methods of detecting outliers in univariate data, including those based on standard deviation, Z-score, Tukey's interquartile range, adjusted box plots, generalized box plots, and the median rule (Bruffaerts, Verardi, & Vermandele, 2014; Seo, 2006). The standard deviation method is accurate if the distribution of the data is known and if the data is reasonably symmetric (Seo, 2006). The Z-score method is only applicable to normally distributed data (Aggarwal, 2017; Seo, 2006). Tukey's method is less sensitive to the distribution of the data, however, when the data are skewed, the method tends to falsely detect some values as outliers (Hubert & Vandervieren, 2008). Hence, an appropriate transformation may be used to normalize the data. We used log transformation to convert the heavily positively skewed unit price variable into a symmetric shape, following UNSD protocols (Food and Agriculture Organisation, United Nations Statistics Division, International Trade Centre, 2019). Moreover, Tukey's fence method may not be suitable for small sample sizes, 10 or below (Jones, 2018). We detected outliers within 581 commodities separately. There were only 17 commodities which had trade flows below 11.

Hubert and Vandervieren (2008) introduced adjusted box plot method and Bruffaerts et al. (2014) developed generalized box plot method for detecting outliers in skewed data. We explored the adjusted box plot method for detecting outliers in the unit price and compared with the number of outliers detected by Tukey's (1977) method in log unit price. Tukey's method detected 13,177 outliers in the log unit price variable and the adjusted box plot method detected 23,874 outliers in the unit price variable. The adjusted box plot method accounts for the skewness, but not for the heavy tail, which is the case in our data set (Hubert & Vandervieren, 2008). Due to the simplicity and the accuracy of Tukey's method and the complexity in the adjusted box plot method and the generalized box plot method, we selected Tukey's fence method for detecting the outliers in log unit price (see also Food and Agriculture Organisation, United Nations Statistics Division, International Trade Centre, 2019).

Assumption of less error in value than quantity

Official trade value ($) data are used to calculate taxes and duties, and as consequence subject to considerable control and attention. In contrast, quantities are not always reported and are often estimated and hence more prone to error. With this in mind, we tested the assumption that there was more error in quantity data than value ($) data by identifying outliers in each for the entire dataset, using Tukey's fences ($k = 1.5$) on logged data as described above. Marginally more quantity data (0.8%, 2,566 records) than value data (0.5%, 1,700 records) were recognised as outliers. Importantly, however, 52% of trade flows with quantity error, compared to 5% of data points with value error, were also considered unit price errors. Greater total error, and correlation with unit price error in the quantity data, combined with a significant number of trades with completely implausible quantities and unit prices (Table 2), provided support for revising quantities rather than values. It was not, however, feasible to correct quantity error using quantity data only because outliers would be recognised as large and small trades only, and the results biased towards PICTs that trade in either large or small quantities of each commodity. Based on the assumption that trade values were reliable, we used unit price data to identify outliers. Outlier quantities were adjusted by dividing the value ($) of the trade record by the imputed median unit price. In instances where the trade flow was between two PICTs, the exporting PICT was used to calculate non-outlier median because there was a greater likelihood of more trade flows in a given sample frame (e.g. export of fish).

Sensitivity of outlier detection to normality assumption

Outliers were detected using Tukey's fence method within each of the 581 commodities. The data in 220 commodities were normally distributed (Kolmogorov Smirnov tests). Figure A7.1 presents the distributions of eight commodities whose log unit prices were non-normal. The box plots show that the distributions were not heavily skewed.

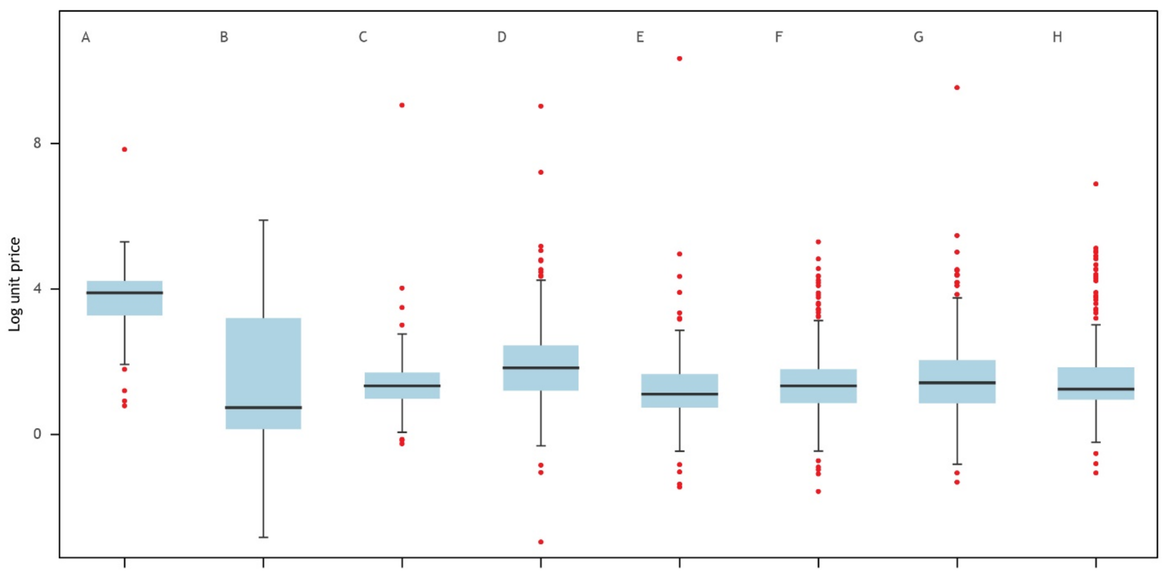


Figure A7.1 Box plots of eight of the commodities with non-normal log unit price according to the K-S test. A = Poultry: live, fowls of the species *Gallus domesticus*, weighing not more than 185 g; B = Poultry: live, ducks, geese, turkeys and guinea fowls, weighing more than 185 g; C = Meat: of bovine animals, carcasses and half-carcasses, fresh or chilled; D = Meat: of bovine animals, cuts with bone in (excluding carcasses and half-carcasses), fresh or chilled; E = Meat: of bovine animals, carcasses and half-carcasses, frozen; F = Meat: of bovine animals, boneless cuts, frozen; G = Meat: of swine, hams, shoulders and cuts thereof, with bone in, fresh or chilled; H = Meat: of swine, n.e.s. in item no. 0203.1, fresh or chilled.

Sensitivity of outlier detection to Tukey's k

We tested the sensitivity of results to the *k* = 1.5 convention by comparing the number and quantity of outliers recognized by incrementally increasing k from 1 to 3.0. We did not identify a step change in the number of identified outliers across the gradient of *k* (Figure A7.2), which might occur if, for example, a large portion of the error occurred due to mis-reporting quantity as kilograms instead of tonnes. We therefore followed the k = 1.5 convention. Example distributions are shown in Figure A7.3.

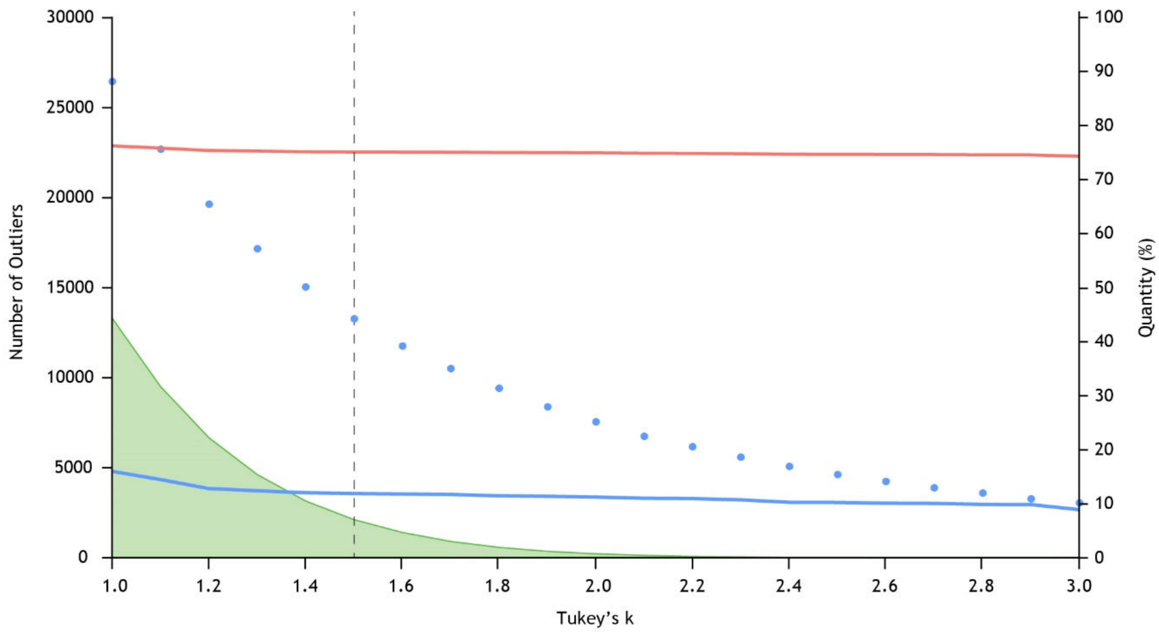


Figure A7.2. Outliers detected as a function Tukey's k. The dashed blue line shows the number of outliers identified in log(unit price) of trade flows within HS6 at different levels of Tukey's k; the solid lines shows the % quantity in those outliers and for all trade flows (red) and with the 20 trade flows with the largest quantities excluded (blue). The green shaded area is the number of outliers detected from a simulated normal distribution. The vertical dashed lines is at K = 1.5, the value used by convention and used in this study.

