Standardization & Balancing
Overall Workflow of the new methodology

Cristina Muschitiello
Food and Agriculture Organization of the United Nations

16 May 2018

Abstract

This vignette provides a description of the Workflow of the Standardization and Balancing Procedure: This represents the process of aggregating and balancing all the accounts of individual products to their primary equivalents. Other documents will provide details of each step of the overall procedure.

Contents

Disclaimer

1 The Food Balance Sheet Framework

1.1 Introduction

A food balance sheet can be defined as an aggregated and analytical data set that "presents a comprehensive picture of the pattern of a country's food supply during a specified reference period."[1] FBS are presented as products accounts, where the quantities allocated to all the sources of total supply must be equal to the quantities allocated to all the sources of utilization. This balance is compiled for every food item consumed within a country at primary commodity equivalent basis, and all of the primary commodity equivalent balances are then combined into a single overall FBS. FBSs are, then, expressed in terms of per capita supply for each food item by dividing by the country's population, with the per capita supplies being expressed both in terms of quantity and, through the application of food conversion factors, in terms of caloric value, protein, and fat content. These per capita estimates of caloric value for individual food products are then summed to obtain the total daily per capita Dietary Energy Supply (DES) of a country.

While FBS are typically only published at the primary commodity equivalent level to facilitate interpretation, they are created starting from information on all the derived products processed by each primary. For example, in most cases, a balance for wheat alone would in most cases include little or no food use, because wheat is commonly processed into flour before it is consumed by humans, and flour is then used to produce various other derived products such as bread, pastries and pasta. Because there is both supply and demand for each of these products (both primary and derived), individual accounts should be kept for both the primary product and all of its derived products. The process of obtaining FBS starting from product's accounts is the so-called standardization. This process starts from individual accounting balances for individual products and leads to the creation of a FBS for the primary commodity equivalent products, passing trough various steps, from filling of missing data to the balancing of the account. In the FBS framework the *Central Product Classification (CPC)* System is used for classifying commodities[2].

All the steps of the standardization process are performed inside the standardization & Balancing module, described in this document.

1.2 The Balancing equation and its variables

At the most basic level, Food Balance Sheets are, like all commodity balances, simple identities. In these identities, the sum of all supply variables is equal to the sum of all demand variables; the two most common identities set domestic supply equal to domestic demand (first equation) or total supply equal to total demand (second equation).

$$P_{ijt} + I_{ijt} - X_{ijt} - \Delta St_{ijt} = Fo_{ijt} + Fe_{ijt} + Lo_{ijt} + Se_{ijt} + IU_{ijt} + T_{ijt} + ROU_{ijt} \tag{1}$$

$$P_{ijt} + I_{ijt} - \Delta St_{ijt} = X_{ijt} + Fo_{ijt} + Fe_{ijt} + Lo_{ijt} + Se_{ijt} + IU_{ijt} + T_{ijt} + ROU_{ijt} \tag{2}$$

where:

- P=Production
- I=Imports
- X=Exports
- S=Stock level
- ΔSt_t = Stock Variation = $St_t - St_{t-1}$

---

[1]For this definition and a more 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.

[2]The CPC represents a comprehensive classification of products into a system of categories that are both exhaustive and mutually exclusive. It is based on a set of internationally agreed concepts, definitions, principles and classification rules. The custodian of this classification is the UNSD. For more information, see the *Guidelines on International Classifications for Agricultural Statistics* and the *UNSD official document on CPC Vesion 2.1*

- *Fo*=Food availability
- *Fe*=Feed
- *Lo*=Losses
- *Se*=Seed
- *IU*=industrial use
- *T*=Tourist consumption
- *ROU*=Residual Other Use.

In the equation the i index runs over all countries, the j index over all commodities, and t over years. Ideally, as many variables as possible should be measured and measurement should take place with a maximum degree of accuracy. At international level, the primary data source that FAO uses to compile the the Supply Utilization Accounts/Food Balance Sheets are the data as collected through the annual *Agriculture Production Questionnaires*. In reality, measured values are mostly limited to variables on the supply side (production, imports and exports), while, on the demand side, most estimates are imputed data. Moreover, some of the variables on the demand side are never measured and are always calculated during the standardization process itself. In particular[3]:

- *Production (P)*: Data on production are data at farmgate level. As data on production are very important for countries, these data are very often survey-based data. Nevertheless, not all countries collect data of production for all commodities. Therefore other data collecion methods are used, like records of private firms and commodity organization. When no other data are available Production is imputed or estimated. Imputation and estimation procedures depends on the specific commodity. There are different for crops and livestock but all based on an *ensemble approach*ˆ[See *Production module* documentation for more details*. Production data in the FBS framework are needed for all the primary commodities and for a set of derived commodities.

- *Import(I) and Export(X)*: Data on Trade are, mainly, official from international trade databases, like UNSD and EUROSTAT, at HS6 commodity level[4]. Official data are integrated with supplementary data having the main aim of filling in all the information hidden by the unrecorded trade and coming, mainly, from trading partners. HS6 classification is more detailed than the used CPC classiffication, therefore, these data are aggregated in CPC commodities before being used in the standardization process[5].

- *Stock Variation (ΔSt_t)*: In the FBS framework, stocks are considered as *changes in stocks* from one time period to the next. Moreover they are considered as a component of supply. Therefore, the - sign indicates that the stock is decreased, which means that the stocks are available as a supply, while the + sign indicates that the stocks have increased and they are, therefore, considered as a utilization of that commodity. Changes in stocks are tipically limited to a short number of commodities, mainly grains, pulses and sugar and, because they are very rarely measured by country, figures are very often imputed or estimatedˆ[For a complete list of stock commodities and for details about the imputation methodology, please refer to specific documentation for stock}. Estimation of Changes in stock is based on opening stocks figures throught an approach that mantain time consistency of data that are available and official[6].

- *Food availability (Fo)*: Food availability is defined as the quantity of any substance that is available for human consumption at the retail level by the country's resident population during a given reference period. Official data of food availability come from questionnaires, industrial ouptut surveys and household consumption or expenditure surveys. When these sources of data are not available, food availability data are imputated or estimated. Not all CPC commodities are Food commodities, as not all

---

[3]For more details on FBS variables please see the latest version of the *Resource Book*

[4]Harmonized Commodity Description and Coding Systems (HS) is an internationalclassification of products held by UNSD. Is made of six-digit level codes and used worldwide for trading data classifications. See official *HS6 UNSD webpage* for mode details

[5]See *Trade module* documentation for more details

[6]All data are marked as *official*, *semi − official* or *unofficial*, depending on the source they come from, throught *flags*. Flag management is one of the core responsibilities of the *Office of the Chief Statistician* Department in FAO. Flags are used from all the estimation procedure for distinguishing between different level of reliability in the data. The most reliable data are used to estimate missing or less reliable data.[this has to be better specified]

commodities are used for human consumption. Food commodities are divided in two main groups: *Food Estimates* and *Food Residual* and estimated differently depending on the pertaining group, respectively as linear or logaritmic function of income elasticity of demand, GDP per capita, and population, or as residual quantity of production and net trade quantities[7].

