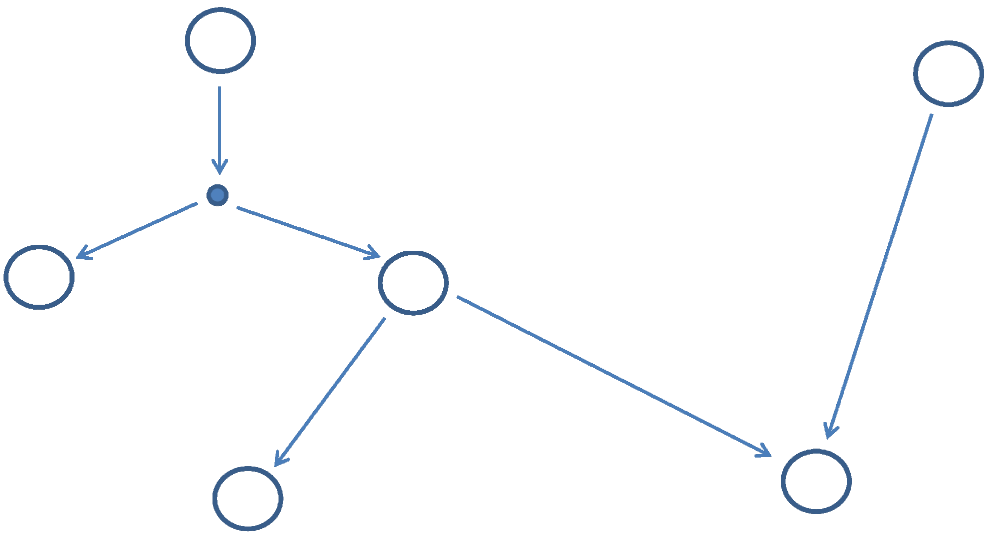
Flexible aggregation of FAO's Supply Utilization Accounts

Maarten van 't Riet

draft: June 29, 2011

The Food Balance Sheets are aggregated commodity accounts based on FAO's Supply Utilization Accounts. This paper describes principles of commodity aggregation and documents choices made for generating the Food Balance Sheets. Both the principes and the specific aggregation are illustrated with an example with real country data.



" *I see, ye can not see the wood for the trees*" - John Heywood, English Proverbs, 1564

June 2011

Flexible aggregation of FAO's Supply Utilization Accounts

Maarten van 't Riet[1]

This paper has been prepared under a consultancy contract with FAO,

May - June 2011.

[1] Maarten van 't Riet was a consultant to FAO in the mid-1990s and is now affiliated with CPB Netherlands Bureau for Economic Policy Analysis.

Flexible aggregation of FAO's Supply Utilization Accounts

Maarten van 't Riet

Introduction

The Food Balance Sheets must be one of FAO's most important products. They consist of a set of aggregate commodity accounts, such as wheat & products, showing production, trade and utilization. In addition they show food availability. The Food Balance Sheets are based on the Supply Utilization Accounts (SUAs), which are the detailed commodity accounts, i.e. wheat and flour of wheat. The SUAs are compiled by FAO's Statistics Division to estimate food availability, expressed in average daily per capita calorie supply. This statistic is central to describing developments in the world food situation.

The Food Balance Sheets (FBS) detail this statistic with information on commodity groups. Creating the FBS from the SUAs amounts to an *aggregation* process where input use of primary products is cancelled out against the production of derived products. Within the Organisation this process is referred to as '*standardisation*'.

Aggregation of data from a detailed product classification to a more comprehensive one is common practice for two reasons. The first reason is to reduce the amount of data, and the number of commodities involved, to a level and size more suited for analytical purposes. The second reason is, that by cancelling out the intermediate production of derived products against the input use of primary products, a clearer view emerges of the domestic availability of a product and its final uses.

Generating the FBS from the SUAs involves a huge and complex computational exercise, which requires execution by the computer. Still, the computer has to be instructed what to do, what computations to perform. These computations have to be specified, and the specifications involve assumptions and choices. These assumptions and choices need to be documented. This paper attempts to document the assumptions and choices made in generating the FBS from the SUAs.

Given that there is degree of freedom on how to aggregate detailed accounts, different aggregations may exist. Within FAO several 'standardisations' of the SUAs exist. Next to the FBS there is, for instance, the aggregation to the commodity classification of the Agriculture Towards 2000 studies. In addition, it may be necessary to revisit existing aggregations to make changes to meet the demands of a changing world. Therefore there is also a need to document the principles of aggregation so that the process is well understood and can be controlled.

The paper consists of three parts. First a general introduction to aggregation of commodity accounts is given. Basic principles and structures are presented with a focus on flexible specification. The second part concerns a specific aggregation: 'standardisation' of the SUAs for the FBS. The focus in this part is on documentation. Reference files contain all the assumptions and choices made in the aggregation. They are supplied in annexes. The third part illustrates both the specific aggregation and the principles. Also an example is given where an aggregate account has a serious imbalance. The cause of this imbalance is found by backtracking the aggregation procedure.

This consultancy paper complements an earlier paper titled 'The activity dimension of FAO's Supply Utilization Accounts'. This companion paper identifies repeated problems with 'standardisation' of the SUAs as inconsistent recording of the processing flows. What is required to ensure consistent accounting is explicit recognition of the dimension of agricultural processing activities. This paper adds to the claim of the importance of the activity dimension: specification of 'standardisation' starts with it.

Part I: Principles of flexible aggregation

1 Introduction to the principles

This part presents the principles of flexible aggregation of commodity accounts. The focus on flexiblility is necessary because within FAO there are already several different aggregations of the SUAs. And these existing aggregations may need to be revised every now and then to accommodate changes in which agricultural and food products are percieved to be of importance. Understanding of the principles should make interventions in the aggregation procedure possible.

This part of the paper is fairly abstract. Illustration follows in the next parts.

Section 2 maps out aggregation of commodity accounts in vector equations. Aggregation makes use of the processing, or production, relations underlying the accounts. Section 3 presents how aggregation works out in basic processing structures. And in section 4 matters are rounded up by addressing the processing structure in its entirety. The required reference files for an aggregation are identified.

The work in this part of the paper builds on earlier work of Keyzer and van ’t Riet (1993).

2 Aggregation in vector notation

This section presents aggregation of commodity accounts in vector notation, repeating material from the previous paper ‘The activity dimension of FAO’s Supply Utilization Accounts’. First the variables that contain the processing structure underlying the accounts are specified. Examples of this structure are given in section 3. Next, aggregation of commodity accounts, ‘standardisation’ in FAO parlance, is shown to be no more than a mapping from one classification to another. Subsequent sections in this paper concern a procedure to generate the elements of the transformation matrix of that mapping. Examples of simple aggregations will show that mappings that cancell out processing use against intermediate production are what we are looking for.

2.1 *Processing structure*

For a given country and given year the supply and utilization accounts may be written as:

(2.1) $$y_i^0 + y_i + m_i = x_i + v_i + c_i + f_i + o_i$$

for $i = 1, .. , \mathrm{m}$, where

y_i^0	:	primary production,
y_i	:	intermediate production,
m_i	:	imports,
x_i	:	exports,
v_i	:	input into food processing sectors,
c_i	:	availability for food use,
f_i	:	feed use,
o_i	:	other categories (stock mutations, waste, seed use, industrial use).

The set of commodities ($i = 1, .. , \mathrm{m}$) is meant to cover all food consumption by human beings for nutrition**.** The FAO Commodity List amounts to such a classification.

Agricultural and food products are related through processing activities ($k = 1,.., \mathrm{n}$).
A detailed list of all possible agricultural processing activities must be available.
It should be based on FAO’s commodity trees. See annex 1, introduced in part II.

The volumes (quantities) involved in these activities would ideally be represented within an input-output structure, like in input-output analysis, where the balances are cast in a matrix form:

(2.2) $$\sum_k A_{ik} q_k = v_i$$

(2.3) $$\sum_k B_{ik} q_k = y_i \qquad \text{for} \quad i = 1,..,\mathrm{m} \quad , \quad \text{where}$$

A_{ik} : input coefficient of commodity *i* for activity *k*,

B_{ik} : output coefficient of commodity *i* for activity *k*,

q_k : intensity of activity *k*.

Total utilization of commodity *i* as input into food processing sectors, v_i, is the sum of what is being used by the distinct processing activities. Separate variables for this breakdown of total input use are:

(2.4) $$A_{ik} q_k = v_{ik}$$: input use of commodity *i* by processing activity *k*,

(2.5) $$\sum_k v_{ik} = v_i$$

Variable q_k is the level of operation, or intensity, of activity *k*, and B_{ik} and A_{ik} are the output and input coefficients respectively. B_{ik} is the output of commodity *i* per unit of processing activity *k*, say wheat milling, and A_{ik} is the input of commodity *i* per unit of processing activity *k*.

While domestic production of derived commodities (wheat flour, butter etc.) appears through the y_i's, domestic supply of farm commodities (wheat, milk) enters as given primary production through y_i^0. Hence, either y_i or y_i^0 will usually be zero.