Distribution of outliers

Figure A7.3 presents the cumulative quantity of outliers in eight commodities. Here outliers are ranked in descending order of the quantity and then the cumulative quantity is calculated. The majority of quantity is accounted by the first few outliers. For example, in baked goods (blue line), 15 of 56 outliers account for the majority of the total quantity outliers in that commodity. The illustrative examples are mostly for commodities traded in small quantities hence the small numbers and quantities involved. Outliers in canned fish and vegetables n.e.s. are much more evenly distributed.

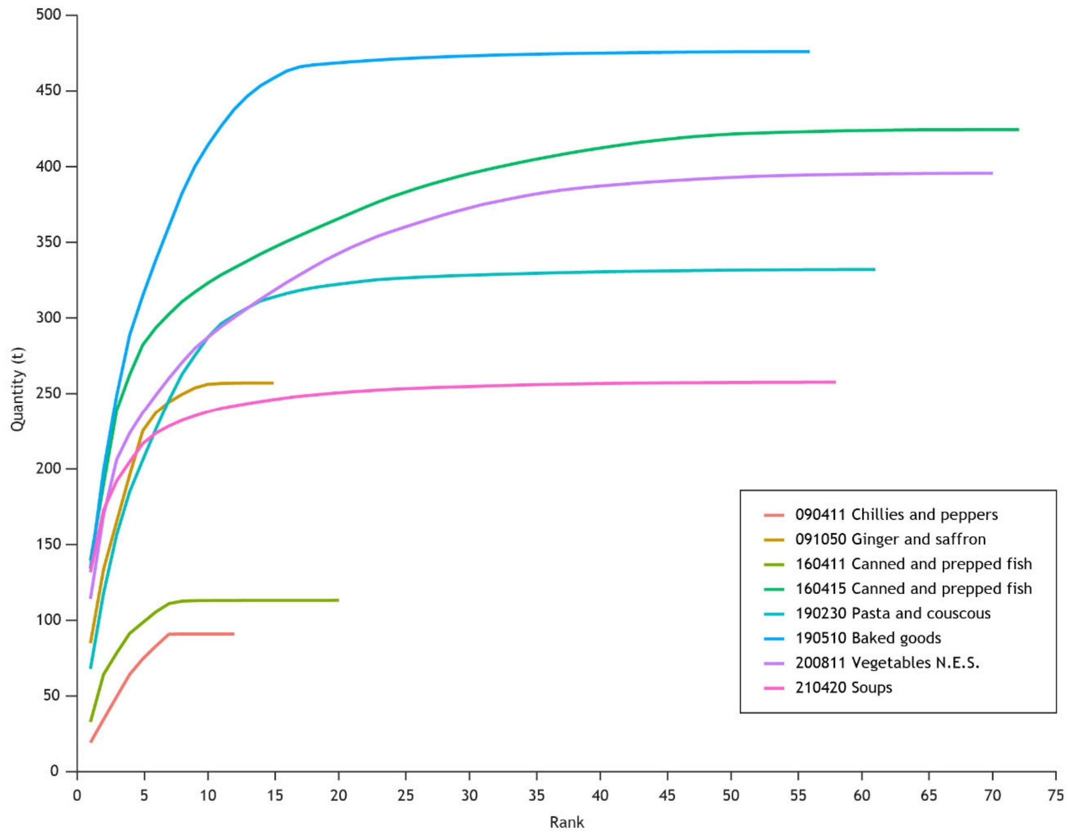


Figure A7.3 The cumulative quantity of outliers ranked in the descending order of the quantity.

Sensitivity of results to imputation method

Once unit prices were identified as outliers using the Tukey's method described above, the corresponding quantities were then corrected using a median unit price estimated from the distribution of non-outlier unit prices. We compared 10 methods to estimate median unit price, of increasing granularity and complexity of rule structure:

Method 1: all trade flows in the data set at HS6 were pooled to estimate the median of non-outliers and that median imputed for all outliers in the dataset at that HS6, irrespective of sample size. This method produces one median unit price for each of the 581 food commodity. This method is comparable to the use of standard unit prices (SUP) to infer quantities.

Method 2: all trade flows at HS6 from the same PICT were pooled to estimate the median of non-outliers in the sample and that median was imputed for each PICT at that HS6, irrespective of sample size. In 1.2% of instances the outlier was the only trade flow for that PICT in the dataset; in these instances the median of non-outliers from all PICTs and years was imputed.

Method 3: all trade flows at HS6 from the same YEAR were pooled to estimate the median of non-outliers in the sample and that median was imputed for each YEAR at that HS6, irrespective of sample size. In 0.3% of instances the outlier was the only trade flow for that YEAR in the dataset; in these instances the median of non-outliers from all PICTs and years was imputed.

Method 4: all trade flows at HS6 from combinations of PICT * YEAR were pooled to estimate the median of non-outliers in the sample and that median was imputed for the PICT * YEAR combination irrespective of sample size. In 15.3% of instances the outlier was the only trade flow for that PICT * YEAR combination in the dataset; in these instances the median of non-outliers from all PICTs and years was imputed.

Method 5: as per Method 4, but with a sample of n ≥ 20 used to estimate the median. If this sample size rule was not satisfied for a single year then years were added in increments of two to a maximum of 21 years. If the sample size remained <20 at this point then the median was estimated from all years for that PICT, irrespective of sample size (see also Method 2). The outlier year was taken as the mid-year if the sample (e.g. if three years was required to achieve n ≥ 20 for outliers in 1997 then the sample was drawn from 1996, 1997 and 1998). As a consequence, imputation of outliers towards either end of the time series, notably the first and last years, occasionally required sample years 'outside' the dataset (before 1995 or after 2018). If this occurred then outliers were imputed using the median for all years in the PICT at that HS6. In 1.2% of instances the outlier was the only trade flow for that HS6 for that PICT in the dataset; in these instances the median of non-outliers from all PICTs and years was imputed.

Method 6: as per method 5 but with a minimum sample size of n ≥ 15.

Method 7: as per method 5 but with a minimum sample size of n ≥ 10.

Method 8: as per Method 5 in requiring a minimum sample size of 20, but using a different rule if the sample size rule cannot be satisfied by years within the dataset. As with Method 5, if the rule was not satisfied for a single year then years were added in increments of two, with the outlier year the middle year of the sample. In this method, however, 'out of scope' years (before 1995 or after 2018) were used as dummy variables and the median calculated from those years in scope. This artefact was greatest toward either end of the times series (Figure A8.5). The effect of this rule was to use years closest to the outlier year to estimate the median and therefore minimize the impact of systemic change in unit price through time. In 1,964 (15% of 13,177 outliers) instances there was fewer than 20 non-outlier data points within the PICT across all years. In these instances the process was repeated using non-outlier data from all PICTs. Finally, in 21 instances the minimum sample size could not be reached from all PICTs and the median unit price of the sample within HS6 across all PICTs, irrespective of sample size was used.

Method 9: as per method 8 but with a minimum sample size of n ≥ 15.

Method 10: as per method 8 but with a minimum sample size of n ≥ 10.

The alternative imputation methods yielded similar estimates of trade flows for individual PICTs and for the region (Table A7.1). There were, however, distinct differences between PICTs in the effect of imputation, irrespective of imputation method. These consistent and dramatic differences suggest systematic error in quantity within the primary data for a limited set of countries including Kiribati, Marshall Islands, Papua New Guinea, Samoa and Tuvalu. Further, quantity estimates across all PICTs were altered as a result of imputation. The overall percentage change is minimal in most instances, but could be influential for specific commodities within PICTs.