- *Feed (Fe)*: Feed demand is increasing because of the increase of income in developing countries. Animal feed may vary among countries due to the difference in livestock and the diversity of commodity used for livestock's rations among different countries. Official feed demand data might be available from specific questionnaires. Even when available, these data need to be cross-checked against livestock availability in terms of requirements. When official data, and also other sources of semi-official data, are not available, feed data are estimated as a function of livestock availability and livestock feed demand in terms of energy and protein requirements, in accordance with an inventory of the potential feed supply's products of any country.

- *Seed (Se)*: Official seed data may come from agricultural surveys, while other sources of data might be found in some technical publication. When data are not available these are estimated as a function of a seeding rate and a sown area in the following year.

- *Tourism Consumption (T)*: Tourism consumption is considered here as a separate utilization variable, while in the past it was included in the "other utilization" catchall cathegory. Official data for this variable are rare and may come from tourism offices or collected by tourism boards trought surveys. UNWTO is an alternative source of data or other authorities. Imputations and estimations of tourist food are made as a function of food figures.

- *Industrial Use (IU)*: This variable refers to utilization of any food items in any non-food industry. Non-food use of food commodities is growing and is highly context and country specific. For this reason there are not, at the moment, suggestions on how to impute and estimate missing figures. As a consequence, Industrial data available for Food Balance Sheets are only those coming from Official or unofficial sources. At the moment the data used comes from USDA and from questionnaries.

- *Loss (L)*: FAO has developed the Global Food Loss Index (GFLI) that focuses on the supply-side aspects of improving the efficiency of global food supply chains. The index is based on a set of primary commodities that are key in agricultural production systems, including crops, livestock, and fisheries. In order to track losses without compounding production variability, losses are expressed as a percentage and are aggregated using fixed quantities and prices. The primary data source that FAO uses for compiling GFLI are loss factors as collected by Questionnaires. Other sources are publications and reports from subnational reports, academic institutions, international organizations and so on. The missing data are imputed using a hyerarchical model based on commodity groups[8].

- *Residual and other use (ROU)*: ROU is used to capture categories of products that do not follow in any other category and that might be considered "not important" for the FBS scope. Normally, these residual commodities are different from country to country and for this reason they fall in this variable. RoU are set not to be higher that 5% of total supply and are calculated at-post as absorbing element, in the sense that it absorbs part or all of the imbalance that may exist, at FBS commodity aggregate level, after the standardization process. Any imbalance bigger than 5% of totla supply is balanced thrugh a balancing mechanism that will be later specified.

One more variabel has to be mentioned which is not included in the function, because is somehow inluded in *Production*, but is very important in the process of creating FBS. This is the *Food Processing*.

- *Food Processing (FP)*: This variable represents the amount of the availability of a commodity that enter a manufacturing process to be tranformed in a derived commodity. Food processing is not officially measured, nor collected via official sources. This variable is entirely calculated during the standardization process by applying extraction rates to the amount of production of the derived commodity. This will be better specified in section XXX of the present document.

---

[7]For a complete list of food commodities of the two typologies and for details about the imputation methodology, please refer to specific documentation for Food availablity

[8]ask for links to a proper documentation

2 The Workflow

2.1 Data Pull

The process starts by considering the initial SUA for each commodity, 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 method, 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 specific *module* (Figure 1).
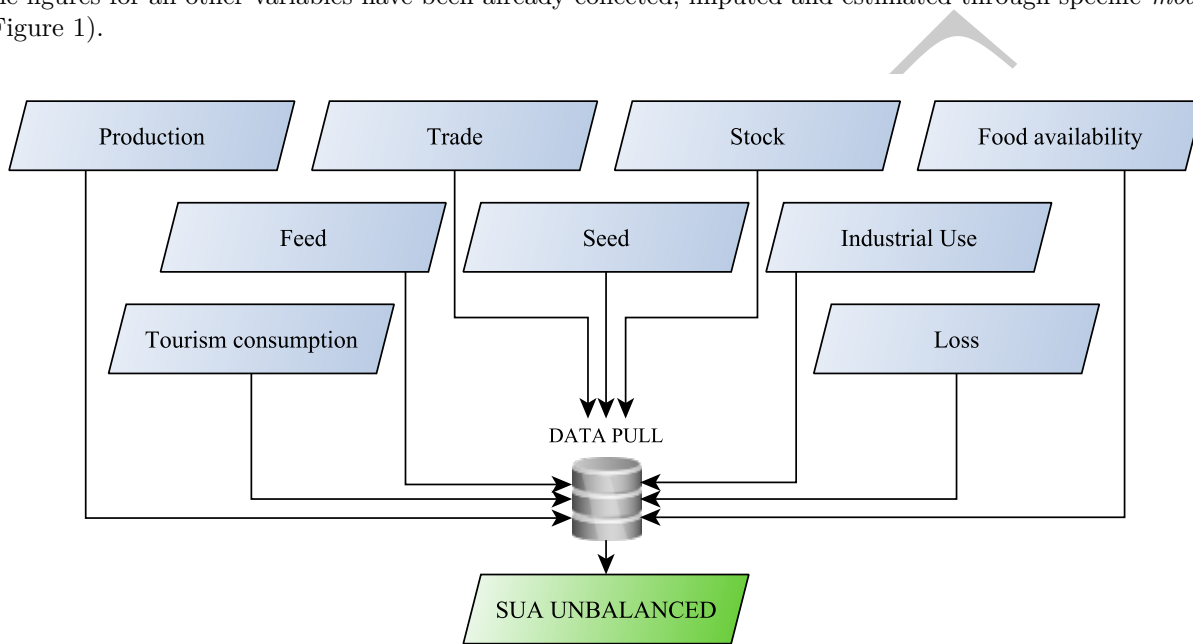

Figure 1: Data Pull from datasets containing data for each separate variable

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)*[9].Output data of a module may become input data of another module, this 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 module have run and have produced reliable data or each variable. The detailed description of the workflow for the execution of al the module of the FBS is described in a separate document[10].

2.2 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

---

[9]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.

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

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.

