

Appendix: `complete_tf_cpc` module

Marco Garieri

Alexander Matrunich

Christian A. Mongeau Ospina

Bo Werth

Food and Agriculture Organization of the United Nations

4 January 2017

This document gives a faithful step-by-step sequence of the operations performed in the `complete_tf_cpc` module. For a narrative version of the module's approach, please see its main document.

Parameters

- **year**: year for processing.
- **out_coef**: coefficient for outlier detection, i.e., the **k** parameter in the *Outlier Detection and Imputation* section.
- **hs_chapters**: specific HS chapters that are downloaded (this parameter can not be set by the user as it is provided by Team B/C and hardcoded). The HS chapters are the following:

01 02 03 04 05 06 07 08 09 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 33 35 38 40 41 42 43 50 51 52 53

Input Data

Supplementary Datasets

- **hsfclmap3**: Mapping between HS and FCL codes extracted from MDB files used to archive information existing in the previous trade system (Shark/Jellyfish). This mapping is provided by a separate package: <https://github.com/SWS-Methodology/hsfclmap>
- **adjustments**: Adjustment notes containing manually added conversion factors to transform from non-standard units of measurement to standard ones or to obtain quantities from traded values.
- **unsdpartnersblocks**: UNSD Tariffline reporter and partner dimensions use different list of geographic area codes. The partner dimension is more detailed than the reporter dimension. Since we can not split trade flows of the reporter dimension, trade flows of the corresponding partner dimensions have to be assigned the reporter dimension's geographic area code. For example, the code 842 is used for the United States includes Virgin Islands and Puerto Rico and thus the reported trade flows of those territories. Analogous steps are taken for France, Italy, Norway, Switzerland and US Minor Outlying Islands.
- **fclunits**: For UNSD Tariffline units of measurement are converted to meet FAO standards. According to FAO standard, all weights are reported in tonnes, animals in heads or 1000 heads and for certain commodities, only the value is provided.
- **comtradeunits**: Translation of the **qunit** variable (supplementary quantity units) in Tariffline data into intelligible unit of measurement, which correspond to the standards of quantity recommended by the *World Customs Organization* (WCO) (e.g., **qunit**=8 correspond to *kg*). See: <http://unstats.un.org/unsd/tradekb/Knowledgebase/UN-Comtrade-Reference-Tables>

- **EURconversionUSD**: Annual EUR/USD currency exchange rates table from SWS.

Extract Eurostat Combined Nomenclature Data

1. Download raw data from SWS, filtering by **hs_chapters**.
2. Remove non-numeric codes for reporters/partners/commodities.
3. Keep only **stat_regime=4**.
4. Use standard (common) variable names (e.g., **declarant** becomes **reporter**).
5. Convert ES geonomenclature country/area codes to FAO codes.
6. Remove reporters with area codes that are not included in MDB commodity mapping area list.
7. Map HS to FCL.
8. Remove unmapped FCL codes.
9. Add FCL units.
10. Specific ES conversions: some FCL codes are reported in Eurostat with different supplementary units than those reported in FAOSTAT, thus a conversion is done.

Extract UNSD Tariffline Data

1. Download raw data from SWS, filtering by **hs_chapters**.
2. Remove non-numeric commodity codes.
3. Identical combinations of reporter / partner / commodity / flow / year / qunit are aggregated.
4. Use standard (common) variable names (e.g., **rep** becomes **reporter**).
5. Tariffline M49 codes (which are different from official M49) are converted in FAO country codes using a specific conversion table provided by Team ENV.
6. European countries are removed (will be replaced by ES data).
7. Area codes not mapping to any FAO country code are removed.
8. Re-imports become imports and re-exports become exports.
9. Map HS to FCL.
10. Remove unmapped FCL codes.
11. Add FCL units.
12. General TL conversions: some FCL codes are reported in Tariffline with different units than those reported in FAOSTAT, thus a conversion is done.
13. Specific TL conversions: some commodities need a specific conversion.
14. If the **quantity** variable is not reported, but the **weight** variable is and the final unit of measurement is tonnes the **weight** is used as **quantity**
15. Aggregate UNSD Tariffline Data to FCL.

Combine Trade Data Sources

1. Application of “adjustment notes” to both ES and TL data.
2. Convert currency of monetary values from EUR to USD using the `EURconversionUSD` table.
3. Combine UNSD Tariffline and Eurostat Combined Nomenclature data sources to single data set.
 - TL: assign `weight` to `qty`
 - ES: assign `weight` to `qty` if `fclunit` is equal to `mt`, else keep `qty`

Outlier Detection and Imputation

1. Unit values are calculated for each observation at the HS level as ratio of monetary value over quantity `value / qty`.
2. Outlier detection by using the logarithm of the unit value.
3. Imputation of missing quantities and quantities categorized as outliers by applying the method presented in the *Outlier Detection and Imputation* section. The `flagTrade` variable is given a value of 1 if an imputation was performed.
4. Aggregate values and quantities by FCL codes.
5. Map FCL codes to CPC.
6. Map FAO area codes to M49.

Mirror Trade Estimation

1. Obtain list of non-reporting countries as difference between the list of reporter countries and the list of partner countries.
2. Swap the reporter and partner dimensions: the value previously appearing as reporter country code becomes the partner country code (and vice versa).
3. Invert the flow direction: an import becomes an export (and vice versa).
4. Calculate monetary mirror value by adding (removing) a 12% mark-up on imports (exports) to account for the difference between CIF and FOB prices.

Flag management

Note: work on this section is currently in progress.

- `observationStatus`:
 - Reporting countries:
 - * `X` if `flagTrade` is zero (i.e., no imputation) and FCL unit != “\$ value only”
 - * `I` if `flagTrade` is non-zero (i.e., imputation) and FCL unit != “\$ value only”
- Non-reporting countries: `E`
- `flagMethod`:
 - Reporting countries:
 - * `<BLANK>` if `flagTrade` is zero (i.e., no imputation) and FCL unit != “\$ value only”
 - * `e` if `flagTrade` is non-zero (i.e., imputation) and FCL unit != “\$ value only”

- Non-reporting countries: **e**

Output for SWS

1. Filter observations with FCL code **1181** (bees).
2. Filter observations with missing CPC codes.
3. Rename dimensions to comply with SWS standard, e.g., **geographicAreaM49Reporter**
4. Calculate unit value (US\$ per quantity unit) at CPC level if the quantity is larger than zero
5. Transform dataset separating monetary values, quantities and unit values in different rows.
6. Convert monetary values, quantities and unit values to corresponding SWS element codes. For example, a quantity import measured in metric tons is assigned **5610**.
7. Overwrite **flagMethod** for mirrored quantities: **e** becomes **c**
8. Add **flagMethod i** to unit values