Table A7.1. Percentage change (+/-) in total quantity within PICT following imputation for each of the ten described imputation methods.

	Imputation Method									
PICT	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Cook Islands	0.36	0.40	0.33	0.32	0.39	0.35	0.35	0.39	0.35	0.28
Federated State of Micronesia	-0.03	0.09	0.00	0.06	0.10	0.11	0.11	0.08	0.09	0.09
Fiji	-1.92	-1.92	-2.09	-1.94	-1.85	-2.01	-2.00	-1.84	-1.98	-2.00
French Polynesia	-0.19	-0.27	-0.17	-0.22	-0.26	-0.26	-0.23	-0.24	-0.24	-0.23
Kiribati	-37.65	-37.62	-37.73	-37.65	-37.65	-37.66	-37.70	-37.68	-37.69	-37.70
Marshall Islands	-20.55	-20.16	-20.59	-20.55	-20.14	-20.14	-20.14	-20.14	-20.12	-20.14
Nauru	-0.27	-0.39	-0.33	-0.24	-0.40	-0.40	-0.40	-0.28	-0.28	-0.43
New Caledonia	-3.76	-3.80	-3.77	-3.80	-3.80	-3.80	-3.79	-3.79	-3.79	-3.79
Niue	2.43	3.23	2.19	2.63	3.23	3.23	3.26	2.26	2.94	3.28
Palau	-3.73	-3.71	-3.79	-3.68	-3.71	-3.69	-3.70	-3.72	-3.71	-3.70
Papua New Guinea	-86.10	-85.94	-86.04	-85.92	-85.94	-85.92	-85.91	-85.95	-85.94	-85.93
Samoa	-40.34	-39.97	-40.00	-39.57	-39.70	-39.43	-39.25	-39.68	-39.38	-39.27
Solomon Islands	-1.98	-2.04	-2.67	-3.00	-2.17	-2.18	-2.40	-2.44	-2.43	-2.41
Tokelau	-3.73	-3.81	-3.35	-3.75	-3.81	-3.81	-3.84	-3.35	-3.35	-3.46
Tonga	-3.97	-3.66	-4.12	-3.68	-3.71	-3.66	-3.63	-3.72	-3.68	-3.67
Tuvalu	-33.79	-33.73	-33.63	-33.73	-33.70	-33.70	-33.68	-33.73	-33.67	-33.63
Vanuatu	-1.85	-1.86	-1.86	-1.85	-1.84	-1.84	-1.85	-1.83	-1.84	-1.85
Wallis and Futuna Islands	0.84	0.89	0.64	0.90	0.91	0.91	0.93	0.52	0.41	0.40
All PICTs	-74.61	-74.47	-74.58	-74.46	-74.47	-74.45	-74.45	-74.48	-74.48	-74.47

The alternative imputation methods also yielded similar estimates of trade flows for individual years (Table A7.2). The final quantity did no differ markedly (<2%) from the data prior to imputation for most years. There was, however, a block of years (2000-2005) that were dramatically altered by all imputation methods. This block is primarily explained by error in Papua New Guinea rice imports.

Table A7.2. Percentage change (+/-) in total quantity within YEAR comparing data prior to imputation methods with each of the ten described imputation methods.

	Imputation Method									
YEAR	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
1995	-0.13	-0.13	-0.12	-0.12	-0.13	-0.13	-0.13	-0.12	-0.12	-0.12
1996	-0.42	-0.40	-0.39	-0.39	-0.40	-0.39	-0.39	-0.36	-0.36	-0.36
1997	1.54	2.15	1.68	2.31	2.22	2.20	1.84	2.60	2.56	1.87
1998	-76.73	-76.41	-76.53	-76.08	-76.31	-75.98	-76.01	-76.03	-75.99	-75.98
1999	-0.63	-0.39	-0.40	1.05	0.28	0.71	1.06	0.28	0.71	1.06
2000	-95.40	-95.35	-95.34	-95.20	-95.26	-95.25	-95.25	-95.26	-95.24	-95.24
2001	-95.02	-94.98	-94.92	-94.87	-94.94	-94.92	-94.92	-94.93	-94.91	-94.91
2002	-96.01	-95.96	-95.95	-95.88	-95.92	-95.92	-95.89	-95.93	-95.91	-95.89
2003	-69.20	-68.81	-68.70	-68.47	-68.67	-68.63	-68.61	-68.74	-68.71	-68.61
2004	-16.04	-15.76	-15.59	-15.33	-15.32	-15.24	-15.20	-15.34	-15.25	-15.21
2005	-5.39	-5.36	-5.15	-5.24	-5.23	-5.18	-5.16	-5.23	-5.18	-5.16
2006	-0.35	-0.20	-0.05	0.19	0.03	-0.06	-0.05	0.03	-0.06	-0.05
2007	-2.22	-2.12	-2.11	-2.02	-2.02	-2.00	-2.00	-2.02	-2.00	-1.99
2008	-1.27	-1.93	-1.77	-2.29	-1.17	-2.31	-2.32	-1.18	-2.32	-2.32
2009	-23.41	-23.39	-23.47	-23.42	-23.45	-23.45	-23.45	-23.45	-23.45	-23.45
2010	-1.02	-0.95	-1.08	-0.97	-1.00	-1.00	-1.01	-1.00	-1.01	-1.01
2011	-1.49	0.73	-2.36	-2.01	-0.31	-0.18	-0.29	-1.98	-1.81	-1.81
2012	2.96	5.37	1.01	1.50	2.87	2.18	2.05	2.77	2.12	2.00
2013	-0.62	-0.63	-0.95	-1.36	-0.82	-0.82	-0.86	-0.84	-0.84	-0.86
2014	-0.64	-0.24	-1.26	-1.10	-0.60	-0.59	-0.60	-0.62	-0.59	-0.60
2015	-1.17	-1.01	-1.42	-1.08	-1.18	-1.18	-1.34	-1.31	-1.35	-1.34
2016	-16.33	-16.35	-16.58	-16.60	-16.69	-16.69	-16.63	-16.70	-16.69	-16.64
2017	-2.72	-2.67	-2.94	-2.86	-2.72	-2.79	-2.76	-2.83	-2.84	-2.83
2018	-2.56	-2.56	-2.66	-2.64	-2.63	-2.63	-2.63	-2.68	-2.66	-2.66
Total	-74.61	-74.47	-74.58	-74.46	-74.47	-74.45	-74.45	-74.48	-74.48	-74.47

Choosing which method to use to create the final database required balancing simplicity (parsimony) with the possibility of bias in analyses of commodities with few trades of small quantity in few PICTs. As analyses become more granular, there is an increasing likelihood that important differences within PICTs would be swamped by trade flows in larger PICTs. For example, because smaller PICTs import smaller volumes, the unit price of imported commodities it likely to be higher. Given the database is intended to be of use for analyses focused on smaller PICTs and specific commodities we lean towards choosing the method that is likely to provide the most accurate outcomes.

Compared to methods 1-3, methods 4-10 account for greater variability in the data as median unit prices were estimated at a more disaggregated and therefore less heterogeneous level. Methods 5-10 may provide more robust estimates as they impose a minimum sample size. The rule structure for methods 8-10 assumes the greatest difference in unit prices resides among HS, then among PICTs then among YEARS, with adjacent years containing more similar unit price data (see Figures A7.4 and A7.5 for examples showing variation in unit price among PICTs and among years). We tested this assumption by calculating the range, variance and standard deviation of the mean unit price using the non-outlier data (Table A7.3). As assumed, HS6 contains the greatest level of data dispersion, followed by PICT then YEAR. PICT and YEAR have similar dispersion compared to HS6. Additionally, YEAR data is ordinal, allowing for imputation from adjacent years. Using unit price data from adjacent years means that more similar data were used for the imputation, than data from all years, adding support to our choice of rule structure. For example, there is an overall trend in increasing unit price within the non-outlier data and individual commodities, such as brown rice have distinct unit price trends through time (Figure S7.5). We are therefore confident that the rule structure of methods 8-10 is the most appropriate for accounting for natural variability in unit price within the data.

Table A7.3. Comparison of descriptive statistics on means of unit price of non-outlier data within HS6, PICT and YEAR.