Supply utilization accounts are typically organized into tables where the SUA for the primary commodity is at the top, and the SUAs for all of the products derived from that commodity follow (Table 1). Commodities in the SUA table are in a parent-child relationship representing a real commodity process. In the example of Table 1 the relationship between commodities is the following:

- Maize is processed in flour, Bran and Breakfast Cereals,
- Flour of maize is processed into Starch and Gluten, Wafers, pastry, bread and pasta
- Starch is processed into Glucose and Dextrose,
- Gluten and Bran are processed into some feed and meal,

All these relationships happens at different levels, creating a tree of relationships representing what is called the **Commodity Tree**. In particular, our example represents the commodity tree of Wheat in China in 2014 and is displayed, in its graphical form, in Figure 2. A commodity Tree is the representation of all possible parent-child relationship between commodities at all levels and might be different from country to country. A better descritpion of Commodity tree is given in a separate document[11].
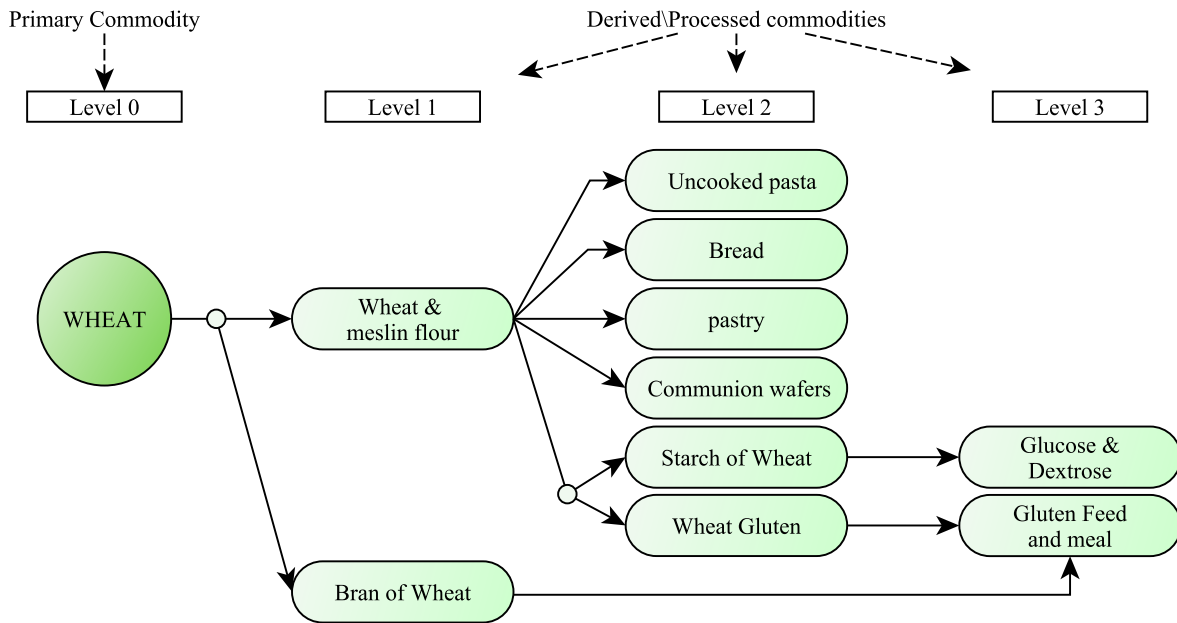

Figure 2: Commodity Tree for Wheat in China mainland 2014

---

[11]The reference document, at the moment is the *tecnical conversion factor* document available in the documentation folder on *GitHub*

Table 1: Unbalanced Sua table (Example China Mainland, Wheat 2014)

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			-

[*] 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

There are unbalanced SUAs for all combinations of commodity tree/country/year. These tables are the starting poin of the *Standardization and Balancing* process, which starts with the "Filling" of the missing figures in the SUAs.

2.3   The *Sua Filling*

Main rules and the rationale

The Standardization, i.e. the conversion of the variables of any child commodity in primary-equivalent commodity, requires all variables of the SUAs to be filled (exept for ROU, which is a residual variable and is imputed at the very last step of theprocess), "all variables" meaning "all variables that are supposed to be filled". Indeed, not all variables have to exist for all commodities. Consider maize as an example: if there is no figure of *food availability* for maize, that does not mean that there is a missing figure, because maize is rarely used for human consumption directly. However, in some country people do eat raw maize and, therefore, for those countries a food figure is expected and, if missing, that has to be taken into account.

A simplifided workflow of *Sua Filling* is reported in Figure 3.
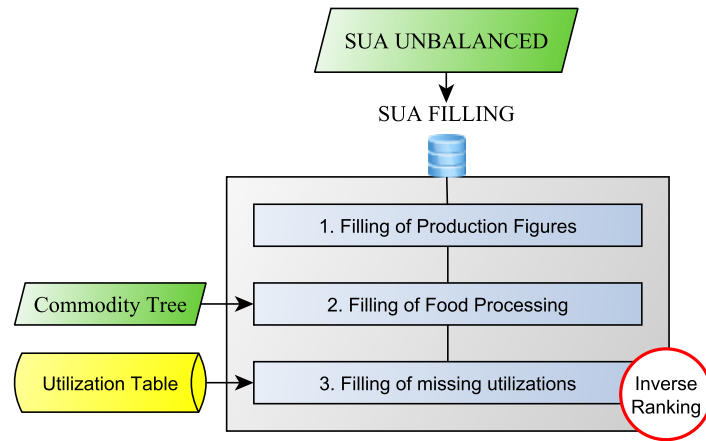
Figure 3: The Sua filling simplified workflow

Food Processing needs to be computed first, which does not come from any module. For the food processing to be correclty computed, a check on the other variables as to be performed, in order to be sure that the food processing might be correclty imputed. In particular, Production is checked first, as the *Food Processing* figure of any commodity is based on the production one. Then *Food processing* can be computed and, after that, other missing figures, if existing, are filled.

This last step requires two major and not straightforward steps, representing each a separate issue, for solving which, an algorithm has been developed that responds to the following rationale and rules:

1. *Identify which are the missing values that have to be filled.*
2. *Decide how to fill those missing values.*
3. *Preserving reliability of sources of data*

Not all the variables (supplies and utilizations) are to be filled:

- Trade (import and Export) data are shared with countries and published by FAO. Therefore these figures have a separate process of validation that makes them reliable enough not to be questioned during the standardization process.

8

- Production figures of primary commodities are also published by FAO. As previously mentioned, production values can have various flags which assign different degrees of reliability to these figures. The *sua filling* treats production figures accordingly to flags.
- Stock figures are also not touched during the process of filling. This decision was taken because the standardization process works year by year, while *stock variation* is a time-based variable that have to be always treated looking at the data over time. The cross year approach is incompatible with this characteristic. Any inconsistent or missing value for stock found would require to go back to the module imputing figures and check for possible errors. If no errors can be found, a manual correction would be required.

Information about the remaining variables, which are all utilizations, is taken from FBSs and SUAs of past years. FBSs published with the previous methodology and their corresponding SUAs can give information about which are the utilizations supposed to be active for every single commodity in each country.

Therefore, these information have to be given externally, as the process is automated and there is non-human knowledge intervening in training the machine that is performing the calculations. The previous methodology used for compiling FBS made massive use of manual interventions while the new methodology tries to avoid it, but taking advantage of the work and effort of the country experts that compiled the FBSs in the past. Past information are embedded in the new methodology trought the use of the *Utilization table* which helps idenifying which variables have to be filled, commodity by commodity, while the amount of the to be assiged to sigures is identified through a set of rules, known as *Sua filling*. This procedure makes use of the imbalance between supply and utilization values for each commodity and fills the figures that have been recognized as missing from the *Utilization table*. The filling procedure assigns figures using an ad-hoc algorithm, called **Inverse Ranking**, designed for dimensioning the figures in accordance with the importance they had in the past FBSs, as suggested also from the *Utilization table* through ranks associated to each variable. As a consequence, all the commodities entering the sua filling procedure are eventually balanced.

