The Biomass Study
Task 5: Integrated Modeling Framework

R-Tool for Output Processing

Jorge Ortigosa
Claudia Bulgheroni

Summary

The Biomass project pursues the setting of a knowledge base on the biomass sector in Europe. This includes the setting up of an Integrated Modeling Framework (IMF) for the analysis of future EU biomass supply and demand under different scenarios, as well as its potential environmental, economic and social effects. The IMF is composed by 5 models: the *Carbon Budget Model* (CBM), the *Global Forest Trade Model* (GFTM), the *Land Use Integrated Sustainability Assessment platform* (LUISA), the *Common Agricultural Policy Regionalised Impact* (CAPRI) and *The Integrated MARKAL-EOFM System for the EU* (EU-TIMES).

Each of the 5 models composing the IMF have been developed separately, to be applied independently and not as a part of the IMF. This means that there are great differences among them, for instance in the structure of the outputs, the nomenclature and the spatial scale and regional sub-divisions considered. Not to mention the differences in the basic assumptions and in the data sources, such as FAO, Eurostat, or guidelines applied, such as those from the IPPC for the definition of harvested wood products. Therefore, if any insight into the future of the biomass sector is to be obtained from the outputs of these models, it is necessary to find a common ground for all of them, which, at the same time is harmonized both with the terminology used to define the scenarios, and with the IPCC methodology for the calculation of different parameters.

With this purpose, and given the huge amounts of data involved in this project, it was decided to develop a tool in R language (R-Tool), which would harmonize the outputs of the models, complete them if necessary, introduce external data (e.g. LCA, emission factors) and calculate the values of a set of environmental, social and economic, indicators to assess the impact of the scenarios considered.

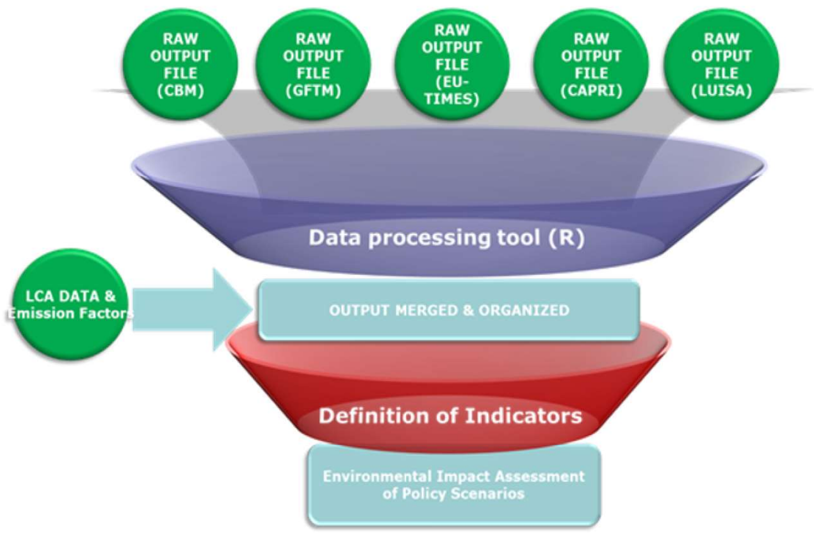

Figure 1: R-Tool Scheme

However, since the only outputs currently available are those from CBM and GFTM, the work has focused on the development of the R-Tool utilities for the forest sector. In this sense, the following tasks have been accomplished so far:

- A good insight into the models workings, in particular about the definitions of the categories considered in each of them, was gained through the review of the literature available and meetings with the persons in charge of running them.
- A correspondence between the models parameters, and between them and those used by FAO and the IPCC was established.
- When this direct correspondence did not exist, modifications were implemented, and missing parameters were calculated out of the existing data, in order to guarantee it.
- Since no set of indicators had been agreed upon so far, an extensive literature review on the most common environmental indicators used to monitor the forest sector was carried out.
- Based on the mentioned review, and taking into account the currently available data, a set of environmental indicators was chosen and the code to calculate them was developed (i.e. Forest C-Stock (Living Biomass, Death Organic Biomass, Soil Organic Biomass), Forest CO2 Absorption, Forest CO2-equivalent emissions, Harvested Wood Products C-Stock, Harvested Wood Products CO2 emissions).

Since the final indicators set is still under definition, the indicators were selected so that they would be useful regardless the definitive indicator set. Therefore, only basic performance and pressure environmental indicators were chosen. Possible next steps, provided no new outputs will be available, might include the validation of the indicators values obtained so far, the expansion of the set of indicators or the introduction of external parameters in the calculations (e.g. emission factors).

Contents

List of Figures

List of Tables

Nomenclature

AGB	Above Ground Biomass
AR	Afforestation and Reforestation
BGB	Below Ground Biomass
C	C-Content in dry matter [g of C/g]
$CAPRI$	Common Agricultural Policy Regionalised Impact
CBM	Carbon Budget Model
CF	Conversion Factors
D	Deforestation
DB'	Tree species density [g $/m_3$]
DOM	Death Organic Matter
EC	European Commission
EOL	End of Life
$EU - TIMES$	The Integrated MARKAL-EOFM System for the EU
FAO	Food and Agriculture Organization of the United Nations
$FAOSTAT$	FAO Statistics service
$FAWS$	Forest Area Available for Wood Supply
FW	Fuel Wood
HL	Half-Lives
HWP	Harvested Wood Products
IMF	Integrated Modelling Framework
$IPCC$	International Panel on Climate Change

IRW	Industrial Roundwood
JRC	Joint Research Center
k	Decay Factors
LB	Living Biomss
LM	Land Management
$LUISA$	Land Use Integrated Sustainability Assessment platform
$LULUCF$	Land Use Land Use Change and Forest
OW	Other Wood
PFP	Primary Forest Products
PP	Primary Products
PPB	Paper and Paperboard
$SFWP$	Semi-finished harvested wood products
SOM	Soil Organic Matter
SPU	Spatial Unit
$SPUID$	Spatial Unit Identifier
StW	Stem Wood
SW	Sawnwood
VS	Veneer Sheets
Wbp	Wood-based panels

Chapter 1

Introduction

1.1 Background

The Biomass Study has its origin in a mandate given to the Joint Research Center (JRC) by a number of European Commission (EC) services[1] in 2015 to establish a knowledge base on the biomass sector in Europe, comprehending biomass flows, supply and demand on a long-term basis. This involved, among other things, the analysis of the prospective biomass supply and demand in the short, medium and long-term, under different scenarios and the assessment of their potential social, environmental and economic effects[2]. In order to do so an *Integrated Modeling Framework* (IMF), composed by 5 models was set up: the *Carbon Budget Model* (CBM), the *Global Forest Trade Model* (GFTM), the *Land Use Integrated Sustainability Assessment platform* (LUISA), the *Common Agricultural Policy Regionalised Impact* (CAPRI) and *al* (EU-TIMES). These models will be used to make predictions of future biomass supply and demand in Europe from 2015 onward.

[1] Directorate Generals Agriculture and Development (AGRI), Climate Action (CLIMA), International Cooperation and Development (DEVCO), Energy (ENER), Environment (ENV), Internal Market, Industry, Entrepreneurship and SMEs (GROW), Maritime Affairs and Fisheries (MARE), Mobility and Transport (MOVE), Regional and Urban Policy (REGIO), Research and Innovation (RTD), Secretariat-General (SG) and Trade (TRADE) . See [2]

[2] *Mandate on the provision of data and analysis on biomass flow, supply and demand by JRC on a long-term basis:* `https://biobs.jrc.ec.europa.eu/biomass-assessment-study-jrc`

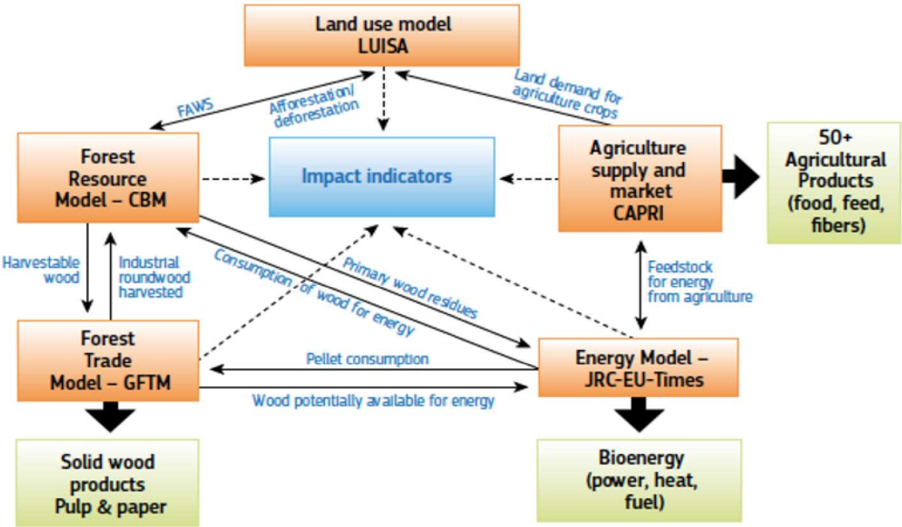

Figure 1.1: The Biomass Study Integrated Modelling Framework

1.2 Objectives

The impact assessment will be carried out through a set of indicators to be calculated out of the combined outputs of the models and data bases. However, since the models were developed independently, to be used separately and not combined with the others, the nomenclature used differs from one to another, as well as the structure of the outputs, the period of time covered etc. Moreover, in some cases the names categories considered in the outputs do not match or only do it partially the categories considered by the most commonly used methodology for the calculation of the indicators (e.g. IPCC Guidelines[7]), and therefore they have to be complemented or the appropriate categories have to be calculated out of them. Finally, the huge amount of data they produce in each run poses a challenge in terms of data management. For these reasons, it was decided to create a tool (henceforth *R-tool*) using the programming language R[3]. which should be able to perform the following tasks:

- Extract the information of interest from the outputs.
- Re-shape and re-structure it to fit a standard format.
- Complete information, aggregate categories and calculate the values for a pre-defined set of indicators.

[3]R is a free software environment and programming language specifically developed to handle great amounts of data in all kind of formats. For the present work the version R i386.3.5.1 has been used with R-Studio as a command interface. For more information about R see `https://www.r-project.org/`

1.3 Tool Structure

The R-tool is composed by a number of packages, one for each of the five models composing the IMF. In addition to this, a number of extra packages may have to be developed, for instance to combine the already processed outputs of the different models or to extract required historical data from different data sources such as FAOSTAT.

Each package is composed by a number of functions fulfilling different tasks. In general, all the packages with the exception of the one processing the historical data, contain the following functions:

- *read function*: interprets the models′ outputs, makes the first assignation of names, and creates an object on which the following functions can easily work. This function was thought as a sort of filter between the raw model outputs and the functions conforming the core of the packages, and therefore, in case the a output changes , it will have to be modified, so that it keeps producing the same results.
- *prepare function*: further processes the outputs, adding categories when missing, or complementing those that required it using historical data or other sources. The resulting object is ready to be used to calculate the indicators corresponding to the model.
- *analysis 1 function*: calculates one or several of the the indicators selected.
- *analysis 2 function*: calculates one or several of the the indicators selected.

  ....
- *analysis n function*: calculates one or several of the the indicators selected.
- *results function*: reshapes the outputs of the *analysis* functions according to a pre-defined format.
- *graphs function*: represents the outputs of the previous functions.

1.4 Limitations

As it is now, the IMF presents a number of limitations that should not be overlooked. For instance, no model accounts for the emissions produced by the imported materials and raw products before entering Europe or considers the construction or material sector, thus neglecting the substitution potential of the wood products. Finally, the wood products End-of-Life stage (EOL), is not considered at any point.

In addition to these, there are other limitations related to the current status of the study. At the moment of writing this report only the outputs of two of the five models composing the IMF, namely CBM and GFTM, are available. Therefore, only the packages related to those models could be developed. For this same reason, it is difficult to know what the outputs of the remaining models would contain, and therefore it is highly probable that the already developed packages will have to be be modified to some extent in the future to make them compatible with the new outputs.

When it comes to the individual models it is necessary to point out that:

- Between 2015 and 2030 CBM only considers one land type, namely the Forest Area available for Wood Supply (FAWS), without making any distinction between the different kinds of forested land, such as reforested area, afforested area, protected area, High Biodiversity Value area etc. From 2030 these an other distinctions will be included, and it will be possible to make specific calculations for each of them.
- In [8, 7, 1] four kinds of forest activities are considered within the Land Use, Land Use Change and Forest (LULUCF) sector: Afforestation and Reforestation (AR), Deforestation (D) and Land Management (LM). So far, CBM runs do not consider the first two, and therefore all the removals fall within the last category.

1.5 Indicators

The effects that the different policies on the biomass sector may have on the environment, the society and the economy are to be assessed by means of a set of indicators of different types (i.e. impact, pressure, performance). However,, no such a set of indicators has been defined for the Biomass Study so far. Besides, as it has been mentioned in the previous section, only the outputs of the models referring to the forest sector have been made available. For this reason, a literature review on the most common indicators used to assess environmental impacts in the forest sector had to be carried out(see appendix A), resulting in the selection of a set of indicators. The criteria followed to choose the indicators, were, in the first place, the availability of the data required to calculate them, and, in the second, their future utility regardless the final set of indicators.

All the indicators selected fall within the *pressure* and *performance* indicators category as defined in [3].

Forest Carbon Stock

Definition: Tonnes of carbon in the main forest carbon pools (living biomass above ground, living biomass below ground, dead wood and soils) as defined in [7].

Forest CO_2 Absorption

Definition: Tonnes of CO_2 absorbed from the atmosphere by forested areas annually.

Forest CO_2-equivalent emissions

Definition:Tonnes of CO_2-equivalent emitted by the forest to the atmosphere annually. Note that, as it was explained in the previous section, only land management contribution will be so far.

Forest Removals

Definition: Tonnes of carbon removed from the forest annually.

Harvested Wood Products (HWP) Carbon Stock

Definition: Tonnes of carbon stored in semi-finished harvested wood products (SFWP), namely Sawn-wood, Paper and Paperboard and Wood-pulp [8].