	Mean	Range (max-min)	Variance	St. Dev.
HS6 (n = 581)	8.64	231.11	416.15	20.40
PICT (n = 18)	5.89	6.81	3.01	1.73
YEAR (n = 24)	6.43	4.06	1.59	1.26

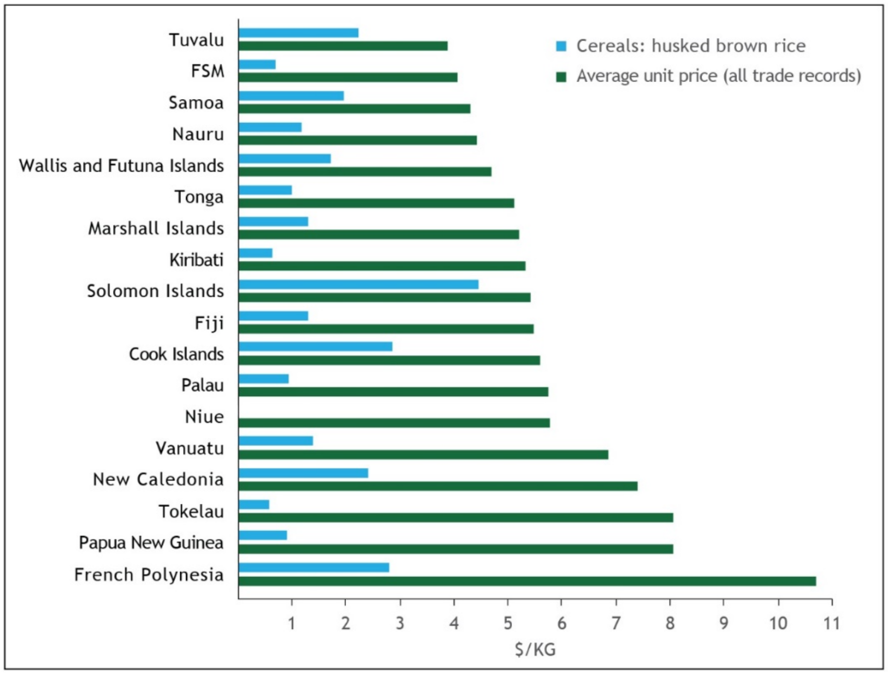


Figure A7.4. Within PICT mean unit price of husked brown rice (HS100620) and all trade records for non-outliers. Variation between PICTs highlight the importance of nesting imputation within PICTs where there is adequate sample size within HS6.

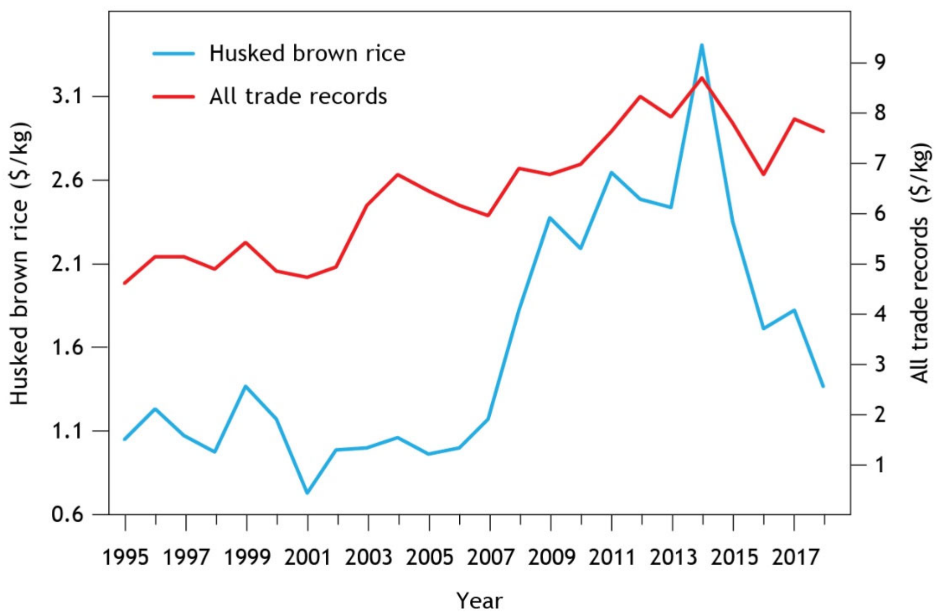
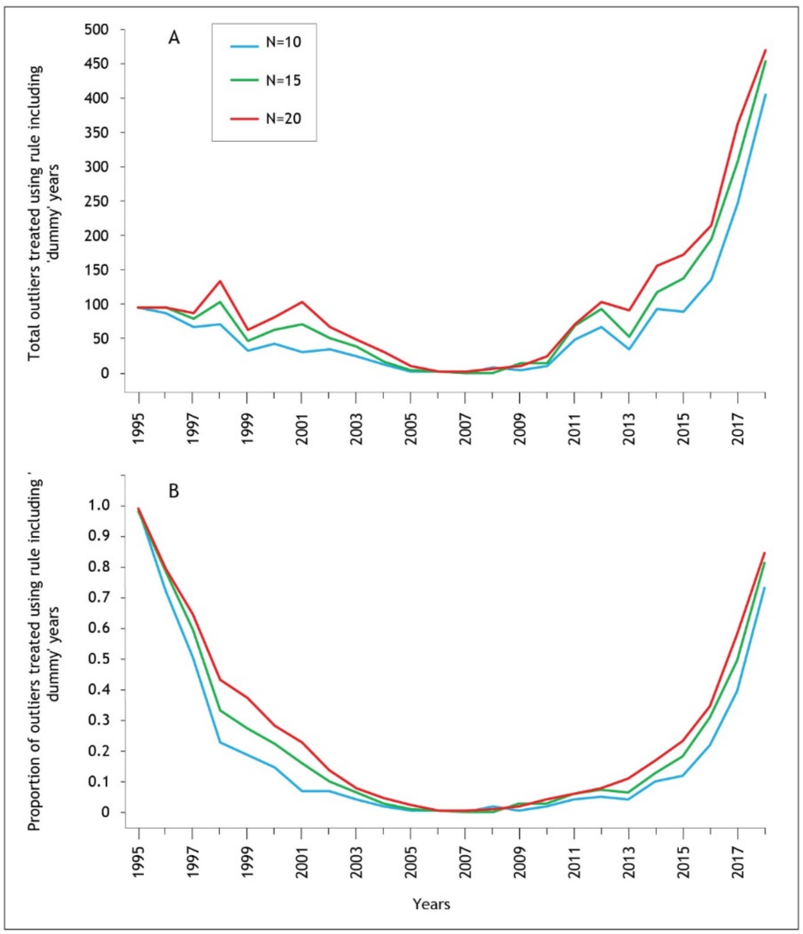


Figure A7.5. Within year mean unit price of husked brown rice (HS100620) and all trade records for non-outliers. Temporal trends highlight the importance of using data from adjacent years to impute updated quantities.

Sensitivity of results to sample size used to estimate median unit price

The final decision in choosing a method was minimum sample size (10, 15, or 20). We observed very little difference between methods 8, 9, and 10 in adjusted quantities within PICT and within YEAR (Tables A7.1, A7.2). We also reviewed the effects of each of the sample sizes on the total number, and proportion, of outliers corrected using a rule that involved dummy years (Figure A7.6). There was little difference among sample sizes in total number or proportion of outliers. We used a minimum sample size to 20 (method 8), balancing adequate sample size with imputing using the most similar data.

Figure A7.6. Total (A) and, proportion (B) of, outliers imputed using rules that included 'dummy' years. Each trend in both A and B shows differences between the different sample sizes used in each method (method 8, N = 20, method 9, N = 15, method 10, N = 10). For example, across all three methods, ~100% of outliers in 1995 were imputed using a rule that included dummy years.



Validation of the variability assumptions in the outlier detection and outlier treatment methods

We detected outliers of log unit prices within HS6 codes assuming there is more variability within hs6 codes than within PICTs or YEARs. The primary method of treating outliers is 'method 8' which starts by calculating the median within HS6, PICT and YEAR combination to replace an outlier. Here, we assumed that a significant variability is captured within this combination. We conducted an analysis of variance to check the validity of these two assumptions.

Since there are 581 HS6 codes, we sampled all trade flows of the 10 HS6 codes with the largest volumes for fitting a linear regression model for predicting log unit prices from a model that includes the three-way interaction of hs6, PICT and YEAR. The regression model can be written as:

$$logUnitPrice = \beta_0 + \beta_1 HS6 + \beta_2 PICT + \beta_3 YEAR + \beta_4 HS6 * PICT + \beta_5 HS6 * YEAR + \beta_6 PICT * YEAR + \beta_7 HS6 * PICT * YEAR \quad \text{(Model 8.1)}$$

Three variables, 'HS6', 'PICT' and 'YEAR' were treated as categorical variables. The ANOVA output is given in Table A7.4. According to the sum of squares values, the largest variability is due to HS6 which confirms our assumption for detecting outliers within HS6. The second largest variability is by the three-way interaction which strengthen our choice for the outlier treatment method. Third largest source of variability is YEAR, however we did not impute medians disregarding the commodity of the outliers. Therefore, any outlier that was not treated by a median calculated considering the combination, HS6*PICT*YEAR were then treated by a median calculated within HS6 and YEAR which also captured a significant variability in the log unit price.