Again, separate variables can be used to represent the breakdown of intermediate production:

(2.6) $$B_{ik} q_k = y_{ik}$$: production of commodity *i* by processing activity *k*,

(2.7) $$\sum_k y_{ik} = y_i \qquad , \quad \text{for those } i \text{ for which} \quad y_i^0 = 0,$$

Examples of processing activities and their representation are given below.

As the present SUA-methodology does not recognize processing activities with joint inputs, each activity has a single input commodity. The unit intensity of a processing activity can be defined as the processing of one volume unit of the single input commodity. The variable q_k then expresses the number of volume units processed of the single input commodity of activity *k*. All non-zero input coefficients can then be taken to be 1. Examples follow with the basic structures in section 3.

2.2 *Vector notation*

In a formal way standardisation of commodity accounts can be defined as a transformation from one commodity classification to another. Let the original set of m accounts be given by:

(2.8) $$y^0 + y + m = x + v + c + f + o$$

The variables are the same as before but are now seen as vectors of dimension m.
Let the processing relations be defined by:

(2.9) $$Aq = v$$

(2.10) $$Bq = y$$

Given a list of n identified activities, A is an (m by n) input matrix, B n (m by n) output matrix , q is an n -vector of activity levels. The equations above are the same as the first three in this section.

The transformation to a new classification, of say r classes, can be defined as a (r by m) matrix T . Its elements T_{ij} denote the contribution of the *j*-th original account to the *i*-th new account. Standardisation then takes the form of pre-multiplication by matrix T of the set of accounts:

(2.11) $$T(y^0 + y + m = x + c + f + o)$$

The result of this operation is a new set of r aggregate accounts.

Since the number of SUA-commodities is large it is cumbersome to enter all the coefficients T_{ij}, especially considering that aggregations are country and year specific. Therefore a procedure is required that generates the coefficients T_{ij}.

2.3 *Simple examples*

When all detailed accounts are expressed in the same quantity unit, say Metric Tons, straightforward addition may be considered meaningful; element by element, the quantities of all accounts are added. Alternatively, cereals could be grouped together, and all fruits together, etcetera.
The mapping then would look like something as follows, the rows identifying the aggregates.

(2.12) $$T = \begin{pmatrix} 111 & 000000 & .. & 00 \\ 000 & 111111 & .. & 00 \\ ... & .... & .. & .. \\ 000 & 000000 & .. & 11 \end{pmatrix}$$

Often weighted addition makes more sense. This involves pre-multiplication with a single row-vector.

(2.13) $$T = p$$

Row-vector p could denote domestic commodity prices.

Aggregation then means conversion of quantities to values and adding up. That would give an account with the total value of primary agricultural production, imports, etcetera. This agricultural product account would include the value added of the processing activities, at least as far as the agricultural and food products are concerned: $py - pv$ or $pBq - pAq$.

Weighted addition can also be performed with the caloric contents of the commodities as weights p . The element 'food' of the single aggregate account would then represent the total amount of calories available for food consumption for the year and country under consideration. Other elements have a similar meaning. What remains in the account is a term, $py - pv$, denoting the loss of calories in all processing activities.

In other cases aggregation makes most sense when the underlying commodity flows are eliminated. Or, in other words, when input use is cancelled out against intermediate production:

(2.14) $$Ty - Ty = 0 \quad ; \quad TAq = TBq$$

This basically implies expressing quantities of commodities in equivalents of related commodities up, or down, the processing structure (or commodity trees).

3 Basic structures and backward calculation

The cancelling out of intermediate production against input use is illustrated in this section for a number of basic structures. Quantities of derived products are expressed in equivalents of the products they originate from. This is up the processing chain and is therefore referred to as backward aggregation or backward calculation. Aggregations are shown that satisfy condition 2.14.

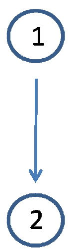
The basic structures shown are: i) single input – single output, ii) alternative outputs, iii) joint inputs, iv) joint outputs, v) alternative inputs, vi) second level processing and vii) cycles.

The first, most simple, case is worked out in some more detail. The last two structures provide a transition to dealing with the processing structure in its entirety.

3.1 *Single input – single output*

The most simple case of a production structure is the one with a single input and a single output commodity. This structure can be depicted in a figure where the circles represent commodities, and their accounts, and the arrow gives the direction of the input-output relation.

Figure 1: single input – single output



Processing 1 kilogram of the input commodity will give β kilograms of the derived commodity. This information can be covered by the matrices A and B. They concern two commodities and only one processing activity, hence the matrices are (2x1).

(3.1) $$A = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \qquad B = \begin{pmatrix} 0 \\ \beta \end{pmatrix}$$

The vectors of processing use and intermediate production are as follows.

(3.2) $$v = \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = Aq = \begin{pmatrix} 1 \\ 0 \end{pmatrix} q_1 = \begin{pmatrix} q_1 \\ 0 \end{pmatrix}$$

(3.3) $$y = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = Bq = \begin{pmatrix} 0 \\ \beta \end{pmatrix} q_1 = \begin{pmatrix} 0 \\ \beta q_1 \end{pmatrix}$$

The processing relation is, of course, summarized by $y_2 = \beta v_1$: production of the derived product equals the amount of the primary product inputted times the technical conversion factor.
The other accounting identities are the two commodity balances .

(3.4) $$y_2 = \beta v_1$$

(3.5) $$y_1^0 + m_1 = x_1 + v_1 + c_1 + f_1 + o_1$$

(3.6) $$y_2 + m_2 = x_2 + c_2 + f_2 + o_2$$

Backward aggregation means expressing quantities of the derived product in equivalents of its parent product. For this a conversion factor of $1/\beta$ is required. The matrix of the aggregation mapping takes the following form.

(3.7) $$T = \begin{pmatrix} 1 & 1/\beta \end{pmatrix} \ , \qquad \text{which implies} \qquad TB = TA = 1 \quad !$$

Observe that the aggregation coefficients are independent of the actual activity levels.

The new account of the aggregate commodity reads as follows.

(3.8) $$y_1^0 + 1/\beta\, y_2 + m_1 + 1/\beta\, m_2 = x_1 + 1/\beta\, x_2 + v_1 + c_1 + 1/\beta\, c_2 + f_1 + 1/\beta\, f_2 + o_1 + 1/\beta\, o_2$$

And the processing and derived production cancell out : $1/\beta\, y_2 = v_1$.

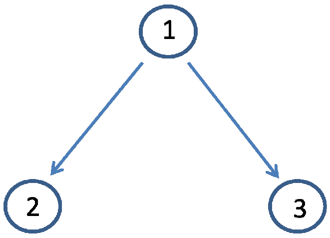
(3.9) $$y_1^0 + m_1 + 1/\beta\, m_2 = x_1 + 1/\beta\, x_2 + c_1 + 1/\beta\, c_2 + f_1 + 1/\beta\, f_2 + o_1 + 1/\beta\, o_2$$

This is the basic idea. Below other processing structures are discussed, though in less detail.

3.2 *Alternative outputs*

The next basic structure is where a single product can be processed in two, or more, separate processing activities to give alternative ouput commodities. They are found aplenty in the SUAs.

Figure 2: alternative outputs



There are now two activities and three commodities, requiring (3x2) input and output matrices.

(3.10) $$A = \begin{pmatrix} 1 & 1 \\ 0 & 0 \\ 0 & 0 \end{pmatrix} \;,\; B = \begin{pmatrix} 0 & 0 \\ \beta_2 & 0 \\ 0 & \beta_3 \end{pmatrix}$$

The aggregation mapping for backward calculation is as follows.

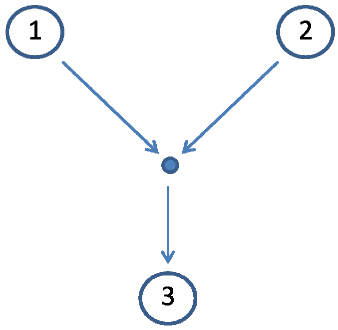
(3.11) $$T = \begin{pmatrix} 1 & 1/\beta_2 & 1/\beta_3 \end{pmatrix} \;, \qquad \text{giving} \qquad TB = TA = \begin{pmatrix} 1 & 1 \end{pmatrix}$$

Cancelling out input use against intermediate production now involves two activities. And, as this structure is just two times the single input – single output case, again the aggregation coefficients do not depend on the activity levels.

3.3 *Joint inputs*

Consider an activity that requires two simultaneous inputs in a process to derive another product. This is the case of joint inputs, which are not encountered in the SUAs. However, it is instructive to see how easy backward aggregation is interpreted in this case.

Figure 3: joint inputs



There is one activity and backward calculation means that the derived product will be expressed in equivalents of both joint inputs.

(3.12) $$A = \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ 0 \end{pmatrix}, \quad B = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$$

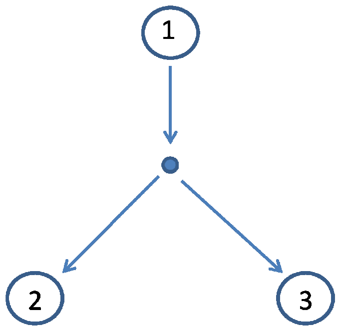
(3.13) $$T = \begin{pmatrix} 1 & 0 & \alpha_1 \\ 0 & 1 & \alpha_2 \end{pmatrix}, \quad \text{meaning that} \quad TB = TA = \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix}$$

The conversion factors are precisely the input coefficients. A single unit consumed of the derived product is equivalent to the simultaneous consumption α_1 units of primary product 1 and α_2 units of primary product 2. This interpretation is straightforward. This is in contrast with what is found with the next structure.

