Food Balance Sheet workflow in the Statistical Working System

Cristina Muschitiello
Food and Agriculture Organization of the United Nations

24 May 2018

Abstract

This vignette provides a description of the workflow and dependencies of operations in the Statistical Working System for the production of Food Balance Sheets.

Contents

List of Tables

List of Figures

Disclaimer

This Working Paper should not be reported as representing the official view of the FAO. The views expressed in this Working Paper are those of the author and do not necessarily represent those of the FAO or FAO policy. Working Papers describe research in progress by the authors and are published to elicit comments and to further discussion.

This paper is dynamically generated on May 24, 2018 and is subject to changes and updates.

The Overall Workflow of The FBS

The process of creating FBSs starts by collecting all data for the different variables of the Food Balance Sheet Framework[1]. The data of each variable are generally checked and imputed in time series. The set of operations required for creating/checking time series of data for each variable is called *module*. A ***module***, in the FBS Framework, is an R-script, written by an R-developer and integrate inside the ***Statistical Working System (SWS)***[2] by means of *plugins*. There is at least one module (there might be more) for each variable of the FBS. Each module produces figures that are collected in a dataset inside the SWS for future uses or publication. Output data of a module may become input data of another module, this circumstance creating a precise sequence for the execution of a complete FBS.

Definitions

B. Standardization and Balancing

1 The Overall Workflow

The *Standardiation & Balancing* process is presented in Figure 1. It involves 5 main steps and requires a some auxiliary information table.

The 5 steps are:

1. Data Pull,
2. Sua Filling,
3. Standardization,
4. Balancing,
5. FBS aggregation.

The additional information's tables are:

- *Utilization Table*,
- *Zero-Weight* table,
- *cut* table,
- *Fbs Tree.*

[1] For definitions and an extended description of the motivation behind the development of FBS, see FAO, 2001, *Food Balance Sheets: A Handbook*, available at: http://www.fao.org/docrep/003/X9892E/X9892E00.HTM. Accessed on 19 January 2017. Moreover see *Standardization & Balancing for Food Balance Sheet Calculation*, in the *Standardization & Balancing* module's documentation on *GitHub*

[2] SWS is an internal Working System providing a platform for statisticians and statistical clers to collect, collate, validate and correct data. Moreover, the platfors supports the possibility of performing imputations of data based on statisticians' knowledge and development.

All the steps and the table will be described across the present document, after having introduced the the basic equation of FBS and the *Commodity tree* which represents the structure of the set of commodity involved in the Food Balance Sheets.