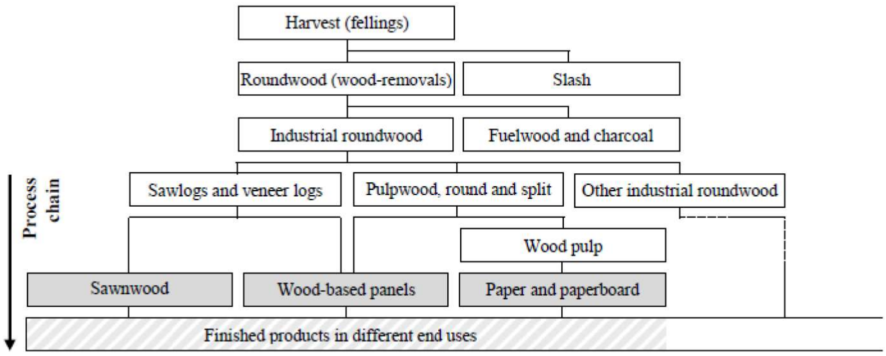

Figure 1.2: Simplified FAO forest wood products classification [8]

HWP CO_2 equivalent emissions

Definition: Tonnes of CO_2-equivalent emitted by the HWP to the atmosphere annually.

These last two indicators were evaluated following the IPCC Tier 2 methodology [8], which is composed by a number of steps:

1. Estimation of the annual fraction of domestic primary products (PP), that is Industrial Roundwood (IRWc and IRWnc), and Pulpwood (PULP), over their total annual consumption. This is done by dividing the difference between production and exports by the sum of production plus imports minus exports, which is equal to the consumption.

$$f_{PP}(i) = \frac{PP_P(i) - PP_{EX}(i)}{PP_P(i) + PP_{IM}(i) - PP_{EX}(i)} \tag{1.1}$$

   To calculate IRW domestic fraction, it is necessary to calculate coniferous and non-coniferous specific domestic product fractions. In order to do so, the first step is calculating the IRWc and IRWnc weights:

$$w_{IRWc}(i) = \frac{IRWc_{Cons}(i)}{IRWc_{Cons}(i) + IRWnc_{Cons}(i)} \tag{1.2}$$

   Then, together with the domestic product fractions calculated with eq. 1.1, a value for the IRW domestic fraction can be calculated:

$$f_{IRW}(i) = f_{IRWc}(i) * w_{IRWc}(i) + f_{IRWnc}(i) * w_{IRWnc}(i) \tag{1.3}$$

2. In case different activities and land types are considered (i.e Afforested land , Deforested land, Reforestation, Forest Management) an additional factor that accounts for the fraction of the total harvest due to the different activities must be calculated as follows:

$$f_{Activity}(i) = \frac{Harvest_{Act}(i)}{Harvest_{Total}(i)} \tag{1.4}$$

   Then, the annual domestic fraction of any PP due to a particular activity could be calculated as follows:

$$f_{PP_{Ac}}(i) = PP(i) * f_{PP}(i) * f_{Ac}(i) \tag{1.5}$$

3. Which primary product domestic fraction is to be used to calculate the SFWD domestic product fractions can be derived from Fig. 1.2. Thus, f_{IRWc} and f_{IRWnc}, will be used to calculate the coniferous and non-coniferous sawnwood (SW) domestic fractions respectively, f_{IRW} to calculate wood-based panels′ (Wbp) and f_{IRW} together with f_{PULP} to calculate paper and paperboard′s (PPB). Note that no activity factor is included here.

$$f_{SWc}(i) = SW_{Prod}(i) * f_{IRWc}(i); f_{SWnc}(i) = SW_{P}(i) * f_{IRWnc}(i) \tag{1.6}$$

$$f_{SW}(i) = f_{SWc}(i) + f_{SWnc}(i) \tag{1.7}$$

$$f_{Wbp}(i) = Wbp_{Prod}(i) * f_{IRW}(i) \tag{1.8}$$

$$f_{PPB}(i) = PPB_{Prod}(i) * f_{IRW}(i) * f_{PULP}(i) \tag{1.9}$$

4. To calculate the actual annual carbon *inflow* into the HWP pool, the SFWP quantities, which are normally given in m_3 (SW, Wbp) or tonnes (PPB) have to be converted to tonnes of carbon. To do so the conversion factors (CF) provided for each SFWP provided in [8] will be used.

$$Inflow_{SW} = f_{SW}(i) * CF_{SW} \tag{1.10}$$

Depending on the period covered, it may be necessary to estimate the carbon inflows into the HWP pool due to the different SFWPs also in the period previous to the first FAO records (in general 1961, although it depends on the country). In [8] it is assumed that carbon stock in 1900 for every SFWP is equal to 0 tonnes, and that from that year until the first year with historical records the SFWPs carbon inflow is constant and equal to the average of the carbon inflow in the first five years of the historical period.

$$\text{For t=1900 (i=0)} \implies C_{SFWP}(t) = 0 \text{ tonnes.} \tag{1.11}$$

$$\text{For i} \in (0{,}60)^{4} \implies Inflow_{SFWP}(i) = \frac{\sum_{i=61}^{66} Inflow_{SFWP}(i)}{5} \tag{1.12}$$

5. Once the annual carbon *inflow* for each SFWP have been calculated, in order to calculate the SFWPs carbon stock in each year, it is necessary to estimate the annual carbon *outflows* from the the different SFWPs′ pools due to the carbon decay. It is assumed this decay happens according to a first-order decay function, which variables are the annual carbon *inflow* and the decay factors (k), which are specific to each product and are related to the corresponding half-lives by $k = \frac{ln(2)}{HL}$.

$$C_{sfwp}(i{+}1) = e^{-k_{sfwp}} {*} C_{sfwp}(i) + \left[\frac{(1 - e^{-k_{sfwp}})}{k_{sfwp}}\right] {*} Inflow_{sfwp}(i) \tag{1.13}$$

$$\Delta C_{SFWP}(i) = C_{SFWP}(i+1) - C_{SFWP}(i) \tag{1.14}$$

From the previous equation it can be derived that the annual CO_2 emissions from each SFWP pool would be given by:

$$Emissions(i) = (1 - e^{-k}) * C(i) + \left[1 - \frac{(1 - e^{-k})}{k}\right] * Inflow(i) \tag{1.15}$$

Finally, summing up the figures for the different SWFPs′ it is possible to estimate the total HWP C-Stock, HWP C-Stock change and HWP CO_2 emissions.

1.6 Preliminary Steps

Given the amount of information and the great number of files to be managed, the way these are saved and called is of the utmost importance. Related to this, there are a few steps that have to be carried out before starting using the tool.

[4] Assuming that the historical FAO records begin in 1961

- Setting up the main directory and saving the direction as a global variable.
- Creating sub-directories within the main directory for the different packages and within them, folders for the models′ outputs and the intermediate results of the functions.
- Creating a specific sub-directory to host a number of files containing extra information required by the functions to work

The second and the third of these steps only have to be carried out the first time the functions are used, unless the directories structure is deleted, whereas the first has to be performed in each session, so that the functions know where the main directory is. In order to automatize to the extent possible these processes an additional R package *firststeps* was developed. It contains three functions, corresponding to the three tasks described above:

setpath

Before running this function the user must have created a main directory and an empty file of any kind inside it, for, although this function does not require an input, it asks the user to choose a file inside the directory to become the main directory. Then, the function automatically sets the path to this directory as an environmental variable `path` (it will appear on the global environment box on the upper right corner) to be used by the functions in the other packages.

```
setpath<-function(){
path<-file.choose()
path«-dirname(path)
}
```

structure

Once the path to the main directory is set, the present function builds sub-directories for each package inside it and folders for the outputs and results of the models and functions respectively inside the corresponding packages′ directories. For that purpose *structure* requires two vectors as inputs: one containing the names of the packages considered and a second containing the names of the folders to be created within the packages′ directories.

In the following example, directories for four packages (`packages<-c("CBM","GFTM" ,"FAOSTAT","JoinData")`) are created, each of which containing two folders (`folders<-c("Outputs","Results")`) .

```
models<-c("CBM","GFTM","FAOSTAT","JoinData")
```

```
folders<-c("Outputs","Results")

   structure<-function(m,f){
modelspaths<-paste(path,pack,sep="/")
lapply(modelspaths,function(x) dir.create(x))
modfol<-expand.grid(pack,folders)
folderpaths<-paste(modfol[,1],modfol[,2],sep="/")
folderpaths<-paste(path,folderpaths,sep="/")
lapply(folderpaths,function(x) dir.create(x))
}
```

addfiles

Finally, the third function of the package creates a folder `General Parameters` within the main directory, and fills it with files that are necessary for the functions to work. These files contain for instance, the column names of the GFTM raw outputs files, the SFWP half-lives and conversion factors or a list with the FAO items of interest.

```
+---AllData
    |   path
    |
    +---CBM
    |   +---Outputs
    |   \---Results
    +---FAOSTAT
    |   +---Outputs
    |   \---Results
    +---General Parameters
    |       .Rhistory
    |       AllCountries.rds
    |       ConversionFactors.rds
    |       ElementsGFTM.rds
    |       HalfLives.rds
    |       ItElGFTM.rds
    |       ItemsGFTM.rds
    |       MainElementsFAO.rds
    |       MainElementsGFTM.rds
    |       MainItemsFAO.rds
    |       UnitsGFTM.rds
    |
    +---GFTM
    |   +---Outputs
    |   \---Results
    \---JoinData
        +---Outputs
        \---Results
```

Figure 1.3: Directory Tree

The result of running these functions would be the directory tree in Fig. 1.3:

All the functions in the different packages have been developed according to the structure built by these three functions. This means that, they will extract the models′ raw outputs in the *Outputs* folder of the corresponding directory, will look for the files with additional information in the *General Parameters* directory and will save the intermediate and final results in the *Results* folders. Of course this is only one of the many possible directories structures. The user is encouraged to alter or completely change it at will.

Finally, there is a number of packages, not included in the downloaded R version, which are necessary for the tool functions to work (see appendix B). Packages only have to be installed once but they have to be loaded each session, so that the tool functions can make use of the functions contained in them. What follows is an example on how to install and load package *dplyr*, which contains, among other the *mutate* function[5].

```
>install.packages(dplyr)

>library(dplyr
```

1.7 General naming rules

In general the following guidelines have been adopted:

- FAO nomenclature is applied. This implies:
  - Harvested wood products (HWP) are classified in *coniferous* and *non-coniferous*, instead of *coniferous* and *broad-leaves*, or *softwood* and *hardwood* as they are in CBM.
  - FAO′s designations for the products, *items*, and for and the trade, consumption and production figures, *elements* are used. Thus, an item-element pair could be for instance *Plywood consumption* or *Industrial Round-wood coniferous production.*
  - GFTM country names were substituted by the FAO equivalent *Czech Republic –> Czechia.*
- To avoid problems with names formed by several words, such as *Sawnwood coniferous* or *United Kingdom*, the spaces between words have been systematically substituted by an hyphen, "-", e.g.*Sawnwood-coniferous, United-kingdom.* However, when for instance, item-elements couples are created the space between both elements will be substituted by an underscore "_".

  Thus for instance, the sawnwood coniferous production figures may be found under the *Sawnwood-coniferous_Production* header.
- When an function which does not belong to any of the basic packages included in the downloaded version is mentioned, the packages it comes from is indicated using the `::` mark, e.g. `dplyr::mutate`.

[5] Note that packages are usually downloaded from the internet and therefore a connection is required to install them. Once installed the user may load them as many times as necessary without need of a connection to the internet.

These rules, as well as the names used for the functions and for the different objects within the functions themselves have been chosen with the objective of making the understanding and use of the functions as easy as possible and can be naturally changed anytime.

Chapter 2

CBM Package

2.1 The Model

The Carbon Budget Model of the Canadian Forest Sector (CBM-CFS3) [6] is a semi-empirical model for carbon-dynamics in forestry and land-use change developed by the Department of Natural Resources of Canada, also known as Natural Resources Canada (NRCan) [1]. It applies the IPCC Good Practice Guidance Tier 3 approach [7] for reporting on carbon stock and carbon stock changes resulting from Land Use, Land-use Change and Forestry (LULUCF) and can be applied at stand (communities of trees that are homogeneous enough to be treated as an unit), landscape as well as at national level.

The forested area of interest is divided into a number of *Spatial Units* (SPU) identified by a number (*SPUID*), resulting from the intersection between administrative and climatic divisions (the latter defined by its temperature and precipitations [citation]). Each SPU contains several forest stands, characterized by a number of *classifiers* (Forest Type, Management Type, Conifer or Broad-leave etc).Each stand is thus described as a *ClassSets*, which will be assigned an identification number, *UserDefdClassSetsID*. The particular values these *classifiers* may take will be used to link each *ClassSets* with other CBM′s parameters and actions.

