**DOCUMENTATION MITERRA-PORTUGAL**

# 1. Main\_function

This code includes all the main functions that are going to be used throughout all the estimates, e.g. GNB, gaseous N losses, runoff. These are very basic, and whose main goal is simply to search for the main activity data folder, select specifically the data of any estimate, read it, store the files to then disaggregate by year. Moreover, these also include one sub-function to subset all the data, which is then used in the main function to merge all the data.

## check\_folder

This function checks whether the last two indexes of a specific folder\_path have the required “\\”. If not, these are added.

## set\_maindata\_folder

This function specifies the folder with all the activity data. By default, it is set to “C:\\Users\\cms\\OneDrive\\Major project\\MITERRA-PORTUGAL\\Activity data\\”, but this can be changed using the parameter *path*.

## select\_maindata\_pattern

Selects any subfolder within the main activity data, according to any estimate. These can range from the GNB (“GNB”), N-NH3, N-NOx, Runoff.

This is the main starting point for any calculation, as it compiles the main path where all the required spreadsheets are.

## print\_loop\_files

Prints all the “.csv” files within a specified folder\_path.

## store\_folder\_files

Similar to function 1.4 print\_loop\_files, but instead of simply printing the files, it stores in a vector their fullpath.

This is the starting point to read these files and further disaggregated them per year.

## disaggregate\_year

This is one type of a main function, i.e. one function that is heavily reliant on a sub-function (see 1.6.1).

An input parameter is the function 1.5 store\_folder\_files, which are then looped. These are then compared to the specified year (e.g. 1999) and stored in a db vector.

### year\_prefix

It receives as input one year. By default, it should contain 4 digits but as this is obvious, I did not include an exception/error/solution. Basically it substrings the year to its last two digits. 1999 🡪 99

## store\_file\_names

Very similar to function 1.5 store\_folder\_files, but instead it only stores the filenames within a specified path. Additionally, it receives the same input as store\_folder\_files, which is basically the activity data path specified by the user, e.g. GNB, N-NH3.

## read\_disagg\_files

Receives 3 different inputs:

1. The output of store\_folder\_file, which is the fullpath of the spreadsheets within a specified sub activity data folder, disaggregated by year using function 1.6 disagg\_year.
2. The fullpath of the sub activity data folder.
3. The specified full-year (e.g. 1999).

The function then returns the name of the files ready to be estimated. This is particularly relevant for e.g., the GNB.

## merge\_data

Two different datasets are merged according to a common ID. These can be then subset based on an “input\_subset”, which can be either in column names or column numbers. This is disaggregated according to the sub-function 1.7.1 which subsets the merged dataset.

### subset\_data

Subsets a dataset based on specified conditions (see above).

# 2. GNB functions

The GNB module has two main functions: compute GNB and compute NUE, although the latter is optional. There are 4 sub-functions that are needed to calculate the GNB and NUE.

## 2.1 gnb\_compute

As the name indicates, this calculates the GNB. As input, it receives a list disaggregated per year, which is the output of the main\_functions: (i) store\_file\_gnb, (ii) disaggregate\_year and read\_disagg\_files. This output is a list with two indexes, each one corresponding to a file of a year. Moreover, gnb\_compute also checks for irrigation.

All inputs and outputs from a specified list index are summed and whose difference is the GNB. All these are aggregated into a main dataframe with tot\_inputs, tot\_outputs and the gnb.

### 2.1.1 gnb\_element\_id

This sub-functions receives as input the same as the main function, i.e. read\_disagg\_files. Its goal is to check the name of each list index. If a name contains the substring “Input”, it returns a vector with the index of the input in the first position and the opposite to the “Outputs”.

### 2.1.2 gnb\_organize

This function is used to create the main dataframe used in gnb\_compute. As parameters, it receives one dataframe to aggregate the specified data, particularly the tot\_sum of either input and output, receiving as the second parameter the remaining one. Further, the col\_name must be specified as to modify the dataframe.

### 2.1.3 gnb\_sum\_input

This is analogous to the sub-fucntion gnb\_sum\_output. Its inputs are the output of read\_disagg\_files and irrigation. It checks whether the latter is TRUE or FALSE and calculates the total sum of inputs accordingly.

Basically, it calls the sub-function 2.1.1 to compute the index of the input and output file from the list with both. It then specifies the input (or the output) by calling this index to the list and further converts it into a dataframe using the same column names.

If irrigation is to be accounted, it then sums the column 4 to 10. Otherwise, it only sums 4 to 9.

## 2.2 nue\_compute

The output of gnb\_compute is a input-parameter to nue\_compute, and further calculates the ratio of outputs and inputs as a percentage.

# 3. GASEOUS functions

GASEOUS\_functions developes wide functions to apply for all gaseous emissions. Several pieces of code work, though they are still in development as they can be still further improved.

## print\_gaseous\_sources

General print function that displays all the major pathway losses for each gaseous emission and respective sources.

## gaseous\_select\_source

This function allows to specify which sub-folder of gaseous N losses the users wants to call (i.e. application, housing, storage). This calls the global function select\_maindata\_pattern and then selects the source path accordingly.

#gaseous\_select\_source(‘NH3’, source=’Application’)

## gaseous\_main\_fullpath

This is a sub-function called in gaseous\_store\_filepath. This creates a list whose length is in accordance to the type of gaseous emissions. User has to specify whether this related to NH3 emissions or not, as it is the only gaseous source with a length of 3. Within each list\_index, the fullpath of the said pathway is stored and can be accessed as cond[[i]].