3.4 *Joint outputs: weights!*

The next structure is where a single activity gives simultaneous, or joint, outputs. These are the flour and bran produced by milling, or the oil and cake that come from crushing seeds. Also, slaughtering an animal will give the meat, offals and hides or skin at the same time.

Figure 4: joint outputs



A single activity, three commodities.

(3.14) $$A = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \qquad B = \begin{pmatrix} 0 \\ \beta_2 \\ \beta_3 \end{pmatrix}$$

With respect to backward calculation it must be observed that, while the conversion factors can be based on the output coefficients, weighting is required to avoid double counting. To see this, imagine β_2 units of derived product 2. Expressing this in equivalents of product 1 with factor $1/\beta_2$ will give 1 unit. But if there would in addition be β_3 units of product 3 that would give another unit of product 1, making a total of 2, whereas only one unit is needed to produce these amounts of the derived products.

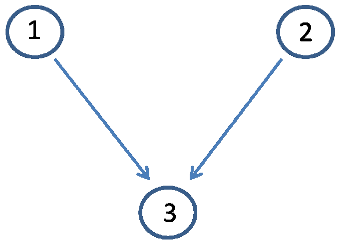
(3.15) $$T = \begin{pmatrix} 1 & \tfrac{1}{\beta_2} w_2 & \tfrac{1}{\beta_3} w_3 \end{pmatrix}, \quad \text{with } w_2 + w_3 = 1 \qquad \text{giving} \qquad TB = TA = 1$$

The weights to be used need to be specified and are a matter of choice.
Sometimes one of the outputs may be identified as the main product (say flour) and the other outputs as mere by-products (bran and germ). Then only the main output (flour) could be backward calculated. This amounts to giving the other outputs weight zero. This is the approach taken for the FBS.
Other aggregations have used caloric content to construct weights.
Again the conversion coefficients do not depend on the actual activity levels.

3.5 *Alternative inputs: shares!*

The next structure is where a derived product may originate from different alternative sources. In the SUAs bread may be made from flour of wheat or from rye. Beverages from distilled alcohol can have all sorts of cereals and fruits as alternative inputs.

Figure 5: alternative inputs



Two activities, three commodities.

(3.16) $$A = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix}, \qquad B = \begin{pmatrix} 0 & 0 \\ 0 & 0 \\ b_{31} & b_{32} \end{pmatrix}, \qquad q = \begin{pmatrix} q_1 \\ q_2 \end{pmatrix} = \begin{pmatrix} v_1 \\ v_2 \end{pmatrix}$$

The production of the derived products has two components:

(3.17) $$y_3 = y_{31} + y_{32} = b_{31}q_1 + b_{32}q_2 = b_{31}v_1 + b_{32}v_2$$

Backward calculation now involves shares. Only a specific share of the derived account can be added to each of the originating products. These shares are taken to be the part of the domestic production of the derived product. The shares s add up to 1 and thus double counting is again avoided.

(3.18) $$T = \begin{pmatrix} 1 & 0 & \frac{1}{b_{31}} s_1 \\ 0 & 1 & \frac{1}{b_{32}} s_2 \end{pmatrix}, \quad \text{where} \quad s_1 = \frac{y_{31}}{y_3} = \frac{y_{31}}{y_{31} + y_{32}} = \frac{b_{31}q_1}{b_{31}q_1 + b_{32}q_2} = \frac{b_{31}v_1}{b_{31}v_1 + b_{32}v_2}$$

$$\text{and} \quad s_2 = \frac{b_{32}v_2}{b_{31}v_1 + b_{32}v_2} \qquad \text{giving} \qquad TBq = TAq = q = \begin{pmatrix} v_1 \\ v_2 \end{pmatrix}$$

Here the conversion coefficients do depend on the actual processing levels. The pattern of domestic production encountered for the country in the year under consideration determines the aggregation.

Sometimes however there will be no domestic production of a derived product that may originate from different sources. This is when its supply comes from net imports. Default shares are then needed to specify along which activities the product is to be backward calculated and to what extent.

Ideally the imported derived products should be backward calculated respecting the way the product was produced in the exporting country. That would generate consistency at world scale. The availability and quality of trade data organised by country of origin and country of destination is however such that this precision is not foreseeable in the near future. However, the accounts of derived products could be split in a part of domestic origin and a part of foreign origin. The second split part could then be backward calculated according to default shares, derived from production at world scale. This could enhance the consistency and stability of the aggregations. This is a possible improvement over the current practice where the whole account is aggregated according to the pattern of domestic production.

3.6 *Second level processing*

A derived product can also be an input of another activity: this gives a second level of processing.

Figure 6: a derived product as input

Two activities, three commodities, but two inputs and two outputs.

(3.19) $$A = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix}, \qquad B = \begin{pmatrix} 0 & 0 \\ \alpha & 0 \\ 0 & \beta \end{pmatrix}$$

It is easily seen that the conversion matrix is as follows.

(3.20) $$T = \begin{pmatrix} 1 & 1/\alpha & 1/\alpha\beta \end{pmatrix} \qquad \text{and} \qquad TB = TA = \begin{pmatrix} 1 & 1/\alpha \end{pmatrix}$$

It can be seen to consist of two steps: first eliminating commodity 3 and then commodity 2.
Two aggregation matrices can be defined that together perform the task.

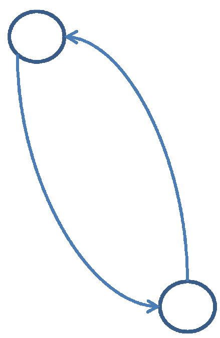
(3.21) $$T^1 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 1/\beta \end{pmatrix}, \qquad T^2 = \begin{pmatrix} 1 & 1/\alpha \end{pmatrix}, \qquad T = T^2 T^1$$

And thus a more complex structure with high levels of processing can also be backward calculated to the primary products.

3.7 *Cycles*

A final structure is the cycle. This is illustrated for two commodities below where the second level of processing has as output the commodity that was input in the first place.

Figure 7: a cycle



It should be clear that cycles are problematic for backward aggregation.
The stepwise reduction of the total structure suggested before assumes a finite number of steps.
Cycles must therefore be removed in the specification of an aggregation.

Cycles do exist in the SUAs (whey to dry whey and vice versa) because of missing joint inputs, such as water. In production theory a requirement of irreversibility of production is often imposed.

4 Specification

Aggregation of commodity accounts comes down to a transformation from one commodity classification to another. The new classification is assumed to be smaller so that it is more managable, but it will be based on information from the more detailed classification. Specification of an aggregation is the definition of the new classification and how it is related to the original one. The specification must contain sufficient information to generate the transformation mapping T.

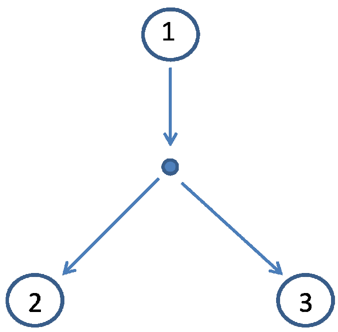
The main steps of defining the new classification are deciding on a subset of the original commodities and specifying further aggregation. The selected subset of commodities are the so-called ‘targets’. The other original commodities are expressed in equivalents of the target commodties, using the underlying processing structure. Further aggregation involves the grouping of target commodities.

Eliminating products from the detailed classification is governed by the technical factors of the processing activities, as already illustrated in section 3 for isolated basic structures. Before the structure in its enirety can be addressed two more elements need to be introduced.

4. 1 *Forward aggregation and ‘cuts’*

Recall the straightforward interpretation of the backward calculation of a product to its joint inputs as opposed to the more awkward aggregation of joint outputs, requiring weights, as described in subsections 3.3 and 3.4. There is a mirror image of this when forward aggregation is considered: quantities of a primary product may be expressed in its joint outputs simultaneously. This is a choice often made for oilcrops. An unit imported, or domestically harvested, of an oilcrop is easily expressed in both its oil and cake equivalents.

Figure 8: joint outputs again



A single activity, three commodities. And forward calculation.

(4.1) $$A = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \qquad B = \begin{pmatrix} 0 \\ \beta_2 \\ \beta_3 \end{pmatrix}, \text{ and}$$

(4.2) $$T = \begin{pmatrix} \beta_2 & 1 & 0 \\ \beta_3 & 0 & 1 \end{pmatrix} \qquad \text{such that} \qquad TB = TA = \begin{pmatrix} \beta_2 \\ \beta_3 \end{pmatrix}$$

Next to backward and forward calculation there is a third and final option of how to treat a processing activity in an aggregation: it can be ignored. This implies that the processing relation remains underlying the accounts of the aggregate commodity classification. Ignoring an activity in aggregation is referred to as a 'cut'. Cuts are applied in two circumstances. One is when a product fundamentally changes it characteristics when processed. An example is making distilled alcoholic beverages from cereals or fruits. The second circumstance is when the structure contains cycles.