[1] `https://www.nrcan.gc.ca/home`

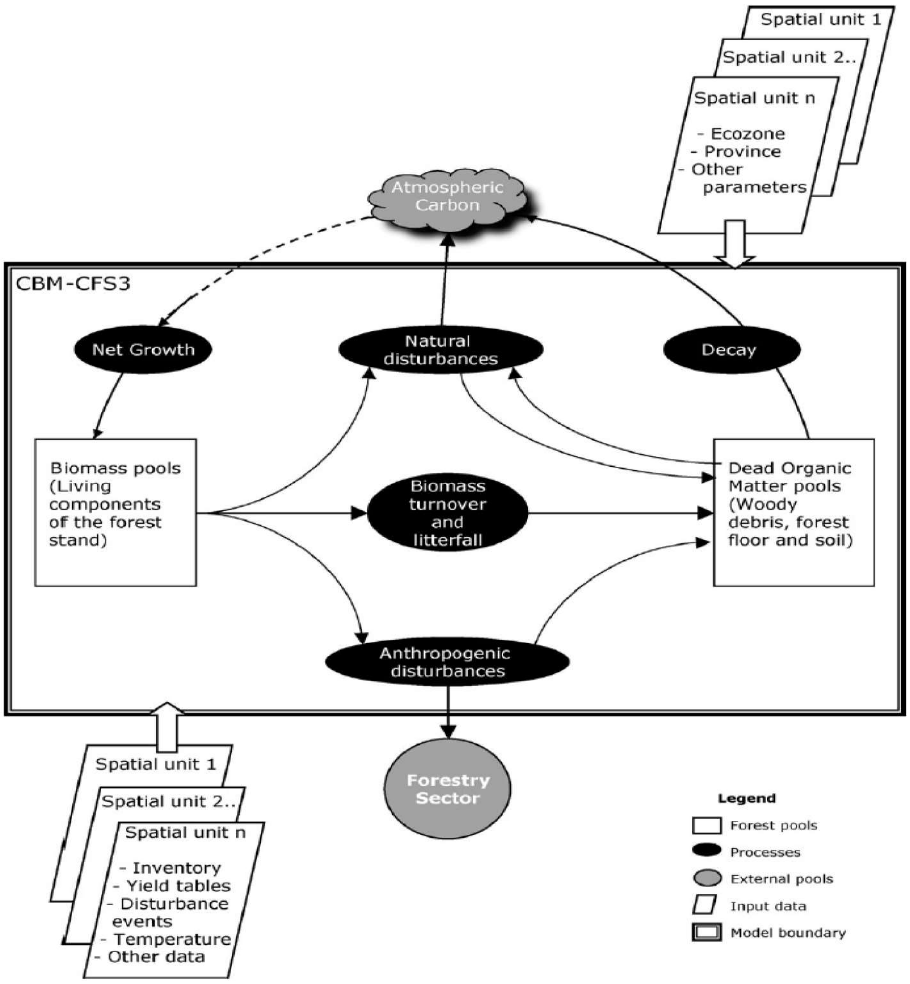

Figure 2.1: CBM General Scheme [4]

Each stand contains in turn a number of carbon pools representing the stems, branches, leaves, death matter etc etc. Given an initial carbon distribution among the different pools, the model calculates all the carbon exchanges among the different carbon pools and between these and the atmosphere as well as the final carbon distribution for each pre-defined *time step*. These carbon fluxes are the result of the natural processes occurring in the forest (leaves fall, death matter decay, CO_2 absorption etc), and other events, such as fires and plagues, but also cuts and thinning for instance. The latter are are called *disturbances* in CBM, each has an identification number and is characterized by a *disturbance matrix* which contains information about the stands to be affected and the carbon fluxes it will produce (for more information see [4])). Note that the absence of *disturbances* is a disturbance itself, with a *disturbance matrix* and an identification, which accounts for forest natural process.

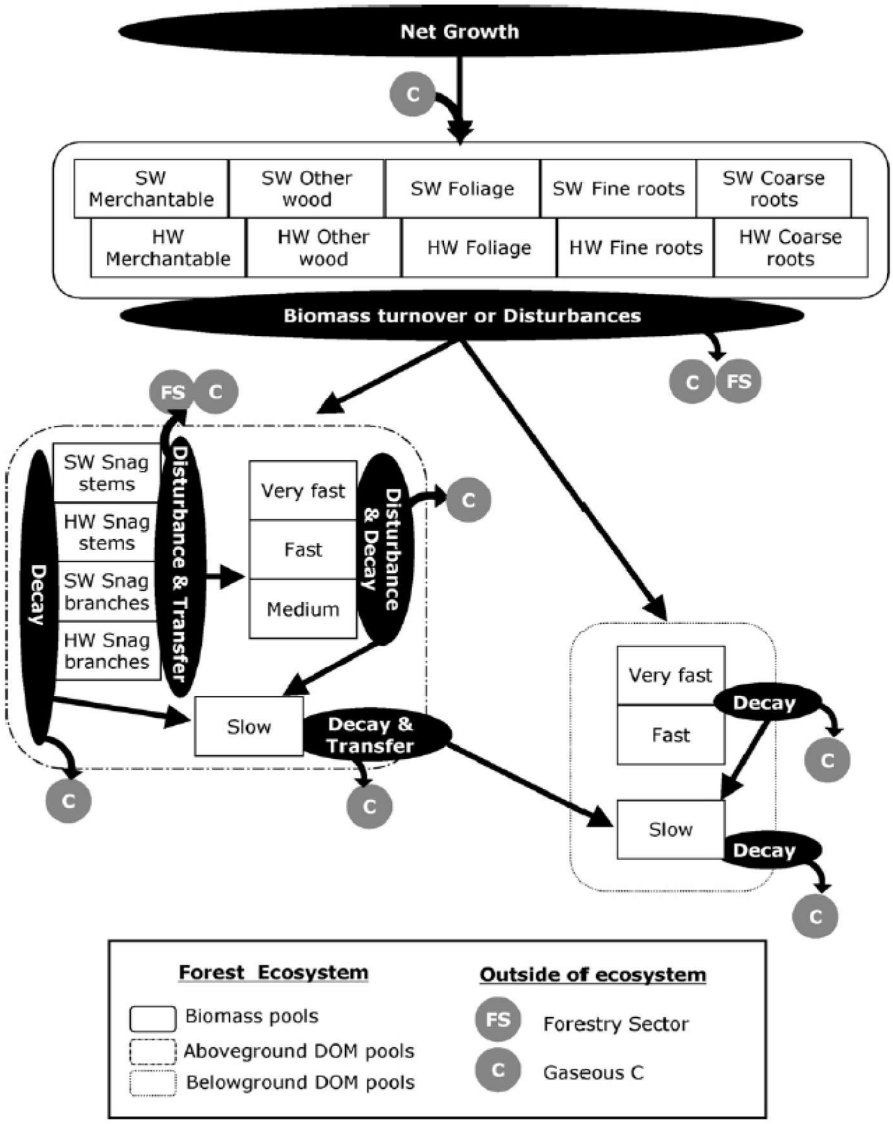

Figure 2.2: Forest Carbon Fluxes [4]

Within the Biomass Study, country-specific CBM runs have been produced for the 2000-2100 period, although the results from 2030 onward shall be neglected. The starting year of these runs depends on the country, thus, for instance 1998 is the starting year for Spain, while 2000 is the one for Slovakia. The time step used is one year for every country.

2.2 The Functions

The CBM package contains four functions developed to be applied on the CBM raw outputs for one specific country and scenario.

2.2.1 *readCBM*

As it was said in section 1.3, this function has the only purpose of extracting the information of interest from the the raw output files. So far CBM′s raw outputs are provided as Microsoft Access Database (.mdb) files (to be stored in the **...\CBM\Outputs** directory, where the function will look for them). Each of them corresponding to a particular run and containing a number of tables, which in turn contain multiple columns.

The function *readCBM* takes as an input the whole name of the file containing the CBM raw outputs to be processed including the extension as in the following example:

```
>readCBM("CBMOutput1.mdb")
```

Reads it, extract the tables below as data.frames and stores them in a list which is in turn saved as a RDS file[2]. Each table is composed by a number of columns with a header, which identifies the data category displayed below:

- `tblUserDefdClassSets`: contains all the information about the stands or (*ClassSets*) that integrate the forest. Of all the columns conforming the table, only *UserDefdClassSetID* and *Name*, which contain respectively the *ClassSets′* identifiers and the set of *classifiers* values characterizing each of them as a string (e.g: "CC,AA,SK00,H,E,34,Con") will be used.
- `tblDisturbanceType`: contains information about the different kinds of disturbances. Only two columns are kept, *DistTypeID* and *DistTypeName* , which are two different sets of identifiers for the disturbances. Through this table the correspondence between both sets is established, so that, regardless which set is being used, the corresponding identifier in the other set can be easily obtained. None of both matches the default naming system described in [4])
- `tblPoolIndicators`: contains the information about the carbon stored in each forest carbon pool for each stand and time step in tonnes of carbon.
- `tblFluxIndicators`: contains the information about all the carbon fluxes occurring in each stand organized per time step and disturbance in tonnes of carbon. This means, that at a certain time step, the same disturbance may produce several carbon fluxes, and they all appear in the same row.

[2] The exact names of the tables in the CBM Output file may change. It is recommended to agree them with the team running the model

- `tblSilvTreatments`: contains the information on the effects of the human-caused disturbances on the different stands depending on their characteristics. This information will be used to classify the removals between Industrial Round-wood and Fuel Wood. Finally, in this table the disturbance identifiers are displayed under the *DistTypeID* column, although they follow the same naming system as in *DistTypeName*. For this reason the column was renamed after the latter.
- `Coefficients`: contains information about the density (DB) and dry-matter carbon content (C) of each tree specie considered. It varies from country to country even for the same specie.

To know more about the meaning of each of the categories cover by the different tables see [4]).

2.2.2 *prepareCBM*

This function transforms the tables extracted from the CBM outputs into objects out of which the selected indicators can be easily calculated. Thus, *prepareCBM* takes the list created by *readCBM* as an input, extracts the tables stored within as data.frames and performs the following operations on them:

```
>prepareCBM("CBMTables.rds")
```

1. Split `Name` strings in `tblUserDefdClassSets` creating one column for each of the classifiers, and names the columns according to the meaning of the classifier. The same names were imposed to the corresponding columns in `tblSilvTreatments`

   ```
   ClassifiersNames<-c("S","FT","R","MT","MS", "CU","C/NC")
   ```
   [3]

2. The new `tblUserDefdClassSets` and `Coefficients` are merged according to the *Forest Type*(FT) classifier, which indicates the main tree specie conforming each stand. The resulting data.frame, displays for each stand, the tree species′ specific values along with the stand′s classifiers values.

   ```
   ClassSets<-merge(x=tblUserDefdClassSets,y=Coefficients,
   by.x="FT",by.y="Species")
   ```

[3] *Status, Forest Type, Region, Management Type, Management Strategy, Climatic Unit, Conifers/Broad-leaves*

3. The `tblFluxIndicators` is merged with `tblDisturbanceType` (according to the *DistTypeID*), then with `ClassSets` (according to the *UserDefdClassSetID*) and finally with `tblSilvTreatment` (according to `MS,FT,MT, C/NC,DistTypeName`). The result is the `CFlux` data.frame, which contains for each time step and disturbance, the carbon fluxes occurring, information about the disturbance that originated it, information about the stands in which they occur and all the information about the classification of the possible removals.

4. In the recently created `CFlux` data.frame the categories *HardProduction, SoftProduction,*CO_2 *Production, and DOMProduction* correspond to the different types of wood that can be extracted from the forest, stem-wood (Non-coniferous and coniferous), branches and other non stem-wood, and snags. A set of eight new categories are added to `CFlux`, by converting these to volume units and group them. This is done using the function `dplyr::mutate`

```
CFlux<-CFlux %>% dplyr::mutate(
Vol_Merch_B=HardProduction/DB/C,
Vol_Merch_C=SoftProduction/DB/C,
Vol_SubMerch=CO2Production/DB/C,
Vol_Snags=DOMProduction/DB/C,
StemWoodRemovals=Vol_Merch_B+Vol_Merch_C,
StemWoodRemovals_C=StemWoodRemovals/DB/C,
OtherWoodRemovals=Vol_SubMerch,
OtherWoodRemovals_C = OtherWoodRemovals/DB/C
)
```

5. `ClassSets` and `CFlux` are saved as *.rds* files in the *CBM/Results* folder

2.2.3 *ForestCStock*

This function takes again the list created in *readCBM* as an input and extracts the data.frames corresponding to the `tblFluxIndicators` and `tblPoolIndicators` tables. Out of them it calculates the amount of carbon stored in each of the forest main carbon pools (as defined in [8, 7]) and the amount of carbon absorbed and emitted to the atmosphere as described in section 1.5 for each time step. To do so, it aggregates multiple of the categories in the input tables representing C-Pools and forest-atmosphere (F-A) C-Fluxes. The first step in this process is creating vectors composed by the names of the categories to be aggregated as follows:

•**<u>Carbon Pools</u>**

- *Hardwood Above Ground Biomass (=HW_Merch+HW_subMerch+HW_Foliage+HW_Other)* and *Softwood Above Ground Biomass(= SW_Merch+SW_subMerch+SW_Foliage+SW_Other).*

```
AGB_HW<-paste("HW",c("Merch","subMerch","Foliage","Other"), sep="_-
")
AGB_SW<-paste("SW",c("Merch","subMerch","Foliage","Other"), sep="_-
")
AGB<-c(AGB_HW,AGB_SW)
```

- *Hardwood Below Ground Biomass (=HW_ Merch+HW_ subMerch+HW_-Foliage+HW_ Other)* and *Softwood Below Ground Biomass(= SW_ Merch+SW_-subMerch+SW_ Foliage+SW_ Other).*

```
BGB_HW<-paste("HW",c("Coarse","Fine"),sep="_")
BGB_SW<-paste("SW",c("Coarse","Fine"),sep="_")
BGB<-c(BGB_HW,BGB_SW)
```

- *Living Biomass(LB)=Above Ground Biomass(ABG)+Below Ground Biomass(BGB)*

```
LB<-c(AGB,BGB)
```

- *Death Hardwood(=HW_ StemSnag+HW_ BranchSnag)* and *Death Softwood(=SW_ StemSnag+SW_ BranchSnag).*

```
DW_HW<-paste("HW",c("StemSnag","BranchSnag"),sep="_")
DW_SW<-paste("SW",c("StemSnag","BranchSnag"),sep="_")
DW<-c(DW_HW,DW_SW,"FastBG","Medium")
```

- *Litter=VFast_ AG, Fast_ AG, Slow_ AB*

```
Litter<-paste(c("Fast","Slow","VFast"),"AG",sep="_")
```

- *Death Organic Matter (DOM)=Death Wood (DW) + Litter*

```
DOM<-c(DW,Litter)
```

- *Soil Organic Matter (SOM)= VFast_ BG+Slow_ BG*

```
SOM<-paste(c("Slow","VFast"),"BG",sep="_")
```

Then the values in the columns referred by the vectors *LB,DOM* and *SOM* are summed up:

```
MainPools<-tblPoolIndicators %>% dplyr::mutate(
LB=rowSums(select(.,LB)),
AGB=rowSums(select(.,AGB)),
DOM=rowSums(select(.,DOM)),
SOM=rowSums(select(.,SOM))
)
```

Note that AGB was also explicitly calculated, for it will be used later on,separated from the other living biomass pools.