Table A7.4. Analysis of Variance output for Model 8.1.

Source of Variance	**Df**	**Sum of squares**	**Mean Squares**	**F-value**	**P-value**
HS6	9	1759	195.5	147.6	<0.001
PICT	17	645.5	38	28.7	<0.001
YEAR	23	1298.7	56.5	42.6	<0.001
HS6:PICT	110	708.4	6.4	4.9	<0.001
HS6:YEAR	207	887.9	4.3	3.3	<0.001
PICT:YEAR	350	866.7	2.5	1.9	<0.001
HS6:PICT:YEAR	929	1558.7	1.7	1.3	<0.001
Residual	4183	5539.8	1.3		

— Appendix 8: Summary of quantities (t) post-imputation for the 10 tested methods. Here we show all PICTs and all PICTs minus Papua New Guinea to highlight the influence of error in rice trade records.

	All PICTs	Without PNG
Total number of observations	311,474	280,524
Number of outliers (by HS, k = 1.5)	13,177	11,018
Total quantity (t) before treating outliers	314,669,653	45,664,272
Total quantity (t) in Method 1	79,887,786	42,416,393
Total quantity (t) in Method 2	80,319,871	42,445,112
Total quantity (t) in Method 3	80,002,212	42,373,474
Total quantity (t) in Method 4	80,364,926	42,413,738
Total quantity (t) in Method 5	80,348,576	42,463,875
Total quantity (t) in Method 6	80,385,042	42,437,599
Total quantity (t) in Method 7	80,397,028	42,438,306
Total quantity (t) in Method 8	80,313,878	42,457,914
Total quantity (t) in Method 9	80,316,051	42,437,074
Total quantity (t) in Method 10	80,346,798	42,436,590

— Appendix 9: Summary of outliers treated by rules within each of the 10 tested methods. PICTs and all PICTs minus Papua New Guinea included to highlight the variable influence of Papua New Guinea across imputation rules.

Method 1

Imputation rule	Number of outliers treated	
	All PICTs	**Without PNG**
By HS	13177	11018

Method 2

Imputation rule	Number of outliers treated	
	All PICTs	**Without PNG**
By HS, PICT	13016	10876
By HS	161	142

Method 3

Imputation rule	Number of outliers treated	
	All PICTs	**Without PNG**
By HS, YEAR	13141	10983
By HS	36	35

Method 4

Imputation rule	Number of outliers treated	
	All PICTs	**Without PNG**
By HS, PICT, YEAR	11168	9239
By HS	2009	1779

Method 5

Imputation rule	Number of outliers treated	
	All PICTs	**Without PNG**
By HS, PICT, 1-YEAR & N>=20	572	498
By HS, PICT, 3-YEAR & N>=20	3172	2622
By HS, PICT, 5-YEAR & N>=20	1978	1607
By HS, PICT, 7-YEAR & N>=20	1214	983
By HS, PICT, 9-YEAR & N>=20	698	593

By HS, PICT, 11-YEAR & N>=20	505	382
By HS, PICT, 13-YEAR & N>=20	309	245
By HS, PICT, 15-YEAR & N>=20	174	145
By HS, PICT, 17-YEAR & N>=20	96	84
By HS, PICT, 19-YEAR & N>=20	73	73
By HS, PICT, 21-YEAR & N>=20	42	34
By HS, PICT, 23-YEAR & N>=20	10	12
By HS, PICT	4173	3598
By HS	161	142

* 1-YEAR for all years between 1995 – 2018, 3-YEAR for all years between 1996 – 2017 ….., 23-YEAR for 2006 and 2007.

Method 6

Imputation rule	**Number of outliers treated**	
	All PICTs	**Without PNG**
By HS, PICT, 1-YEAR & N>=15	1160	1016
By HS, PICT, 3-YEAR & N>=15	3773	3065
By HS, PICT, 5-YEAR & N>=15	2007	1629
By HS, PICT, 7-YEAR & N>=15	1112	920
By HS, PICT, 9-YEAR & N>=15	664	551
By HS, PICT, 11-YEAR & N>=15	388	310
By HS, PICT, 13-YEAR & N>=15	231	194
By HS, PICT, 15-YEAR & N>=15	173	133
By HS, PICT, 17-YEAR & N>=15	84	72
By HS, PICT, 19-YEAR & N>=15	39	39
By HS, PICT, 21-YEAR & N>=15	22	25
By HS, PICT, 23-YEAR & N>=15	4	5
By HS, PICT	3359	2917
By HS	161	142

Method 7

Imputation rule	Number of outliers treated	
	All PICTs	Without PNG
By HS, PICT, 1-YEAR & N>=10	2367	2014
By HS, PICT, 3-YEAR & N>=10	4400	3561
By HS, PICT, 5-YEAR & N>=10	1766	1461
By HS, PICT, 7-YEAR & N>=10	918	762
By HS, PICT, 9-YEAR & N>=10	510	409
By HS, PICT, 11-YEAR & N>=10	320	253
By HS, PICT, 13-YEAR & N>=10	191	141
By HS, PICT, 15-YEAR & N>=10	77	64
By HS, PICT, 17-YEAR & N>=10	51	35
By HS, PICT, 19-YEAR & N>=10	28	22
By HS, PICT, 21-YEAR & N>=10	11	11
By HS, PICT, 23-YEAR & N>=10	7	8
By HS, PICT	2370	2135
By HS	161	142

Method 8 (method used in Stage 4.2 to generate the Pacific Food Trade Database)

Imputation rule	Number of outliers treated	
	All PICTs	Without PNG
By HS, PICT, 1-YEAR & N>=20	572	498
By HS, PICT, 3-YEAR & N>=20	3247	2700
By HS, PICT, 5-YEAR & N>=20	2124	1750
By HS, PICT, 7-YEAR & N>=20	1370	1129
By HS, PICT, 9-YEAR & N>=20	865	731
By HS, PICT, 11-YEAR & N>=20	707	566
By HS, PICT, 13-YEAR & N>=20	510	427
By HS, PICT, 15-YEAR & N>=20	372	332
By HS, PICT, 17-YEAR & N>=20	279	238
By HS, PICT, 19-YEAR & N>=20	247	206
By HS, PICT, 21-YEAR & N>=20	216	176
By HS, PICT, 23-YEAR & N>=20	158	128
By HS, PICT, 25-YEAR & N>=20	127	109

By HS, PICT, 27-YEAR & N>=20	97	69
By HS, PICT, 29-YEAR & N>=20	74	56
By HS, PICT, 31-YEAR & N>=20	56	40
By HS, PICT, 33-YEAR & N>=20	56	42
By HS, PICT, 35-YEAR & N>=20	54	38
By HS, PICT, 37-YEAR & N>=20	29	23
By HS, PICT, 39-YEAR & N>=20	17	18
By HS, PICT, 41-YEAR & N>=20	13	13
By HS, PICT, 43-YEAR & N>=20	10	9
By HS, PICT, 45-YEAR & N>=20	8	7
By HS, PICT, 47-YEAR & N>=20	5	4
By HS, 1-YEAR & N>=20	731	616
By HS, 3-YEAR & N>=20	835	733
By HS, 5-YEAR & N>=20	166	155
By HS, 7-YEAR & N>=20	89	63
By HS, 9-YEAR & N>=20	43	46
By HS, 11-YEAR & N>=20	22	26
By HS, 13-YEAR & N>=20	9	13
By HS, 15-YEAR & N>=20	9	7
By HS, 17-YEAR & N>=20	13	1
By HS, 19-YEAR & N>=20	3	2
By HS, 21-YEAR & N>=20	4	3
By HS, 23-YEAR & N>=20	5	6
By HS, 25-YEAR & N>=20	4	6
By HS, 27-YEAR & N>=20	2	1
By HS, 29-YEAR & N>=20	2	2
By HS, 31-YEAR & N>=20	0	1
By HS, 35-YEAR & N>=20	1	1
By HS, 37-YEAR & N>=20	1	1
By HS, 39-YEAR & N>=20	2	2
By HS, 41-YEAR & N>=20	2	1
By HS	21	23

* All year ranges (1-YEAR to 47-YEAR) are considered in all years from 1995 – 2018.