4.2 *Specification by aggregation directives*

One way of specifying a complete aggregation is by assigning an aggregation directive to each of the possible processing activities of the detailed accounts. There are three options: 'b', 'f' and 'c', or backward, forward and cut. Most often the choice will be for 'b', expressing derived products in primary product equivalents. But the availability of the other options creates a whole spectrum of possible aggregations.

Observe that specification of 'standardisation' by activities is another compelling argument to take the activity dimension serious.

Assigning a directive to an activity, not a cut, yields a number of aggregation commands:
"convert commodity j to commodity i with factor F, weight W and share S."

When the directive is 'backward' then commodity j is the output of the activity and i the input, and when the activity has joint outputs there will be a command for each output.

Factor F is determinded by the io-coefficients of the activity, the actual ones used for the domestic processing in the country that year, or is determined by specified default io-coefficients.
Weights W must always be specified exogenously, and they can be zero.
Shares S follow from the actual pattern of domestic processing or are supplied as defaults.

The transformation coeffient, or conversion factor M is the product of the three: M = F * W * S.

Notice that this need not yet be a final transformation coefficient T_{ij}, because commodity i may not be a target, see subsection 3.6. Repeated application of the aggregation commands is then required.[2]

4.3 *Target commodities and further aggregation*

For consistency in the shares of the conversions a commodity cannot be backward and forward calculated at the same time. Eventually there are three options for the commodties of the original clasification: they are backward calculated, they are forward calculated or they are target commodities.

The second condition for a consistent aggregation is that the repeated application of the commands must be finite. Therefore cycles must be excluded. Cuts can be used for this.

After a specification by means of aggregation directives has been checked for consistency, all commodities have an aggregation label: 'b', 'f' and 't'. Those with the former two labels will be eliminated, their information added to the target commodities.

A specific category of target commodities are the by-products of an activity which have been assigned weight zero: they are not backward calculated and are retained at target level.

The aggregation can then proceed with the targets. Some of these can be grouped together according to some criterion. The result will be a further reduced set of commodities. This involves simple addition, see subsection 2.3, and is not further discussed. The files that govern this further aggregation are of course part of the specification of the aggregation. Their content is typically a matter for the data management authority of the organisation.

4.4 *Defaults and other control files*

The compostion of aggregate items in terms of targets are an example of control files of an aggregation. The are a number of reference files that must be available to execute an aggregation. Some concern classifications or dimensions, others concern quantitive information. Consider the next list of broad categories of reference files.

---

[2] This process can be described with matrices M which are (m x m), which can be multiplied in any order (!) to yield the final (r x m) matrix T by taking the zero rows out.

i. the initial commodity classification
ii. the classification of processing activities
iii. default io-coefficients of the processing structure
iv. the specification of the desired aggregation,
v. including additional quantitative information

The reference files i. and ii. are required to have a data set of commodity accouts to begin with. Default io-coefficients, i.e. generic elements of the matrices A and B, are probably a help in constructing the detailed accounts and are needed for the conversion factors when there is no domestic production of a derived product. These technical coefficients could be year and country specific.
A list of aggregation directives for each activity must be available, and checked for consistency.
From this follows a list of aggregation labels for the commodities, identifying target products.
Next, the weights for backward calculation of joint outputs must be specified, as must be the default shares for alternative inputs.
Finally there are the correspondence files for further aggregation, possibly involving weights.
Examples of actual files will be given in part II.

4.5 *Conluding remarks*

This completes the discussion of the principles of aggregation of commodity accounts. The focus has been on flexible specification. The backgroud of this focus is the fact that different 'standardisations' of FAO's SUAs do exist, serving different purposes. And these aggregations may also be revisited, and adapted to meet new demands of a changing world.

The core of the existing aggregations is backward calculation, but, in principle, expressing groundnuts in peanut butter equivalents is possible, if this would suit the purpose for which an aggregation is performed.

A specific aggregation of the SUAs, the one for the FBS, is illustrated in the next parts of the paper.

Finally, executing an aggreation requires the availability of explicitly recorded processing.
That this condition is not immediately met was described in the paper 'The activity dimension of FAO's Supply Utilization Accounts'. Gaining control over the 'standardisation' operation begins with the processing activities.

Part II: A specification: the FBS aggregation of the SUAs

5 Introduction to the specification

This part concerns a specific aggregation: aggregation of the SUAs for the Food Balance Sheets. Five categories of reference files have been identified to be addressed in a documentation of an aggregation. They are discussed in the next two sections. Specific reference files are attached as annexes. These reference files contain all the assumptions and choices made in the aggregation.

Apart from the reference files, which constitute the main part of the documentation, the consequences of the assumptions and choices made are shown in a summary table. This summary table, **annex 0**, lists for each SUA-commodity what role it plays in the aggregation process. Concerning the other end of the process, **annex 9** shows for each FBS- item its default composition. The last annex**, annex 10**, accounts for the origin of the other annexes.

This part is based on a consultancy note titled 'Aggregation of FAO's Supply Utilization Accounts for the Food Balance Sheets' (1997).[3] The 'standardisation' presently performed at FAO is still based on the reference files, and specification, presented in that note of 1997.[4] There has never been a formal validation of this specification by a data management authority of the Organisation.[5]

Specification of an aggregation involves choices which should be guided by the purpose for which the aggregation exercise is performed. The Statistics Division of FAO had two aggregations of the SUAs: the Demand and the Supply method. Both were stored in a data domain called the Commdity Balances. The philosophy of the Demand method (CBD) is to present the standardised information ordered in commodity groups that best reflect how the products are consumed. The selected commodities therefore include their derived products which fall in the same commodity group, but exclude those which, through processing, change their nature and become part of another commodity group, e.g. while groundnuts are shown under oilcrops, the groundnut oil is listed under vegetable oils. The information presented as such is therefore oriented for demand analysis.

[3] This note was handed to Sami Zarqa, then Chief Basic Data Unit, on April 14, 1997.
[4] Changes since then in the detailed commodity classification (FCL), or in the processing relations actually applied, will be a cause of potential error in 'standardisation'. I have seen minor changes in the further aggregation. These cannot contribute to imbalances.
[5] The previous 'standardisation'(AGROSTAT) was never documented: it was hidden in computer code.

By contrast, the other standardisation (CBS) is more oriented towards supply analysis. The Food Balance Sheets for the largest part correspond with the CBD. For a more detailed discussion of the FBS, CBD and CBS see a note titled ‘The Food Balance Sheets and the Commodity Balances - Demand and Supply method’, (1998).

The aggregation discussed in this, and the next, part concern the CBD.
This therefore also applies to the reference files, and the generated information, found in the annexes.
The annexes marked with an asterix are reference files.

Annex 0	:	Summary table (SUA - items to 'targets')	
Annex 1	:	Input - Output list	*
Annex 2	:	Output - Input list	*
Annex 3	:	default extraction rates	*
Annex 4	:	'autocuts'	*
Annex 5	:	zero weights	*
Annex 6	:	Default conversion factors	
Annex 7	:	FBS selection	*
Annex 8	:	'target' products to FBS items	
Annex 9	:	default composition of FBS items	
Annex 10	:	Accounting for the origin of the annexes	

6 From SUA items to the 'targets'

Below the first step in the aggregation of the SUAs for the FBS will be documented. The five categories identified in subsection 4.4 will each be addressed, with an emphasis on the specification of the aggregation and resulting in a set of default conversion factors.

6.1 *The initial commodity classification*

The initial data set are the SUAs and the initial product classification are the SUA-items found in the FAO Commodity List . This classification is documented in "Definition and classification of commodities" (draft), FAO, Rome, 1993.

There are 806 potential SUA items, from Wheat (15) 1599 Fish Total Value (1599).
All SUA items are listed in a summary table that presents information on the aggregation.
This table, **annex 0**, will be introduced after the specification of the aggregation.

6.2 *The classification of processing activities*

The specification of the production structure, i.e. the complete set of all identified processing relations, is done by means of the Input-Output list.

6.2.1 The format of the Input-Output list

The IO-list is built up by two types of lines: I-lines, for input-commodities, and O-lines, for output commodities. Jointly derived products appear on the same O-line. An O-line, together with the preceding I-line define a processing activity. Therefore the O-lines contain a code for the activity. These codes are constructed as follows: the first 4 digits are the ICS commodity code for the input-commodity, the next 4 digits are the commodity code of one of the output-commodities, the 9th digit is usually 0, but is also used to distinguish between activities for which the first 8 digits are identical. The symbol which appears in the first column of the 0-lines is a control character for aggregation, which is discussed later. The IO-list is attached as **annex 1**.

6.2.2 The Output-Input list

There are quite a number of products that may be produced by more than one processing activity. The IO-list can be inverted to give an Output-Input list. Each derived product (output) is followed by the activities from which it may originate. Since, presently, each processing activity has a single input-commodity, these are listed as well.

The symbol * is used to indicate an activity with joint outputs.

The complete OI-list is attached as **annex 2**.

The values on each I-line, all 0 or 1, have significance for aggregation and will be discussed later.

6.2.3 Some numbers

There are 903 distinct activities identified in the production structure.

These activities involve 586 products, either as input to or output from a processing activity.

There are 418 derived products, of which 147 can originate from more than one activity.