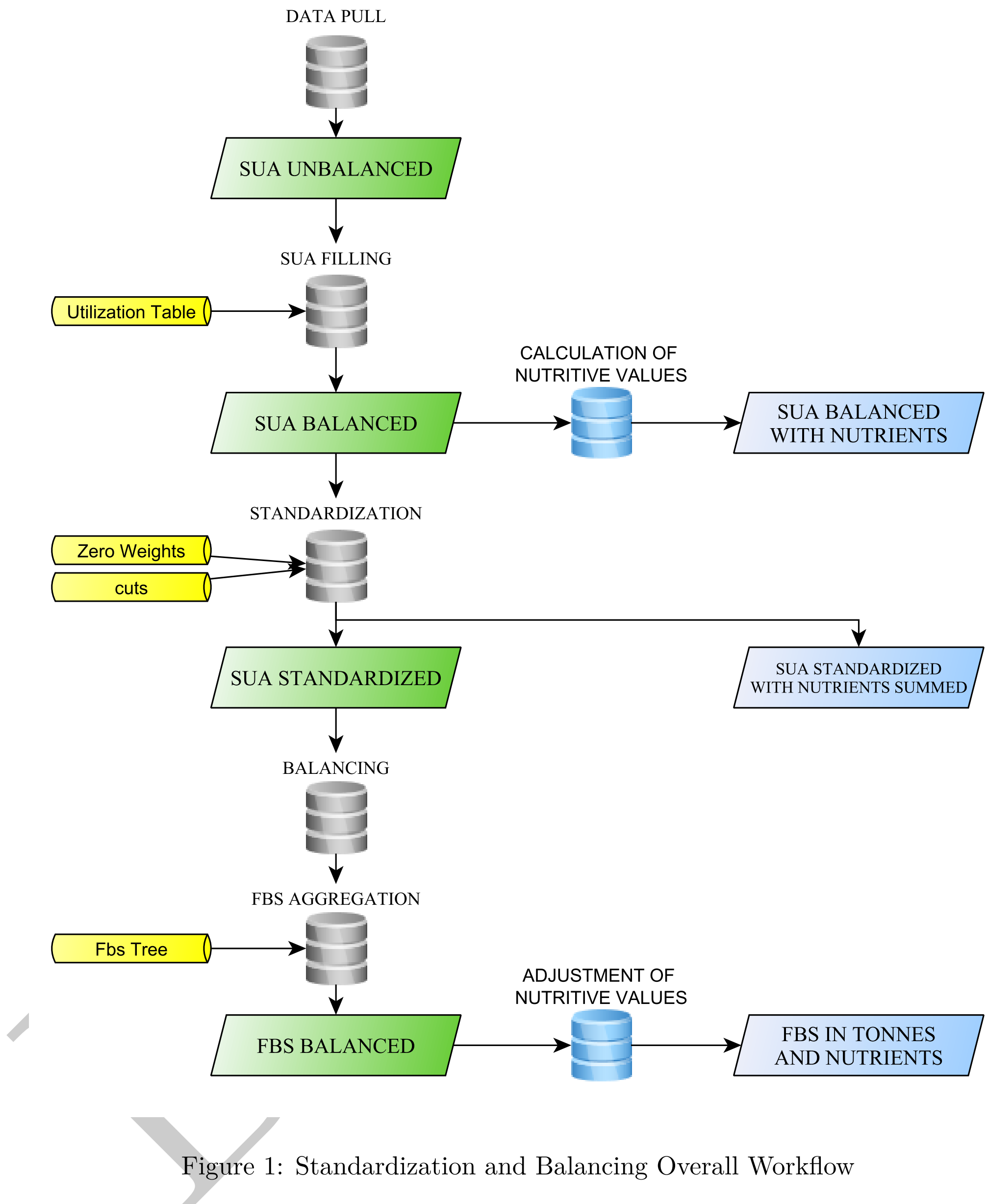

Figure 1: Standardization and Balancing Overall Workflow

Commodity Tree

The process of combining commodity balances for creating Food Balance Sheets is based on a structured and clear set of relationships between commodity given by the *Commodity tree*. The majority of the commodities are produced from one (or more) commodity (/ies), called *parent* commodity(/ies), and/or are themselfes parent of one (or more) *child* (*children*) commodity (/ies). These structure creates an intense and articulated network of relationships at different levels: primary commodities, like crops, are *parent* commodities and, also, *zero-level* commodities from which *children* commodities of *level-1* are produced, which are in turn, used to produce other commodities of a gradually *"lower"* level. In commodtity trees, the bigger the level number, the lower the processing level. There are as many commodity trees as the number of process chains in a country. Fundamental characteristics of commodity trees are[3]:

1. Each commodity tree is represented as a flowchart of the kind presented in Figure 2 where:
   - ***nodes*** represent commodities,
   - ***edges*** represent production processes ,
   - ***joints*** indicate where a single production process creates more that one commodity. These commodities are, then, called *by-products* or *co-products*.