•**Carbon Fluxes**

C-fluxes will be aggregated according to two different criteria: their source(either living or death biomass) and the kind of substance emitted (CO,CO_2 or CH_4).

- *Emissions from Living Biomass (=Bio_ CO2Emission+Bio_ COEmission+Bio_-CH4Emission)*

  ```
  Bio_Em<-paste("Bio",c("CO2Emission","COEmission","CH4Emission"),
  sep="_")
  ```

- *Death Organic Matter Emissions(=DOM_ CO2Emission+DOM_ COEmission+DOM_ CH4Emission)*

  ```
  DOM_Em<-paste("DOM",c("CO2Emission","COEmission","CH4Emission"),
  sep="_")
  ```

- Total CO_2, CO and CH_4 Emissions

  ```
  CO_Em<-paste(c("DOM","Bio"),"COEmission",sep="_")
  CO2_Em<-paste(c("DOM","Bio"),"CO2Emission",sep="_")
  CH4_Em<-paste(c("DOM","Bio"),"CH4Emission",sep="_")
  ```

  Then, in analogy to what happened with the C-Pools the values in the columns referred by the vectors *BioEm, DOMEm, COEm, CO2Em* and *CH4Em* are summed up.

  ```
  ForestEm<-tblFluxIndicators %>% dplyr::mutate(
  CO2=rowSums(select(.,CO2_Em)),
  ```

```
CO=rowSums(select(.,CO_Em)),
CH4=rowSums(select(.,CH4_Em)),
CO2Tot_Em=CO2*(44/12)+(CH4*(16/12)*28),
CTot_Em=CO2+CO+CH4,
CBioTot=rowSums(select(.,Bio_Em)),
CDOMTot=rowSums(select(.,DOM_Em)),
CO2BioTot=rowSums(select(.,Bio_Em)*c(44/12,0,(16/12)*28)),
CO2DOMTot=rowSums(select(.,DOM_Em)*c(44/12,0,(16/12)*28))
)
```

Note that in CBM, C-Fluxes are given in tonnes of carbon. Therefore, in order to calculate the amount of CO_2 absorbed and emitted by the forest, these quantities have to be converted to CO_2 or CO_2 equivalent, as in the case of CH_4. [4].

- *C-Absorbed(=GrossGrowth_AG+GrossGrowth_BG)*

```
CAbs<-c("GrossGrowth_AG","GrossGrowth_BG")
```

Two new categories resulting of aggregating the categories included in *CAbs* are added to the `FEmissions` data.frame:

```
ForestEm<-ForestEm %>% dplyr::mutate (
CTot_Abs=rowSums(select(.,CAbs)),
Co2Tot_Abs=rowSums(select(.,CAbs))*(44/12)
)
```

The result of these operations is two data.frames `ForestEm` and `MainPools`, which respectively contain all the information in `tblFluxIndicators` and `tblPoolIndicators`, plus a set of new categories which account for the carbon fluxes between the forest and the atmosphere and the forest main carbon pools.

Finally, the function sums up all the C-Stocks in the main pools and F-A C-Fluxes by *TimeStep*, producing two new data.frames, with the same categories and as many rows as time steps there are. Then the two new data.frames will be merged matching the respective *TimeStep* columns. The results is the *Forest Carbon Stock and Fluxes* per time step (`FCSF_TS`) data.frame in which each row displays the sum of all the pools and fluxes at a particular *TimeStep*.

```
ForestEm_TS %>% dplyr::group_by(TimeStep) %>% dplyr::summarise_all(sum)
```

[4] https://www.ipcc.ch/publications_and_data/ar4/wg1/en/ch2s2-10-2.html

```
MainPools_TS<-MainPools %>% dplyr::group_by(TimeStep) %>%
dplyr::summarise_all(funs(sum))
FCSF_TS<-merge(MainPools_TS,ForestEm_TS,by=TimeStep)
```

All the new data.frames are saved as .rds files in the *CBM/Results* folder.

2.2.4 *ForestHWP*

The last function in the CBM packages takes `CFlux` and `FCSF_TS` as inputs and allocates the wood removed from the forest in volume (*Vol_Merch, Vol_SubMerch* and *Vol_Snags*) to one of the Primary Forest Products (PFP) types, Industrial Round-wood and Fuel-wood (for both coniferous and non-coniferous). This allocation is carried out according to a general rule (see Table 2.1) which allocates the removals in each `CFlux` row to the different PFP categories according to category *HWP* value.

HWP	Industrial Round-wood	Fuel-wood
IRW	*Vol_Merch*	*Vol_SubMerch+Vol_Snags*
FW	-	*Vol_Merch+Vol_SubMerch+Vol_Snags*

Table 2.1: Wood removals general PFP allocation rule

However, depending on the country and the period, there may be exceptions to this rule. For instance, for Slovakia and other countries, between 2000 and 2030, two exceptions to this rule will have to be introduced. First, if *DistTypeID*=3 then *Vol_SubMerch*=0 regardless the value of *HWP* [5]. And if *DistTypeID*=9 and *HWP*= IRW_C or IRW_B, *Vol_Snags* will be allocated to *IRWc* or *IRWnc* instead of to *FWc* or *FWnc* [6]. The first exception can be applied using the following command:

```
CFlux[CFlux$DistTypeID==3,]["Vol_SubMerch"]<-0
```

Like in `ForestCStock`, the first step in this process is creating and filling the new categories, which correspond to the four PFP types *IRWc, FWc, IRWnc* and *FWnc*, total Industrial Round-wood (*IRW*), total Fuel-wood (*FW*) and total removals (*TotRemovals*).

[5]Definition in [4]): *Commercial thinning of merchantable trees resulting in a 10%–75% reduction in biomass carbon (amount chosen by the user) and transfer of the merchantable carbon to the forest products sector.The remaining carbon removed from the biomass pools is transferred to the respective DOM pools.*

[6]Definitions in [4]):*A generic disturbance causing a 5–95% reduction or mortality in biomass carbon (amount chosen by the user), which is transferred from the biomass pool to respective DOM pools.*

Besides, a few extra categories are created to account for the origin of the removals, either *stem-wood* (StW) or *other wood*(OW) [7]. In order to implement all these changes the `dplyr::mutate` was applied on `CFlux` as follows:

- First a set of 4 auxiliary categories, *Ic, Inc, Fc* and *Fnc* are created in `CFlux`. They take the values `TRUE` or `FALSE` depending on the row′s value for the *HWP* parameter (respectively *IRW_ C, IRW_ B, FW_ C* or *FW_-B*). Besides, an additional auxiliary category *DT9* is created to introduce the second exception to the general rule explained before. *DT9* is `TRUE` when *DistTypeID*=9 and `FALSE` if not.

```
HaWoPr<- CFlux %>% dplyr::mutate(
Ic=grepl("IRW_C",HWP),
Inc=grepl("IRW_B",HWP),
Fc=grepl("FW_C",HWP),
Fnc=grepl("FW_B",HWP),
```

- Then, the just created auxiliary categories are used to allocated the wood removals to the right PFP category according to the general rule, but also taking into account the second exception.

```
IRWc=Ic*((!DT9)*Vol_Merch_C+DT9*(Vol_Merch_C+Vol_Snags)),
IRWnc=Inc*((!DT9)*Vol_Merch_B+DT9*(Vol_Merch_B+Vol_Snags)),
IRW=IRWc+IRWnc,
FWc=Ic*(Vol_SubMerch+(!DT9)*Vol_Snags)+Fc*(Vol_Merch_C+
Vol_SubMerch+Vol_Snags),
FWnc=Inc*(Vol_SubMerch+(!DT9)*Vol_Snags)+Fnc*(Vol_Merch_B+Vol_-
SubMerch+Vol_Snags),
FW=FWc+FWnc,
TotRemovals=IRW+FW,
```

- Finally, the extra categories which should account for the origin of the wood removals, and therefore of the PFP are created and filled.

```
IRWc_StW=Ic*Vol_Merch_C,
IRWc_OW=Ic*DT9*Vol_Snags,
IRWnc_StW=Inc*Vol_Merch_B,
IRWnc_OW=Inc*DT9*Vol_Snags,
IRW_StW=IRWc_StW+IRWnc_StW,
IRW_OW=IRWc_OW+IRWnc_OW,
FWc_StW=Fc*Vol_Merch_C,
FWc_OW=Ic*(Vol_SubMerch+(!DT9)*Vol_Snags)+
Fc*(Vol_SubMerch+Vol_Snags),
FWnc_StW=Fnc*Vol_Merch_B,
```

[7] Note that, as it was said in section 1.6, FAO nomenclature is applied, that is coniferous instead of Softwood and non-coniferous instead of broad leaves (B) and Hardwood

```
FWnc_OW=Inc*(Vol_SubMerch+(!DT9)*Vol_Snags)
+Fnc*(Vol_SubMerch+Vol_Snags),
FW_StW=FWc_StW+FWnc_StW,
FW_OW=FWc_OW+FWnc_OW,
TotRemovals_OW=IRW_OW+FW_OW,
TotRemovals_StW=FW_StW+IRW_StW)
```

The result is a new data.frame `HaWoPr`, that contains all the categories included in `CFlux` plus the new ones. Grouping it by *TimeStep*, and combining it with the previously created `FCSF.TS` data.frame, a number of new magnitudes which could provide more insight into the forest sectors and the effects of the different policies can be easily calculated. For instance, the wood removals-wood available ratio, which has been traditionally used as an indicator of management intensity, defined as follows:

$$Man.Intensity(i) = \frac{Wood\, removed\, from\, the\, managed\, area\, in\, \mathrm{i}}{Wood\, available\, in\, the\, managed\, areaat\, \mathrm{i}} \tag{2.1}$$

where i is the time step, the numerator would correspond to `HaWoPr` *TotRemovals* category and the denominator would be given by AGB in `FCSF.TS`. Below you may find the code lines that would carry out this calculations:

```
HaWoPr_TS<-HaWoPr %>% dplyr::group_by(TimeStep) %>% dplyr::summarise_-
all(funs(sum))

Comb_TS<-merge(HaWoPr_TS,FCSF_TS,by="TimeStep")

Comb_TS[,"Removals/Available"]<-Comb_TS[,"TotRemovals"]/Comb_TS[,"AGB"]
```

2.3 Representation

Finally the categories contained in the created data.frames can be represented in graphs (see Fig. 2.3 and Fig. 2.4) ,using the different possibilities R offers, such as the function `ggplot2::ggplot` function, which will have to be customized for each case (see the example below), or in a map (see Fig. 2.5) if an univocal correspondence between *SPUID* or *tblUserDefdClassSets* and a geographical unit system cab be established (e.g.NUTS1, NUTS2 or NUTS3[8]) using GIS (e.g. QGIS[9]).

```
graphsCBM<-function(Data,objects,period){

   setwd(paste(path,"CBM/Results",sep="/"))

   Data$TimeStep<-c(2000:2100)

   Data<-Data[(Data$TimeStep %in% period),c("TimeStep",objects)]

   Data<-cbind(TimeStep=period,Data)

   Data<-reshape::melt(Data,id.vars="TimeStep")

   ggplot(Data,aes(x=TimeStep,y=value,color=variable))+geom_line(size=1.5)+
labs(x="Years",y="C-Stock[tonnes]")+
theme(legend.title=element_text(size=8,face="bold"),legend.text=element_-
text(size=6),axis.text=element_text(size=6), axis.title=element_text(size=8,face="bold"))+
scale_color_hue(name = "C-Pool")+
scale_x_continuous(breaks = seq(from=2000,to=2030,by=5))+
scale_y_continuous(breaks = seq(from=0, to=3.0e+09,by=5.0e+08))
}
```

[8] Geographical units used by Eurostat:https://ec.europa.eu/eurostat/web/nuts/background

