A method for cleaning trade data for regional analysis: The Pacific Food Trade Database (version 2, 1995-2018)

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

Citation:

Brewer, T.D., Andrew, N.L., Sharp, M.K., Thow, AM., Kottage, H., Jones, S. (2020). A method for cleaning trade data for regional analysis: The Pacific Food Trade Database (version 2, 1995-2018). Pacific Community working paper. *

*A comprehensive version of this research is in progress for peer review publication.

Acknowledgements

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. N. Troubat and A. Dunn provided insightful and extensive feedback on the manuscript. We are indebted to in-country experts from PICTs for their contribution to stage 2.1 of the cleaning process. 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 [ACIAR projects FIS/2016/300, FIS/2018/155].

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 the analysis and interpretation of trade flows among countries. The frequently used source of globally standardised data, the United Nations Comtrade database, is appropriate for coarse analysis but has limitations caused by the availability and accuracy of information reported by nations. Exploratory analysis of food trade flows with and among 18 countries in the Pacific region revealed significant and consequential errors in UN Comtrade and the derived CEPII BACI databases. We describe a stepwise cleaning process 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 and imputation of unit prices and the revision of quantity in each trade flow. For the period 1995–2018, a total of 4,634 (1.5 %) trade flows contained categorical errors in some combination of exporter * importer * commodity and were re-categorized or deleted. 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 the error. The imputation process reduced the total dataset from 314,669,653 t to 80,313,878 t. While the 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.

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. 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 et al., 2011). Increased incidence of diet related non-communicable disease presents a profound public health burden for many Pacific countries. Development of robust trade data is central for identifying trends in availability of different foods and their potential effects on public health.

Trade datasets are rarely complete and plagued with inaccuracies (Ortiz-Ospina, Beltekian, & Roser, 2019). Consequently, analysis of smaller datasets focused on individual commodities or countries, particularly in low- and middle-income countries (LMICs) are fraught with challenges. The Pacific region is an exemplar of unreliable food trade data. In many analyses, Pacific Island Countries and Territories 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 better economic and health outcomes for the region.

Here we describe the development of a food trade database for 18 PICTs in the Pacific region. This work is part of a broader analysis on food security and nutrition in the Pacific region that draws on tables of the nutrition composition of foods and national household acquisition and consumption surveys, among other sources to generate and integrated understanding of the Pacific food system. The appearance of COVID-19 has further increased demand for reliable and timely food trade data as governments scramble to analyse the implications of disrupted international and domestic supply chains (Farrell et al., 2020).

Sources for international trade data and rationale for this analysis

National trade data held by PICTs were explored for their suitability; however, due to our inability to acquire sufficiently long and consistent time-series across all PICTs and concerns surrounding insufficiently granular classification, it was necessary to use other primary data sources. Alternative data sources include a range of international organizations, notably United Nations Comtrade (https://COMTRADE.un.org), International Trade Centre (http://www.intracen.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.

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 current 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 trades in the Pacific region. Both datasets are wholly reliant on information provided to UN Comtrade by countries. BACI, however, may also introduce error because trade flows from mirror data are equally open to mis-reporting. Error generated from the use of mirror data could be exaggerated for small PICT economies (e.g. Tuvalu) when error is introduced from their larger trading partners (e.g. Australia). Sources of error and differences in estimates among data sources can be attributed to a range of factors including incorrect attribution of 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 et al., 2019; UN, 2008).

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.

Methods

Stages in database development

The described method comprised step-wise process in four 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.

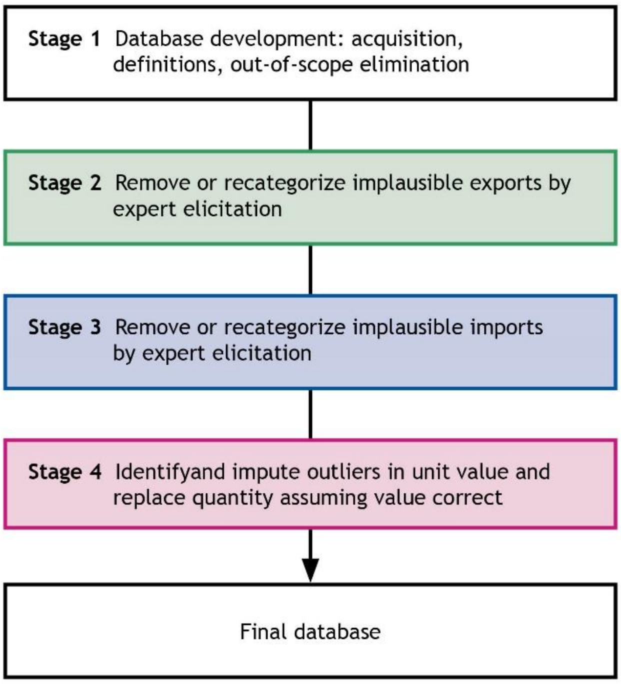

Figure 1. Sequence of stages in database development and cleaning. See text for additional description of sub-stages.