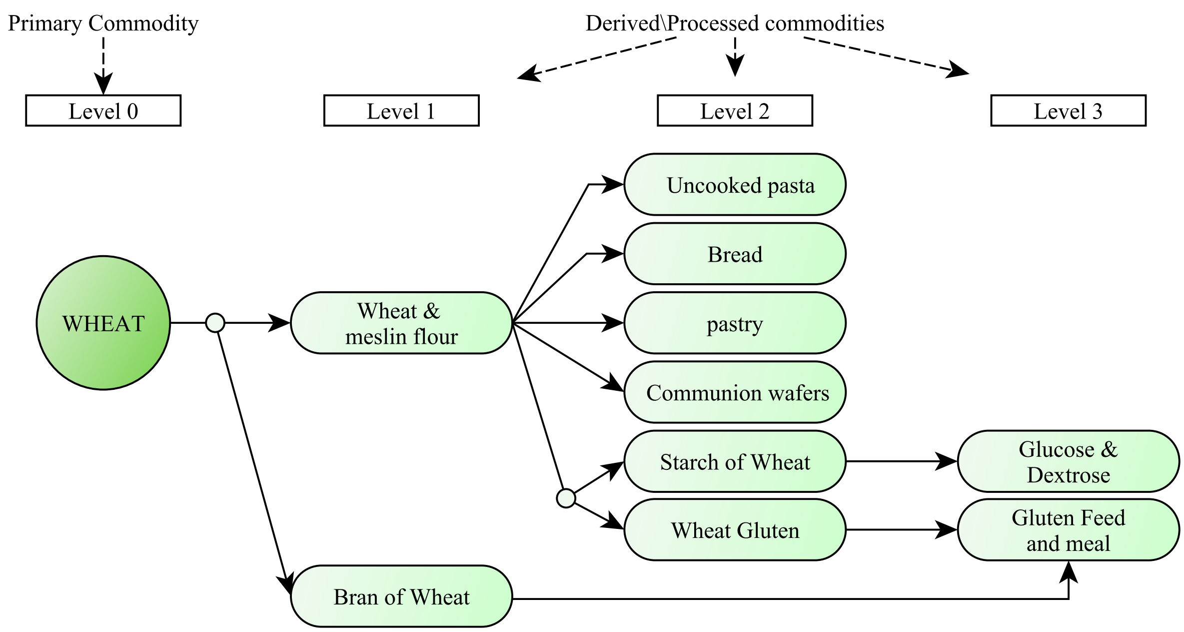

Figure 2: Commodity Tree for Wheat in China mainland 2014

2. Not all the countries have the same production processes, because countries have different technologies and primary products availabilities. Therefore, the commodity trees are not the same across countries.

3. if a production process is active in a country, this is expressed throught the existence of a conversion factor called ***extraction Rate***. An Extraction rate (eR) represents how much amount of the child commodity is produced from 1 unit of parent commodity. It is expressed as a ratio of the processed product obtained from the processing of the parent/originating product.

4. Some child commodities can be produced starting from more that one parent commodity. A second conversion factor exists representing the amount of a child commodity that is produced from each parent commodity. This conversion factor is called ***Share***. Shares represent the amount of the child commodity that is produced from the specified parent and are expressed as a ratio. Shares are generically defined as:

$$s_{cp} = \frac{availability_{p(c)}}{\sum_{p=1}^{A} availability_{p(c)}} \tag{1}$$

   where $availability_{p(c)}$ is the availability of each parent p of child c expressed in terms of c (in *child equivalent*).

[3]For a more detailed description of commodity trees, please see the specific documentation. The reference document, at the moment is the *tecnical conversion factor* document available in the documentation folder on *GitHub*

Commodity trees are presented in tables like Table 10, which represents the same example of Figure 2. In the table each production process is represented in a separate row.

There are some concepts linked to the *Commodity tree* framework:

- ***Proxy-Primary*** commodities. These are a set of commodities that are children of other commodities but, because they are important in representing the food availability of a country, are not aggregated to their primary commodities, but are kept separated. These commodities are *cut* from the tree of the primary commodity/ies and, if they can be processed in other products, have their own commodity tree. The name *proxy-primary* is assigned because they are considered as primary-commodities in the *Standardization & Balancing* process.
- ***No-Tree*** commodities. These are *zero-level* commodities that are primary commodities and are never processed in other products. As they are not involved in any production process, there is no tree associated to them. Notice that, even a commodity that is included in the commodity tree of one Country, might be a No-tree commodity for another country or another year, if no production processes have been activated for that specific Country or year.

2 Data Pull

The process of creating FBSs starts by considering the initial commodity balance for each CPC commodity, either primary and derived, with the different variables of the equation (as listed and briefly described in the previous section) filled with figures as available from official or other sources and from imputation and estimation methods, when applied. In other words, the process starts by pulling figures inside a so-called *Sua Unbalanced*. In this initial account, food processing and ROU figures are not available (because they, by default, will be measured during the process), whereas the figures for all other variables have been already collected, imputed and estimated through a specific *module* (Figure 3).

A ***module***, in the FBS Framework, is an R-script, written by an R-developer, for the execution of a set of operations (either data import, manipulation, imputation or estimation) required for compiling the time series of one variable. There is at least one module (there might be more) for each variable of the FBS. Each module produces figures that are collected in a dataset inside the ***Statistical Working System (SWS)***[4].Output data of a module may become input data of another module, this circumstance creating a precise sequence for the execution of a complete FBS. In the present document, we are analyzing the workflow of the Standardization process as starts after all the modules have run and have produced reliable data or each variable. The detailed description of the workflow for the execution of all the modules of the FBS will be given in a separate document.[5].

the Initial Sua Unbalanced

After pulling all data, the process of compiling Food Balance Sheets is a non-complete supply-utilization account. The non-completeness of the SUAs is due to different reasons: first, as already said, some variables are not collected, nor estimated before the process begins, second there is the model for industrial use that does not impute or estimate data, but just collects data from different sources, this opening the strong possibility not to have values where they are supposed to exist and, also, not guaranteeing consistency of data over time. Thirdly, modules might, sometimes, fail in the imputation, because of the strong complexity and structural diversity of the input data.