[9] https://www.qgis.org/en/site/

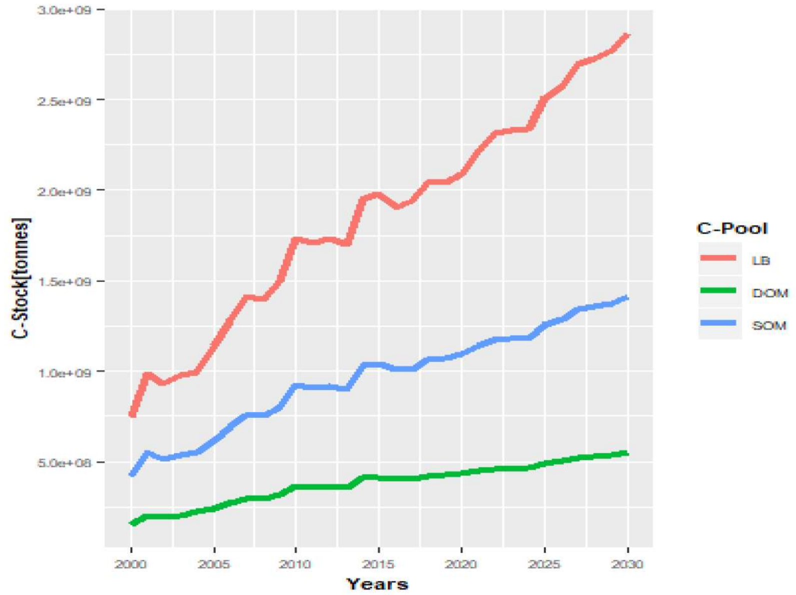

Figure 2.3: Forest carbon pools in Slovakia

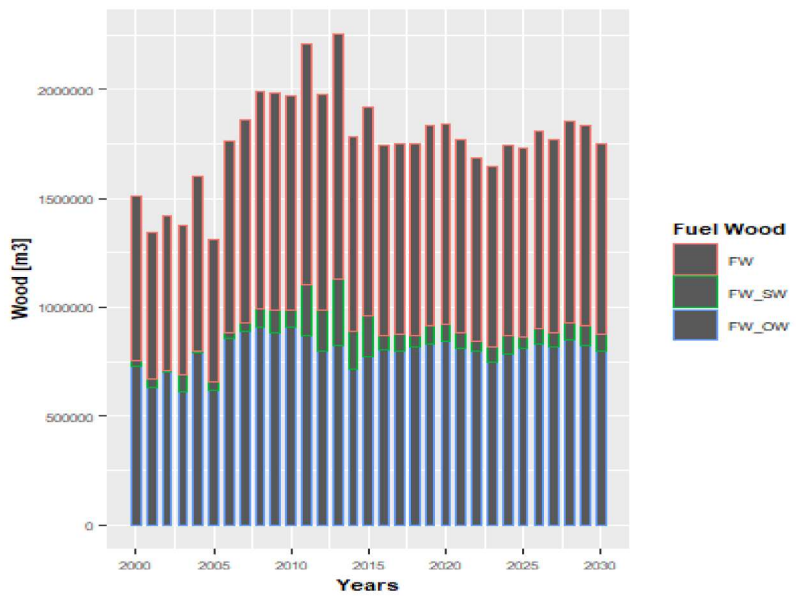

Figure 2.4: Fuel Wood by origin in Slovakia

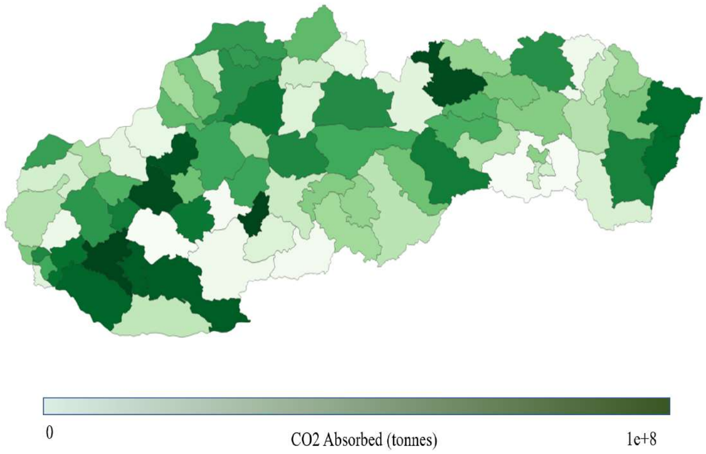

Figure 2.5: Absorbed CO_2 by Spatial Unit in 2020

Chapter 3

FAOSTAT Package

3.1 The Data Base

FAOSTAT is FAO′s statistic service[1] which makes available data on the agriculture and forestry sectors, from as early as 1961, although this depends on country and item. This data will be used to calculate the historical value of the indicators of interest (prior to 2015), and complement the outputs of the IMF models when necessary. FAOSTAT periodically updates the information available to be downloaded from its website. In order to guarantee coherence with other projects within the JRC, and in particular within the Biomass Study, the FAO Yearbook of Forest Products 2016 (released on the 20^{th} of December 2017), is taken as standard data source for the FAOSTAT Package[2]. These releases can be downloaded as *.csv* files in which the information is displayed as follows:

Area Code	Area	Item Code	Item	Element Code	
11	*Austria*	*1873*	*Wood-based panels*	*5516*	
...	...	...	...	...	
Element	**Year Code**	**Year**	**Unit**	**Value**	**Flag**
Production	*1989*	*1989*	*m3*	*1633000*	*A*
...	...	...	...	...	

Table 3.1: FAOASTAT data

[1] `http://www.fao.org/faostat/en/#home`
[2] Note that this may change in the future

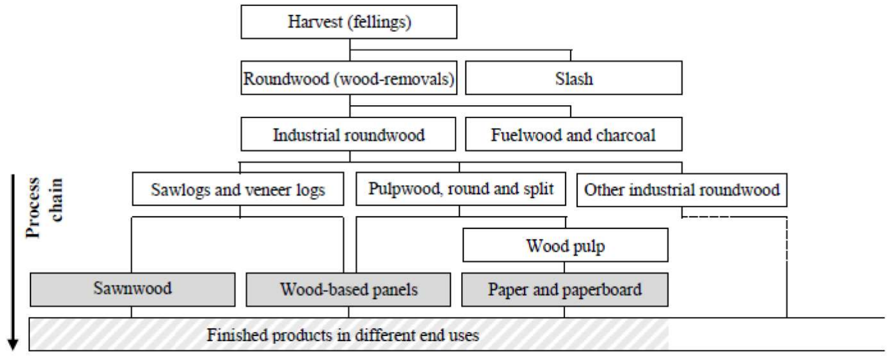

Figure 3.1: Simplified FAO forest wood products classification [8]

3.2 The Functions

The FAOSTAT package contains four functions, which, unless the reference FAOSTAT release is changed, only have to be run once, for their goal is the creation of auxiliary files that complement or complete the models′ outputs. The application of these functions requires, as a previous step, the definition of the items (e.g. Industrial Round-wood, Veneers sheets etc) and elements (e.g. Production, Exports etc) of interest. To do so two vectors, each of them containing the names of the elements and items of interest, should be created and saved in the **...\General Parameters** directory.

```
MainItemsFAO<-c("Industrial roundwood, coniferous","Wood pulp",
"Industrial roundwood, non-coniferous" ,"Sawnwood", "Plywood",
"Sawnwood, coniferous","Sawnwood, non-coniferous all","Veneer sheets",
"Wood-based panels","Paper and paperboard")

MainElementsFAO<-c("Production","Import Quantity","Export Quantity")
```

3.2.1 *readFAOSTAT*

Following the structure presented in 1.3, this function looks for the downloaded FAOSTAT file (usually a *.csv*) in the **...\FAO\Outputs** directory) and extracts the information in the shape of a data.frame with as many columns or categories as table 3.1. Then it performs the following operations:

1. Selects the categories of interest:

```
FAOData<-FAOData[c("Area","Element","Item","Unit","Year","Value")]
```

2. Selects the objects of interest within the categories *Item* and *Element*:

   ```
   FAOData<-dplyr::filter(FAOData,Element %in% ME, Item %in% MI)
   ```

3. Reshapes the data.frame, turning the objects in years into columns so that, each row contains all the values available for a particular set of country, item and element

   ```
   FAOData<-reshape::cast(FAOData,Area+Item+Element+Unit Year)
   ```

In addition to this, since FAOSTAT does not consider "Rest of the World" (ReWo) and "Rest of Latin America" (ReLaAm) regions, which GFTM does, the present function calculates the figures of the different elements and items for these two areas by adding up those of the countries conforming them[3]:

```
FAODataReWo<-subset(FAOData,(Area %in% ReWo))

FAODataReLaAm<-subset(FAOData,(Area %in% ReLaAm))

FAODataReWo[is.na(FAODataReWo)]<-0

FAODataReLaAm[is.na(FAODataReLaAm)]<-0

ReLaAmSum<-FAODataReLaAm[c("Object.Code",years)] %>% group_by(Object.Code)
%>% summarise_all(funs(sum))

ReLaAmSum2<-merge(FAODataGer[c("Item","Item.Code","Element","Element.Code",
"Object.Code","Unit")],ReLaAmSum,by="Object.Code")

ReLaAmSum2$Area<-"ReLaAm"

ReWoSum<-FAODataReWo[c("Object.Code",years)] %>% group_by(Object.Code)
%>% summarise_all(funs(sum))

ReWoSum2<-merge(FAODataGer[c("Item","Item.Code","Element","Element.Code",
"Object.Code","Unit")],ReWoSum,by="Object.Code")

ReWoSum2$Area<-"ReWo"

FAOData<-rbind(FAOData,ReWoSum2,ReLaAmSum2)
```

[3]The vectors containing them, `ReWo` and `ReLaAm` are to be found in the **...\General Parameters** directory.

Note, that, in order to do so, the NA values, very common specially in the early years of the period covered by the FAO, were substituted by 0. This prevent the results of the sums to be NA (which is the case if just one of the values is NA), but leads to a result that does not take into account the contribution from those countries for which the correspondent value was missing at that time. This can of course be changed at any time.

After running the function, the first columns of the resulting data.frame may look like this:

Area	**Element**	**Item**	**Unit**	**1961**	**1962**	**...**
Germany	*Production*	*Plywood*	*m3*	*661000*	*677800*	...

Table 3.2: *readFAOSTAT* resulting data.frame

3.2.2 *prepareFAOSTAT*

This function takes the data.frame resulting from the application of *readFAOSTAT* to the file downloaded from FAOSTAT (called `FAOData0` here) as input and carries out a number of modifications on the names used in the data.frame that will make it easier to handle.

For instance, it substitutes the empty spaces and remove the commas in elements, items and country names,

```
FAOData<-FAOData %>%
dplyr::mutate_at(.funs=funs(gsub(" ","-",.)),
.vars=c("Area","Item","Element"))
%>%
dplyr::mutate_at(.funs=funs(gsub(",","",.)),
.vars=c("Area","Item","Element"))
```

,replaces long and cumbersome names by abbreviations like *IRWc* instead of *Industrial-roundwood-coniferous* or completes the FAOSTAT data by creating NA-filled rows for those items for which FAOSTAT does not provide information on specific elements and countries[4].

```
FAOData<-tidyr::complete(FAOData,crossing(Area,Item,Element))
FAOData1<-dplyr::filter(FAOData,(Item %in%
```

[4]This is done by calculating all the possible combinations between the existing *Area,Item* and *Elements*, and adding one NA-filled row for those missing. This expansion takes *m3* as default unit and therefore this has to be modified for those cases in which the unit is other

```
c("Wood-pulp","Paper-and-paperboard"))) %>% dplyr::mutate(Unit="tonnes")
FAOData2<-dplyr::filter(FAOData,!(Item %in%
c("Wood-pulp","Paper-and-paperboard"))) %>% dplyr::mutate(Unit="m3")
FAOData<-rbind(FAOData1,FAOData2)
```

In addition to this, the function also calculates the *Consumption* and *Trade* figures for every *Area* and *Item*, and for all the years considered. Finally, out of the original, it creates four different data.frames which contain the *Production, Consumption, Exports* and *Imports* figures respectively.

```
FAODataExp<-FAOData[FAOData$Element=="Export-Quantity",]

FAODataImp<-FAOData[FAOData$Element=="Import-Quantity",]

FAODataPro<-FAOData[FAOData$Element=="Production",]

FAODataTrade<-cbind(FAODataImp[c(1:4)],(FAODataExp[Years]-FAODataImp[Years]))

FAODataCon<-cbind(FAODataPro[c(1:4)],
(FAODataPro[Years]-FAODataTrade[Years]))

FAODataTrade$Element<-factor("Trade")

FAODataCon$Element<-factor("Consumption")
```

The outputs of *prepareFAOSTAT* are the modified data.frame still containing all FAOSTAT data (`FAOData1` here), and four data.frames containing only the data on consumption, imports, exports and production (here `FAODataCon`, `FAODataImp`,`FAODataExp` and `FAODataPro` respectively).

3.2.3 *FAOCoeff1*

In GFTM the wood-based panels (*Wbp*) category does not include veneer sheets (*VS*). However FAO wood-based panels does, and so does to the IPCC in its definition of the semi-finished wood products to be considered when calculating the harvested wood products C-Stock [8]. For FAO and GFTM data to be comparable and to apply the IPCC methodology on GFTM data, *Wbp* category has to include *VS's* contribution. The present function takes *prepareFAOSTAT* output containing all the information (`FAOData1` here) to calculate the historical mean *VS-Wbp* ratio, which will be used to estimate the missing *VS* contribution to *Wbp* in the GFTM outputs according to the following expression[5]:

[5] A different method, using the Veneers sheets-Plywood ratio instead was also considered. However this ratio proved to be less stable than the *VS-Wbp* ratio, was thus discarded.

$$Wbp_{GFTM_0} = Fiberboard_{GFTM} + Particleboard_{GFTM} + Plywood_{GFTM} \tag{3.1}$$

$$Wbp_{GFTM} = Wbp_{GFTM_0} * (1 + VS/Wbp_{FAO}) \tag{3.2}$$

What follows are the code lines that make the calculation. Here, `period` comprehends the 2010-2015 period, however, this can of course be modified anytime.

```
FAODataVeneer<-FAOData[FAOData$Item=="Veneer-sheets",]

FAODataWbp<-FAOData[FAOData$Item=="Wood-based-panels",]

VenWbp<-merge(FAODataWbp,FAODataVeneer,by=c("Element","Area","Unit"))

Coeff<-rowMeans(VenWbp[paste(period,".y",sep="")]
/VenWbp[paste(period,".x",sep="")])

Coeff[is.na(Coeff)]<-0

VenWbpCoeff<-cbind(FAODataWbp[c(1:4)],Coeff)
```

The output of this function is a data.frame `VenWbpCoeff`, with a value for the *VS-Wbp* ratio for each country and element.

3.2.4 *FAOCoeff2*

GFTM does not produce exports and imports figures, but only trade figures. For further calculations it is of interest to know the values of both magnitudes for each country and item. *FAOCoeff2* takes all five *prepareGFTM* outputs as inputs and out of them calculates the Exports-Production and the Imports-Consumption ratios for every country and item, and selects which of them is more reliable and therefore more suitable to be used on GFTM raw outputs. The calculation of GFTM imports and exports figures for a particular country and item will be carried out in two different ways depending on the ratio used:

$$Apparent\ Consumption = Production - Trade \tag{3.3}$$

$$Trade = Exports - Imports \implies \begin{cases} Exports = Trade + Imports \\ Imports = Exports - Trade \end{cases} \tag{3.4}$$

$$Exports = Trade + Consumption * \left(\frac{Imp}{Cons}\right)_{FAO} \tag{3.5}$$

or

$$Imports = Production * \left(\frac{Exp}{Prod}\right)_{FAO} - Trade \tag{3.6}$$

The first step is calculating the mean ratios for each country and item in the 2010-2015 period[6]. This requires to create beforehand two lists: a list (here `FAODataSep`) that has as objects the four data.frames resulting from *prepareFAOSTAT* containing the FAOSTAT data separated by *Element*and a second list, here `ExpImpCoeff`, with three empty data.frames where the Exports-Production, Production-Consumption and Imports-Consumption ratios will be stored as they are created.

```
for (i in (1:3)){
mix<-merge(FAODataSep[[i]],FAODataSep[[i+1]],by=c("Item","Area"))
if (names(FAODataSep)[i]=="Con"){
mix<-merge(FAODataSep[[i+1]],FAODataSep[[i]],by=c("Item","Area"))}
mix[period]<-mix[paste(period,".x",sep="")]/mix[paste(period,".y",sep="")]
mix[mix==Inf|is.na(mix)]<-0
mix["Mean"]<-apply(mix[period],1,mean,na.rm=TRUE)
mix["SD"]<-apply(mix[period],1,sd,na.rm=TRUE)
mix["Var.Coeff"]<-mix["SD"]/mix["Mean"]
mix[is.na(mix)]<-0
ExpImpCoeff[[i]]<-mix[c("Area","Item","Mean","SD","Var.Coeff")]}

ExpImpCoeff<-do.call(cbind,ExpImpCoeff)
```

The resulting data.frame `ExpImpCoeff` displays for each country the mean Exports-Production, Production-Consumption and Imports-Consumption ratios, with their corresponding standard deviations and variation coefficients.