Not all the commodities enter the process the same way. In particular, a distinction is made between primary commodties and derived commodities:

- Primary commodities are normally the commodities for which more reliable figures are produced. Moreover, FAO often publish information about primary commodities, therefore values are not expected to be modifyed during a computational process.
- Derived commodities are more affected by modifications during this process, according to the flag of each figure.

Therefore, the steps of *Sua filling* process affect the SUA differently from line to line. Figure 4 shows the general Workflow of the sua filling highlighting the distinction between primary and derived commodities.
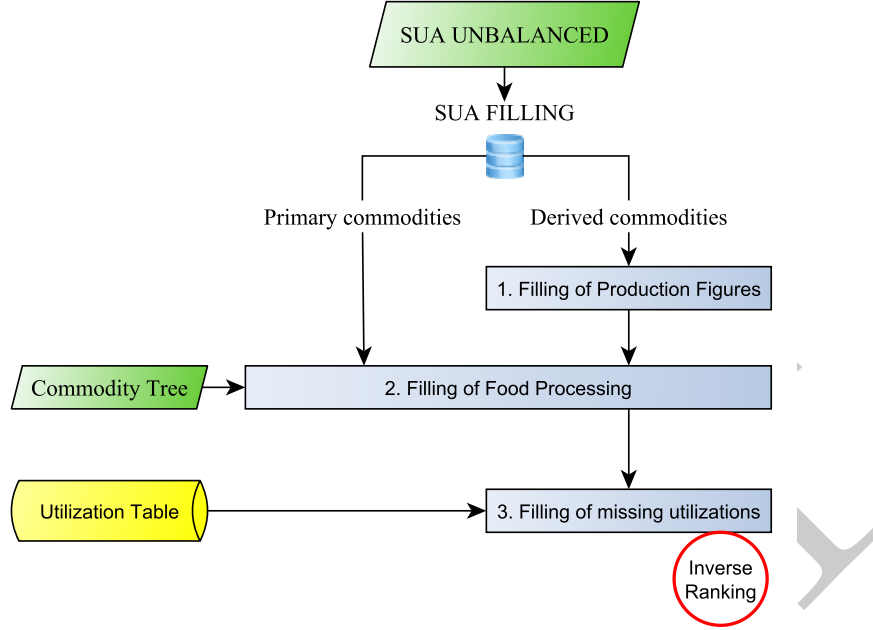
Figure 4: The Sua filling divesified workflow

Step 1: Filling of production figures

Only Derived commodities are involved in this process. The sua filling detect the Production figures that need to be filled or modified, by looking at the following *Imbalance*:

$$Imb_1 = P_{ijt} + I_{ijt} - X_{ijt} \tag{3}$$

where Imb_1 is the Imbalance and is enumerated because is different from a second Imbalance that will enter into the process later. Here:

- If $Imb_1 < 0$ there is no supply enough for the export and this is intepreted as the need for creating/incrementing the Production figure. The new production figure (P_{ijt}^*) is computed as:

$$P_{ijt}^* = P_{ijt} + Imb_1 \tag{4}$$

- If $Imb_1 >= 0$ there is enough supply and, therefore, no need for changing the production figure.

The reason for excluding other Variables from the computation of the imbalance here is that, at this step of the process, Trade variables are the most reliable and those for which, to the highest level of probability, there are all the figures filled. In Table 2 *Imb1* of our example is reported. Notice that there are 3 rows with negative values, but one of them is considered Primary ommodity. Indeed, *Mixes and dough for the preparation of baker's wares* is a commodity never processed and, therefore, is considered primary commodity and added to wheat in the FBS item *Wheat and produts*. Consequently production of only 2 commodity is changes, as shown in Table 3.

Table 2: Sua table with Imb1 (Example China Mainland, Wheat 2014)

itemName	P	I	X	DSt	Fo	FP	Fe	Se	T	IU	L	ROU	Imb1
Wheat	126,208,400	2,971,249	957	1,120,565		-	29,181,617	4,277,567		2,985,279	2,713,000	-	*129,178,692*
Wheat and meslin flo	70,500,000	33,055	188,674		67,300,000	-			-17,345			-	*70,344,381*
Mixes and doughs for		6,497	38,072		0	-	0		0			-	*-31,575*
Other Fructose and S	126,277	3,659	162,324		0	-			0			-	*-32,388*
Starch of Wheat	239,816	11,035	40,311			-	172,196			7,919		-	*210,541*
Wheat Gluten	25,580	877	117,373			-	0					-	*-90,916*
Communion wafers	13,263	8,796	5,822		16,241	-			-4			-	*16,237*
Uncooked pasta	1,415,692	12,520	22,550		1,405,661	-			-362			-	*1,405,661*
Food Preparations of		69,686	21,977		47,709	-			-12			-	*47,709*
Bran of Wheat	21,414,279	156,359	2,200		16,500,000	-	4,827,244		-4,252			-	*21,568,438*
Gluten Feed and Meal	793,740	160,231	529,333			-						-	*424,638*
bread	15,485	2,897	4,210		14,175	-			-3			-	*14,172*
pastry	193,950	89,593	117,630		165,914	-			-43			-	*165,914*

[*] 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

Table 3: Sua table with Production filled/incremented (Example China Mainland, Wheat 2014)

itemName	P	I	X	DSt	Fo	FP	Fe	Se	T	IU	L	ROU	Imb1
Wheat	126,208,400	2,971,249	957	1,120,565		-	29,181,617	4,277,567		2,985,279	2,713,000	-	129,178,692
Wheat and meslin flo	70,500,000	33,055	188,674		67,300,000	-			-17,345			-	70,344,381
Mixes and doughs for		6,497	38,072		0	-	0		0			-	-31,575
Other Fructose and S	***158,665***	*3,659*	*162,324*		*0*	*-*			*0*			*-*	*0*
Starch of Wheat	239,816	11,035	40,311			-	172,196			7,919		-	210,541
Wheat Gluten	***116,496***	*877*	*117,373*			*-*	*0*					*-*	*0*
Communion wafers	13,263	8,796	5,822		16,241	-			-4			-	16,237
Uncooked pasta	1,415,692	12,520	22,550		1,405,661	-			-362			-	1,405,661
Food Preparations of		69,686	21,977		47,709	-			-12			-	47,709
Bran of Wheat	21,414,279	156,359	2,200		16,500,000	-	4,827,244		-4,252			-	21,568,438
Gluten Feed and Meal	793,740	160,231	529,333									-	424,638
bread	15,485	2,897	4,210		14,175	-			-3			-	14,172
pastry	193,950	89,593	117,630		165,914	-			-43			-	165,914

[*] 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

Step 2: Filling of food processing

When Production has been checked, created or incremented, *Food Processing* can be calculated. Food processing is the amount of the supply of a commodity processed into derived commodities. *Food processing* is calculated for all the parent commodities as sum of the food processing of all the possible derived commodities. For example, Wheat in China is processed into Flour of Wheat, Bran of Wheatand Breakfast Cereals: Flour and Bran are produced during the same production process (they are called *co-products* or *by-products*), therefore the same amount of wheat is used for producing all of them, Breakfast Cereals comes from a separate production process instead. Therefore, *Food processing* of Wheat is given as sum of the amount of wheat's supply used for priducing Flour and Bran + The amount used for producing Breakfast Cereals. Calculation of Food processing happens by level, then after having calculated the amount for wheat, the amount of supply of flour of wheat used for producing Starch, Gluten, bread and the other derived of wheat flour becomes the *Food processing* of flour of Wheat and so on, for any subsequent level.

