Appendix: `complete_tf_cpc` module

Marco Garieri
Alexander Matrunich
Christian A. Mongeau Ospina
Bo Werth
Food and Agriculture Organization of the United Nations

9 March 2018

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.
- `hs_chapters`: 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', '43', '50', '51', '52', '53'

Download auxiliary tables

- `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 corresponds to *kg*). See: http://unstats.un.org/unsd/tradekb/Knowledgebase/UN-Comtrade-Reference-Tables
- `EURconversionUSD`: Annual EUR/USD currency exchange rates table from SWS.
- `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.
- `fclcodes`: List of valid FCL codes.
- `livestockweights`: List of valid FCL codes.
- `hs6standard`: HS6standard will be used as last resort for mapping.
- `hsfclmap4`: Additional mapping between HS and FCL codes (extends `hsfclmap`).
- `hsfclmap`: Mapping between HS and FCL codes extracted from MDB files used to archive information existing in the previous trade system (Shark/Jellyfish). This mapping table contains (identifier: `hsfclmap5`) also some "corrections" to the original mapping found in the MDB files. These are contained in the `correction_*` variables (e.g., `corrections_fcl`), and if for a given HS range one or more of these variables are non-missing they will replace the original corresponding variable (e.g., if `corresponding_fcl` is non-missing, it will replace `fcl`). Missing HS to FCL links in the MDB files are mapped by Team B/C and stored in a table (identifier: `hsfclmap4`) that will extend the original mapping table. *[Note: for reference, the actual name of the initial mapping table is `hsfclmap3`; the naming convention of these tables should probably be made more logical or, at least, more easily identifiable.]* The resulting mapping table gets subsetted with the condition that the`startyear` and `endyear` of the HS to FCL links should satisfy the condition: *startyear* <= *year* <= *endyear*.

- `force_mirroring`: Datatables for those reported that need to be treated as non-reporters as mirroring is required.
- `corrections_table`: Table with corrections applied during the validation process.
- `unsdpartnersblocks`: UNSD Tariffline reporter and partner dimensions use different list of geographic are 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.

Download raw data and basic operations

1. Download Eurostat data (ES) available in `trade-input-data:ce_combinednomenclature_unlogged_YEAR` datatables (label: "EU Commission - Combined Nomenclature YEAR").
2. Download Tariff line data (TL) available in `trade-input-data:ct_tariffline_unlogged_YEAR` datatables (label: "UNSD Tariffline YEAR").
3. Keep only `stat_regime` = 4 in ES.
4. Remove European-aggregated data (i.e., totals) from ES.
5. Use standard (common) variable names (e.g., `declarant` becomes `reporter`) in ES and TL.
6. Filter HS codes of interest, i.e., codes that do not participate in further processing. Such solution drops, e.g., all HS codes shorter than 6 digits.
7. Remove non numeric reporters / partners / hs codes from ES and TL.
8. Use standard (common) variable types in ES and TL.
9. Apply specific HS corrections. Some HS codes in some countries need specific HS corrections. As on 2018-03-08 only a subset of HS codes for a given TL reporter are corrected (tonnes were reported instead of kilograms, so raw data was multiplied by 1000).
10. Convert ES geonomenclature country/area codes to FAO codes.
11. TL M49 codes (which are different from official M49) are converted in FAO country codes using a specific conversion table (`unsdpartnersblocks`) provided by Team ENV. Then, these M49 codes are converted to FAO codes.
12. Remove invalid reporters (i.e., keep countries/areas that existed in the year considered).
13. Force mirroring, i.e., remove countries that appear as official reporters so that the mirroring procedure will estimate their data.
14. Remove ES reporters from TL.
15. Apply explicit corrections to the HS-FCL mapping.
16. Extend the `endyear` for those combinations of `area` / `flow` / `fromcode` / `tocode` for which `endyear` < `year`.
17. Add additional codes that were not present in the HS-FCL original mapping file.
18. Keep HS-FCL links for which `startyear` <= `year` & `endyear` >= `year`
19. Generate HS to FCL map at HS6 level (if switched on)
    1. Universal (all years) HS6 mapping table.