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. 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 lens through which to recognize implausible trade flows, but deletions were not be made on the basis of ‘quantity’ *per se*. As examples, consider these implausible trade flows from the raw BACI download (cleaning stage 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)

Stage 1: Database development

Stage 1.1: Data acquisition

Data were downloaded from http://www.cepii.fr/CEPII/en/bdd_modele/presentation.asp?id=1 on 25 March, 2020. The download contained all trade flows from 1995 to 2018 for all countries 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 undergoes a major review every five years to remain relevant resulting in code changes. In this application we used HS92 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 for non-human consumption were removed. To determine whether a commodity was for human consumption we used HS sub-heading definitions in conjunction with heading and chapter definitions. Where further clarification was needed, we cross referenced with the CPC v.1.1 (UN, 2002). HS sub-headings within the following HS Chapters were retained: 01-04, 07-12, 15-22, 24 and 25 (salt only, HS250100). Excluded exceptions at HS4 within these Chapters were: 1209 - seeds for sowing; 1211 - plants and parts used in perfumery and industry; 1505 – wool grease, lanolin; 1802 – cocoa husks and other waste. Although not considered a food 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, and the latter because it is administratively part of Australia.
3. All 20,661 trade flows for 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 records for American Samoa and Guam for the years 1995–1999. American Samoa, Commonwealth of Northern Mariana Islands and Guam were retained in the database as non-PICTs.

At completion of Stage 1.3, the data 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 records was US$79.6 billion and the total quantity was 315,337,498 t.

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

Table 1. Total records for included PICTs at completion of database development. 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.

Stage 2: Removing and adjusting 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 a 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 were instructed to categorize 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 at this point or 2,692 implausible trade flows (Figure 2). The total quantity removed was 344,583 tonnes, 0.63% of all PICT export quantity in the database at this point. The total value removed was US$192,174,000 or 0.45% of all PICT export value in the database at this point. Examples of trade flows removed 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)

Several PICTs, notably Fiji, Samoa and Solomon Islands import foods from outside the region and re-export them to smaller PICTs (e.g. Fiji to Kiribati and Samoa to Tokelau). 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 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 associated exporting PICT. These records were retained within the database with exporter name changed to 'Unknown'. This stage reinstated all 305 between PICT trade flows corresponding to 27,460 t (Figure 2). 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')
- 18 t of wheat flour from Kiribati to Nauru (in this instance the record was not altered)

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.