The next step is selecting among the ratios calculated which is the one to be use on GFTM Outputs to calculate the Exports and Imports figures. The criteria to do so can be resumed as follows[7]:

If $\frac{Production}{Consumption}$>1.05–>*Export/Production* ratio

If $\frac{Production}{Consumption}$<0.9–>*Imports/Consumption* ratio

[6] The election of this period of time responds to the requirement of using updated ratios, which value can be applied on GFTM predictions for the future. Events like economic crisis have a big influence on trade figures, and therefore it is not recommended to use pre-event data to treat post-event predictions.

[7] This criteria was agreed with the responsible of running the GFTM model

If $0.9 \leq \frac{Production}{Consumption} \leq 1.05 -> \textit{The ratio with the lowest variation coefficient}$

This criteria was introduced in the function through the following lines of code:

```
ExpImpCoeff<-ExpImpCoeff %>% dplyr::mutate(
A=(ProCon.Mean>1.05),
B=(ProCon.Mean<0.9),
C=(ProCon.Mean<=1.05 & ProCon.Mean>=0.9),
D=(ExpPro.Var.Coeff>=ImpCon.Var.Coeff),
Coeff1=(A*ExpPro.Mean)+(B*ImpCon.Mean)+(C*((D*ImpCon.Mean)+
((!D)*ExpPro.Mean))),
Coeff2=(Coeff1==ExpPro.Mean)*ImpCon.Mean+(!(Coeff1==ExpPro.Mean))*ExpPro.Mean,
Crit1=(Coeff1==ExpPro.Mean))
```

The result is a data.frame, here `ExpImpCoeff`, with five columns or categories: *Area,Item,Coeff1,Coeff2* and *Crit1*. Category *Coeff1* contains the value of the ratio to be used, either Exp-Pro or Imp-Cons, according to the criteria described above for each country and item. Category *Coeff2* displays the alternative ratio, and category *Crit1* is *TRUE* when the ratio to be used for that particular country-item couple is Exp-Pro and *FALSE* when it is Imp-Cons.

Chapter 4

GFTM

4.1 The Model

GFTM, the *Global Forest Trade Model* is an economic model of the global forest-based sector[5]. It is an equilibrium trade-based model for the forest sector with provides projections on production and trade of ten final products[1], four intermediate products[2] and four primary products [3] for 48 countries and sub-regions of the world.

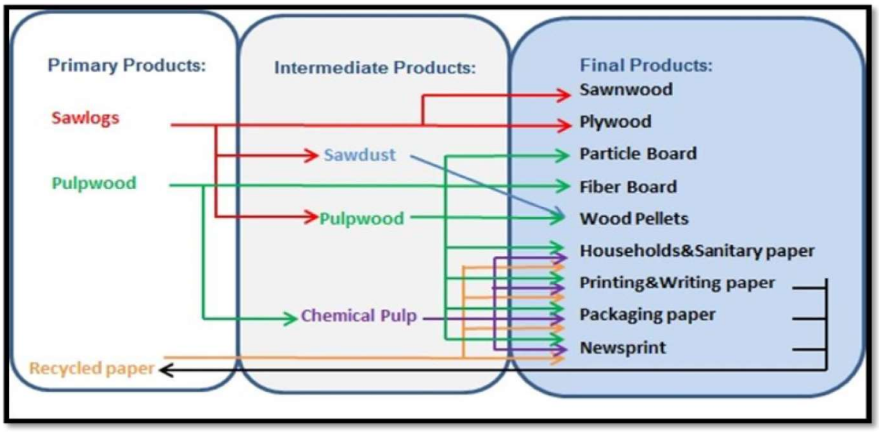

Figure 4.1: Forest industry transformation process in GFTM [5]

[1] sawn-wood coniferous and non-coniferous, plywood, particle board, fibreboard, wood pellets, household & sanitary paper, printing & writing paper, newsprint, packaging paper

[2] chemical pulp, recovered paper, coniferous and non-coniferous sawdust

[3] coniferous and non-coniferous sawlogs and pulpwood

Each GFTM run produces a MATLAB file (*.mat*) containing a 43x23 matrix with the predicted figures for 43 item-element pairs in 23 different countries or sub-regions[4] under a particular scenario (see Annex for the item-element pairs covered). The data contained in each output is considered to be valid for a five years period. In order to make it easier to manage the raw output files, and the application of the functions, a naming system has been established. So, for instance, the file containing the predictions for the period 2015-2019 under scenario *S0* will be renamed as *Y2015S0.mat*.

4.2 The Functions

Unlike in the previous packages, in this case, since the modifications to be implemented on the data.frame resulting from *readGFTM* were several, instead of just one *prepare* function, it was decided to create a specific function to perform each modification, so they can be carried out separately and the intermediate results can be analyzed. To illustrate the evolution of the intermediate output names, the GFTM output for the 2015-2019 period under S0 scenario *Y2015S0.mat* will be used as an example through the next sections. Finally all the *prepare* functions are written to take the GFTM raw output file name without extension as input, following the example just presented this would be *Y2015S0*.

4.2.1 *readGFTM*

As its equivalent functions in the CBM and FAOSTAT packages, this function takes GFTM raw output files as inputs, extracts the data and saves it in a more manageable form, a data.frame in this case. In addition to this it also assigns column names, that corresponds to the countries and sub-regions covered by GFTM, and adds three new columns, *Item, Element* and *Unit* that provide the meaning for the values in the rows. The column names (`AC`) and the three extra columns have been previously defined and saved in **...\General Parameters**[5].

Item	Element	Unit	Austria	Belarus	...	...
Sawnwood C	*Production*	*m_3*	*9,558,528*	*2,561,784*	...	...
Sawnwood C	*Trade*	*m_3*	...	...	...	...

Table 4.1: *readGFTM* resulting data.frame

[4] As indicated in 1.6 FAO′ country names have been adopted. As for the sub-regions, such as Northern Africa or Latin America, the equivalence between FAO and GFTM definitions was checked, with the GFTM responsible

[5] The exact correspondence between the raw data columns and rows and the country and item-element-unit names was provided by the GFTM responsible. It is recommended to check with them whether these correspondence still holds or not

Since this process will have to be carried out on GFTM output files for multiple periods and scenarios, it is recommended to name the intermediate outputs so that it can be easily understood which data they contain. Here, for instance if the GFTM raw file was *Y2015S0.mat*, *readGFTM* input would be *Y2015S0* and the resulting data.frame will be named `GFTM1_Y2015S0`.

4.2.2 The *prepareGFTM* functions

prepareGFTM1

GFTM outputs do not include all the elements for all the items considered. For instance, primary products (Saw-logs and Pulpwood) consumption is missing, and for the recovered paper only trade is shown. As it was explained in 3.2.2, for further calculations it is convenient to have all the elements of all the items covered. In this respect, the role of *prepareGFTM1* is, departing from *readGFTM* outputs, first calculate the primary products consumption out of the trade and production figures.

```
pp<-expand.grid(c("Sawlogs","Pulpwood"),c("-coniferous","-non-coniferous"))

AddItems<-c(paste(pp[,1],pp[,2],sep=""),"Wood-pulp")

for(i in AddItems){
D<-GFTMData[!(GFTMData$Element=="Consumption" & GFTMData$Item==i),]
Prod<-dplyr::filter(D,Element == "Production" & Item == i)
Trad<-dplyr::filter(D,Element == "Trade"& Item == i)
new<-cbind((Prod[AC]-Trad[AC]),Item=i,Element="Consumption",Unit=Prod["Unit"])
GFTMData<-rbind(D,new)
}
```

Then, introduce two NA-filled rows representing recovered paper production and consumption. Like in 3.2.2 this is made by calculating all the possible combinations of the *Elements* and *Items* columns and adding those that were missing to the work data.frame.

```
GFTMData<-tidyr::complete(GFTMData,crossing(Item,Element))

GFTMData[GFTMData$Item=="Recovered-paper","Unit"]<-"tonnes"
```

prepareGFTM2

The semi-finished forest wood products (SFWP) considered in [8] to calculate the Harvested Wood Products C-Stock are not considered as such by GFTM. However their values can be calculated by adding up those of their components (see below).This second *prepare* function takes *prepareGFTM1* output, which following the example given above it would be GFTM2_Y2015S0, and calculates all the SFWP elements.

SFWP=Wood-based-panels, Paper-and-paperboard and Sawnwood

Wood-based panels(Wbp)=*Fibreboard*(FB)+*Plywood*(Plw)+*Particleboard*(PB)[6]

Paper and Paperboard(P&PB)=*Newsprint*+*Printing & Writing paper*(P&W)+ *Household & Sanitary paper*(HH$Sanitary)+*Other Paper & Paperboard*(OP&PB)[7]

Industrial Round-wood=Saw-logs + Pulp-wood

Sawnwood=Sawnwood coniferous+Sawnwood non-coniferous.

It addition to this it also calculates total Sawnwood and Industrial Roundwood figures. The latter is of special relevance, for it is one of the two harvested wood products considered by CBM, and therefore one of the connection points between both models.

```
for (j in c("Production","Trade","Consumption"){
w<-filter(GFTMData,Element == j)
lista<-list(grepl("Sawnwood-coniferous|Sawnwood-non-coniferous-all",w$Item),
grepl("Plw|PB|FB",w$Item),
grepl("Sawlogs-non-coniferous|Pulpwood-non-coniferous",w$Item),
grepl("Sawlogs-coniferous|Pulpwood-coniferous",w$Item),
grepl("Newsprint|P&W|OP&PB|HH&Sanitary",
w$Item))
y<-lapply(lista,"*", w[AC])
k<-lapply(y,function(x) replace(x,is.na(x),0))
z<-lapply(k,function(x) colSums(x))
m<-matrix(unlist(z),nrow = 5,ncol=length(AC),byrow=TRUE)
namesm<-data.frame(Item=SFWP[1:5],Element=c(rep(j,5)),
Unit=c(rep("m3",4),"tonnes"))
m<-setNames(data.frame(m),AC)
GFTMData<-rbind(GFTMData,cbind(m,namesm))}
```

[6]Note that, as it was explained in 3.2.3, GFTM does not include Veneers sheets. Its contribution to the Wood-based panels is added latter on.

[7]It includes packing paper.

The resulting data.frame, `GFTM3_Y2015S0` here, contains both, the original GFTM items plus the ones that have just been calculated.

prepareGFTM3

As mentioned in section 3.2.4, GFTM do not produce exports and imports figures, but only trade figures. In order to calculate the imports and exports elements for the items covered, the present function uses `ExpImpCoeff`, the data.frame created for this purpose by the FAOSTAT package function, *FAO-Coeff2* (see section 3.2.4). In order to make this calculation, the function goes through two previous steps:

1. The data.frame resulting from *prepareGFTM2* is divided into three data.frames according to the *Element* category (production, trade and consumption). A list is created which objects are these data.frames.

```
GFTMDataList<-list(subset(GFTMData,Element==ME[1]),
subset(GFTMData,Element==ME[2]),subset(GFTMData,Element==ME[3]))
```

Item	Element	Unit	Austria	Belarus	...
Sawnwood C	*Prod.*	*m3*	*9,558,528*	*2,561,784*	...
Sawnwood C	*Trade*	*m3*	...	...	...

⇓

Item	Element	Unit	Austria	Belarus	...
Sawnwood C	*Prod.*	*m3*	*9,558,528*	*2,561,784*	...
Pulpwood C	*Prod.*	*tonnes*	...	...	...

Item	Element	Unit	Austria	Belarus	...
Sawnwood C	*Trade*	*m3*	*2,725,737*	*341,249*	...
Pulpwood C	*Trade*	*tonnes*	...	...	...

Item	Element	Unit	Austria	Belarus	...
Sawnwood C	*Cons.*	*m3*	*6,832,792*	*2,220,534*	...
Pulpwood C	*Cons.*	*tonnes*	...	...	...