The formula for calculating FP is based on the production of the child product:

$$FP_{p_{ijt}} = \sum_{c=1}^{C} \left(\frac{P_{c_{ijt}}}{eR_{p \to c}} \right) \times s_{cp} \times w_c \tag{5}$$

where:

- $FP_{p_{ijt}}$ is *Food processing* of the generic parent p.
- $c = 1, 2, 3...C$ are the C children c of parent p.
- $P_{c_{ijt}}$ is the Production of Child c.
- $eR_{p \to c}$ is the *extraction Rate* from parent p to child c. The Extraction rate represents how much amount of the child commodity is produced from 1 unit of parent commodity[12]. It is expressed as a ratio of the processed product obtained from the processing of the parent/originating product.
- s_{cp} is the *share* of child c from parent p. Shares represent the amount of the child commodity that was produced from the specified parentand are expressed as a ratio Indeed, many products can be obtained from more than one parent (i.e. the commodity fruiti juice) and this requires to define the ratio coming from each parent. Shares are defined as:

$$s_{cp} = \frac{availability1_{p(c)}}{\sum\limits_{p=1}^{A} availability1_{p(c)}} \tag{6}$$

  where $availability1_{p(c)}$ is the availability of each parent p of child c expressed in terms of c (as say in *child equivalent*). Is is enumerated as 1 because a slighlty different availability will be found in a following step:

$$availability1_{p(c)} = (P_{p_{ijt}} + I_{p_{ijt}}) \times eR_{p \to c} \tag{7}$$

- w_c is the *weight* of child c. This parameter is used for identifying and treating co-product. Indeed, if the production of two or more co-prodiucts is used for producing *Food processing* of the parent commodity producing them, the result would be a Food processing twice the size of what it should be. This would happen because the two co-products are produced during the same process, in other words, fromt he same amount of the parent commodity. For this to be taken into account, a weight is used, which is 1 for the commodity that have to be considered for determining Food processing of parent commodity and 0 for the child commodity that are co-products (equation **??**). Please notice that, the co-products are still important in the standardization, because they contribute in the creation of the calories availability of the country, therefore, these commodities (Also called *zero weight* commodities) are multiplied by 0 only when quantities are treated, they will instead be considered when calories will be taken into account:

---

[12]FBSs are normally expressed in tonnes.

$$\begin{cases} w_c = 1 & \text{if boht quantity and calories have to be standardized} \\ w_c = 0 & \text{if only calories have to be standardized} \end{cases} \tag{8}$$

Table 4 contains the parent/child relationships, extraction rates and shares for *Food processing* computation of che wheat/China example.

Table 4: Parent/child eR and shares

ParentName	ChildName	eR	share	weight
Wheat	Wheat and meslin flo	0.78	1.00	1
Wheat	Bran of Wheat	0.22	1.00	0
Wheat and meslin flo	Uncooked pasta, not	1.00	1.00	1
Wheat and meslin flo	bread	1.00	1.00	1
Wheat and meslin flo	pastry	1.00	1.00	1
Wheat and meslin flo	Starch of Wheat	0.75	1.00	1
Wheat and meslin flo	Wheat Gluten	0.08	1.00	0
Wheat and meslin flo	Communion wafers, em	1.00	1.00	1
Starch of Wheat	Other Fructose and S	1.00	1.00	1
Wheat Gluten	Gluten Feed and Meal	1.00	0.33	0
Bran of Maize	Gluten Feed and Meal	1.00	0.67	0

This table reproduces the relationships of Figure 2. Shares lower than 1 mean that the child commodity might be produced also from other parents and, therefore, the reported ratio represents the ratio of that commodity that is produced from the reported parent. For example, *Gluten Feed and Meal* is produced from *Wheat Gluten* and from *Bran of maize*. The sum of the shares associated to this child (0.33 and 0.67) is equal to 1.

Table 4 shows the SUA table after the two steps just described. Notice that, in the specific example, there was no need for the creation/incrementation of Production on any derived commodity because $Imb1$ is positive for all commodities.

14

Table 5: Sua table with Food Processing filled (Example China Mainland, Wheat 2014)

itemName	P	I	X	DSt	Fo	FP	Fe	Se	T	IU	L	ROU
Wheat	126,208,400	2,971,249	957	1,120,565		**90,384,615**	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	**1,958,146**			-17,345			-
Mixes and doughs for		6,497	38,072		0		0		0			-
Other Fructose and S	158,665	3,659	162,324		0				0			-
Starch of Wheat	239,816	11,035	40,311			**158,664**	172,196			7,919		-
Wheat Gluten	116,496	877	117,373			**0**	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			-

* 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

From Table 4 and Table 5 here is, as example, the computation of *Food processing* of Wheat,Flour of wheat and Starch of wheat:

$$FP_{wheat_{2014}} = (70,500,000/0.78) \times 1 \times 1 + (21,414,729/0.8) \times 1 \times 0 = 90,384,615 \qquad (9)$$

$$FP_{flourW_{2014}} = (1,415,692/1) \times 1 \times 1 + (15,486/1) \times 1 \times 1 + (193,950/1) \times 1 \times 1 + \\ (239,816/0.75) \times 1 \times 1 + (116,496/1) \times 1 \times 0 + (13,263/1) \times 1 \times 1 = 1,958,146 \qquad (10)$$

$$FP_{starchW_{2014}} = (158,665/1) \times 1 \times 1 + (21,414,729/0.8) \times 1 \times 0 = 158,665 \qquad (11)$$

Step 3: Filling of other missing utilizations

After *Food processing* figures have been computed, the other variables have to be checked and either filled, increased or decreased. Only derived commodities enter this step. Primary commodities are excluded because for these commodities, the figures are normally more reliable and complete and, very often protected. Only the final balancing may affect Primary commodity figures. This step is based on the value of the following Imbalance:

$$Imb_2 = P_{ijt} + I_{ijt} - X_{ijt} - \Delta St_{ijt} - Fo_{ijt} - Fe_{ijt} - Lo_{ijt} - Se_{ijt} - IU_{ijt} - T_{ijt} \qquad (12)$$