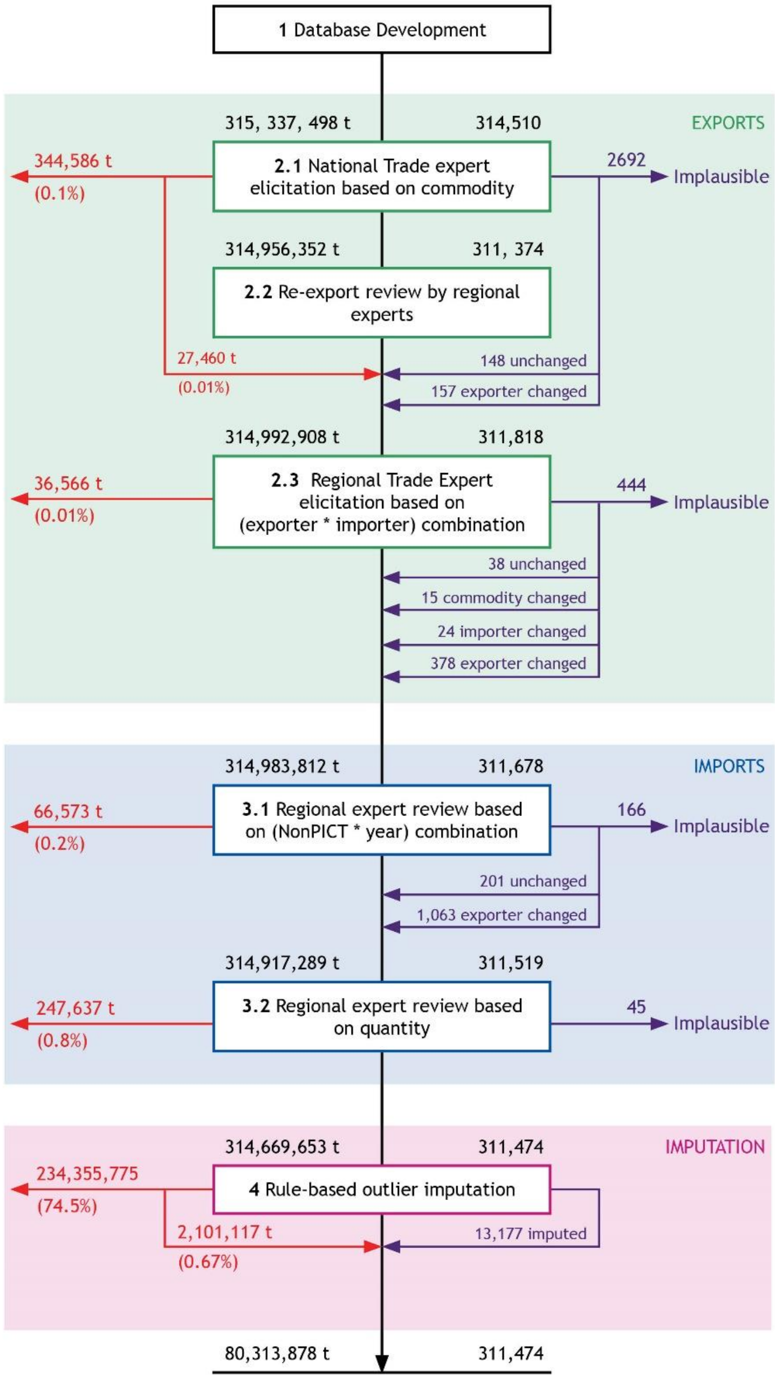

Figure 2. Outcomes at each stage of database development and cleaning. See text for additional description of sub-stages. The numbers (purple text) and quantities (red text) of trade flows reviewed and either returned or deleted are indicated. Quantities and numbers of trade flows in black refer to trade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

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 of cases the experts differed in their judgement – in these instances inclusion or exclusion of the record was decided upon on the weight of evidence (e.g. supplementary verification by other experts), with a bias toward inclusion if doubt existed. Other anomalies identified in this stage required a change either in the name of the importing countries of 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.

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. 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

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 blocks of trade flows of improbable commodities in one or several years 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). 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. 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 overall cleaning process, 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 is not included separately and is likely to comprise the majority of trade records 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 exceeds Australia's total net annual rice production of around 800,000 t, and equates to roughly 10 t per capita available for consumption in PNG
- 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 (see Figure 6 and Figure 7 below for graphical examples of the impacts of outliers on trends in trade flows). 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 (FAO, 2019). Below we describe the two steps in recognizing and treating outliers.

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 (FAO, 2019). All available trade data (1995-2018) for each HS6 code were pooled across all PICTs and years. This approach maximized 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 (FAO, 2019; Tukey, 1977) which, if the data were perfectly normally distributed, would result in ca. 0.7% of observations being recognized as outliers (Jones, 2018). Following sensitivity analysis, we set k at 1.5. Example distributions are shown in Figure 3.