Table 4.2: *Original data.frame and data by element value*

2. The three objects of the list are merged into a single data.frame, each of which rows contains the production, consumption and trade figures for a particular item in all the countries considered.

```
GFTMDataSep<-merge(GFTMDataList[[1]],merge(GFTMDataList[[2]],
 GFTMDataList[[3]],by=c("Item","Unit")),by=c("Item","Unit"))
```

Item	Elem.	Unit	Austria	-	Elem.x	Austria.x	-	Elem.y	Austria.y	-
SW C	*Prod.*	*m3*	*9,558,528*	-	*Trade*	*2,725,737*	-	*Cons.*	*6,832,792*	-
PW C	*Prod.*	*tonnes*	-	-	*Trade*	-	-	*Cons.*	-	-

Table 4.3: *Elements data.frames merged by item*

The modifications implemented in the two previous steps make it easy for the function to calculate exports and imports, for, as it can be seen in the table above, all the values of a particular item can be found now in the same row. The code lines below create two empty data.frames to accommodate the new elements, calculates the export and import values and finally put all the data.frames (the three original data.frames plus the two new ones) together in the same one.

```
GFTMDataExp<-GFTMDataImp<-data.frame(row.names = 1)

for(j in AC){
C<-GFTMDataSep[c("Item",paste(j,c("",".x",".y"),sep=""),"Unit")]
colnames(C)<-c("Item",ME,"Unit")
C<-merge(C,dplyr::filter(ExpImpCoeff,Area==j)) %>%
dplyr::mutate(Exp1=Production*Coeff1,
Imp1=Exp1-Trade,
Imp2=Consumption*Coeff2,
Exp2=Trade+Imp2,
Exp=(Crit1*Exp1)+(!Crit1)*Exp2,
Imp=(Crit1*Imp1)+(!Crit1)*Imp2)
CExp<-setNames(C[c("Item","Unit","Exp")],nm=c("Item","Unit",j))
CExp["Element"]<-"Export-Quantity"
CImp<-setNames(C[c("Item","Unit","Imp")],nm=c("Item","Unit",j))
CImp["Element"]<-"Import-Quantity"
GFTMDataExp<-merge(GFTMDataExp,CExp)
GFTMDataImp<-merge(GFTMDataImp,CImp)
}

GFTMDataList<-append(list(GFTMDataExp,GFTMDataImp),GFTMDataList)

GFTMData<-do.call(rbind,GFTMDataList)
```

prepareGFTM5

As it was explained in 3.2.3, GFTM Wood-based panels (Wbp) figures do not include Veneers sheets (VS). This last *prepare* function adds the *VS* contribution to *Wbp* according to equation 3.1 and equation 3.2 and using for that purpose the `VenWbpCoeff` data.frame obtained through the FAOSTAT package *FAOCoeff1* function.

```
VenWbpCoeff<-reshape::cast(VenWbpCoeff,Item+Element+Unit Area)

VenWbpCoeff<-VenWbpCoeff[intersect(colnames(GFTMData),
colnames(VenWbpCoeff))]

GFTMWbp<-merge(VenWbpCoeff,GFTMData,by=c("Item","Element","Unit"))

GFTMWbp[AC]<-GFTMWbp[paste(AC,".y",sep="")]*(1+GFTMWbp[paste(AC,".x",sep="")])

GFTMData[GFTMData$Item=="Wood-based-panels",]<-
GFTMWbp[c("Item","Unit","Element",AC)]
```

4.2.3 *joinGFTM*

The purpose of this last function is joining all the data.frames resulting of applying the above described functions on multiple GFTM raw files for the same scenario but for different periods (e.g. *Y2015S0.mat, Y2020S0.mat, Y2025S0.mat etc*). The inputs of this function are the scenario common to all the data.frames to be joined (e.g. `scenario`<-"S0") and the initial years of the data contained in them (e.g. `years`<-c(2015,2020,2025)). Using this information, the function looks for the corresponding data.frames in **...\GFTM\Results**, and join them. It also modifies the data.frame structure, so that, now, each row contains the data for a particular item-element-area set, over the years covered by the data.frames used, which are now displayed as columns (see table in next page).

```
for (y in years){
GFTMFile<-grepl("GFTM5",list.files())*
grepl(paste(scenario),list.files())*
grepl(paste(y),list.files())
File<-list.files()[which(GFTMFile %in% 1)]
output<-readRDS(File)
output<-reshape::melt(output,measure.vars = AC)
output<-data.frame(output,rep(output["value"],4))
colnames(output)<-c("Item","Unit","Element","Area",c(y:(y+4)))
```

```
AllData[[paste(y)]]<-output}

GFTMDataAll<-purrr::Reduce(merge,AllData)
```

Like in the previous cases it is recommended to save the resulting data.frame under a name that makes reference to the data it contains, here a combination of the initial years of the periods considered and the scenario were used, thus, following the example used throughout the chapter, the final data.frame output would be named `GFTMDataAll_2015_2020_2025_S0_.rds`.

Item	**Unit**	**Element**	**Area**	**2015**	**2016**	**...**	**2024**
Sawnwood C	*m3*	*Prod.*	*Austria*	*9,558,528*	*9,558,528*	...	*9,021,438*
Pulpwood C	*tonnes*	*Trade*	*Belarus*	...	...	...	...

Table 4.4: *e.g.*`GFTMDataAll_2015_2020_2025_S0_.rds`

Chapter 5

JoinData

The goal of this last package of functions is bringing FAO and GFTM data together to calculate all the indicators related with the carbon stock in the harvested wood products (HWP). For this purpose it uses FAOSTAT and GFTM packages outputs (see subsection 3.2.2 and 4.2.2), and follows step by step, the IPCC methodology [8], described in section 1.5. Like in chapter 4, GFTM outputs from 2015 until 2019 under scenario S0 are taken as an example to illustrate the operations performed by the functions and the names used to designate the intermediate outputs. The outputs files are named after the order in which the function that produces them is run, so for instance, *readJD* output is `JList`, *HWPCStock1* output is `JList1`, *HWPCStock2* output is `JList2` and so on.

5.1 The Functions

Due to its different role, it does not directly work with the outputs of a model, the structure of the presente package is a bit different from that of the previous ones. It still has the usal *read* function, but lacks any *prepare* function, and has 6 *analysis* functions instead, each of which calculates new magnitudes to be added to the original data. Each function′ output is used as input in the next.

5.1.1 *readJD*

Unlike the *read* functions in the previous packages, this one does not take a model′s output as input, but FAOSTAT and GFTM final outputs, which are respectively, acording to the names used here, the `FAOData1` and `GFTMAll_-2015_2020_S0_` data.frames.

In addition to this it also requires the user to specify the period covered by the data being provided. In general this period goes from 1961 (the first year for which there is FAO data) until the last year covered by the GFTM outputs being analysided, in this case 2024. The function reads the *.rds* files given as an input and merge them according to the *Item, Area, Element* and *Unit* columns or categories. Note that, although there may be data available until the present year, here only the FAO data from 1961 until 2014 is selected. This is done to avoid the overlapping between the FAO and the GFTM data.

```
JData<-merge(FAO,GFTM,by=c("Item","Area","Element","Unit"))
```

Item	**Area**	**Elem.**	**Unit**	**1961**	**...**	**2024**
IRWc	*Austria*	*Prod.*	*m3*	*9,441,000*	...	...
IRWnc	*Croatia*	*Cons.*	*m3*	*NA*	...	...

Table 5.1: `JData`

Then, the *Item,Element* and *Unit* columns are melted with each other to form an new column called *HWP*, the data contained int `JData` is splited in multiple data.frames according to the value of *Area*, and these are in turn stored as objects in a new created list called `JList`.

```
JData<-within(JData,HWP<-paste(Item,Element,Unit,sep="_"))

JData<-JData[c("HWP","Area",years)]

JList<-split(JData,JData["Area"],drop=TRUE)
```

Finally, each of the data.frames in `JList` is transposed so that the *item-element-unit* sets become the new column headers and the years the row names (see table below).

```
CN<-JList[[1]][,"HWP"]

JList<-lapply(JList,function(x) as.data.frame(t(x[,-c(1,2)])))

JList<-lapply(JList,"colnames<-", CN)
```

	IRWc_Production_m3	**IRWc_Export-Quantity_m3**	**...**
1961	*9144000*	*NA*	**...**
1962	...	...	..

Table 5.2: `JList` object

5.1.2 *HWPCStock1*

The first of the six *analysis* functions takes `JList` as input and calculates for each of the objects contained in it the domestic fraction for *IRWc, IRWnc* and *Wood-pulp* according to eq. 1.1 in section 1.5. The new created magnitudes *FIRWc,FIRWnc* and *Fpulp* are added as columns to each of the data.frames.

```
CN<-colnames(JList[[1]])

DP<-c("IRWc","IRWnc","pulp")

P=grepl("Production",CN)
E=grepl("Export-Quantity",CN)
C=grepl("Consumption",CN)

for (i in DP){A=grepl(i,CN)
JList<-lapply(JList,function(x) x %>%
dplyr::mutate(fr = (x[,CN[A&P]]-x[,CN[A&E]])/x[,CN[A&C]]) %>%
dplyr::mutate(fr = replace(fr,fr<=0,0)) %>%
dplyr::rename(!!paste("F",i,sep = ""):=fr))}
```

5.1.3 *HWPCStock2*

This second *analysis* takes *HWPCStock1* list output (here `JList1`) as input, and calculates *IRWc* and *IRWnc* weight factors (see eq. 1.2) for each data.frame in the list. Then, and using the weight factors in combination with the domestic fractions estimated by *HWPCStock1* it determines the value of the averaged *IRW* and *Pulp* domestic fractions for each list object according to eq. 1.3. Like in *HWPCStock1*, here the new magnitudes (*IRWc_ WF,IRWnc_ WF,FIRWAv* and *FpulpAv*) are also added to the corresponding data.frames as columns.

```
JList<-lapply(JList,function(x) x %>%
mutate(SumIRWCon = rowSums(.[,CN[grepl("^IRW.*Consumption",CN)]]))
%>%
mutate_at(vars(CN[grepl("^IRW.*Consumption",CN)]),funs(WF=./SumIRWCon))
%>%
rename_at(vars(colnames(.)[grepl("WF",colnames(.))]), paste(DP[1:2],
"WF",sep = "_")) %>%
mutate_at(vars(IRWc_WF,IRWnc_WF), replace(.,.>1,1)) %>%
mutate(FIRWAv=(FIRWc*IRWc_WF)+(FIRWnc*IRWnc_WF),FpulpAv=Fpulp*FIRWAv))
```

5.1.4 *HWPCStock3*

Using the simple and averaged domestic products fractions for *IRW* and *Pulp*, calculated by *HWPCStock1*, it applies eq 1.6- 1.9 to calculate the domestic products fractions of all the semi-finished wood products (SFWP). The table below show which fraction has been used on which product.

SFWP	Domestic Product Fraction
Wood-based panels(Wbp)	*FIRWAv*
Paper-and-paperboard(PPB	*FpulpAv*
Sawnwood-coniferous(SWc)	*FIRWc*
Sawnwood-non-coniferous(SWnc)	*FIRWnc*

Table 5.3: *SFWP-Domestic Product Fraction correspondence*

```
CN1<-colnames(JList[[1]])

JList<-lapply(JList, function(x) x %>%
dplyr::mutate(SWc_DP = FIRWc*x[,"Sawnwood-coniferous_Production_m3"],
SWnc_DP = FIRWnc*x[,"Sawnwood-non-coniferous-all_Production_m3"],
SW_DP = SWc_DP+SWnc_DP,
Wbp_DP = FIRWAv*x[,grepl("based.*Production",CN1)],
PPB_DP = FpulpAv*x[,"Paper-and-paperboard_Production_tonnes"]))
```

5.1.5 *HWPCStock4*

As it was explained in section 1.5, in order to estimate the evolution of the carbon stock in the harvested wood products sector in the years between 1900[1] and 1961 (first FAO data), it is assumed that the carbon inflow in each SFWP is constant and equal to the average of the carbon inflow during the first 5 years of the historical period (1961-2014). The present function converts all the magnitudes calculated in the previous functions to tonnes of carbon (referred as *Inflows* or *Ifw* from now on) and calculates the averaged inflows for all the SFWP (see eq. 1.10- 1.12).

```
CF<-readRDS("ConversionFactors.rds")

HL<-readRDS("HalfLives.rds")

DP<-c("IRWc","IRWnc","pulp")
```

[1] 1900 is the year proposed by the IPCC to begin the carbon accounting in the HWP sector [8, 7]

```
SF<-c("SW","SWc","SWnc","Wbp","PPB")

SFDP<-paste(SF,"DP",sep = "_")

SFIfw<-paste(SF,"Ifw",sep = "_")

JList<-lapply(JList,function(x) {x[,SFIfw] = t(t(x[SFDP])*CF[(CF$DP
%in% SFDP),"CCF"]);x})

AvInflow<-lapply(JList,function(x){x=colMeans(na.omit(x[SFIfw])[c(1:5),]);
x=as.data.frame(rbind(x,x/HL[,"k"][match(attr(x,"names"),HL[,"Ifw"])]))})
```

To perform these tasks the function requires some additional information, namely SFWD carbon conversion factors and half-lives. This information had been extracted beforehand from [8] and stored as data.frames in the **...\General Parameters** directory as `CF` and `HL` respectively. Note that, as it was explained in section 1.5, in absence of country-specific data, the IPCC Tier 2 method for the calculation of the harvested wood products carbon stock [8] is being followed. If at any moment country-specific data is made available and thus it is decided to switch to the Tier 3 method, these two sets of parameters will have to be modified.

TABLE 2.8.1
DEFAULT CONVERSION FACTORS FOR THE DEFAULT HWP CATEGORIES AND THEIR SUBCATEGORIES

HWP categories	Density (oven dry mass over air dry volume) [Mg / m³]	Carbon fraction	C conversion factor (per air dry volume) [Mg C / m³]	Source
Sawn wood (*aggregate*)	0.458	0.5	0.229	1
Coniferous sawnwood	0.45	0.5	0.225	2
Non-coniferous sawnwood	0.56	0.5	0.28	2
Wood-based panels (*aggregate*)	0.595	0.454	0.269	3
Hardboard (HDF)	0.788	0.425	0.335	4
Insulating board (Other board, LDF)	0.159	0.474	0.075	5
Fibreboard compressed	0.739	0.426	0.315	6
Medium-density fibreboard (MDF)	0.691	0.427	0.295	4
Particle board	0.596	0.451	0.269	4
Plywood	0.542	0.493	0.267	7
Veneer sheets	0.505	0.5	0.253	8
	(oven dry mass over air dry mass) [Mg / Mg]		(per air dry mass) [Mg C / Mg]	
Paper and paperboard (*aggregate*)	0.9		0.386	9

[1] Calculated from the weighted average of coniferous and non-coniferous sawnwood production volumes (FAOSTAT average of the years 2006-2010) of the countries as listed in Appendix of the Annex of Decision 2/CMP.7
[2] IPCC 2003, Appendix 3a.1
[3] Calculated from the weighted average of included subcategories of the production volumes (FAOSTAT average of the years 2006-2010) of the countries as listed in Appendix of the Annex of Decision 2/CMP.7
[4] Rüter and Diederichs (2012)
[5] Derived from Environmental product declarations EPD-GTX-2011111-E, EPD-KRO-2009212-E and EPD-GTX-2011211-E provided by IBU e.V. (http://bau-umwelt.de/hp550/Insulating-materials.htm)
[6] Calculated from 50% of HDF and 50% of MDF
[7] Derived from Wilson and Sakimoto (2005) and basic density for non-coniferous species listed in the table above
[8] Calculated from 50% sawnwood (Coniferous) and 50% of sawnwood (Non-Coniferous)
[9] Calculated from the weighted average of included subcategories of the production volumes (FAOSTAT average of the years 2006-2010) of the countries as listed in Appendix of the Annex of Decision 2/CMP.7, including information derived from Fengel and Wegener (1984), Paulapuro (2000), Gronfors (2010) and industry information.