Method 9

Imputation rule	Number of outliers treated	
	All PICTs	Without PNG
By HS, PICT, 1-YEAR & N>=15	1160	1016
By HS, PICT, 3-YEAR & N>=15	3866	3162
By HS, PICT, 5-YEAR & N>=15	2168	1780
By HS, PICT, 7-YEAR & N>=15	1279	1075
By HS, PICT, 9-YEAR & N>=15	839	706
By HS, PICT, 11-YEAR & N>=15	599	496
By HS, PICT, 13-YEAR & N>=15	385	326
By HS, PICT, 15-YEAR & N>=15	332	275
By HS, PICT, 17-YEAR & N>=15	233	190
By HS, PICT, 19-YEAR & N>=15	163	134
By HS, PICT, 21-YEAR & N>=15	132	109
By HS, PICT, 23-YEAR & N>=15	116	82
By HS, PICT, 25-YEAR & N>=15	80	70
By HS, PICT, 27-YEAR & N>=15	67	55
By HS, PICT, 29-YEAR & N>=15	65	51
By HS, PICT, 31-YEAR & N>=15	52	44
By HS, PICT, 33-YEAR & N>=15	40	29
By HS, PICT, 35-YEAR & N>=15	39	30
By HS, PICT, 37-YEAR & N>=15	27	22
By HS, PICT, 39-YEAR & N>=15	24	21
By HS, PICT, 41-YEAR & N>=15	19	17
By HS, PICT, 43-YEAR & N>=15	5	4
By HS, PICT, 45-YEAR & N>=15	6	4
By HS, PICT, 47-YEAR & N>=15	2	1
By HS, 1-YEAR & N>=15	701	627
By HS, 3-YEAR & N>=15	514	467
By HS, 5-YEAR & N>=15	126	89
By HS, 7-YEAR & N>=15	50	55
By HS, 9-YEAR & N>=15	21	23
By HS, 11-YEAR & N>=15	16	13
By HS, 13-YEAR & N>=15	9	5
By HS, 15-YEAR & N>=15	9	7

By HS, 17-YEAR & N>=15	2	5
By HS, 19-YEAR & N>=15	8	4
By HS, 21-YEAR & N>=15	2	4
By HS, 23-YEAR & N>=15	1	2
By HS, 25-YEAR & N>=15	2	2
By HS, 27-YEAR & N>=15	2	2
By HS, 29-YEAR & N>=15	1	2
By HS, 31-YEAR & N>=15	0	1
By HS, 35-YEAR & N>=15	0	1
By HS, 37-YEAR & N>=15	1	1
By HS	14	9

Method 10

Imputation rule	**Number of outliers treated**	
	All PICTs	**Without PNG**
By HS, PICT, 1-YEAR & N>=10	2367	2014
By HS, PICT, 3-YEAR & N>=10	4529	3683
By HS, PICT, 5-YEAR & N>=10	1942	1625
By HS, PICT, 7-YEAR & N>=10	1074	911
By HS, PICT, 9-YEAR & N>=10	647	535
By HS, PICT, 11-YEAR & N>=10	441	365
By HS, PICT, 13-YEAR & N>=10	295	237
By HS, PICT, 15-YEAR & N>=10	178	158
By HS, PICT, 17-YEAR & N>=10	143	110
By HS, PICT, 19-YEAR & N>=10	117	91
By HS, PICT, 21-YEAR & N>=10	94	63
By HS, PICT, 23-YEAR & N>=10	81	65
By HS, PICT, 25-YEAR & N>=10	64	62
By HS, PICT, 27-YEAR & N>=10	43	39
By HS, PICT, 29-YEAR & N>=10	51	40
By HS, PICT, 31-YEAR & N>=10	41	26
By HS, PICT, 33-YEAR & N>=10	47	31
By HS, PICT, 35-YEAR & N>=10	17	13
By HS, PICT, 37-YEAR & N>=10	19	18

By HS, PICT, 39-YEAR & N>=10	11	12
By HS, PICT, 41-YEAR & N>=10	11	7
By HS, PICT, 43-YEAR & N>=10	10	12
By HS, PICT, 45-YEAR & N>=10	5	4
By HS, PICT, 47-YEAR & N>=10	1	1
By HS, 1-YEAR & N>=10	560	529
By HS, 3-YEAR & N>=10	255	228
By HS, 5-YEAR & N>=10	63	72
By HS, 7-YEAR & N>=10	25	21
By HS, 9-YEAR & N>=10	12	13
By HS, 11-YEAR & N>=10	10	12
By HS, 13-YEAR & N>=10	5	2
By HS, 15-YEAR & N>=10	6	5
By HS, 17-YEAR & N>=10	1	2
By HS, 19-YEAR & N>=10	1	0
By HS, 21-YEAR & N>=10	3	1
By HS, 23-YEAR & N>=10	0	3
By HS, 25-YEAR & N>=10	2	1
By HS, 27-YEAR & N>=10	0	1
By HS, 29-YEAR & N>=10	2	2
By HS, 31-YEAR & N>=10	2	2
By HS	2	2

— Appendix 10: Specific changes for each commodity set reviewed in stage 5, structured by chapter.

Chapter 01: Live animals

All 743 records in this chapter were deleted from the database because it was assumed that, of the small quantity of live animals traded within the region, the vast majority was traded for breeding purposes rather than direct consumption. The assumption was confirmed in discussion national experts. The total quantity of HS01 deleted from the database was 8,107 t, including 5,970 t of live poultry and 1,250 t of bovine animals.

Chapter 02: Meat and edible meat offal

Chapter 2 contains frozen and fresh whole, and cuts, of bovine, sheep, poultry, swine, and other. Separate quantity matrices were reviewed for each of these 5 animal groups for each PICT. No commodities within Chapter 2 were deleted in their entirety. 1 record had exporter name changed, 42 records had quantity changed, and 209 records were deleted. At the end of stage 4, the total tonnage of HS02 was 3,119,613 t across 17,766 records. At the end of stage 5, the total tonnage of HS02 was 3,078,420 t across 17,556 records.

Chapter 03: Fish, crustaceans, molluscs, aquatic invertebrates

Chapter 3 was also disaggregated to a number of groups; 'baitfish', 'coastal', 'invertebrates', and 'other'. Tuna; fresh, frozen, and chilled, was not reviewed. A significant volume of fish is imported to various PICTs to be used as bait in oceanic fisheries. Three commodities in particular, frozen sardines and pilchards (HS030371), frozen mackerel (HS030374), and frozen cuttlefish and squid (HS030749). A regional fisheries expert, William Sokimi, and M. McCoy, a regional tuna fishery expert, were consulted on the data for these categories. There was a high correlation between the presence of a longline fishery in a country/year and the important of large quantities of baitfish. It was concluded that the great majority of the tonnage of these commodities imported to the PICTs was for bait, so these were removed from the database.

Additionally, there were a large number of records of Pacific Salmon (HS030212, HS030310, HS030541) export from the region. Pacific Salmon is neither wild-caught or farmed within PICTs included in this study, or likely to be re-exported in significant volumes, so was assumed to be an attribution error. In this instance we deleted all records (n = 196; 53,788 t) of export from PICTs, assuming that importing countries applied the correct classification attribution in their reporting. The other major set of deletions were various tuna and frozen fish, n.e.s. trades between PICTs and Mauritius. Tuna fishery expert, M. McCoy suggested that these records (n = 50; 43,719 t) would primarily relate to Pacific flagged vessels operating in the Indian Ocean.

Cleaning Chapter 03 resulted in change of exporter name for 1 record, change of quantity for 4 records, and deletion of 2,754 records. Of the deleted records, 2,216 were baitfish, 196 were Pacific Salmon, 50 were trades with Mauritius, mostly comprising tuna exports, and 292 were for other reasons determined through detailed matrix review. At the end of stage 4, the total tonnage of HS03 was 9,059,170 t across 29,673 records. At the end of stage 5, the total tonnage of HS03 was 8,146,879 t across 26,919 records. Most of this reduction is explained by the removal of baitfish.

Chapter 04: Dairy, eggs and honey

Chapter 4 was disaggregated to two groups, dairy and eggs, and honey. All HS 6-digit codes were retained. Dairy data was relatively clean across PICTs. Within this chapter 129 records were deleted (11,317 t) due to implausibility, and 7 records had changes to quantity. At the end of stage 4, the total tonnage of HS04 was 1,363,979 t across 15,598 records. At the end of stage 5, the total tonnage of HS04 was 1,350,256 t across 15,469 records.

Chapter 07: Vegetables

Chapter 7 was reviewed as a single group. No commodities were entirely removed. 30 records were deleted, 3 records had their quantity changed, and 1 record had both quantity and value ($) changed. At the end of stage 4, the total tonnage of HS07 was 2,201,001 t across 20,120 records. At the end of stage 5, the total tonnage of HS07 was 2,182,081 t across 20,090 records.