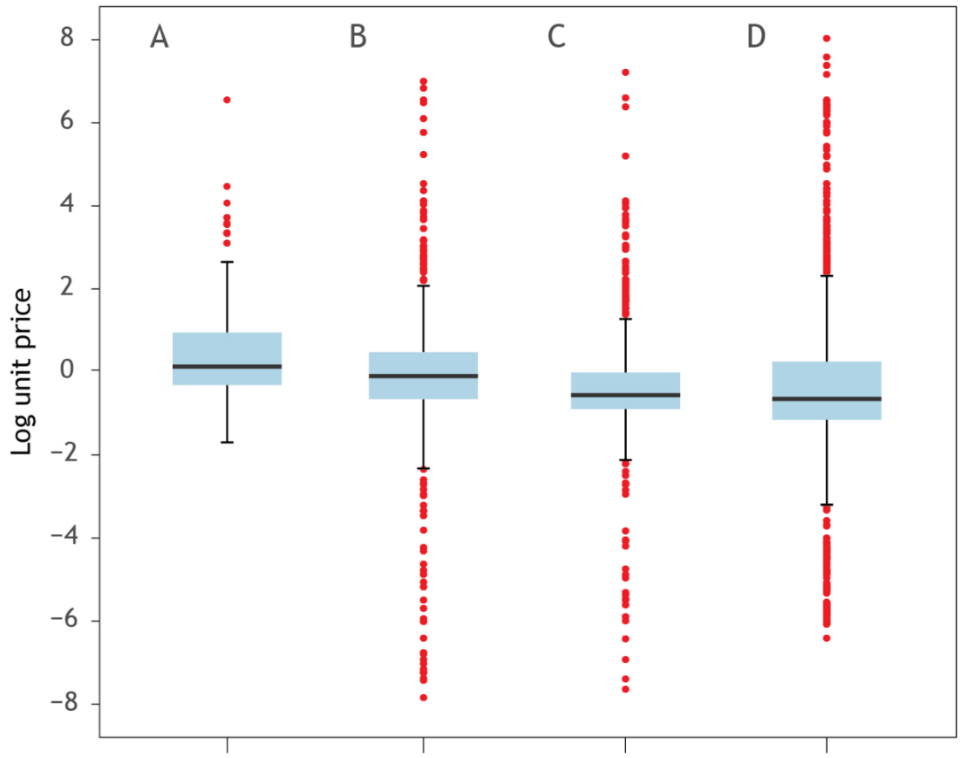

Figure 3. Illustrative box and whisker plots of log unit price of commodities showing trade flows identified as outliers at k = 1.5. Outliers shown as red dots. A = Vegetable roots and tubers: sweet potatoes, with high starch or inulin content (HS071420); B = Cereals: rice, semi-milled or wholly milled (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 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.

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

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

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 (FAO, 2019), and the most suitable method used here. Medians at HS6 were estimated from combinations of PICT * YEAR with a minimum sample size of N ≥ 20. 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.

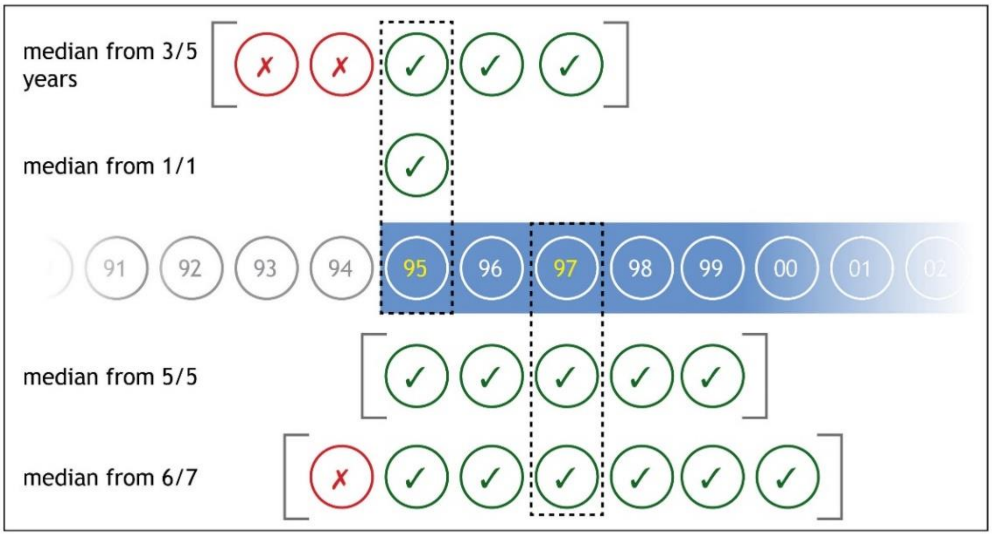

Figure 4. Illustrative examples of sample years to estimate medians of unit price in 1995 (above the timeline) and 1997 (below). The middle year in the sample contains the outlier (in dashed rectangle). Years before 1995 are outside the scope of the dataset. Sample sizes were calculated from in-scope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

Results

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 (PFTD) with the same commodities contained within the BACI and Comtrade datasets (Figure 5). 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 5A). 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 5B) is mostly driven by records of copra and palm nut and kernel. BACI methods ameliorate the observed volatility of Comtrade data. Similarly, Comtrade quantity estimates exhibited dramatic volatility across the remaining commodities (Figure 5C). This volatility is partly addressed by BACI, including increased mean estimates through the incorporation of mirror data. The PFTD (this analysis) 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 5, 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 6). 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.