Table 6 reports $Imb2$ for all the lines of the SUAs of our example.
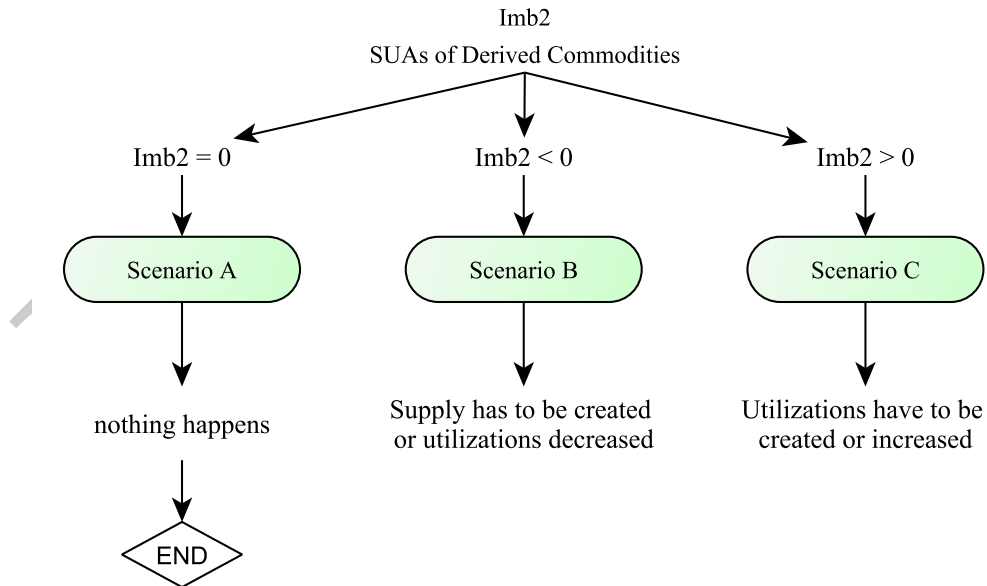
Three different scenarios con be found (Figure 5):

Figure 5: Scenarios of the sua filling

16

Table 6: Sua table with Imb2 (Example China Mainland, Wheat 2014)

itemName	P	I	X	DSt	Fo	FP	Fe	Se	T	IU	L	ROU	Imb2
Wheat	126,208,400	2,971,249	957	1,120,565		90,384,615	29,181,617	4,277,567		2,985,279	2,713,000	-	*-1,483,951*
Wheat and meslin flo	70,500,000	33,055	188,674		67,300,000	1,958,146			-17,345			-	*1,103,580*
Mixes and doughs for		6,497	38,072		0		0		0			-	*-31,575*
Other Fructose and S	158,665	3,659	162,324		0				0			-	*0*
Starch of Wheat	239,816	11,035	40,311			158,664	172,196			7,919		-	*-128,239*
Wheat Gluten	116,496	877	117,373			0	0					-	*0*
Communion wafers	13,263	8,796	5,822		16,241				-4			-	*0*
Uncooked pasta	1,415,692	12,520	22,550		1,405,661				-362			-	*363*
Food Preparations of		69,686	21,977		47,709				-12			-	*12*
Bran of Wheat	21,414,279	156,359	2,200		16,500,000		4,827,244		-4,252			-	*245,447*
Gluten Feed and Meal	793,740	160,231	529,333									-	*424,638*
bread	15,485	2,897	4,210		14,175				-3			-	*0*
pastry	193,950	89,593	117,630		165,914				-43			-	*42*

* 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

Please notice that, besides export and stock Variations, that are excluded from this step for the reasons previously explained, also Food Procesisng is excluded from the next steps, because FP is related to the production figures of the children and to modify it, would detach this very important parent-children link.

- *Imbalance = 0 (Scenario A)*

In this scenario the Imbalance is null, this meaning that there is enought supply for the existing utilizations. In This case the SUA line is balanced and nothig happens for the commodity. In Table 6 this is the case of *Other Fructose and syrup* and *wheat Gluten.*

- *Imbalance < 0 (Scenario B)*

If this happens, it means that there is not enough supply for the commodity to be used in the way the SUA figures suggest. First of all it has to be rembered that, as said in a previous section, trade figures are never changed during the filling process, this meaning that, if there is a need for interveening on supply, only production figures might be taken into account. In this case, as shown in Figure 6 a distinction as to be made as of the Production figure is protected or not.

- If Production figure is not protected, it is interpreted as if there was not enough information for producing a correct production number, therefore, this figure is changed, i.e. created, if missing, or incremented.
- If Production figure is protected, it should never be touched, therefore the algoritm tries to solve the eccess of utilization by reducing all the existing utilizations by 30% (except Exports[13]). If this is enough to cover the Utilization, the balancing line for that commodity is balanced and the process can go to the next step. After the reduction by 30%, the imbalance $Imb2$ is recalculated and it can give two different results:

  – If this imbalance is still negative a warning is given that tells that there is an unsolvable problem and that all figures have to be checked because either the production protected figure is wrong, or the Trade or some utilization have to be changed.

  – If the imbalance becomes positive, the line falls into the Scenario C that will be explained shortly.

Starch of Wheat belongs to this very last case. Indeed, it has a negative Imbalance therefore its utilizations (except those utilizations that are completely excluded from this process) are reduced by 30% of their values (see Table 7). After that happens, the Imb2 is recalculated and it is still negative, equal to $74,204$ (this value is not reported in the table). In the following step, as the production is not protected, the new value of production is equal to:

$$P_{StarchW} = 239,816 + 74,204 = 314,020 \tag{13}$$

Remember that the commodity *Mixes and dough for the preparation of baker's wares* is considered here as a primary commodity, therefore it will not balanced even if it has a negative imbalance.