Figure 5.1: SFWD Carbon Conversion Factors[8]

TABLE 2.8.2 TIER 2 DEFAULT HALF-LIVES[141] OF HWP CATEGORIES	
HWP categories[142]	**Default half-lives (years)**
Paper	2
Wood panels	25
Sawn wood	35

Figure 5.2: SFWD Half Lives[8]

In this case, besides the usual list, `JList4` in this case, the function also produce a second output which is a list, with as many objects as the former, each of which contains the SFWP averaged inflows for one country.

SW_Ifw	SWc_Ifw	SWnc_Ifw	Wbp_Ifw	PPB_Ifw
1181843	*1134876*	*32757,99*	*349027*	*638179,6*
59676372	*57304808*	*1654092,59*	*12588487*	*1841397*

Table 5.4: *Austria SFWP* `AvInflows`[2]

5.1.6 *HWPCStock5*

This function takes the outputs of `HWPCStock5`, `JList4` and `AvInflow` as inputs and performs the following operations:

- To each data.frame of the list it adds as many empty rows as years between 1900 and the first historical data, normally 1961 or 1990, depending on the country and fills them with the information in `AvInflow`.

```
JList<-lapply(names(JList),function(x) {
JList[[x]]<-rbind((JList[[x]][c(1:61),]*NA),JList[[x]]) ;
JList[[x]][c(1:(sum(is.na(JList[[x]][SFIfw]))/length(SFIfw))),SFIfw]
<-AvInflow[[x]][1,] ;
JList[[x]]}
```

[2]The second row corresponds to $\frac{AvInflow}{ln(2)/HL}$

- To each data.frame of the list it adds 9 new columns, three for each SFWP, containing the parameters e^{-k}, $\frac{1-e^{-k}}{k}$ and k[3], which will be used latter on to calculate the SFWP C-Stock each year.

  ```
  JList<-lapply(JList,function(x) x %>%mutate_at(
  vars(SFIfw),funs(k=rep(1,length(.))*HL[HL$Ifw==deparse(substitute(.)),"k"]))
  %>%
  mutate_at(vars(contains("k")),funs(ek=exp(-.)))
  %>%
  mutate_at(vars(contains("ek")),funs(Ik = (1-.)/(-log(.,exp(1))))))
  ```

5.1.7 *HWPCStock6*

This last function takes as input a list, here `JList5`, which objects already contain all the magnitudes required to estimate the C-Stock, C-Stock variations and CO_2 emitted by the Harvested Wood Product sector according to eq. 1.14 from 1900, through the years covered by FAO until the last year considered by the GFTM outputs used, which would be 2024 following the example presented at the beginning of this chapter.

First it defines the names of the columns that are going to be used:

```
SF<-c("SW","SWc","SWnc","Wbp","PPB")

SFDP<-paste(SF,"DP",sep = "_")

SFIfw<-paste(SF,"Ifw",sep = "_")

SFIfw_k_ek<-paste(SFIfw,"k","ek",sep="_")

SFIfw_k_ek_Ik<-paste(SFIfw,"k","ek","Ik",sep="_")
```

Then, the names of the columns that will contain the magnitudes to be calculated are filled with 0s:

```
SFC<-paste(SF,"C",sep="_")

SFCV<-paste(SF,"CV",sep="_")

SFEm<-paste(SF,"Em",sep="_")
```

[3] $k = \frac{ln(2)}{HL}$

Finally it applies the equations in section 1.5:

```
JList<-lapply(JList,function(x) {x[SFEm]<-x[SFCV]<-x[SFC]<-0 ;
for (j in (2:dim(x)[1])){
x[j,SFC]<-x[(j-1),SFC]*x[j-1,SFIfw_k_ek]+x[j-1,SFIfw]*x[j-1,SFIfw_-
k_ek_Ik]
x[j-1,SFCV]<-x[j,SFC]-x[j-1,SFC]
x[j-1,SFEm]<-(x[(j-1),SFC]+x[j-1,SFIfw]-x[j,SFC])*(-44/12)};
x["TotalC"]=rowSums(x[SFC]);
x["TotalCV"]=rowSums(x[SFCV]);
x["TotalCO2"]=x["TotalCV"]*(-44/12);
x["TotalEm"]=rowSums(x[SFEm]*(-44/12));
x})
```

Note that, apart from calculating the magnitudes for each of the SFWD it also sums them up to obtain the figures for the whole HWP sector per year, such as the total HWP C-Stock (*TotalC*), the total HWP C-Stock variations (*TotalCV*), the total CO_2 absorbed by the HWP sector (*TotalCO2*) and the total CO_2 emitted by the HWP sector due to the decay of the products (*TotalEm*).

5.2 Representation

The magnitudes calculated by the package functions can be represented in all imaginable ways, using different R packages such as `ggplot` or additional software like QGIS[4]. It is recommended to create standard `graph` functions that take for instance, the magnitudes, countries and periods of time as inputs and, to filter the data to be represented. What follows is just an example of a possible `graph` function:

```
graphsHWP<-function(Data,countries,objects,period){
setwd(paste(path,"JoinData/Results",sep="/"))
Data<-readRDS(Data)
Data<-Data[countries]
Data<-lapply(Data,function(x) subset(x,(Year %in% period)))
Data<-lapply(Data,"[",objects)
Data<-cbind(period,do.call(cbind,Data))
CNGraph<-rep(countries,length(objects))
CNGraph<-expand.grid(objects,countries)
CNGraph<-c(paste(CNGraph[,1],CNGraph[,2],sep=""))
colnames(Data)<-c("Year",CNGraph)
```

[4]https://www.qgis.org/en/site/

```
Data<-melt(Data,id.vars="Year")
colnames(Data)[colnames(Data)=="variable"]<="Countries"
ggplot2::ggplot(Data,aes(x=Year,y=value,color=variable))
   +geom_point(shape=20,size=2)+
   labs(x="Years",y="C-Stock[tonnes C]")+
   ggtitle("Semi-finished HWP C-Stock")+
   theme(plot.title = element_text(size=17,face="bold",hjust = 0.5))+
   scale_color_discrete(name = "Countries")+
   scale_x_continuous(breaks = seq(from=period[1],
   to=period[length(period)],by=10))+
   scale_y_continuous(breaks = seq(from=0, to=3.7e+08,by=40000000))+
   geom_vline(aes(xintercept =1961),color="black",linetype="dashed")+
   geom_vline(aes(xintercept =2015),color="black",linetype="dashed")+
   annotate("text",x=1930,y=3.7e+08,label="Constant Inflow \n
   (IPCC)",size=4,fontface=2)+
   annotate("text",x=1990,y=3.7e+08,label="FAO Data",size=4,fontface=2)+
   annotate("text",x=2025,y=3.7e+08,label="Model \n
   (GFTM)",size=4,fontface=2)
}
```

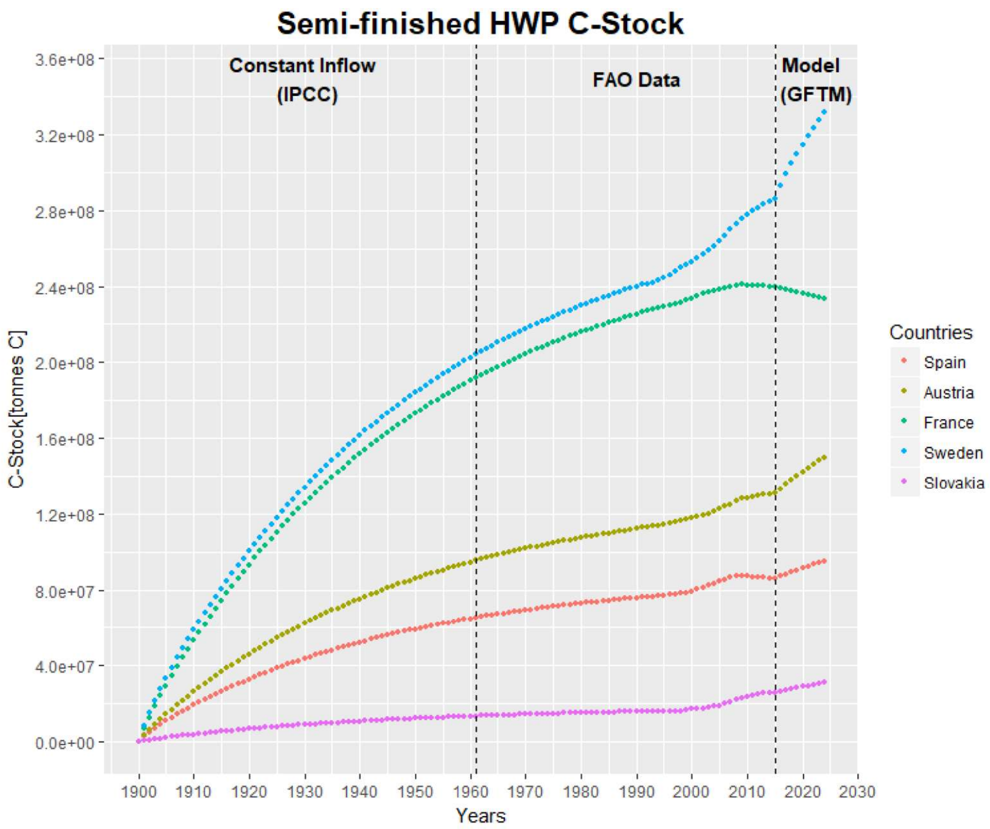

Figure 5.3: SFWD HWP C-Stock (1900-2030)

Bibliography

[1] Agostini A., Manfredi S., Torres de Matos C., Marelli L., and Giuntoli J. Biomass Study 2015/16. Pursuant to the Mandate on the provision of data and analysis on a long-term basis on biomass supply and demand by the JRC. Technical report, 2018.

[2] Camia A., Robert N., Jonsson R., Pilli R., García-Condado S., López-Lozano R., van der Velde M., Ronzon T., Gurría P., M'Barek R., Tamosiunas S., Fiore G., Araujo R., Hoepffner N., Marelli L., and Giuntoli J. Biomass production, supply, uses and flows in the European Union. first results from an integrated assessment. Technical report, EC-JRC, 2018.

[3] J. Giuntoli, .A Agostini, L. Marelli, S. Manfredi, and C. Torres de Matros. Biomass Study 2015/16: Environmental Assessment of impacts. deliverable D3.2. Technical report, JRC, 2016.

[4] Kull S.J.and Rampley G.J., Morken S., Metsaranta J., Neilson E.T., and Kurz W.A. *Operational-Scale Carbon Budget Model of the Canadian Forest Sector (CBM-CFS3) Version 1.2: USERS GUIDE*, 2016.

[5] F. Jonsson, R.and Rinaldi and J. San Miguel Ayanz. The Global Forest Trade Model-GFTM. Technical report, EC-JRC, 2015.

[6] W.A. Kurz, C.C. Dymond, T.M. White, G. Stinson, C.H. Shaw, G.J. Rampley, C. Smyth, B.N. Simpson, E.T. Neilson, J.A. Trofymow, J. Metsaranta, and title=CBM-CFS3: a model of carbon-dynamics in forestry and land-use change implementing IPCC standards. journal=Ecological Modelling volume=220(4) pager=480-504 year=2009 Apps, M.J.".

[7] J. Penman, M. Gytarsky, T. Hiraishi, T. Krug, D. Kruger, R. Pipatti, L. Buendia, K. Miwa, T. Ngara, K. Tanabe, and F. (Eds.) Wagner. Good Practice Guidance for Land Use, Land-Use Change and Forestry. Technical report, IPCC, 2003.

[8] Hiraishi T., Krug T., Tanabe K., Srivastava N., Baasansuren J., and Fukuda M. andTroxler TG. 2013 Revised Supplementary Methods and Good Practice Guidance Arising from the Kyoto Protocol. Technical report, IPCC, 2014.

Appendix A

Indicators Literature Review

Sources
Recebio Task 4: Resource efficiency implications for the scenarios: http://ec.europa.eu/environment/enveco/resource_efficiency/index.htm#bioenergy
Forest Bioeconomy, a new scope for sustainability indicators: https://www.efi.int/sites/default/files/files/publication-bank/2018/efi_fstp_4_2016.pdf
State of Europe's Forests 2015: https://www.foresteurope.org/docs/fullsoef2015.pdf
OECD Environmental Indicators: https://www.oecd.org/site/envind/indicators/
OECD Database: https://stats.oecd.org
FAO Global Forests Resource Assessment 2015: http://www.fao.org/3/a-i4793e.pdf
European Timber Regulation: http://ec.europa.eu/environment/forests/timber_regulation.htm
ITTO Criteria and indicators for the SM of tropical forests: http://www.itto.int/direct/topics/topics_pdf_download/topics_id=4872&no=1&disp=inline
European Environment Agency Indicators Data Base: https://www.eea.europa.eu/data-and-maps/indicators
The European Forest Sector Outlook Study II (UNECE/FAO,2011): https://www.unece.org/fileadmin/DAM/timber/publications/sp-28.pdf

Appendix B

Additional Packages

As it was explained in section 1.6, besides the packages included in the downloadable version of R, a number of additional packages are required for the functions to run. What follows is a list of these packages, and which of their functions have been used to develop the R-tool:

Package	**Functions**	**Where it is used**
RODBC	*odbcConnectAccess*	*readCBM (2.2.1)*
dplyr	*mutate,filter,group_by*	*readFAOSTAT(3.2.1), prepareGFTM(4.2.2)*
reshape	*cast,melt*	*readFAOSTAT (3.2.1), joinGFTM(4.2.3)*
ggplot2	*ggplot*	*graphsHWP(5.2)*
R.matlab	*readMat*	*readGFTM(4.2.1)*
tidyr	*complete*	*prepareFAOSTAT(3.2.2)*
purrr	*Reduce*	*joinGFTM(4.2.3)*

Table B.1: *Additional packages*