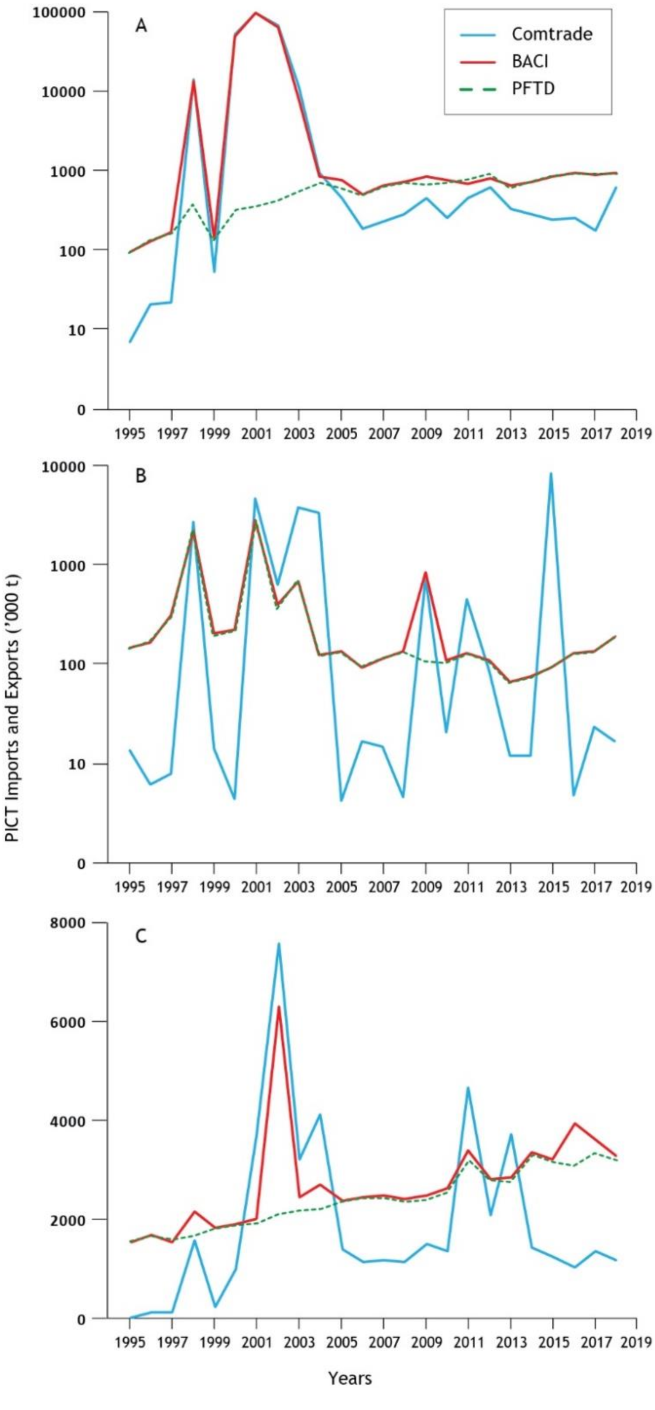

Figure 5. Selected examples of differences among databases in the quantity of food (t) traded (sum of imports and exports). UN Comtrade and CEPII-BACI are global public databases, PFTD is the research database developed in this paper by cleaning the BACI database. **A)** HS Chapter 10 (Cereals), **B)** HS Chapter 12 (Oil seeds, oleagic fruits etc.), and **C)** All other HS Chapters included within the PFTD. Comtrade data includes exports and imports reported by PICTs with the 'World'.