---

[13]Stock is changed in this case, because this is just a proportional reduction of values and does not need a time series analysis
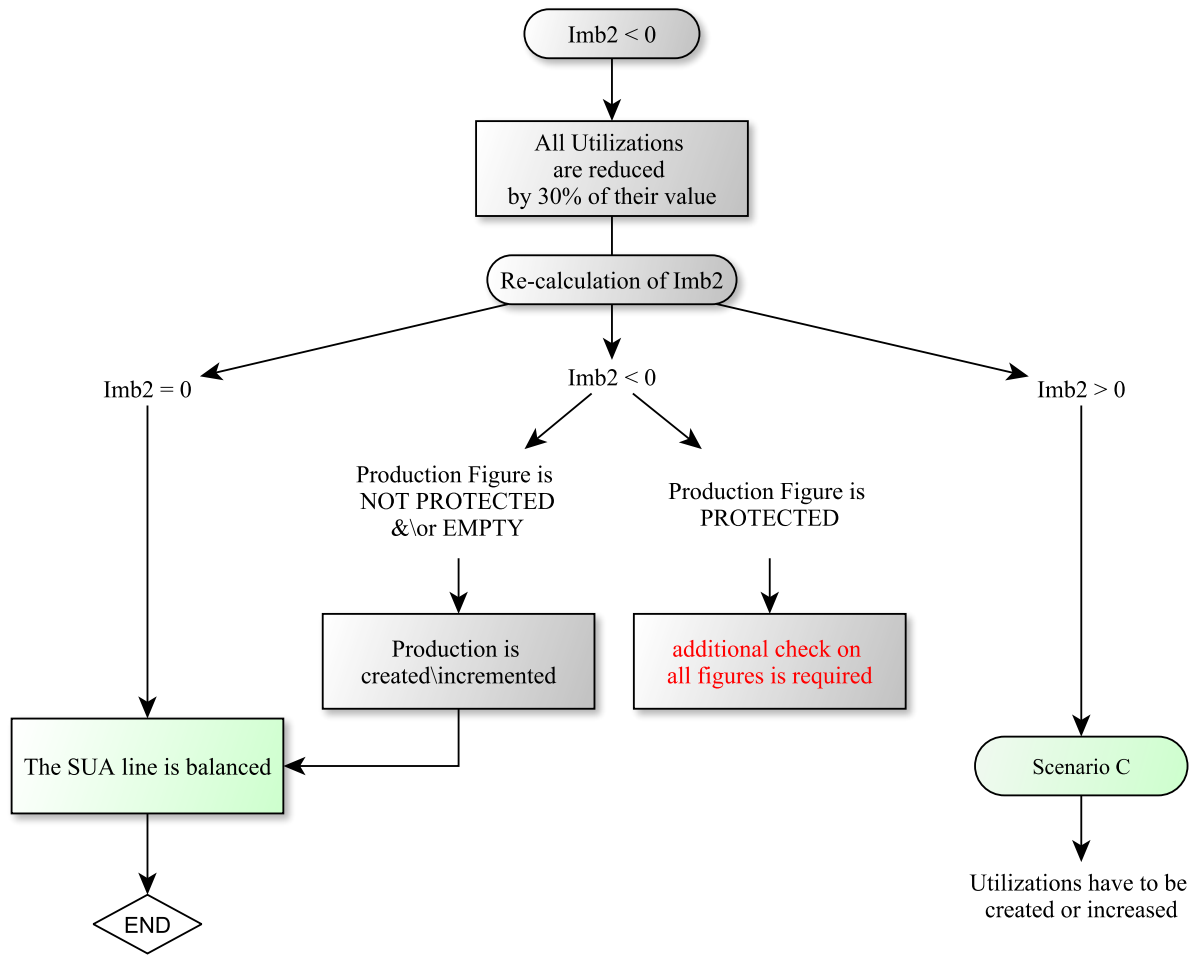
Figure 6: Workflow of Sua Filling when there is a negative imbalance

Table 7: Sua table with utilizations filled (Example China Mainland, Wheat 2014)

itemName	P	I	X	DSt	Fo	FP	Fe	Se	T	IU	L	ROU	X.1
Wheat	126,208,400	2,971,249	957	1,120,565		90,384,615	29,181,617	4,277,567		2,985,279	2,713,000	-	-1,483,951
Wheat and meslin flo	70,500,000	33,055	188,674		**68,372,646**	1,958,146			**-17,621**			-	0
Mixes and doughs for		6,497	38,072		0	0			0			-	-31,575
Other Fructose and S	158,665	3,659	162,324		0				0			-	0
Starch of Wheat	**314,020**	11,035	40,311			158,664	**120,537**		**5,543**			-	0
Wheat Gluten	116,496	877	117,373		0	0						-	0
Communion wafers	13,263	8,796	5,822		16,241				-4			-	0
Uncooked pasta	1,415,692	12,520	22,550		**1,406,024**				**-363**			-	0
Food Preparations of		69,686	21,977		47,709				-12			-	12
Bran of Wheat	21,414,279	156,359	2,200		**16,689,929**		**4,882,810**		**-4,301**			-	0
Gluten Feed and Meal	793,740	160,230	529,333				**424,637**					-	0
bread	15,485	2,897	4,210		14,175				-3			-	0
pastry	193,950	89,593	117,630		**165,957**				**-42.75**			-	0

* 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

- *Imbalance > 0 (Scenario C)*

When $Imb2 > 0$ there is an excess of supply that has to be distributed trought utilizations. Here:

- A *Utilization Table* tells which figures are supposed to be filled.
- The Imbalance is distributed among variables according to a, so called, *Multiple Filler approach* which make use of the *Inverse Ranking Rule.*

$->$ **The *Utilization table***

Utilization table is a Country/commodity table telling which Utilizations historically have been active for each commodity and assigning a Rank and an inverse rank to each utilization based on its mean value over the period 2000-2013. Table 8 shows the Utilization table for *Wheat and meslin flour* in China Mainland.

Table 8: Utilization Table Wheat and meslin Flour China Mainland 2014

AreaM49	Element	ItemCpc	rank	rankInv
1248	FP	23110	2	3
1248	Fo	23110	1	4
1248	DSt	23110	4	1
1248	X	23110	3	2

Where 1248 is the m49 code for China Mainland and 23110 is the cpc code of Wheat and Meslin Flour. The table tells the following information:

1. The reported commodity in China, as been historically used:
   - for producing other commodities (*Food Processing*),
   - for human consumption (*Food*),
   - for increasing stocks (*Stock Variation*),
   - for exports (*exports*).
2. The biggest amount of availability of the reported commodity was used for Human consumption, the second big for processing other commodity, then for exports and, in the end, for increasing stocks.

These informations are combined with the amount of *Imb2* for filling the SUAs of each derived commodity according to a set of rules all falling into the, so called, *Multiple Filler approach.*