Chapter 08: Fruits and nuts

Chapter 8 was reviewed as a single group. No commodities were entirely removed. 28 records were deleted. At the end of stage 4, the total tonnage of HS08 was 478,018 t across 17,351 records. At the end of stage 5, the total tonnage of HS08 was 473,586 t across 17,323 records.

Chapter 09: Coffee, tea, mate, and spices

Chapter 9 was reviewed as a single group. A significant number of changes were conducted for this chapter, primarily due to incorrect HS code attribution and incorrect reporting, particularly for tea and coffee, and errors likely associated with reporting of different codes by trade partners. In total 68 records were assigned different HS codes, mostly changing from coffee husks and skins (HS090130) to coffee and substitutes (HS090140). Quantity was changed for 38 records, and 479 records were deleted as implausible. At the end of stage 4, the total tonnage of HS09 was 1,862,040 t across 18,306 records. At the end of stage 5, the total tonnage of HS09 was 1,719,342 t across 17,829 records.

Chapter 10: Cereals & Chapter 11; Milling products

Because the major staples imported to the Pacific, rice and wheat, span these two chapters they were considered a single set for the purposes of review. Within the entire set we reviewed four separate groups, rice and rice flour, wheat and wheat flour, other grains, and other milling products. Due to significant cleaning required, each of the four groups are outlined separately below.

Wheat and wheat products

Wheat and wheat flour included seven HS6 commodity codes dominated, in terms of tonnage, by HS100190 (meslin and wheat other than durum) and HS110100 (wheat or meslin flour). None of the seven commodities were removed from the database. The most significant change to wheat and wheat flour commodities was substitution of Australian export records to major PICTs with flour mills (FJI, PNG, NCL, SLB with Australian Bureau of Statistics data for HS100190 (meslin and wheat). This change was made due to an error in trade records in both Comtrade and BACI data (see main text for elaboration). At the end of stage 4, the total tonnage of wheat and wheat products was 7,959,482 t across 2,634 records. At the end of stage 5, the total tonnage of wheat and wheat products was 9,418,402 t across 2,636 records; a negligible change to the number of records, but a significant change in tonnage.

Rice and rice flour

Rice and rice flour included six unique HS6 commodities, dominated, in terms of tonnage, by HS100630 (rice, semi-milled or wholly milled). Rice and rice flour presented significant challenges. It was clear, during high-resolution inspection in stage 5, that there were major errors in rice data for a number of PICTs, most significantly in Fiji, Papua New Guinea, Solomon Islands, Vanuatu, and Western Samoa. For these countries the official Comtrade records, BACI data, and the PFTD data to end of stage 4 had major irregularities including significant gaps in data across years. For example, imports to PNG in the years 1997, 1998, 1999 were 487 t, 182,692 t, and 1,311 t respectively. Rice production in PNG was negligible during this time so does not explain the difference. Similarly, Solomon Islands data showed virtually zero rice imports up to and including 2008, then around 30,000 – 45,000 t in proceeding years. Solomon Islands does not produce meaningful quantities of rice.

Most of the observed error was corroborated by United Nations Comtrade, so substitution with unadjusted Comtrade data was not tenable. Numerous alternate data sources were reviewed in an attempt to triangulate and generate plausible estimates. Fiji Bureau of Statistics data were reviewed for Fiji import records. Numerous sources including M. R. Bourke and Harwood (2009), International Merchandise trade records from Australian Bureau of Statistics, and International Food Policy Research Institute estimates (Schmidt & Fang, 2021) were reviewed for Papua New Guinea. R. Bourke et al. (2006) and International Merchandise trade records from Australian Bureau of Statistics were reviewed for Solomon Islands. Vanuatu national statistics office data was reviewed for Vanuatu. All of these potential data sources differed significantly from Comtrade and stage 4 PFTD data, and varied among themselves. In the absence of plausible nationally reported data, the most suitable source was Food and Agriculture Organisation statistics derived from their detailed trade matrix (https://www.fao.org/faostat/en/#data/TM). Consequently, the PICTs of concern, Fiji, PNG, Solomon Islands, Vanuatu, and Western Samoa, rice data across HS100610, HS100620, HS100630, and HS100640 was substituted with FAOSTAT estimates across all years. FAOSTAT data did not include information on exporting country, so revised estimates are each aggregates for each HS6 within each year for the PICTs of concern.

Additional cleaning was conducted on rice data including deletion of a series of records of importation of broken rice to Federated States of Micronesia. This deletion was executed after review of the data by a representative of the Federated States of Micronesia Division of Statistics. As for wheat, some of these changes have resulted in dramatic changes to the total tonnage of trade, highlighting the need for great caution when setting policy based on global data. At the end of stage

4, the total tonnage of rice and rice products was 6,157,508 t across 3,281 records. At the end of stage 5, the total tonnage of rice and rice products was 6,943,236 t across 2,365 records.

Other Cereals

Other cereals contained nine unique HS6 codes. Three of these were removed from the database during this stage of cleaning. Maize (corn) seed (HS100510) was removed because it was confirmed that this was primarily used for planting. Grain sorghum (HS100700) and buckwheat (HS100810) were removed because it was determined that it was rarely consumed in the Pacific and where imported was done so for stock feed. The remaining data was relatively clean with only some deletions such as a highly anomalous import record of 30,000 t of maize from Serbia to Republic of Marshall Islands in 1999. At the end of stage 4, the total tonnage of other cereals was 733,700 t across 1,471 records. At the end of stage 5, the total tonnage of other cereals was 220,646 t across 1,082 records.

Other Flours

Other flours contained 27 unique HS6 commodities, none of which were removed at this stage. While there were some anomalies and a number of clearly implausible records, these commodities had relatively clean and stable data. At the end of stage 4, the total tonnage of other flours was 660,599 t across 6,249 records. At the end of stage 5, the total tonnage of other flours was 631,329 t across 6,181 records.

Chapter 12: Oil seeds and fruits

Chapter 12 was reviewed as a single group. Two commodities were removed from the database. Palm nut and kernels (HS120710) were removed because they are not for human consumption in their unprocessed state, and the large quantities exported from Papua New Guinea and Solomon Islands could skew interpretations. Castor oil seeds (HS120730) were removed because they are not primarily for human consumption. One record in this chapter had the exporter name changed, one record had the importer name changed, two records had quantity changes and 166 records were deleted. Of the deleted records, 155 were HS120710 and two were HS120730. At the end of stage 4, the total tonnage of HS12 was 8,692,850 t across 4,310 records. At the end of stage 5, the total tonnage of HS12 was 2,994,648 t across 4,144 records.

Chapter 15: Fats and oils

Chapter 15 was reviewed as a single group. Three commodities were removed from the database. Castor oil (HS151530) and Tung oil (HS151540) were removed because they are not primarily for human consumption. Animal or vegetable fats and oils and their fractions (HS151800) was removed because it is defined as inedible mixture. Four records in this chapter had quantity changes and 541 records were deleted. Of the deleted records, 83 were HS151530, 13 were HS151540, and 403 were HS151800. At the end of stage 4, the total tonnage of HS15 was 15,211,801 t across 15,602 records. At the end of stage 5, the total tonnage of HS15 was 14,913,018 t across 15,061 records.

Chapter 16: Meat and seafood preparations

Chapter 16 was reviewed as three separate commodity groups; meat preparations, fish preparations, and invertebrate (shellfish, octopus etc.) preparations. No commodities were removed from the database. 17 records in this chapter had exporter name changed, one record had importer name changed, 2 records had quantity changed, and 70 records were deleted. At the end of stage 4, the total tonnage of HS16 was 1,769,283 t across 20,456 records. At the end of stage 5, the total tonnage of HS16 was 1,757,430 t across 20,386 records.

Chapter 17: Sugar and sugar confectionary

Chapter 17 was reviewed as a single group. No commodities were removed from the database. This chapter presented some challenges because, as with chapter 9, there were numerous instances where commodity definitions are similar, resulting in importer and exporter reporting the same trade as a different commodity. This simple error in reporting highlights the potential ramifications of using global datasets without considering the associated assumptions. The consequence of this is an effective doubling of the reported quantities. Determining instances of this required numerous approaches including converting import tonnages to per capita availability to ascertain whether the total imports were plausible, and reviewing national records. Where there were instances of clear double-reporting, the least plausible record was deleted. Additionally, the database included very large quantities of sugar beet export from Fiji, mostly to USA. These records were not identified when checked against Fiji national statistics, and some were consequently deleted. At the completion of cleaning Chapter 17, two records had exporter name changed, 16 records had HS code changed, 133 records had quantity changes, and 632 records were deleted. At the end of stage 4, the total tonnage of HS17 was 9,618,882 t across 11,334 records. At the end of stage 5, the total tonnage of HS17 was 9,393,012 t across 10,702 records.