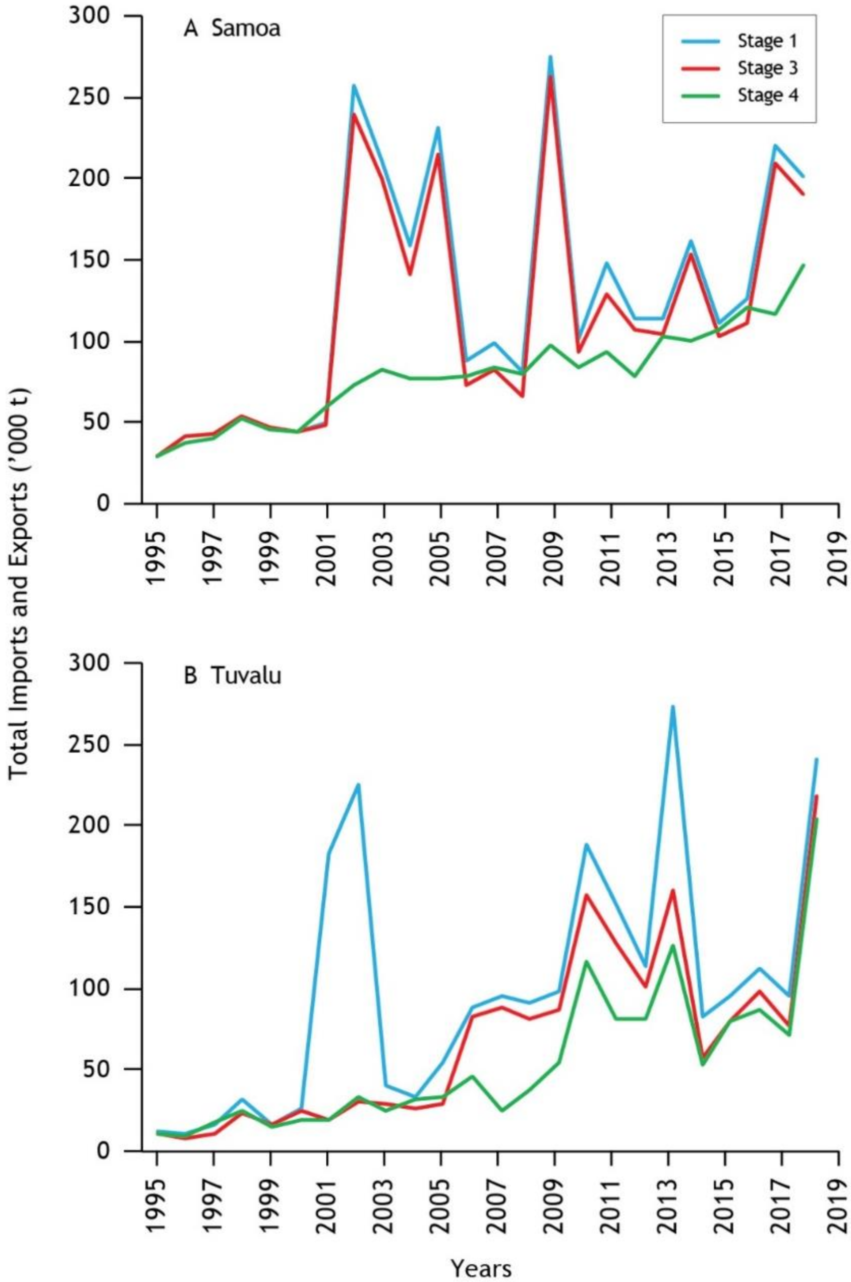

Figure 6. Total quantity (t) imported to and exported from Samoa and Tuvalu through time. Each line shows quantity trends at the end of each stage of either data preparation or cleaning. Stage 1 is BACI data, reduced to the relevant data set, as described in Stage 1.

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. 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.

PICT	Quantity	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

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

There was significant variation in the effect of stage 2 and 3 cleaning on commodity groups (Table 4). 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.

Commodity Chapter (HS2)	Quantity	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 product 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, ne (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

Table 4. Percentage of data [quantity (t), count (number) of trade flows, and value US$000’] removed in cleaning stages 2 & 3 by Commodity Chapter. Chapters are ranked in decreasing order of percentage change in quantity.

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 PNG and Samoa (Table 5). The most effected cells were rice in PNG, beverages in Tuvalu, and sugar in RMI. 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 PNG.

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

Table 5. 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 4 and Appendix 2 for HS Chapter definitions and inclusions.

Discussion

Policy Implications

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).

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).

Given their cross-sectoral value-add, robust trade data, such as those presented in this paper, also offer an opportunity to assess multiple dimensions of the impact of shocks, including implications of COVID-19 for food systems. Early reports from the Pacific indicate that impacts on trade are one of the mechanisms through which COVID-19 and associated responses may contribute to food insecurity in the region (Farrell et al., 2020). Regional trade datasets will be vital for informing policy responses by enabling the rapid generation of detailed evidence regarding the nature of the impact of such shocks.