6.2.4 Cycles

The production structure knows two cycles, A to B to A, which have to be taken care of in aggregation. These cycles are Whey (903), Fresh to Dry Whey (900) and vice versa, and the same thing for Oils Animal (1168) to Tallow (1225).

6.3 *Default io-coefficients of the processing structure*

All activities have the characteristic of having only one input. Therefore the unit-level of operation of the activities can be chosen to be the processing of one unit of the single input commodity. The rates of transformation then can be called extraction rates.

Derived products that originate from more than one activity should have for each activity a distinct extraction rate. However, the previous computer system for the SUAs, ICS, as well as the present FAOSTAT Working System, do not support activity dependent extraction rates. Therefore, the data will only contain a single rate for each derived product.

The default values of the extraction rates used in the aggregation are also a single value for each derived product, irrespective of the activity. These default values are listed in **annex 3**.

For all derived products not listed there, its extraction rate will have as default value 100%.

The default io-coeffients could in principle be year and country specific. Certainly when the timeseries to be aggregated cover more than 40 years, some change in the technical conversion factors is to be expected.

6.4 *The specification of the desired aggregation, including additional information*

6.4.1 Aggregation directives

The specification of an aggregation begins with assigning an aggregation directive to each of the processing activities. The control character in the first column of the O-lines of the IO-list is used for that purpose. These characters can be 'b' for backward, 'f' for forward and 'c' for cut.
See the io-list in **annex 1**.[6]

In addition to assigning aggregation directives in the IO-list, derived products can be forced to become a target of the aggregation. When a derived product is present in a control file called 'autocuts', all the activities that produce that product will be assigned the directive 'c'. This overwrites what has been specified in the IO-list. The control file 'autocuts' is attached as **annex 4**.

The exercise of aggregation of the SUAs for the FBS has a focus on nutrition. That focus is reflected in the choice to keep alcoholic beverages separate and not have them backward calculated to the fruits, cereals and other products from which they are made.

Most activities, 712 of 903, have backward as directive. There are 187 cuts and 4 'f'-s.

The four cases of forward calculation are special cases.
These are the dummy commodities Cane Sugar (158) and Beet Sugar (159) to Sugar, Raw (162).
And Kapok Seed In Shell (311) to Kapokseed Shelled (312), Seed Cotton (328) to Cottonseed (329).

Apart from the cuts for the alcoholic beverages, there are cuts between the oilcrops and their oils, and between animals and meat.

Notice that also the activities '90009030' and '116812250' have been assigned the aggregation directive 'cut'. This has been done to avoid cycles in the aggregation directives.

6.4.2 Aggregation labels for the commodities

The specified aggregation is consistent: each commodity is either forward calculated, backward calculated or is a target. Of the 586 commodities involved in the production structure there are the four special cases of forward, and there are 289 commodities that go backward. That leaves 293 targets.

[6] The IO-list is therefore a double reference file: for the activity dimension and for the specification.

In the AGROSTAT 'standardisation' weighted backward calculation of joint outputs did not take place. All by-products, such as the brans of cereals, were being kept separate.

As discussed in subsection 3.4, the relative importance of joint output can be specified by giving weights. When the assigned weight is zero that product is not backward calculated, and becomes a target commodity of the aggregation. The list of by-products with weight zero is given in **annex 5**.

A number of by-products, such as the cake of oilcrops, had already been designated as targets. The list in annex 5 adds (60 - 25 =) 35 new target commodities.

6.4.3 Summary of target commodities

Starting with 806 SUA-items, and observing that there are 586 products involved in the production structure, the other (806 - 586 =) 220 items must be targets.
However, 43 items are not included in the Grand Total, aggregate code 1701. They are not involved in the structure and are excluded from aggregation. That leaves (220 - 43 =) 177 automatic targets. Specification of the aggregation adds (293 + 35 =) 328 targets, which gives a total of (177 + 328 =) 505 target commodities.

The summary table in **annex 0** presents a list of all SUA-items, indicating whether they are target, backward or forward calculated or excluded.

The table gives the SUA-items in ascending order. After the item-code and name follows an 0/1 indicator for inclusion in the Grand Total (1701). Then there are two columns that describe the role of the commodities in the production structure. A product can be an input-commodity, an output-commodity, both or none of the two. Then follows a column for the aggregation labels, which are 'T' for target, 'B' for backward and 'F' for forward. The next column indicates those products with weight zero. The other information in the table is explained later.

6.4.4 Default shares

The production structure specifies that Bread (20) may be produced from Flour of Wheat (16) or from Rye (71). When in an aggregation exercise an account for bread is found which only contains imports and consumption, no domestic pattern of production is available to determine how the quantities of bread should be backward calculated. Therefore default shares are required, see also subsection 3.5.

The default shares are specified in the Output-Input list, **annex 2**.[7]

In principle any shares could be specified, like 80% to Flour of Wheat and 20% to Rye.
However, single parent products have been selected as default destinations.
This is reflected in the values found on the I-lines which are either 1 or 0.[8]

The default shares complete the first step of the specification of the aggregation. When the dataset to be aggregated solely consists of trade data, no country specific information is available and default conversion factors must be generated. These conversion factors are found in **annex 6**.

---

[7] The OI-list is generated from the IO-list and is thus not an independent referenc file. It is edited to contain the default shares. These are then stored in a separate file, see Van't Riet (1997b).

[8] The default shares are typically a reference file that, with some frequency, must be examined for its up-to-dateness.

7 From the ‘targets’ to the FBS selection

Below the second step in the aggregation of the SUAs for the FBS will be documented. This step is a further aggregation of the target products as mentioned in subsection 4.3.

7.1 *The FBS selection*

In the world of AGROSTAT, ‘standardisation’ of the SUAs resulted in the data domains CBD and CBS, respectively Commodity Balances Demand method and Supply method. From these domains the items for the Food Balance Sheets publication were selected.
Annex 7 gives the FBS selection, with the CBD/CBS item codes and the breakdown of the included aggregates.

7.2 *From ‘target’ products to CBD/CBS items*

The correspondence between the ‘target’ products and CBD/CBS items is in most cases a matter of recoding. In other cases it is simple addition of products. The correspondence file is given in **annex 8**.
The destination of the target products had already been included in the summary table.
In addition, annex 8 specifies the weights in the conversion. In most cases this weight is 1.
Paddy Rice is multiplied by 0.667 to be expressed in milled equivalent.
Special cases are those were the weight is zero. This means that only the nutrients (calories, protein and fat) are being added, not the quantities of the commodity balance.

7.3 *All products to CBD/CBS items*

Finally, combining the default conversion factors with the simple aggregation of ‘targets’ to the CBD/CBS items the default composition of the CBD/CBS items can be generated. This combination of annexes 6 and 8 is given in **annex 9**.

7.4 *Concluding remarks*

This concludes the specification of the aggregation. The references files are part of the documentation as are the generated files with the conversion factors, and compositions for default situations.
Aggregation of the accounts for an actual country will have the default factors overwritten with country specific information. This is illustrated in part III.

The reference files are the controls for the operation. They start with classifications of the initial data set and continue to determine the mapping to the desired, more managable, dataset. Responsibility for their content should lie with the data management authority of the organisation.

Part III: An example of an aggregation for a country

8 Introduction to the example

From principles to specification we now come to the execution of an aggregation. Real SUA data are used. Brazil is the example country. There are no special reasons why Brazil was selected other then both being not too complex and not too simple with regard to the underlying processing. Also, no special problems were known. No further reference to the country will be made.

Execution of an aggregation for an example country with the explanations given in this document constitutes a ‘worked example’. It proves that the methodology does perform what is is supposed to perform. Possible ‘aggregation errors’, i.e. imbalances at the aggregate level, must be explainable. First the operation and results are presented following the order of the aggrgeation procedure. Next an error is found out by backtracking the procedure.

This part is about inspecting data files. They are not included in this document: there are just too many files and they are too large. The files are provided in a separate zip-file.

9 Following the procedure

This section follows the sequence of steps in the aggregation from the initial data to the final set of aggregated data. There is a script called **mvtrCBD** that was executed after extraction of the data. Here the results are discussed. Together with the complete set of generated files this constitutes the desired 'worked example'.

9.1 *The initial data set …*

The initial data set are the SUAs for Brazil for the year 1961 up until the year 2008. The data were extracted from the database on May 3, 2011, from a set referred to as the parallel system.[9]

A simple printing program generates a file called **sua.prt** . From that file, see the accounts for the commodity Wheat (15) for the last 4 of the 48 years. The accounts for these years are balanced.

```
 ** Brazil         **
 ** WHEAT          **      2005          2006          2007          2008
 **   15        21 **

  PRODUCTION         4658790.      2484848.      4114060.      6027131.
  IMPORTS            4988139.      6530502.      6638019.      6032698.
  FROM STCKS         1000000.F     2300000.F     -500000.F    -1050000.F
  EXPORTS             156571.       652102.       104477.       645140.
  FEED                200000.*      300000.*      100000.*      200000.*
  SEED                150579.C      157680.C      202762.C      207564.C
  WASTE               319408.C      339460.C      322562.C      361795.C
  PROCESSED          9820371.B     9866107.B     9522277.B     9595330.B
```

The printing program generates a file, **rwcflex.out**, with warnings on imbalances.