[4]SWS is an internal Working System providing a platform for statisticians and statistical clers to collect, collate, validate and correct data. Moreover, the platfors supports the possibility of performing imputations of data based on statisticians' knowledge and development.

[5][report the document when it will be available]

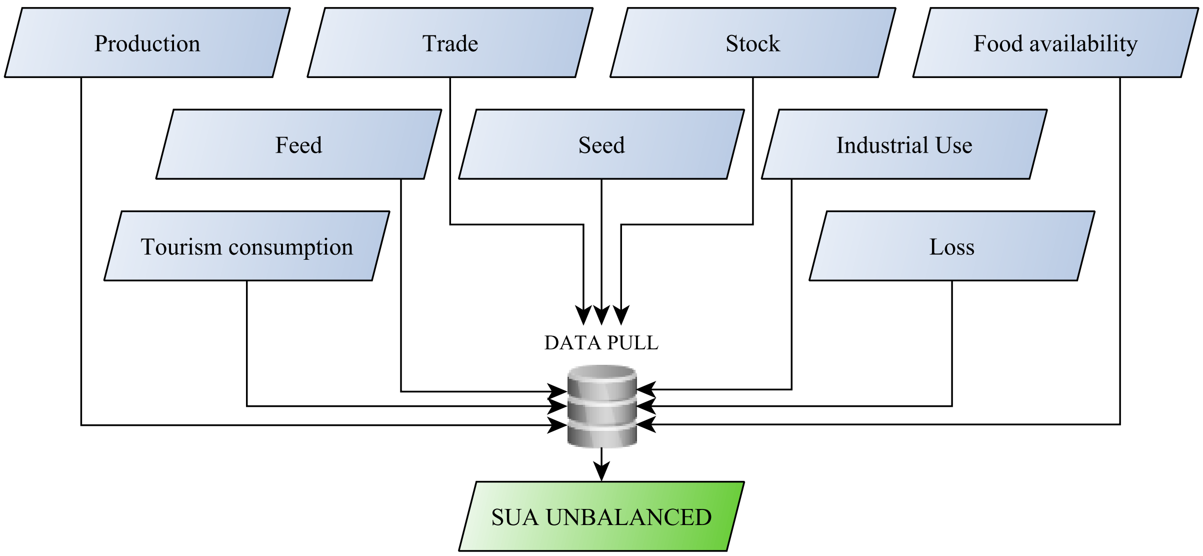

Figure 3: Data Pull from datasets containing data for each separate variable

Table 1: Unbalanced Sua table - China/Wheat/2014 example

itemName	P	I	X	DSt	Fo	FP	Fe	Se	T	IU	L	ROU
Wheat	*126,208,400*	*2,971,249*	*957*	*1,120,565*		-	*29,181,617*	*4,277,567*		*2,985,279*	*2,713,000*	-
Wheat and meslin flo	70,500,000	33,055	188,674		67,300,000	-			-17,345			-
Mixes and doughs for		6,497	38,072		0	-	0		0			-
Other Fructose and S	126,277	3,659	162,324		0	-			0			-
Starch of Wheat	239,816	11,035	40,311			-	172,196			7,919		-
Wheat Gluten	25,580	877	117,373			-	0					-
Communion wafers	13,263	8,796	5,822		16,241	-			-4			-
Uncooked pasta	1,415,692	12,520	22,550		1,405,661	-			-362			-
Food Preparations of		69,686	21,977		47,709	-			-12			-
Bran of Wheat	21,414,279	156,359	2,200		16,500,000	-	4,827,244		-4,252			-
Gluten Feed and Meal	793,740	160,231	529,333			-						-
bread	15,485	2,897	4,210		14,175	-			-3			-
pastry	193,950	89,593	117,630		165,914	-			-43			-

[a] P=Production, I=Import, X=Export, DSt=Delta Stock, Fo=Food Availability, FP=FoodProcessing, Fe=Feed, Se=Seed, T=Tourism Consumption, IU=IndustrialUse, L=Loss, ROU=Residual and other uses

S