Chapter 18: Cocoa and cocoa preparations

Chapter 18 was reviewed as a single group. No commodities within Chapter 18 were deleted from the database. 4 records had changes to quantity and 30 records were deleted. At the end of stage 4, the total tonnage of HS18 was 1,420,045 t across 8,966 records. At the end of stage 5, the total tonnage of HS18 was 1,418,876 t across 8,936 records.

Chapter 19: Cereal, flour, starch, milk preparations…

Chapter 19 was reviewed as a single group. No commodities within Chapter 19 were deleted from the database. One record had exporter name changed, two records had changes to quantity, and 58 records were deleted. At the end of stage 4, the total tonnage of HS19 was 1,300,317 t across 25,792 records. At the end of stage 5, the total tonnage of HS19 was 1,291,407 t across 25,734 records.

Chapter 20: Preserved fruit and vegetables

Chapter 20 was reviewed as a single group. No commodities within Chapter 20 were deleted from the database. Six records had changes to HS codes, 11 records had changes to quantity, and 45 records were deleted. At the end of stage 4, the total tonnage of HS20 was 1,219,506 t across 38,421 records. At the end of stage 5, the total tonnage of HS20 was 1,200,054 t across 38,372 records.

Chapter 21: Miscellaneous food preparations

Chapter 21 was reviewed as a single group. No commodities within Chapter 21 were deleted from the database. Two records had changes to exporter name, 5 records had changes to quantity, and 34 records were deleted. At the end of stage 4, the total tonnage of HS21 was 1,066,226 t across 21,979 records. At the end of stage 5, the total tonnage of HS21 was 1,063,668 t across 21,944 records.

Chapter 22: Beverages

Chapter 22 was reviewed as a single group. Ethyl alcohol (HS220720) was deleted from the database because it is not for human consumption. 11 records had changes to exporter name, 4 records had changes to importer name, 15 records had changes to quantity and 476 records were deleted. Of the deleted records 400 were ethyl alcohol. At the end of stage 4, the total tonnage of HS22 was 5,646,175 t across 26,107 records. At the end of stage 5, the total tonnage of HS22 was 5,595,018 t across 25,631 records.

Chapter 24: Tobacco and manufactured tobacco substitutes.

While not strictly food or beverages, tobacco products are consumed, and play a significant role in the health outcomes for populations so this chapter is retained in the database. Chapter 24 was reviewed as a single group. No commodities within Chapter 24 were deleted from the database. 7 records had changes to quantity and 58 records were deleted. At the end of stage 4, the total tonnage of HS24 was 82,153 t across 3,548 records. At the end of stage 5, the total tonnage of HS24 was 77,331 t across 3,490 records.

Chapter 25: Salt, sulphur…

From this chapter, only salt (HS250100 salt (sodium chloride) including solution, salt water) existed within the database at the end of stage 4. Isolating salt for human consumption from other uses was problematic, notably because large quantities were imported to some PICs to be used by the oceanic purse seine fleets in brine immersion systems to freeze tuna at sea (see DeBeer et al. 2019 for review). Importation of large quantities of salt was correlated with PICs with purse seine transhipment ports such as Tarawa (KIR), Majuro (FSM), and Wewak, Lae, Madang, and Rabaul (PNG; Mike McCoy unpublished analysis). In 2018, for example, 7,864 t of salt was imported to Kiribati, which is equivalent to ca. 68 kg per person. The picture was further complicated by observations that salt imported for brining tuna was instead used for human consumption in some ports (Mike McCoy pers. comm.). These complexities, along with those created by other uses of salt, such as in the manufacture of plastic bottles and in stock feeds meant that was considerable uncertainty in correctly attributing salt imports to use. Salt (HS250100) was therefore removed from the database pending further in-depth analysis of salt in the Pacific food system. A total of 1,768 records and 683,416 tonnes were removed from the database, representing 0.57% of trade flows and 0.85% of total quantity at the end of stage 4.

— Appendix 11: Commodities excluded from HS92 Chapters HS01 – HS25 in the development of the database. Some entire chapters were excluded for reasons outlined below. Other Chapters were completely retained. Some exclusions occurred during stage 1, and other exclusions occurred at stage 5 when more knowledge was available on the role of specific commodities. Complete HS92 commodity descriptions can be found through the BACI download page, accessible here: http://www.cepii.fr/CEPII/en/bdd_modele/presentation.asp?id=37

HS Chapter code	Commodities excluded	Excluded commodity description (abbreviated)/reason for chapter exclusion
HS01	All excluded	Primarily for breeding purposes and not direct consumption
HS02	All retained	
HS03	HS030110	Fish: live, ornamental
	HS030371	Fish: sardines etc., frozen
	HS030372	Fish: haddock, frozen
	HS030373	Fish: coalfish, frozen
	HS030374	Fish: mackerel etc., frozen
	HS030749	Molluscs: cuttle fish and squid, frozen, dried, salted or in brine
HS04	All retained	
HS05	All excluded	Animal products not for human consumption
HS06	All excluded	Plant products not for human consumption
HS07	HS070110	Vegetables: seed potatoes, fresh or chilled
HS08	All retained	
HS09	All retained	
HS10	HS100700	Cereals: grain sorghum
	HS100810	Cereals: buckwheat
	HS100830	Cereals: canary seed
HS11	All retained	
HS12	HS120710	Oleaginous fruits: palm nuts and kernels, whether or not broken
	HS120720	Oil seeds: cotton seeds, whether or not broken
	HS120730	Oil seeds: castor oil seeds, whether or not broken
	HS120792	Oleaginous fruits: shea nuts (karite nuts)
	HS120911	Seed: sugar beet seed, of a kind used for sowing
	HS120919	Seed: beet seed, (excluding sugar beet), of a kind used for sowing
	HS120921	Seed: lucerne (alfalfa) seed, of a kind used for sowing
	HS120922	Seed: clover (trifolium spp.) seed, of a kind used for sowing
	HS120923	Seed: fescue, of a kind used for sowing
	HS120924	Seed: kentucky blue grass seed, of a kind used for sowing
	HS120925	Seed: rye grass seed, of a kind used for sowing
	HS120926	Seed: timothy grass seed, of a kind used for sowing
	HS120929	Seed: of forage plants, other than beet seed, n.e.s. in item no. 1209.2
	HS120930	Seed: of herbaceous plants cultivated principally for their flowers
	HS120991	Seed: vegetable seed, of a kind used for sowing
	HS120999	Seed: n.e.s. in heading no. 1209, of a kind used for sowing

	HS121110	Liquorice roots used primarily in perfumery etc.
	HS121120	Ginseng roots, used primarily in perfumery etc.
	HS121190	Plants and parts n.e.s. used primarily in perfumery etc.
	HS121300	Cereal straw and husks: unprepared, whether or not chopped..
	HS121410	Lucerne (alfalfa) meal and pellets
	HS121490	Forage products including swedes, mangolds, fodder roots etc.
HS13	All excluded	Saps and gums not for human consumption
HS14	All excluded	Plant products not for human consumption
HS15	HS150510	Animal fats and oils: wool grease, crude
	HS150590	Animal fats and oils: wool grease (other than crude) and fatty substances
	HS151530	Vegetable oils: castor oil and its fractions, whether or not refined
	HS151540	Vegetable oils: tung oil and its fractions, whether or not refined
	HS151800	Animal or vegetable fats and oils and their fractions
	HS151911	Stearic acid
	HS151912	Oleic acid
	HS151913	Tall oil fatty acids
	HS151919	Industrial monocarboxylic fatty acids n.e.s. in heading no. 1519
	HS151920	Acid oils from refining
	HS151930	Industrial fatty alcohols
	HS152010	Glycerol (glycerine): crude: glycerol waters and glycerol lyes
	HS152090	Glycerol: n.e.s. in heading no. 1520, including synthetic
	HS152110	Vegetable waxes (other than triglycerides): whether or not refined
	HS152190	Waxes, other than vegetable, whether or not refined or coloured
	HS152200	Degras: residues resulting from the treatment of fatty substances etc.
HS16	All retained	
HS17	All retained	
HS18	HS180200	Cocoa: shells, husks, skins and other cocoa waste
HS19	All retained	
HS20	All retained	
HS21	All retained	
HS22	HS220720	Ethyl alcohol and other spirits: denatured, of any strength
HS23	All excluded	Bran and sharps residue, and oil cake not for human consumption
HS24	All retained	
HS25	All excluded	Salt primarily used for tuna fishery