See the next lines from this file.[10]

```
WARNING : IMBALANCE FOR CTR.           21  FOR COM.          109   INFANT FOOD   :
  61:      -9167.      -11452.      -8113.      -7260.
```

Thus the account for Infant Food (109) is not balanced.

This can be checked in the file **sua.prt** :

```
 ** Brazil         **
 ** INFANT FOOD    **      2005          2006          2007          2008
 **  109        21 **

  IMPORTS                618.          816.         1217.         1591.
  EXPORTS               9785.        12268.         9330.         8851.
  * IMBAL.             -9167.       -11452.        -8113.        -7260.
```

[9] A single minor change was made to the original data on May 30.

[10] The '61' at the beginning of the second line denoting 1961 is incorrect.

There are net exports of Infant Food (109) from Brazil for these years. Clearly it is difficult, if not impossible, to know and record what this product is made from. The imbalances however, will re-appear at the aggregate level.

9.2 *… with the processing flows*

Underlying the set of detailed accounts are the processing flows. Unfortunately these are not explicitly recorded in the current working system for the SUAs. This fact is discussed in the companion paper 'The activity dimension of FAO's SUAs'. Also the need to recover the flows, i.e. make them explicit, is stated there. When this is done a new print file, **ioflows.prt**, can be generated to give the accounts with the flows. See the following (balanced) accounts for 5 commodities. This is an example of alternative outputs, see subsection 3.2. The breakdown of processing shown for commodity 16 is not readily available in the present FAOStat Working System.

```
** Brazil       **
** FLOUR WHEAT  **
**   16      21 **                 2005         2006         2007         2008
   INPUT                       9820371.     9866107.     9522277.     9595330.
  +WHEAT       [  1500162]     9820371.     9866107.     9522277.     9595330.
    OUTPUT COEF                   7200.        7200.        7200.        7200.
   EXTR.RATE                      7200.        7200.        7200.        7200.
   PRODUCTION                  7070667.     7103597.     6856040.     6908638.
   IMPORTS                       43523.      147957.      640423.      695253.
   EXPORTS                        1335.        1375.        2108.        1919.
   WASTE                        142284.      145031.      149929.      152078.
   PROCESSED                     41140.       44900.       40500.       50000.
  +MACARONI    [  1600180]          0.           0.           0.        5000.
  +BREAD       [  1600200]        490.         400.         500.         500.
  +PASTRY      [  1600220]      29150.       32000.       28000.       29000.
  +WAFERS      [  1601100]      11500.       12500.       12000.       15500.
   FOOD                        6929431.     7060248.     7303925.     7399894.

** Brazil       **
** MACARONI     **
**   18      21 **                 2005         2006         2007         2008
   INPUT                            0.           0.           0.        5000.
  +FLOUR WHEAT [  1600180]          0.           0.           0.        5000.
    OUTPUT COEF                  10000.       10000.       10000.       10000.
   EXTR.RATE                     10000.       10000.       10000.       10000.
   PRODUCTION                       0.           0.           0.        5000.
   IMPORTS                       12750.       13421.       19000.       16617.
   EXPORTS                        4649.        1477.        2061.       17861.
   FOOD                           8101.       11944.       16939.        3756.

** Brazil       **
** BREAD        **
**   20      21 **                 2005         2006         2007         2008
   INPUT                          490.         400.         500.         500.
  +FLOUR WHEAT [  1600200]        490.         400.         500.         500.
    OUTPUT COEF                  12000.       12000.       12000.       12000.
   EXTR.RATE                     12000.       12000.       12000.       12000.
   PRODUCTION                      588.         480.         600.         600.
   IMPORTS                          83.         172.         119.          92.
   EXPORTS                         660.         595.         592.         579.
```

```
   FOOD                              11.          57.         127.         113.

** Brazil        **
** PASTRY        **
**   22      21 **                 2005         2006         2007         2008
   INPUT                          29150.       32000.       28000.       29000.
  +FLOUR WHEAT [  1600220]        29150.       32000.       28000.       29000.
    OUTPUT COEF                   10000.       10000.       10000.       10000.
   EXTR.RATE                      10000.       10000.       10000.       10000.
   PRODUCTION                     29150.       32000.       28000.       29000.
   IMPORTS                         7825.        4975.        5874.        7155.
   EXPORTS                        36917.       36929.       33506.       35845.
   FOOD                              58.          46.         368.         310.

** Brazil        **
** WAFERS        **
**  110      21 **                 2005         2006         2007         2008
   INPUT                          11500.       12500.       12000.       15500.
  +FLOUR WHEAT [  1601100]        11500.       12500.       12000.       15500.
    OUTPUT COEF                   11000.       11000.       11000.       11000.
   EXTR.RATE                      11000.       11000.       11000.       11000.
   PRODUCTION                     12650.       13750.       13200.       17050.
   IMPORTS                          898.        1165.        1326.        1622.
   EXPORTS                        13044.       14456.       14020.       17836.
   FOOD                             504.         459.         506.         836.
```

Warnings should now include inconsistencies in the processing. The file **rpflex.out** lists only one warning, a ‘processing error’, see the next lines.

```
** WARNING : PROCESSED COMM.          664  : COCOA BUTTER
   PROCESS.ERR.                    9150.        7300.        6000.        5500.
```

This error arises from a convention in the construction of the SUAs. Cocoa Butter (664) is a joint input, with Cocoa Powder (665) in the production of Chocolate (666). But the working system cannot accommodate joint inputs. Therefore the processing of Cocoa Butter (664) does not seem to go anywhere. See for this Commodity Tree no. 46 in The ISC User’s Manual, page 171.

For more information on the processing the file **recov.out** can be inspected. It is an output file of the program that recovers the processing. This current file is somewhat inaccessible because it contains debugging information. Still some interesting information can be obtained from it. For instance, the file reports that for the SUAs of Brazil 189 different processing activities are detected (and temporarily stored, outside the working system!).

In addition some recurring warnings are found: the commodities Fruit Pulp Feed (628) and Dregs from Brewing and Distilling (654) are deliberately removed as joint outputs from the processing structure so as not to complicate the recovering proces. This results in fairly innocent warnings. Also the warning for the Cocoa Butter (664) appears standardly.

9.3 *Aggregation structure*

Taking distance from the actual data we turn to the specification of the aggregation. The aggregation directives by activity, see subsections 4.2 and 6.4.1, are given in annex 1 and the file **LST\iobas22.lst** . Consider the next activities with Wheat (15) as input.

```
                I      15   WHEAT
b      1500161   O      16   FLOUR WHEAT     17   BRAN WHEAT      19   GERM OF WHEA
b      1500162   O      16   FLOUR WHEAT     17   BRAN WHEAT
b      1500210   O      21   BULGUR
b      1500410   O      41   BREAKF CERLS
b      1500490   O      49   MALT BARLEY
b      1501130   O     113   CER PREP NES
b      1506320   O     632   ALC NON-FOOD
b      1506340   O     634   BEV DIS ALC
```

The aggregation directives in the first column above may however be overwritten by autocuts, see subsection 6.4.1 again and the file **autocuts.lst** . Both alcoholic products (632 and 634) are listed in the file. In the file **BIN\dirarc.ags** it can be observed that the activities involving these outputs are assigned the aggregation directive 'c' for cut. This part is therefore commodity driven.
The file **BIN\dirarc.ags** lists all directives. See the next small part of this file.[11]

```
   1 :    1500161 :B
   2 :    1500162 :B
   3 :    1500210 :B
   4 :    1500410 :B
   5 :    1500490 :B
   6 :    1501130 :B
   7 :    1506320 :C
   8 :    1506340 :C
```

Next there are the subsequent commodity lables for aggregation which can be found in a file called **BIN\comlab.ags** . See again a small part.

```
  15   WHEAT          LABEL : T
  16   FLOUR WHEAT    LABEL : B
  17   BRAN WHEAT     LABEL : B
  18   MACARONI       LABEL : B
  19   GERM WHEAT     LABEL : B
  20   BREAD          LABEL : B
  21   BULGUR         LABEL : B
  22   PASTRY         LABEL : B
  23   WHEAT,STARCH   LABEL : B
  24   WHEAT GLUTEN   LABEL : B
  26   WHEAT FER BE   LABEL : T
```

[11] The first column of this small part of the file shows that internal to the programs there is a simple numbering of the processing activities.

Bran of Wheat (17) and Germ of Wheat (19) logically have been assigned the lable ‘b’ for backward. However, they have also been assigned weight zero, see subsection 3.4 and the file **LST\wghts.lst** . See also annex 5 to part II. Because of this they are retained as targets, see subsection 4.3, so that possible food consumption will not be excluded. This doubleness is reflected in the Summary table, annex 0 to part II.

File **BIN\comlab.ags** gives more info: there are only 4 ‘forwards’, see subsections 4.1 and 6.4.1. The domain CBS, Commodity Balances – Supply Method, contains far more lables ‘f’. This is because in that particular aggregation the oilseeds are expressed in their oil and cake equivalents.

9.4 *Aggregation commands*

The the actual data file with processing is confronted with the aggregation structure. This leads to the aggregation commands. They can be found in a file called **ag111.out**, which is, alas, also cluttered with debugging information.