$->$ **The *Multiple Filler Approach* and the *Inverse Ranking Rule***

This approach distribute the *Imb2* across utilizations, by the ules reported in Figure 7.
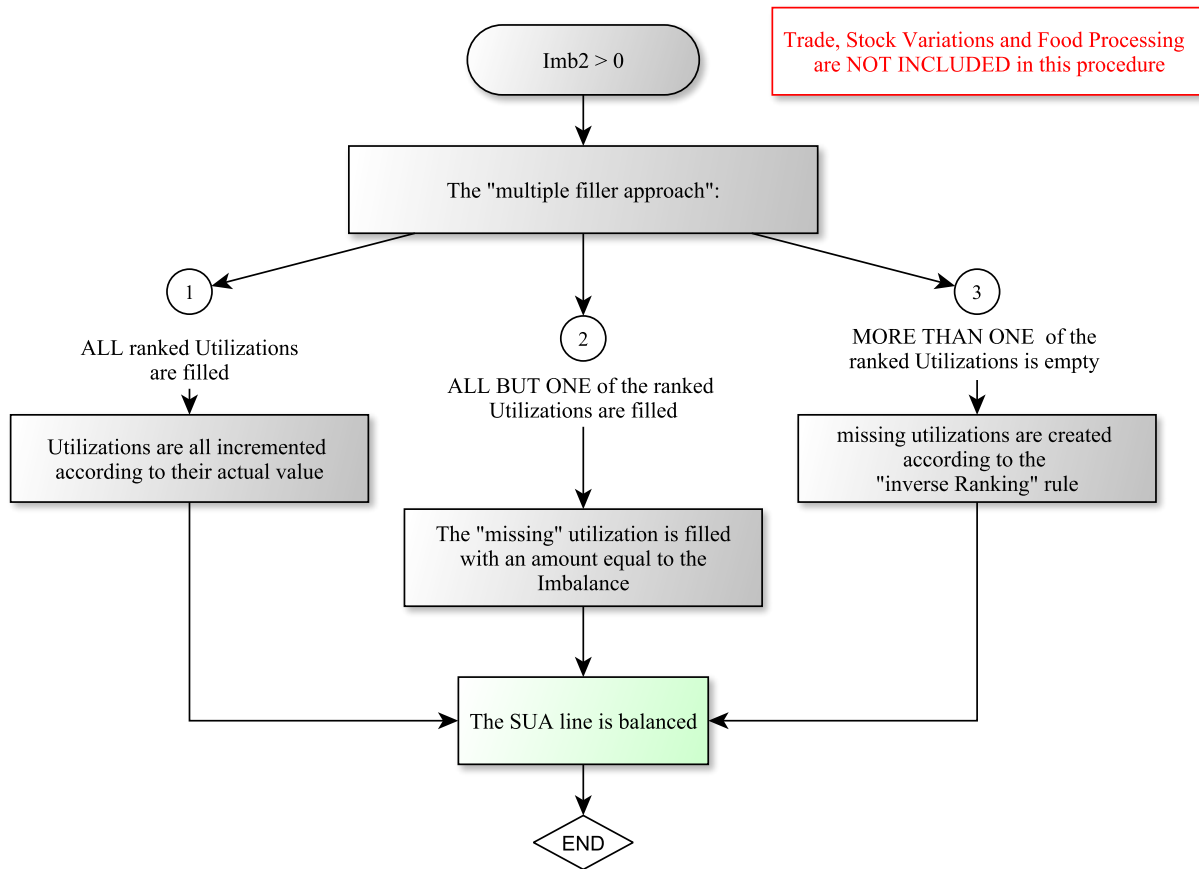
Figure 7: Workflow of Sua Filling when there is a positive imbalance

As reported in the Figure, Three cases may happen:

1. If all the ranked utilizations are already filled, all of them are increased proportionally to their values. *Bran of Wheat* in our example, belongs to this case. Utilization table for Bran of Wheat is reported in Table 9

Table 9: Utilization Table Bran ofWheat China Mainland 2014

AreaM49	Element	ItemCpc	rank	rankInv
1248	Fe	39120.01	2	2
1248	Fo	39120.01	1	3
1248	X	39120.01	3	1

According to this table, there are 3 Utilizations that are suppose to be active for this commodity: *feed*, *food* and *exports*. *Exports* is excluded from this step, therefore, food and feed reamain. Both Variables are filled in the SUA line of This Commodity, therefore, all of them are increased proportionally to their values.

2. If all the ranked utilizations are already filled but one, the all imbalance goes to that commodity. In our Example, this is the case of *Gluten Feed and meals* (Table 9).

3. Finally, if more that one utilization is empty, the so called ***Inverse Ranking rule*** is activated. This rule assigns part of the imbalance to the ranked utilizations proportionally to the ranks. This happens

by making use of the inverse rank of all the utilizations that are "activable" and not excluded from the procedure. The rule works for any number of ranks and independently from the number of utilizations that have to be excluded.

The *Inverse ranking rule* starts from the Utilization table with the addition of a column containing the *rank weigth* Wr, as reported in the following table:

element (Ut_i)	rank (r_i)	Inverse rank (Ir_i)	weight Rank (Wr_i)
Ut_1	r_1	Ir_1	Wr_1
...	...	...	...
Ut_i	r_i	Ir_i	Wr_i
...	...	...	...
Ut_n	r_n	Ir_n	Wr_n

where:

- Ut_i is the generic utilization, with $i = 1, 2, ...n$
- r_i is the rank of utilization Ut_i, with $i = 1, 2, ...n$
- Ir_i is the Inverse rank of utilization Ut_i, with $i = 1, 2, ...n$ and is equal to:

$$Ir_i = max(r) - r_i + 1 \tag{14}$$

- Wr_i is the rank weight of utilization Ut_i and is equal to 0 or 1 depending on the element being, respectively, exluded or not from the sua filling (Table 10).

Table 10: Rank Weight of the elements of FBS

Utilizations	Wr
X	0
DSt	0
Fo	1
FP	0
Fe	1
Se	1
T	1
IU	1
L	1
ROU	0

The value $Value_i$ assigned to the utilization of each commodity is calculated as a rank based proportion pR_i of the total $Imb2$ according to the following equation:

$$Value_i = Imb2 \times pR_i \tag{15}$$

with pR_i depending on $\sum\limits_{i=1}^{n} Wr_i$:

$$
\begin{cases}
pR_i = 0 & \text{when} \quad \sum\limits_{i=1}^{n} Wr_i = 0 \\
pR_i = \frac{Ir_i \times we_i}{\sum\limits_{i=1}^{n}(Ir_i \times Wr_i)} & \text{when} \quad \sum\limits_{i=1}^{n} Wr_i > 0
\end{cases}
\tag{16}
$$

As an example consider the SUA line of *Cocoa Beans* in Madagascar in 2014 (Table 11) and the Utilization table of the same country/commodity combination (Table 12)

23

Table 11: Sua line of Cocoa Beans Madagascar 2014 - before Sua Filling

itemName	P	I	X	DSt	Fo	FP	Fe	Se	T	IU	L	ROU	Imb2
Cocoa beans	10,865	36	8,326	-	-	-	-	-	-	-	743	-	*1,832*

Table 12: Utilization table for Cocoa Beans Madagascar

Ut	r	Ir	Wr
FP	4	2	0
Fo	2	4	1
IU	3	3	1
DSt	5	1	0
X	1	5	0

Besides the excluded elements, there are 2 utilizations empty: *food* and *Industrial Use*. The Values of these two elements is calculated as folows:

$$pR_{Fo} = \frac{4 \times 1}{0 + 0 + 4 + 3 + 0 + 0} = 4/7 = 0.57143$$
$$Value_{Fo} = 1,832 \times 0.57143 = 1,047$$

(17)

$$pR_{IU} = \frac{3 \times 1}{0 + 0 + 4 + 3 + 0 + 0} = 3/7 = 0.42857$$
$$Value_{IU} = 1,832 \times 0.42857 = 785$$

(18)

Table 13: Sua line of Cocoa Beans Madagascar 2014 - after Sua Filling

itemName	P	I	X	DSt	Fo	FP	Fe	Se	T	IU	L	ROU	Imb2
Cocoa beans	10,865	36	8326	-	*1,047*	-	-	-	-	*785*	743	-	*0*