Limitations

It is possible that, despite overall improvement in data quality, the PFTD method introduces a marginal level of error. Specifically, some instances of outlier identification in stage 4.1 might have been value errors rather than quantity errors. However, given the relative outlier rate for each quantity and value, individually, it is unlikely that this error occurred over a large portion of the 13,177 trade records identified as outliers. Further, development of the PFTD involved the removal of a number of trade records assumed to be incorrectly attributed to PICTs. Conversely, it is probable that within Comtrade there were trade flows that should have been attributed to PICTs but are not, due to country attribution error. Therefore, the PFTD likely under-represents trade flows. There is, however, no reason to assume that specific PICTs, years, or commodities are disproportionally under- or over-represented.

The database also contains minor structural errors carried over from Comtrade and BACI. First, Comtrade and BACI do not report trade flows valued at less than $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 to Comtrade, the database will not include records of trade between 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 database, 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.

Some clusters of commodities are also problematic in their interpretation. For example, trade flows for tuna as occurring in Comtrade and carried through to the PFTD may not accurately characterize trade in tuna. Tuna is frequently caught by foreign vessels, and significant transhipment occurs, so the movement of tuna between countries is poorly captured in trade data. Similarly, not all commodities in the PFTD are necessarily for human consumption. For example, mackerel, pilchards, and sardines, while suitable for human consumption, might also be imported to the Pacific to be used as bait in the tuna fishery. Another example is palm nut and kernel which is traded in large quantities, but the edible portion is only a fraction of its total quantity. Cleaning of tobacco data revealed a uniquely high error rate, suggesting a structural issue with reporting and the need to be cautious in interpreting patterns in tobacco trade for the region.

Future priorities

Database improvements will be ongoing, as errors are identified in the course of more granular analyses of individual commodities and at the level of PICTs. A series of smaller revisions will be completed at the HS sub-heading * PICT level. Because medians were imputed at the sub-heading *

PICT level the dataset is relatively modular, with little interdependence in revised quantities among commodities or PICTs, it will be possible to perform specific adjustments without compromising the database. These analyses will use the best available data, such as reliable in-country data and commodity expert knowledge, to empirically replace trade flows that can be shown to be incorrect. We will not, however, repeat Stage 4 to re-estimate non-outlier medians and re-impute the full dataset. Priority areas for more granular analysis include: (i) cereals, (ii) export cash crops, (iii) fish including tuna, (iv) animal source foods, and (v) case studies for Solomon Islands, Vanuatu and Kiribati. Additionally, this version of the database will be revised by repeating the full process described here when post-COVID-19 trade flows in 2020 become available from BACI.

Disclaimer

The Pacific Food Trade Database described herein was developed as a research tool and does not constitute an official record of trade flows among countries in the Pacific region. The views expressed are those of the authors and not their respective institutions.

Declaration of Competing Interest

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

References

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

Brooks, J., & Matthews, A. (2015). Trade dimensions of food security.

Cowling, K., Thow, A. M., & Porter, K. P. (2018). Analyzing the impacts of global trade and investment on non-communicable diseases and risk factors: a critical review of methodological approaches used in quantitative analyses. *Globalization and Health, 14*(1), 53.

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.

FAO, U., ITC. (2019). Trade data processing and validation methodology. *United Nations Department of Economic and Social Affairs*.

Farrell, P. T., AM. Tutuo Wate, J. Nonga, N. Vatucawaqa, P. Brewer, T. Sharp M. Farmery, A. Trevena, H. Reeve, E. Eriksson, H. Gonzalez, I. Mulcahy, G. Eurich, JG. and Andrew, NL. . (2020). COVID-19 and Pacific food system resilience: opportunities to build a robust response. *Food Security, https://doi.org/10.1007/s12571-020-01087-y*.

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*: Univ of California Press.

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.

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*. Retrieved from

Kerr, W. A. (2011). The role of international trade in achieving food security. *Estey Journal of International Law and Trade Policy, 12*(1753-2016-141239), 44-53.

Ortiz-Ospina, E., Beltekian, D., & Roser, M. (Producer). (2019, 26/09/2019). Trade and Globalization. [Report] 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.

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. doi:10.1080/03670244.2010.524104

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.

UN. (2002). *Central Product Classification (CPC) Version 1.1*. Retrieved from

UN. (2008). *International Merchandise Trade Statistics: Supplement to the Compilers Manual*.

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.