Simple

See the first command found, the values given are for the year 2008.

```
CTR.   21 :   16 FLOUR WHEAT   BACKWARD TO    15 WHEAT          WITH :
 SHARE:     1.0000  WEIGHT:     1.0000  FACT:     1.3889   MULT:     1.3889
```

The commands are built up as Mult = Share * Weight * Fact.

The situation of Wheat (15) and Flour of Wheat (16) is in principle that of weighted backward calculation because of the presence of joint outputs. But the commodities Bran and Germ of Wheat (17 and 19) have weight zero. Therefore this situation reduces to the most simple case. The factor is the inverse of the extraction rate for the year 2008: 1 / 0.72 = 1.3889 .

Weights

There is a command, or rather two, with non-zero weights.[12]

```
CTR.   21 :   46 BARLEY,PEARL  BACKWARD TO    45 POT BARLEY     WITH :
 SHARE:     1.0000  WEIGHT:     0.5612  FACT:     1.8182   MULT:     1.0204
CTR.   21 :   48 BARLEY FL GR   BACKWARD TO    45 POT BARLEY    WITH :
 SHARE:     1.0000  WEIGHT:     0.4388  FACT:     2.3256   MULT:     1.0204
```

[12] This was a surprise to me, these joint outputs must have escaped my attention in the mid 1990s.

Barley, Pearled (46) and Barley Flour and Grits (48) are joints outputs from a processing activity that has Pot Barley (45) as input. This activity (4500460) has directive ‘b’.

The commodities (46 and 48) have not been assigned weights, the file **LST\wghts.lst** .

Then the weights are constructed from the extraction rates, which may be found in the accounts.

Inspecting the file **sua.prt** we find that there are no accounts for commodities 46 and 48 for the years 2005 up to 2008. In earlier years there have been some imports.

As no country and year specific extraction rates are found, default values are used.

They are, see annex 3, 0.55 for Barley, Pearled (46) and 0.43 for Barley Fl Gr (48).

The weights are then: 0.55 / (0.55 + 0.43) = 0.5612 and 0.43 / 0.98 = 0.4388 .

The factors are the inverse of the (default) extraction rate and the conversion factor (Mult) to Pot Barley (45) is then the product of the weight and the factor.

Of course, since there are no accounts for the last four years this command will do nothing there.

Shares

When shares differ from 1, then also more than 1 command must follow.

```
CTR.   21 : 1043 LARD          BACKWARD TO  1037 FAT PIGS      WITH :
 SHARE:     0.6643  WEIGHT:     1.0000  FACT:     1.0669   MULT:     0.7087
CTR.   21 : 1043 LARD          BACKWARD TO  1040 PIG BUT FAT   WITH :
 SHARE:     0.3357  WEIGHT:     1.0000  FACT:     1.0669   MULT:     0.3581
```

In the file **ioflows.prt** the account for Lard (1043) is found.

```
 ** Brazil       **
 ** LARD         **
 ** 1043      21 **                  2005         2006         2007         2008
    INPUT                         373340.      372804.      411644.      417046.
   +FAT PIGS    [103710430]       243340.      232804.      271644.      277046.
     OUTPUT COEF                    9284.       10105.        9321.        9373.
   +PIG BUT FAT [104010430]       130000.      140000.      140000.      140000.
     OUTPUT COEF                    9284.       10105.        9321.        9373.
    EXTR.RATE                       9284.       10105.        9321.        9373.
    PRODUCTION                    346600.      376700.      383700.      390900.
    EXPORTS                         2482.        3215.        3532.        2287.
    FOOD                          344118.      373485.      380168.      388613.
```

The shares are computed as the contribution to production. It is unlikely that the output coefficients would be the same for the two different origins, but the current database allows only for storage of one ‘technical conversion factor’. Thus the shares can also computed on the basis of the ‘input’.

The share for activity (103710430) becomes 277046 / 417046 = 0.6643 .

The other share then equals 1 - 0.6643 = 0.3357 .

Skimmed Milk (888) is an interesting case: shares for backward calculation to the same product.

```
CTR.   21 :  888 SK MILK COWS  BACKWARD TO   882 COW MILK      WITH :
 SHARE:     0.9723  WEIGHT:     1.0000  FACT:     1.0438   MULT:     1.0149
CTR.   21 :  888 SK MILK COWS  BACKWARD TO   882 COW MILK      WITH :
 SHARE:     0.0277  WEIGHT:     1.0000  FACT:     1.0438   MULT:     0.0289
```

Skimmed milk is produced in two distinct processing activities, one where cream is made from whole cow milk, and the other where butter is made. Cream (885) and Butter (886) have zero weights, so they are retained as targets. This fits the focus of the Commodity Balances - Demand method: processed products with different characteristics than the originating product are kept separate.

9.5 *Targets*

The aggregation commands may involve addition to commodities which are not targets. For instance, Pot Barley (45) from above is backward calculated to Barley (44). So, also Barley, Pearled (46) and Barley, Flour & Grits (48) must end up with the account for the target Barley (44).
Repeated application of the commands is required. After a (finite) number of rounds a set of final aggregation commands to the targets results. These can be found in the file **new.nam** .

```
  21   45   POT BARLEY      to   44   BARLEY                1.4286
  21   46   BARLEY,PEARL    to   44   BARLEY                1.4577
  21   48   BARLEY FL GR    to   44   BARLEY                1.4577
```

The conversion factor for commodities 46 and 48 to 44 are the product of two commands, see 3.6 :

$$1.0204 * 1.4286 = 1.4577$$

For the barley products these numbers are, accidentally, default values, but this file does contain the country specific values.

After the succesive application of the commands the target commodities remain. They can be printed. See the first account from file **target.prt** .

```
 ** Brazil       **
 ** WHEAT        **     2005          2006          2007          2008
 **   15      21 **

  PRODUCTION        4658790.      2484848.      4114060.      6027131.
  IMPORTS           5079332.      6765193.      7566425.      7036475.
  FROM STCKS        1000000.      2300000.      -500000.     -1050000.
  EXPORTS            233393.       726295.       175191.       745587.
  FEED               200000.       300000.       100000.       200000.
  SEED               150579.       157680.       202762.       207564.
  WASTE              517025.       540892.       530797.       573014.
```

```
PROCESSED              0.            0.            0.            0.
FOOD             9636191.      9823198.     10169163.     10284463.
OTHER UT.            935.         1975.         2570.         2977.
```

Even though commodity code 15 is printed here, the commodity involved is no longer wheat itself, it is now wheat and products. Production should of cource be the same as in the detailed account (and it is). Food consumption was not seen in the detailed account, see 9.1, now it is there.

RwcflexTAR.out gives the warnings of imbalances at the target level. Inspection of this files gives a quick & dirty assesment of the aggregation: pretty clean, there are no huge imbalance for the important products.[13]

9.6 *Further aggregation*

The next step is further aggregation, where there is relabling of accounts and grouping together. Consider the next part from reference file **LST\BDStep1.lst** , or annex 8.

```
2511  WHEAT              15  WHEAT              1.00
2511  WHEAT              17  BRAN WHEAT         0.00
2511  WHEAT              19  GERM WHEAT         0.00
2511  WHEAT              24  WHEAT GLUTEN       0.00
2511  WHEAT             114  MIXES AND DO       0.00
2511  WHEAT             115  FOOD PREP.FL       0.00
```

The zero weights for the targets other than Wheat (15) mean that only the nutritive values associated with the availability for food consumption (element 141) are added. See also annex 0. These nutritive values are calories, protein and fat. They are not shown in the examples here.[14]

The results from the further aggregation can be seen in **cbd.prt** and the warnings in **rwcoptCBD.out** .

Finally there is a merging with the results from the CBS aggregation. This is found in **fbs.prt** and warnings in **rwcopt.out** .

[13] My simple print programs do not accommodate specific item types. Warnings for products such as Hen Eggs No (1067) are therefore false.

[14] Observe that the treatment of the calories cs. is fundamentally different from the treatment of the quantities of products. Calories can be added and need no conversion!

10 Backtracking

This section presents an example of error tracking. The error is hoped to be explainable.

The first sizable imbalances after further aggregation are found for Oil from oilcrops n.e.s. (2586). See in **rwcoptCBD.out** .

```
WARNING : IMBALANCE FOR CTR.           21  FOR COM.        2586   OILCR NES OI :
  61:     -12419.     -14576.     -17712.       -280.
```

In **cbd.prt** the account is as follows.

```
 ** Brazil        **
 ** OILCR NES OI **     2005          2006          2007          2008
 ** OILCROPS, OTHER OIL      **
 ** 2586      21 **

  PRODUCTION         156175.       134812.       135287.       139182.
  IMPORTS             15715.        14854.        19197.        23222.
  EXPORTS             78608.        71268.        69707.        43084.
  PROCESSED            9150.         7300.         6000.         5500.
  OTHER UT.           96551.        85674.        96489.       114100.
  * IMBAL.           -12419.       -14576.       -17712.         -280.
```

Apart from the imbalance there is also non-zero processed in the account, though this is not the same amount. Let us backtrack the aggregation procedure and explain both.

First have a look at the default composition of FBS items, from annex 9.