2. Current year specific HS6 mapping table.

Specific operations on Eurostat data

1. Add variables that will contain flags. (Note: flags are set in various steps in the code. Please, refer to the "Flag Management in the Trade module" document.)
2. Remove in ES those reporters with area codes that are not included in MDB commodity mapping area list.
3. Map HS codes to FCL.
    1. Extract HS6-FCL mapping table.
    2. Extract specific HS-FCL mapping table.
    3. Use HS6-FCL or HS-FCL mapping table.
    4. Use HS6 standard for unmapped codes.
4. Remove unmapped FCL codes (i.e., transactions with no HS to FCL link).
5. Add FCL units, i.e., the target units of measurement of the items (e.g., tonnes, heads).
6. Specific conversions: some FCL codes are reported in Eurostat with different supplementary units than those reported in FAOSTAT, thus a conversion is done.

Specific operations on Tariff line data

1. Do mathematical conversions on specific `qunits`: 6 (pairs), 9 (thousands), and 11 (dozens) become 5 (units), by multiplying them by 2, 1000, and 12, respectively.
2. Identical combinations of `reporter` / `partner` / `commodity` / `flow` / `year` / `qunit` are pre-aggregated.
3. Add variables that will contain flags. (Note: flags are set in various steps in the code. Please, refer to the "Flag Management in the Trade module" document.)
4. Area codes not mapping to any FAO country in the HS to FCL mapping codes are removed.
5. Re-imports become imports and re-exports become exports.
6. Map HS codes to FCL.
    1. Extract HS6-FCL mapping table.
    2. Extract specific HS-FCL mapping table.
    3. Use HS6-FCL or HS-FCL mapping table.
    4. Use HS6 starndard for unmapped codes.
7. Remove unmapped FCL codes (i.e., transactions with no HS to FCL link).
8. Add FCL units, i.e., the target units of measurement of the items (e.g., tonnes, heads).
9. General conversions: some FCL codes are reported in Tariffline with different units than those reported in FAOSTAT, thus a conversion is done.
10. Specific conversions: some FCL codes are reported in Tariff line with different supplementary units than those reported in FAOSTAT, thus a conversion is done.

11. If the `weight` variable is available and the final unit of measurement is tonnes then `weight` is used as `quantity`.

Combine Trade Data Sources

1. Convert currency of monetary values from EUR to USD using the `EURconversionUSD` table (required only for ES).
2. Convert data in thousands of dollars.
3. Combine Tariff line and Eurostat data sources in a single data set:
   - TL: assign `weight` to `qty`.
   - ES: assign `weight` to `qty` if `fclunit` is "mt", else keep `qty`.
4. Unit values are calculated for each observation at the HS level as ratio of monetary value over quantity: $uv = value/qty$.

Imputation

1. Imputation of missing quantities by applying the method presented in the *Missing Quantities Imputation* subsection of the *faoswsTrade: `complete_tf_cpc` and `total_trade_CPC` modules* document (*Standardization, editing and outlier detection* section). The `flagTrade` variable is given a value of 1 if an imputation was performed.

Additional operations

1. Separate flags.
2. Aggregate values, quantities, and flags by FCL codes.
3. Map FCL codes to CPC (version 2.1).
4. Map FAO area codes to M49. Countries with FAOSTAT code 252 ("Unspecified") are converted to M49 code 896 ("Other nei").

Mirror Trade Estimation

1. Create a table with the list of reporters and partners combined as areas and count the number of flows that the areas declare as reporting countries. The partners that never show up as reporters or the reporters that do not report a flow will have a number of flows equal to zero and will be mirrored.
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 aggregation

Flags are aggregated as mentioned in the *Flags* section in the main documentation or, more in depth, in the "Flag Management in the Trade module" document.

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. Keep officially reported weight in kilograms for livestock. Besides of quantities in "heads" or "1000 heads" if a country reported also the weight, it will be kept and saved to SWS.
6. Use corrections set by analysts during the validation process.
7. Transform dataset separating monetary values, quantities and unit values in different rows.
8. Convert monetary values, quantities and unit values to corresponding SWS element codes. For example, a quantity import measured in metric tons is assigned `5610`.
9. Generate metadata for corrections.
10. Removed "protected" data from the module's output.
11. Remove transactions saved on SWS that are not generated by the module.

Save data

Finally, data is saved in the `completed_tf_cpc_m49` dataset of the `trade` domain.