The commodities marked with an * are found in the data.

```
  2586  OILCROPS, OTHER OIL          264  KARIT NT BUT
  2586  OILCROPS, OTHER OIL          266  OIL CAST BNS     *
  2586  OILCROPS, OTHER OIL          276  TUNG OIL         *
  2586  OILCROPS, OTHER OIL          278  JOJOBA OIL
  2586  OILCROPS, OTHER OIL          281  OIL SAFFLWR
  2586  OILCROPS, OTHER OIL          297  OIL POP SD
  2586  OILCROPS, OTHER OIL          306  VEG TALLOW
  2586  OILCROPS, OTHER OIL          307  STILLING OIL
  2586  OILCROPS, OTHER OIL          313  OIL OF KAPOK
  2586  OILCROPS, OTHER OIL          334  OIL LINSEED      *
  2586  OILCROPS, OTHER OIL          337  OIL HEMPSD
  2586  OILCROPS, OTHER OIL          340  OIL VG OR NS     *
  2586  OILCROPS, OTHER OIL          664  COCOA BUTTER     *
  2586  OILCROPS, OTHER OIL         1241  LIQUID MARG
  2586  OILCROPS, OTHER OIL         1242  MARGARINE        *
  2586  OILCROPS, OTHER OIL         1273  CASTOR OIL H
  2586  OILCROPS, OTHER OIL         1274  OILS BOILED      *
  2586  OILCROPS, OTHER OIL         1275  OILS HYDROGN     *
```

Next look at the target products to FBS items, with weights, from annex 8.

```
2586  OILCR NES OI      266  OIL CAST BNS       1.00
2586  OILCR NES OI      276  TUNG OIL           1.00
2586  OILCR NES OI      278  JOJOBA OIL         1.00
2586  OILCR NES OI      281  OIL SAFFLWR        1.00
2586  OILCR NES OI      297  OIL POP SD         1.00
2586  OILCR NES OI      306  VEG TALLOW         1.00
2586  OILCR NES OI      307  STILLING OIL       1.00
2586  OILCR NES OI      313  OIL OF KAPOK       1.00
2586  OILCR NES OI      334  OIL LINSEED        1.00
2586  OILCR NES OI      337  OIL HEMPSD         1.00
2586  OILCR NES OI      340  OIL VG OR NS       1.00
2586  OILCR NES OI      664  COCOA BUTTER       1.00
```

Then see default conversion factors (to targets), from annex 6, for a set of higher order derived products. Also the conversions to Oil from linseed (334) and Oil from castor beans (266) are relevant because they end up in aggregate commodity 2586.

```
- 1241  LIQUID MARG     to  340  OIL VG OR NS         1.0000
- 1242  MARGARINE       to  340  OIL VG OR NS          .8333     *
- 1273  CASTOR OIL H    to  266  OIL CAST BNS          .9091
- 1274  OILS BOILED     to  334  OIL LINSEED          1.0000     *
- 1275  OILS HYDROGN    to  340  OIL VG OR NS          .8333     *
```

Now we know which commodities from the initial SUA data to look for. We need them with the recovered processing, i.e. from **ioflows.prt** (one more reason to have this readily available).

```
** Brazil       **
** OIL CAST BNS **
**  266      21 **                  2005         2006         2007         2008
    PRODUCTION                     70200.       48700.       43300.       52300.
    IMPORTS                           32.          10.        3738.        6744.
    EXPORTS                        11782.        4343.         746.         237.
    OTHER UT.                      58450.       44367.       46292.       58807.

** Brazil       **
** TUNG OIL     **
**  276      21 **                  2005         2006         2007         2008
    PRODUCTION                        55.          55.          50.          50.
    IMPORTS                          500.         670.           0.           0.
    EXPORTS                            0.           0.           1.           0.
    OTHER UT.                        555.         725.          49.          50.

** Brazil       **
** OIL LINSEED  **
**  334      21 **                  2005         2006         2007         2008
   PRODUCTION                     3400.        3600.        3800.        4300.
   IMPORTS                         339.         665.         586.         483.
   EXPORTS                          13.          10.         521.          13.
   OTHER UT.                      3726.        4255.        3865.        4770.

** Brazil       **
** OIL VG OR NS **
**  340      21 **                  2005         2006         2007         2008
   PRODUCTION                     4000.        4000.        5800.        5900.
   IMPORTS                         572.         436.        1335.        1323.
```

```
   EXPORTS                                 923.         2689.         4446.         1891.
   OTHER UT.                              3649.         1747.         2689.         5332.


** Brazil       **
** COCOA BUTTER **
**  664      21 **                        2005          2006          2007          2008
   PRODUCTION                            78520.        78457.        82337.        76632.
   IMPORTS                                   0.            0.            1.            6.
   EXPORTS                               39199.        36577.        32744.        25997.
   PROCESSED                              9150.         7300.         6000.         5500.
   PROCESS.ERR.                           9150.         7300.         6000.         5500.
   OTHER UT.                             30171.        34580.        43594.        45141.
```

Here we find a reservation for processing that is not accounted for. The warning for this was already found in **recov.out** and was discussed in subsection 9.2 . It is an old SUA convention.

```
** Brazil       **
** MARGARINE    **
** 1242      21 **                        2005          2006          2007          2008
   INPUT                                393000.       398000.       400000.       408000.
  +OIL MAIZE   [  6012420]               38000.        38000.        40000.        45000.
    OUTPUT COEF                          12595.        12513.        12575.        12500.
  +OIL SOYABEAN[ 23712420]              355000.       360000.       360000.       363000.
    OUTPUT COEF                          12595.        12513.        12575.        12500.
  +OIL,PALM KER[ 25812420]                   0.            0.            0.            0.
    OUTPUT COEF                          12595.        12513.        12575.        12500.
   EXTR.RATE                             12595.        12513.        12575.        12500.
   PRODUCTION                           495000.       498000.       503000.       510000.
   IMPORTS                                   0.            0.            4.            1.
   EXPORTS                               20353.        14372.        24868.        34548.
   FOOD                                 474647.       483628.       478136.       475453.
```

Though Margarine (1242) is a potential contributor to the aggregate Oil from oilcrops n.e.s. (2585) in the case of Brazil this is superceded by the recorded processing origins: Margarine (1242) comes from Maize oil (60) and from Soybean oil (237). In some earlier years it may have originated from Oil of palm kernels (258). So, Margarine (1242) will not be backward calculated to commodity 340 and will not end up in Oil from oilcrops n.e.s. (2586).

```
** Brazil       **
** OILS BOILED  **
** 1274      21 **                        2005          2006          2007          2008
   IMPORTS                                 874.         1679.         1292.         2348.
   EXPORTS                               11331.        14276.        19600.         8545.
  * IMBAL.                              -10457.       -12597.       -18308.        -6197.

** Brazil       **
** OILS HYDROGN **
** 1275      21 **                        2005          2006          2007          2008
   IMPORTS                               16078.        13673.        14694.        14782.
   EXPORTS                               18432.        16048.        13979.         7681.
   * IMBAL.                              -2354.        -2375.          715.         7101.
```

Then there are these two higher level derived products with large imbalances.
Here is the source of the imbalance at aggregate level!

From **new.nam**, see 9.5, the actual final aggregation commands for Brazil 2008 can be found. They are the defaults seen above.

```
  21 1274  OILS BOILED     to  334  OIL LINSEED          1.0000
  21 1275  OILS HYDROGN    to  340  OIL VG OR NS         0.8333
```

Then we can start computing:
- 6197 * 1.000 = - 6197 to commodity 334 and 7101 * 0.8333 = 5917 to commodity 340,
and both are mapped with weight 1 to aggregate commodity 2586, which gives the sum of -280.
And this is the imbalance found at the aggregate level.

References

FAO (1986), The ICS Users' Manual

FAO (1987), The AGROSTAT Codebook

FAO (1993), 'Definition and classification of commodities', (draft)

Fischer, G. and U. Sichra (1983), 'The Aggregation of the Agricultural Supply Utilization Accounts', IIASA WP-83-42, Austria, Laxenburg

Keyzer, M.A. and M. van 't Riet (1993), 'Checking and aggregating FAO's Supply Utilization Accounts', SOW-VU WP-93-06, Centre for World Food Studies, Amsterdam

Van't Riet, Maarten (1997a), 'Aggregation of FAO's Supply Utilization Accounts for the Food Balance Sheets', consultancy note, Rome, April 1997

Van't Riet, Maarten (1997b), 'Documentation of the programs and procedures for the io-edits and standardisation', Rome, April 1997

Van't Riet, Maarten (1998), 'The Food Balance Sheets and the Commodity Balances - Demand and Supply method', documentation note, Rome, May 1998

Van't Riet, Maarten (2011), 'The activity dimension of FAO's Supply Utilization Accounts', consultancy paper, The Hague, May 2011

Veenendaal, P.J.J. (1991), 'Aggregation of FAO's supply utilisation accounts to ECAM's exchange commodity list', LEI-DLO, The Hague, mimeo

Annexes to Part II: a separate file **annexes-part-II.pdf**

Files with Part III: a separate zip-file **files-with-part-III.zip**

The reference files are all found in the subdirectory LST, the result files are in the main directory.